



Real-Time Event Based Predictive Modelling for
Industrial Control and Monitoring

A thesis submitted for the degree of
Doctor of Philosophy

by

FUTRA ZAMSYAH BIN MD FADZIL

Department of Electronic and Computer Engineering
Brunel University London
February 2020

Abstract

The purpose of this research work is to presents a novel event-based predictive modelling technique, namely, Event Modeller Data Analytic (EMDA), applicable for a large-scale real-time complex system. Borrowed from the Event Tracker and Event Clustering method, EMDA continuously estimates and builds a correlation map between system input (triggered data) and output (event data) parameters while predicting system failure based on machine performance metrics. With the aid of advanced machine learning models, EMDA can potentially predict linear and non-linear problems, thereby improving rapid decision-making for system engineering problems. For proof of concept, EMDA was used to analyse the mystery of an escalating harmonic failure trend in one of the Malaysian power plants. Analysis of the harmonic parameter in their Continuous Ship Unloader machines indicates that the power quality is stable as per IEEE standards; however, in practice, repetitive harmonic failures occur and the reasons remain unknown. The hypothesis associated with this research is that: *"A fault in a power system distribution could be influenced not only by internal events but also by external events such as environment and climate change"*. In addition to the conventional method used by the power system engineers, we challenge the body of knowledge in the subject area by exploring and potentially incorporating external variables that may influence the state of the system. Software-In-the-Loop application was developed using the National Instrument LabVIEW. The purpose of this deployment was to test and validate the concept and to demonstrate whether the correlation analysis was in synchronisation with the latent knowledge (KPI) translated into the system. EMDA was also used as a tool to visualise the occurrences of the system parameters and its KPIs with a predictive analytical approach to data. This research conducts extensive experimental work on both industrial and synthetic data to evaluate the proposed method. The results of the study reveal that in addition to the known parameters that may affect harmonic filter performance, there is one new parameter that shows a reasonable correlation with performance. The previously unknown parameter is the humidity of the operational environment having a significant impact on the occurrence of harmonic failure. This proved the hypothesis set in the underpinning research endeavour presented in this thesis. By controlling the humidity of the operational environment and deploying EMDA, the state transition and trends were accurately predicted. The results of this research can help power generation plants to devise adaptive strategies to optimise the performance of plants.

Acknowledgements

I would like to acknowledge the contributions of the following individuals and groups for their help and support throughout the development of this thesis:

1. My supervisors Dr Alireza Mousavi and Professor John Stonham who have provided me guidance and support.
2. The JEV Power Plant Superintendent, Mr Abdul Razak Jamin and the Coal Handling Head of Department, Mr Asmadee Mat Jusoh who have to provide me with space and opportunity to render my service with the company.
3. Mr Khairul Azad Azmi, Mr Abdul Fatah Abd Halim and all the team member of Electrical & Instrumentation at JEV Power Plant, who provide me with the experiment data and support me during my research.
4. My beloved wife and family who have inspired me in throughout my life.
5. Members of the Systems Engineering Research Group (SERG) at the College of Engineering, Design and Physical Sciences, Brunel University who provides an inspiring atmosphere of passion and joy in the workplace.
6. The administrative staff at the College of Engineering, Design and Physical Sciences, Brunel University, without whose aid and support this research could not have been completed.
7. Finally, I would like to express my sincere thanks to all of my friends who have encouraged and support me during my PhD Journey.

This thesis is dedicated with endless love and gratitude to my wife, Farrah Nadia and my son, Futra Iskandar, who has always been there for me. Not to forget the two individuals who are important to my life, parents Md Fadzil and Dr Sharifah Meriam who have always been inspired, supportive, encouraging despite the distance, patient and worked hard to see that I succeed in life. Also a special appreciation to my mother in law, Rohana Kamarudin who have been supporting me during my PhD journey.

Contents

1	Introduction	1
		1
1.1	Research Background	1
1.2	Research Motivation	5
1.3	Research Context	5
1.4	Research Aim, Objectives & Questions	6
1.5	Research Hypothesis	7
1.6	Research Design	8
1.7	Main contribution of the research work	9
1.8	Structure of the thesis	9
2	Literature Review	11
2.1	Overview	11
2.2	Power System	11
2.2.1	Power System Stability	12
2.2.2	Power System Fault	13
2.3	Power Quality Disturbance	14
2.3.1	Transient	14
2.3.1.1	Oscillatory Transient	14
2.3.1.2	Impulsive Transient	14
2.3.2	Short Duration Voltage Variation	15
2.3.2.1	Interruption	15
2.3.2.2	Voltage Sags	15

2.3.2.3	Voltage Swell	16
2.3.3	Long Duration Voltage Variation	16
2.3.3.1	Over-voltage	16
2.3.3.2	Under-voltage	16
2.3.3.3	Sustained Interruption	17
2.3.4	Voltage Unbalance	17
2.3.5	Waveform Distortion	17
2.3.5.1	DC Offset	17
2.3.5.2	Harmonics	18
2.3.5.3	Inter-harmonics	18
2.3.5.4	Notching	18
2.3.5.5	Noise	19
2.3.6	Voltage Fluctuation	19
2.3.7	Power Frequency Variation	19
2.4	Troubleshooting Power System Quality	20
2.4.1	Power Quality Instrument	20
2.4.2	Power Quality Trouble-shooter Device	21
2.4.3	Power Quality Troubleshooting Methods	21
2.5	Signal Processing Techniques	22
2.5.1	Feature Extraction	22
2.5.1.1	Fourier Transform	23
2.5.1.2	Wavelet Transform	24
2.5.1.3	Stockwell Transform	26
2.5.1.4	Hilbert Huang Transform	27
2.5.1.5	Gabor Transform	28
2.5.1.6	Other Miscellaneous Transform	28
2.5.2	Feature Selection	29
2.5.2.1	Genetic Algorithm	29
2.5.2.2	Particle Swarm Optimisation	30
2.5.2.3	Ant Colony Optimisation	31
2.6	Sensitivity Analysis Techniques	31

2.6.1	Feature Extraction	32
2.6.1.1	Input Variable Selection	33
2.6.1.2	Sensitivity Analysis	33
2.6.1.3	Event Tracking	34
2.6.2	Feature Selection	34
2.6.2.1	Event Clustering	34
2.7	Classifiers Techniques	35
2.7.1	Supervised Learning	36
2.7.2	Unsupervised Learning	38
2.7.3	Hybrid Learning	39
2.7.4	Deep Learning	40
2.7.5	Support Vector Machine	42
2.7.6	Fuzzy Expert System	43
2.7.7	Other Classifiers	43
2.8	Summary Analysis	43
2.9	Summary	45
3	Event Modeller Technique	46
3.1	Overview	46
3.2	Predictive Analytic	47
3.2.1	Predictive models	48
3.2.2	Big Data Analytics	49
3.2.3	Predictive Analytics Workflow	51
3.3	Event Modeller Data Analytics Technique	53
3.4	Stage 1: Access and Explore Data	54
3.5	Stage 2: Event Modeller Technique	55
3.5.1	Event Modeller Basic Parameter	57
3.5.2	Exclusive-NOR Functionality	58
3.5.3	Event Modeller Basic Assumption	58
3.5.3.1	Assumption 1: Delays	58
3.5.3.2	Assumption 2: Thresholds	58
3.5.3.3	Assumption 3: Homogeneity of the data series	59

3.5.3.4	Trigger-Event detection	59
3.5.3.5	Sample scan size	59
3.5.3.6	Average sensitivity analysis weight	59
3.5.3.7	Event Modeller Limit	59
3.5.3.8	False negative test	60
3.5.4	Rank Order Clustering	60
3.5.4.1	Rank Order Clustering Algorithm	60
3.5.5	Event Modeller Algorithm	62
3.5.6	Key Performance Indicator	65
3.5.6.1	Time-based Key Performance Indicator	65
3.5.6.2	Energy and Emission Based Key Performance Indicator	69
3.5.6.3	Event Modeller Key Performance Indicator Transfer Function	70
3.6	Stage 3: Develop Predictive Models	71
3.6.1	Decision Tree	73
3.6.2	Multi-Layer Perceptron Neural Network	75
3.6.3	Naïve Bayes	76
3.7	Summary	77
4	Industrial Case Study	78
4.1	Overview	78
4.2	Power Generation Industry	80
4.2.1	Coal Handling Plant	80
4.2.2	Super Critical Boiler	81
4.2.3	Steam Turbine	81
4.2.4	Power Generator	82
4.3	Continuous Ship Unloader Machine Breakdown	82
4.3.1	Continuous Ship Unloader Overview	83
4.3.1.1	Bucket Elevator	84
4.3.1.2	Rotary Feeding Table	85
4.3.1.3	Boom Conveyor	85
4.3.1.4	Main Slewing & Hoisting	86
4.3.1.5	Central Chute & Portal Conveyor	87

4.3.1.6	Travel Gear	87
4.3.2	Continuous Ship Unloader Electrical System	88
4.3.2.1	Incoming Transformer	88
4.3.2.2	System Protection	88
4.3.2.3	Motors	89
4.3.2.4	REGEN & Control Drives System	89
4.3.2.5	DC-Bus Distribution System	90
4.3.3	System Architecture for Continuous Ship Unloader	91
4.3.3.1	Hardware	91
4.3.3.2	Software	92
4.3.3.3	Network Communication	93
4.4	Power Quality Assessment Initiatives	94
4.4.1	Power Quality Measurement Setup	95
4.4.2	Power Quality Assessment Results	99
4.4.3	Power Quality Assessment Summary	106
4.5	Industrial Data Collection Setup	107
4.5.1	Limitation on the Existing Infrastructure	107
4.5.2	Installation of New Relative Humidity & Temperature Room Sensor	109
4.5.3	Upgrading the Analogue Output for the Multi-function Meter	109
4.5.4	Network Scheduling	109
4.5.5	Data Logger Configuration	110
4.6	Summary	112
5	Event Modeller on Software in the Loop	113
5.1	Overview	113
5.2	System Development Implementation	114
5.2.1	Software in the Loop Requirement	114
5.2.2	Software in the Loop Design	115
5.2.3	System Integration	117
5.2.4	System Testing	118
5.3	Methodology	119
5.3.1	Experimental Methodology	120

5.3.2	Dataset Arrangement	120
5.3.2.1	Event Dataset	121
5.3.2.2	Trigger Dataset	121
5.3.2.3	Disturbance Dataset	122
5.4	Results and Discussion	123
5.4.1	Pre-Disturbance Stage	123
5.4.2	k-Disturbance Stage	125
5.4.3	Post-Disturbance Stage	127
5.4.4	Overall Results	127
5.5	Summary	130
6	Event Modeller on Industrial Data	131
6.1	Overview	131
6.2	System Development Implementation	132
6.2.1	New Software-in-the-Loop Requirements	133
6.2.2	Software in the Loop Design	133
6.2.3	System Integration	135
6.2.4	System Testing	136
6.3	Methodology	136
6.3.1	Experimental Methodology	136
6.3.2	Data Collection	137
6.3.3	Dataset Arrangement	138
6.3.4	Event Dataset	140
6.3.5	Assumption on Total Harmonic Distortion Data	142
6.3.6	Triggered Dataset	143
6.3.7	Data Discretisation	144
6.3.8	Key Performance Indicator Parameter	144
6.3.9	Predictive Model Dataset	146
6.3.9.1	Pattern 1 - Standby, No Abnormality and Good Weather	148
6.3.9.2	Pattern 2 - Standby, No Abnormality and Bad Weather	150
6.3.9.3	Pattern 3 - Standby, Downtime and Good Weather	150
6.3.9.4	Pattern 4 - Standby, Downtime and Bad Weather	153

6.3.9.5	Pattern 5 - Operating, No Abnormality and Good Weather	153
6.3.9.6	Pattern 6 - Operating, No Abnormality and Bad Weather	156
6.3.9.7	Pattern 7 - Operating, Downtime and Good Weather	156
6.3.9.8	Pattern 8 - Operating, Downtime and Bad Weather	160
6.4	Results and Discussion	160
6.4.1	Experiment 1: Machine Failure Analysis with Key Performance Indicators	161
6.4.1.1	Experiment Strategy	161
6.4.1.2	Relationship between the observed data against Event Modeller	163
6.4.1.3	Relationship between the observed data against Key Performance Indicators	169
6.4.1.4	Discussion	175
6.4.2	Experiment 2: Machine Failure Analysis with Predictive Model	178
6.4.2.1	Experiment Strategy	178
6.4.2.2	Classification Results	179
6.4.2.3	Computational Effort Results	181
6.5	Summary	183
7	Conclusion	184
7.1	Overview	184
7.2	Summary of Dissertation	184
7.3	Contribution	187
7.3.1	Contribution as a System Engineering Tools	187
7.3.2	Contribution in Power Quality Disturbance	188
7.4	Limitation	189
7.4.1	Total Harmonic Data	189
7.4.2	Data Benchmarking	189
7.4.3	Data Quality	190
7.4.4	Reports	190
7.5	Future Works	191
7.5.1	Wide Variety of Industrial Dataset	191
7.5.2	Real-time Cloud Services	191
7.5.3	Other Machine Learning Algorithm	192

Appendix A	222
Appendix B	232

List of Figures

1.1	Proposed Framework for Real-Time System Modelling	8
2.1	Different Signal Processing Methods in Feature Extraction	23
2.2	Different Optimisation Techniques in Feature Selection.	29
2.3	Different Input Variable Selection and Sensitivity Analysis in Feature Extraction.	32
2.4	Different Artificial Intelligence Techniques in Classifiers.	36
3.1	Predictive Analytic Workflow	52
3.2	Event Modeller Data Analytics Technique Overview	53
3.3	Step 1: Access and Explore Data	55
3.4	Event Modeller Techniques	56
3.5	Machine Part Incidence Matrix	61
3.6	Incident Matrix Row Ranking	61
3.7	Incident Matrix Column Ranking	62
3.8	Combining all Incidents Matrix	63
3.9	Reliability Bath-tub Curve Image: courtesy of [1]	66
3.10	KPI Categorisation	67
3.11	Automatic Translation of KPI Transfer Function	71
3.12	Classification using Machine Learning	73
4.1	Power Generation Supply Chain, Image: courtesy of Australian Energy Market Operator [2]	79
4.2	Coal Fired Power Plant Overview, Image: courtesy of [3]	80
4.3	Continuous Ship Unloader Overview	83
4.4	CSU Machine Single Line Diagram	89

4.5	CSU Machine Control System	92
4.6	Power Quality Analyser Loggers with Clamp CT's and Flexible Voltage Probes	95
4.7	Point of Common Coupling for the 1250 KVA Main Incoming Transformer . . .	97
4.8	Point of Measurement for the VVVF Drives Input (Master REGEN)	98
4.9	Total Harmonic Distortion (THDV & THDI) at Main Incoming (1250kVA Transformer Secondary)	100
4.10	Harmonic Spectrum at Main Incoming (1250kVA Transformer Secondary) . . .	101
4.11	Total Harmonic Distortion (THDV & THDI) at CSU Machine VVVF Drives Input	104
4.12	Harmonic Spectrum at CSU Machine VVVF Drives Input	105
4.13	Panel Temperature and Humidity Sensor Installation	108
4.14	E-House Temperature and Humidity Sensor Installation	108
4.15	Network Scheduling for ControlNet	111
5.1	Overview of System Development	115
5.2	Snippets of Event Modeller VI - Block Diagram	116
5.3	Overview of System Integration Architecture	118
5.4	Snippets of System Integration VI - Block Diagram	119
5.5	Overview of Experimental Methodology	120
5.6	Snippets of Event Modeller VI - Block Diagram (Pre-Disturbance Stage) . . .	124
5.7	Snippets of Event Modeller VI - Block Diagram (k-Disturbance Stage)	126
5.8	Snippets of Event Modeller VI - Block Diagram (Post-Disturbance Stage) . . .	128
6.1	Overview of System Development	132
6.2	Snippets of Event Modeller VI- Block Diagram (Enhancement)	134
6.3	Control System Architecture	134
6.4	Snippets of System Integration VI - Block Diagram (Enhancement)	135
6.5	Event Modeller Data Analytics Technique	137
6.6	Overall Data Tabulation for CSU Machine Activity	138
6.7	Monthly Data Tabulation for CSU Machine Activity	139
6.8	CSU Machine Key Performance Indicator	145
6.9	CSU Machine Energy Efficiency KPI	147

6.10	CSU Machine - Pattern 1 (Standby, No Abnormality and Good Weather)	149
6.11	CSU Machine - Pattern 2 (Standby, No Abnormality and Bad Weather)	151
6.12	CSU Machine - Pattern 3 (Standby, Downtime and Good Weather)	152
6.13	CSU Machine - Pattern 4 (Standby, Downtime and Bad Weather)	154
6.14	CSU Machine - Pattern 5 (Operating, No Abnormality and Good Weather)	155
6.15	CSU Machine - Pattern 6 (Operating, No Abnormality and Bad Weather)	157
6.16	CSU Machine - Pattern 7 (Operating, Downtime and Good Weather)	158
6.17	CSU Machine - Pattern 8 (Operating, Downtime and Bad Weather)	159
6.18	Tabulation of CSU Machine Failure vs REGEN Failure	162
6.19	Experiment Strategy for KPIs Analysis	163
6.20	Rank Order Clustering - Cluster 1	165
6.21	Rank Order Clustering - Cluster 2	166
6.22	Rank Order Clustering - Cluster 3	167
6.23	Rank Order Clustering - Cluster 4	168
6.24	Electrical Variables vs KPIs	171
6.25	Motor Variables vs KPIs	172
6.26	Environment Variables vs KPIs	173
6.27	Experiment Strategy for Predictive Modelling Analysis	179
A.1	Snippets of CSU Machine KPI	222
A.2	Snippets of CSU Machine KPI	223
A.3	Snippets of Energy Efficiency KPI	224
A.4	CSU Machine Event Modeller - Main Overview	225
A.5	CSU Machine Event Modeller - Main Setting	226
A.6	CSU Machine Event Modeller - SLD and Energy Efficiency	227
A.7	CSU Machine Event Modeller - Event Data Trending	228
A.8	CSU Machine Event Modeller - Event Clustering	229
A.9	CSU Machine Event Modeller - Key Performance Indicator	230
A.10	CSU Machine Event Modeller - Machine Positioning & Status	231

List of Tables

2.1	Summary Analysis of Different Techniques	44
3.1	Summary of Basic Description for Event-Modeller	57
3.2	Exclusive-NOR Functionality	58
3.3	Summary of Predictive Model Steps	72
3.4	Criteria in Selecting the Classifiers	74
4.1	Motor List for CSU Machine	90
4.2	Control System Configuration Hardware List for CSU machine	93
4.3	Load Description and Time Interval for Power Quality Measurement at CSU Machine	96
4.4	Data Tabulation for CUL-1 Main Incomer	102
4.5	Data Tabulation for CSU VVVF Master Drives Input	102
4.6	Data Tabulation for CSU VVVF Slaves 1-REGEN Drives Input	103
4.7	Data Tabulation for CSU VVVF Slaves 2-REGEN Drives Input	103
5.1	Event Data Simulation Parameters	121
5.2	Triggered Data Simulation Parameters	122
5.3	Event Data Disturbance Parameters	122
5.4	k-Disturbance Signal Setting	123
5.5	CSU Real-time Data Simulation Results Based on Disturbance	129
6.1	Event Data Benchmark Configuration & Average Value for CSU Machine	140
6.2	Event Data Upper Limit & Lower Limit for CSU Machine	141
6.3	Triggered Data Parameters	143
6.4	Data Discretisation Strategy	144

6.5	Key Performance Indicator Parameters	146
6.6	Description of Machine Status & Formula for KPI	148
6.7	Predictive Model Pattern Logic	150
6.8	Summary of Sensitivity Index	169
6.9	Summary of Electrical Variables	170
6.10	Summary of KPIs	170
6.11	Summary of Motor Variables	174
6.12	Summary of Environment Variables	175
6.13	Experiment Details for CSU Machine with 15% Holdouts	180
6.14	Classification Result for CSU Machine with 15% Holdouts	180
6.15	Experiment Details for CSU Machine with 30% Holdouts	181
6.16	Classification Result for CSU Machine with 30% Holdouts	181
6.17	Time Span Computational Effort Results	182
6.18	CPU Utilisation Computational Effort Results	182
7.1	Research Question	185
B.2	Test Plan for Event Modeller Development	233
B.3	Summary of System Operator Daily Log Sheet for CSU Machine	234
B.4	Conversion factor for Coal Combustion	235
B.5	96 Input Variable List	235
B.6	24 Output Variable List	236

Acronyms

AC Alternating Current

ACB Air Circuit Breaker

AIS Artificial Immune Systems

ANN Artificial Neural Network

ANOVA Analysis Of Variance

API Application Program Interface

CBM Condition Based Monitoring

CNN Convolution Neural Networks

CT Current Transformer

CSU Continuous Ship Unloader

DAQ Data Acquisition

DAE Deep Auto Encoder

DBN Deep Belief Networks

DC Direct Current

DCS Distributed Control System

DES Discrete Event System

DSTP Data Socket Protocol

EventiC Event Clustering

ED Event Data

EDIM Event-Driven Incidence Matrix

ECP Electrostatic Precipitator

EMDA Event Modeller Data Analytic

ET Event Threshold

FDILA Flexible Data Input Layer Architecture

FLS Fuzzy Logic Systems

FTP File Transfer Protocol

HIL Hardware in the Loop

HTTP Hypertext Transfer Protocol

IF Interconnected Facilities

IP Internet Protocol

ISO International Standards Organization

IVS Input Variable Selection

JEV Jimah Energy Ventures

KPI Key Performance Index

LD Ladder Diagram

LDVV Long Duration Voltage Variation

LFO Light Fuel Oil

MCC Motor Control Centre

OSI Open Systems Interconnection

PCC Point of Common Coupling

PHIL Power Hardware-in-Loop

PLC Programmable Logic Control

PQ Power Quality

REGEN Regenerative Drives

RNN Recurrent Neural Network

RT-HIL Real-Time Hardware in the Loop

R3M Real-time Model Matching Mechanism

SA Sensitivity Analysis

SCADA Supervisory Control and Data Acquisition

SDVV Short Duration Voltage Variation

SERG System Engineering Research Group

SIL Software in the Loop

SoS System of System

SR Stacker Reclaimer

SS Search Slot

SVM Support Vector Machines

TCP Transmission Control Protocol

TD Trigger Data

THD Total Harmonic Distortion

TNB Tenaga Nasional Berhad

TT Trigger Threshold

UAT Unit Auxiliary Transformer

UPS Uninterrupted Power Supply

VVVF Variable Voltage Variable Frequency

Chapter 1

Introduction

1.1 Research Background

Power utilities are recognised as the backbone of industry and modern society, supporting several other critical infrastructures such as transportation, communications, military, water and etc. [4, 5]. They play an important role in distributing and generating reliable energy in order to ensure global business stability and sustainability. The power sector is viewed as the key sector that significantly contributes to the growth of country's economy [6]. In fact, to maintain economic development, continuous supply of power is required at an affordable price. Power utilities around the world would have to guarantee their infrastructure has the capacity and capability to provide uninterrupted electricity while maintaining its integrity in sustaining the environment.

With increasingly growing fossil-fueled power plants around the world, there have been many environmental concerns, notably CO_2 emissions and greenhouse gas (GHG_4). Statistic shows that fossil fuel emission from electricity generation accounts for 42% of the total global CO_2 emission [7]. In fact, the world depends on fossil fuels to meet its daily energy needs. Although most governments increasingly embrace renewable energy sources such as solar, wind and geothermal, fossil fuels still dominate the world's source of energy for several reasons. The main reason is that fossil fuels produce significant amounts of energy per unit mass, which is one of the cheapest fuel sources on the planet. Also, its Calorific Value (CV) is one of the best compared to any other petrol. Fossil fuels are available globally and may not be depleting soon. Furthermore, the carbon and hydrogen molecules in fossil fuels make them highly stable,

rendering easier storage, transfer and transport. Most significantly, fossil fuels are capable of producing huge amount of energy at a single location, hence they are widely found in utility sites.

Even though renewable energy research has been an active field since the last decade, this environmentally friendly technology requires a huge amount of capital expenditure and a significant amount of transition time. The situation differs from one region to another due to the different rates of market growth. Developed countries, such as the United Kingdom, have been steadily generating and using renewable energy since 2000. This is in line with the EU Renewable Energy Directive, which requires the EU as a whole to produce 20% of its energy from renewable sources by 2020 [8]. In 2017, the UK generated 29.3% of its electricity from renewable sources. The gas share is the largest at 40.4%, followed by nuclear at 20.8%. The remainder are 6.7% from coal and 2.9% from other fuels [9].

In contrast, developing countries, such as Malaysia, which has a steady increase in energy demand, electrical generation are mainly dominated by gas (43.5%) and coal (42.5%), accounting for 86% of the country's energy mix. The remaining share is hydro (13%) followed by other fuels at 1% while renewable energy is only at 0.2% [10]. Although Malaysia has large renewable energy resources from solar, hydro and biomass, the Energy Commission has decided to replace the 6,256 MW obsolete power plant with more than 10,000 MW over the next decade, with 92.30% still being dominated by gas and coal. This indicates that the country is still committed to fossil-fuel option.

Even though Malaysia has been promoting renewable energy in its Five-Fuel Diversification Strategy since 2001, research has shown that the percentage of renewable energy implementation is still low due to high installation costs but low generation capacity [11]. Much renewable energy operates on a small scale; therefore, it can only be considered as a supplement to the installed capacity. Malaysia receives up to four to six hours of sunlight on daily average, which promises a solar energy resource but requires a large area to generate a reasonable amount of electricity. Meanwhile, hydro needs a large area, which would lead to the displacement of inhabitants from the flooded areas. However, as with other countries around the world, the Malaysian government had to reconsider its plan to build and operate a nuclear power plant since the Fukushima nuclear disaster in 2011. Other renewable energy options, such as biomass, solid waste and biogas plants, show some growth, but currently, only six plants

operate with a total capacity of 24.7 MW. This suggests that renewable energy is still not being significantly developed [12].

With new generation projects expected to contribute to overall system reliability and efficiency improvement, existing plants will be maintained to ensure reserved capacity is kept at 30%, to cover for any sudden system grid shortfalls due to system instability, plant outages or fuel shortages. The system operator must ensure that the plants are available, maintain a strict quality standard and provide services at a competitive cost, while at the same time reducing the risk for energy shortages in the system grid, which may affect national safety.

Three main criteria that could sustain the performance of these plants are system stability, system security and system reliability. The concept of power system stability and security relates to the ability of the power generators to maintain synchronisation with the system grid without any disruption, while the system can instantaneously revert to its steady-state operation when disruptions are introduced into the system. Reliability, on the other hand, is determined by the performance of the entire supply chain, which includes the supply of fuel and raw materials, the availability of machines and the distribution of power.

Knowing that machine availability is one of the key aspects, any failure that occurs during the power generation process will cause the system grid to be interrupted. Monitoring of Power Quality (PQ) has been the focus of research and development for many years. The main focus is to protect the equipment and minimise the losses while increasing the levels of operational safety. However, systems are becoming more complex as they are being further developed for improvement. Huge numbers of control drives and other non-linear loads have been installed to satisfy the demands of modern lifestyles. It has led to instability of the power system, creating high noise levels in the system grids and decaying the electrical distribution system. Even worse, climate change has an impact on the environment that the system operates in [13]. In some cases, particularly for hot countries, the electrical distribution system requires air conditioning to protect the electrical equipment in the substation. In normal circumstances, if the air conditioning fails, it may surpass the permitted set point and may not protect the equipment. When the global temperature rises, the potential rate of failure also increases. Research has shown that the long-term average global temperature is increasing, which has a significant negative impact on the performance of the machines [13].

PQ has become a significant issue for a modern power industry in order to protect elec-

trical and electronic equipment by identifying the sources of disturbances and providing an appropriate solution to mitigate them [14, 15]. The key reasons for PQ disturbances are the widespread use of switching circuits, capacitors, inverters and converters, particularly in renewable energy sources, and the implementation of Industrial Internet of a Thing (IIoT) and Smart Home, leading to unbalanced lighting control loads and excessive use of computer and data processing equipment in daily life [16]. PQ disruptions cause various problems for power suppliers and customers, such as malfunctions, instability, short lifespan and loss of electrical equipment. Harmonic disturbances, on the other hand, contributed to power losses in transmission lines, power transformers and rotating machines.

In order to maintain a good PQ in the power system, it is important to consider the sources and causes of PQ disturbances in order to appropriately mitigate them. Traditional PQ monitoring methods used by utilities are typically based on visual inspections, which are indeed laborious and time-consuming. A highly automated hardware-and software-based monitoring system is therefore required to provide sufficient information on the overall system, understand the primary sources of disturbances, seek better solutions and predict future disturbances. Various Artificial Intelligence (AI) techniques offer a better solution to automatically classify PQ disturbances. AI is a cognitive science with rich research activities in the fields of Image Processing, Natural Language Processing (NLP), Robotics, Machine Learning (ML), Deep Learning (DL), Expert Systems, Fuzzy Sets and Heuristic Search [17]. Among these methods, Expert System, ML and DL are commonly used to solve power system problems [18]. ML uses approaches from neural networks, statistics, operations research and physics to find hidden insights into data without explicitly programmed where to look or what to infer. DL uses massive neural networks with several layers of processing units, taking advantage of advances in computing power and improved training techniques to learn complex patterns in large volumes of data [19]. There are four types of ML/DL techniques; Supervised Learning, Unsupervised Learning, Semi-Supervised Learning and Reinforcement Learning with Supervised Learning is the most common technique for detecting PQ issues.

Previously, conventional methods use manual configurations and visual inspection to monitor the quality of the power supply in the system. This method was too difficult to interpret and time-consuming. Later, an automated classification system, which uses the signal processing technique was developed and currently advancing with various Machine Learning and Deep

Learning approach. Various combinations of features extraction and classifiers have proven to locate PQ disturbances accurately while using synthetic data. However, this is not the case with real industry data. Real industry data are complex and much more complicated when there are external or environmental factors that could potentially influence the state of the system. This creates a gap in existing techniques, which motivates authors to investigate the relevance of the environmental parameter to the issue of PQ disturbances. In addition, with the current emergence of big industrial data, system engineer must ensure that the system is robust and capable of performing ML/DL techniques in real-time applications. Disturbance data exists in the order of microseconds, which greatly enhances the record data [20]. In consequence, it will burden learning classifiers to run unimportant parameters that result in high computation. The authors, therefore, suggest a method that could minimise the dimensional by selecting the most relevant data to be trained automatically, while incorporating a new unknown parameter that was previously considered irrelevant in the system state.

1.2 Research Motivation

The motivation for this research project is to contribute to new knowledge of the system engineering learning method to the discovery of unknown or external parameters that could affect the entire system of complex behaviour. As mentioned by [21], "As technology continues to improve, systems are becoming more complex as they are developed. Scientists can access more data and are looking for better ways to interpret and ultimately solve complex problems". This is consistent with this research learning method in finding a better way to solve complex systems by developing Software-In-The-Loop (SIL) framework, to measure and test various parameters that cause the system to be ineffective.

1.3 Research Context

The machine learning techniques has been tremendously improved in line with the technology within the system engineering computational learning theory and study of pattern recognition. The algorithm that has been developed is able to recognise large data acquisition and complex system and able to make decision based on learning experience of the complex data. There are many researchers who have been working on machine learning techniques and their

specific application. [22] has applied the machine learning techniques in a real-time inferencing mechanism which uses the knowledge provided by the learning algorithm, to decide which alarms to block when there is a string of alarm sequence obtained via the SCADA from a power distribution system. [23] has been working on Machine-Learning Approaches to Power-System Security Assessment in extracting and synthesising relevant information and reformulate it in a suitable way for decision making. [24] has introduced the pattern learning approach for Dynamic Security Analysis (DSA) classification using neural network to obtain system security status without running time-consuming time domain simulation. Reflecting to the ineffectiveness system issues in specific case study at JEV power station, this machine learning techniques are motivated to identify the missing parameter on system of the system (SoS) which could validate and verify the abnormality in the system. The existing controllers are able to access the entire input/output modules throughout the system hence are able to initiate the event tracking purposes. The event tracking data will be communicating with the Software in the Loop system in Real-Time for learning purposes.

1.4 Research Aim, Objectives & Questions

The aim of this research is to propose a machine learning framework architecture that automates and integrates real-time data acquisition and raw data analysis processes, accessing internal and environment parameter that caused harmonic phenomena in power generation system. To address the aim of this research, the following objectives are established:

1. To capture and to allow for visualisation of system architecture randomness in real-time using available data from the server.
2. To calculate the correlation index (weight) for each parameter using the basic logic of the number of coincidence in a time-span; interpreting changes in the values of input-output (I/O) data as I/O events.
3. To reduce the dimensionality of the I/O events by detecting if the I/O events coincide, and cluster them as related events according to its rank.
4. To investigate the status of the system state by calculating the Key Performance Indicator defined in manufacturing system literature; linking with the I/O events in Software-

In-the-Loop (SIL) environment framework.

5. To develop a data analytics platform, to train, test and validate the data using Machine Learning algorithm; and compare the performance.
6. To test and validate the proposed system using synthetic data; to prove the capability and the advantages of the proposed framework and later examine the outcome with a real-industry data.

Two research questions are posed to stipulate the study:

- What are the internal and environmental events within the power system distribution that affect harmonic filter performance? Internal events, is referring to dynamics and interactions within plant machinery, and environmental event, is for fluctuations in temperature, humidity and pollution in the plant vicinity.
- Is the bathtub theory of harmonic filters performance a reliable mechanism for predicting its performance at the these periods: (1) an “early failure” (burn-in) period, where the hazard function decreases over time, or (2) a “random failure” (useful life) period, where the hazard function is constant over time, or (3) a “wear-out” period, where the hazard function increases over time.

1.5 Research Hypothesis

The hypothesis that underpins this research is: *"A fault in a power system distribution will be influenced not only by internal events but also by external events such as environment and climate change"*. Power system engineer utilised signal processing approach, to locate the fault solely based on the time domain and the frequency domain of the electrical signal. We challenged this by looking at all possible knowledge that is available within the defined system, including the conventional approach. The key objective of this thesis is to provide an optimal practical and technical platform to demonstrate the interrelationship between the power system parameter and other knowledge available in the system state. Hence, a CSU machine with frequent harmonic failure was used as a case study to test the hypothesis.

1.6 Research Design

To address the stated research questions, the study reported in this thesis employed a real-time system modelling design. This is a two-phase design where the researcher investigates the root causes of harmonic phenomena using synthetic data before moving to real industrial data. The design aims at generalising the findings of the first phase through the use of system modelling. This approach enables the researcher to test the applicability of the proposed method before applying it to big data analytics. Figure 1.1 presents the proposed framework for real-time system modelling.

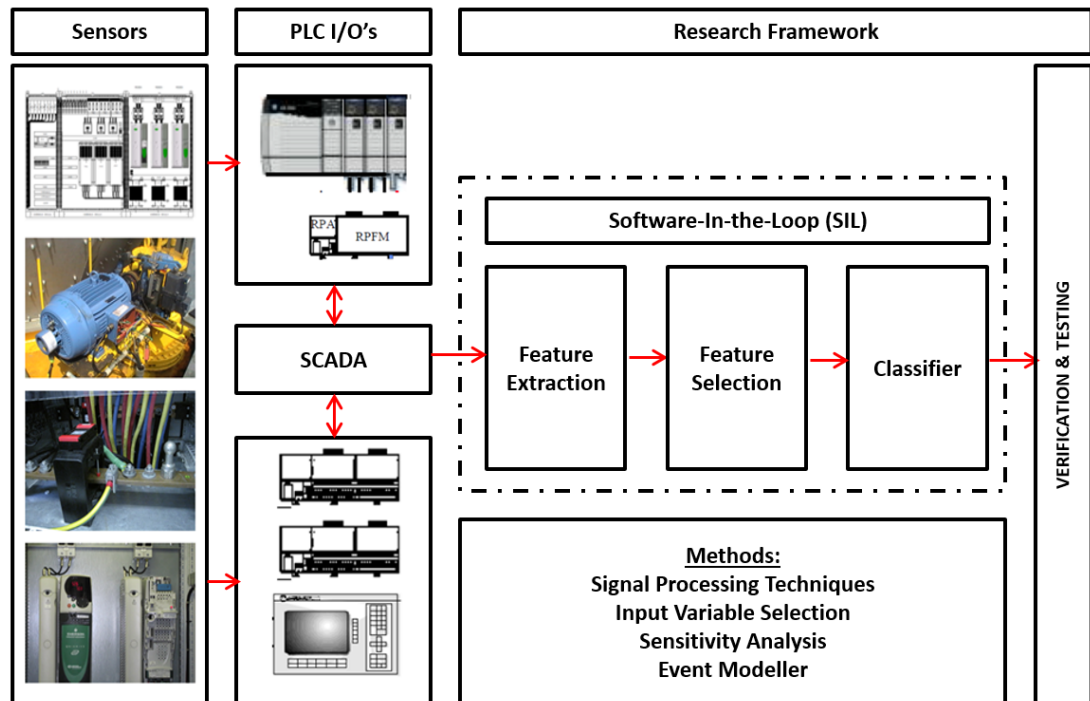


Figure 1.1: Proposed Framework for Real-Time System Modelling

1.7 Main contribution of the research work

The potential practical applications of this research are as follows:

1. Formulate new knowledge information in the diagnosis of PQ disturbances, thus incorporating the influence of the environmental parameter such as temperature, humidity and wind speed in the intelligent system loop.
2. Develop a novel and scalable architecture that links the embedded system with the machine learning algorithm and predicts the system state pattern for real-time decision-making tool.
3. The discovery of known and unknown parameters and their relationship within the state of the system, and group them in order of its importance in real-time, hence providing system status based on the Key Performance Indicators.
4. A robust system engineering tools and knowledge engineering technique to meet the challenges of the dynamic, autonomous, adaptive and self-organising embedded systems, and, seamless/secure interaction of the embedded system/cyber-physical systems with their environment.

1.8 Structure of the thesis

This thesis is organised in the following Chapters:

- **Chapter Two**

This chapter provides a literature review on the main concepts of PQ disturbances, feature extraction, feature selection and classifier that are used within this research. The subject areas: Power system and PQ disturbances issues are first described. Then the signal processing techniques and sensitivity analysis techniques, which are suitably used in analysing PQ disturbance data, are discussed. Next, the relevant classifiers techniques for machine learning are demonstrated. Finally, a summary on analysis using different techniques is presented.

- **Chapter Three**

This chapter proposed the Event Modeler Data Analytic (EMDA) technique to investi-

gate industrial control and monitoring issues. First, predictive analytics big data issues are explained. The proposed EMDA technique, which includes three stages: (1) access to plant data, (2) Event Modeller Technique, and (3) the development of a predictive model, is further elaborated. The selected predictive model is finally explained.

- **Chapter Four**

This chapter explains the industrial case study that is experiencing an escalating trend in harmonic failure. The chapter provides a general overview of the power plant, followed by a detailed explanation of the target machine. Next, the results of the PQ assessment will be discussed. Finally, the data collection setup and the counter-measure of the limitation in the existing plant are presented.

- **Chapter Five**

This chapter extends the industrial case study in Chapter 4 by applying the event modeller technique using synthetic data. The purpose of the SIL is to assess the suitability and applicability of event modeller technique in power system environment. The chapter discusses the development of a real-time application using the National Instrument LabVIEW, followed by the introduction of a k-disturbance signal. The chapter concludes with the result of the synthetic data.

- **Chapter Six**

This chapter focuses on the implementation of real industry data using EMDA technique. The chapter enhanced the SIL framework, which was built to meet the needs of industrial data applications in Chapter 4. It further describes the arrangement of the data set for the CSU machine. Eight (8) predictive model patterns of the CSU machine state were established and labelled. The chapter concludes with the experiment's results and elaborates the main findings of this study.

- **Chapter Seven**

This chapter concludes the thesis by discussing the contributions of this research to the knowledge body. In addition, the implications of the findings to the PQ disturbance environment is discussed. The chapter ends by identifying research limitations and possible research extensions.

Chapter 2

Literature Review

2.1 Overview

The previous chapter presented a detailed view of the research background and identified the research gap that the study aims to address. This chapter reviews relevant researches in the area including the problem domain and techniques used within this research. The chapter is presented in nine main sections. Section 2.2 introduced power system in general followed by PQ disturbance in Section 2.3. This section aims to provide different types of PQ disturbance. In reviewing the technique to troubleshoot PQ problems, Section 2.4 reviews the traditional technique followed by the conventional signal processing technique in Section 2.5. The existing techniques help to reduce irrelevant input on PQ disturbance using Input Variable Selection and Sensitivity Analysis are reviewed in Section 2.6 followed by the available classifiers technique in Section 2.7. Section 2.8 compares all the techniques followed by Section 2.9 which summarised the chapter and its place in the thesis.

2.2 Power System

The term power system in the utility company is the ability of the generator to provide electricity to the grid (load) at the desired quality with minimum interruptions as possible. The power system is commonly subjected to disturbances during operation. [25] mentioned that power systems are examples of complex systems in which generators and loads are dynamically interconnected. Hence, they can be seen as networked systems, where each bus is a node in

the network. This shows that the relationship between the generator system and the grid system all the way to the supply chain are critical in achieving the overall power system stability. Most of the industry in any part of the world has modernised and invested in high technology devices which are mainly controlled by a drive system to achieve higher efficiency performance. [26] agreed with this issue as they mentioned that “with increasing utilisation of high power-electronic converters, power-system modelling and stability analysis become ever more challenging”. This is because the utilisation of the high power-electronics converters creates more disturbance and noise in the system, which could lead to instability of the power system.

2.2.1 Power System Stability

For any power utilities company supplying energy to the grid system, three (3) main criteria that could sustain the performance of the energy distribution are the system reliability, system security and system stability. The system reliability is the probability of a system to provide a desired function under specific operating conditions during its lifetime. The systems are able to consistently operate at the required parameters without reaching the undesired state or condition. To attain system reliability, the system must be secured most of the time, during and post fault periods. The system security refers to the ability to withstand contingencies without interrupting the system function. It is the measurement of the system robustness to its contingency. To attain system security, the system must be stable and concurrently run in time varying attributes. This system stability is referring to the ability of the system to sustain its operation and remain stable even though a disturbance has been introduced in the system. This instability and insecurity could lead to catastrophic consequences of system disturbances such as blackouts. In the power utilities environment, the concept of power system stability and security relates to the ability of the generators to maintain synchronisation with the grid system without any interruption and the system is able to return to the point of steady-state operation if a system disturbance has been introduced. [26] define the power-system stability as traditionally analysed by the phasor theory, which describes the slow dynamics of power systems, where electromagnetic transients have negligible effect. Ideally, when dealing with power system stability issues, a modelling technique is one the leading techniques in simulating an adequate time-domain system modelling which includes the system behaviour

of the past, present and future. These models components consist of mathematical relations comprising physical behaviour of the components. The past system behaviour is based on historical data, experience data, disturbance occurrence, and the intrinsic characteristic that contributes to improvement of operating conditions. The present system behaviour refers to the real-time calculation analysis, which includes the estimation of eliminating bad data from the measurement and includes the estimation of un-defined information. Power flow analysis is one of the examples which calculate the real-time power flow distribution of the system. [27] recommend that power flow is the most accurate approach to model the steady-state behaviour of balanced three-phase electrical power systems. Two main methods are: (1) Gauss-Seidel, and (2) Variation process: which is either a Newton-Raphson (N-R) method, or Jacobian method. Finally, the future time system behaviour referring to analysing the supposed system by simulation in order to provide decision for the system development plan, system operation and emergency control strategy. It is crucial to analyse the power system output under different operating conditions to ensure stable operation of the power system. Analysis involves studies such as load flow, and both steady state and transient stability. The subject of stability thus encompasses modelling, computation of load flow in the transmission grid, stability analysis under both steady-state and disturbed conditions and appropriate controls to enhance stability.

2.2.2 Power System Fault

[28] mentioned that faults, dynamic operations, or nonlinear loads in power system often cause various types of power quality disturbances which include voltage sags, voltage swells, switching transients, impulses, notches, flickers and harmonics. These internal disturbances could lead to a power system fault that could affect the entire system. However, it is likely that there is also the possibility of a power system fault due to external parameters, which could also lead to system failure. This is in line with the research objectives of this study that is to further explore the root cause of the ineffectiveness of the harmonics filter. The external parameters may include temperature, humidity, lightning, dust particle, and etc. [29] argues that the existing techniques are capable to automatically identify and classify various types of distribution-level power quality disturbances, however, they do not provide any information on the locations of the disturbance sources. This motivates them to identify the solution in making a decision on the location of the disturbance which includes the information on

changes in Disturbance Power (DP) and Disturbance Energy (DE) as energy tends to flow toward the disturbance source. The problem of fault diagnosis, (i.e. detection, isolation and identification) of such nonlinear power networks is formulated as a comprehensive sensing or sparse signal recovery problem.

2.3 Power Quality Disturbance

One critical aspect in power quality studies is the ability to perform automatic power quality monitoring and data analysis while another principal aspect of a power quality study is the coordination between the power system behaviour and the equipment performance. As discussed in the power system fault, there are seven (7) types of power quality disturbance which will be discussed in detail on the following.

2.3.1 Transient

The term “transient” means lasted for a short duration. Power system defined transient as a rapid disturbance in the system for a short period of time which created an undesirable distortion and ranked as the most damaging of all disturbance [30]. The power transient is divided into two parts which are oscillatory transient and impulsive transient [31].

2.3.1.1 Oscillatory Transient

An oscillatory transient occurs when there is a repetitive change in the steady-state condition in a short duration of time which tends the power signal to alternately swell and shrink rapidly. This kind of disturbance occurs in an inductive and capacitive load mainly in motors and capacitor bank, which has repetitive switching behaviour [30]. In modern industries, most of the motors are operated with variable speed drives for controlling purposes and these drives will be tripped with overvoltage alarm indicating the occurrence of an oscillatory transient [31].

2.3.1.2 Impulsive Transient

Impulsive transient in the steady-state condition occurs as a spontaneous rise of high voltage or high current level in either a positive or negative direction in a short duration of time. This

kind of disturbance is always referring to mainly lightning strikes, followed by poor grounding, inductive loads switching, utility fault clearing and Electrostatic Discharge [31]. The consequences of this disturbance will be as bad as physical damage since the electromagnetic field created by the lightning will be conducted to the equipment and the structure of the machines.

2.3.2 Short Duration Voltage Variation

Short Duration Voltage Variation (SDVV) refers to a slight change or difference in the steady-state condition in a short period of time which leads to three types of disturbance which are Interruption, Voltage Sags and Voltage Swell. This kind of disturbance always occurs due to a high current drawn by large loads or poor wiring connection [31].

2.3.2.1 Interruption

Interruption in a power system refers to a voltage loss or loss of supply in the steady-state system which could happen either instantaneously (0.5 to 30 cycles), momentarily (30 cycles to 2 seconds), or temporarily (2 seconds to 2 minutes) depending on the duration [31]. This classification of duration is in line with the IEEE-1250 Standard [32], in which the value of voltage drop is below 10% of nominal voltage. This interruption is usually caused by lightning strikes, animal or tree interrupting the distribution line, destructive weather and equipment failure. This type of disturbance could lead to data loss or corrupted, which could lead to a very costly impact.

2.3.2.2 Voltage Sags

Voltage sags refers to a decrease or a reduction of voltage from the steady-state levels at a given frequency in a short duration of 0.5 cycles to 1 minute. This kind of disturbance always occurs in the system faults or large energy switching condition like large motors start-up and power transformer energising [31]. This disturbance will likely cause a significant voltage drop to the rest of the circuit. Among power quality problems, voltage sags are typically the most common.

2.3.2.3 Voltage Swell

Voltage swell is the alternate of Voltage Sag which refers to an increase of voltage from the steady-state levels at a given frequency for the short duration of 0.5 cycles to 1 minute [31]. This kind of disturbance always occurs in a high-impedance neutral connection, sudden reductions in large loads or single-phase faults on a three-phase system. The consequences of this disturbance could be observed from the data errors, the flickering of lights, degradation of electrical contacts, semiconductor damage in electronics and insulation degradation .

2.3.3 Long Duration Voltage Variation

Long Duration Voltage Variation (LDVV) is referring to a slight change or difference in the steady-state condition in a longer period of time which leads to four types of disturbance which are Over-voltage, Under-voltage, Sustained Interruption and Voltage Unbalance [31].

2.3.3.1 Over-voltage

Over-voltage is the extended of Voltage Swells in long term Voltage Swell which refers to an increase of voltage for more than a minute [31]. This kind of disturbance always occurs due to incorrect power transformer tap setting while the actual loads have been reduced accordingly [33]. Consequently, this disturbance will definitely create a high current draw and cause unnecessary tripping of downstream circuit breakers while the equipment will be stressed and overheat.

2.3.3.2 Under-voltage

Under-voltage is the extent of Voltage Sag in a long term that caused a decrease or reduction of voltage, for more than a minute duration [31]. This kind of disturbance always occurs during large load switching on, or switching a capacitor bank off, as well as overloaded circuits application [33]. The consequence of this disturbance is overheating in motors and can lead to the failure of non-linear loads. If this Under-voltage scenario is extended for a longer period, it may be due to a severe equipment fault or configuration issues which need to be rectified.

2.3.3.3 Sustained Interruption

A sustained interruption is the longer period of interruption, which refers to a voltage loss or loss of supply in the steady-state system, which could happen for more than 2 minutes. All this classification of duration is in line with the IEEE-1250 Standard [32], which the values of voltage drop are below 10% of nominal voltage [31].

2.3.4 Voltage Unbalance

A voltage unbalance fault occurs when the voltages of the three phases are not equal [31]. This kind of disturbance is always referring to the source of supply. If one of the phases has less than 5% of the rated voltage, the system is said to be unbalanced. The consequence of this disturbance is excessive heat when the supply is connected to the motor component. The motor controller tends to result in intermittent failure if there is a greater imbalance in the disturbance.

2.3.5 Waveform Distortion

A steady-state power system represents an ideal balanced voltage or current sinewave graph shape, without any distortion. However, this ideal waveform may get affected by some distortions from any of five categories which are direct current offset, harmonic distortion, inter-harmonics, notching and noise [31]. The consequences of this disturbance could cause damage or disruption to IT equipment.

2.3.5.1 DC Offset

A Direct Current (DC) offset refers to the presence of this DC Voltage in the AC system, which leads to overheating and saturation of the transformer, and probably reduces the transformer life expectancy. The problem of this DC offset is mainly due to the disturbance in the power electronics converters [31]. The consequence of this disturbance could cause the transformer not delivering full power to the load, and the subsequent distortion of the waveform could lead to further instability in the electronics of the transformer.

2.3.5.2 Harmonics

Harmonics distortion refers to the corruption in the fundamental of a sine wave at frequencies that are multiples of the fundamental [31]. The complete harmonic spectrum detailing the phase angles and magnitudes of the harmonic components clearly illustrates the harmonic distortion levels. Harmonic distortion may also be described by the Total Harmonic Distortion (THD) as an effective measure [34]. The symptom of the harmonics problem includes overheating of the transformer, the neutral conductors and other electrical distribution equipment, as well as the circuit breaker tripping and the loss of synchronisation timing. Consequences of the harmonic issues may lead to losses in the form of heat with associated loss of efficiency. It may also contribute to higher noise levels in transformers and motors while the harmonic currents and voltages may lead to ground fault circuit interrupters falsely tripping. Other loads that contribute to this problem include variable speed motor drives, lighting ballasts and a higher rated UPS system. Methods used to mitigate this problem include over-sizing of neutral conductors, installation of K-rated transformers and harmonic filters.

2.3.5.3 Inter-harmonics

Inter-harmonics refer to the waveform distortion resulting from the signal imposed on the supply voltages by electrical equipment such as frequency converters, induction motors and arcing devices. Inter-harmonics consist of voltage or current frequency components that are not integer multiples of the supply system operating frequency. The consequences of these Inter-harmonics may result in flickering of fluorescent and arc lighting, as well as computers [31].

2.3.5.4 Notching

Notching refers to a periodic voltage disturbance caused by power electronic devices such as variable frequency drives and rectifiers commutated from one phase to another during normal operation [31]. Voltage notching characterises an unusual case that falls between harmonics and transients. In a steady-state system, notching could be distinguished by the harmonic spectrum of the affected voltage. This disturbance could lead to system fault, data loss and transmission problem issue [34].

2.3.5.5 Noise

Noise in the waveform is referring to undesirable voltage or current signal on the power system, which needs to be resolved using an appropriate component such as line conditioners, filters and isolation transformers. A signal exceeding 200 kHz is defined as noise [31]. Noise may also be defined as any undesirable power signal distortion, which is neither transients nor harmonics distortion. The noise can be generated by power electronics devices, control circuits, arc welder, switching power supplies and radio transmitter. There is some other condition that could lead to noise which is poor grounded sites and data corruption. Noise can cause technical equipment problems such as data errors, equipment malfunction, long-term component failure, hard disk failure, and distorted video displays. Data corruption is one of the most common results of noise .

2.3.6 Voltage Fluctuation

Voltage fluctuation refers to a systematic variation of the waveform voltage or a series of random voltage changes in small dimensions, i.e. 95 to 10% nominal at low frequency, generally 25Hz [35]. Any load exhibiting significant current variation can cause voltage fluctuation. This disturbance could lead to flickering of incandescent lamps [31]. Moreover, the Arc furnaces are the most common cause of voltage fluctuation on the transmission and distribution system. Removing the offending load, relocating the sensitive's equipment, or installing power line conditioning or UPS devices are methods to resolve the problem.

2.3.7 Power Frequency Variation

Power frequency variation refers to deviations occurring in the fundamental frequency, usually 50 or 60 Hz. The frequency is based on utility generators' rotational speed [31]. The fundamental frequency is set within strict limits under normal steady-state operation, but departures from these limits may be due to faults affecting the bulk transmission system. Frequency variation is more common, especially when the generator is heavily loaded; the IT equipment is frequency-tolerant and generally not affected by minor shifts in the local frequency generator. This disturbance may cause a motor to run faster or slower to match the frequency of the input. This would cause the motor to run inefficiently and lead to additional heat and motor degradation through increased motor speed and/or additional current draw.

2.4 Troubleshooting Power System Quality

When dealing with power quality issues, a single power system could be observed in different aspects depending on the problem that arises in the system. [28] define troubleshooting the Power System Quality as the activities of detecting and classifying power quality events, characterising and locating events, studying equipment sensitivity, and modelling of the system and equipment that are closely related and interdependent. The power quality event could be observed from the data acquisition and physical inspection of the equipment in line with the Condition Based Monitoring (CBM) data. Whenever an abnormality was found, a power quality trouble-shooter will collect sufficient measurement to characterise and locate the abnormalities. Later it will observe the sensitivity of the system by varying some parameters in order for them to model the characteristics into waveform to compare against the types of disturbance discussed earlier. [29] highlight the two common disturbance types are capacitor switching and voltage sag, where it is possible to determine whether a disturbance originates in front of or behind a recording device. In this scenario, power quality troubleshooting is a little bit tricky and need certain tacit knowledge and experience in evaluating the waveform to ensure it is closely related to the abnormality faced.

2.4.1 Power Quality Instrument

In the modern industry, most of the troubleshooting analysis required an instrument to measure the parameters in order to make a decision. This instrument is suggested to be calibrated to ensure all data are accurate to give explicit information on the problem. [36] mentioned that PQ troubleshooting instrument must have two operation modes which are Real-time Mode and Post-Processing Mode. For real-time mode, the data are collected and classified accordingly while in Post-processing Mode, all data are further analysed to determine the source and severity of the disturbances. These operation modes are important to ensure that the troubleshooting processes are systematic as trouble-shooter has the option of changing the mode depending on the criticality of the initial findings. In addition to that, five important features related to PQ troubleshooting instrument, as introduced by [36], were essential for all trouble-shooters as guidelines for them to perform the job. The features includes: (1) Availability of raw waveform data for further analysis as some existing PQ monitors have limited capability

to export raw waveform data, making them unsuitable for a number of troubleshooting applications, (2) Flexible data acquisition schemes in which the raw waveform should be collected in the user defined formats, (3) Flexible snapshots to ensure the reliability of event captured. (4) Disturbance source detection which is able to determine if a disturbance originates from upstream or downstream of the metering point and (5) Easy verification of instrument setup with a diagnosis function that can check if the instrument connection and probe ratios are appropriately established.

2.4.2 Power Quality Trouble-shooter Device

Modern technology in the areas of Power Quality troubleshooting has been improved from the classical point to point measurement to the data acquisition processing software which reduces the troubleshooting time and increase the possibility of determining the problem arose in the system. [36] has suggested three number of ways to construct a PQ trouble-shooter device which includes: (1) Developing the entire system using digital signal processors or similar hardware which has unlimited control over the system. The developer is free to customize the troubleshooting features specifically to the problem, but it has to start from the very basic hardware level and required expertise and time to develop the system. (2) Combining the custom hardware with the existing commercial Data Acquisition (DAQ) card. This is a little bit more flexible as the developer could customise its own signal-conditioning circuit hardware on the existing commercial DAQ card but attention to detail is required to ensure the DAQ card connection are accurately established. (3) By using the completely commercial DAQ hardware and software development platform which takes full advantage of the commercial signal-conditioning and DAQ hardware. It significantly reduces hardware development effort, but the developer has to comply with the specification/limitation provided by the vendors.

2.4.3 Power Quality Troubleshooting Methods

The characteristic of undesired power system is subjective, which required troubleshooting methods. [36] mentioned that some existing PQ monitors report did not have waveform snapshot, only depending on processed data. The point is that the true conditions experienced by the system may not be revealed since the data may have been manipulated by a series of built-in algorithms. Therefore, [36] have agreed with five troubleshooting methods to observe

the characteristics of the Power Quality disturbance, which includes:

1. Waveform snapshots which provide the trouble-shooter recordings of signal waveforms and sufficient raw data for further analysis in either auto (discretion) or manual mode (user-specified).
2. Data Trending Analysis which requires trouble-shooter to trend the data analysis using a statistical summary tool (histogram plot) or trends correlation among themselves and later conclude the preliminary findings.
3. Automatic Disturbance Capture using Power Quality Instrument which will provide information in terms of type and frequency of the disturbance.
4. Disturbance-Source Detection which will locate the direction of the Power Quality disturbance origin.
5. Verification of Instrument Setup by having a self-diagnosis feature for verification of instrument connection and later to ensure the magnitude, polarity, and phase sequence of the input signals are consistent with the user's configuration. The instrument setup is verified by displaying line power flow, phase sequence, degree of voltage imbalance, and a few other parameters.

2.5 Signal Processing Techniques

The process of interpreting the type of PQ disturbance involved three (3) layers which include Feature Extraction, Classification and Feature Selection [37]. Dealing with pattern recognition in PQ involves both traditional and modern signal processing methodologies. Many researches were conducted since 2000 (Scopus) in line with the increase of nonlinear load applications in daily usage. The process of classification is highly critical in reducing the different possibilities of root causes in any application. To make the study relevant to PQ problems, various types of disturbances need to be classified according to magnitude and duration [38].

2.5.1 Feature Extraction

Feature extraction is the classification process in PQ, which involves a different algorithm for different problems. The main challenge of these classification processes is to obtain an

accurate result with minimum computational time. Furthermore, it is frequently mentioned that it is critical to apply practical techniques to solve these issues in real-time. Figure 2.1 is an overview of the available feature extraction techniques in signal processing methods.

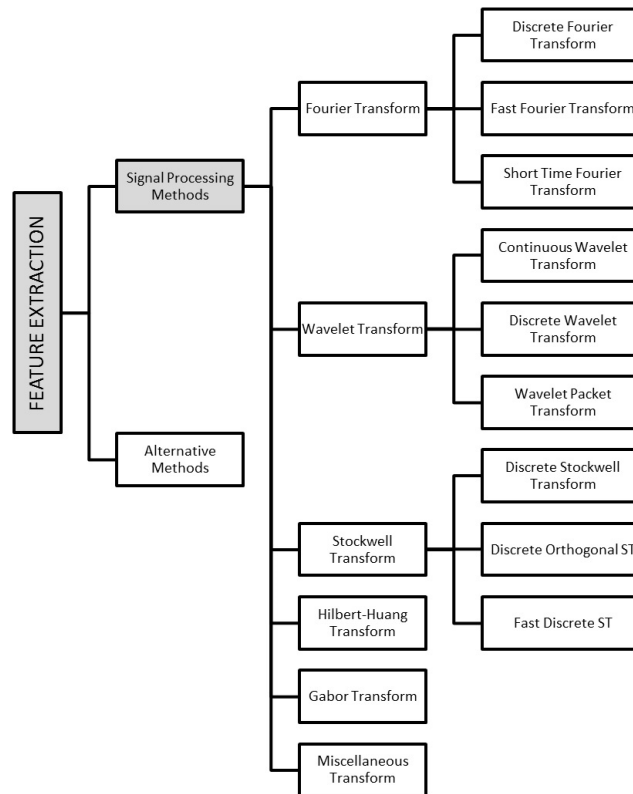


Figure 2.1: Different Signal Processing Methods in Feature Extraction

2.5.1.1 Fourier Transform

The most widely used computational algorithm for stationary signals is the Fourier Transform (FT) which extracts the time domain signal into specific frequency domains in two dimensions, frequency and amplitude. Three main algorithms that use FT are Discrete Fourier Transform (DFT), Fast Fourier Transform (FFT) and Short Time Fourier Transform (STFT). The difference between DFT and FFT is the extraction speed. DFT converts the sinusoidal signal into frequency. However, both methods lack information about time. STFT resolves the timing issue by dividing the signal into windows with segments and localising it in a different time dimension. Unlike DFT and FFT, STFT works for non-stationary signals. The fundamental operation of STFT is to slice up the signal into suitable overlapping time segments while per-

forming Fourier analysis on each slice to ascertain the frequencies contained in it [39]. The drawback of STFT is the constant window size for all frequencies in the signal, thus only producing good resolution either in time or the frequency domain depending on the window size [40]. Note that STFT is inappropriate for non-stationary PQ disturbances due to the uncertainty principle of incompatibility between time and frequency. However, the generalised version of Fourier transform known as Fractional FT (FRFT) was recently introduced, offering additional order control to visualise the disturbance and extract features from time, frequency and intermediate domains which make them comparable with Stockwell Transform [41] which will be further discussed later.

2.5.1.2 Wavelet Transform

Unlike FT, the Wavelet Transform (WT) algorithm provides three-dimensional information (frequency, amplitude and time). It is suitable for both non-stationary and stationary signals. It also provides both real and complex output compared to FT, which only provides complex output. The signal is mainly converted into a various scales of a short term waveform called "mother wavelet." In [42], an overview and comparison of commonly used wavelets resulting in the Daubechies wavelet are presented as the most widely used mother wavelets. The Daubechies wavelet transform performs very well in analysing PQ. The main advantage of WT over STFT is it uses short windows at high frequencies and long windows at low frequencies, thus narrowly monitoring the features of non-stationary signals [40]. The WT results in windowing segmentation techniques and is able to hold the characteristic of the signal in different frequency bands. It is sensitive to signal irregularity but not quite sensitive to the regular signal behaviour, but it is suitable for detecting and extracting disturbance of different type of PQ disturbance. Three main algorithms that use WT in PQ applications are Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT) and Wavelet Packet Transform (WPT). The main difference between the CWT and DWT is the scale parameter [43]. The CWT discretizes scale more finely than the DWT. The CWT typically uses exponential scales with a base smaller than 2, while the DWT always uses exponential scales with the base equal to 2. On the other hand, WPT is a generalization of wavelet decomposition that offers a richer signal analysis [43].

A review in [37] posits the wide use of WT in solving different PQ problems including

disturbance detection, feature extraction, data compression and data denoising of PQ disturbances. For disturbance detection, [37] has briefly highlighted the success of researchers in detecting and localizing non-stationary PQ disturbance signal using different techniques; Multi-Resolution Analysis (MRA) [44], and Multiresolution Signal Decomposition (MSD) [45]. A wide area of feature extraction has been using CWT and DWT for PQ classifications which later on extended to the hybrid approach of DWT-FFT for fuzzy expert decision making [46], Wavelet Network (WN) for larger classification [47], Adaptive Wavelet Network (AWN) for dynamic environment [48], Multi-Wavelet Transform (MWT) for transient disturbance [49], and Wavelet MRA (WMRA) for online classifications [50]. In addition, the Wavelet Packet Transform based on Hidden-Markov Model (WPT-HMM) [51] and Wavelet Packet Energy (WPE) based on Multiclass Support Vector Machine (MSVM) [52] had also been used for power distribution classification purposes.

It was mentioned in [53] that with the advancement of PQ monitoring equipment, there were issues on data storage and communication when dealing with a big amount of PQ data while and noise that could affect the recognition rate of the classifiers. Therefore, the integration of Spline WT and Radial Basis Function (RBF) in [54] was reported to be successful in data compression while denoising schemes which use WT are capable of suppressing the noise riding.

[55] suggested that a WT-SVM algorithm is a powerful tool in any real-time applications. The wavelet-based Global Disturbance Ratio (GDR) index used in [56] for Optimal Multi-Event Classification, results in high computation simplicity, accuracy and adaptability with a large potential for online classification. This research is in line with the current study to select a lower computational time approach. In [57], a hybrid approach combining WT and HT using simplistic mathematics has been presented by the author to detect and classify different types of PQ events. In [58], the author presented a new mother wavelet for electromagnetic transients applications using an adaptive wavelet algorithm based on the best classification results. [59] investigated the performance of PQ event using WT and Weighted Extreme Learning Machine finding and observed that these WT-WELM methods are robust and have better classification of speed and higher recognition accuracy. In spite of the success of WT, there are few PQ problems that cannot be resolved with WT including Voltage Swell (Heavy load switching or capacitor bank switching) [60].

2.5.1.3 Stockwell Transform

Another leading classification method in the PQ assessment is the Stockwell Transform (ST) [61]. It is a hybrid of the STFT and WT which provides the time-frequency representations with frequency-dependent resolution. The key feature of this algorithm is its ability to detect, localize and quantify the disturbance in both low frequency and high-frequency components which comprise most of the PQ disturbances such as voltage sag, voltage swell, momentary interruption, oscillatory transients, notches, harmonics with sag and swells. Moreover, it is also capable of providing a better time-frequency representation of a signal to be applied as inputs to intelligent classifiers. The three leading ST algorithm are Discrete Stockwell Transform (DST), Discrete Orthogonal Stockwell Transform (DOST) and Fast Discrete Stockwell Transform (FDST).

DST has high computational cost due to its high redundancy on its large size data representation [61] & [62]. To overcome this drawback, DOST was introduced [40]. DOST sets on the orthonormal basis function which samples the DST with non-redundancy while retaining the original phase properties. It has multiple scales and a referenced phase. However, it is quite challenging during the interpreting stage as it provides coarse time-frequency representations with its frequency resolution proportionally scaled to the logarithm of the frequency [63]. In other words, the zero redundancy has made the signal interpreting becomes difficult due to insufficient information in the signal. On the other hand, the FDST algorithm which was introduced in [64] proves to be low computational cost and uses fewer resources.

A review by [18] reports the success of ST in PQ analysis including the hybrid approach of ST-Dynamic [65] in reducing run-time, the ST-MRA techniques in performing variable window changing [66], and hybrid of windowed FT and ST to extract distinctive features of PQ signals [67]. ST has also been used as a feature extraction method for different classifiers such as Probabilistic NN [68]; Fuzzy Expert System [69]; Rule Base [70]; Decision Tree [64], [71]; Fuzzy Decision Tree [72]; and ANN [73].

ST is a popular method amongst analysts and researcher such as [74] who investigated the performance of the hybrid decision tree and Fuzzy C-means clustering algorithm based on ST, demonstrating higher efficiency in comparison with other methods. [75] presented a rule-based ST as an extraction tool, and Adaptive Boost with decision stump as a weak classifier for PQ assessment to be better than SVM and Decision Tree. [76] suggested the hybrid of ST-WT with

PNN as an intelligent classifier to effectively detect PQ events in noisy environments. Recently, [77] introduced a modified S-Transform (MST) achieving desired time–frequency resolution for various PQ disturbances signals and investigated its performance with the standard ST using three classifiers (BP, SVM and ELM). Results showed that MST performed better than ST in low SNR while ELM has higher accuracy, faster and independence from a training set.

2.5.1.4 Hilbert Huang Transform

Hilbert Huang Transform (HHT) is also among the latest algorithms used in the PQ disturbance assessment by extracting non-stationary signals. HHT is the combination of time-frequency analysis of Empirical Mode Decomposition (EMD) and instantaneous estimation of Hilbert Transform (HT). It estimates the frequency, amplitude and phase. The EMD is a technique for detecting and localizing the transient features including the periodical notch, voltage dip, and transient oscillatory disturbance by generating the Intrinsic Mode Function (IMF) from a non-stationary signal [78]. The orthogonal signal produced by HT is phase shifted by 90 degrees from the original signal. The HT features are proven to be less sensitive to noise level [79].

A review in [80] discusses the original methods [81], and the applications of classifiers e.g. Radial Basis Function [79]; Relevance Vector Machine [82]; and Support Vector Machine [83], [84], [85]. The HHT has also been presented in a mathematical morphology [86] and multi-disturbance complex [87] while in [88], the HHT was improvised orthogonally for Voltage Flicker Analysis. [89] investigated the classification performance of voltage sag sources in an EMD-SVM algorithm and compared it with WT-PNN and EMD-PNN. The results show that the EMD-SVM scores the highest efficiency and accuracy.

[73] investigated the performance of HHT-PNN with different classifiers and extraction methods, highlighting the persistent demand for intelligent techniques to perform well in single stage and multiple PQ events. The latest research adapting HHT was carried out in [90] where comparisons were made between feature extraction performances with ST in detecting voltage sag in the power grid. The conclusion was that in real-time analysis, ST is not capable of detecting noise as much as HHT.

2.5.1.5 Gabor Transform

Gabor Transform (GT) [91] on the other hand is the extended version of STFT which uses Gaussian function as a window while the resulting function is again transformed with FT to derive the Time-Frequency analysis. The Gaussian function has the same properties of the time and the frequency domains. It was found effective in monitoring the frequency variation of a signal as time varies [92] and have less computational time. However, it has a drawback in a trade-off between time and frequency resolutions, caused by the fixed width of the window. Nevertheless, GWT which is the hybrid of GT and Wigner distribution function [93] has improved the time and resolution at the same time and covers the cross-term problem of WDT and the low clarity of GT.

The application of GT in PQ prediction was reported in [94]. A new synthetic methodology integrate GT with Probabilistic Neural Network (PNN) demonstrating high classification capability in noisy conditions. In [95], the author presented a combination of GT-ANN methods for transient fault identifications. The method confirmed that GT was able to detect and localize signal disturbance. [96] proposed a new method for online detection using DGT-SVM, resulting in a reduction in features of the disturbance while minimizing the execution time and memory for training and testing purposes.

[97], introduced the hybrid of finite impulse response window with GT (FIR-DGT) for the detection and combination of SVM with Type-2 Fuzzy Kernel (T2FK-SVM) as a classifier leading to the reduction of computational cost and improvement in accuracy due to fast learning and execution. In [98], the author investigated the performance of the [97] with other methods demonstrating high overall accuracy.

2.5.1.6 Other Miscellaneous Transform

Apart from the above techniques, some other extraction techniques have played significant roles in the PQ detection and classifications including Kalman Filter [99], [100]; Extended Kalman Filter (EKF) [101] and Unscented Kalman Filter (UKF) [102]. Meanwhile, [103] and [104] introduced Time-Frequency ambiguity plane with kernel techniques using Time Frequency Representation (TFR) and [105], [106] used two dimensional time-time representation TT Transform. Other techniques such as Chirplet Transforms [107]; Mathematical Morphology (MM) [108]; Slant Transform [109]; and Teager Energy Operator (TEO) [110] have proposed

methods to extract certain disturbances. Other extraction techniques such as Spectral Kurtosis [111]; Principal Curves (PC) based on Principal Component Analysis (PCA) [112]; and Sparse Signal Decomposition (SSD) [15] have made important progress in feature extraction techniques.

2.5.2 Feature Selection

Feature selection is a process of selecting the most useful and relevant data to ensure low redundancy in classifying the PQ Disturbance. Based on the extraction data, optimisation techniques such as Genetic Algorithm (GA), Particle Swarm Optimisation (PSO) and Ant Colony Optimisation (ACO) are the common techniques used to reduce the level of redundancy and complexity in the data-set while contributing to high accuracy classification and reducing the overall computational time. The main challenge of this process is to achieve the shortest computational time. Figure 2.2 is an overview of the relevant feature selection techniques.

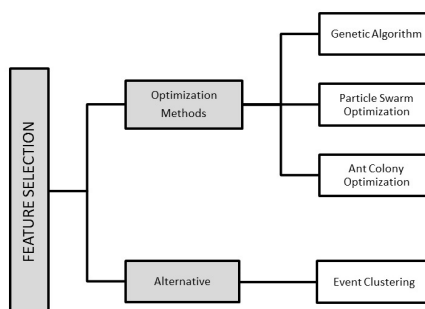


Figure 2.2: Different Optimisation Techniques in Feature Selection.

2.5.2.1 Genetic Algorithm

The GA is inspired by the natural evolution (for e.g. mutation, selection and crossover) which uses an adaptive heuristic search method in finding the optimum solution. The process starts by initiating and evaluating each candidate of a population stochastically, and later to select the parent while recombining (mutation) the pairs of parents to form a new population. The process (iteration) is competitively repeated until it satisfies the optimised global solution by eliminating poor solutions using generic parameters (crossover and mutation) based on a heuristic.

[113] presented an optimisation technique for selecting the best features out of 128 features obtained from WPT using GA and Simulated Annealing (SA) and produced highly accurate grid search methods while the classification of SVM classifier has significantly reduced the computational time. [114] investigated the performance of Extension Genetic Algorithm (EGA) with K-means and fuzzy c-means demonstrating the high accuracy (96%) of EGA and its advantage in removing off the extension of experience rule.

[115] presented a meta-heuristic Micro-Genetic Algorithms (MGA) to detect and classify electric power disturbances with low computational, simplicity and reduction in resources. [116] combines the optimisation performance of GA (structure optimisation) and Harmony Search (parameter optimisation) in the classification of the time-series sequential data in training the HMM structure. The results show that the optimised HMM has a higher probability in classifying the global maximum compared with the un-optimised HMM.

2.5.2.2 Particle Swarm Optimisation

Particle Swarm Optimisation (PSO) was firstly developed by Jim Kennedy, inspired by the simulation of social behaviour based on bird flocking, fish schooling and swarming theory. This optimisation tool uses the concept of sharing information in adjusting the velocity and particle position based on individuals' and the neighbouring particle (members of the population) experiences. The algorithm starts by identifying the location and velocity of every particle. Later, each particle will evaluate the objective function and update the best solution to find the best global solution. The iteration will be continuous until it obtains the best position.

The PSO was applied to a number of engineering applications including PQ. PSO is a powerful heuristic searching algorithm for finding the optimal parameters. PSO has some key advantages over other optimisation techniques in terms of search capabilities, memory of a good solution, algorithmic simplicity, simple implementation, low computational effort, high accuracy, high speed and convergence [117].

[118] used the PSO algorithm to specify the optimal parameters for the membership function parameter in a fuzzy system to detect and classify PQ disturbances. The results showed great capability and high accuracy under noisy conditions. PSO has also been used to distinguish between the types of sag, interruption and swelling in the deep learning-based classification. The robustness of the PSO in dealing with noise is also noteworthy [119]. [120]

introduced the hybrid of fuzzy C-means and Adaptive PSO to optimise the cluster centre of features into distinct groups resulting in maximum classification accuracy. The accuracy performance of APSO and GA techniques were also compared, showing that GA-based clustering yield better results in some disturbance pattern; however, it is computationally expensive compared to APSO.

2.5.2.3 Ant Colony Optimisation

Ant Colony Optimisation (ACO) is one of the most recent optimisation techniques. The ACO is mainly used in dynamic applications such as circuit switch networks, and packet-switched routing network while having the potential to retain memory among the entire colony. ACO is an effective optimisation tool which helps in refining data clusters by optimising the Euclidean distance of each data points from the centre. [121] presented a Hybrid of ACO (HACO) and fuzzy C-means clustering algorithm in generating a decision tree for classifying various power signal disturbance classifications resulting in robust pattern classifier and high pattern recognition accuracy compared with other optimisation methods such as PSO, GA and classic ACO.

2.6 Sensitivity Analysis Techniques

In spite of the successful progress in signal processing methods, there are still demands for new, hybrid or improvised ways to solve PQ control and classification problems. Furthermore, it is vital to include external parameters in the data to observe any other impact which could link to the PQ problems. In this section, the relevance of input variable selection and sensitivity analysis within the context of time-critical and computation resource-limited dimensionality reduction problems [122] are investigated. It may challenge the ways of feature extraction from the modern complex systems and infrastructure.

Real-time data acquisition is materialised by industrial controller system and supervisory control and data acquisition (SCADA) system providing big data on the machine state, instrumentation values, alarms and sequence of events. This requires large database memory to store, process and for communication capabilities to guarantee continuous monitoring and control. As systems get more complex, the amount of data collected from multiple sources

grows; subsequently, it is more challenging to identify the most useful and relevant parameters of performance and system state. This could worsen the curse of dimensionality problem in a large data set. Eventually, it needs to find a way on how to eliminate input variables that have the least impact on the system rapidly, computationally cheap and reliable.

2.6.1 Feature Extraction

In the next section, details related to the existing techniques that have been developed in recent years in the context of Input Variable Selection (IVS) and Sensitivity Analysis is presented. Figure 2.3 shows a summary of feature extraction techniques using input variable selection and sensitivity analysis methods.

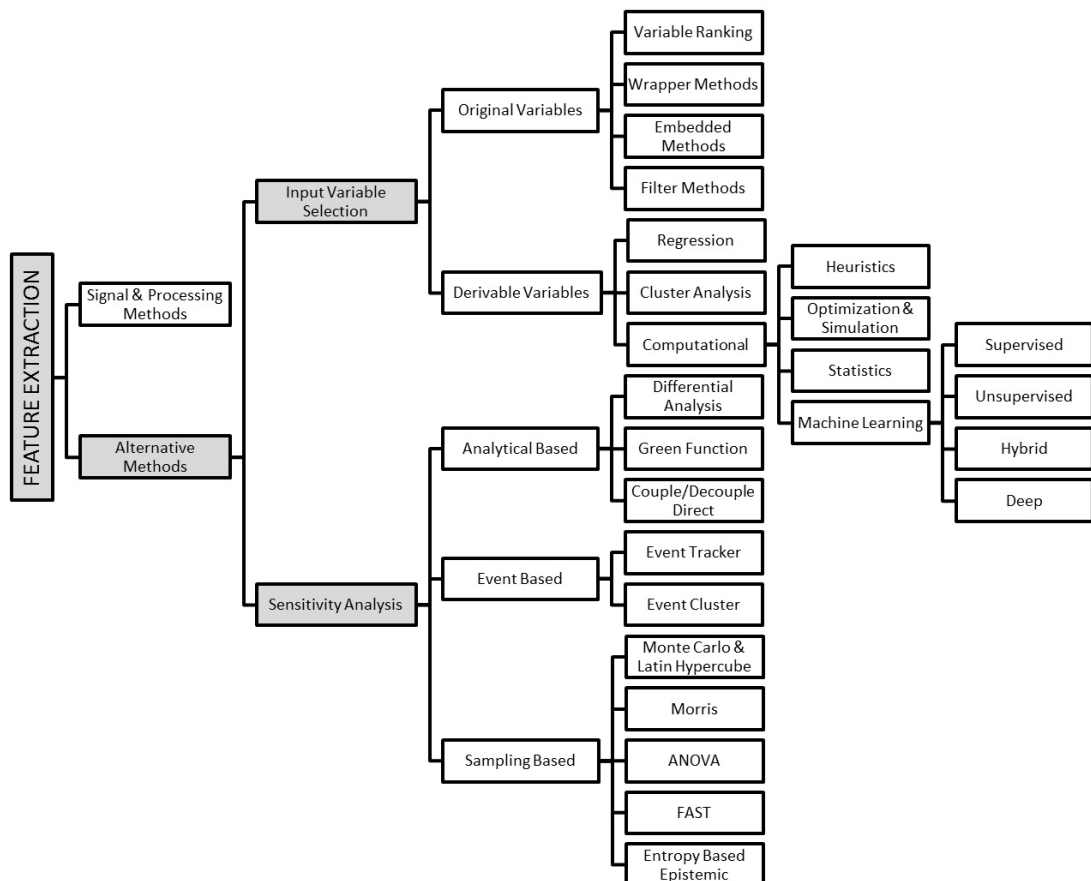


Figure 2.3: Different Input Variable Selection and Sensitivity Analysis in Feature Extraction.

2.6.1.1 Input Variable Selection

A relevant learning algorithm is needed to evaluate a large scale complex system. The task of selecting input variables is largely dependent on the discovery of relationships within the available data to identify suitable predictors of the model output [123]. The objective of input variable selection (IVS) is three-fold: improving the prediction performance of the predictors, producing faster and more cost-effective predictors and providing a better understanding of the underlying process that generate the data [124]. A comprehensive review by [122] has highlighted the IVS from (1) Primary or Original Variable Methods (OVIVS) and (2) Secondary or Derived Variable Methods (DVIVS) perspectives. The major distinctions between those two lies in the issue of keeping variable intact and only deciding on their redundancy, or in contrast, transforming them to new subsets of variables.

The DVIVS relates to data storage and the processing of new variables. These issues contribute to the computational overhead of DVIVS, making them less attractive to use in a time constraint (real-time) data-driven decision support system. In contrast with DVIVS, OVIVS does not add burden to the time-critical data integration. It is hard to find an IVS method that could fulfil (1) promptness, (2) accuracy, (3) computational efficient (in terms of time and resource critical applications). The modern and evolving complex systems have huge data acquisition capabilities making IVS an essential technique in extracting relevant information to manage interrelations and dynamics of a component within the environment questionable. The machine operation itself has massive data changes while the other operations such as key performance index data, power system data, and environment data will cause the system even more complicated with the increased number of system parameters. Therefore, the process of identifying suitable predictors may take a longer time and may or may not produce accurate results due to its complexity.

2.6.1.2 Sensitivity Analysis

Sensitivity Analysis (SA) in the process of engineering is minimising computational overhead by eliminating the input variables that have the least impact while focusing on the more significant variables by exploring the type of relationship between the input and output parameter of the system.

The review by [122] categorised sensitivity analysis into (1) Analytical Based which com-

prises of Differential Analysis, Green Function, Couple/De-Couple Direct and (2) Sampling-Based which includes Monte Carlo and Latin Hypercube Morris, Analysis of Variance (ANOVA), Fourier Amplitude Sensitivity Test (FAST), Time uncertainty, and Entropy-Based Epistemic. Each of these methods has its own advantages and disadvantages which vary in terms of result accuracy, scalability, heterogeneity and derived variable or original variables.

2.6.1.3 Event Tracking

It has been discussed in [122], [21], [125], [126] that most of the sensitivity analysis techniques are inefficient due to their computational constraint, time-consuming and high dependency on historical data. However, this Event Tracking method uses an input-output occurrence matrix populated at ‘pre-defined’ time intervals, a link between the data of actual event (Event Data) to the cause of the triggered events (Trigger Data) by using data mapping concept.

2.6.2 Feature Selection

The success of the Feature Selection techniques in detecting the most useful and relevant data does not stop the researcher(s) from exploring other methods which could further optimise the system. One of the heuristic methods recently introduced is Event-tracking [122]. Event tracker produces an input-output occurrence matrix that needs to be optimised by grouping it according to its interrelationship and internal dynamics of the component within the ecosystem. This could be achieved by performing clustering methods which will be further discussed as follows.

2.6.2.1 Event Clustering

The principle of interrelation causal events of “Event Tracker” was further extended to the “Event-Clustering” in real-time data modelling by [125]. This method intends to tackle control and stability operation in a large and complex system. It uses the Rank Order Clustering (ROC) which was initially introduced by King (1980) [127] by rearranging the row and column of a matrix in the iterative manner of decreasing value order. However, these methods are compounded by the assumption of highly similar groups of data, and it will be placed into a mutually exclusive block. The group will either be replaced by a new value representing the group (clumping) or assigned as a unique type of label (portioning) [128]. Furthermore, real-

time clustering always assumes that the changes in the input may trigger events. A detailed step by step of implementing the Event Clustering algorithm is discussed in [21] and [125].

2.7 Classifiers Techniques

The application of Artificial Intelligence (AI) as classifiers in PQ has had exponential growth in the last ten years. It has a relation to human thinking that automates the decision making and problem-solving based on its learning capabilities [53]. Artificial Neural Network (ANN) is among the most popular and influential classes of machine learning algorithms. The ANN model has been applied not only in PQ but various applications of pattern classification, pattern recognition, optimisation, prediction, automatic control and function approximation (Regression). The motivation of this ANN technology has always been to develop an AI system that is capable of acquiring knowledge like a human brain, by extracting pattern through a sequence of real value activation of this neuron learning and its inter-neuron weight strength connection.

The revolution of ANN started from 1940's as a computational model using threshold logic [129]. Later in 1960's the perceptron convergence theorem [130], and limitation of simple perceptron was discovered [131] but challenging to be practised with the X-OR problem. Later in the 1980's, significant progress was achieved with the content such as addressable memory system of Recurrent NN [132], [133] and the back-propagation learning algorithm for multi-layer perceptron's [134] which solved the X-OR problem [135].

Later in 1990, the Unsupervised Learning was given a new life to NN technology while it has consistently improved the previous Supervised Learning until it has won many official pattern recognition awards. Some other machine learning progress are SVM [136], Long Short-Term Memory [137], Convolutional Neural Network (CNN) [138] and Deep Belief Networks [139]. A comprehensive review by [140] has highlighted the detailed overview of a neural network and deep learning, respectively. Figure 2.4 shows a summary of classifiers using different AI Techniques.

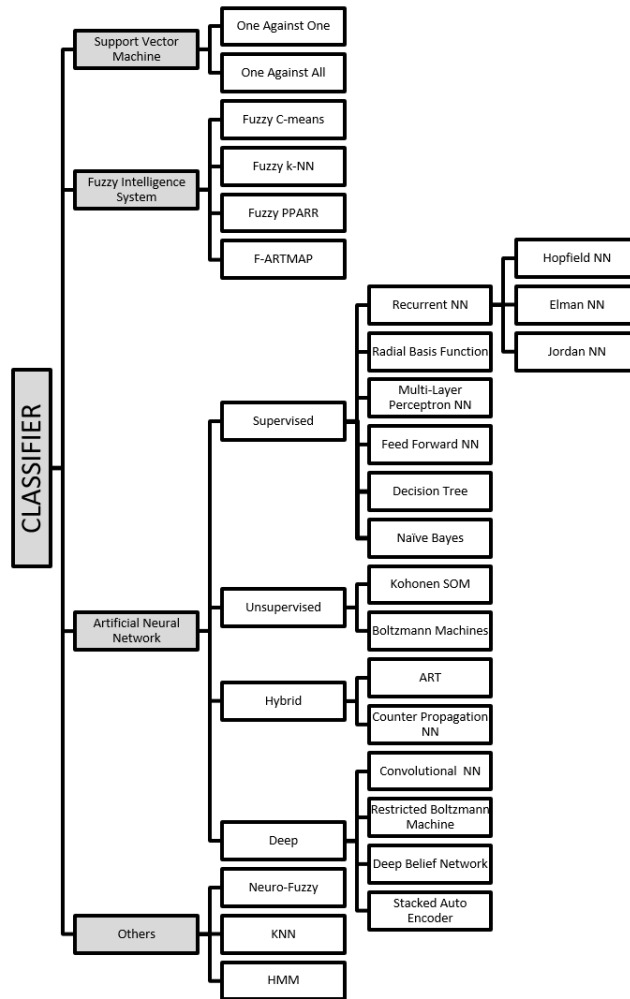


Figure 2.4: Different Artificial Intelligence Techniques in Classifiers.

2.7.1 Supervised Learning

The traditional Feed-Forward networks which consist of Multi-Layer Perceptron's (MLPNN) and Radial Basis Function Network (RBS) are widely used in pattern recognition, speech recognition, system identification, medical diagnosis and fault detections. Both MLPNN and RBF structures are only static approximators. [141] has successfully used MLPNN to classify the PQ disturbance using four output nodes representing the type of disturbances: voltage sag, voltage swell, transient and harmonic distortions. [142] revealed that ANN was more accurate than fuzzy logic as the classification tool for various PQ disturbances but can be slower and very complex. [143] has implemented ANN to classify PQ disturbances, achieving

more than 90% accuracy for all type of disturbances.

On the other hand, the RBF neural network is a forward network model with good performance, global approximation, low training time capabilities and free from the local minimum problems [144], [145]. The structural simplicity and training efficiency of RBFN have made it a nonlinear mapping capability which is able to augment new training data without re-training [146]. [147] has shown that RBF techniques require less sampled data for harmonic assessment. [145] has effectively estimated the fundamental, fifth harmonic and seventh harmonic components in converter waveforms using RBF. Since the feed-forward neural networks cannot retain information about the infinite past, their performances in identifying nonlinear dynamical system will not be as good as those of the recurrent neural networks.

The Recurrent MLP's is a special kind of feed-forward neural network with additional memory neurons and local feedback. It can deal with nonlinear time-varying systems and capable of attenuating noise by interacting with signals using their own dynamics [148]. Two types of the most common RNN are the Elman NN and Jordan NN. Both RNNs used the back-propagation learning algorithm for training. During the training phase, the error is propagated backwards from the output layer to adjust the weights of all feed-forward connections while keeping the constant weight [149]. The main difference for both RNN is that the Elman NN has information feed-back from the hidden layer, whereas the Jordan NN uses feedback from the output layer. The self-connections of the context nodes in the Elman network also make it sensitive to the history of input data which is very useful in dynamic system modelling [150]. A common problem of RNN mainly comes from inappropriate network structure and parameters. An undersized network is unable to generate a satisfying performance, whereas an oversized network might have poor generalisation capability [148]. [151] used RNN on PQ data which have problems related to sag and swell. The outcome shows excellent performance in terms of accuracy, sensitivity and specificity in comparison with FFBP neural network. [152] applied the Jordan type NN to track maximum power point for the wind speed controller without requiring the knowledge of wind speed, air density or turbine parameters.

Decision tree (DT) is a kind of classifier similar to tree structure. DT has the advantages of simple structure, convenient expansion and fast classification speed. DT is usually used to recursively select the best feature and segment of the training data according to the feature [153]. Knowing that SVM is affected by kernel function, [154] combined SVM and binary DT

classifier that reduces the number of decisions made in the testing phase for classifying power quality disturbances. [64] examined FDST based on DT classifier with 40 dB noise power which obtained the best classification accuracy among other classifier such as PNN, expert system, ANN, neuro-fuzzy, and SVM. Recently, [153] found that DT classified faster and more accurately compared to SVM and KNN. When using mathematic algorithm, DT could classify multiple PQ disturbance with low computational effort [155].

Naïve Bayes (NB) is a probabilistic classifier with strong independent assumptions. It is based on Bayesian theorem, which assumes the presence of a particular feature of a class, which is unrelated to the presence of any other feature. Key advantage of using NB is that it requires small amount of training data to estimate the parameters for classification. [156] found NB works efficiently in recognising the features when pairing with fast Time-Time transform. [157] found NB and k-Nearest Neighbors (KNN) are comparative robust when dealing PQ events. Recently, [158] found that KNN and NB perceived the accuracy of 96% in identifying and diagnosing the harmonic sources in the power system. On the other hand, [159] recently trained NB to develop a classifier that is capable of classifying PQ signals with 100% accuracy.

2.7.2 Unsupervised Learning

The Self-Organizing Feature Map (SOM) developed by [160] is easily trained and has attractive properties such as topological ordering and good generalisation. It implements a characteristic non-linear mapping from the high-dimensional space of input signals onto a typically 2-dimensional grid of neurons [161], [162]. The analysis using SOM provides accurate results with a significant saving in computation time, reduction of the online computational requirements. The advantage of SOM is that it tries to identify significant patterns in the disturbance feature vectors and assign them to different disturbance classes [163]. It can be used for clustering data without explicit knowledge of the input data. SOM technique is chosen mainly due to its ability to analyse and classify data with a high level of complexity. SOM is an excellent tool for the visualisation of high dimensional data. [164] and [165] used a novel approach for PQ disturbance classification using Self Organizing Learning Array (SOLAR) system with Wavelet transform to classify PQ disturbance. Based on the outcome, they highlighted three advantages over a typical neural network: data-driven learning, local interconnections and

entropy-based self-organisation. In [166], a PQ disturbance recognition based on S-transform and SOM was used to recognise the types of power disturbance signals with a great noise. Meanwhile, [167] used SOM to locate the PQ source of disturbance including transient disturbances. In [168], the author proposed a SOM-BP hybrid network to further separate and locate fault diagnosis for wind power generation application.

The Boltzmann Machines [169] is widely used in combined optimisation and associated memory problems which solve the process scheduling issue. It can move (temporarily) towards the worst state to escape from local traps, and it uses a probability rule to update the state of a neuron and its energy function [170]. The energy function hardly ever falls into a local minimum which made it very useful for solving combinatorial problems and topological observability problem [171]. The architecture of a Boltzmann machine is arranged in a two dimensional array. The neurons within each row are fully connected, as well as the neurons in each column. The self-connection within the neuron is a positive value while the weight of the connection between neurons is a negative value. The values of the weights are chosen such that the global energy of the network reflects the extent of the consensus of the neurons state [172]. The Boltzmann machine is more efficient in solving the topological observability problem compared to Hopfield. The learning algorithm is very slow in networks with many layers of feature detectors, but it is fast in "Restricted Boltzmann Machines" which have a single layer of feature detectors. To solve learning problem, Boltzmann machines make many small updates to their weights, and each update requires them to address many different search problems [173].

2.7.3 Hybrid Learning

The Adaptive Resonance Theory [174] is widely used in pattern recognition, face recognition, and data clustering. It is capable of classifying a large, highly-dimensional and time-varying set of input data [175]. The learning is completed when all the patterns are grouped into homogeneous clusters. Contrary to SOM, the initial number of clusters and cluster centres are not specified in advance, but the clusters are allocated incrementally [176]. An ART network can be considered as a vector classifier; it classifies an input pattern (or vector) depending on the number of stored patterns it most resembles. The ART network contains four types: ART1, ART2, ART3 and Fuzzy ART. The ART-1 uses a supervised clustering algorithm and is

usually applied to binary input patterns. The ART-2 uses unsupervised clustering algorithm and are typically referred to either analogue or binary input patterns. It has fast learning capabilities, and easily adaptable to a non-stable environment while managing to precisely and automatically decide the number of clusters. The ART3 is applied to the biochemical mechanism of a transmitter in the human brain system to a neural network to carry out the hierarchical search [177]. Fuzzy ART is the combination of ART1 features and fuzzy set theory [178]. In [177], a gas turbine unit used ART2 networks to correctly diagnose fault with rapid training. The outcomes were highly reliable as it indicated the onset of developing faults while it was flexible in providing the mechanism for predictive maintenance requirements. [179] has presented a pattern recognition algorithm based on five types of ART to detect and classify High Impedance Fault (HIF) in a power distribution network. The results have shown that Fuzzy ART and Fuzzy ARTMAP have the highest accuracy while ART1 has the worst outcome.

2.7.4 Deep Learning

Deep learning is inspired by the architecture depth of the brain. Hinton, Osindero & Teh (2006) [139] made a breakthrough in solving the failure of training deep multi-layer neural network by introducing Deep Belief Network (DBN) in which the learning algorithm trains one layer at a time. Later, [180] introduced the Auto-Encoders. Both concepts exploit the unsupervised learning algorithm of Restricted Boltzmann Machine. The advancement of deep network study gained popularity for solving classification and regression tasks. In addition, there are also some new areas which include dimensionality reduction, modelling textures, modelling motion, object segmentation, information retrieval, robotics, natural language processing and collaborative filtering.

The Convolutional Neural Networks (CNN) is a leading deep learning model in detection, segmentation and recognition of objects and regions in images. It is inspired by the biological architecture, which has multiple layers of neuron collections with learn-able weights and biases. The network consists of three layers of convolution: max-pooling, rectified linear unit (ReLU) and local normalisation, followed by a fully connected layer and a linear classifier at the top [181]. The outputs are overlapped to obtain a better representation of the original objects [182]. CNN uses local connections to extract spatial information efficiently and shared weights

to significantly reduce the number of parameters. CNN exploits the local correlation using local connectivity between the neurons of nearby layers. The power of CNN depends on the connections (weights) of the network; hence, it is vital to find a set of proper weights. The CNN is also widely used in computer vision, natural language processing [183], Optical Character Recognition (OCR) and speech recognition [184]. Some key advantage of this method is that it tends to provide better generalisation when facing computer vision problems, and it exploits the local correlation using local connectivity between the neurons of near layers. Recently, [185] has used CNN and Softmax regression to classify power quality signals and compare with another technique of autoencoder and SVM, which results are superior than those obtained by an Auto-encoder.

Auto-encoders are simple learning circuits which aim to transform inputs into outputs with the least possible amount of distortion. It has been introduced in line with the Hebbian learning rules to address the problem of backpropagation in unsupervised learning globally. Later, the Auto-encoders are stacked and trained bottom-up in the form of RBM followed by a supervised learning phase to train and tune the entire architecture.

The Restricted Boltzmann Machines (RBM) is a generative stochastic artificial neural network that can learn using distribution probability over its set of inputs [186]. It works in an unsupervised manner with one visible and one hidden layer. The network is an undirected graphical model in which the neurons in one layer that are all connected to the neurons in a second layer [187]. In [188], the author highlights the efficiency of RBM tool in dimensionality reduction of input data [189]. It is also used in classification, regression, collaborative filtering, feature learning and topic modelling. In most application, it became the fundamental building blocks of the Deep Belief Networks (DBN). RBM's have three parameters: connection weights, visible biases, and hidden biases. [187] briefly elaborate RBM calculation flow, which consists of two repeating steps. When the input data is added to the visible layer, the hidden layer will be calculated using the visible layer's value. Next, the visible layer is calculated using the sampling results from the hidden layer. This step was repeated to update the formula. To the best knowledge of the author, no research has been done in the area of PQ classification using RBM.

Deep Belief Networks (DBN) was developed by [173] as a pre-training tool for a deep neural network. It comprised two parts: layer-wise greedy unsupervised learning pre-training part

and supervised the fine-tuning process part. The pre-training part is a stack or (multi-hidden-layered) of RBM which reconstruct the input data when learning the new representation with no label information provided. The fine-tuning part is the deep auto-encoder together with the Softmax regression layer [190], which encodes the inputs to new representations and then decodes the representation to the reconstruction of the inputs [191]. The nodes of any single layer do not communicate with each other. By applying DBN for pre-training and fine-tuning, the computational effort has been improved. It avoids random guesses at starting point and reduces training duration and final training error. However, as dimensionality increases, the computational effort also rapidly increases [192]. DBN is widely used to recognise, cluster and generate images, voices, text, video sequences and motion-capture data, data fitting, recognition and classification. In [193], DBN is used as a semi-supervised classifier. Recently, DBN has made a significant breakthrough in solving the overfitting problem caused by the conventional backpropagation algorithm [194]. To the author's best knowledge, no research has been done in the area of PQ disturbance using DBN.

2.7.5 Support Vector Machine

Support Vector Machine (SVM) was first introduced by [136] as a machine learning tool for classification and regression analysis. In general, the SVM is used to determine the best hyperplane that divides any two different classes into two. The capability of the SVM algorithm produces better classification performance compared to the ANN. In linear classification, the SVM algorithm will separate the samples in the hyperplanes into two classes (Class 1 and Class 0) while setting up a distinct margin between the two classes. Margin refers to the sum of a minimum distance between the training data set with the separating hyperplane. Compared to ANN, the activation function in SVM uses the kernel functions to transform the nonlinear separable dataset into a new high dimensional feature space. A wide range of Kernel functions includes Linear Kernel, Polynomial Kernel, Radial Basis Kernel and Sigmoid Kernel. SVM is widely applied in the large classification problems due to its capabilities of handling large feature vector dimension and better generalization properties.

2.7.6 Fuzzy Expert System

Fuzzy set theories consider human expert knowledge to simplify decision making for highly complex problems. It generally makes approximations on uncertainties based on binary logic, by mapping the object to the membership function values in the set. The application of natural language terms in Fuzzy sets was introduced by [195] based on the original idea of [196] which mentioned about the ranges in logic truth values. In classification studies, the Fuzzy Expert System (FES) uses fuzzy sets along with the fuzzy rule base, which offers human reasoning capabilities to classify the elements.

2.7.7 Other Classifiers

Apart from the above techniques, some other classifiers have played a significant role in the PQ detection and classifications activity which includes Hidden Markov Model (HMM) [197], [198]; Probabilistic Neural Network (PNN) [68], [81], [199], [200], [201], [202], [76]; Optimized ANN [203], Adaptive Boost (AdaBoost) [75], Extreme Learning Machine (ELM) [77] and Weighted Extreme Learning Machine (WELM) [59].

2.8 Summary Analysis

The summary analysis on different methods and techniques in proposing the most suitable data analytic techniques are illustrated as per Table 2.1. Six main criteria have been set to compare between the available techniques which include: (1) Scalability of the data, (2) Original or Derived based approach, (3) Support Heterogeneity Data, (4) Producing Accurate Results, (5) Global Solution and (6) cheap computational time. The observation reveals that most of the techniques are not scalable and have long computational time. This is mainly due to the properties of the variables which requires more effort to handle a growing amount of data which is not selective. Even though most of the techniques produce accurate results and is a global solution, the solution data may not include the external parameters which may also potentially be the contributing factor to the problem. Meanwhile, producing a real-time solution is an advantage in order to make a fast decision to avoid further loss. Therefore, to the author's best knowledge, the event modeller technique is capable of addressing this issue. With the help of a suitable Machine Learning techniques, the synergy of these two techniques

Table 2.1: Summary Analysis of Different Techniques

Methods	Category	Scalable Data	Original/Derived	Heterogeneity	Accurate Results	Global Solution	Computational Time
Signal Processing Methods							
Discrete Fourier Transform	Fourier	No	Derived	No	No	No	Long
Fast Fourier Transform	Fourier	No	Derived	No	No	No	Long
Short Time Fourier Transform	Fourier	No	Derived	Yes	No	No	Long
Continuous Wavelet Transform	Wavelet	No	Derived	Yes	Yes	No	Short
Discrete Wavelet Transform	Wavelet	No	Derived	Yes	Yes	No	Short
Wavelet Packet Transform	Wavelet	No	Derived	Yes	Yes	No	Short
Discrete Stockwell Transform	Stockwell	No	Derived	No	Yes	Yes	Long
Discrete Orthogonal Stockwell	Stockwell	No	Derived	No	Yes	Yes	Short
Fast Discrete Stockwell Transform	Stockwell	No	Derived	Yes	Yes	Yes	Short
Hilbert-Huang Transform	Hilbert-Huang	No	Derived	Yes	Yes	Yes	Short
Gabor Transform	Gabor	No	Derived	Yes	Yes	Yes	Short
Kalman Filter	Others	No	Derived	Yes	Yes	Yes	Short
Extended Kalman Filter	Others	No	Derived	Yes	Yes	Yes	Short
Unscented Kalman Filter	Others	No	Derived	Yes	Yes	Yes	Short
Time Frequency Representation	Others	No	Derived	Yes	Yes	No	Long
TT-Transform	Others	No	Derived	Yes	Yes	Yes	Short
Chirplet Transform	Others	No	Derived	No	No	No	Long
Mathematical Morphology	Others	No	Derived	Yes	Yes	Yes	Short
Slant Transform	Others	No	Derived	Yes	Yes	Yes	Short
Teager Energy Operator	Others	No	Derived	Yes	Yes	Yes	Short
Spectral Kurtosis	Others	No	Derived	Yes	Yes	Yes	Short
Principal Curves	Others	No	Derived	Yes	Yes	Yes	Short
Sparse Signal Decomposition	Others	No	Derived	Yes	Yes	Yes	Short
Input Variable Selection Techniques							
Variable Ranking	Original Variables	No	Original	Yes	No	Yes	Long
Wrapper Methods	Original Variables	No	Original	Yes	No	Yes	Long
Embedded Methods	Original Variables	No	Original	Yes	Yes	Yes	Long
Filter Methods	Original Variables	No	Original	Yes	No	Yes	Long
Regression	Derivable Variable	No	Derived	No	Yes	No	Long
Clustering	Derivable Variable	No	Derived	No	No	No	Long
Heuristics	Computational	No	Derived	Yes	No	No	Long
Optimization & Simulation	Computational	No	Derived	Yes	No	No	Long
Statistics	Computational	No	Derived	Yes	No	Yes	Long
Data mining	Computational	No	Derived	No	No	Yes	Long
Sensitivity Analysis Techniques							
Differential Analysis	Analytical S.A.	No	Original	Yes	Yes	Yes	Long
Green Function	Analytical S.A.	No	Original	Yes	Yes	Yes	Long
Couple/De-Couple Direct	Analytical S.A.	No	Original	Yes	Yes	Yes	Long
Monte Carlo and Latin Hypercube	Sampling Based S.A	No	Original	Yes	Yes	Yes	Long
ANOVA	Sampling Based S.A	No	Original	Yes	Yes	Yes	Long
Fourier Amplitude Sensitivity Test	Sampling Based S.A	No	Original	Yes	Yes	Yes	Long
Time uncertainty	Sampling Based S.A	No	Original	Yes	Yes	Yes	Long
Entropy Based Epistemic	Sampling Based S.A	No	Original	Yes	Yes	Yes	Long
Event Tracker	Event Based S.A	Min	Original	Yes	Yes	Yes	Short
Event Clustering	Event Based S.A	Min	Original	Yes	Yes	Yes	Short
Classifiers Techniques							
Feed-Forward Neural Networks	Supervised	No(1)	Derived	Yes	Yes	No	Long
Multi-Layer Perceptron (MLPNN)	Supervised	No(1)	Derived	Yes	Yes	No	Long
Decision Tree (DT)	Supervised	No(1)	Derived	Yes	Yes	No	Long
Naïve Bayes (NB)	Supervised	No(1)	Derived	Yes	Yes	No	Long
Radial Basis Function	Supervised	No(1)	Derived	Yes	Yes	No	Long
Hopfield Neural Network (RNN)	Supervised	No(1)	Derived	Yes	Partially	No	Long
Elman Neural Network (RNN)	Supervised	No(1)	Derived	Yes	Partially	No	Long
Jordan Neural Network (RNN)	Supervised	No(1)	Derived	Yes	Partially	No	Long
Self-Organizing Feature Map	Unsupervised	No(1)	Derived	Yes	Yes	Yes	Medium
Boltzmann Machines	Unsupervised	No(1)	Derived	Yes	Yes	Yes	Long
Convolutional Neural Networks	Deep Learning	No(2)	Derived	Yes	Yes	Yes	Medium
Restricted Boltzmann Machine	Deep Learning	No(2)	Derived	Yes	Yes	Yes	Medium
Deep Belief Networks (DBN)	Deep Learning	No(2)	Derived	Yes	Yes	Yes	Medium
Stacked Auto Encoder (SAE)	Deep Learning	No(2)	Derived	Yes	Yes	Yes	Medium
Adaptive Resonance Theory 1	Hybrid	No(1)	Derived	Yes	Yes	Yes	Long
Adaptive Resonance Theory 2	Hybrid	No(1)	Derived	Yes	Yes	Yes	Fast

will suggest a new methodology framework a fast decision to avoid further loss. Therefore, to the author's best knowledge, the event-based analysis is capable of addressing this issue, which will be further discussed in the next chapter.

2.9 Summary

This chapter reviews the existing literature on PQ disturbances, with special emphasis on troubleshooting techniques in both traditional and modern technology. Modern technique is progressing well with simulation data, but there is a gap in the handling real-time industry data which need to sort-out the dimensionality reductions problem. Replying to the key research question about including the environment parameter in the system state analysis, finding a proper research methodology that could deploy the relationship between all input variables with the system KPIs with time constraints and low computational effort is the main objective.

Borrowing from the signal processing methods, the author is looking at adapting the event-based technique with the same approach with the Feature Extraction and Feature Selection framework. The technique which comprises of Event Tracker and Event Clustering, are categorised as the only sensitivity analysis method that competently works with complex systems with regards to heterogeneity and a large number of input variables in real-time [122]. Having said that, a new approach in real-time sensitivity analysis methodology will be introduced in Chapter 3 which will build a map of correlation between system input and output parameters.

With the correlation data in hand, the importance of applying machine learning on the data was discussed based on the literature. The existing ML techniques available in the literature, such as classification, which is to be used in the proposed approach of this research, have shown promising results. Furthermore, the widely used supervised learning such as Multi-Layer Perceptron Neural Network, Decision Tree and Naïve Bayes, which are the focal part of this research are to be explored in this work. Thus, contributing new knowledge in predictive modelling of a large-scale complex system. For proof of concept, an industrial case study which experiencing significant technical and economic benefits will be discussed in later chapters.

Chapter 3

Event Modeller Technique

3.1 Overview

As mentioned in the previous chapter, there are various ways to troubleshoot PQ disturbance problems. The conventional method uses manual configurations and visual inspection to monitor the quality of the power supply in a system. This method was too difficult to interpret and time-consuming. Later, the signal processing method was introduced to automatically classify the PQ problem. Although this technique has been progressing well with various artificial intelligence learning techniques, it doesn't reflect well to a real industry data which is complex and having external influence such as machine utilisation, energy efficiency and environment factor. Having considered this influence factor in the assessment, the system is getting more complex as it is. There is a need to deploy this information in a dynamic platform that connects the embedded system with its Key Performance Indicator (KPI) to a Discrete Event Simulation (DES) running in real-time. This platform will be trained using machine learning tools for classification purpose. The main challenge is to produce a real-time critical accurate knowledge that represents the system's state, at a minimum cost. Two main objectives are (1) to minimise the computational overhead by eliminating the input variables having the least impact and (2) focus on the most significant variables by designing a DES framework that considers all possible factors that could lead to a PQ disturbance problem in a single scan. The gaps in the existing methods reviewed in Chapter 2 have been identified, proposing a novel Event Modeller Data Analytics technique that could illustrate the research method in this thesis.

This chapter is divided into seven (7) sections. Section 3.2 aims to discuss the predictive analysis concept in general. This is followed by section 3.3, which aims to provide an overview of the event modeller technique with the associated predictive learning techniques using event based sensitivity analysis approach. Section 3.4 intends to discuss on how data is accessed and explored. Section 3.5 explains the event modeller technique and how it integrates with KPI transfer function. Section 3.6 discusses predictive model development using various machine learning technique and why the chosen techniques over other techniques are explained. Finally, Section 3.7 provides a summary of the chapter and its place in the thesis.

3.2 Predictive Analytic

Predictive analytics has received a lot of attention in recent years due to advances in supporting technology, particularly in the areas of big data and machine learning [204–206]. Predictive analytics is the branch of the advanced analytics used to make a prediction about unknown future events. It uses a number of data mining, predictive modelling, and analytical techniques to bring together the information from historical data, business process, IT and management team, to make predictions about future and thus optimise it for business benefits [207].

Predictive analytics has been around with different names. The term “data mining” and “knowledge discovery” has been used by commercial and academic consecutively, to describe the processes involved in creating predictive models [208]. The core of predictive analytics relies on capturing relationships between explanatory variables and the predicted variables from past occurrences and exploiting them to predict the unknown outcome. It is important to note that the accuracy and usability of the prediction results will depend significantly on the level of data analysis and the quality of assumptions [209].

As this research is posed to predict the variables that caused machine failures and deal with big data analysis in real-time, a predictive analytics method is chosen to leverage on the information provided by the machine, while gaining new insight for PQ disturbance fundamental research. The goal of this research framework is to integrate the tools and techniques for predictive analytics and data visualisation with a domain-specific modelling environment that makes PQ disturbance problem specification becomes easier. Knowing that big data has a big implication on the predictive model process, it is important to define the term "predictive

model" and "big data analytics" and how it is associated in the following sub-sections.

3.2.1 Predictive models

Predictive models are used to predict behaviour or values for new occurrences. It involves data mining or statistical probability to estimate the future outcome using historical data or transaction data [210]. In general, it extracts data from a massive database and uses this data to predict future events or behaviour with the help of advanced algorithm such as neural network, decision trees, linear regression, logistic regression, ridge regression, time series, ANOVA and support vector machine to find the hidden pattern. The core element of predictive models is the predictor, a variable that can be measured for an individual or an entity to determine various risks and opportunities [211]. For example, a manufacturing plant may consider the machine's KPI such as energy efficiency, machine availability and environment index as a predictor to determine the performance of the machine operations. Once data has been collected for the relevant predictors, a statistical model is formulated. Various applications use a predictive model to predict future event such as manufacturing [212], financial prediction, Internet of Things [213, 214], machinery [215], health and energy management system [206]. The use of this predictive model has steadily increased over the past several years [216]; however, there is still a lack of capitalisation of the data available for it to make accurate decision [217]. There are four steps in predictive models as follows:

1. Descriptive Analytics. Describing the raw data from multiple sources, to give a valuable insight into the past without knowing the reason.
2. Diagnostic Analytics. Drilling down the historical data to find out the dependencies of why it happened.
3. Predictive Analytics. Predicting future trends using the findings from descriptive and diagnostic models by telling what is likely to happen.
4. Prescriptive Analytics. Making a decision to eliminate the future problem by prescribing what action to advise the stakeholder.

Predictive models are often discussed in the context of big engineering data such as data from sensors, instruments, data acquisition, servers and connected systems in a big data

arrangement. This would make the prediction model very challenging in predicting accurate future events. [218] view big data as one of the most escalating IT problems in the next decade due to the enormous demands in different area such as science, industrial, military, medicine, ecology, safety and others. The following subsection will further discuss the challenges in handling big data.

3.2.2 Big Data Analytics

The increase in manufacturing data has made it challenging to conduct process analysis. There is a need to manage this massive amount of data to ensure heterogeneity data is captured for predictive analytics purposes. Many organisations collect data from different sources, which make it very challenging to analyse data using traditional data management which is expensive and time-consuming [219]. Big data refers to massive, high growth and diversified information assets that can bring a lot of valuable information. It involves more than just managing the volume of data [220]. Earlier, most researchers categorised data to the three pillars of words which are volume, variety and velocity. Volume refers to the size or enormous amounts of available data [221]. Variety refers to different types of data such as a vast range of systems and sensors, or formats such as text, images, audio and video. Velocity refers to the speed at which the data is collected and processed, such as real-time data or data stream from different sources [222]. Therefore data with huge volume, high velocity, and great variety can be referred to big data.

Recently, big data are associated with veracity, validity, and volatility [223]. Veracity refers to noise, biases or abnormality in the data. This is to ensure the data is clean by performing pre-processing stage. Validity refers to the quality of the data to be analysed. Valid data is the key in making the right decision, producing a clear and accurate pattern or knowledge while mining the data. Volatility, on the other hand, refers to data optimisation. High volatility data makes the data relevant and available to be analysed. In today's real-time scenario, a massive amount of data is collected continuously, and it is vital to make sure the data selected are relevant and valid, to get high accuracy results. The data is valid if it is converted into useful information [224]. Big data gathered from different sources such as mobile phones, social media feeds, IoT devices, databases, servers, and engineering applications. The data is meaningless unless a scientific algorithm is implemented to extract valuable information from

the voluminous data to make accurate decisions. The process of extraction and analysis using this scientific algorithm is called Big Data Analytics. A term “Big data analytics”, is a set of advanced technologies designed to work with large volumes of heterogeneous data. It uses complex data mining algorithms in high-performance processors [225]. Big Data Analytics is used to process a large number of data sets to uncover hidden patterns, market trends, customer preferences and much other useful information that can be helpful for organisations to make decisions to enhance their business. With Big data analysis, it is possible to process the data very quickly and efficiently, which was not possible using traditional business intelligence solutions.

There is a huge challenge in processing big data as follows:

1. Data sampling issue. Different data type makes it difficult to sample it.
2. Data extraction issue. To take out useful data from overall big data is crucial and challenging.
3. Data structuring issue. The performance of the analytic process may vary due to this data structuring process.
4. Data interpretation issue. The set of data is normally referred to as a group, and it is very challenging to cluster the same groups.
5. Data complexity issue. Some data represent various format and highly complex which affects performance and response time of the analytic process.
6. Data storage issue. Data storage issue is due to a large scale data, as most storage devices are insufficient to store incoming data from various resources, so data storage is becoming a challenge in big data.
7. Infrastructure issue. Infrastructure of an organisation can be a challenge for big data analytics; a security breach of the organisation can directly affect the security and privacy of data in big data analytics.
8. Data privacy issue. Data privacy focuses on the use and processing of individual data, such as the development of policies in place to ensure the proper use of personal information accessed and exchanged by the user. It can have a negative impact on the privacy

of consumers while affecting business growth, sales performance and other competitive disadvantages.

9. Data management issue. Data management is a crucial issue in big data because it can affect performance, response rate and usability of a big data analytics solution.

Recently, cloud computing provides an efficient infrastructure to cater for this complexity [226]. Cloud computing offers access to data storage, processing and analytics on a more scalable, flexible, cost-effective and even secure basis than can be achieved with an on-premises deployment. These characteristics are essential when data volumes are growing exponentially, to ensure storage and processing resources are available as needed, as well as to get added value from that data. However, due to some restricted confidential data, some company doesn't allow the manufacturing plant data to be shared in cloud computing.

It is useful to indicate the types of data involved in manufacturing. There are two types of manufacturing data, known as structured data and unstructured data. Structured data is the most common data using as a pre-defined data format collected from the manufacturing operation such as field sensors, SCADA system, Enterprise Resource Planning (ERP) and Manufacturing Execution System (MES). Unstructured data, on the other hand, are data collected from an external system such as email and document management system. However, it is difficult to extract and analyse process information from unstructured data [227].

As this chapter is intended to propose a predictive modelling technique in handling manufacturing complex big data, there is a need to solve the big data problem using an alternative method which filters the output data into smaller scale, that have an influence to any uncertain input data rather than analysing the entire raw data. Before proceeding further on the proposed method, it is important to highlight the general workflow of a predictive analytics technique, to observe how the proposed method fits and to build a predictive model that could help system engineer solve complex plant data.

3.2.3 Predictive Analytics Workflow

Figure 3.1 illustrates a typical workflow of a predictive analytics process which will be discussed further in the following steps:

1. Step 1: This stage involved importing data from various sources, such as real-time

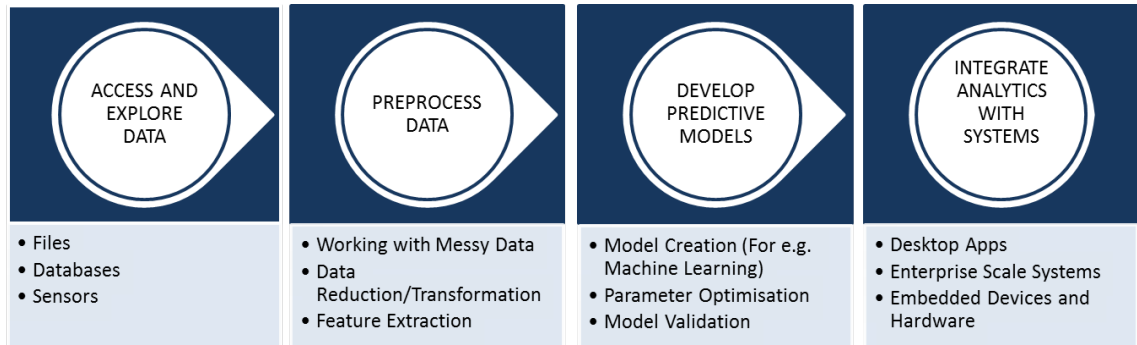


Figure 3.1: Predictive Analytic Workflow

SCADA system, databases, web archives and spreadsheets. The sources contain manufacturing structured data which has been converted into a formatted file (for e.g. .txt or .csv) for the pre-processing stage.

- Step 2: The data access in the previous stage is explored, extracted and selected. Different data sources are combined and cleaned by removing outliers. A good, clean source of data is important to get a good chance of pattern or relationships. This stage is equivalent to the signal processing technique and sensitivity analysis technique reviewed in Chapter 2.
- Step 3: Develop an accurate predictive model based on the aggregated data using statistics, curve fitting tools, or machine learning to predict an event or outcome. This stage involved splitting the data into two parts, one is used to create an analytic model and the remaining is used to test and validate the training model. The complexity of this stage is highly dependent on how the data is pre-processed in the previous stage.
- Step 4: The final step in the predictive model workflow is to integrate the model into the main production system. Using desktop application or built-in SCADA system, the model could forecast an outcome at some future state or time, based on changes to the real-time data inputs to automate decisions and business processes.

Before introducing the Event Modeller Data Analytics technique in the next section, it will be necessary to note that the proposed technique only cover the first three steps of this workflow. The final step is reserved for future works when the predictive model is tested and validated. This is to ensure the most effective model is chosen to warrant accurate prediction.

3.3 Event Modeller Data Analytics Technique

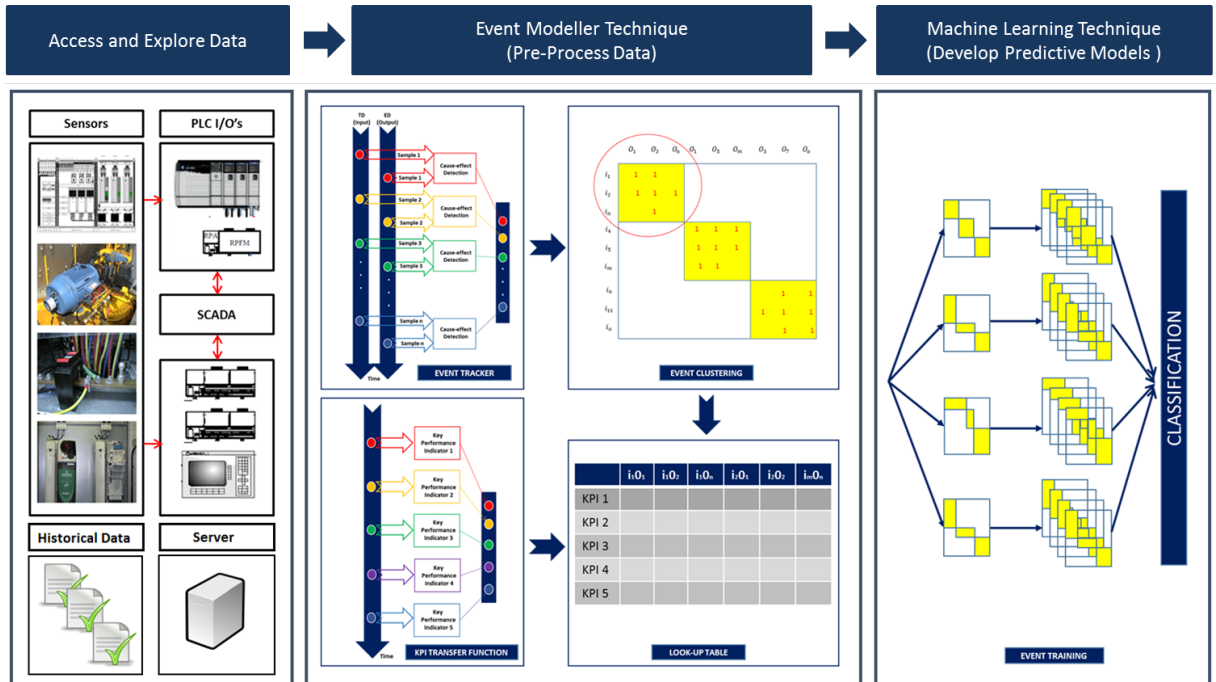


Figure 3.2: Event Modeller Data Analytics Technique Overview

Sensitivity Analysis is an important technique in measuring the degree of influence that an independent variable has on its uncertainty inputs [228]. It has the potential to minimise the computational overhead by eliminating the input variables that have the least impact on the system [126, 229–233]. However, this technique is found to be very challenging in finding a true representation, especially when there is uncertainty in the input-output variables [234]. The competency of this sensitivity analysis was reviewed in Chapter 2 and understood that most of the technique is computationally constraint, time-consuming and highly dependent on historical data. [235] recommends a model-free solution which later [236] introduced Event Tracker to solve the sensitivity analysis problem in high complexity systems. It is capable of capturing the cause-effect relationships between triggers (input variables) and events (output variables) within a specified period of time.

In most recent studies, [233] has introduced EventiC to solve the sensitivity analysis problem more intelligently, by removing all logical boundaries of isolation that exist in complex systems with the principle that every acquirable knowledge affects the output unless proven otherwise. The motivation for this research is to provide a solution for filling the gap between

the existing sensitivity analysis techniques with the machine learning techniques, contributing new knowledge in predictive modelling in a large-scale complex system. The proposed Event Modeller Data Analytic technique built a map of correlation between system input and output parameters while predicting the system fault based on the machine performance metrics. Having reflected to the predictive modelling workflow discussed in Section 3.2.3, the proposed Event Modeller Data Analytics technique is divided into three steps, as illustrated in Figure 3.2. The main steps are (1) accessing the plant data, (2) pre-processing the data and (3) developing predictive models based on the data. Details of each step will be further discussed in the following sections.

3.4 Stage 1: Access and Explore Data

This is the most important step in any research or assessment in predictive modelling technique. It is debated to find the best and accurate data fit to represent the system state. Most industries such as power plant, manufacturing, automotive, and aerospace reveal that the process of evaluating accurate models in real-time is expensive and an extremely time-consuming process [233]. In addition, dealing with big data as explained in Section 3.2.2 makes it more challenging. Engineering data emanated from the control system (i.e. electromechanical device, electrical drives, transmitter, sensors, motors, actuators, alarm) proliferates via the industrial Supervisory Control and Data Acquisition (SCADA) system. An efficient sensitivity analysis technique is required to sort and reduce this big data into a smaller scale, which helps data to be trained in a faster and less computational environment. Figure 3.3 illustrates a common flow in accessing data in industrial system architecture. In this development, data were extracted in two main modes known as real-time data and historical data. The real-time data runs simultaneously with the SCADA system, providing direct access to the system via the Open Platform Communication (OPC) server. For e.g., a level transmitter with a specific tag_name updated the bunker level measurement in both local network (SCADA System) and system development (Event Modeller) via the OPC server connection. The main advantage of having this mode is that the system could predict system failure in real-time. Alternatively, the historical data logged in the data acquisition server is converted into a .csv file format, an excel file which could be easily accessed by both development software and system

engineer. The main advantage of having the historical data mode is that whenever there is a communication problem during operation, it could still run in offline mode. However, the main challenge is to ensure all input-output variables in the main controller are linked to the system development with the same functionalities. Errors in linking the tag_name on either side will cost error to the system analysis, which mislead the prediction process. To avoid this error, it is recommended to use the same tag_name for both sides, and thorough connectivity test has to be in place during the development period.

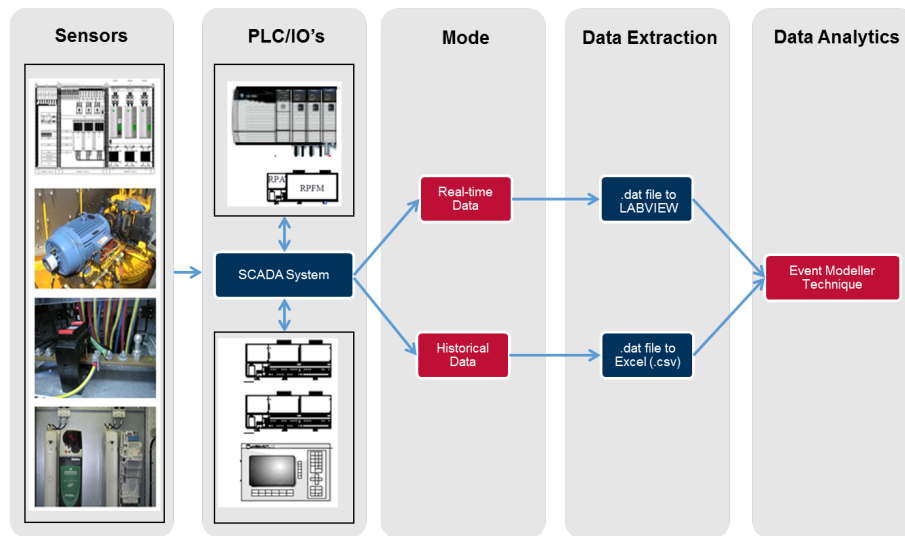


Figure 3.3: Step 1: Access and Explore Data

3.5 Stage 2: Event Modeller Technique

The event modeller technique is designed to evaluate the relationship between the actual events (Output Data) to the cause of the triggered events (Input Data) using a data mapping concept. It groups the high correlation system parameters in the form of matrices and places them into mutually exclusive blocks. This creates an input-output relationship, which is considered both internal and external factors. One significant difference between the proposed event modeller and other traditional data modelling techniques is how the input data make an assumption on its output data. The traditional method assumes the input-output relationship as a true representation of a known data series, while the event modeller technique makes no assumption about it. This makes the proposed event modeller technique become neutral in making a judgement of the input-output relationship. Figure 3.4 illustrates the event modeller

techniques which comprises of Event Tracker, Event Clustering and a Look-up table.

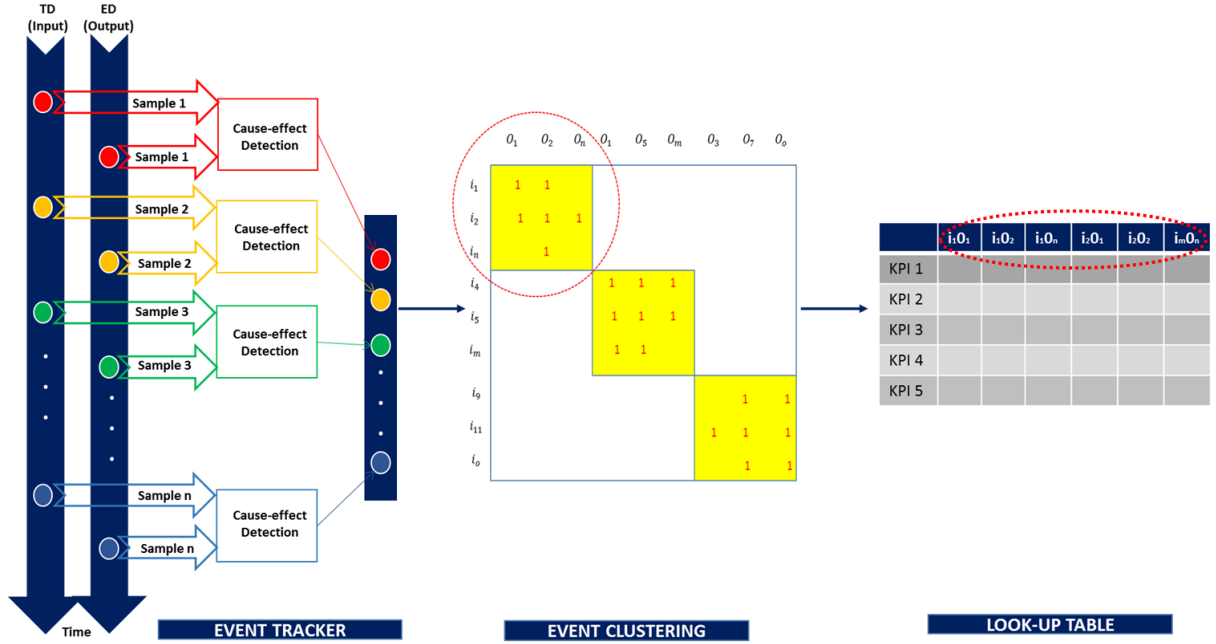


Figure 3.4: Event Modeller Techniques

Industrial data are complex. In this system development, event tracker [126] use cause-effect detection to dynamically track the complexity of each triggered-event relationship. It generates a sensitivity index which measures the impact of the relationship between the triggered data and the event data pairs. High sensitivity index score will be indicated while the lesser impact relationship will be eliminated. This will reduce the computational effort while achieving dimensional reduction. The Event Tracker is a computationally efficient technique that focuses on the state changes of the involved system components. It merely takes a snapshot of the system states, which helps engineers in observing system performance [237].

Event Clustering is designed to improve the real-time sensitivity analysis, by automatically re-arranging the input-output relationship in rank order of its importance and relevance. This technique interprets the changes in the value of input-output data at the given level, detecting the coincidence and finally groups it as a related event. The process of calculating the number of coincidence occurs at a specified scan rate time interval, to ensure a relationship weight is established for modelling and control purposes [233]. Despite filtering the unimportant relationship occurs between the input-output relationship, event clustering has the potential to identify new influencing parameters that were previously thought irrelevant, which make it

unique and interesting to improve the data quality.

A look-up table is an array of data to map input values to output values. It is used to transform the input data into a more desirable output format. The lookup table allows replacing run-time computation or logic circuitry with a simpler array indexing operation [238]. Retrieving a value from a look-up table is faster rather than examining it from the whole database. With the advantage of the logical separation of data, it makes it relevant to prepare the data for machine learning purposes. In this system development, the look-up table is used to match the outcome of the event clustering with the industrial key performance indicator expressed within the system state operation.

3.5.1 Event Modeller Basic Parameter

Before proceeding further into the event modeller technique, it is vital to explain the basic parameters of the event modeller technique, which are borrowed from [126]. Event modeller defines an input and output occurrence matrix [+ -] at pre-specific time intervals. This matrix subsequently describes the relationships between causes that trigger events (trigger data) and the actual events (event data), enabling the construction of a discrete event framework for sensitivity analysis. A short description of discrete event systems, together with the definitions and impact of trigger data and event data, are summarised in Table 3.1.

Table 3.1: Summary of Basic Description for Event-Modeller

Description	Definition	Impact
Discrete Event System	The disparate occurrence of events in a specified time span.	Any changes in the input/output will change the system state.
Triggered Data (TD)	Any input variable whose value transition registered an event.	Any Individual or combination of input variables may have different effects on different system outputs ED = {TD1, TD2,..., TDn}
Event Data (ED)	The series of data that represent the state of the system at a given time	Any Individual or combination of input variables may have different effects on different system outputs ED = {TD1, TD2,..., TDn}
Triggered Threshold (TT)	A given numerical value set point that the values of TD series based on experts/historical data	The fluctuations in the TD series that are interpreted as triggers are determined in comparison with the TT.
Event Threshold (ET)	A proportion or percentage of an overall range of values of TD series over the time scale	The fluctuations in the ED series that are interpreted as triggers are determined in comparison with the ET
Search Slot (SS)	A fixed time slot within which batches of TD and ED are captured	The fixed time slot(scan rate) is determined by experts.
Analysis Span (AS)	The time span within which a period of sensitivity analysis occurs.	Comprised of a number of consecutive SS.

3.5.2 Exclusive-NOR Functionality

Having defined the Event Modeller basic parameter in the previous section, it is worth noting the Exclusive-NOR logic that is widely used in the Event Modeller technique. The output value of this logic can be set to 0 and 1 depending on the state of the input/output. If both or none of the input/output event data is triggered, the value is 1; otherwise, it is 0. This logic made Event Modeller a unique technique, which eliminates all the logical boundary restrictions present in a complex system, with the principle that any acquired knowledge or data (input) influences the output unless it is not demonstrated. In addition, the technique is not only capable of filtering unwanted data but is capable of including information that was thought to be irrelevant to the system. Table 3.2 present the table truth for an Exclusive-NOR logic.

Table 3.2: Exclusive-NOR Functionality

Input 1	Input 2	Output
0	0	1
0	1	0
1	0	0
1	1	1

3.5.3 Event Modeller Basic Assumption

There are a number of assumptions that need to be made before implementing the Event Modeller. These assumptions are borrowed from [239] and can be listed as:

3.5.3.1 Assumption 1: Delays

The delay between EDs and the respective TDs is negligible, and all TDs lead to a specific ED (for all intent and purposes instantaneous).

3.5.3.2 Assumption 2: Thresholds

The triggers and event thresholds are a pre-specified range of signal fluctuation for every data series and determined the system expert, which remains fixed within sampling time. Thresholds are usually based on a percentage of a signal's real value which has to be assumed as an event. For example, a signal with a value of 100 units and a 1% threshold is detected as an event if its value exceeds 101 or decreases 99 units at the next analysis span (sample).

3.5.3.3 Assumption 3: Homogeneity of the data series

The event data series is assumed to be stationary covariance (mean and variance remain constant during the analysis span).

3.5.3.4 Trigger-Event detection

Equation 3.1 and 3.2 indicate the relationship between each event triggered by input at t and $t - 1$ with respect to changes in output. Each change to the output in a given time span can be expressed as an event and the positive value of the inputs as triggers, thus output can be defined as Event Data (ED). Both $Input_t$ and $Input_{t-1}$ can be considered as Triggered Data (TD).

$$if(Input_t - Input_{t-1}) \geq \Theta \xrightarrow{Trigger} TD_t \quad (3.1)$$

$$if(Output_t - Output_{t-1}) \geq \Psi \xrightarrow{Event} ED_t \quad (3.2)$$

3.5.3.5 Sample scan size

The system expert defines the sample size (i.e. the number of samples that construct the incident matrix). The data series does not have a maximum sample size, but commonly 250 samples are taken as the minimum sample size. The data is then transferred to the EventiC algorithm to construct the incidence matrix.

3.5.3.6 Average sensitivity analysis weight

In order to find the average SA weight, all diagonal ROC matrices for complete sample numbers must be applied. The normalised weight of each input variable is the output coefficient of the system.

3.5.3.7 Event Modeller Limit

Event Modeller Limit (EML) or also known as Cut-Off Threshold is a mechanism to deduct the less important input variables and is in the range $0 \leq EML \leq 1$ [126]. For example, when $EML = 0.95$, all inputs with an average SA weight of less than 0.95 or 95% are deducted.

3.5.3.8 False negative test

A false-negative test is carried out in addition to a cut-off threshold, to ensure the inputs are not unnecessarily discounted. In order to confirm these eliminations, a false negative test must be performed by a system expert analyst.

3.5.4 Rank Order Clustering

The Rank Order Clustering (ROC) technique proposed by King (1980) [127] uses matrix manipulation techniques to rearrange the rows and columns of the matrix in an iterative manner. The approach ultimately results in a matrix structure, in which both rows and columns are arranged in order of decreasing value in a finite number of steps. The determination of clusters of occurrence in block diagonal format is an efficient algorithm. This method is limited as it is based on the premise that data classes will be highly similar and grouped into blocks which are mutually exclusive. In the cluster analysis method, a group of data values are ‘similar’ according to a ‘similarity criteria’. They can either be replaced by a new value representing the group (clumping) or assigned a unique type of label (partitioning) [128, 240].

The Event Modeller technique embraced the ROC technique, to establish a cause-effect grouping of system inputs (originating from sensors/actuator’s) and outputs (system performance indicators).

3.5.4.1 Rank Order Clustering Algorithm

Turning now to the ROC algorithm, step-by-step implementation of the ROC method is explained as follows.

1. Step 1: Populate machine-part incident matrix (MPIM), where elements are presented as “0” or “1”. Zero (0) indicates no operation and one (1) indicates an active operation. Parts are arranged in columns and machines are in rows.
2. Step 2: A weight for each row i and column j (in a m by n matrix) are calculated using Equation 3.3 [127].

$$Row_i : W_i = \sum_{j=1}^m a_{ij} 2^{m-j} \quad (3.3)$$


$$Column_j : W_j = \sum_{i=1}^m a_{ij} 2^{n-i} \quad (3.4)$$

Input	Out1	Out2	Out3	Out4	Out5
In1	0	1	1	0	1
In2	0	1	1	0	1
In3	0	1	1	0	1
In4	1	0	0	1	0
In5	0	1	1	0	1
In6	1	0	0	1	0
In7	1	0	0	1	0
In8	0	1	1	0	1

Figure 3.5: Machine Part Incidence Matrix

- Step 3: Read the series of 1 and 0s from left to right in the matrix as a binary number 2^i (0 to $i - 1$ number of rows). Rank the rows in the order of decreasing values. In the case of a tie, rank the rows in the same order as they appear in the current matrix.
- Step 4: Numbering from top to bottom, is the new order of rows the same as the rank order determined in the previous step.
- Step 5: Reorder the rows in the part-machine incidence matrix by listing them in decreasing rank order.

Input	Out1	Out2	Out3	Out4	Out5	Decimal Value	Rank
In1	0	1	1	0	1	13	4
In2	0	1	1	0	1	13	5
In3	0	1	1	0	1	13	6
In4	1	0	0	1	0	18	1
In5	0	1	1	0	1	13	7
In6	1	0	0	1	0	18	2
In7	1	0	0	1	0	18	3
In8	0	1	1	0	1	13	8




Input	Out1	Out2	Out3	Out4	Out5	Decimal Value	Rank
In4	1	0	0	1	0	18	1
In6	1	0	0	1	0	18	2
In7	1	0	0	1	0	18	3
In1	0	1	1	0	1	13	4
In2	0	1	1	0	1	13	5
In3	0	1	1	0	1	13	6
In5	0	1	1	0	1	13	7
In8	0	1	1	0	1	13	8

Figure 3.6: Incident Matrix Row Ranking

- Step 6: In each column of the matrix, read the series of 1s and 0s from the top to the bottom of the binary number 2^j (0 to $j - 1$ number of rows). Rank the columns in order of decreasing value. In the case of a tie, rank the columns in the same order as they appear in the current matrix.
- Step 7: Numbering from left to right, is the current order of columns the same as the rank order determined in the previous step.
- Step 8: Reorder the columns in the part-machine incidence matrix by listing them in decreasing rank order, starting with the left column.

Input	Out1	Out2	Out3	Out4	Out5
In4	1	0	0	1	0
In6	1	0	0	1	0
In7	1	0	0	1	0
In1	0	1	1	0	1
In2	0	1	1	0	1
In3	0	1	1	0	1
In5	0	1	1	0	1
In8	0	1	1	0	1
Decimal Value	224	31	31	224	31
Rank	1	3	4	2	5



Input	Out1	Out4	Out2	Out3	Out5
In4	1	1	0	0	0
In6	1	1	0	0	0
In7	1	1	0	0	0
In1	0	0	1	1	1
In2	0	0	1	1	1
In3	0	0	1	1	1
In5	0	0	1	1	1
In8	0	0	1	1	1
Decimal Value	224	224	31	31	31
Rank	1	2	3	4	5

Figure 3.7: Incident Matrix Column Ranking

3.5.5 Event Modeller Algorithm

The existing Event Modeller technique proposed by [233] is improved for optimisation purposes. The following are the step by step implementation of the algorithm.

Step 1: Set Event Modeller Limit (EML) in %.

Step 2: Set Threshold Setting (Th) in %.

Step 3: Assign Upper Limit (UL_{Th}) and Lower Limit (LL_{Th}) with the following equation:

$$UL_{Th} = \frac{100 + Th}{100} \quad (3.5)$$

$$LL_{Th} = \frac{100 - Th}{100} \quad (3.6)$$

Step 4: Populate First Triggered Data (TD_1)

Step 5: Multiply First Triggered data with Upper Limit,

$$TD_{1UL} = TD_1 \cdot UL_{Th} \quad (3.7)$$

Step 6: Multiply First Triggered data with Lower Limit,

$$TD_{1LL} = TD_1 \cdot LL_{Th} \quad (3.8)$$

Step 7: Populate Next Triggered Data (TD_2).

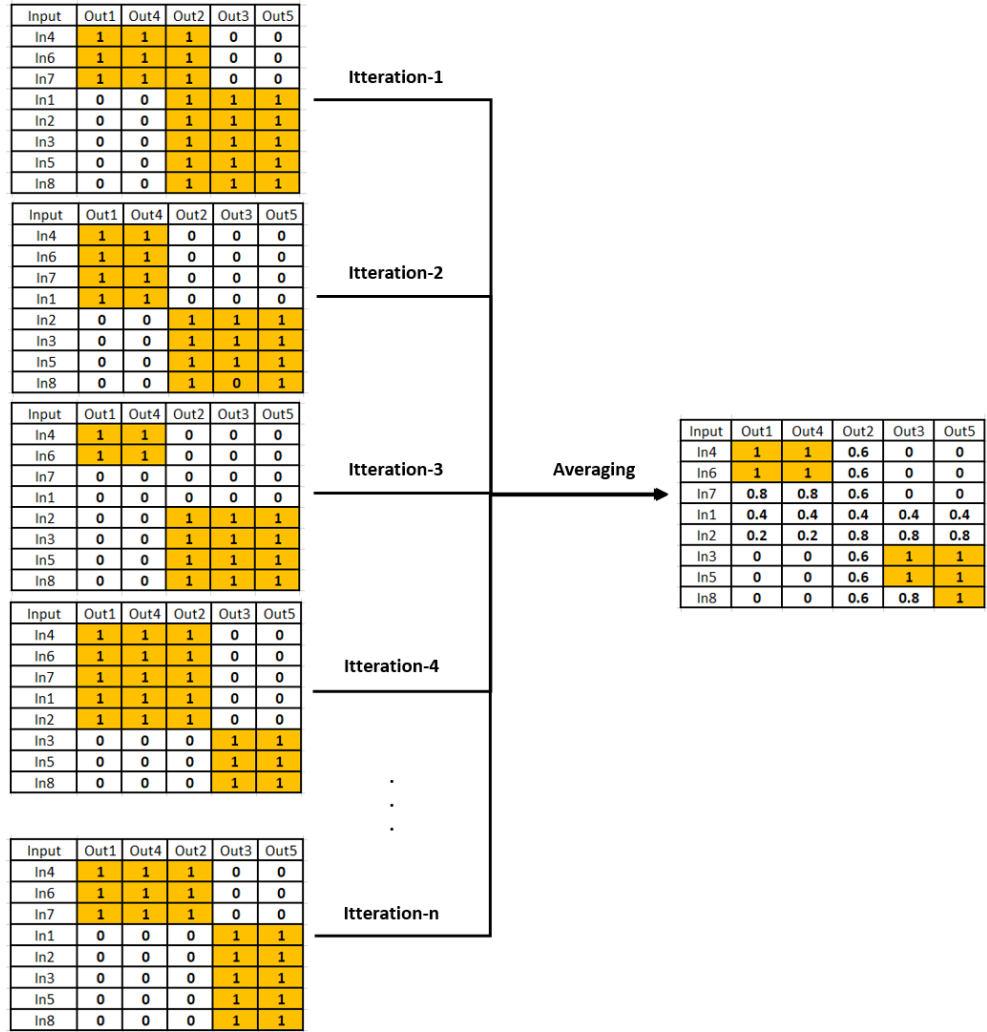


Figure 3.8: Combining all Incidents Matrix

Step 8: Evaluate the different between First Triggered Data (TD_1) and Next Triggered Data (TD_2):

$$TD_{diff} = TD_1 - TD_2 \quad (3.9)$$

Step 9: Determine the final output for the first pair (TD_{Output}):

$$\begin{aligned}
 & \text{if } (TD_{diff} > TD_{1UL}) \Rightarrow TD_{Output} = 1 \\
 & \text{else if } (TD_{diff} < TD_{1LL}) \Rightarrow TD_{Output} = 1 \\
 & \text{else } \Rightarrow TD_{Output} = 0
 \end{aligned} \quad (3.10)$$

Step 10: Populate First Event Data (ED_1).

Step 11: Get First Event Data Benchmark (ED_{1B}).

Step 12: Get First Event Data Setting (ED_{1S}).

Step 13: Assign First Event Data Upper Limit (ED_{1UL} and Lower Limit (ED_{1LL} with the following equation:

$$ED_{1UL} = ED_{1S} + ED_{1B} \quad (3.11)$$

$$ED_{1LL} = ED_{1S} - ED_{1B} \quad (3.12)$$

Step 14: Populate Next Event Data (ED_2).

Step 15: Evaluate the different between First Event Data (ED_1) and Next Triggered Data (ED_2):

$$ED_{diff} = ED_1 - ED_2 \quad (3.13)$$

Step 16: Determine the final output for the first pair (ED_{Output}):

$$\begin{aligned} & \text{if } (ED_{diff} > ED_{1UL}) \Rightarrow ED_{Output} = 1 \\ & \text{else if } (TD_{diff} < ED_{1UL}) \Rightarrow ED_{Output} = 1 \\ & \text{else } \Rightarrow TD_{Output} = 0 \end{aligned} \quad (3.14)$$

Step 17: Repeat Step 4 to Step 9 for Next set of Triggered Data (Input).

Step 18: Repeat Step 10 to Step 16 for the Next set of Event Data (Output).

Step 19: Repeat Step 17 to Step 18 to complete first batch of Input-Output data.

Step 20: Populate the first batch of input-output event coincidence matrix with binary weighting values of exclusive NOR function.

Step 21: Populate the second batch of input-output event coincidence matrix with binary weighting values of exclusive NOR function.

Step 22: Average each input-output event coincidence.

Step 23: Sort rows of the resultant binary matrix into decreasing order of their decimal weights.

Step 24: Repeat steps 20 to 23 for every column.

Step 25: Repeat steps 20 to 24 until the position of each element in each row and column does not change.

Step 26: A weight for each row i and column j (in a m by n matrix) is calculated using Equation 3.3 for row and Equation 3.4 for column.

3.5.6 Key Performance Indicator

Modern manufacturing process produces huge data within the control system. This data will be meaningful if it is translated into a performance measure. Metrics used to characterise such performance measures are referred to as Key Performance Indicators (KPIs). The use of existing shop-floor to measure and monitor industrial KPIs has been a trend. [241] used OLE which integrates with DES modelling capabilities to measure the KPIs for a brewery industry. [242] used data-driven scheme of KPIs prediction and diagnosis for hot strip mill industry. [243] proposes an analytics solution for calculating statistical KPIs in the Human Machine Interface (HMI) layer. Some HMI products have built in libraries which allow Standard KPI calculation but it is not suitable for all types of industry. Therefore, there is a need to translate a suitable KPI that suit the type of operation. Basic KPIs are calculated directly from the output operation data, and they serve as the foundation for Overall performance KPIs [244].

The reliability of a machine is indirectly proportional to the period it is working by going through all phases of product life cycle management. The first stage is the early failure period, where no preventive maintenance policy is needed, and the failure is corrected on an occurrence basis. The next phase is the useful life where failure occurs randomly, and the Maintenance Policy is determined by the equipment's operating condition. The final stage is the wear out where a preventive maintenance policy for a fixed time is needed as the probability of a malfunctioning item will increase over time. A bath-tub curve in Figure 3.9 presents the three-phase of this reliability period [1].

3.5.6.1 Time-based Key Performance Indicator

Time-based KPIs are data related to time duration, defining activities associated with production and maintenance. The general viewpoints for measuring time base KPIs are machine, op-

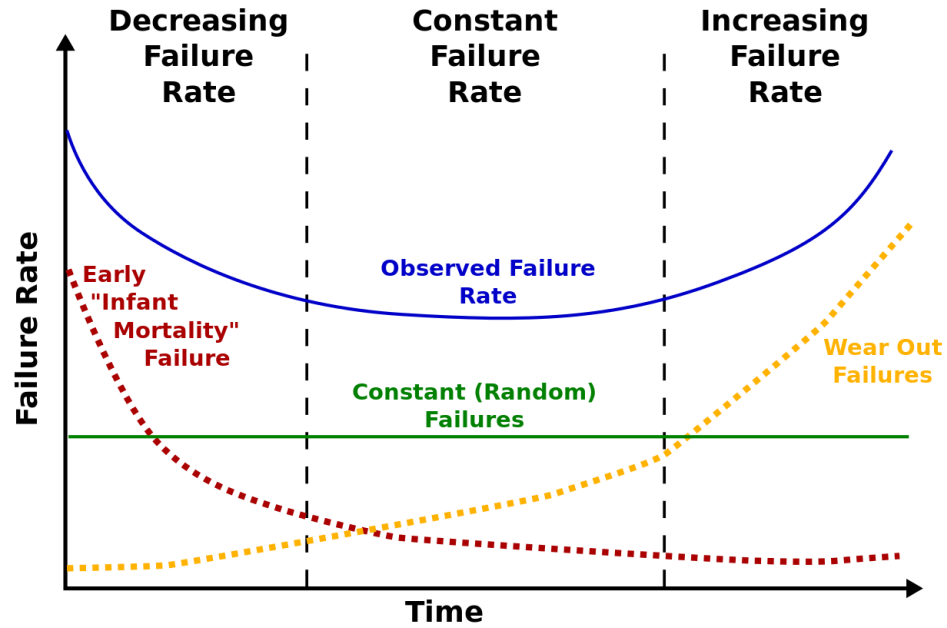


Figure 3.9: Reliability Bath-tub Curve Image: courtesy of [1]

erator or order production. However, in practice, actual production time differs from planned production time. This is because actual production time incorporate machine breakdowns, shift changeover and quality deterioration. Thus, the calculation of KPI has to consider all these factors to reflect accurate metrics. Hence it is important to understand the KPI inter-relationship. A hierarchical structure for KPI categorisation proposed by [244] is shown in Figure 3.10.

The time-based KPIs studied in this project are Availability, Instantaneous Utilisation, Schedule Utilisation and Performance. Considering a piece of equipment operating life, the time probability until a breakdown or dysfunction occurs is known as Mean Time Between Failures (MTBF) that is directly proportional to the number of years a machine has been functioning as determined by the following equation:

$$MTBF = \frac{\text{Operating Life of a Number of Items}}{\text{Total Number of Failures}} \quad (3.15)$$

Similarly, the time probability of failed system, under maintenance proceedings, is restored to operable conditions within the downtime period is known as Mean Time To Repair (MTTR) which is a measure of Maintainability. MTTR is directly proportional to the Total Repair

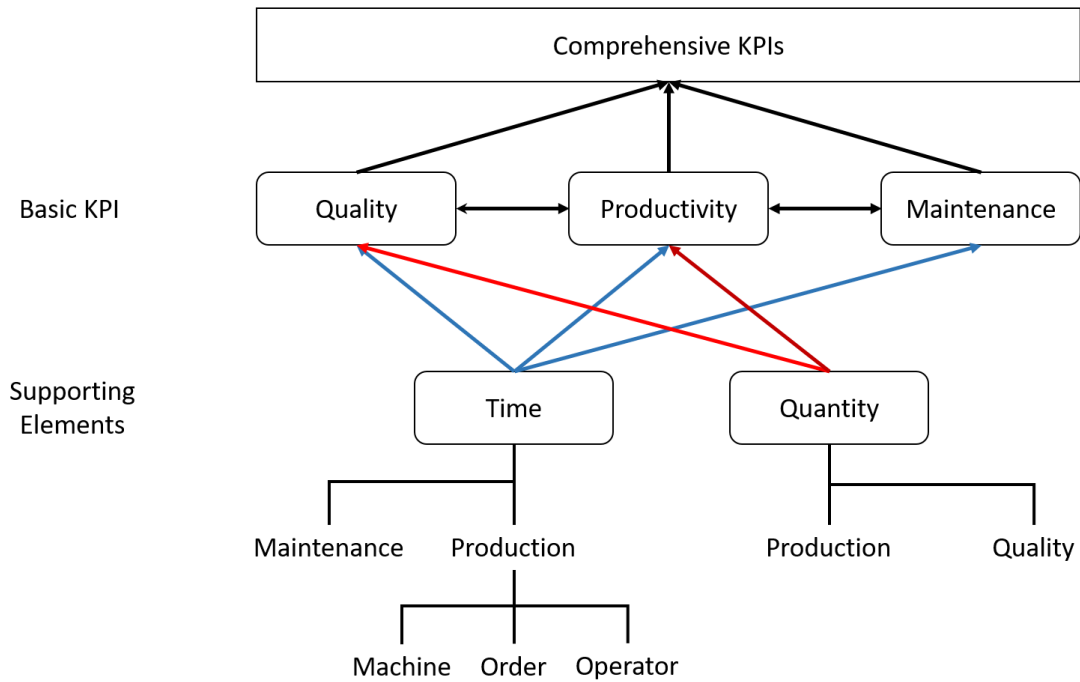


Figure 3.10: KPI Categorisation

Time defining by the following equation:

$$MTTR = \frac{\text{Total Repair time of a Number of items}}{\text{Total Number of Failures}} \quad (3.16)$$

Availability is defined as the relationship between Maintainability and Reliability by the following Equation:

$$\text{Availability} = \frac{MTBF}{MTBF + MTTR} \quad (3.17)$$

Alternatively, availability could also be calculated as the ratio of Run Time to Planned Production Time. It takes into account all events that halt planned production long enough where it makes sense to track a reason for being down (typically several minutes).

$$\text{Availability} = \frac{\text{Run Time}}{\text{Planned Production Time}} \quad (3.18)$$

Run Time is simply Planned Production Time less Stop Time, where Stop Time is defined as total time where the manufacturing process was intended to be running but was not met due

to Unplanned Stops (e.g., Breakdowns) or Planned Stops (e.g., Changeovers). [245]

$$\textit{Run Time} = \textit{Planned Production Time} - \textit{Stop Time} \quad (3.19)$$

The Instantaneous Utilisation is defined by the following Equation:

$$\textit{Instantaneous Utilisation} = \frac{\textit{Busy Time}}{\textit{Available Time}} \quad (3.20)$$

This indicator accounts statistics on a machine's usage at discrete intervals in time by pondering the Utilisation as a function of time. This metric considers all resources in the manufacturing system including all non-scheduled periods a workstation was not planned for production. On the other hand, Schedule utilisation reports the cumulative average utilisation over the period that the resource was scheduled in the system [246].

$$\textit{Schedule Utilisation} = \frac{\textit{Busy Time}}{\textit{Planned Production Time}} \quad (3.21)$$

Performance takes into account anything that causes the manufacturing process to run at less than the maximum possible speed (including both Slow Cycles and Small Stops). Performance is the ratio of Net Run Time to Run Time. It is calculated as:

$$\textit{Performance} = \frac{\textit{Ideal Cycle Time} \times \textit{Total Count}}{\textit{Run Time}} \times 100\% \quad (3.22)$$

Ideal Cycle Time is the fastest cycle time that your process can achieve in optimal circumstance. Therefore, when it is multiplied by Total Count the result is Net Run Time (the fastest possible time to manufacture the parts). Since rate is the reciprocal of time, Performance can also be calculated as:

$$\textit{Performance} = \frac{\frac{\textit{Total Count}}{\textit{Run Time}}}{\textit{Ideal Run Rate}} \times 100\% \quad (3.23)$$

Quality takes into account parts made that do not meet quality standards, including parts that need to be reworked. OEE Quality is similar to First Pass Yield, in that it defines Good Parts as parts that successfully pass through the manufacturing process for the first time without any rework. Quality is calculated as:

$$Quality = \frac{Good\ Count}{Total\ Count} \times 100\% \quad (3.24)$$

OEE takes into account all losses, resulting in a measure of truly productive manufacturing time. It is calculated as:

$$OEE = Availability \times Performance \times Quality \quad (3.25)$$

If the equations for Availability, Performance, and Quality are substituted in the above and reduced to their simplest terms the result is:

$$OEE = \frac{Good\ Count \times Ideal\ Cycle\ Times}{Planned\ Production\ Time} \quad (3.26)$$

3.5.6.2 Energy and Emission Based Key Performance Indicator

To complement the Time-Based KPIs discussed earlier, it is useful to look at the energy consumption and emission contribution that could potentially harm the environment. In general, energy consumption can be calculated by multiplying the motor rating (kW) with the duration it operates. There are three states of motor known as run, idle and stop. During running state, the motor is capable to operate at its 100% loading while in idle state, the motor operates at 25% of its loading. Obviously, there is no loading when it stops.

$$Busy\ EC = \frac{Busy\ Time}{Motor\ Rating \times No\ of\ Motors \times 3600} \quad (3.27)$$

$$Idle\ EC = \frac{Idle\ Time}{Motor\ Rating \times No\ of\ Motors \times 3600 \times 0.25} \quad (3.28)$$

Thus, to calculate Total Energy Consumption, both busy and idle has to be added.

$$Total\ EC = Busy\ EC + Idle\ EC \quad (3.29)$$

Having the total energy consumption, it is easier to calculate the carbon footprint which consist of Carbon Dioxide Emissions ($kgCO_2$), Methane gases ($kgCH_4$), Nitrous oxide gases ($kgNO_2$) and Carbon Dioxide Equivalent ($kgCO_2e$), courtesy of UK Greenhouse Gas Reporting: Conversion Factors 2017, [247] which will be discussed in Equation 3.30 to Equation 3.33

as follows:

$$\text{Carbon Dioxide Gases (CO}_2\text{)} = \text{Total EC} \times 0.448581 \quad (3.30)$$

$$\text{Methane Gases (CH}_4\text{)} = \text{Total EC} \times 0.000412 \quad (3.31)$$

$$\text{Nitrous Oxide Gases (NO}_2\text{)} = \text{Total EC} \times 0.002339 \quad (3.32)$$

$$\text{Carbon Dioxide equivalent Gases (NO}_2\text{)} = \text{Total EC} \times 0.451331 \quad (3.33)$$

As a key performance indicator, the energy consumption ratio can be calculated with the following equation:

$$\text{Energy Consumption Ratio} = \frac{\text{Total EC}}{\text{Total Production}} \quad (3.34)$$

$$\text{Green House Gases Ratio} = \frac{\text{kgCO}_2\text{e}}{\text{Total Production}} \quad (3.35)$$

3.5.6.3 Event Modeller Key Performance Indicator Transfer Function

Considering the time-based KPI, energy-based KPI and environment-based KPI in the previous subsection, it is essential to integrate the Event Modeller technique with these KPIs for data analytics purposes. In EMDA, KPI plays the role of a human, monitors the manufacturing line in real-time. All machine activity such as in-service, breakdowns, and failure does impact the KPI metrics. Having tailored it with the event modeller input-output relationship, rational information from the system state could be generated in the form of look-up table as shown in Figure 3.11. Accessing both data from KPI and Event Modeller Technique in a single look-up table platform, the machine learning algorithm could retrieve this packeted data in predicting future events. At this proof of concept stage, both are running independently. For future work, if both systems are tested to be compatible, it could be run simultaneously for a real-time predictive model. So far this section has discussed on the second step of the

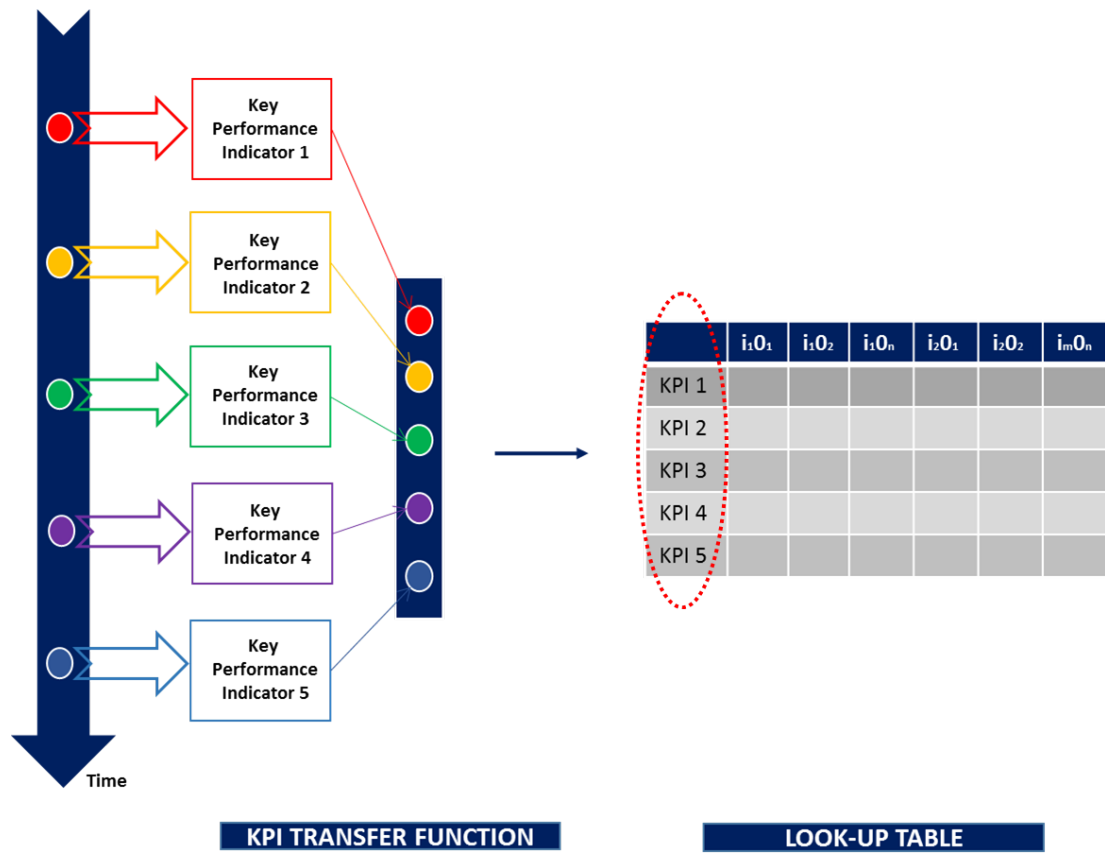


Figure 3.11: Automatic Translation of KPI Transfer Function

proposed EMDA technique which focused on the Event Modeller, KPI and the integration of the Event Modeller and KPI. The following section will discuss the third step of the proposed EMDA technique, further elaborating the machine learning model that has been selected to be used in this research.

3.6 Stage 3: Develop Predictive Models

Predictive modelling incorporates mathematical techniques, machine learning and data mining to explore knowledge in order to make predictions about unknown future events. A mathematical approach uses an equation-based model that describes the phenomenon under consideration while machine learning and data mining uses a "black box" approach, which requires an extensive simulation effort to train model input from the output to make a prediction. Some of the applications include predictive maintenance [248], fault diagnosis [249], control system [250], and quality control [251]. Predictive modelling is often performed using

curve and surface fitting, time series regression, or machine learning approaches. Regardless of the approach used, the process of creating a predictive model is the same across methods. The steps are summarized in the following Table 3.3.

Table 3.3: Summary of Predictive Model Steps

Steps	Description
Step 1	Clean the data, to eliminate outliers and treating missing data.
Step 2	Identify a parametric or non-parametric predictive modelling approach to use.
Step 3	Pre-process the data into a suitable form for the chosen modelling algorithm.
Step 4	Specify a subset of the data to be used for training the model.
Step 5	Train, or estimate, model parameters from the training data set.
Step 6	Conduct model performance or goodness-of-fit tests to check model adequacy.
Step 7	Validate (Testing) predictive modelling accuracy on data not used for calibrating the model.
Step 8	Use the model for prediction if satisfied with its performance.

As explained in the introduction of this Chapter, the objective of this research is to deploy machine learning platforms that could predict the variables causing machine failure. Machine failure could be due to different reason, such as climate change, device failure, system interruptions, or operator mishandling. Thus a suitable supervised learning model has to be decided. In supervised learning, the training sets are composed of known input data and response values. The training sets feed the machine learning system that tries to generalise the function involved in mapping new data sets [217]. Then it follows up with testing sets, to validate the model accuracy. There are numerous applicable classification methods, such as Support Vector Machine, Logistic Regression, Random Forest, Naïve Bayes, Decision Tree and artificial neural networks. However, a choice of classifiers is crucial in order to provide a meaningful solution to the problem. In terms of precision and speed, each model has its own advantages and disadvantages.

Figure 3.12 illustrates the EMDA classification technique. The packet data from the look-up table is transferred to the predictive model platform. Once the data is exported, all steps in Table 3.3 are implemented according to the modelling approach. There are various supervised learning technique can be employed within the EMDA platform as discussed in Section 2.7. A choice of classifiers is crucial in order to provide a meaningful solution as there is no one algorithm works well for every problem. As an introduction to EMDA technique, three classification modelling approaches have been chosen, namely, Decision Tree (DT), Multi-layer Perceptron Neural Network (MLPNN) and Naïve Bayes (NB). Four main criteria have been set

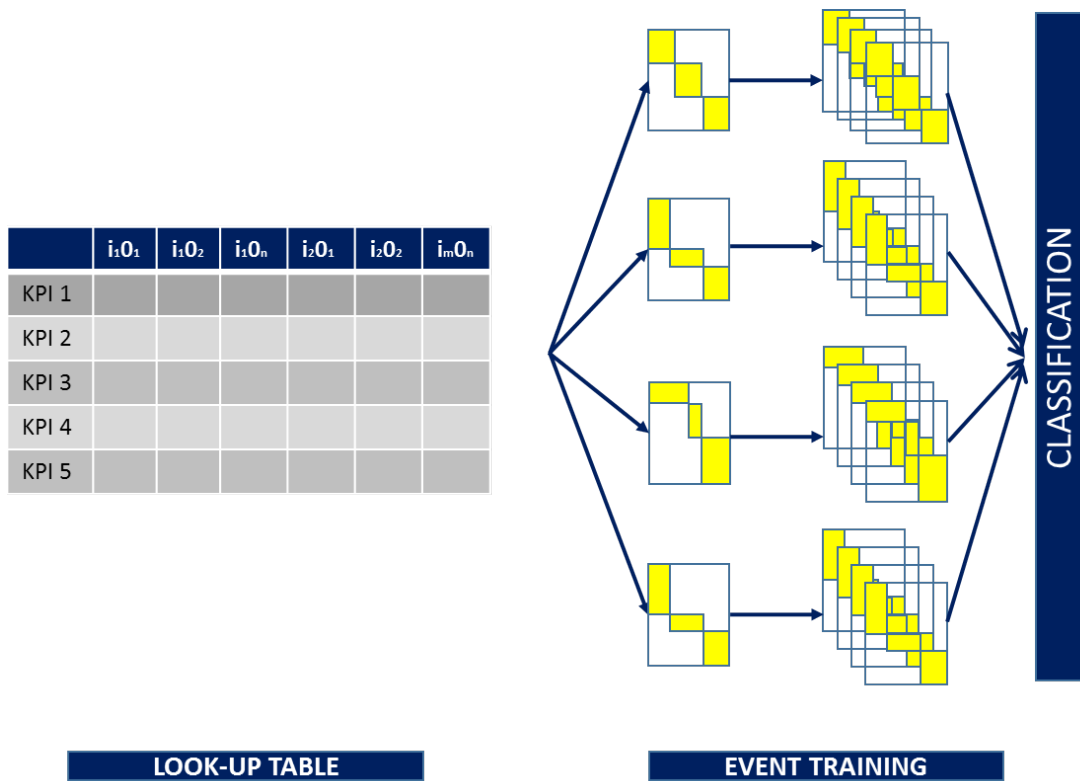


Figure 3.12: Classification using Machine Learning

to justify the selection of classifier techniques which include (1) Type of the Learning, (2) Type of the Data, (3) Complexity and (4) Support Real-Time which will be further discussed in the following Table 3.4.

3.6.1 Decision Tree

DT model recursively separates data samples into branches to construct a tree structure in order to improve the accuracy of the classification. Every tree node is either a leaf node or a decision node. All decision nodes are split, to test the values of some data attribute functions. Each branch of a decision node corresponds to another test result. Every leaf node has a class label attached to it.

DT classifiers are simple and straightforward. The classifier can handle a variety of input data that is nominal, numeric and textual, capable of processing invalid datasets or missing values, and is a high-performance classifier with a limited number of efforts [252]. The decision tree is developed in two steps to learn the model using the training dataset and to test the

Table 3.4: Criteria in Selecting the Classifiers

Description	DT	MLPNN	NB
Type of Learning	Supervised Learning	Supervised Learning	Supervised Learning
Type of Data	Numerical and categorical data	Numerical and categorical data	Binomial and categorical data
Complexity	Able to handle multi-output problems	Capability to learn non-linear models	Super Simple
Real-time resource	Yes but with limitation.	Yes. Capability to learn models in real-time	Yes. Capability to be updated frequently.

model using the test dataset to evaluate the model. In a training dataset, the decision tree is constructed by obtaining significant attributes (variables) to be the root and nodes of the tree. Significant attributes are computed using entropy and information gain (for C4.5 tree type) as in Equation 3.36 and 3.38

$$E(C) = - \sum_{i=1}^c P_i \log_2 p_i \quad (3.36)$$

Where C in Equation 3.36 denotes the computation of entropy of a target variable. Meanwhile, Equation 3.37 is used to compute the entropy of a target variable (C) with a condition of an attribute (X). While in Equation 3.38 is to obtain gain information which indicates the significant attribute to be a node of the tree.

$$E(C, X) = \sum_{C \in X} P(C) E(C) \quad (3.37)$$

$$Gain(C, X) = E(C) - E(C, X) \quad (3.38)$$

Selection of an attribute to be a node in the tree is based upon the highest value of gain information (Equation 3.38). A branch with entropy of 0 will be a leaf node while a branch with entropy more than 0 needs further splitting. The general algorithm for building DT is as follows:

1. Start with the entire training subset and a vacant tree

2. If all training samples are of the same class label c at the current node n , then the node becomes a leaf node with label c
3. Or else, select the splitting attribute x that is the most important in separating the training samples into different classes. This attribute x becomes a decision node
4. A branch is created for each individual value of x , and the samples are partitioned accordingly
5. The process is iterated recursively until a certain value of specified stopping criterion is achieved

3.6.2 Multi-Layer Perceptron Neural Network

Artificial Neural Network (ANN) is a machine learning approach which mimics the functioning of the nervous system. It is usually referred to as biologically motivated and highly sophisticated analytical techniques. They are capable of modelling extremely complex non-linear functions. The neural network consists of several neuron layers. The number of layers depends on the best fit for the problem model studied. When the neural network is fed with data, back-propagation or forward propagation algorithms will run to update the connection strength and reduce the cost function. ANN are analytical techniques based on cognitive system learning processes and brain neurological functions, capable of predicting new patterns (on specific attributes) from other patterns (on the same or other attributes) following a method of so-called learning from existing data [253].

Back-propagation MPLNN is the most popular ANN architecture. MLPNN is known to be an efficient function approximation medium to predict and classify problems. The structure of MLPNN is organised into layers of input, output and hidden layers of neurons. When at least one hidden layer exists, the network's actual computations are then processed. Each neuron in the hidden layer adds its input attributes to x_i after multiplying them by the respective connection weights to w_{ij} and computes its output to y_j using the Activation Function (AF) of that amount. AF can be either threshold function, a sigmoidal, hyperbolic tangent, or radial basis function:

$$y_i = f\left(\sum W_{ij}X_i\right) \quad (3.39)$$

where f is the activation function. Back-Propagation (BP) works by presenting each input pattern to the network where the estimated output is computed by performing weighted sums and transfer functions. The sum of the squared differences between the desired and the estimated values of the Mean Square Error, MSE output neurons is defined as:

$$MSE = \frac{1}{2} \sum_j (y_{dj} - y_j)^2 \quad (3.40)$$

where y_{dj} is the desired value of output neuron j and y_j is the estimated output of that neuron. In Equation 3.39, each weight w_{ij} is adjusted to reduce the MSE of Equation 3.40 as fast as possible. BP applies a weight correction to reduce the difference between the network estimated outputs and the desired ones; i.e., the neural network can learn and can thus reduce the future errors [254, 255]. BP is easy to implement and works well in general. Nonetheless, it has a slow convergence approach and can be trapped in the local minima [163]. Another drawback of the MLPNN models is that they require the initialisation and adjustment of many individual parameters in order to optimise their performance.

3.6.3 Naïve Bayes

Probabilistic approaches to classification are a common machine learning task. Examples include Bayesian-based classifiers [256, 257], which performance is good in terms of accuracy. In recent work, the classifier's predictive performance (accuracy 90.3%) has proven to be the best in predicting the chemical Ames mutagenicity with 5-fold cross-validation [258]. The Naïve Bayes (NB) classifier is particularly appropriate when the dimensionality of the inputs (variables) is high [259]. NB is a classification algorithm for binary (two-class) and multi-class classification problems. The NB algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given data set based on the Bayes's Theorem with the conditional independence assumptions. Bayes' Theorem is formulated as [260]:

$$P(c|t_i) = \frac{P(c)P(t_i|c)}{P(t_i)} \quad (3.41)$$

where c is a class of the target variable (C)($c \in C$)

$P(c|t_i)$ is the posterior probability of class (target) given predictor (attribute). $P(c)$ is the

prior probability of class. $P(t_i|c)$ is the likelihood which is the probability of predictor given class. $P(t_i)$ is the prior probability of predictor.

3.7 Summary

Event Modeller Data Analytic (EMDA) was introduced in this chapter as a technique to predict the pattern of system problem in a time-constrained complex system. Borrowed from the EventTracker and EventiC principle of the interrelation of causal events, the EMDA technique cohesively blended with translated KPI which makes systems more intelligent in classifying the state of the machine that deals with real-time events. For testing and validating purpose, three machine learning models, which include Decision Tree (DT), Multi-Layer Perception Neural Network (MLPNN) and Naïve Bayes has been chosen to integrate with the EMDA technique. This technique embraced the predictive analytics workflow, consisting of four steps. However, EMDA only progressed up to step 3, which left step 4 for future works. Details of each step and how it is integrated within the proposed EMDA technique has been discussed extensively in this chapter. The basic parameter of DES, including all algorithm and mathematical function used within the research is also presented in this chapter. As proof of concept, this technique will be used to solve the industrial case study in Chapter 6.

Chapter 4

Industrial Case Study

4.1 Overview

The power industry is essential to the industrial world, transporting energy along the supply chain to residential, commercial, government infrastructure and to consumers for daily life activity. Power plant generates electricity and convert them into High Voltage (HV) before transfer it to the system grid. This can reduce the energy losses while transmitting it in long distance. When the energy reach the substations, it is then convert into Low Voltage (LV) before it distributed to consumers. Figure 4.1 shows a general power generation supply chain.

The reliability of the individual power plant is determined by the performance of the entire supply chain, which includes the performance of the equipment, the quality of the generated power and control lines, as well as the supply of materials. The performance of the machinery is one of the highest priorities in operation and maintenance of the power station as it has direct impact on the financial, reputation, security and safety of the plant. The performance of a Continuous Ship Unloader (CSU) machine in a Malaysian Power Plant is being used as a case study in studying the failures of harmonic filters. The implication of these frequent failures has increased machine downtime and disruption to the national grid while reducing the personnel safety and plant availability. To solve the problem, an effort to analyse both internal and environment parameter that may have a significant impact on the system is highly desirable. A set of Key Performance Indicators (KPI) are selected for continuous monitoring of the machines and to determine the effect of internal and environmental events that may demonstrate high correlations with failures of harmonics filters. The aim is to

propose a Real-Time Software in the Loop (SIL) framework that identifies the inputs which have a significant impact on the ineffectiveness of the machine power system. The judgement between useful/useless and relevant/irrelevant for the performance affecting events (input) depends on the speed and the quality of the separation process.

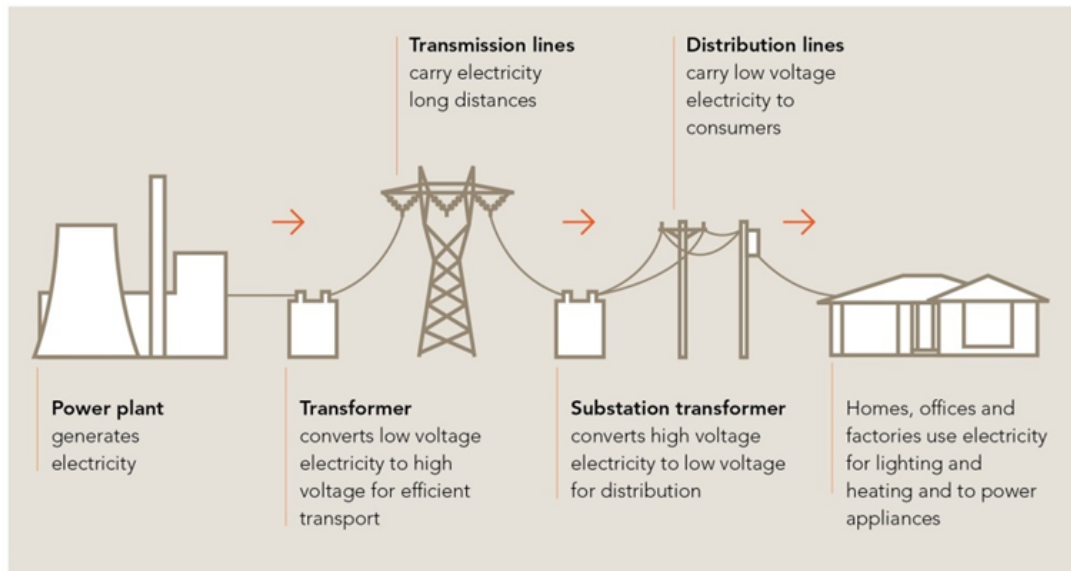


Figure 4.1: Power Generation Supply Chain, Image: courtesy of Australian Energy Market Operator [2]

The hypothesis to be tested is that the implementation of EMDA could lead to (a) offering operation manager the key performance indicator status of the plant in real-time (b) providing system engineer with known and unknown parameter that could lead to power system failure. (c) decision aid system in making necessary action before fault occurs again in the future. (d) suggesting ways to schedule an efficient operation activity and (e) providing information to tune the machine.

This chapter aims to address the industrial case study and how the EMDA technique introduced in the previous chapter can be used to predict future fault patterns. The remainder of this chapter is divided into six main sections. Section 4.2 provides a general overview of a power generation industry. Section 4.3 describes the detailed breakdown of the target machine. In this section, sufficient information is presented to ensure that the reader understands the operation of the system. Section 4.4 brings out the result of the PQ Assessment Initiatives to highlight the gap in this study. The chapter then discusses the industrial data collection set-up in Section 4.5. Chapter 4 is finally summarised in Section 4.6.

4.2 Power Generation Industry

The main objective of this section is to gain a sufficient understanding of the power generation main processes with its sub-processes which have a significant relationship to the Key Performance of the whole plant availability. The coal-fired power plant is a type of power plant that uses the combustion of coal to generate electricity. In general, this plant is driven by three main component which are generator, steam turbine and supercritical boiler which supported by Coal Handling Plant and other Balance of Plant (BOP)'s. The conversion of this coal into electricity is a multi-faceted process:

1. Coal Unloading and pulverising process feed via Coal handling Plant
2. Combustion of pulverised coal in Super Critical Boiler
3. Extraction of thermal energy from pressurised steam to rotate Steam Turbine
4. Generation of electricity in Power Generators

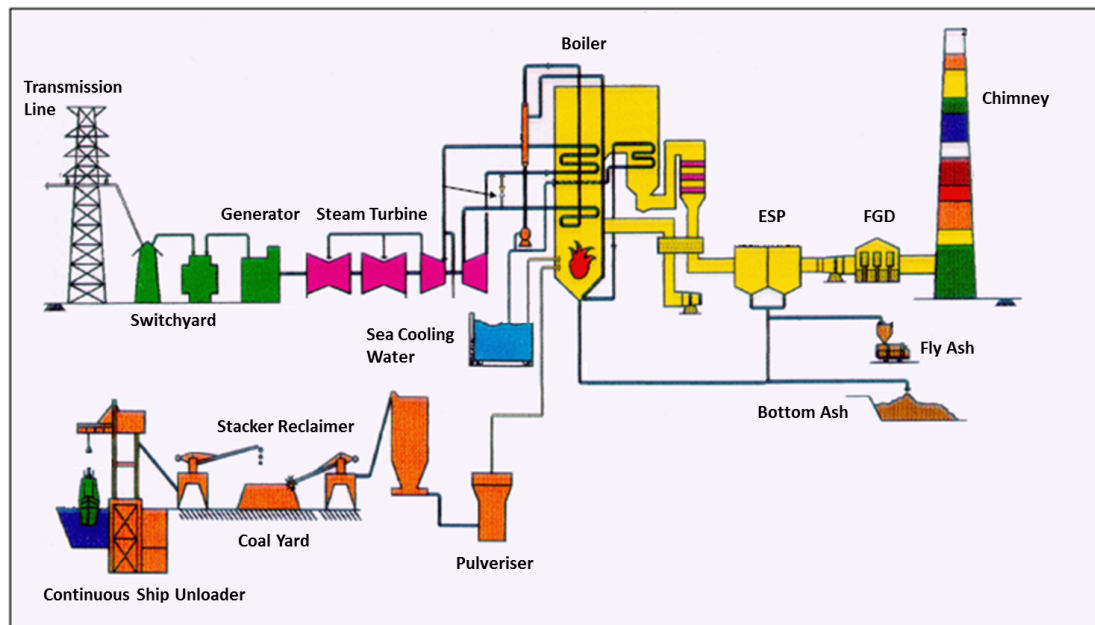


Figure 4.2: Coal Fired Power Plant Overview, Image: courtesy of [3]

4.2.1 Coal Handling Plant

Coal is unloaded from the vessel and transferred to the coal yard via a series of belt conveyors. The Stacker Reclaimer machines then ferry the incoming coal at the coal yard to the selected

stockpiles, which is later being reclaimed and transported to the coal bunker by another series of conveyor belt. Before entering the bunker, the coal will pass through another necessary segregation process at the coal crusher house, where coal with the size of less than 50mm will be crushed and directly sent to the bunkers by the belt conveyors. The rejected coal (sized more than 50mm) will be sent back to the coal yard. The coal retained inside the bunker will be re-crushed into pulverised fuel in a pulveriser. Primary air is being used to convey the pulverised fuel into the furnace located inside the boiler for a smooth combustion process to take place.

4.2.2 Super Critical Boiler

During the boiler start-up, Liquefied Fuel Oil (LFO) is used until stable firing is achieved at approximately 30% of the generator load. Then the pulverised fuel from the Coal Handling Plant is used to firing between 30% till 100% load. Secondary air from Force Draft Fan is used as combustion air. Demineralised water is supplied to the boiler steam drum and thousands of water tubes inside the boiler. The heat generated from the combustion process turns water into steam inside the steam drum.

The steam is then moved to a series of superheaters while unconverted steam inside the steam drum will be recirculated to water wall. Bottom ash and fly ash will be the by-product of this combustion process. The bottom ash will be sent to Bottom Ash Bin before being sent out to ash pond. Meanwhile, fly ash in the flue gas will be filtered by Electrostatic Precipitator (ESP) and stored in a silo before transporting to the cement industry. SO_2 inside the flue gas is then treated inside the Flue Gas Desulphurization (FGD) before being released to the environment through the chimney.

4.2.3 Steam Turbine

Steam emitted from the steam drum will pass through a series of superheaters before entering the High Pressure (HP) steam turbine. The steam released from here will then be relayed to the reheater positioned inside the boiler, before being channelled into the Intermediate Pressure (IP) and Low Pressure (LP) Turbine. The high energy from the steam turns and spins the turbine blades at a nominal speed of 3000 rpm. The exhausted steam released from the LP Turbine is reverted to the condenser and regenerated into water through a condensation

process.

4.2.4 Power Generator

A turbine rotates the generator rotor and hence creates an electric current. The voltage produced at 26kV and stepped up to 500kV by the generator transformer. The electricity later enters the transmission line and exported to the grid system. Once the electricity is dispatched to the grid, it will then be distributed to the consumers.

4.3 Continuous Ship Unloader Machine Breakdown

As was pointed out in the introduction to this chapter, the CSU machine is experiencing frequent harmonic filters failures. This frequent failure is linked to the application of control drives and other nonlinear loads in the industry have resulted in the deterioration of PQ in the electrical distribution system. In general, PQ is the measurement of the voltage, current and frequencies in a stochastic power system event, to ensure that the system is stable and reliable as per IEEE standards which allow the loads to be operated without disturbing, damaging or reducing the performance, efficiency and life expectancy [261].

The harmonic issue in line with other PQ disturbance (e.g., voltage sags, voltage swells, switching transients, impulses, notches, and flickers) has been vigorously debated as the main reasons for machine failure, operation downtime and injury to the personnel [262]. It has been widely agreed that the three (3) main criteria that could sustain the performance of the electrical distribution system are reliability, security, and stability. The system reliability refers to the consistency of operating at the optimum condition without reaching any undesired state. System security is the level of robustness required to withstand its integrity, while system stability refers to the ability of the system to maintain its operation and remain stable even though disturbance has been introduced in the system [263].

In the context of this study, power system stability and security relate to the ability of the CSU machine to consistently supply stable voltage to the Regenerative (REGEN) drives without any interruption or fault, while if there is any, it is capable of restoring to its steady state efficiently. Original Equipment Manufacturer (OEM) has ensured the electrical system is designed based on its total connected load and has the capacity to deal with high-frequency

solid-state switching devices such as control drives and other non-linear loads. The disadvantages of this non-linear loads are that it generates harmonics currents while degrading the electrical system and deviating from the sinusoidal waveform. To overcome the presence of these PQ disturbances, the OEM has designed the system with a harmonic filter, also known as "EMC filters" to cancel out the harmonic distortion in the system by diverting it into the low impedance path. Before proceeding to discuss further on the harmonic issue, it will be necessary to explain the key features of a CSU machine.

4.3.1 Continuous Ship Unloader Overview

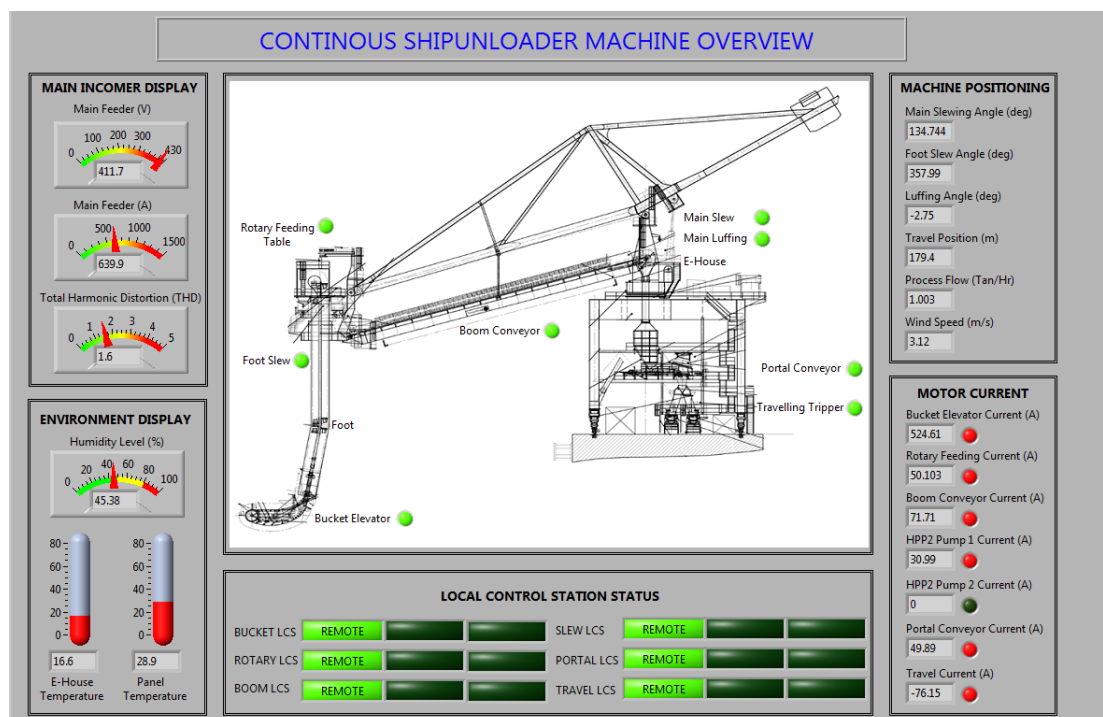


Figure 4.3: Continuous Ship Unloader Overview

The CSU serves to handle power station coal and is designed to unload ocean-going vessels from 35 000 DWT up to 150 000 DWT in load capacity. The material to be unloaded is reclaimed by the bucket elevator in the ship's hold and transferred to the rotary feeding table which conveys the material further to the boom belt conveyor via the central chute system and hopper. The materials discharged from the boom conveyor are transferred to the portal belt conveyor, from where it is further transported to either one of the two jetty belts by the help of a two-way chute.

To unload residual material, provision is made for the use of a front loader which is lowered into the ship's hold to feed the bucket elevator by scooping and pushing the material from the ship's bottom and the corners. The transportation of the front loader will be carried out by means of the luffing, slewing and travel gears. For that reason, the front loader will be fixed at the connecting beam to a suspension device and lowered into and subsequently, to be lifted out of the hold. The connecting beam for the front loader will be attached at the lower part of the elevator tube. The weight of the front loader cannot exceed 18 tons. The main component in CSU machine is divided into 7 main areas as follows:

1. Bucket Elevator
2. Rotary Feeding Table
3. Boom Conveyor
4. Main Slewing & Hoisting
5. Electrical House (E-House)
6. Central Chute & Portal Conveyor
7. Travel Gear

4.3.1.1 Bucket Elevator

Bucket elevator is connected to the CSU to be lifted and lowered such that the axis of the bucket elevator will be in the vertical direction over the whole lifting range. The vertically movable bucket elevator foot is connected to the carrying tube via the guide rollers and the vertical cylinder. In order to increase the average unloading capacity, taking into account the unloading of the remainder, the lower part of the bucket elevator features a bent design that allows the areas under the hatch wings to be reached.

The chain speed is limited to 1.8 m/s during normal unloading operation. The running direction of the bucket elevator can be changed for inspection purposes or for removing bulky foreign pieces. The chain speed during inspection works is about 0.30 m/s. The chain speed (1.8 m/s) for unloading operation as well as changing of running direction (chain speed about 0.3 m/s) can be controlled either from the remote control station or the operator cabin. The chain speed for inspection (about 0.3 m/s), in unloading and backward direction, can be

controlled via the local control station. The bucket elevator is driven by an electromechanical drive unit which consists of electric motors, shaft mounted gearbox with coupling & brake and chain wheels. The mounted-on bucket elevator drives are shrink-fitted to both sides of the drive shaft. The driveshaft rests on the steelwork using two pillow blocks.

The bucket elevator is connected to the vertical-holding platform by an external toothed large-diameter anti-friction bearing. In the "out of operation" position, the bucket elevator foot is positioned over the runaway of the unloader. The bucket elevator can operate within a slewing range of 360° , which is endless. The exact slewing position of the bucket elevator foot is registered via an encoder being in mesh with the ball-race slewing connection using a pinion.

The Bucket Elevator slewing gear drive unit consists of a planetary gearbox with flange-mounted multi disc brake and with a flange-on-hydraulic motor. A drive pinion, mounted on the drive shaft, meshes with the external toothing of the slewing ring and provides slewing motion.

4.3.1.2 Rotary Feeding Table

The buckets reclaim the material to be handled in the lower area of the bucket elevator foot and transfer it to the rotary feeding table. The discharge of the material approximately starts at the peak of the drive chain wheel and is completed after a slewing movement covering an angle of 30° - 45° . With the rotary feeding table, the coal will be transported in the rotation direction and discharged by the removal arm via the movable chutes onto the boom conveyor. In the area of rotary feeding table, probes are installed, which in case of chute or feeding table clogging initiate the bucket elevator drives to switch off. The rotary feeding table is driven by an electromechanical drive unit consisting of two flange-mounted gear motors.

4.3.1.3 Boom Conveyor

Materials discharged from the bucket elevator is directed to the rotary feeding table and transported to the boom belt conveyor. The boom belt conveyor transports the material to be carried out to the central chute, from where it is transferred to the portal belt conveyor. The carrying belt is guided over three-piece troughed idler sets whereas the return belt guides over one-piece and two-piece idlers. Within the material feeding area, the idlers are provided with

buffer rings. In order to improve the belts running characteristics, a training idler station is arranged behind the return as well as in front of the drive pulley. The belt conveyor is equipped with safety devices consisting of emergency pull chord switches, belt sway switches and speed monitoring switch. The drive of the boom belt conveyor is made up of LV motor, hydraulic coupling, flexible coupling, drum brake and gearbox. The gearbox is connected to the pulley by means of a shrink disc. In order to absorb the reaction power resulting from the driven end torque, the support for the frame torque is mounted on the steel profile of the take-up station.

4.3.1.4 Main Slewing & Hoisting

Slewing gear of the CSU machine consists of a double-row slewing connection with external tothing, designed as a double-row roller bearing, and four drive units laterally attached to the slewing connection. The boom is slewed parallel to the runaway and manually locked using a bolt when it is out of position. The locking bolt on the limit switch prevents any accidental start of the slewing gear. Rotary cam limit switches limit the slewing range (maximum $\pm 90^\circ$) of the boom. The final positions of the slewing range are limited by a limit switch in each slewing direction, the slewing movement being detected electronically by means of a resolver.

Bumper arranged at the limit stops absorb the kinetic energy, in case the slewing gear rotates beyond the electric limit stops against the buffer stops. An encoder is used to register the exact slewing position. One slewing gear drive unit consists of a planetary gearbox with a flange-mounted multidisc brake, a flexible coupling and a flanged-on hydraulic motor. A drive pinion, mounted on the drive shaft, meshes with the external tothing of the slewing ring and provides for the slewing motion. At the front end, the boom and the counterweight boom are connected with the vertical-holding platform of the bucket elevator. In the middle of the ship unloader, they are attached to the pylon by heavy-duty hinged bearings in such a way that the axis of the bucket elevator remains in the vertical direction over the entire hoisting length.

The hoisting movement and bucket elevator are affected by one hydraulic cylinder which is centrally mounted under the counterweight boom. The respective power pack (hydraulic) is installed on the slewing platform. The boom hoisting position is detected by an absolute encoder and transmitted to the control unit.

4.3.1.5 Central Chute & Portal Conveyor

The central chute system is designed to transfer the material discharged from the boom belt conveyor to the portal belt conveyor. The system is composed of a stationary part with a hopper arranged at the portal beam and a slewing part connected to the pylon. The central chute system is equipped with inspection doors. The lower parts of the central chute were designed with intermediate hopper. It serves to equalise the material flow and limit the converting rate which will be transferred to one of the jetty conveyors by the portal conveyors. The hoppers were equipped with monitoring devices in order to observe the upper operation limit and the upper emergency limit. In case the filling degree reaches the upper operation limit, the speed of the bucket elevator eases off and further decreased as long as the upper limit remains. In case the upper emergency limit is breached, the bucket elevator will be totally switched off.

The portal belt conveyor reclaims the material discharged from the boom belt conveyor. By the help of a gate, the material flow will be levelled and is transported to the two way chute, from where it is directed to one of the two jetty belt conveyors. The drive pulley, the discharge hood as well as the pre and main cleaner are arranged at the head of the belt conveyor, whereas the return pulley, the take-up device with return hood, the safety cleaner and the skirting completely covering the material transfer zone are located at the rear of the belt conveyor. The magnetic separator is installed above the discharge hood. The drives of the portal belt conveyor consist of LV motor, flexible coupling, drum brake and gearbox. The gearbox is connected to the drive shaft of the pulley by a shrink disc. Absorbing the reaction power resulting from the driven end torque, is by the frame torque support resting on the steel profile of the take-up station. The belt conveyor is equipped with safety devices which consist of emergency pull chord switches, belt sway switches and speed monitoring switch.

4.3.1.6 Travel Gear

The travel gear is driven by 10 geared motors. Each motor moves two runner wheel of a two-wheel bogie. The electric motor is flanged-connected to the gearbox which is screw-fixed to the two-wheel bogie and connected with the pinion shaft. The travel movements of the ship unloader are decelerated on a time-delay basis with the help of a frequency variable motor, where the delay interval can be adjusted. Only at the moment, when the ship unloader stands

still, the multi-disk brakes flanged to the gearboxes are closed.

4.3.2 Continuous Ship Unloader Electrical System

So far, this section has focused on the CSU machine in general. The electrical aspect of the CSU machine will be discussed in the following section. It is important to understand the electrical distribution system of this machine and to have a better understanding on how the PQ assessment initiative is implemented, which will be discussed later in this chapter. Figure 4.4 is the single line diagram for the entire CSU machine.

The main component for the electrical system in the CSU machine is divided into 5 main areas as follows:

1. Incoming Transformer
2. System Protection
3. Motors
4. REGEN & Control Drives System
5. DC-Bus distribution system

4.3.2.1 Incoming Transformer

CSU machine is designed with a 3.3KV/0.438KV Cast Resin Transformer which is more compact and requires less space. This dry type transformer is self extinguishing and maintenance-free. It does not produce toxic gases while only needing air to cool which make it easy to maintain. For monitoring purposes, a Current Transformers (CT's) is installed at the LV side of the incoming voltage to feed into Multi-function SOCOMEC device.

4.3.2.2 System Protection

A Circuit Breaker (CB) is required to protect the CSU machine whenever there is an event of excessive current from an overload or short circuit within the electrical system. This machine is designed with an Air Circuit Breaker (ACB) which is more efficient, less bulky and cheaper in cost. With non-explosive and silent operation behaviour, this type of circuit breaker is suitable for repeatable duty operating characteristics which meets the CSU machine operation

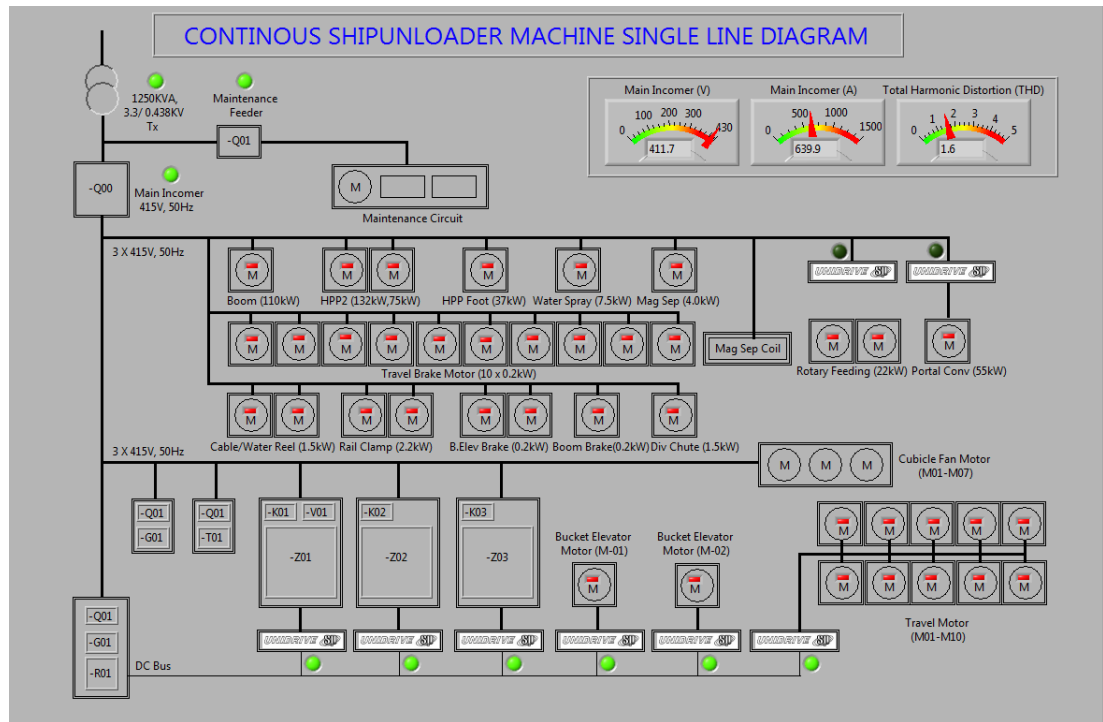


Figure 4.4: CSU Machine Single Line Diagram

demand. 2 types of CB's is used, Maintenance feeder (-Q01) CB which cover the maintenance circuit and other essential load and Main feeder ACB (-Q00) CB rated at 2000 A which covers the entire CSU machine load.

4.3.2.3 Motors

Induction motor is used to drive the CSU machine operation. The construction of the motor is simple but robust and mechanically strong. This type of motor is highly efficient and provides high starting torque to operate the CSU machine capacity. The following Table 4.1 provides the list of motors for the CSU machine.

4.3.2.4 REGEN & Control Drives System

Operating the CSU machines requires precise control. Variable frequency drive controls the movement to meet a different speed requirement of the machine. To operate these drives, a bridge rectifier circuit is required to convert the alternating current (AC) input into a direct current (DC). When bucket elevator is driven downward, energy flow is reversed, that is, from the load, through the motor, back to the drive. This spare energy could be regenerated using

Table 4.1: Motor List for CSU Machine

No	Motor Rating	Quantity
1	0.2 kW Motor	14
2	1.5 kW Motor	3
3	2.2 kW Motor	2
4	4.0 kW Motor	1
5	7.5 kW Motor	1
6	10.9 kW Motor	1
7	11 kW Motor	1
8	22 kW Motor	2
9	37 kW Motor	1
10	55 kW Motor	1
11	75 kW Motor	1
12	132 kW Motor	1

REGEN drives. Regenerative control gives a motor the capability to act as a generator while in operation. Energy is regenerated during the braking process (regenerative braking), and the output energy is supplied to an electrical load. This would also provide higher efficiency levels in the electrical distribution system.

The electrical system must ensure that the power supplied to the drives is clean. Therefore, it requires a power quality filters circuit, which includes Power Contactors (K01), EMC Filter (Z01), Varistor (V01), and Line Reactance Choke (L01). The EMC filter eliminates intolerable electromagnetic disturbance from the power supply, which could harm the drive system. The varistor provides reliable and economic protection against high voltage transients and surges. The Line Reactor (Choke) reduces the risk of damage to the drive from poor phase balance or severe disturbance on the supply network. It also reduces harmonic current emission.

4.3.2.5 DC-Bus Distribution System

The 3-phase supply from the Main feeder -Q00 is distributed to the busbar line. The busbar line supplies energy to the DC-bus distribution, rectifier circuit and other loads including LV motors and cubicle fan motors. The bucket elevator unloads coal from the vessel to the CSU machine used high inertia load. This requires a high voltage supply to move up to 1800 tonne/hr capacity of coal with different speed control. Therefore, the DC bus distribution system needs up to 630 VDC to accommodate the demand. To meet this requirement, the DC bus has to achieve 400 VDC via a bridge rectifier circuit which consists of Circuit Breaker

(Q01), Power Contactors (K04), Bridge Rectifier (G01) and Resistor (R01). Combining this with the REGEN drive output from the power quality filter circuit, the DC-Bus will be able to supply 630 VDC to meet the load demand.

This section has presented the electrical system of a CSU machine which includes system protection, drives system and all electrical devices installed within the machine. The next part of this section will discuss the control system of the machine, which covers the system architecture and the main component of the control system. It is important to understand this system and to have a good overview of how the real-time data collection setup is implemented, which will be discussed at the end of this chapter.

4.3.3 System Architecture for Continuous Ship Unloader

The system architecture for the CSU machine used Programmable Logic Control (PLC) to automate and coordinate the entire machine operation. It reads real-time information from the input sensors and devices, converting it into signals that can be processed by the Central Control Unit (CCU). The CCU then makes an instantaneous decision, to decide on the next status of the machine with respect to the pre-programmed logic stored in the memory. At the same time, the PLC synchronises with Supervisory Control Aided And Data Acquisition (SCADA) system to distribute, monitor, gather and process real-time data through a Human Machine Interface (HMI), by helping operators to control the movement of the machine, thus make important decisions to ensure the machine is safely handled. Having discussed the overall system architecture for the CSU machine, description of the main components of the control system in this machine are as follows:

1. Hardware
2. Software
3. Network Communication

4.3.3.1 Hardware

Hardware refers to the physical elements that make all electronic devices within the PLC networks communicating with each other and physically tangible. The main hardware involves a controller module, input modules, output modules, a communication module, power supply

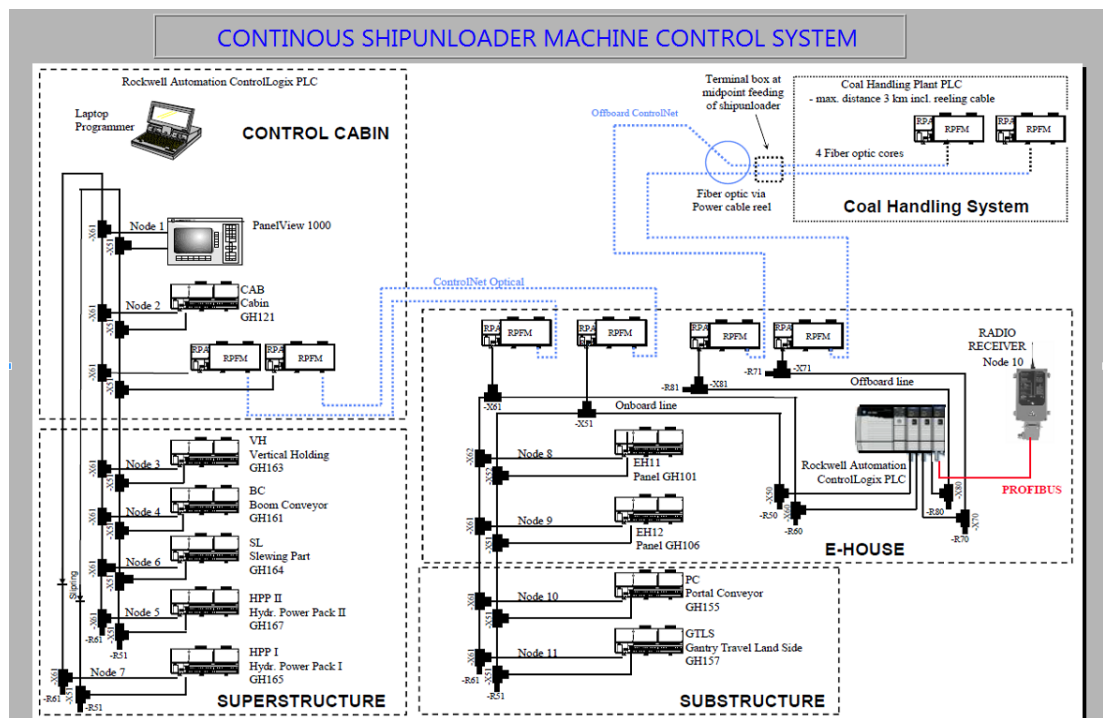


Figure 4.5: CSU Machine Control System

and HMI monitor. Figure 4.5 illustrates the control system configuration for CSU machine. It is divided into four main areas which are Control Cabin, e-house, superstructure and substructure. The CSU machine controller communicates with the main controller at Coal Handling System through a fibre-optic connection via a Medium Distance fibre optic repeater. The controller then synchronised the machine operation within the four main areas with a short distance fibre optic repeater. Alternatively, for a Radio Remote Control (RRC) configuration, the controller communicates with a Profibus Remote I/O scanner, to control the CSU machine remotely at 735MHz frequency. Other hardwares such as Digital and Analogue Module (Input & Output) and ControlNet module are used to link between the sensor and controller. The following Table 4.2 shows the hardware list for the CSU machine.

4.3.3.2 Software

Software is a set of instructions, data or programs used to operate computers and execute specific tasks. Without software, the hardware discussed previously is useless. Rockwell Automation used Studio 5000 Logix Designer or previously known as RSLogix 5000 to program, design and configure the entire Allen Bradley family controller products and other related

Table 4.2: Control System Configuration Hardware List for CSU machine

No	Device	Description	Qty
1	1756-PB72	ControlLogix. Power supply unit. 24V DC.	1
2	1756-L61	ControlLogix Logix5561 Processor 2MB Memory	1
3	1756-CNBR	ControlLogix ControlNet Redundant Media Module	2
4	SST-PFB-CLX	PROFIBUS DP Remote I/O Scanner Module	1
5	1786-RPFM	ControlNet Medium-distance Fiber Repeater	2
6	1786-RPA	ControlNet Modular Repeater Adapter Module~	6
7	1786-RPFS	ControlNet Short-Distance Fiber Module	4
8	1794-IB32	Flex I/O Input Module, 24V DC, 32 Sink Inputs	10
9	1794-OB32P	Output Module, 24V DC, 32 Source Outputs	4
10	1794-IE8	Flex I/O Analog Input Module 8 Single-Ended Inputs~	5
11	1794-ACNR15	FLEX I/O ControlNet Redundant Adapter Module	10
12	1794-OE4	Flex I/O Analog Output Module 12 Bit 4 Single Outputs	2
13	PV1000	PanelView Standard 1000 Graphic Terminals	1
14	1794-IB16	FLEX I/O Input Module, 24V DC, 16 Sink Inputs	19
15	1794-OB16P	Flex I/O 16 Source Outputs Module, 24V DC,	3
16	1794-OB8EP	FLEX Output Module 24VDC 8-Port Electronically	7

devices. Its intuitive programming environment allows users to work collaboratively to design and maintain their systems. On the other hand, Allen Bradley used Studio View Designer or previously known as Panel Builder 32 which allows system engineer to create a control panel application for the standard Panel View operator terminal (HMI). A panel Builder application allows the operator to interact with a logical arrangement such as machine status, alarms, events, indicators and push-button.

4.3.3.3 Network Communication

Network communication is the transmission of data between the main controller with other devices such as PLC, I/O chassis, HMI and drives within the network to exchange data. The physical connection is established using either cable media or wireless media. Allen Bradley uses the ControlNet network to support real-time control of I/O. It is an open industrial network protocol and is managed by “Open DeviceNet Vendors Association” or ODVA. It utilises the Common Industrial Protocol (CIP) for the upper layers of the Open Systems Interconnection or “OSI model” that has seven layers which include physical, data link, network, transport, session, presentation and application. The ControlNet coax media system consists of components, such as the trunk, drop cables, taps, cable connectors, terminating resistors,

nodes and repeaters to create segments, links, and bridges for network communication. Nodes are the physical devices that require an address to function on the ControlNet network. A tap connects these nodes to the coax media system via the trunk cables and the cable connector. A collection of nodes, tap, trunk cables and connector formed a segment which provides a network for a repeater.

Having defined how Allen Bradley apply its network communication, the network topology for the CSU machine use two channels redundant linear bus topology, Channel A and Channel B. There is no overall network parameter that says which channel to use, each node decides on its own. At any time, whether a network has valid or invalid redundant cabling, there can be a mix in which channel any node is listening to; some nodes could be listening on Channel A and some could be listening on Channel B. In the event of a cable failure on one of the channel, the other channel will take over. In terms of a repeater, the CSU machine uses fibre repeater modules with fibre adapter to establish a connection between the main controller in the e-house to the other three locations in the network. The fibre repeater modules send an optical signal through the fibre cable to the next fibre repeater module on the network.

So far this section has focused on the key features of a CSU machine which covers the machine overview, electrical system and control system architecture. The following section will discuss on the power quality assessment initiative to monitor the frequent breakdown incidence as per discussion earlier in this section.

4.4 Power Quality Assessment Initiatives

As the machine mainly controls the operation of bulk handling in three-axis, it is essential to monitor the movement is within the safe limits. These movements are mainly controlled by drive system. Due to the fact that drive system deals with a high frequency solid state switching, it then generates high harmonic currents which disturbs and degrades the electrical system. As mentioned at the beginning of this section, a harmonic filter is installed to reduce, or mitigate harmonics to a tolerable level. When these harmonics are left unattended, it can potentially lead to expensive damage to the component, which results in high machine breakdown and maintenance cost. To take this into account and address this concern, Power Quality Assessment Initiative has been conducted to evaluate Power Quality level and identify

the source fault that contributes to high harmonic failure.

The CSU machine was assessed as per International Standard IEEE 519, 1992, which sets the limit for both harmonic voltages and currents at the Point of Common Coupling (PCC) in a system. It limits voltage THD (Total Harmonic Distortion), defined as the ratio of the RMS value of the harmonic voltage to the RMS value of the fundamental (50Hz) voltage, to a maximum of 5%. Individual voltage harmonic magnitudes are limited to 3% of the fundamental voltage value [264]. The remaining loads were assessed as per International Standard IEC 61000-3-4 setting the limits for the emission of harmonic currents in low-voltage power supply system for the requirement with rated current greater than 16A [265].



Figure 4.6: Power Quality Analyser Loggers with Clamp CT's and Flexible Voltage Probes

4.4.1 Power Quality Measurement Setup

Several measurement tools are available for PQ measurement. PQ analysers are the most commonly used tool to observe real-time readings while collecting historical data for analysis. In most cases, when the analyser is not built in the distribution system, a handheld analyser is used. A PQ analyser is used to continuously monitor and analyse the electricity lines by measuring several electrical parameters. These include AC voltage, AC current, power and frequency to detect disturbance events, such as voltage sags and swells, transients, harmonic distortion, and voltage/current imbalance.

The advantage of using handheld PQ analyser is that it is capable of providing real-time data on the screen. However, it has a limitation in recording long term data for continuous PQ events. In this assessment, a commercial PQ analyser – Fluke 1735 Three Phase PQ logger has been installed at the Point of Common Coupling (PCC) using Clamp Current Transformer (CT's) and Flexible voltage probes to record the associated parameters for up to 45 days. The PCC for the 1250KVA Main Incoming Transformer is located at the secondary part of

Table 4.3: Load Description and Time Interval for Power Quality Measurement at CSU Machine

No	Load Description	Rating	Interval (sec)	Log Duration (min)
1	1250kVA Incoming	3.3 kV/0.438 V	30	1440
2	Maintenance Winch	22 kW X 3	1	15
3	Maintenance Circuit (LHT)		1	5
4	Maintenance Circuit (Floodlight)		1	5
5	Maintenance Circuit (Cabin)		1	5
6	Maintenance Circuit (Motor Heater)		1	5
7	Control Volt. Distribution	2.5 kVA	1	5
8	Rectifier (U01-U03)	20 A X 3	1	30
9	Water Spraying Pump	7.5 kW	1	5
10	Rectifier of Magnetic Separator	10 kW	1	5
11	Magnetic Separator Conveyor	4 kW	1	5
12	Diverter Flap Actuator	1.5 kW	1	5
13	Power Cable Reel	1.5 kW	1	5
14	Water Hose Reel	1.5 kW	1	5
15	Rail Clamp Landside	2.2 kW	1	5
16	Rail Clamp Seaside	2.2 kW	1	5
17	Boom Conveyor	110 kW	1	30
18	HPP II Pump Motor 1	132 kW	1	30
19	HPP II Pump Motor 2	75 kW	1	30
20	HPP I Bucket Elevator foot	37 kW	1	30
21	Bucket Elev Thruster Brake 1-4	0.2 kW X 4	1	40
22	VVVF Drives Input		1	10
23	Bucket Elevator Drives 1	200 kW	1	30
24	Bucket Elevator Drives 2	200 kW	1	30
25	Portal Conveyor Drives	55 kW	1	30
26	Rotary Feeding Table Drives 1	22 kW	1	30
27	Rotary Feeding Table Drives 2	22 kW	1	30
28	Travel Drives	200 kW	1	30
29	Travel Gear Motor 1-10	1.5 kW X 10	1	50
30	Travel Gear Brake 1-10	0.2 kW X 10	1	10

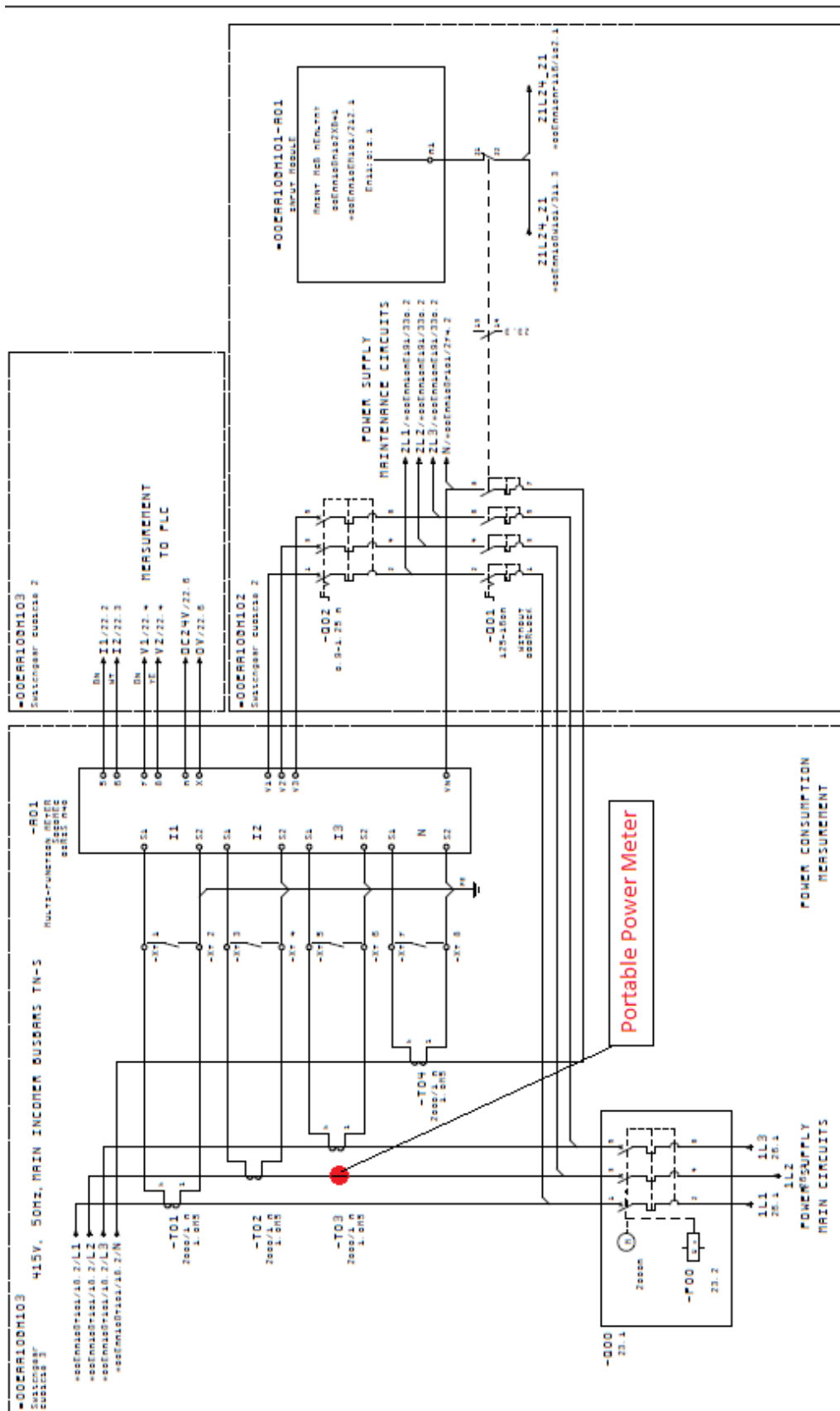


Figure 4.7: Point of Common Coupling for the 1250 KVA Main Incoming Transformer

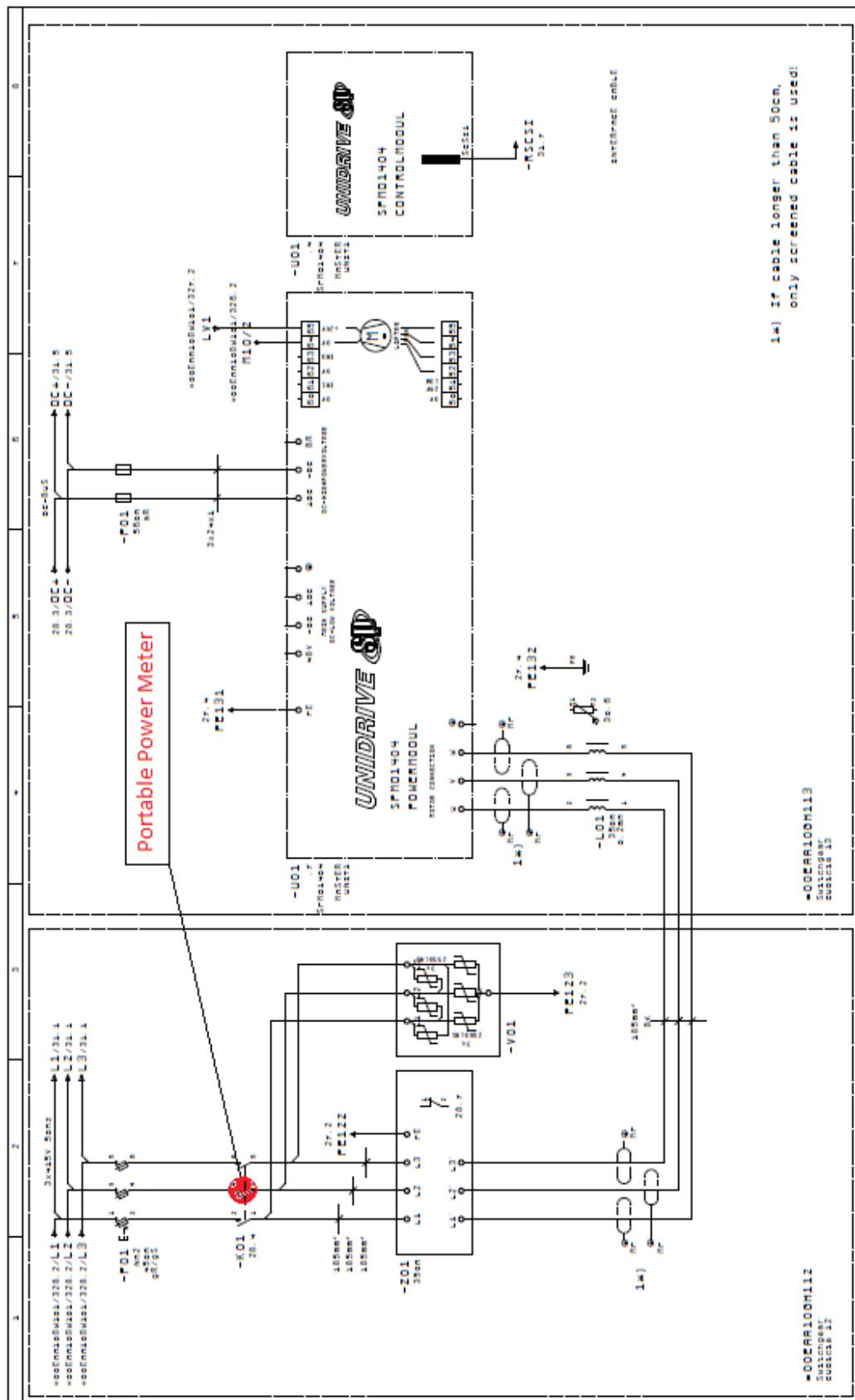


Figure 4.8: Point of Measurement for the VVVF Drives Input (Master REGEN)

the circuit, which is located in the GH103 panel, in line with the SOCOMEC Multi-function meter as per Figure 4.7. On the other hand, the PCC for the VVVF Drives input for the Master-REGEN is located in the GH112 panel as per Figure 4.8. In addition, the voltage probe and clamp CT's were also located at 30 different locations as illustrated in Table 4.3. Each location has different sampling time (log duration) depending on the size of the load. The sampling interval is set at 1 second for short data logging and 30 seconds for long data logging (1-day data).

To fulfil the objectives of this assessment, safety precautions have been taken to ensure all measurement data follows the following steps.

1. Coordination with maintenance engineer for the installation of electrical instruments or measurement instruments.
2. Check all necessary permits are in order. Inform plant owner before starting the works.
3. Make sure all safety gloves and tools are available to hook up a portable power meter
4. Installation of electrical instruments for loads to be measured: (a) Open electrical panel and perform a safety check before start installation works. (b) Install current clamp (CT1, CT2, CT3) to clamp on busbar/cables (L1, L2, L3) for current measurements. (c) Connect L1, L2, L3 and N on the busbar/cables.
5. Connect the auxiliary power supply to a portable power meter. (a) Turn on the power supply to the meter. (b) Check meter for correct settings and configurations.
6. Check meter for correct measurement on the display.
7. Leave portable power meter at the location for measurements. Thereafter, it will be dismantled and transferred to another point for measurement purposes.
8. Inform plant owner upon completion of works.

4.4.2 Power Quality Assessment Results

Potential harmonic instability is either from the incoming supply or the high-frequency switching drives. Therefore, the primary measurement points for this PQ assessment install at the Main Incomers and the VVVF Drives Input. Figure 4.9 and Figure 4.10 shows the results for Total Harmonic Distortion (THD) and Harmonic spectrum, which were recorded for 16 hours within 30-sec intervals at the Main incomers.

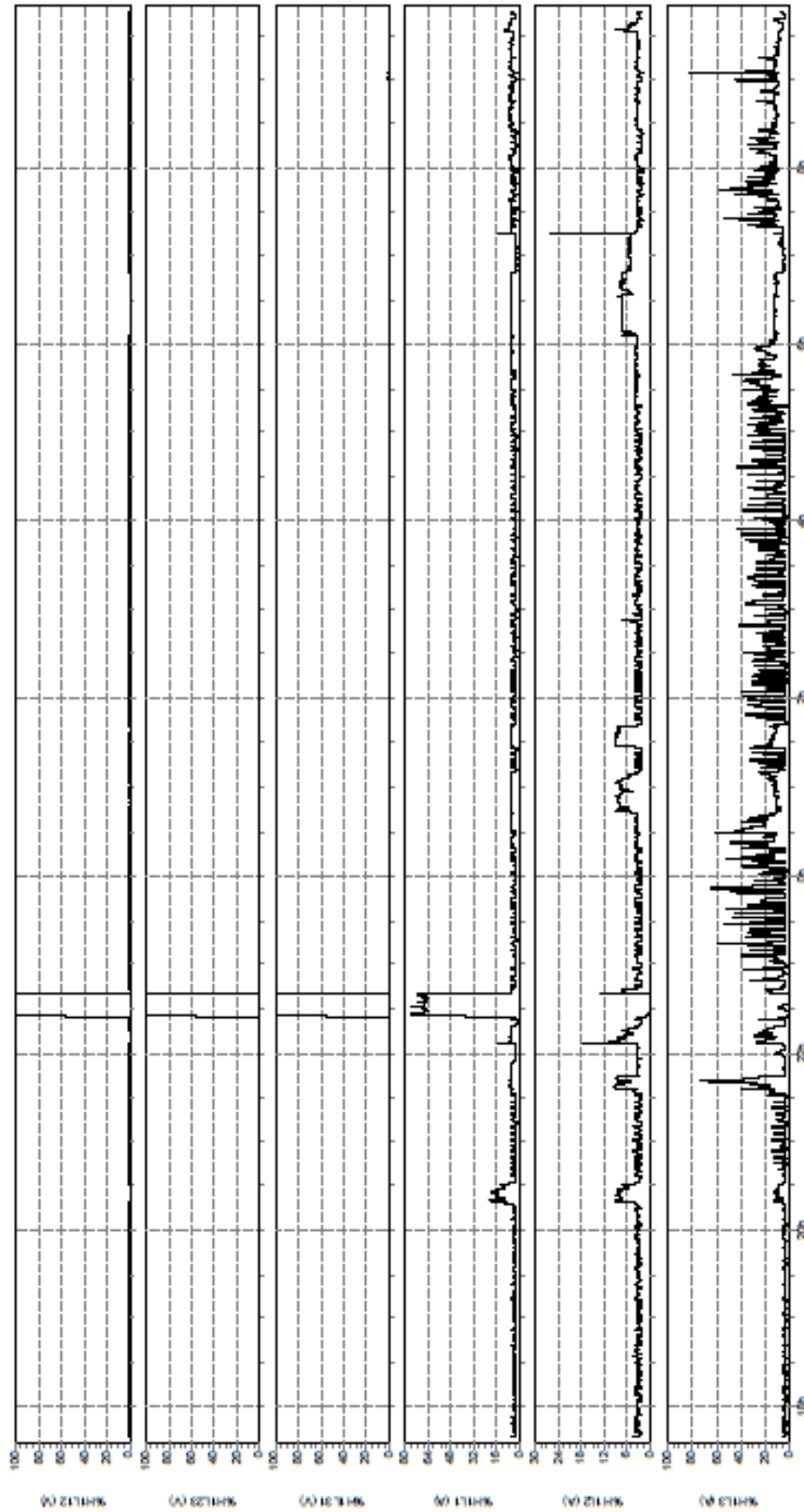


Figure 4.9: Total Harmonic Distortion (THDV & THDI) at Main Incoming (1250kVA Transformer Secondary)

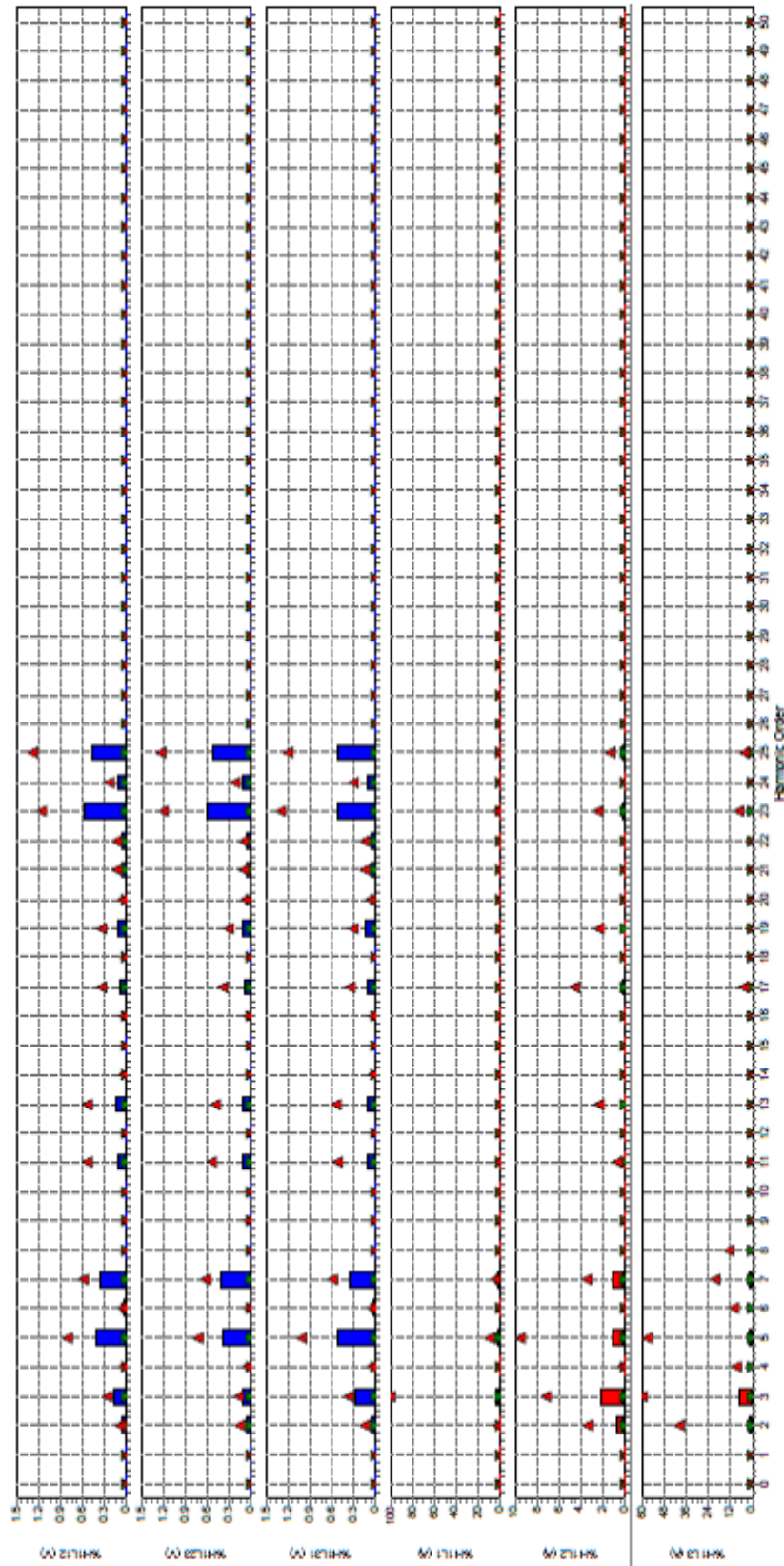


Figure 4.10: Harmonic Spectrum at Main Incoming (1250kVA Transformer Secondary)

Table 4.4: Data Tabulation for CUL-1 Main Incomer

Parameters (Main Incomer)	Units	Minimum	Maximum	Average
Voltage	V_{ab}	385.9 V	424.0 V	411.2 V
	V_{bc}	216.8 V	423.8 V	382.4 V
	V_{ca}	241.8 V	424.1 V	381.5 V
Current	I_a	32 A	1718 A	525.8 A
	I_b	46 A	1870 A	639.9 A
	I_c	30 A	1568 A	217.6 A
Voltage Unbalance	% Unbalance	0.10 %	41.60 %	2.60%
Current Unbalance	% Unbalance	8.80 %	16.70 %	52.80%
Total Harmonic Distortion Voltage (THDV)	$THDv_{ab}$	0.40 %	1.60 %	1.00 %
	$THDv_{bc}$	0.30 %	1.50 %	1.00 %
	$THDv_{ac}$	0.50 %	1.90 %	1.10 %
Total Harmonic Distortion Current (THDi)	$THDi_a$	1.90 %	20.80 %	4.50 %
	$THDi_b$	1.80 %	25.80 %	3.70 %
	$THDi_c$	1.90 %	82.20 %	12.00 %
Total Demand Distortion Current (TDDi)	$TDDi_a$	0.10 %	2.60 %	1.30 %
	$TDDi_b$	0.10 %	1.80 %	1.20 %
	$TDDi_c$	0.10 %	2.20 %	0.80 %
Frequency	f	49.77 Hz	50.13 Hz	49.98 Hz

Table 4.5: Data Tabulation for CSU VVVF Master Drives Input

Parameters (Master-REGEN)	Units	Minimum	Maximum	Average
Voltage	V_{ab}	400.3 V	413.8 V	407.1 V
	V_{bc}	400.5 V	413.8 V	407.1 V
	V_{ca}	399.2V	413.8 V	406.2 V
Current	I_a	32.3 A	267.8 A	134.7 A
	I_b	36.7 A	270.8 A	139.2 A
	I_c	37.2 A	277.7 A	144.1 A
Voltage Unbalance	% Unbalance	0.07 %	0.20 %	0.15 %
Current Unbalance	% Unbalance	2.10 %	8.80 %	3.30 %
Total Harmonic Distortion Voltage (THDV)	$THDv_{ab}$	0.90 %	1.70 %	1.30 %
	$THDv_{bc}$	0.90 %	1.80 %	1.20 %
	$THDv_{ac}$	0.90 %	1.90 %	1.40 %
Total Harmonic Distortion Current (THDi)	$THDi_a$	4.00 %	25.70 %	11.20 %
	$THDi_b$	4.00 %	26.00 %	11.00 %
	$THDi_c$	3.80 %	22.90 %	9.60 %
Frequency	f	49.90 Hz	50.00 Hz	49.98 Hz

Table 4.6: Data Tabulation for CSU VVVF Slaves 1-REGEN Drives Input

Parameters (Slave1-REGEN)	Units	Minimum	Maximum	Average
Voltage	V_{ab}	387.7 V	410.0 V	405.7 V
	V_{bc}	387.7 V	410.1 V	405.9 V
	V_{ca}	386.6 V	409.3 V	404.8 V
Current	I_a	102.8 A	289.7 A	184.2 A
	I_b	109.8 A	296.1 A	190.6 A
	I_c	117.2 A	304.5 A	199.2 A
Voltage Unbalance	% Unbalance	0.12 %	0.19 %	0.17 %
Current Unbalance	% Unbalance	2.60 %	6.60 %	4.10 %
Total Harmonic Distortion Voltage (THDV)	$THDv_{ab}$	1.10 %	1.60 %	1.30 %
	$THDv_{bc}$	0.90 %	1.60 %	1.30 %
	$THDv_{ac}$	1.10 %	1.70 %	1.40 %
Total Harmonic Distortion Current (THDi)	$THDi_a$	3.50 %	10.20 %	5.60 %
	$THDi_b$	3.60 %	10.80 %	5.90 %
	$THDi_c$	3.40 %	8.70 %	5.00 %
Frequency	f	50.00 Hz	50.00 Hz	49.98 Hz

Table 4.7: Data Tabulation for CSU VVVF Slaves 2-REGEN Drives Input

Parameters (Slave2-REGEN)	Units	Minimum	Maximum	Average
Voltage	V_{ab}	399.7 V	409.9 V	405.2 V
	V_{bc}	400.1 V	409.9 V	405.4 V
	V_{ca}	398.7 V	409.0 V	404.3 V
Current	I_a	93.1 A	268.7 A	172.2 A
	I_b	98.7 A	247.2 A	177.1 A
	I_c	103.3 A	277.7 A	181.9 A
Voltage Unbalance	% Unbalance	0.15 %	0.20 %	0.17 %
Current Unbalance	% Unbalance	1.80 %	5.40 %	2.80 %
Total Harmonic Distortion Voltage (THDV)	$THDv_{ab}$	1.10 %	1.70 %	1.30 %
	$THDv_{bc}$	1.00 %	1.60 %	1.20 %
	$THDv_{ac}$	1.00 %	1.70 %	1.40 %
Total Harmonic Distortion Current (THDi)	$THDi_a$	4.10 %	11.20 %	6.50 %
	$THDi_b$	4.00 %	11.20 %	6.40 %
	$THDi_c$	3.80 %	9.10 %	5.60 %
Frequency	f	49.95 Hz	50.05 Hz	50.0 Hz

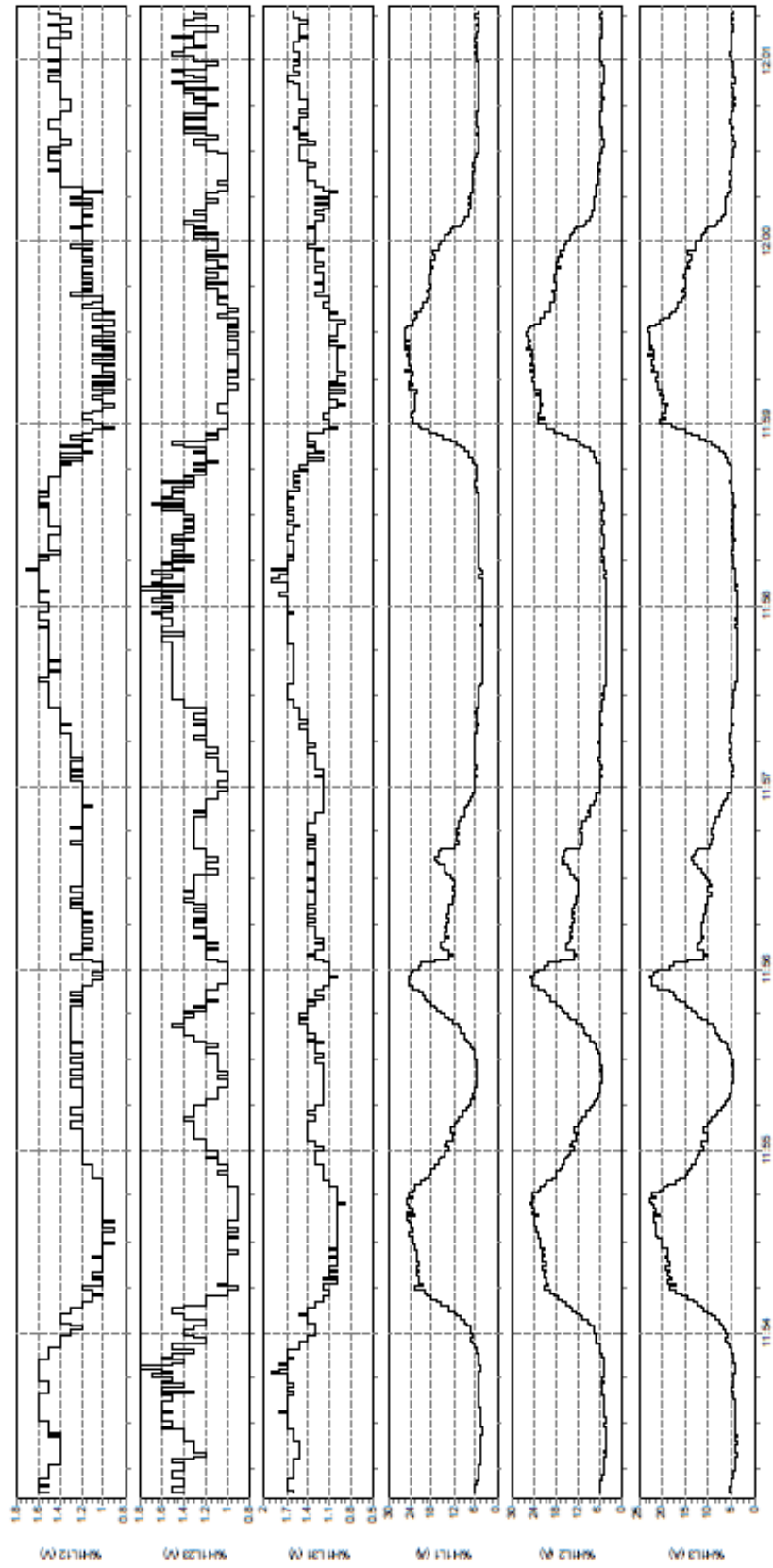


Figure 4.11: Total Harmonic Distortion (THDV & THDI) at CSU Machine VVVF Drives Input

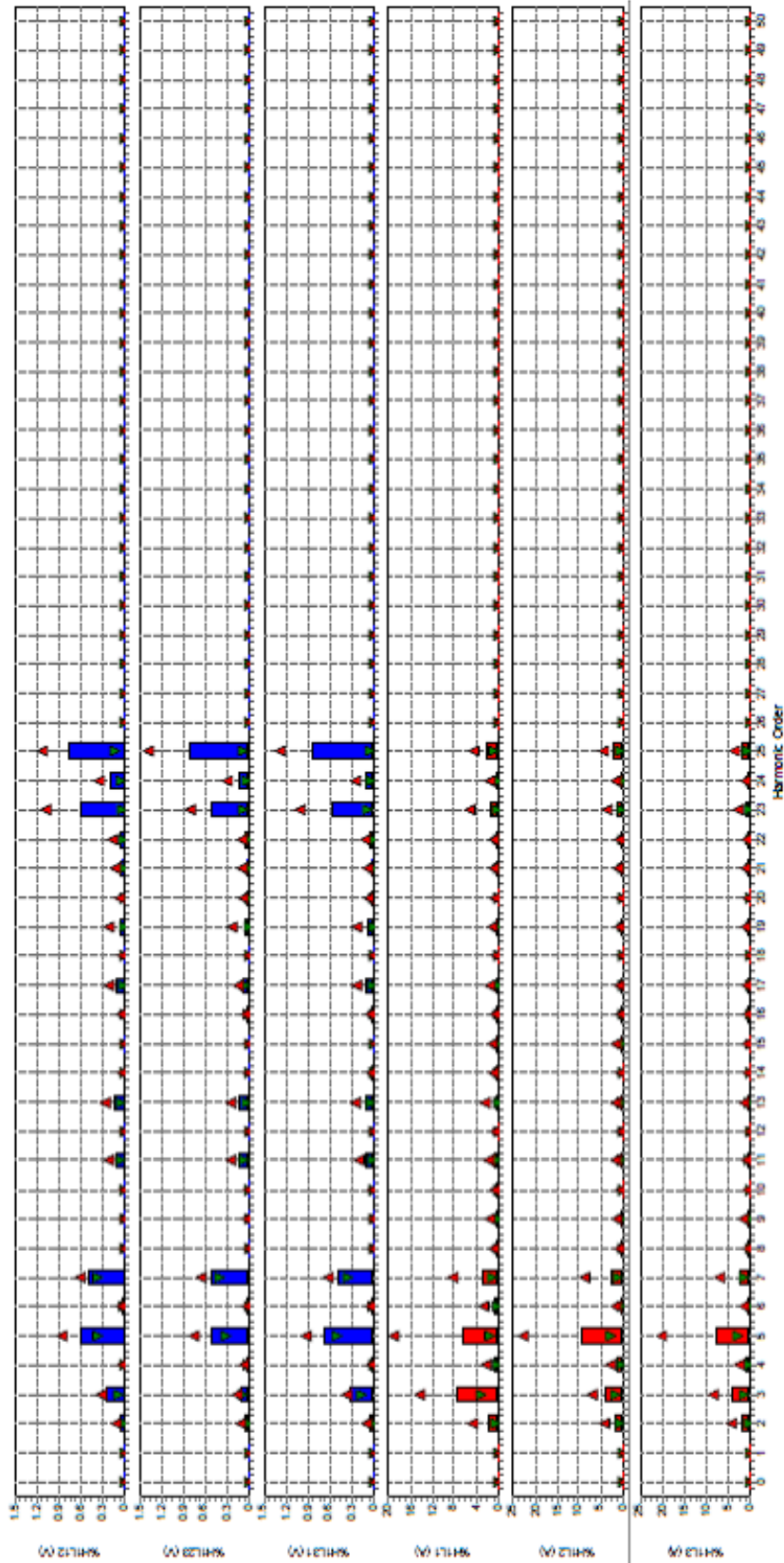


Figure 4.12: Harmonic Spectrum at CSU Machine VVVF Drives Input

From the results, it is proven that the harmonic distortion at the PCC is relatively low, with the majority of the measured period are below 2% for voltage distortion and demand distortion. There is an occurrence of power quality phenomena during this measurement period at PCC. However, the incoming supply voltage was drop at 216.8 V for voltage phase b-c and 214.8 V for voltage phase c-a. This large variation is due to incoming supply V_{bc} and V_{ca} was drop (under-voltage) to approximately 230V for 2.5 hours. Based on the data analysed in Table 4.4, the maximum loading condition at PCC is about 85.9 %, 93.5 % and 78.4 % for each phase respectively. It is also observed that the voltage fluctuated from 400 V to 415 V despite no occurrence of under-voltage. This variation could be an issue to the equipment that has lower voltage tolerance value and sensitive to voltage. Hence, equipment voltage tolerance should be verified in order not to operate beyond its operating condition.

On the other hand, the result for VVVF drives input is shown in Figure 4.11 and Figure 4.12 for both Total Harmonic Distortion and Harmonic Spectrum. From the results, it is proven that the harmonic distortions at the VVVF Master & Slave Drive Input are also relatively low, with the majority of the measured period are below 2%. The voltage and current are relatively constant for all three phases and there is no sign of unbalanced voltage compared to the Main Incomer results. A further PQ analysis is extended to the main load connected to the REGEN drives which includes the Bucket Elevator Motors and Rotary Feeding Table Motors. The result shows that all connected loads meet the IEC Std. 61000-3-4 limits for both voltage and current distortion.

4.4.3 Power Quality Assessment Summary

From previous discussion, it can be seen that the assessment results have indicated that the harmonic distortion at the PCC is relatively small, with a majority of the recorded measurement to be below 2% for voltage distortion and demand distortions. Thus, there is no evidence that could support the existence of Power Quality issues. In terms of harmonics and voltage unbalance, the results indicates that the measurement is within the allowable tolerance. However, the dilemma of high repeatable harmonic filters failures motivates the quest for alternative methods, which could consider external parameters in the system state change of actual event (Event Data) to the cause of the triggered events (Trigger Data) in constraint time [122]. This could be done by adapting the event modeller technique, to group the high

correlation system parameters together, to form an input-output relationship which is not limited to internal parameters only. As such, a homogeneous system of this high correlation will suggest the highest possible root cause of the harmonics problem while eliminating the unimportant parameters in near real-time.

4.5 Industrial Data Collection Setup

In the previous section, it has been explained that there is no evidence of PQ disturbance signal, which results in highly frequent harmonic filter failures at the CSU machine. The following section elaborates on how to consider the event modeller technique, to explore the known and unknown parameter and its relationship within the environment system, which may potentially create a relationship of harmonic failure between them. In order to comply with the existing infrastructure, some modifications are required to ensure all essential information is kept within the SCADA system environment.

4.5.1 Limitation on the Existing Infrastructure

The CSU machine is well equipped with the standard requirement of electrical protection, controls and machine safety. As discussed in Section 4.3.3, controller is a hardware that decides on the status of the machine with respect to the pre-programmed logic in the memory. The controller which synchronised with the SCADA system will monitor the parameters within the network and provides the user with an alarm if there is any abnormality occurring within the tolerance value. Any condition which does not meet the safety requirement shall interrupt the system and alerted the operator to further make a decision. This includes the electrical parameters such as Instantaneous Voltage and Instantaneous Current, which were measured using multi-function meters device, via a current transformer installed within the switchgear panel. Even though the multi-function meter is capable to provide other electrical parameters such as frequency, power, power factors, metering and power harmonics measurement, due to the limitation on the analogue output up to two connections, only voltage and current data is transferred to the controller. There is a need to import the Total Harmonic Distortion (THD) value to the controller, to find a relationship that could lead to harmonic failure within the CSU machine operational state. On the other hand, there is also influence from the environment,

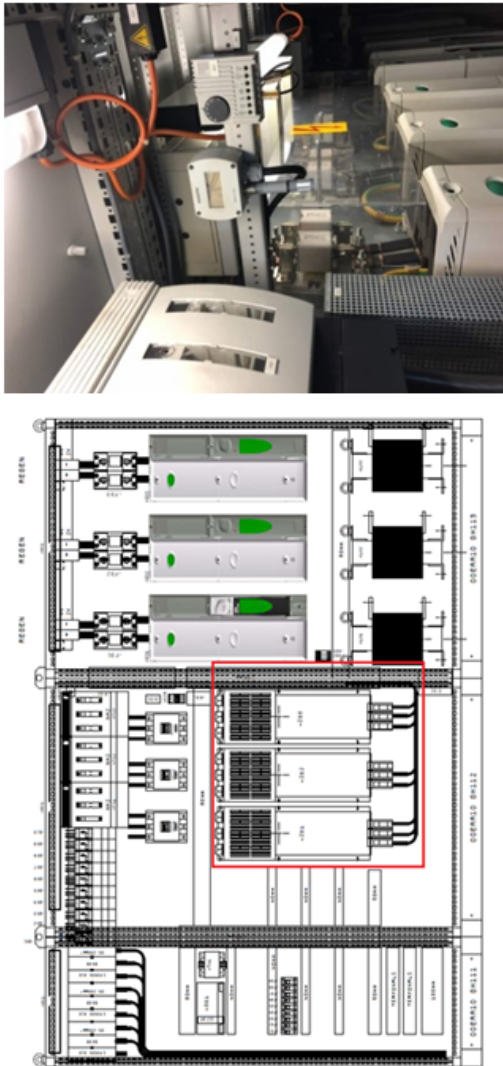


Figure 4.13: Panel Temperature and Humidity Sensor Installation

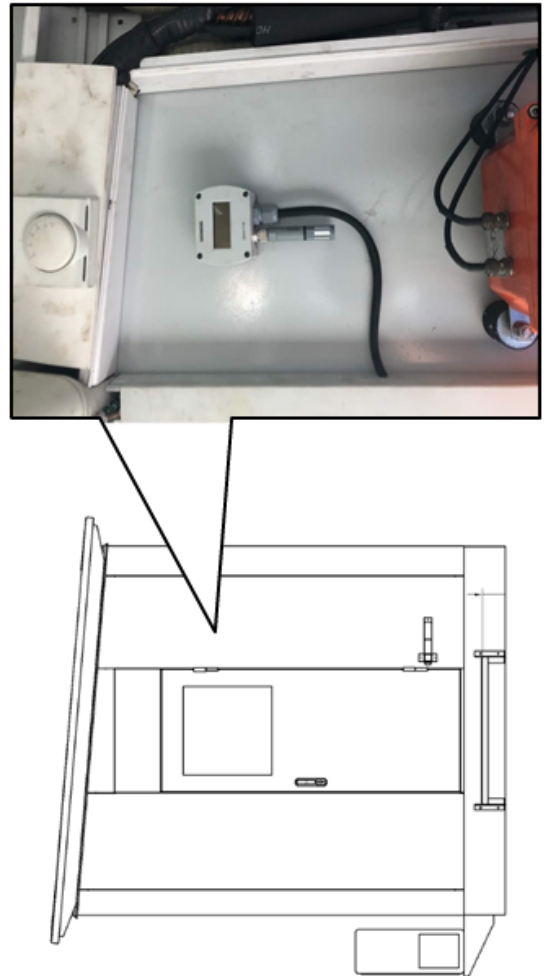


Figure 4.14: E-House Temperature and Humidity Sensor Installation

such as ambient temperature and humidity, which can influence the failure of the harmonic filters. Having all these internal and external parameters within the software in the loop, it shall map the parameters that cause the ineffectiveness of the machine power system. To overcome the limitation on the current infrastructure, two steps prior to the data collection progress has been implemented as follows:

4.5.2 Installation of New Relative Humidity & Temperature Room Sensor

The existing panel in the electrical house as per Figure 4.13 is installed with a Siemens QFA 317 Room sensors for relative humidity and temperature real-time measurement. The sensor is named as panel temperature and is used to measure both humidity percentage and absolute temperature of the electrical panel which contains the harmonic filter circuit and the REGEN drives. The same Siemens QFA 317 sensor is also installed near to the e-House entrance as per Figure 4.14 which is named as E-House temperature and humidity. All these four sensors are connected to the spare analogue input module to provide real-time data to the controller via the ControlNet connection. A tag name is assigned for each sensor and is being updated in the PLC-SCADA data acquisition system. During installation, additional cabling, wiring and termination between the sensor and the analogue input modules were installed during the maintenance window given by the company.

4.5.3 Upgrading the Analogue Output for the Multi-function Meter

The existing panel in GH103 is accommodated with the SOCOMEC A41 Multi-function Meter which provides the measurement data to the PLC. Due to the limitation on the analogue output module from the multifunction meter device, only voltage and current output were hardwired to the PLC input module. Since the primary purpose of this research is to investigate the frequent failure of the harmonic filter, it is necessary to consider installing another analogue output module (4825 0094) to cater for additional 2 output parameters to be hardwired to the PLC.

4.5.4 Network Scheduling

Network scheduling is essential to register/provide information to the ControlNet system that there are additional or changes to the module in the network. Each device on the ControlNet

network is assigned with a unique address. In this project, the Network scheduling is required to configure the replacement of Analogue Input Module from the Digital Input Module.

To configure these devices in the ControlNet networks, it uses the RSNetWorx™ for ControlNet application program. The fastest way to configure the ControlNet network is to let the RSNetWorx for ControlNet software browse the online network for ControlNet devices and automatically add them into your configuration; and then, let the Scan list Configuration Tool automatically configure connections to the target devices. Alternatively, the following steps could also be taken:

1. Browse for the existing online network
2. Configure the ControlNet Network and save them.
3. Start the Scanlist Configuration Tool
4. Define a connection target and save the schedule
5. Diagnose and troubleshoot the online network
6. Schedule network diagnostics

As a result, network scheduling activity has been successful. The newly installed sensors have been powered up and tested. The result shows that the sensors are providing the actual value to PLC Data Acquisition.

4.5.5 Data Logger Configuration

As noted earlier, data collection is an important process for research. It includes various activities which should be considered by the researcher prior to the actual implementation of the research instruments. The data collection strategy consists of offline historical data and online real-time data. Initially, the application needs to have access to historical data for machine learning attributes while existing data is for prediction purposes. However, it has been noted that the existing historical data store in the SCADA system does not have sufficient information needed to run the event modeller. The existing system only stores critical data, while most of the required 24 output variables and 96 input variables are not in the list. Therefore, a continuous PLC Data Logger for the CSU machines has been set up to capture new data. Total of 120 data has been imported to the server using Messaging (MSG)

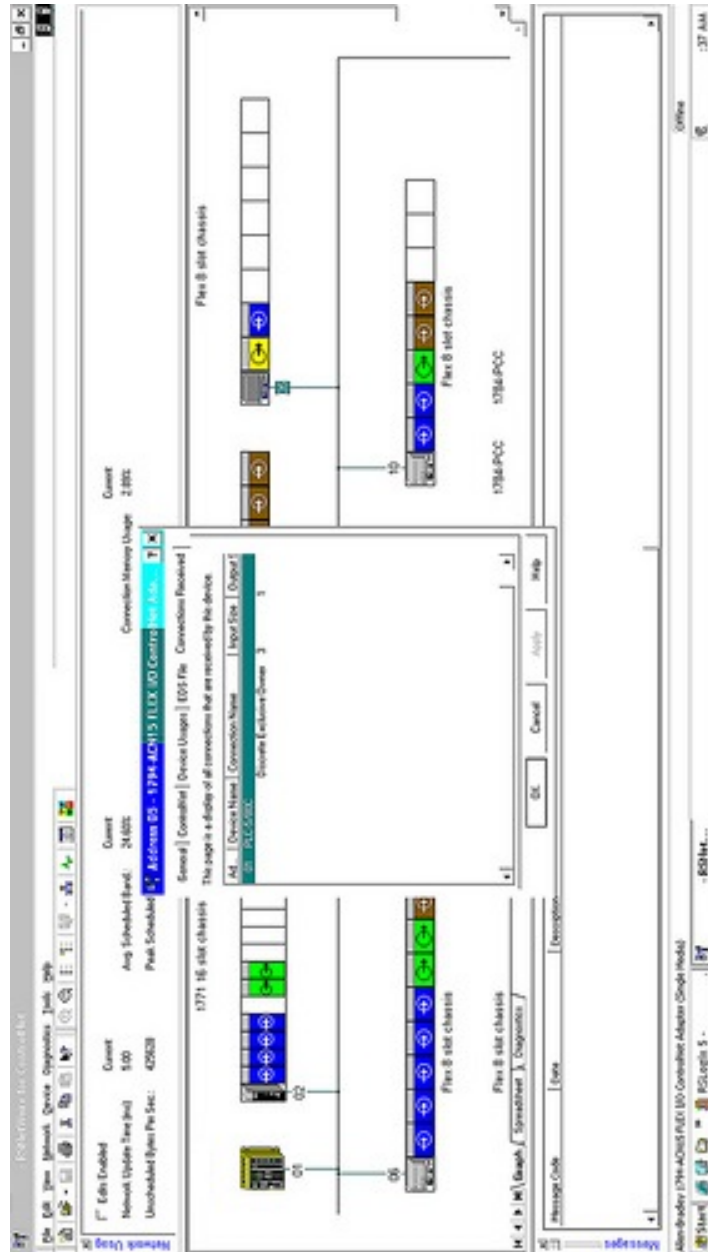


Figure 4.15: Network Scheduling for ControlNet

operand. A modification on the PLC logic has been established to configure this. The data is sampled each day for the period of 14 months from June 2018 till August 2019. Each day, the sampling frequency is for every second, which logged approximately 11,923,476 data. The data was provided in the form of .DAT files before it is converted into an excel sheet by using FactoryTalk View File Viewer software.

4.6 Summary

This chapter provided an overview of the CSU machine, which will be used as a case study to solve the mystery of the escalating pattern of harmonic failure. The outcomes of the PQ assessment have also been discussed in this chapter to demonstrate that there was a gap in the research study. The chapter reviewed the existing system configuration and proposed a data collection set up to collect all system parameters that will be used as proof of concept for the development of the EMDA in the following chapters.

Chapter 5

Event Modeller on Software in the Loop

5.1 Overview

There is a need to propose a Real-Time Software in the Loop (SIL) framework that identifies the inputs which have a significant impact on the efficiency and effectiveness of a power generator. The motivation here is to investigate whether there are other unknown factors that may have an influence on the efficiency and effectiveness of this power generator. Based on the principle of system engineering and advent of big data analytics, an effort is made to investigate the potential unknowns and if possible, calibrate the knowns.

This chapter aims to present a real-time data simulation of CSU machine using event modeller technique. This technique will suggest a homogenised correlation relationship that represents the known and unknown parameters in three different environments which are pre-disturbance environment, k-disturbance environment and post-disturbance environment. A synthetic data is generated to have the same nature and properties as the real CSU machine. In this SIL architecture, eight TD's and eight ED's were selected to evaluate the correlation between them. During the k-disturbance environment, five ED's is replaced with a disturbance dataset, to observe the reaction to the causal relationship. To further examine the accuracy of the outcome results, two types of TD's which include Static TD and Dynamic TD is selected, to decide on a suitable approach for industrial data in Chapter 6. Thus, suggesting the applicability and suitability of this technique in a typical power system setting.

This chapter is divided into five sections. Section 5.2 explains the system implementation. Section 5.3 discusses the synthetic dataset arrangement followed by results and discussions in Section 5.4. Finally, Section 5.5 provides a summary of the chapter.

5.2 System Development Implementation

The essence of successful software development depends on how well the strategies are implemented. Software Development Life Cycle (SDLC) strategy has been proven to be successful and ease the software development task [266]. This includes the sequence of system requirements, design, development, integration and testing stage. Development of this real-time SIL has adopted this strategy to ensure the system development processes are systematic, well defined and trackable.

5.2.1 Software in the Loop Requirement

The purpose of a real-time simulation is to measure the suitability and applicability of the event modeller technique in the power system environment. To ensure the system incorporates the industrial case study which has been discussed in the previous chapter, the requirement specification has been set based on the following:

1. 8 Event Data (ED's). This synthetic event data is simulated using Normal Distribution which represents the CSU Machine Output Data which includes Voltage, Humidity, Harmonics, Slewing Movement, Luffing Movement, Travel Movement, Temperature and Wind Speed.
2. 8 Triggered Data (TD's). This synthetic triggered data is simulated using Normal Distribution which represents the CSU Machine Input data which includes Busy Slew, Busy Luff, Busy Travel, Busy Bucket Elevator, Slewing Motor Run Bit, Hydraulic Motor Run Bit, Travel Motor Run Bit and Bucket Motor Run Bit.
3. Threshold Setting. For the purpose of examining the Trigger Threshold, an arbitrary 5% threshold setting was set, that could later be adjusted.
4. Event Modeller Limit. This setting reduces the complexity of an input-output relationship by its correlation confidence level. An arbitrary 80% was set, that could later be

adjusted.

5.2.2 Software in the Loop Design

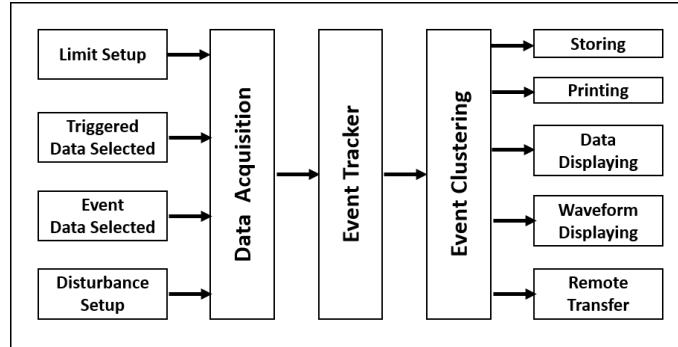


Figure 5.1: Overview of System Development

Figure 5.1 illustrates the overview of the system software development. The SIL is designed to acquire data from different sources determined by the user. Event data such as temperature and voltage is converted into a digital numeric value that can be manipulated by a controller. As such, triggered data that have a causal relationship with this temperature and voltage value is integrated into a data acquisition platform. This platform received other configuration setting such as limit and disturbance setup, to ensure the software development operates within the user specification. On the other hand, the Event Modeller algorithm discussed in Section 3.5 is coded within this software development using National Instrument LabVIEW Professional Development System version 17 (32-bit).

This algorithm is designed to operate in real-time. The Event-Driven Incidence Matrix (EDIM) sorts the rows for inputs and the columns for process outputs. Incidence matrix elements can take a value of 0 and 1. The value is 1 when both or neither of the input/output event data is triggered; otherwise, it is 0. This operation is similar to a logical Exclusive-NOR functionality [21]. Each change to the output in a given time span can be expressed as an event and the positive value of the inputs as triggers, thus output can be defined as Event Data (ED). Both Input t and $t-1$ can be considered as Trigger Data (TD). Equation 5.1 and Equation 5.2 shows the relationship between each event triggered by input at t and input at $t-1$ with respect to changes in output.

$$if(Input_{(t)} - Input_{(t-1)}) \geq \theta \xrightarrow{Trigger} TD_t \quad (5.1)$$

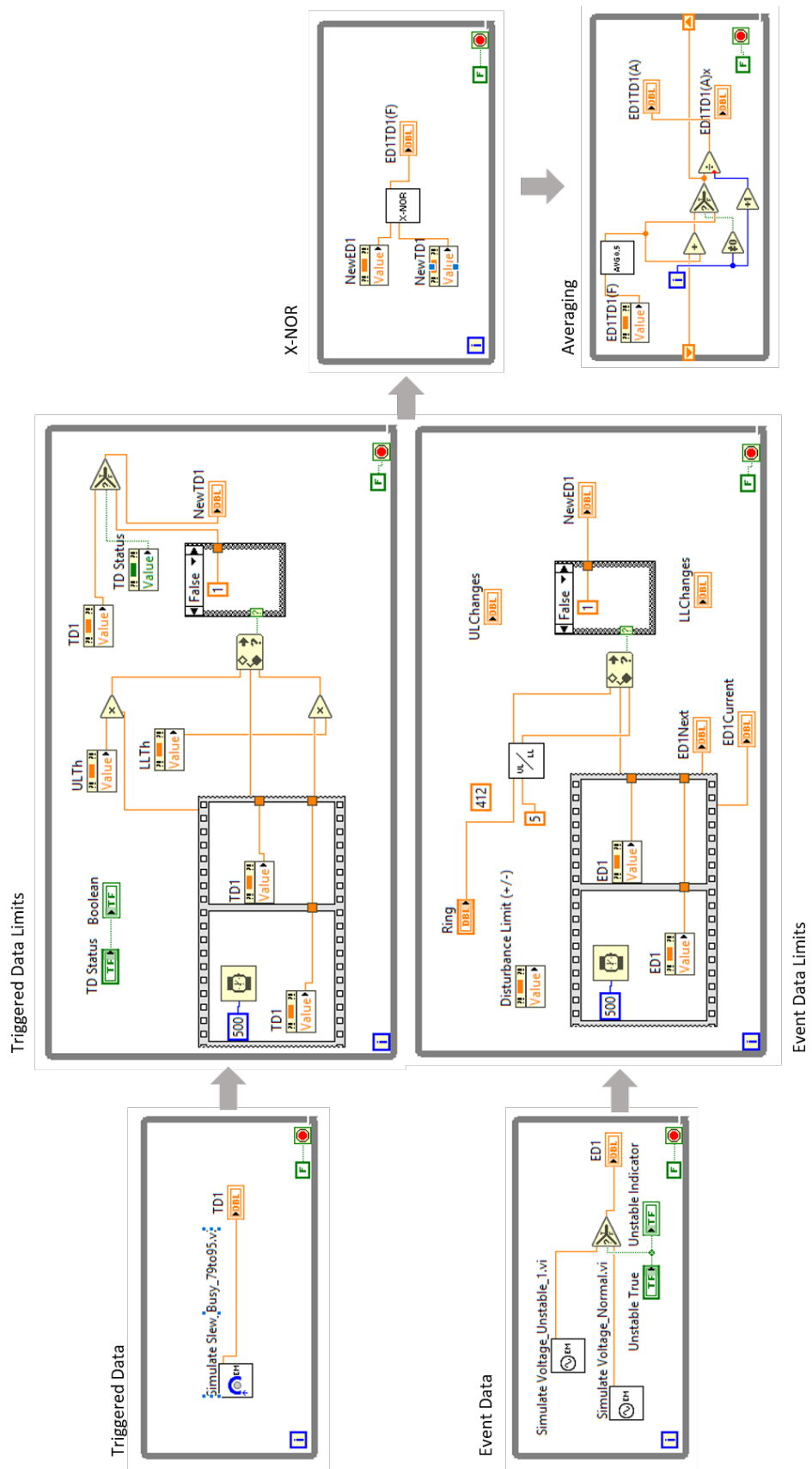


Figure 5.2: Snippets of Event Modeller VI - Block Diagram

$$if(Output_{(t)} - Output_{(t-1)}) \geq \Psi \xrightarrow{Event} ED_t \quad (5.2)$$

Snippets of the Event Modeller are shown in Figure 5.2. As defined in Table 3.1, Triggered Data (TD) is defined as “any input variable whose value transition registered an event”. In contrast, Event Data (ED) is defined as “the series of data that represent the state of the system at a given time”. In this example, the relationship between a simulated Slew TD (TD1) with respect to simulated Voltage ED (ED1) is examined. There are two types of TDs; Static TD and Dynamic TD. Static TD registered the original TD signal from the source while Dynamic TD multiply the changes of TD with a threshold setting defined by the system engineer. Both types of TDs were used in this experiment to investigate which source of data to be selected. This will ensure the event modeller algorithm provides an accurate sensitivity index or weight output. On the other hand, the weight of the simulated Voltage ED are calculated with a suitable range of pre-defined values determine by the system expert. The outcome of both ED1 and TD1 are correlated to each other using an Exclusive-NOR logic, followed by averaging step to update the sensitivity index of this causal-relationship in real-time. This is the example of the first iteration while the system is updated every 1000ms and presented in real-time. To evaluate the system robustness, two types of simulated voltage were introduced; normal dataset and disturbance dataset. The causal-relationship of each input-output pair is stored and updated in the layout display either in absolute value or waveform chart. Alternatively, the system was designed that the output can be print or transfer to other file or software.

5.2.3 System Integration

In order to develop a complex product or large engineering system, it is common practice to decompose the design problem into smaller sub-problems for easier handling [267]. The same strategy has been deployed for this system integration as we used a separate Virtual Instrument (VIs) for each subprocess. This system development borrowed the Flexible Data Input Layer Architecture (FDILA) [236] to combine all real-time data with discrete event simulation. This technique enables system managers to define data set-points within the Real-time Model Matching Mechanism (R3M) without requiring proprietary firmware system.

Figure 5.3 illustrates the overview of system integration architecture for the developed

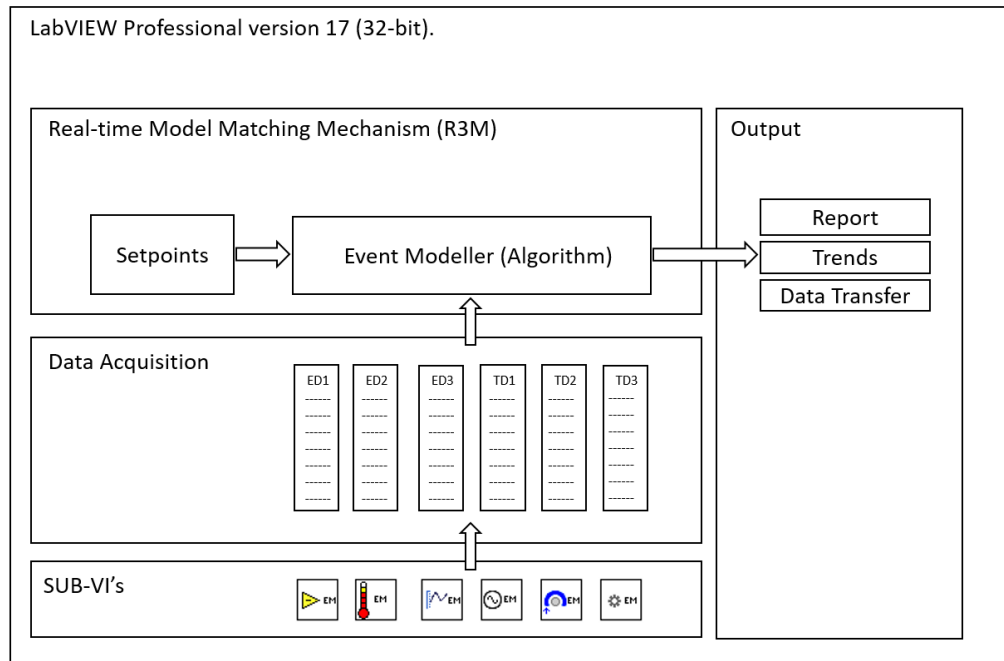


Figure 5.3: Overview of System Integration Architecture

Event Modeller SIL. The SubVI component represents the individual sensor or process data from the engineering layer. This real-time data is registered in a Data Acquisition component synchronously and later link with the R3M configuration. The system operates within LabVIEW software. The communication of the individual components runs via shared variable node, to pass and share data among other components. Snippets of the system integration are shown in Figure 5.4. All data is integrated into a single platform with a standard timestamp. For the purpose of synchronisation, a delay of 500 milliseconds has been set up. This is to avoid race conditions which potentially overlaps or overwrites data when a system attempts to perform two or more operations at the same time.

5.2.4 System Testing

System testing is deployed to check the behaviour of a complete and fully integrated software product based on software requirement specification. The main objective of this testing is to evaluate the system's compliance with the specified requirements, which has been discussed in Section 5.2. A test plan is produced to examine the functionality of the developed application related to the requirement specification. Table B.2 in Appendix B depicts the outcome of the

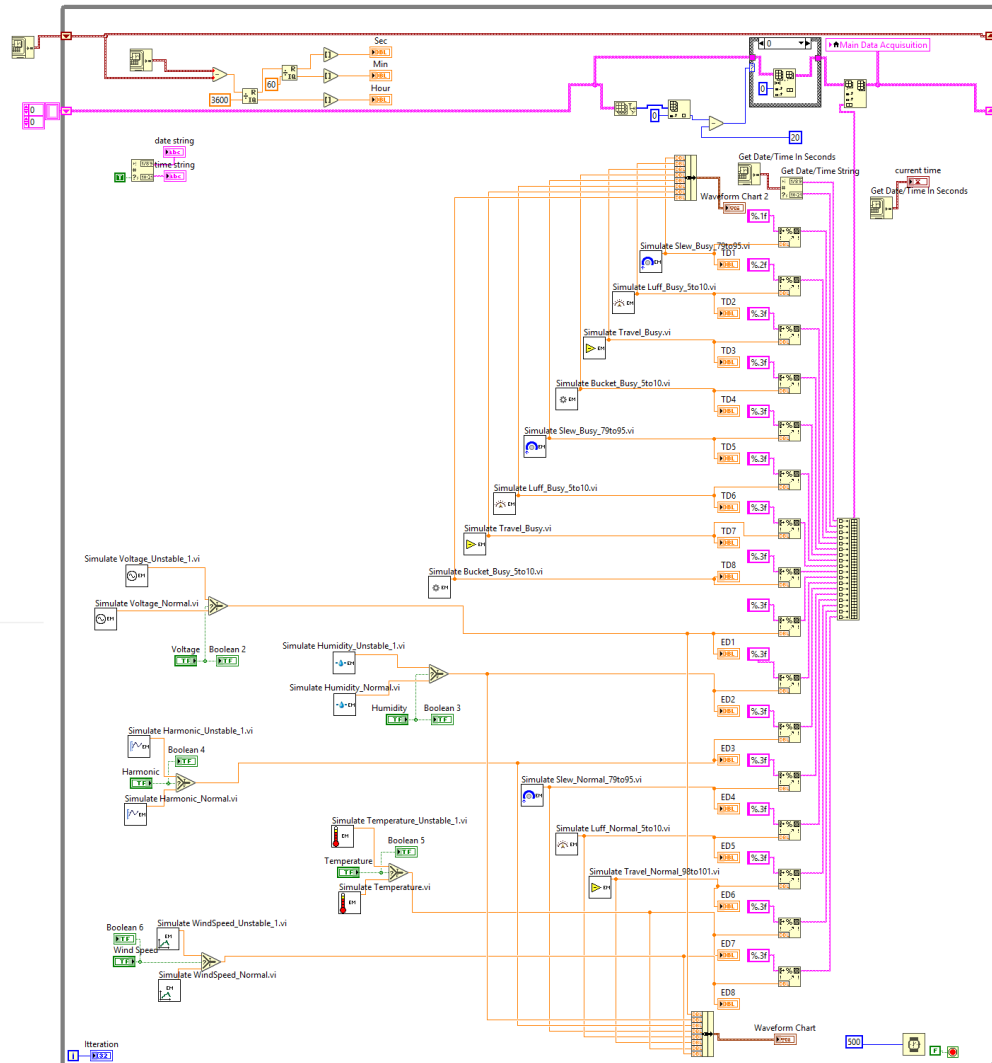


Figure 5.4: Snippets of System Integration VI - Block Diagram

test plan.

5.3 Methodology

Having discussed the system development implementation in the previous section, let's move on to the methodology and dataset arrangement that will be used in this experiment. The industrial system is a complex system. It has various types of data such as boolean, integer and doubles that represent different parameters, which includes electrical variables, machine positioning and environment measures. This requires data sampling from the machine before

it runs on the developed SIL system. The experimental methodology and the arrangement of the dataset will be discussed as follows.

5.3.1 Experimental Methodology

This experiment is intended to test the applicability of the event modeller technique built-in real-time applications. To evaluate this, 16 input/output system parameters of a CSU Machine were sampled from the Allen Bradley PLC-SCADA system using the FactoryTalk view. Once it is proven to be suitable, it will be extended to 120 parameters, which will be discussed further in Chapter 6. Data collection was conducted over a single day shift of a machine operation period, collecting approximately 43,200 lines of 12-hour data samples. The sampling frequency follows [268] by applying the right bandwidth, time constant and settling time. The requirement of the industrial specification has been discussed in Section 5.2.1. Having this data in hand, synthetic data was constructed to have the same nature and properties as the real CSU machine. The synthetic data used in this work are based on the multivariate normal distribution, which is a generalisation of the one-dimensional (univariate) normal distribution to higher dimensions. The dataset is then imported to the SIL system development, running the Event Modeller algorithm in National Instrument LabVIEW software settings (near real-time) as explained in Section 5.2.2. The overview of the experimental methodology is shown in Figure 5.5. Details of the dataset arrangement will be discussed in the following section.

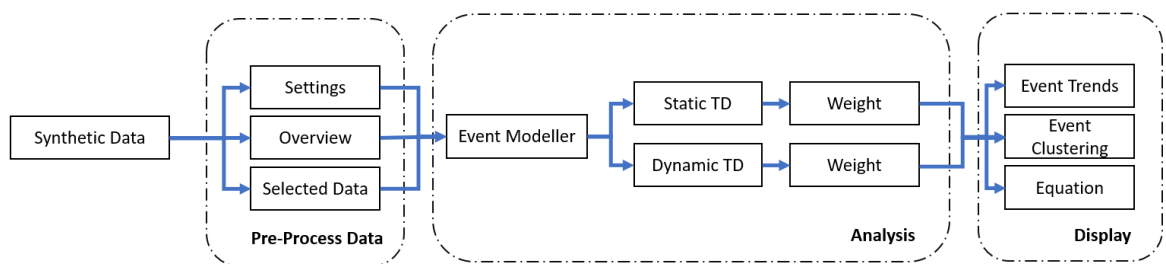


Figure 5.5: Overview of Experimental Methodology

5.3.2 Dataset Arrangement

The dataset arrangement for this SIL is divided into three categories which are Event Dataset, Triggered Dataset and Disturbance Dataset. Details of each dataset will be discussed in the

following subsections.

5.3.2.1 Event Dataset

Event Dataset (ED) is defined as a series of data that represent the state of the system at a given time [126]. Eight simulations ED's incorporating voltage, humidity, harmonics, machine positioning (slewing, luffing and travelling), temperature and wind speed is selected to represent the actual system parameter of a CSU machine. This simulation data is generated by considering three main factors which comprise of sensor data, operational pattern and environment influence. Data from the same sensor will not change much over the range, and it is close to its previous data. A model for a sensor comprises of the mathematical relationship between the input variable data, the prediction data and the threshold setting [269].

Taking this into consideration, the electrical parameter of a CSU machine is generated according to the assessment result in Table 4.4. The positioning data is generated based on the actual operational pattern of a CSU machine in 3-axis, operating at 100 % capacity. The environmental data is generated based on the geography of the installed machine. It considers Malaysia's hot and humid climate throughout the year and the possibility of strong winds as the CSU machine is installed by the seaside. With further advice from the system expert, details of the ED is outlined in the following Table 5.1.

Table 5.1: Event Data Simulation Parameters

No	Description	Normal	
		Min	Max
ED1	Simulated Voltage (V)	409	415
ED2	Simulated Humidity (%)	39	45
ED3	Simulated Harmonic (%)	0.4	1.2
ED4	Simulated Slew (°)	78	95
ED5	Simulated Luffing (°)	5	10
ED6	Simulated Travel (m)	98	101
ED7	Simulated Temperature (°C)	16	26
ED8	Simulated Wind Speed (m/s)	0	8

5.3.2.2 Trigger Dataset

In Discrete Event System, any input variable whose value transition is registered as an event is defined as a Trigger Data (TD) [126]. Eight TD's simulatios embracing the machine status

and motor run feedback is selected to represent the actual system parameter for four main movements. This consists of slewing, luffing, travelling and bucket operation, which have a direct correlation to the ED's described in the previous section. System expert has consulted the machine pattern; thus the generation of this TD dataset is summarised in Table 5.2.

Table 5.2: Triggered Data Simulation Parameters

No	Description	Normal	
		Min	Max
TD1	Busy Slewing Operation	0	1
TD2	Busy Luffing Operation	0	1
TD3	Busy Travel Operation	0	1
TD4	Busy Bucket Elevator Operation	0	1
TD5	Slewing Motor Run Bit	0	1
TD6	Hydraulic Motor Run Bit	0	1
TD7	Travel Motor Run Bit	0	1
TD8	Bucket Motor Run Bit	0	1

5.3.2.3 Disturbance Dataset

The purpose of this simulation is to test the applicability of the Event Modeller algorithm in real-time handling data, thus observing the reaction of the system to the abnormal events. Five ED's signal, representing the internal and environment parameter, has been chosen. This includes voltage, humidity, harmonic, temperature and wind speed. The dataset is simulated based on random fluctuation with 5% disturbance limit from the normal steady-state using the normal distribution. Table 5.3 presents the Event Data Disturbance setting.

Table 5.3: Event Data Disturbance Parameters

No	Description	Disturbance	
		Min	Max
ED1	Simulated Voltage (V)	392	432
ED2	Simulated Humidity (%)	32	52
ED3	Simulated Harmonic (%)	0	3
ED4	Simulated Slew (°)	N/A	N/A
ED5	Simulated Luffing (°)	N/A	N/A
ED6	Simulated Travel (m)	N/A	N/A
ED7	Simulated Temperature (°C)	16	32
ED8	Simulated Wind Speed (m/s)	4	16

5.4 Results and Discussion

The simulation results from simulated data reveal the causal relationship that exists between the input-output parameter by suggesting an equation that represents the system state. It will also discuss the reaction of the event modeller output when a k-disturbance signal is introduced and how it re-bounced when the system is back to normal. A comparison between Static TD and Dynamic TD performance is discussed to determine the best option for further development. In this experiment, 5 ED's synthetic data, which represent the internal and environment parameter as per discussion in the previous section, has been chosen. The data runs in 3 stages with 5 minutes time interval, accumulating up to 15 minutes sampling time. Summary of each stage is explained in Table 5.4.

Table 5.4: k-Disturbance Signal Setting

Description	Stage	Description	Duration
Pre-Disturbance	Warm-up Stage	Normal Range Dataset	5 Minutes
k-Disturbance	Fluctuation of k-events	Disturbance Range Dataset	5 Minutes
Post-Disturbance	Reaction to k-Disturbance	Normal Range Dataset	5 Minutes

5.4.1 Pre-Disturbance Stage

The pre-disturbance refers to the warm-up stage, which is the machine's normal steady state. In this experiment, Static TD has been chosen while both ED and TD Dataset are uploaded in the developed application and runs for 5 minutes. These data is then set to trend synchronously while the Event Modeller output is monitored. As the Event Modeller limit is set at 0.8 (80% Confidence), a blue shaded zone depicted in Figure 5.6 represent a group of high correlation system parameters that have a significant impact on the system state in real-time. After the 5 minutes period, Equation 5.3 reveals the system state during the period.

$$\rho_{0.8} = (ED1, ED8, ED4, ED5, ED7, ED6, ED3) \times (TD2, TD6, TD5, TD3) + (ED2) \times (TD2, TD6) \quad (5.3)$$

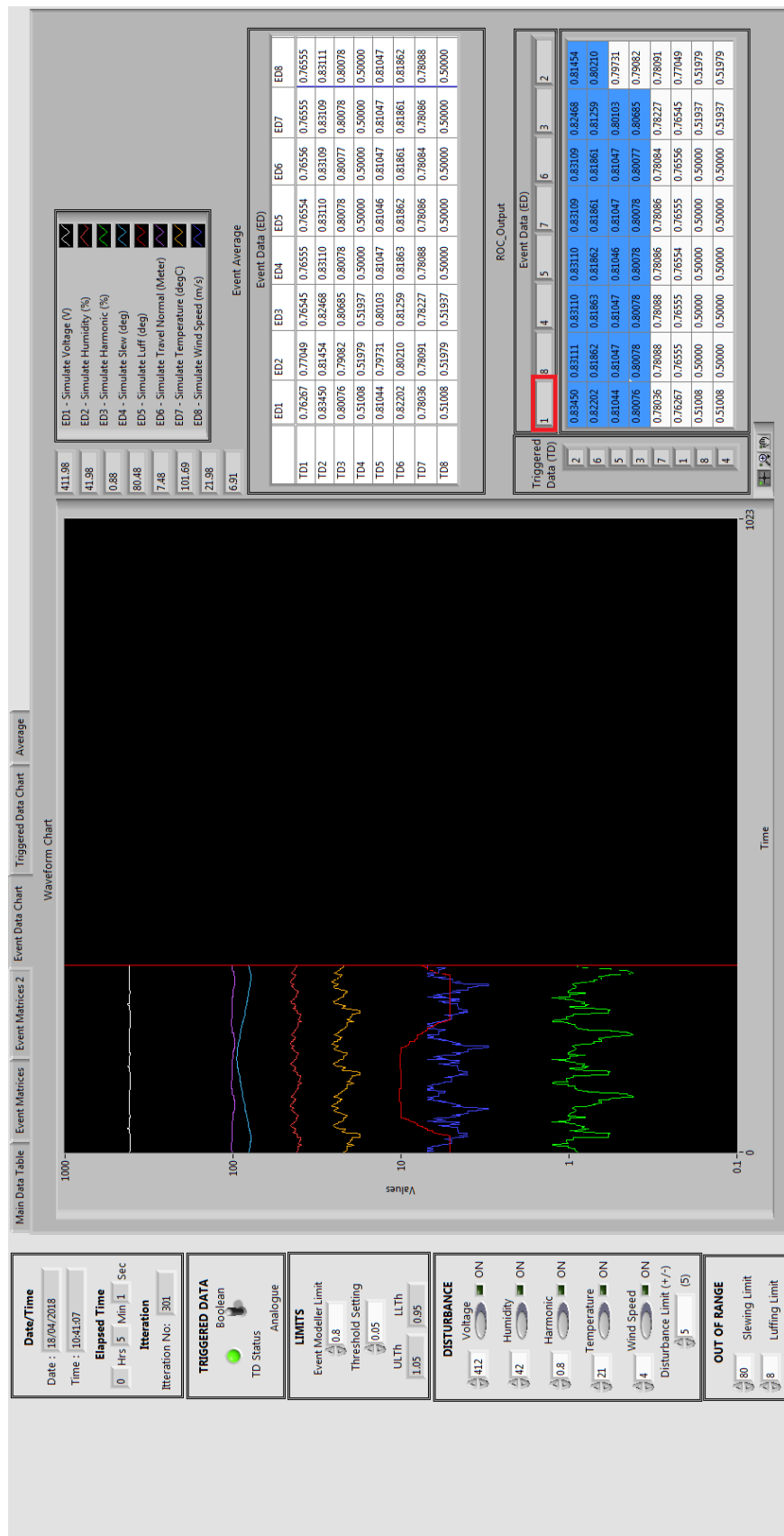


Figure 5.6: Snippets of Event Modeller VI - Block Diagram (Pre-Disturbance Stage)

$$\rho_{0.82} = (ED1) \times (TD2, TD6) + (ED8, ED4, ED5, ED7, ED6, ED3) \times (TD2) \quad (5.4)$$

To reduce the complexity of the equation, the Event Modeller Limit is amended to 0.82 (82% confidence) which is now expressed in Equation 5.4. This suggests that the Simulated Voltage (ED1) has a high correlation with Busy Luffing Operation (TD2) and Hydraulic Motor (TD6). Technically, the Luffing system requires the Hydraulic system for system operation, configuring the logic of the process. Other parameters such as Wind Speed (ED8), Slew angle (ED4) Luffing angle (ED5), Temperature (ED7), Travel movement (ED6) and Harmonic (ED3) also have high correlation with Busy Luffing operation (TD2) which suggest the main activity of the machine is in luffing operation rather than other operation.

5.4.2 k-Disturbance Stage

The k-disturbance refers to the fluctuation of the k-event Data, in such generating disturbance to the system. In this example, voltage disturbance dataset has been chosen and runs for the next 5 minutes. The data trending is performed on all measured parameters, while the reaction of the Event Modeller output is continuously observed. Figure 5.7 presents the outcome of the voltage disturbance after 5 minutes. A new equation which represents the system state is expressed in Equation 5.5. To reduce dimensionality, the input-output relationship has been increased to 82% confidence level, and the outcome expressed in Equation 5.6.

$$\rho_{0.8} = (ED8, ED5, ED4, ED6, ED7) \times (TD2, TD6, TD5) + (ED3, ED2) \times (TD2, TD6) + (ED3) \times (TD3) \quad (5.5)$$

$$\rho_{0.82} = (ED8, ED5, ED4, ED6, ED7) \times (TD2, TD6) + (ED3, ED2) \times (TD2) \quad (5.6)$$

What is interesting in this data is that the sequence of the ED's has changed drastically. The position of Simulated Voltage (ED1) in the causal relationship matrices moved from the first column to the last column. This can be observed from the red marked displayed in Figure 5.7. For e.g., TD2, the sensitivity index drops from 0.83450 to 0.76154, a reduction of 8.742%. This indicates that the introduced voltage disturbance has a significant impact on the system state.

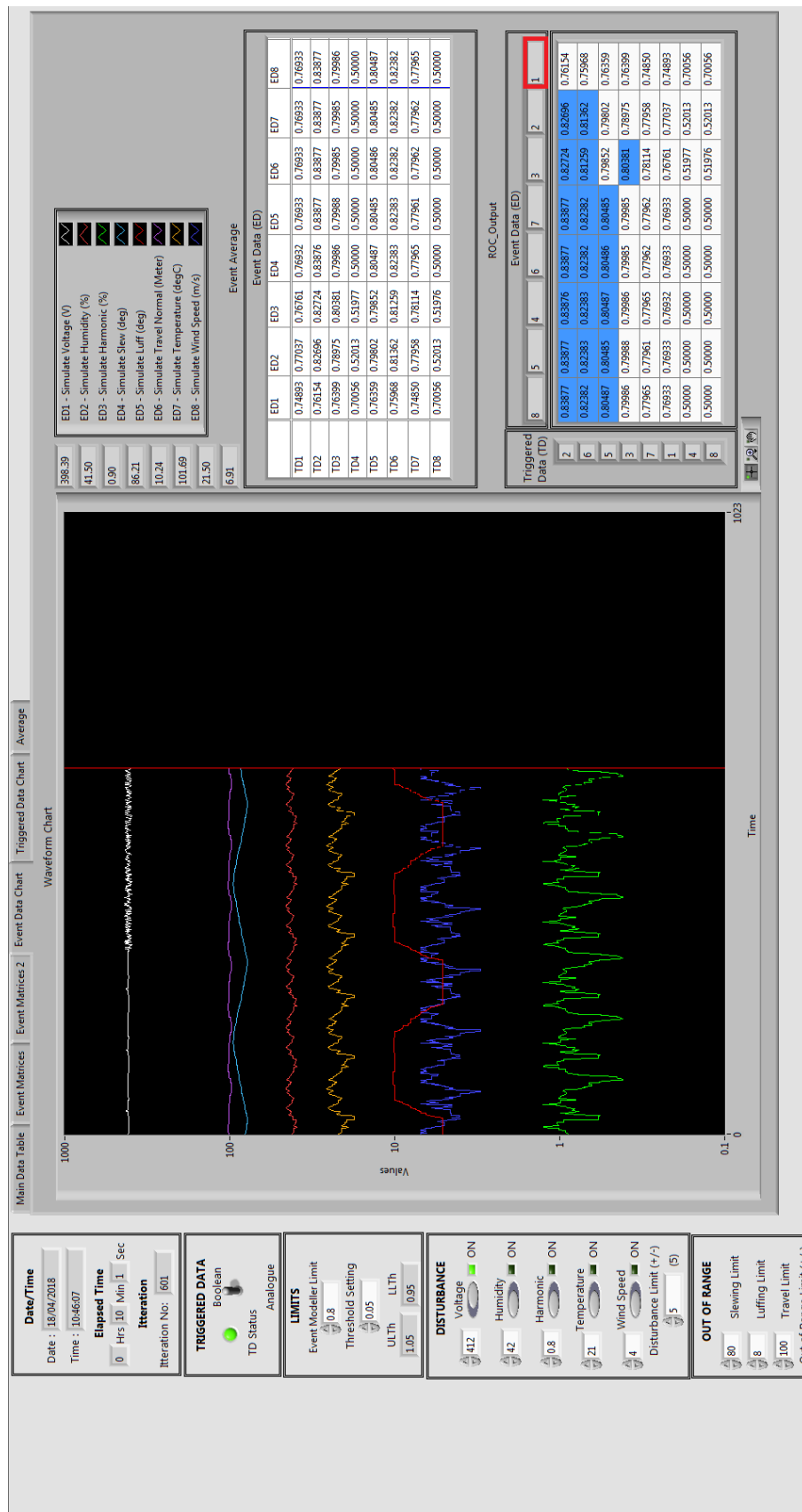


Figure 5.7: Snippets of Event Modeller VI - Block Diagram (k-Disturbance Stage)

5.4.3 Post-Disturbance Stage

The post-disturbance refers to the reaction of the abnormal system back to normal steady state. It is important to see how the sensitivity index settles down when the disturbance has been removed from the system. Figure 5.8 illustrates the outcome of the Event Modeller Output when Voltage disturbance has been removed after 5 minutes. The equations with 80% and 82% confidence level are expressed in Equation 5.7 and Equation 5.8 consecutively.

$$\begin{aligned} \rho_{0.8} = & (ED8, ED5, ED4, ED6, ED7) \times (TD2, TD6, TD5) + \\ & (ED3, ED2) \times (TD2, TD6) + (ED3) \times (TD5, TD3) \end{aligned} \quad (5.7)$$

$$\rho_{0.82} = (ED8, ED5, ED4, ED6, ED7, ED3, ED2) \times (TD2) \quad (5.8)$$

From these results, it is apparent that the sequence of the ED's doesn't change from k-disturbance to post-disturbance, but a significant positive correlation is observed for the disturbed ED. For e.g., TD2, the sensitivity index recovered from 0.76154 to 0.78132, an increase of 2.597%, which indicates a positive reaction from the disturbance system in a short period of time.

5.4.4 Overall Results

The experiment in the previous sections has been repeated for Dynamic TD while the performance between Static TD and Dynamic TD is evaluated to decide on the best option for the real industry data in the next chapter. Table 5.5 compares the results obtained from all ten experiments. To present the findings, the performance is measured from the score of the sensitivity index from two main perspectives, which includes the overall maximum and minimum weight, and the average score for each stage.

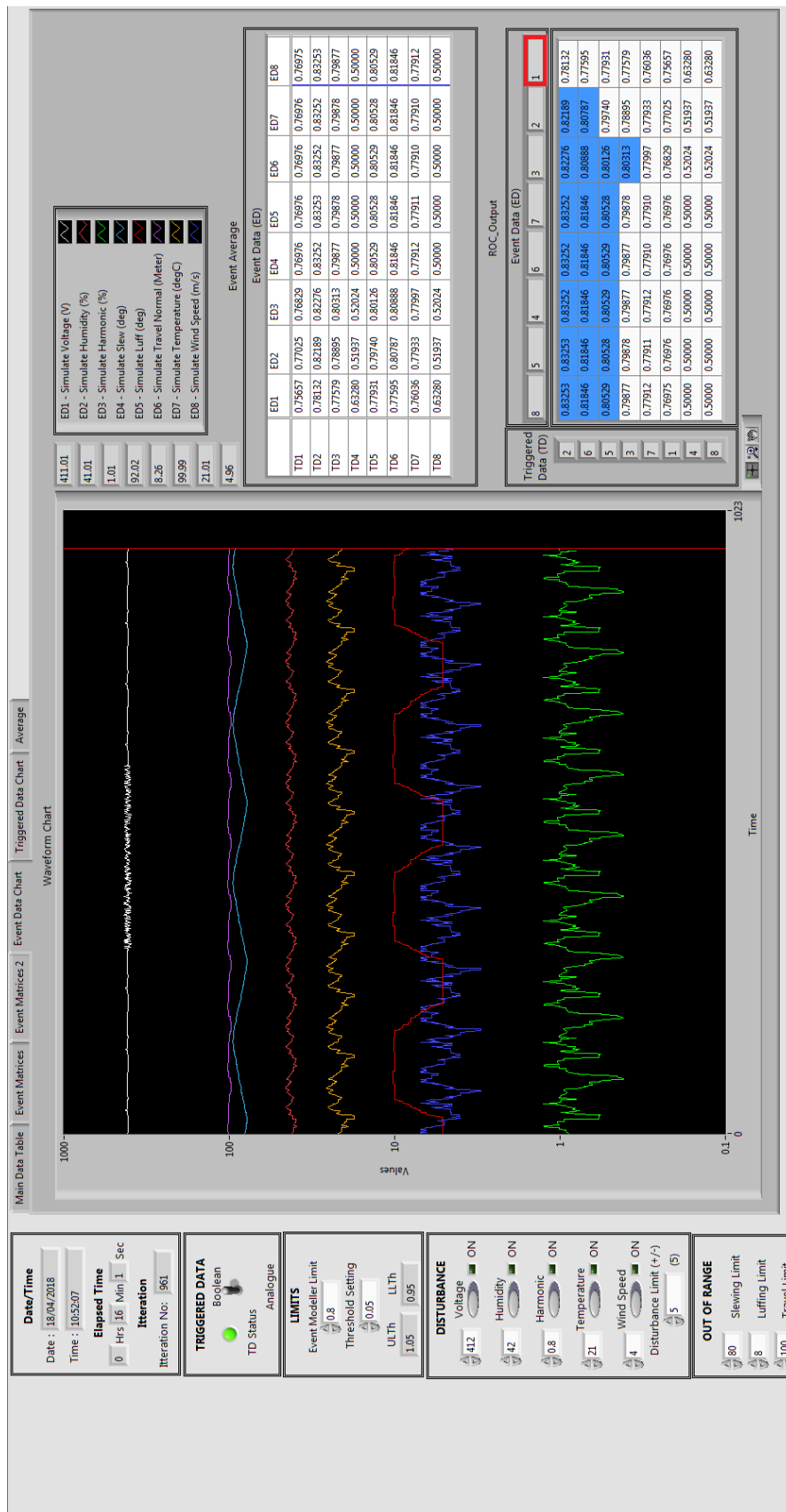


Figure 5.8: Snippets of Event Modeller VI - Block Diagram (Post-Disturbance Stage)

Table 5.5: CSU Real-time Data Simulation Results Based on Disturbance

Description	Voltage		Humidity		Harmonic	
	Static	Dynamic	Static	Dynamic	Static	Dynamic
Overall Maximum Weight	0.83877	0.99966	0.83646	0.99965	0.83702	0.99975
Overall Minimum Weight	0.50000	0.62878	0.50000	0.63719	0.50000	0.63609
k-Disturbance Weight - Pre	0.83450	0.98900	0.82877	0.97900	0.81817	0.97919
k-Disturbance Weight - During	0.76154	0.80654	0.77812	0.83042	0.76374	0.81377
k-Disturbance Weight - Post	0.78132	0.86671	0.80119	0.87890	0.78976	0.86503
% change from Pre to During	8.743%	18.44%	6.111%	15.18%	6.653%	16.89%
% change from During to Post	2.597%	7.460%	2.965%	5.838%	3.407%	6.299%

Description	Temperature		Wind Speed	
	Static	Dynamic	Static	Dynamic
Overall Maximum Weight	0.83604	1.00000	0.83408	1.00000
Overall Minimum Weight	0.50000	0.63128	0.50000	0.63252
k-Disturbance Weight - Pre	0.83063	1.00000	0.84240	1.00000
k-Disturbance Weight - During	0.79774	0.89764	0.78482	0.81037
k-Disturbance Weight - Post	0.80875	0.93103	0.80350	0.86546
% change from Pre to During	3.960%	10.24%	6.835%	18.96%
% change from During to Post	1.380%	3.719%	2.380%	6.798%

It is interesting to note that all ten cases of this study have a common finding. The transition of average sensitivity index (weight) from Pre-Disturbance to k-Disturbance have the same reduction trend with an average of 8.2% reduction for Static TD and 15.94% reduction for Dynamic TD. Alternatively, when the k-Disturbance recovers, there is also a trend of increased weight with an average of 3.218% increment for Static TD and 6.023% increment for Dynamic TD. This finding confirms the interrelation between the triggered data and the k-disturbance event data, where different TD types has different performance. The Dynamic TD has a higher percentage of changes compared to Static TD.

From the other perspective, it has been observed that the minimum weight for Static TD is fixed at 0.5000 while Dynamic TD has different minimum weight scores for each experiment. A possible explanation for this might be that the Dynamic TD captured the changes of input sensitivity rather than a direct Boolean signal from the input itself.

5.5 Summary

This chapter presented a real-time data simulation of Continuous Ship Unloader machine using event modeller technique, to identify the system state of the machine in three different environments which are pre-disturbance environment, k-disturbance environment and post-disturbance environment. The system development implementation and testing were explained followed by the generation of a synthetic data which represents eight TD's and eight ED's of a CSU machine. Two types of TD's including Static TD and Dynamic TD were studied to find the best method for providing accurate sensitivity index or weights. To assess the performance of the developed software in handling system disturbance in real-time, five out of eight ED's has been introduced with synthetic disturbance data while the results of the sensitivity indices were compared. Results have shown that the weight pattern and sequence had been affected when the disturbance signal is introduced, and it recovers when the disturbance signal is removed. This pattern will help system engineers to build a cause-effect relationship of events that represent the current system state conditions. In the next chapter, the system will be integrated with the actual plant data along with some training algorithms to assist system engineer in predicting any of this occurrence.

Chapter 6

Event Modeller on Industrial Data

6.1 Overview

Industrial control systems are becoming more complicated, which makes it very challenging to identify the root cause of a failure when abnormality or failure occurs within the system parameter. Integrated control system such as the one we have in CSU machines requires careful monitoring to ensure the issue of a frequent harmonic failure occurring within the machine is identified and resolved. Analysis of harmonic parameters discussed in Chapter 4 shows that the PQ system is stable. Hence, it is still unclear what actually caused this repeated failure. Applying EDMA technique has the potential to resolve this mystery. SIL system developed in Chapter 5 is able to formulate a multi-variable cause-effect input-output relationship representing the system state. A positive outcome from the findings motivates us to enhance with predictive analytics technique while expanding the complexity of the variables. This chapter presents a novel and dynamic platforms that connect the developed Event Modeller technique with a machine learning system, exploring the integration of this dynamic platform with Key Performance Indicator (KPI), in recognising the system operation pattern for prediction purposes. This whole data analytics technique expects to predict the homogenised system parameter that could present the current state of the system in real-time based on learning knowledge experienced in the past. Chapter 6 is divided into five sections. Section 6.2 explained the system development implementation. Section 6.3 outlines the data arrangement followed by results and discussions in Section 6.4. Finally, section 6.5 provides a summary of this chapter.

6.2 System Development Implementation

Event Modeller was tested in the search for homogenised system parameters, as shown in the previous chapter. Based on the findings of the SIL experiment, it is necessary to further explore the outcome of this technique in complex embedded systems through the use of real-time industrial data. Industrial data are genuine data. This takes into account uncontrollable factors such as noise, vibration, temperature and unforeseen interruption or malfunction during system operation [270]. Running this data along with the developed SIL system will leverage the information provided by the machine while gaining a new unknown parameter that could predict the machine's failure. Figure 6.1 below provides an outline of the proposed system development.

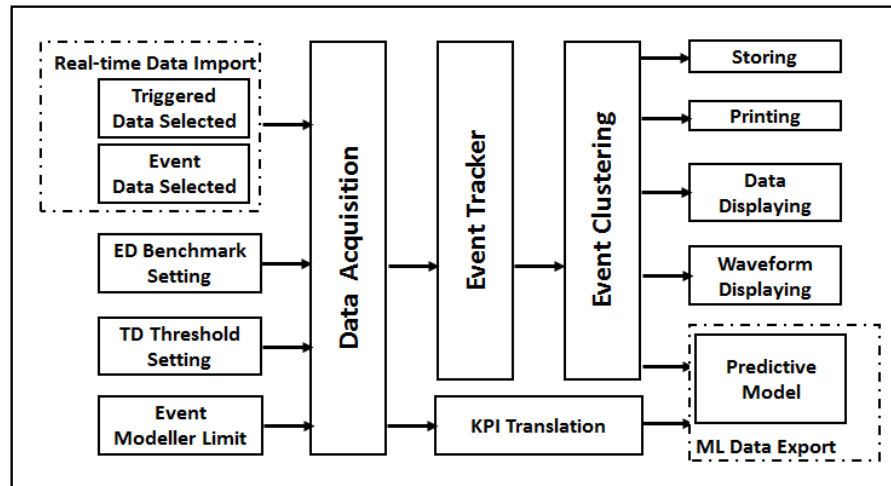


Figure 6.1: Overview of System Development

In comparison to the system development in Chapter 5, a KPI translation is introduced parallel to the event modelling algorithm, which gathers all data together, for exporting to the machine learning technique for predictive modelling purposes. The following sections describe the system development of a standard approach incorporating supervisory control and the DAQ method with predictive modelling. Again the SLDC strategy is adopted to ensure that the execution of this software development is well coordinated.

6.2.1 New Software-in-the-Loop Requirements

As explained in Section 6.1, the aim of this system development is to present dynamic platforms that link the developed Event Modeller technique to the machine learning system, taking into account the performance indicator of the machine. Such information promotes the systematic translation of engineering data into the knowledge management system [271], which could support machine learning analysis. The SIL requirement of the previous chapter is updated to include real industry data as follows:

1. Importing Data. This system is capable of importing real-time / historical data to different file types (.csv,.xlsx,.txt, etc.).
2. 24 Event Data (ED's). Such data are obtained every second for a period of 14 months. Table 6.1 displays the list of CSU Machine EDs by its cluster.
3. 96 Triggered data (TD's). Such data are obtained every second for a period of 14 months. The data was divided into four clusters with 24 TDs for each cluster. Table 6.3 displays the list of CSU Machine TDs by its cluster.
4. ED Benchmark Settings. The user has the option of selecting a benchmark setting for all EDs based on a system expert.
5. Threshold Setting. For the purpose of the Trigger Threshold, an arbitrary 5% was established, which could be adjusted later.
6. Event Modeller Limit. This setting reduces the complexity of the input-output relationship by its confidence level of correlation. An arbitrary of 95% was set, which could be adjusted later.
7. Exporting Data. This system is capable of exporting the output of the event modeller results to different file types (.csv,.xlsx,.txt, etc.). This data will be used in the machine learning algorithm.

6.2.2 Software in the Loop Design

The extended version of this SIL is designed to capture data from the machine in real time. However, due to a company policy that does not allow data to be live, data is collected using

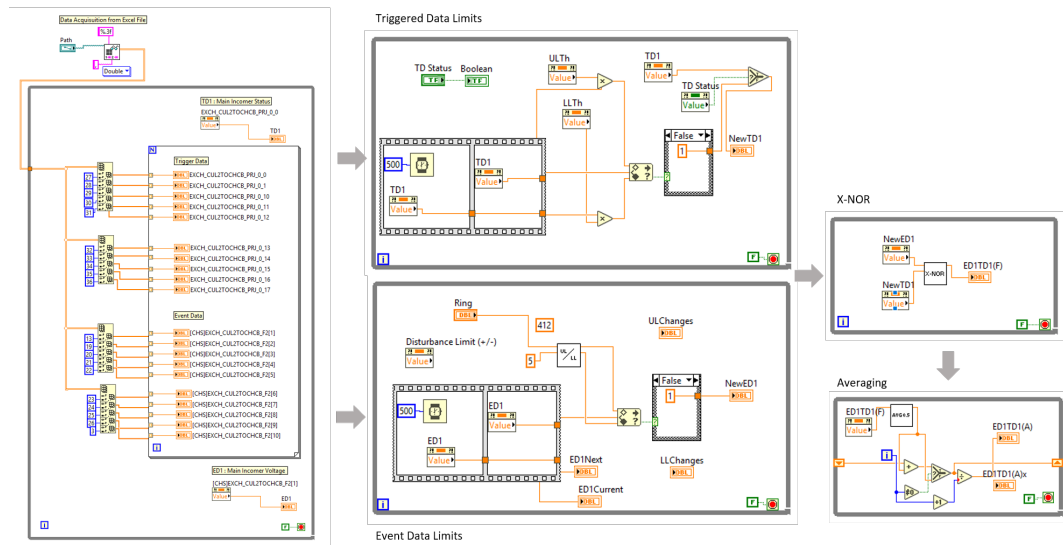


Figure 6.2: Snippets of Event Modeller VI- Block Diagram (Enhancement)

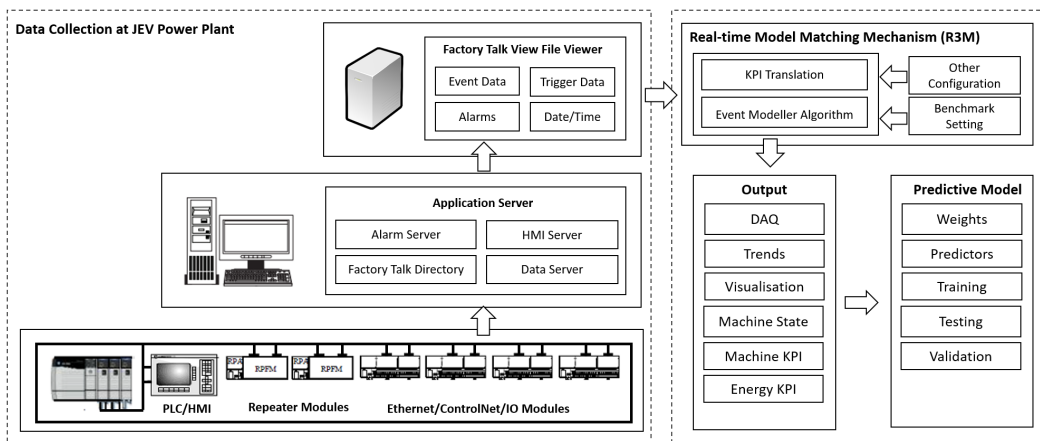


Figure 6.3: Control System Architecture

a data logger for a period of fourteen months from June 2018 to August 2019. Description of the data logger set up can be found in Section 4.5. The event modeller snippets are shown in Figure 6.2. The correlation between Main Incomer Status (TD1) and Main Incomer Voltage (ED1) is explained in this example. Both parameters are obtained from the data acquisition in the form of an extension file.csv. Setting the threshold and setting the benchmark for these parameters were first needed to be determined before running the application. The first TD1 and ED1 batch were multiplied simultaneously by threshold and benchmark-setting. Then, this result was compared to the second batch of TD1 and ED1 to set the index score of 0 (if the result is within the range) or 1 (if the result is out of range). The results of both ED1 and

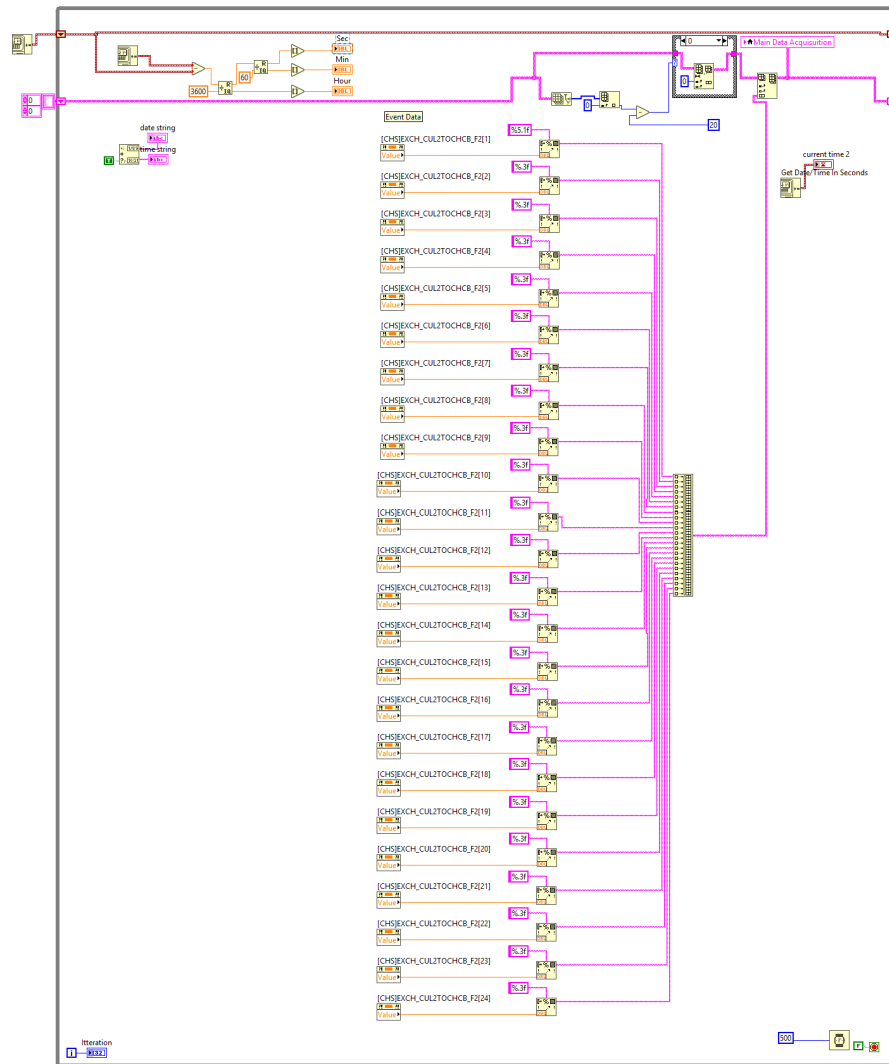


Figure 6.4: Snippets of System Integration VI - Block Diagram (Enhancement)

TD1 are then compared using an Exclusive-NOR logic, followed by an average step to attain the sensitivity index of this causal-relationship in real time.

6.2.3 System Integration

System integration is described as the process of linking systems, devices and programmes together in a common data sharing and exchange architecture [272]. The control system architecture for this project is shown in Figure 6.3. In this system development, field data such as sensors, actuators, limit switches and motors are connected to the programmable logic control (PLC) via input/output modules. In this example, the Allen Bradley controllers

used the Control-Net and Ethernet connections to communicate with each other and update the status of an application server in real-time. The application server stores all information assigned by the engineer, such as date, time, input data, output data, alarms and trends. As the intention is to collect data in real-time, an improvement has been made to the existing system architecture to incorporate all relevant data into the application server. The data were then retrieved using the Factory Talk File Viewer (FTFV) software, before it runs using the Real-time Model Matching Mechanism (R3M) introduced by [236]. R3M has built a bridge between historical data and a real-time simulation that operates based on the actual events. Figure 6.4 shows the snippet of system integration on the R3M platform. In this example, all EDs from industrial data were compiled into data acquisitions for trending purposes.

6.2.4 System Testing

System testing is deployed to check the behaviour of a complete and fully integrated software product based on software requirement specification. The main focus of this testing is to evaluate the system's compliance with the specified requirements which has been discussed in 6.2. A test plan is produced to examine the functionality of the developed application related to the requirement specification. Table B.2 in Appendix B depicts the outcome of the test plan.

6.3 Methodology

Having discussed the implementation of the system development in the previous section, let us move on to the methodology and dataset arrangement that will be used in this R3M application as follows.

6.3.1 Experimental Methodology

The CSU machine that has been chosen for the case study was having escalating trends of harmonic failure. As the control system within the machine is a complex system, it has a large number of variables that could be linked to this event. The dynamic causal relationship between these variables is considered to be the root cause of the failure. To access these variables, raw data (EDs, TDs), translated data (KPIs) and documents (daily reports) are used

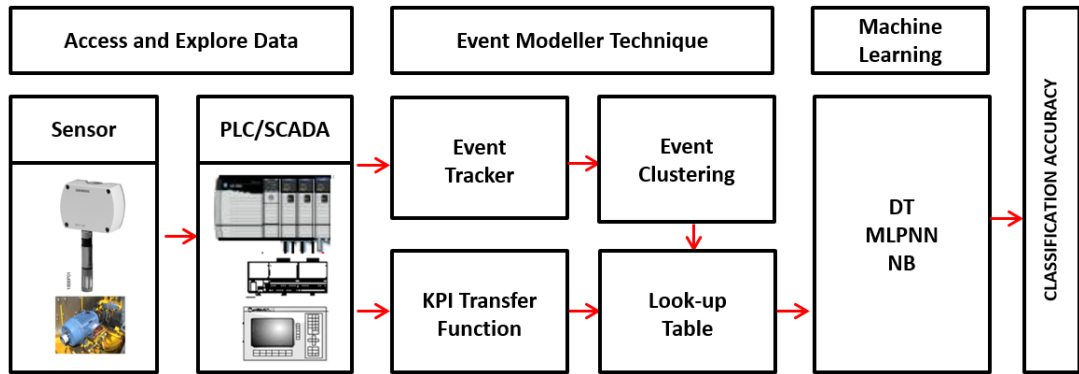


Figure 6.5: Event Modeller Data Analytics Technique

to build a predictive model that could represent the system state according to the pattern. Having the proposed Event Modeller Data Analytics (EMDA) technique in Chapter 3, the intention is to test the applicability of this technique in analysing the mystery of this repetitive harmonic failure. The following Figure 6.5 is the simplified version of the EMDA technique in Figure 3.2, which will be used in this Chapter. Later, two main experiments will be introduced; (1) Machine Failure Analysis with Key Performance Indicators and (2) Machine Failure Analysis with Predictive Model, with each experiment strategy will be discussed further in Section 6.4.1 and Section 6.4.2 consecutively.

6.3.2 Data Collection

In this experiment, the industrial dataset of the CSU machine was collected from the Allen Bradley PLC-SCADA system using the FactoryTalk view. The data was collected and stored in the data logger from June 2018 to August 2019 for 14 months. The collection of this data, which has been discussed in Section 4.5 were customised to suited the practical application of the CSU machine. Hence, to research the association between the different sensors and their triggering point, with a greater focus on events that could potentially cause the system failure. Since the data was collected offline, it is necessary to label it with meaningful information, to validate the data when it is re-run using the proposed method. Information such as plant and equipment status, weather conditions, maintenance work and breakdowns could be extracted from the system operator's daily log sheets. Table B.3 in Appendix B outlined the system operator log sheets for October 2018. The log sheet reports the operating activities such as

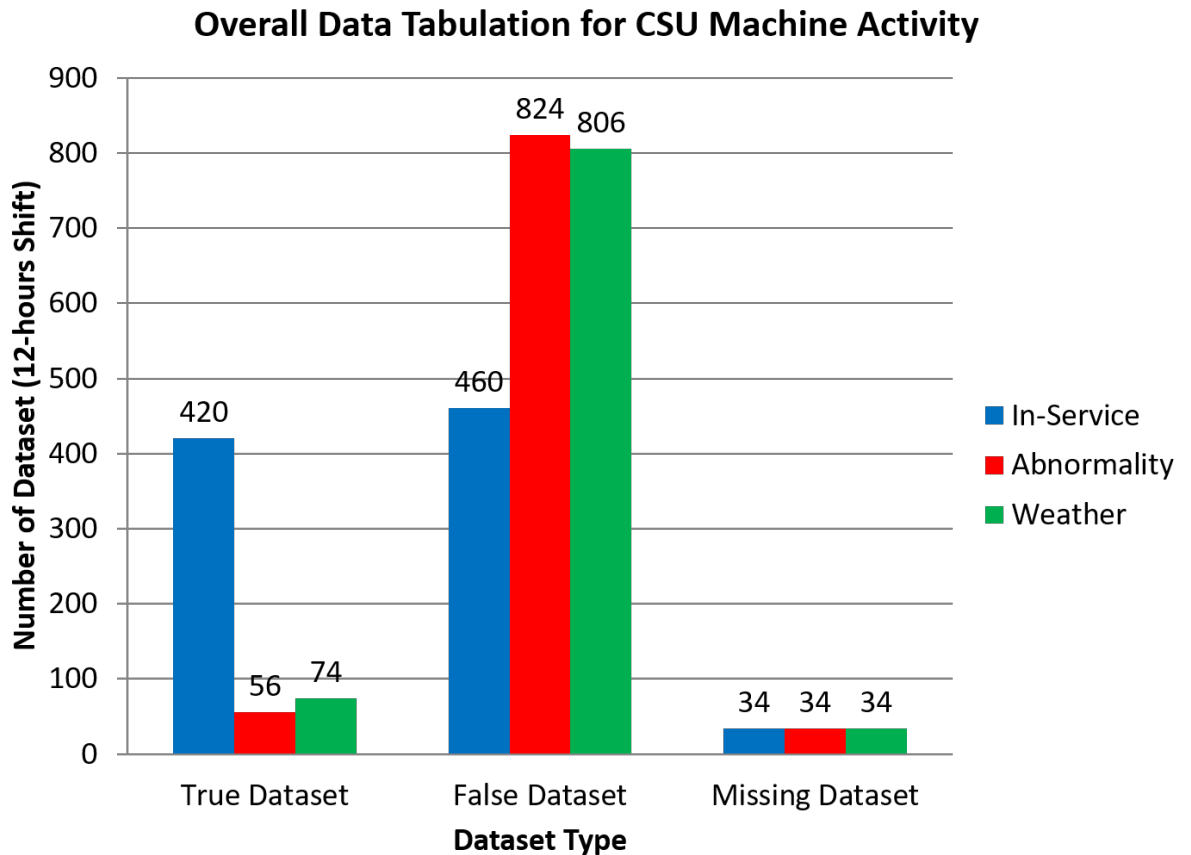


Figure 6.6: Overall Data Tabulation for CSU Machine Activity

dates of operation, log sheet number, CSU status, total unloading, abnormality and weather.

6.3.3 Dataset Arrangement

The complete data tabulation for the CSU machine is shown in Figure 6.6. True dataset and False dataset represent the two truth values of logic and Boolean algebra which associated with the occurrence of the event within the 12-hours shift operation. In contrast, the missing dataset represents the failure to retrieve the data from the system. There are 914 datasets with each dataset containing 120 sensors/seconds. This is equal to 43,200 lines of 12-hour data samples. To help with pattern recognition, the data was divided into three main categories known as in-service, abnormality and weather. In-service refers to the state of the machine that operates at its full capacity. The analysis of the data shows that this state represents 45.95% of the population. The chart also shows that only 56 of the 914 datasets reported abnormalities of the CSU machine during this period. This suggests that only 6.13% of

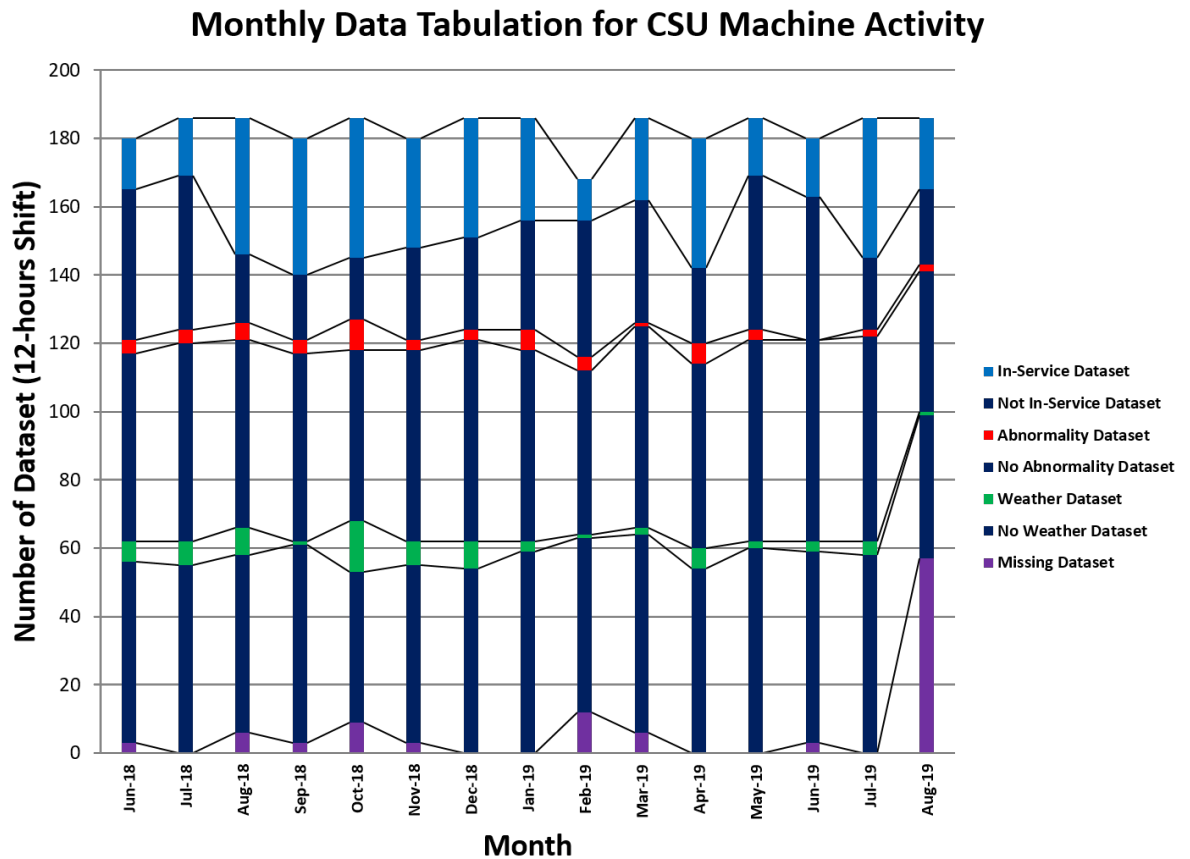


Figure 6.7: Monthly Data Tabulation for CSU Machine Activity

the sample size represents the events of machine failure. On the other hand, 8.09% of the population represents machine operation in adverse weather conditions, while the remaining 3.72% of the population are missing datasets that have not been included in this experiment.

The data set is tabulated according to the month for further analysis of the data sampling. Figure 6.7 outlined the machine activity from June 2018 to August 2019. This chart is quite revealing in several ways. First, the pattern of weather is during Malaysia's rainy seasons. The graph shows that the number of weather events was steadily increasing from September 2018 to December 2018 and a few months before rising again in April 2019. This is supported by Malaysia's weather and climate data assessed in [273], which shows a high amount of rainfall during this season. It is interesting to note that the occurrences of abnormality during the rainy seasons are also at the highest frequency. The data presented here seem to support the assumption that, whenever weather conditions are adverse, the possibility of machine abnormality is higher. This causes a delay in the in-service operation time, requiring more

time to transport the coal to the boiler.

6.3.4 Event Dataset

Table 6.1: Event Data Benchmark Configuration & Average Value for CSU Machine

No	Description	Benchmark				Average Value		
		Set	Limit	UL	LL	Avg	Max	Min
ED1	Main Incomer Voltage (V)	395	5	390	400	395.94	427.46	237.66
ED2	Main Incomer Current (A)	300	300	600	0	275.23	394.68	146.51
ED3	Panel Humidity (%)	20	10	10	30	19.66	25.18	13.02
ED4	Main Incoming THD (%)	2	0.3	2.3	1.8	2.00	2.00	2.00
ED5	Panel Temperature (degC)	30	5	35	25	30.41	34.91	26.06
ED6	E-House Temperature (degC)	20	5	25	15	21.16	33.21	16.09
ED7	Main Slewing Angle (deg)	80	60	20	140	113.47	159.93	67.81
ED8	Travel Position (m)	150	50	200	100	168.31	197.05	133.06
ED9	Foot Slew Angle (deg)	180	180	360	0	199.79	253.24	176.17
ED10	Luffing Angle (deg)	5	15	20	-10	1.84	7.15	-2.97
ED11	Process Flow (Tan/hr)	500	500	1000	0	497.55	930.56	1.59
ED12	Wind Speed (m/s)	6	4	10	2	3.94	5.40	2.47
ED13	Rotary Drive Speed (r/min)	25	15	40	10	27.38	41.71	12.84
ED14	Portal Drive Speed (r/min)	25	15	40	10	29.25	41.82	13.07
ED15	Bucket Drive Speed (r/min)	25	15	40	10	21.43	32.09	10.74
ED16	Travel Drive Speed (r/min)	0	15	15	-15	-0.13	-0.02	-0.28
ED17	Bucket Elevator Current (A)	300	300	600	0	212.34	344.95	81.92
ED18	Portal Conveyor Current (A)	15	15	30	0	8.77	14.20	2.73
ED19	Travel Current (A)	0	200	300	-200	2.58	4.00	1.47
ED20	Boom Conveyor Current (A)	40	40	80	0	36.13	55.05	17.41
ED21	HPP2 Pump 2 Current (A)	40	20	60	20	39.64	51.87	26.68
ED22	HPP2 Pump 1 Current (A)	100	50	150	50	113.47	159.93	67.81
ED23	Rotary Feeding Current (A)	0	10	10	-10	0.00	0.00	0.00
ED24	E-House Humidity (%)	35	10	45	25	33.72	41.86	17.07

Total of 24 ED's have been selected for this study. This includes electrical switch gear parameters such as Main Incomer Voltage (V), Main Incomer Current (A) and Total Harmonic Distortion (THD), which were identified as the main variables in this study. Any extreme events that may harm the system, such as excessive voltage, current or harmonic, will be detected and captured in the event modeller algorithm. Environmental variables such as wind speed, humidity & temperature are also included in the ED list. This is to test the hypothesis that temperature, humidity, and wind speed variations in the vicinity of the plant may also affect the machine's efficiency. Other machine process variables, such as machine positioning, process flow, speed drives and motor running current, are included in the ED

Table 6.2: Event Data Upper Limit & Lower Limit for CSU Machine

No	Description	Upper Limit			Lower Limit		
		Avg	Max	Min	Avg	Max	Min
ED1	Main Incomer Voltage (V)	397.18	427.61	244.61	393.78	427.33	224.70
ED2	Main Incomer Current (A)	973.19	1034.95	938.20	39.36	76.33	5.19
ED3	Panel Humidity (%)	40.74	59.26	20.33	9.91	11.73	6.09
ED4	Main Incoming THD (%)	2.20	2.20	2.20	1.80	1.80	1.80
ED5	Panel Temperature (degC)	37.26	41.95	35.22	21.94	30.90	18.10
ED6	E-House Temperature (degC)	27.49	37.50	17.66	16.66	22.91	14.43
ED7	Main Slewing Angle (deg)	174.12	177.85	148.05	23.95	121.53	3.43
ED8	Travel Position (m)	208.31	215.85	188.59	130.55	161.32	96.65
ED9	Foot Slew Angle (deg)	359.99	360.00	359.97	0.01	0.05	0.00
ED10	Luffing Angle (deg)	26.55	32.58	20.58	-11.28	-7.80	-15.35
ED11	Process Flow (Tan/hr)	1183.75	1618.00	2.00	0.25	1.00	0.00
ED12	Wind Speed (m/s)	9.86	14.75	6.57	0.16	1.23	0.00
ED13	Rotary Drive Speed (r/min)	50.16	50.23	50.10	-0.04	0.08	-0.13
ED14	Portal Drive Speed (r/min)	50.05	50.10	50.03	-0.16	-0.13	-0.26
ED15	Bucket Drive Speed (r/min)	47.20	48.77	46.78	-5.69	0.03	-10.32
ED16	Travel Drive Speed (r/min)	49.76	49.80	49.69	-50.08	-50.03	-50.13
ED17	Bucket Elevator Current (A)	1072.85	1191.39	772.54	0.00	0.00	0.00
ED18	Portal Conveyor Current (A)	83.51	86.17	78.13	-3.30	-1.69	-5.36
ED19	Travel Current (A)	336.77	343.54	279.51	-323.65	-227.33	-347.49
ED20	Boom Conveyor Current (A)	212.31	212.31	212.31	0.17	2.06	0.00
ED21	HPP2 Pump 2 Current (A)	107.23	159.28	79.50	10.15	40.95	0.00
ED22	HPP2 Pump 1 Current (A)	174.12	177.85	148.05	23.95	121.53	3.43
ED23	Rotary Feeding Current (A)	17.24	41.74	0.00	-18.93	-17.77	-20.17
ED24	E-House Humidity (%)	56.32	73.92	31.79	22.02	29.23	13.76

list to validate machine operations. The objective of the event modeller is to capture the cause-effect relationship between the ED's and its TD's beyond its operational benchmark. The benchmark ideally satisfies the following criteria: (a) it is based on real-world raw sensor data from different types of sensor deployments; (b) it contains (natural or artificially injected) faulty data points reflecting various problems in the deployment, including missing data points; and (c) all data points are annotated with the ground truth, i.e., whether or not the data point is accurate, and, if defective, the type of fault [274]. System engineers are therefore required to identify the upper and lower limits for each ED's, in order to eliminate all ED's operating within the normal range. To determine this set of points, a sample of data selected randomly from a group of efficient machine data sets, measuring the average, upper and lower limit of each ED. Table 6.1 and Table 6.2 illustrates the ED's benchmark and limits for CSU machine.

6.3.5 Assumption on Total Harmonic Distortion Data

One of the important event dataset explained in the previous section is ED4 known as Main incoming Total Harmonic Distortion (THD). THD is a measure that reveals how its voltage or current distorts the harmonics in the signal. The parameter is expressed in percentage and is typically determined by taking the root sum of the squares from the fundamental's first five or six harmonics. For example, the formula for calculating the THD in this CSU machine is shown in the following equation with V_n is in RMS voltage.

$$THD (\%) = 100\% \times \frac{\sqrt{V_3^2 + V_5^2 + V_7^2 + \dots + V_n^2}}{V_1} \quad (6.1)$$

In the case of a CSU machine, a panel-mounted multi-function meter is used to measure all energy data, including THD data. The instrument used Equation 6.1 to calculate the THD value and display the measurement through the panel. The initial plan was to import this THD data to the main PLC controller, thus integrating it with other EDs for event modeller analysis purposes. However, some major setbacks that do not allow the data to be available in the data acquisition as discussed in Chapter 4.

Two methods have been used to import this data. The first method is to expand the analogue output module from two to four. While the voltage and current analogue values remain available for data acquisition, the newly introduced THD data does not meet the ED requirement by not having the analogue reading as shown on the panel. The data acquisition only received the binary data which represent the alarm status of THD setpoint. Therefore, the second method using the Modbus communication protocol is introduced.

This method requires a brand new Modbus protocol adapter, to be installed at site with some configuration. Due to the limitation of the maintenance window, the installation of this device has been postponed. Whilst it is installed successfully, the device does not respond to the system. Therefore, an alternative plan has been taken by replacing the ED4 data with a simulated version. Considering the PQ assessment results in Table 4.4 which does not have any evidence of harmonic, the simulated version of the data has been modelled using Equation 6.2, where h_n is the magnitude of the n order harmonic, f_n is the n order harmonic frequency [275].

$$V_a = \sin(2\pi f_1 t + \theta_a) + \left[\sum_{n=3,5,7\dots}^{25} h_n \sin(2\pi f_n t + \theta_n) \right], 1 < n \quad (6.2)$$

6.3.6 Triggered Dataset

Table 6.3: Triggered Data Parameters

No	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Min	Max
TD1	Main Incomer Status	Rotary Drive Enable	Inactive Luff	Idle Luff	0	1
TD2	Maintenance Feeder	Portal Drive Enable	Inactive Foot Slew	Idle Foot Slew	0	1
TD3	REGEN 1 Enable	Bucket Drive Enable	Inactive Slew	Idle Slew	0	1
TD4	REGEN 2 Enable	Travel Drive Enable	Inactive Boom	Idle Boom	0	1
TD5	REGEN 3 Enable	Rotary Feed State	Inactive Rotary	Idle Rotary	0	1
TD6	Boom Motor	Portal State	Inactive Buck	Idle Buck	0	1
TD7	Boom Brake Motor	Slew Remote LCS	Inactive Portal	Idle Portal	0	1
TD8	Rotary Motors	Boom Remote LCS	Inactive Travel	Idle Travel	0	1
TD9	Bucket Motors	Rotary Remote LCS	Inactive Mag Separator	Idle Mag Separator	0	1
TD10	Bucket Brake Motors	Buck Remote LCS	Inactive HPPI	Idle HPPI	0	1
TD11	Portal Motor	Portal Remote LCS	Inactive HPPII	Idle HPPII	0	1
TD12	Travel Motors	Travel Remote LCS	Failure Luff	Busy Luff	0	1
TD13	Travel Brake Motors	Slew Local LCS	Failure Foot Slew	Busy Foot Slew	0	1
TD14	Mag Separator Motor	Boom Local LCS	Failure Slew	Busy Slew	0	1
TD15	HPPI Motor	Rotary Local LCS	Failure Boom	Busy Boom	0	1
TD16	HPPII Motor 1	Buck Local LCS	Failure Rotary	Busy Rotary	0	1
TD17	HPPII Motor 2	Portal Local LCS	Failure Buck	Busy Buck	0	1
TD18	Cable Reel Motor	Travel Local LCS	Failure Portal	Busy Portal	0	1
TD19	Water Hose Reel Motor	Slew Disable LCS	Failure Travel	Busy Travel	0	1
TD20	Rail Clamp Motor 1	Boom Disable LCS	Failure Mag Separator	Busy Mag Separator	0	1
TD21	Rail Clamp Motor 2	Rotary Disable LCS	Failure HPPI	Busy HPPI	0	1
TD22	Diverter Chute Motor 1	Buck Disable LCS	Failure HPPII	Busy HPPII	0	1
TD23	Diverter Chute Motor 2	Portal Disable LCS	Selected BC1A	Busy BC1A	0	1
TD24	Water Spray Motor	Travel Disable LCS	Selected BC1B	Busy BC1B	0	1

The event modeller algorithm looks at the causal effect of the ED's to the TD's. In this study, 96 TD's were selected and divided into four clusters. These clusters represent different types of trigger points. As the main objective of this experiment is to carefully observe the high correlation between input and output, dividing into clusters will help system engineers to develop a unique pattern that represents the state of the system. Cluster details are as described, where Cluster 1 represents the status of the machines and the status of the motor. This cluster validates the electrical component of the running machines. Cluster 2, on the other hand, validates the status of drives and Local Control Station (LCS). This cluster observes the correlation between the modes of operation with the ED's. Meanwhile, the inactive versus failure status for each process are included in Cluster 3. This cluster incorporates the fault or inactive information. Finally, Cluster 4 validates the Idle vs. Busy status to complete the list. Table 6.3 summarised the Trigger Data parameters according to the clusters.

6.3.7 Data Discretisation

Data discretisation refers to the process by which continuous data properties, features or variables are transformed or partitioned into a finite set of intervals with minimal information loss. As plant data is a raw big data from the SCADA system, it is necessary to discrete the data into a smaller range for analysis purposes. In this project, we have divided the data into hourly format and system engineer will be able to select the data based on the following Table 6.4.

Table 6.4: Data Discretisation Strategy

No	Hour	Row Number	No	Hour	Row Number
1	00:00:00	0	13	12:00:00	43202
2	01:00:00	3602	14	13:00:00	46802
3	02:00:00	7202	15	14:00:00	50402
4	03:00:00	10802	16	15:00:00	54002
5	04:00:00	14402	17	16:00:00	57602
6	05:00:00	18002	18	17:00:00	61202
7	06:00:00	21602	19	18:00:00	64802
8	07:00:00	25202	20	19:00:00	68402
9	08:00:00	28802	21	20:00:00	72002
10	09:00:00	32402	22	21:00:00	75602
11	10:00:00	36002	23	22:00:00	79202
12	11:00:00	39602	24	23:00:00	82802

6.3.8 Key Performance Indicator Parameter

The CSU machine is operated 24 hours a day whenever the shipment arrives at the jetty. In practice, coal is normally discharged from the vessel within 72 hours. If the execution breaches the maximum time, there will be a high demurrage cost and that must be avoided. Usually, much effort has been implemented to avoid this penalty, such as performance evaluation [276], which uses plant data to record operating activities and resources while preventing downtime. However, this effort is not made in real-time and requires manpower to monitor the plant. Recently, performance modelling using DES in real-time predictive control has shown promising results. DES model is capable of capturing state variable changes at a discreet time in a highly complex environment [236, 277, 278]. The translation of plant data into management performance metrics uses a combination of DES model and real-time DAQ system.

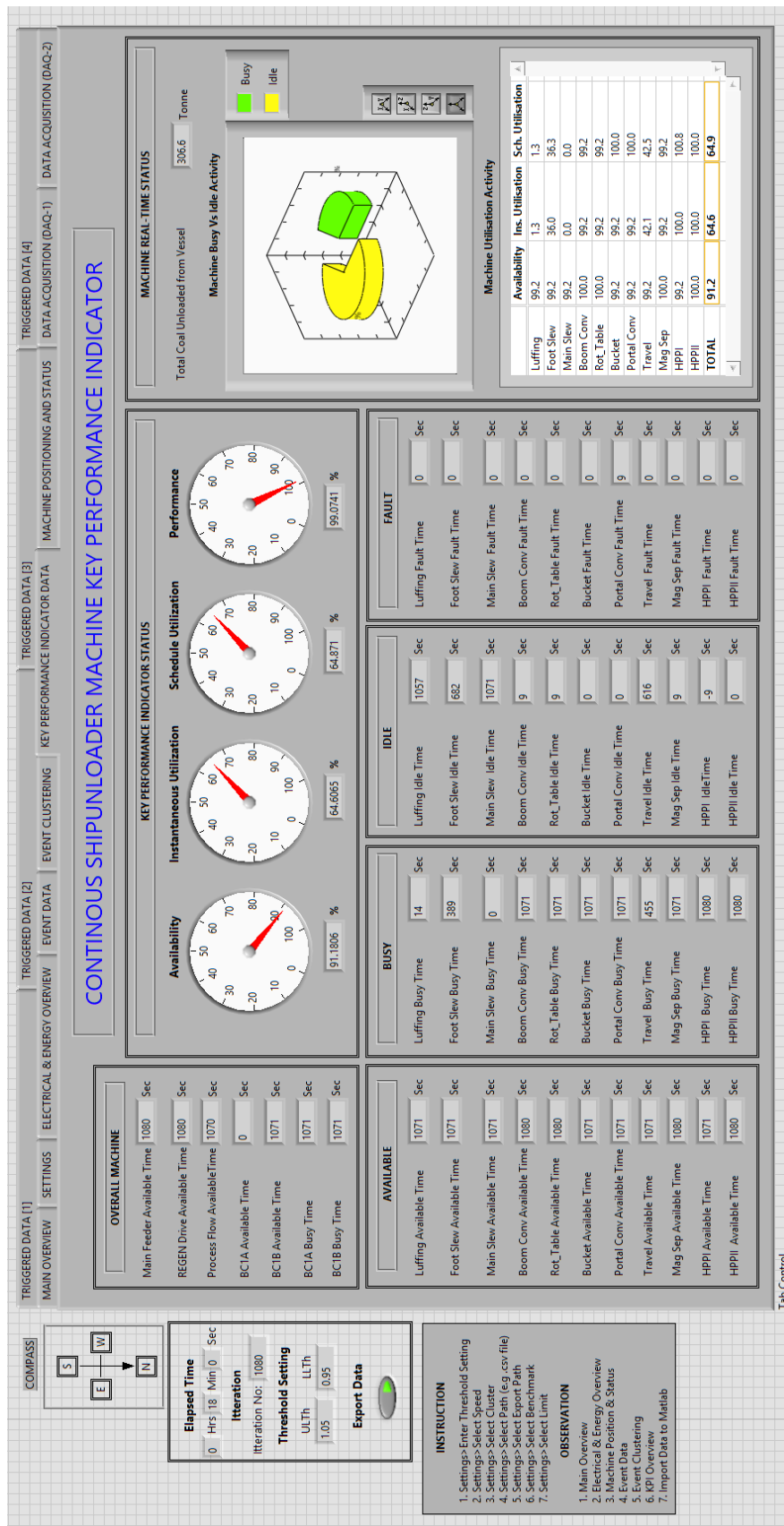


Figure 6.8: CSU Machine Key Performance Indicator

In this section, the translation of plant data into management performance metrics is introduced. It helps engineers to make a quick decision when the performance of the KPI status has reached certain limits. Each performance factor is made up of several measurable indicators. Information for the calculation of KPIs is retrieved from field sensors, actuators, monitors, readers, counters, human inputs and other sources. Measurement indicators are updated in real-time. Figure 6.8 shows the KPI status of the developed CSU machine. The display updates the exact time for available, busy, idle and fault for each main component and previews the KPI and machine utilisation using the performance dashboard.

Table 6.5: Key Performance Indicator Parameters

No	Description	Unit	Min	Max
KPI 1	Availability (A)	%	0	100
KPI 2	Instantaneous Utilisation (IU)	%	0	100
KPI 3	Schedule Utilisation (SU)	%	0	100
KPI 4	Performance (P)	%	0	100
KPI 5	Green House Gases Ratio (GHG)	%	0	100
KPI 6	Energy Consumption Ratio (EC)	%	0	100

Table 6.5 outlined the six KPIs to be used for this experiment. To further understand these six parameters, the CSU system status definition, including the KPI formula used in this calculation, is shown in Table 6.6. Energy consumption is calculated based on the individual motor rating. By multiplying the individual energy consumption with the emission factor, all carbon footprint and environmental impact KPIs can also be calculated. The Energy Efficiency KPI status display of the developed CSU machine is shown in Figure 6.9. The display updates the single-line diagram of the CSU machine and previews the KPI energy efficiency using the performance dashboard. Snippets of the KPI formula can be found in Appendix A.(Figure 1-3)

6.3.9 Predictive Model Dataset

As explained in Chapter 3, predictive modelling is a branch of advanced analytics that used historical data or business processes to predict unknown future events. Supervised machine learning is used to train this model for classification purposes. In order to align the proposed EMDA technique with the system development, the event modeller input-output sensitivity index (weight) was labelled based on the pattern. The KPIs are integrated with the weight via

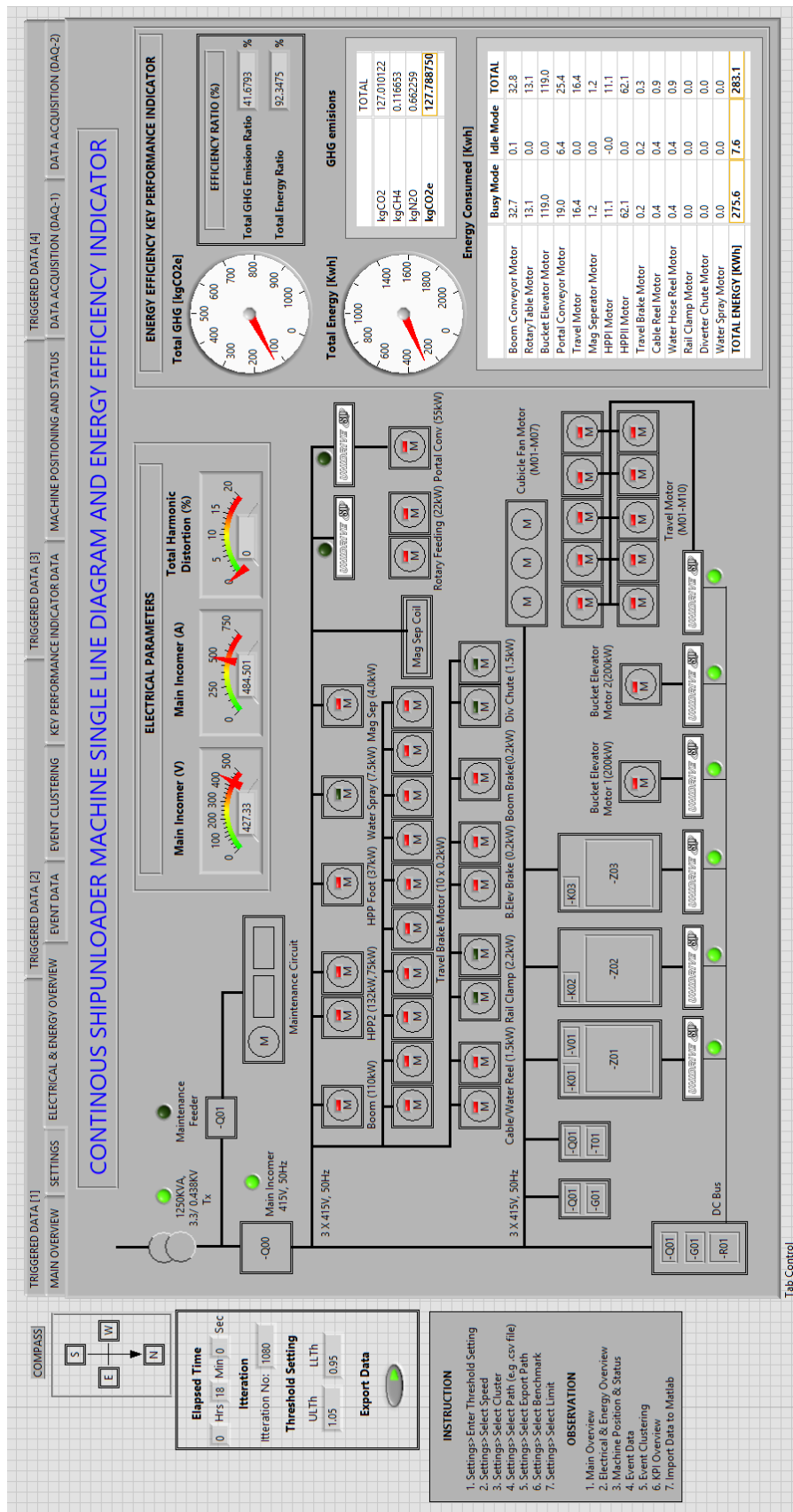


Figure 6.9: CSU Machine Energy Efficiency KPI

Table 6.6: Description of Machine Status & Formula for KPI

Machine Status

Description	Details
Incomer Availability Time	Total time accumulates when main incomer is energised.
Availability Time	Total Time accumulates when machine is healthy but not running.
Busy Time	Total Time accumulates when machine is healthy and running
Fault Time	Total Time accumulates when machine experience fault during running.
Inactive Time	Total Time accumulates when machine is in on disable by operator.
Idle Time	Availability Time - Busy Time
Operating Time	Busy Time + Idle Time + Fault Time + Inactive Time.

Formula for Machine KPI

Description	Formula
Machine Availability	(Available Time/Incomer Available Time)
Machine Instantaneous	(Busy Time/Incomer Available Time)
Machine Schedule Utilisation	(Busy Time/Available Time)
Machine Performance	(Process Flow/Incomer Availability Time)
Total Coal unloaded	(Flow Rate/3600)

Formula for Energy Efficiency KPI

Description	Formula
Busy Energy Consumption	Busy Time/(3600 × kWRating × Number of Motors)
Idle Energy Consumption	Idle Time/(3600 × kWRating × Number of Motors × 0.25)
Total Energy Consumption	(Busy Energy Consumption + Idle Energy Consumption)
Carbon Dioxide Gases (CO ₂)	(Total Energy Consumption × 0.448581)
Methane Gases (CH ₄)	(Total Energy Consumption × 0.000412)
Nitrous oxide Gases (NO ₂)	(Total Energy Consumption × 0.002339)
Carbon Dioxide Equivalent	(Total Energy Consumption × 0.451331)

the look-up table. The table is then exported to a machine learning tool to test the accuracy of the classification. In order to prepare the predictive model dataset, all available data are divided into three main categories known as in-service, abnormality and weather. This 880 dataset is tabulated according to the logic pattern shown in Table 6.7. The next section describes the outcome of the predictive model dataset that uses Cluster 1.

6.3.9.1 Pattern 1 - Standby, No Abnormality and Good Weather

A standby machine with no abnormalities and no weather conditions indicates that the machine is always available, is not overused, is well maintained and has good strategic planning. Based on Table 6.7, this pattern shows the highest number of datasets in the population at 48.75%. Figure 6.10 shows the 4-quadrant result for Pattern 1 dated 12th June 2018 for a

Table 6.7: Predictive Model Pattern Logic

Pattern Description	Logic			Number of Dataset
	In-Service	Abnormality	Weather	
Pattern 1 - Standby, No Abnormality and Good Weather	0	0	0	429
Pattern 2 - Standby, No Abnormality and Bad Weather	0	0	1	12
Pattern 3 - Standby, Downtime and Good Weather	0	1	0	15
Pattern 4 - Standby, Downtime and Bad Weather	0	1	1	4
Pattern 5 - Operating, No Abnormality and Good Weather	1	0	0	330
Pattern 6 - Operating, No Abnormality and Bad Weather	1	0	1	53
Pattern 7 - Operating, Downtime and Good Weather	1	1	0	32
Pattern 8 - Operating, Downtime and Bad Weather	1	1	1	5

period of 4-hours. From the chart, the ROC pattern remains unchanged from the initial experiment. The KPI dashboard remains consistent with zero reading, which shows that the machine is not operated during sampling time.

6.3.9.2 Pattern 2 - Standby, No Abnormality and Bad Weather

Bad weather refers to the rain, the storm, the wind and the combination of these three elements. Since the CSU machine is operated near to the sea, there is always a risk of adverse weather conditions, especially during the monsoon season. Having said that, it is essential to capture this pattern in the predictive model dataset. In theory, while the machine is in standby mode, there will be no operation during bad weather unless it is in between shipments. For example, when the first vessel unloading is completed, the second vessel has to wait for some inspection and documentation before it is allowed to be dock. During this period, the system is set to be on standby. There was 12 dataset available for this pattern. Figure 6.11 shows the 4-quadrant result for Pattern 2 dated 25th July 2018 for the period of 1-hour. In this experiment, the 1-hour sampling time was chosen to respond to bad weather events. From the chart, the ROC pattern remains unchanged from the initial experiment. This indicates that no operation activity carried out during this period. This is supported by information on the KPI dashboard.

6.3.9.3 Pattern 3 - Standby, Downtime and Good Weather

Machine failure could happen at any time, with or without operation. As the CSU machine runs at the jetty according to the shipment's arrival, there are occasions when the machine failure occurs during a non-operating cycle. Although the weather is reported to be normal, the dynamic and the interaction within the machine vicinity will lead to a machine failure.

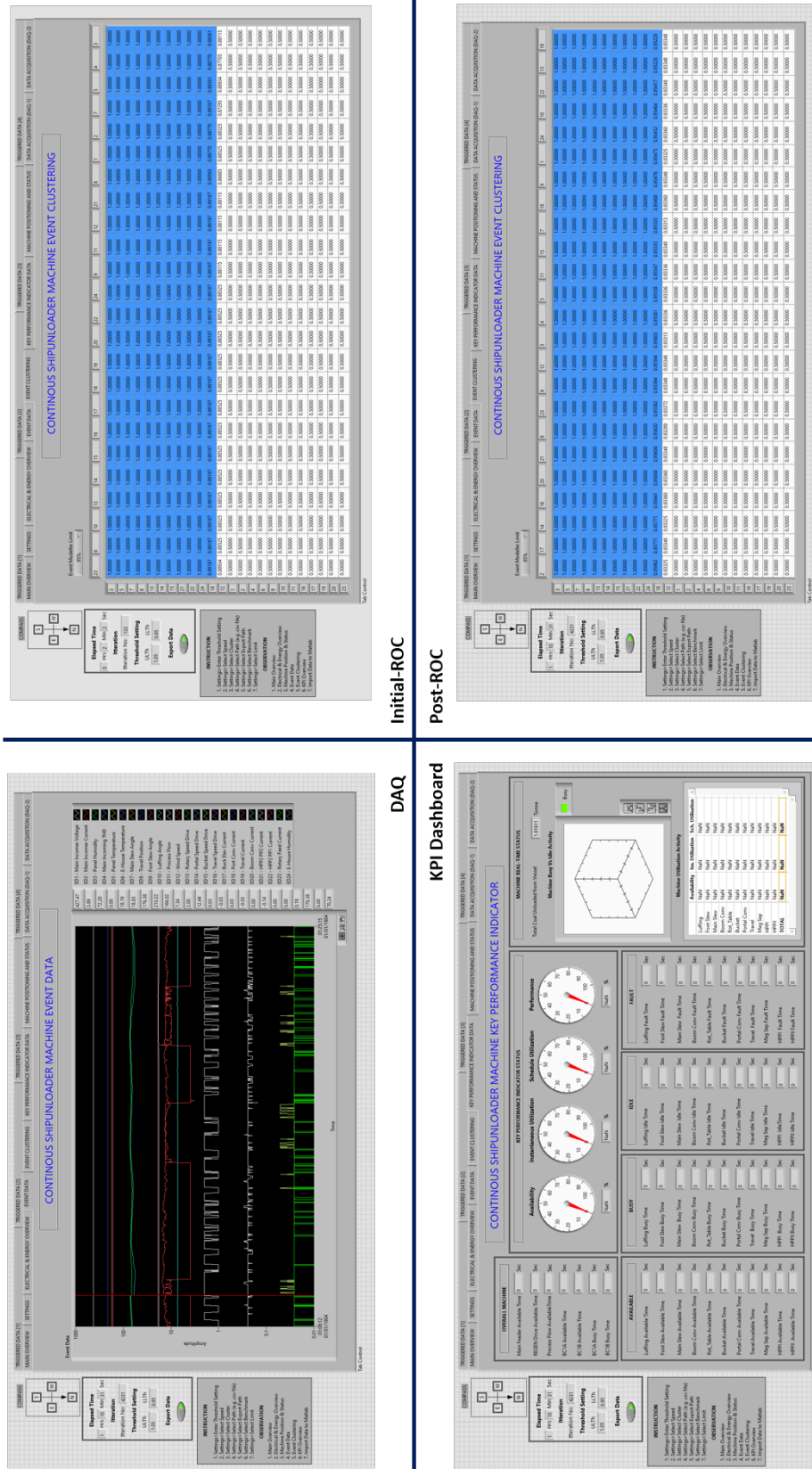


Figure 6.11: CSU Machine - Pattern 2 (Standby, No Abnormality and Bad Weather)

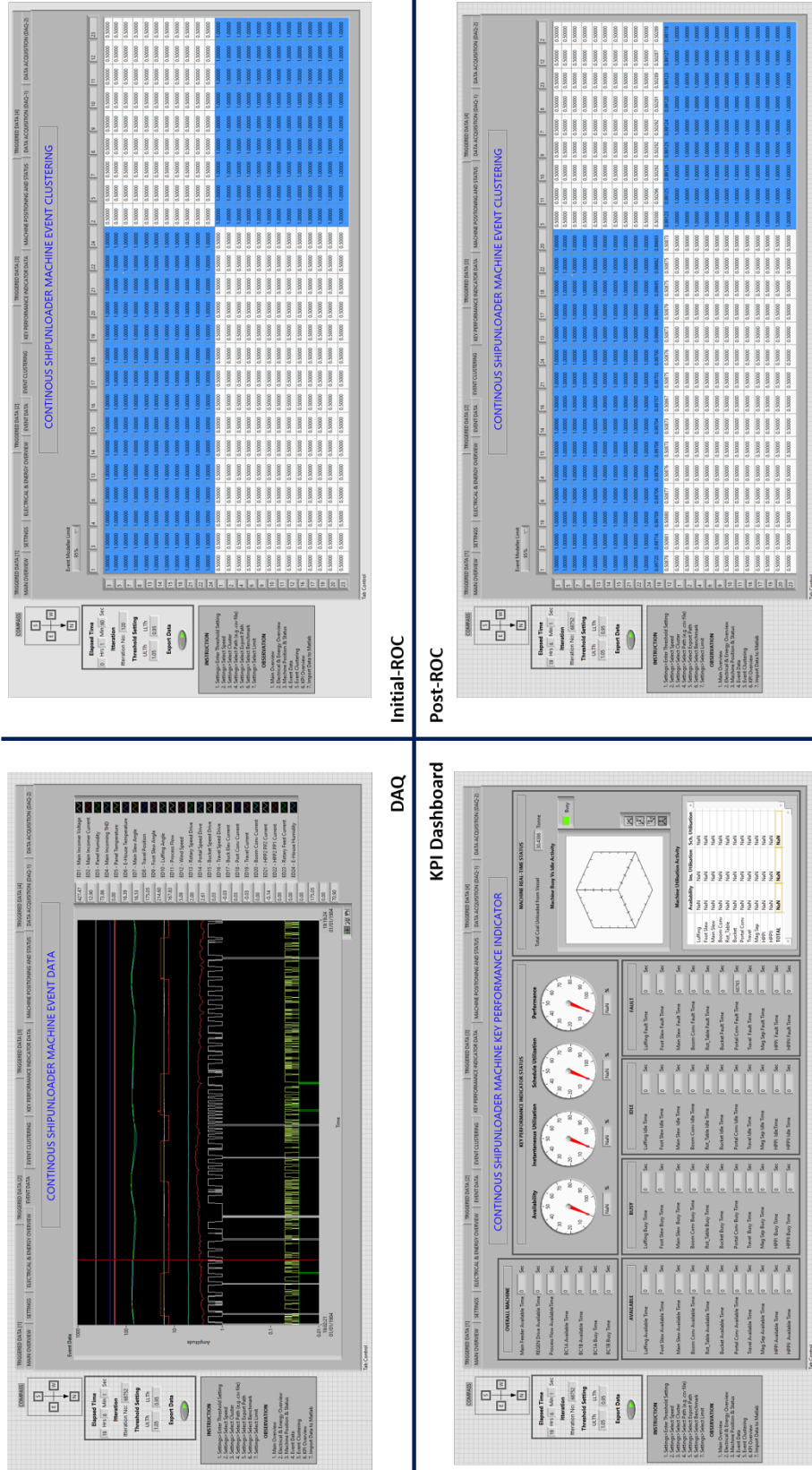


Figure 6.12: CSU Machine - Pattern 3 (Standby, Downtime and Good Weather)

For example, the electrical room of the CSU machine is equipped with an air conditioner to ensure that the room is maintained at a certain temperature. During operation, the choke and electrical drives produce an excessive amount of heat, making the electrical room extremely warm. As a result, an additional air conditioner was switched on to combat the excessive heat. However, if the CSU machine is on standby, the room will be extremely cold and humid if the additional air conditioner is switched on continuously. With high humidity, sensitive electrical devices such as choke and harmonic filters could also fail. There were 15 datasets available that match this pattern. Figure 6.12 shows the 4-quadrant result of Pattern 3 dated 4th January 2019. The pattern of the ROC remains unchanged from the chart, even though it runs for 19 hours. The KPI dashboard does not indicate any reading that proves that the machine is not being operated during sampling time.

6.3.9.4 Pattern 4 - Standby, Downtime and Bad Weather

The coincidence of failure and bad weather at the same time is a bit tricky if it happens in standby mode. However, if the same scenario occurs at the same time in Pattern 2 and Pattern 3, it will lead to a combination factor. Having this combination factor pattern dataset will help system engineers to train supervisory learning with accuracy. Figure 6.13 presents the 4-quadrant result for Pattern 4 dated 28th June 2018 for the 1-hour period. From the chart, the pattern of the ROC does not change much in comparison to the initial state. Again, because the machine is in standby mode, the KPI dashboard remains empty. Only 4 population datasets are available based on Table 6.7.

6.3.9.5 Pattern 5 - Operating, No Abnormality and Good Weather

A CSU machine with continuous and smooth operation indicates that the machine has achieved high availability, reliability and maintenance. This could be revealed by the translation of the Latent Knowledge (KPI) from the sensors and actuators established in the system development process. Having this system state pattern will supervise the machine learning process to train the model to predict the future state of the system. Figure 6.14 shows the 4-quadrant result for Pattern 5 dated 17th July 2018. It is apparent from the ROC that the pattern of the system state has changed slightly. The input-output relationship has been adapted to the system state, leaving the high correlation relationship to be assessed by the system engineer. The KPI

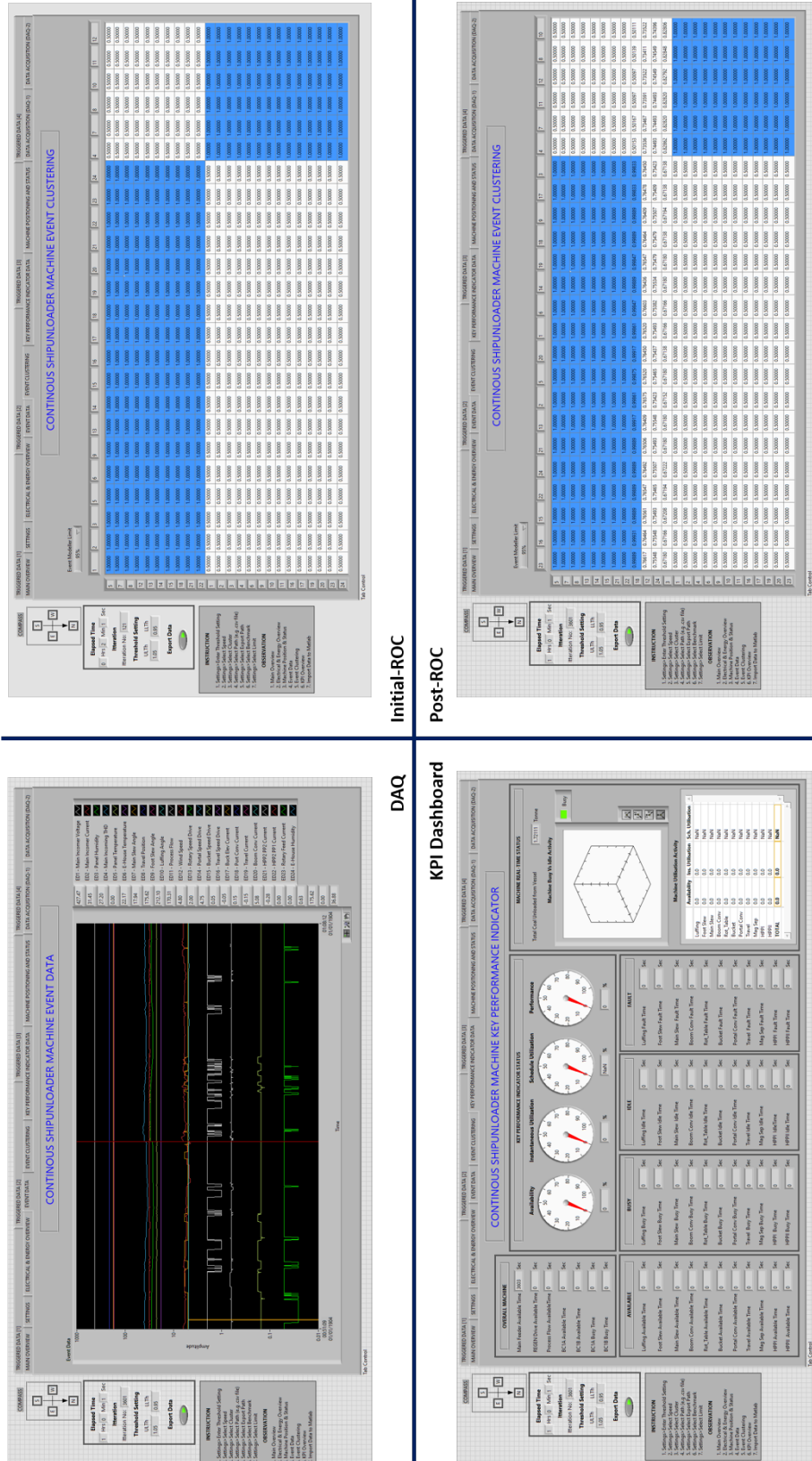


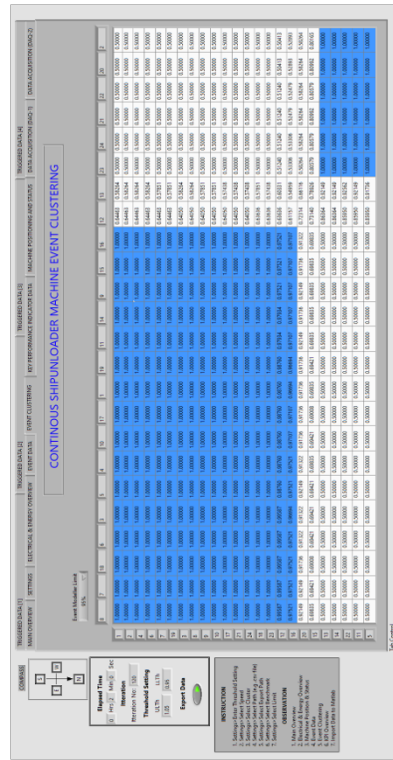
Figure 6.13: CSU Machine - Pattern 4 (Standby, Downtime and Bad Weather)

Initial-ROC

Post-ROC

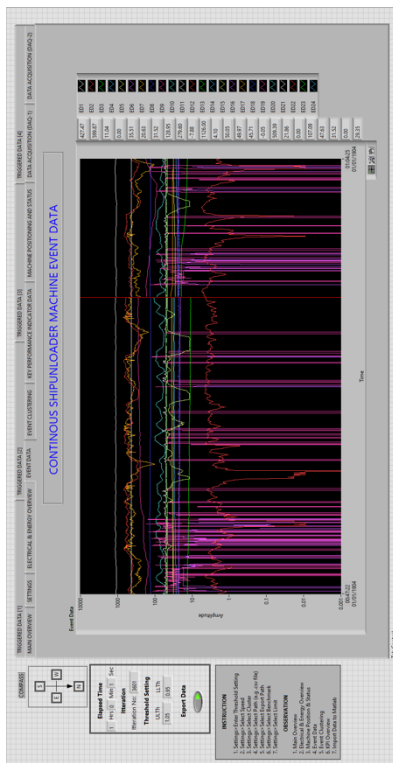
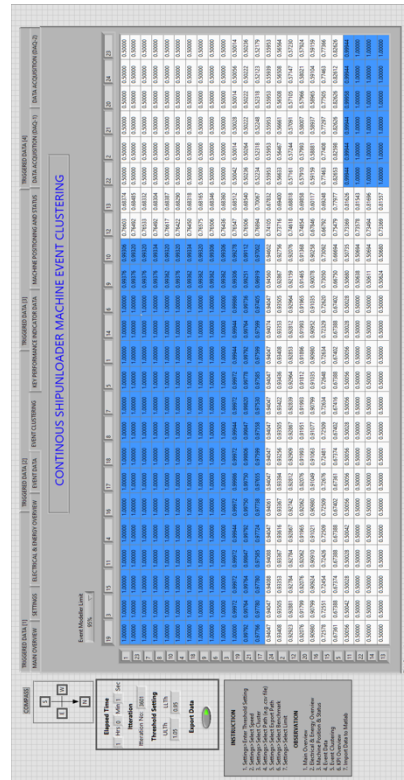
DAQ

KPI Dashboard



Initial-ROC

Post-ROC



DAQ

KPI Dashboard

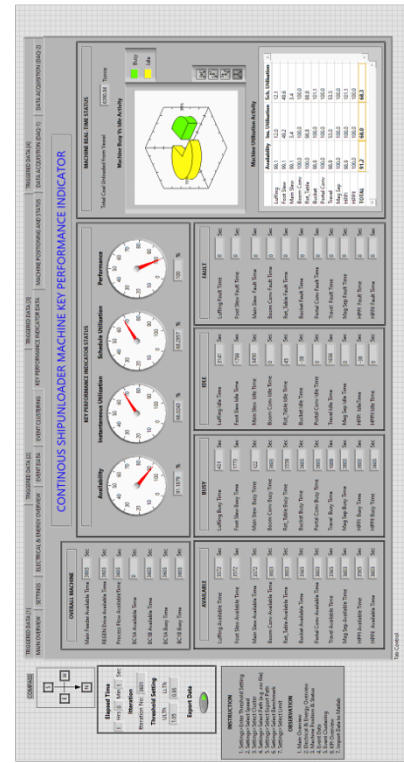


Figure 6.14: CSU Machine - Pattern 5 (Operating, No Abnormality and Good Weather)

dashboard demonstrates that the system is running at its best performance. Performance is rated at 100 %, while Availability and Utilisation are rated at 91.19 % and 68.02 % respectively. This validates the operation of the CSU machine without any interruption and under good weather conditions.

6.3.9.6 Pattern 6 - Operating, No Abnormality and Bad Weather

Moving to the machine operation, which encounters some bad weather while running. The main finding here is whether bad weather influences the system state model. Figure 6.15 presents the 4-quadrant result for Pattern 6 dated 12th December 2018 for the 1-hour period. In this example, the ROC in quadrant 4 reveals that some changes have been made to the system state. It is a change in the shape of the ROC that matters. The activity of the operation is affected by the high correlation variables group. It is clearly stated that the series of triggered data {TD11, TD6, TD7, TD8, TD9 and TD10} are heavily influenced by the Portal Conveyor Current (ED18). There is no evidence of weather conditions affecting the operation of the machine. However, the KPIs indicate that the performance was 33.29% and that the availability was 74.08%. This suggests that the operation of the machine is suspended under bad weather conditions.

6.3.9.7 Pattern 7 - Operating, Downtime and Good Weather

In good weather, the CSU machine operates at full capacity to ensure that the coal is unloaded within the timeframe. In normal circumstances, 72 hours are allocated for each vessel to unload coal. However, a queue or interruption due to machine failure, unavailable spare parts or back to back shipment will force the machine to operate on a non-stop basis. Consequently, it will also contribute to downtime. Figure 6.16 shows the 4-quadrant result for Pattern 7 dated 31th January 2019 for the 1-hour period. In this example, the pattern of the ROC has changed considerably. One of the main findings is that the process flow (ED11) is no longer correlated with other TDs. This indicates that the machine is being stopped. On the other hand, the Boom and Boom Brake motors are also not correlated with other EDs. This supports the previous statement that when the Boom Conveyor stops, which means that the coal was not transported to the central chute. The chart also reveals that the travel and travel brake motor joins the small cluster at the bottom. This indicates that the operator uses the travel

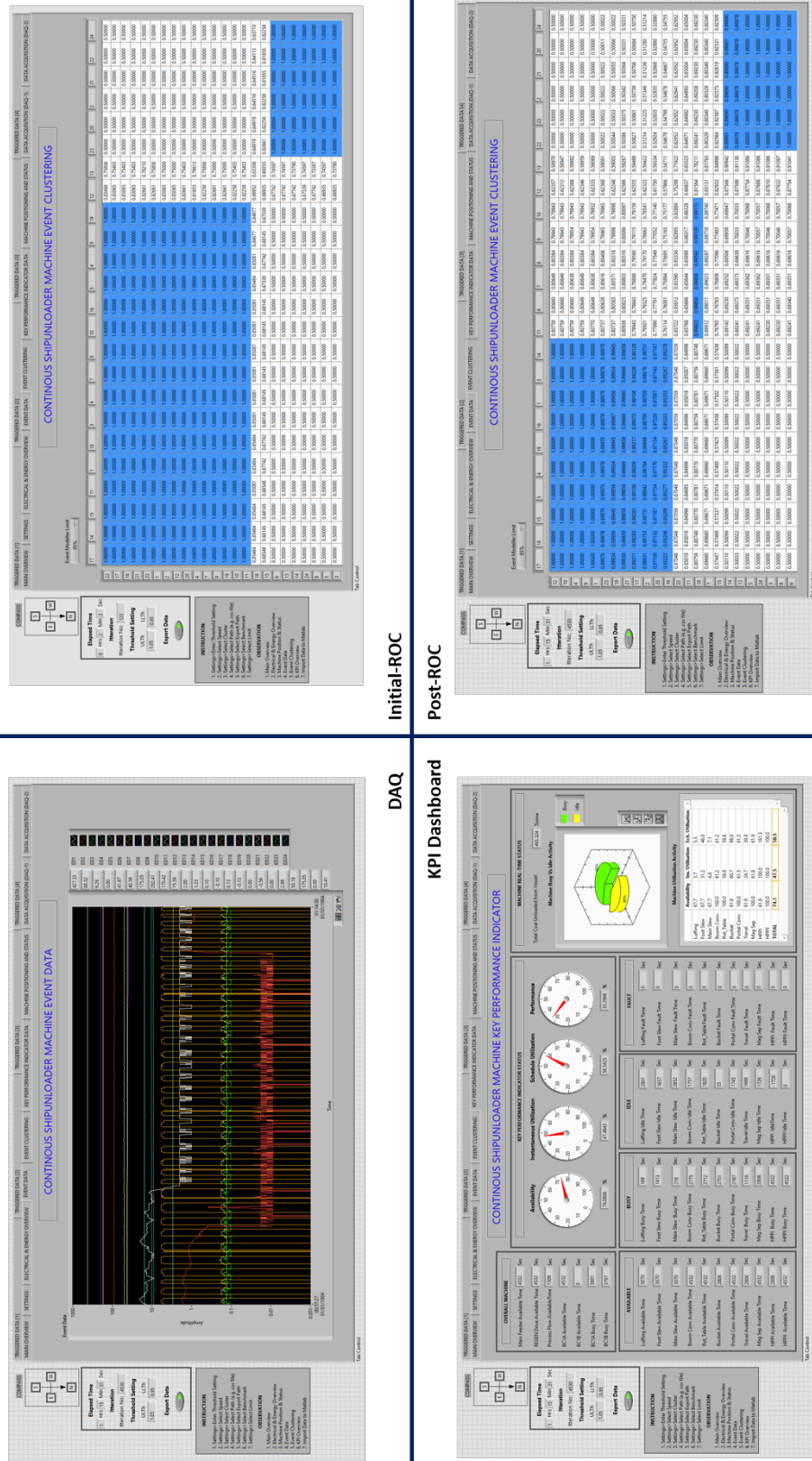
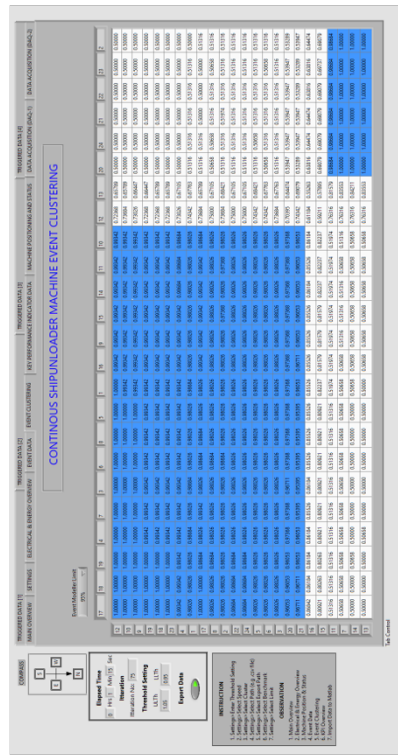
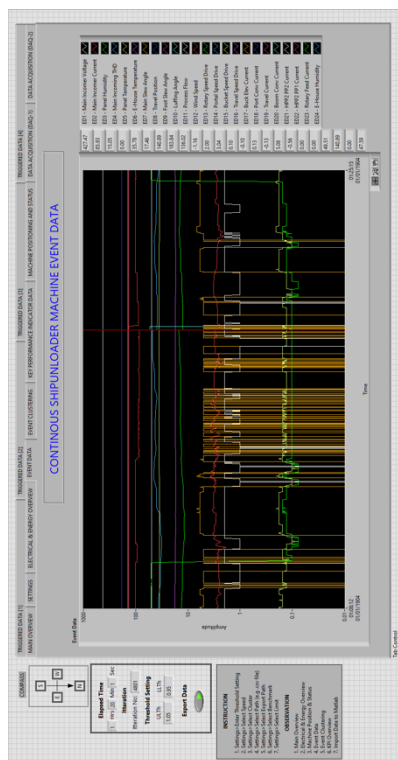
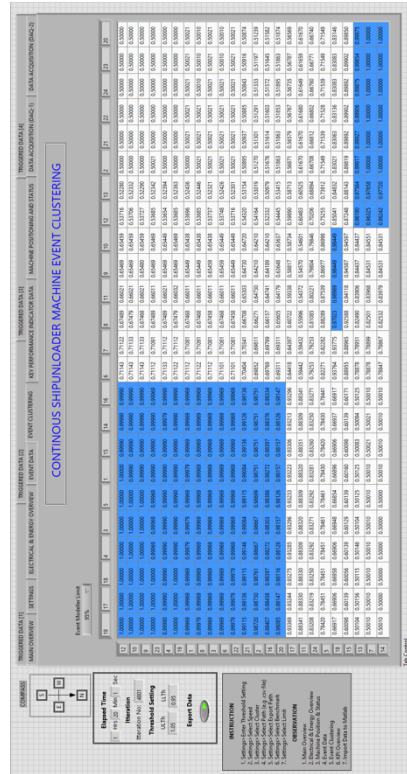


Figure 6.15: CSU Machine - Pattern 6 (Operating, No Abnormality and Bad Weather)



Initial-ROC

Post-ROC



DAQ

KPI Dashboard

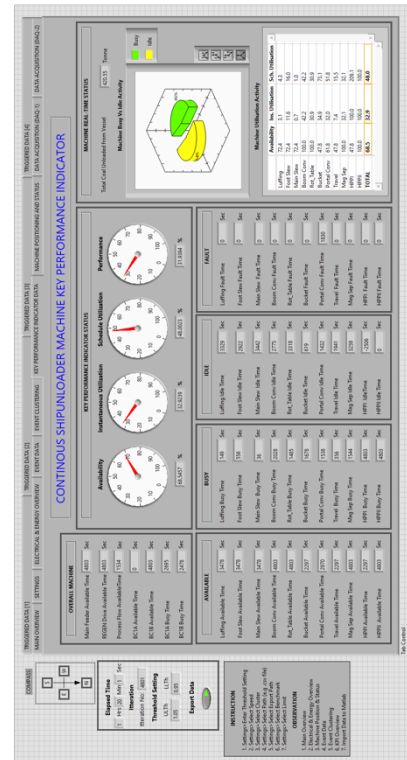
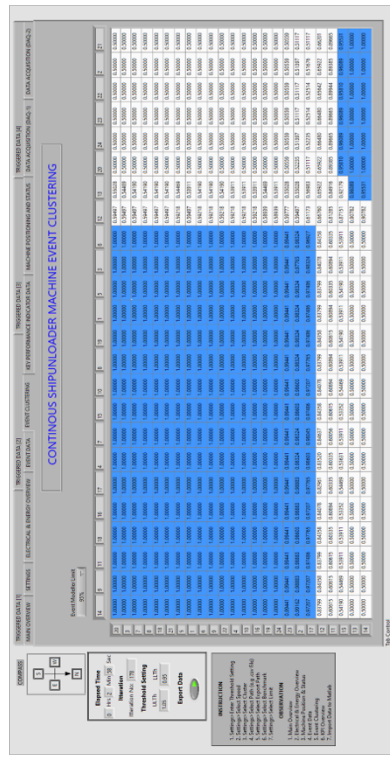
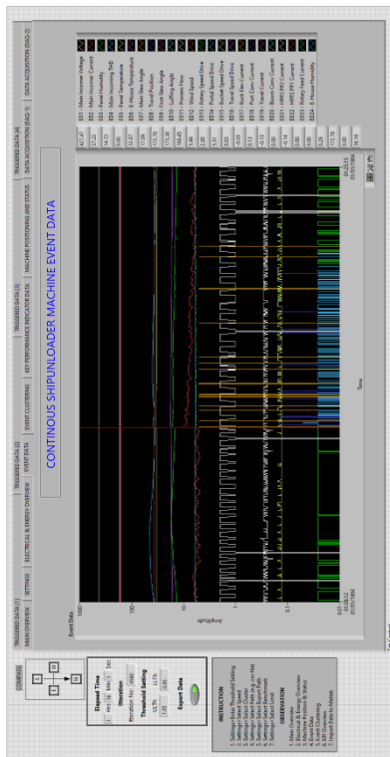
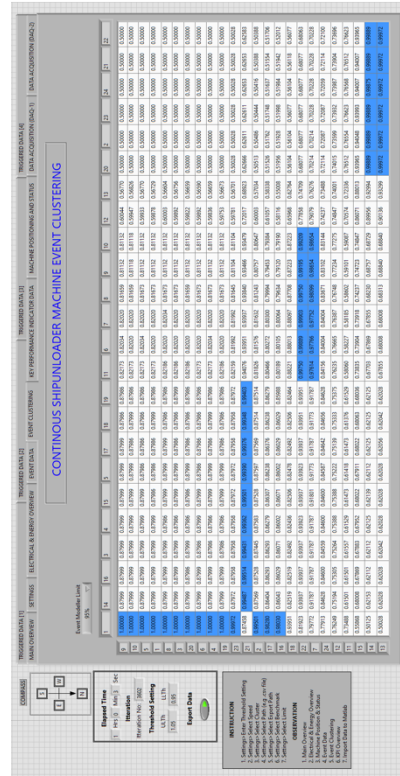


Figure 6.16: CSU Machine - Pattern 7 (Operating, Downtime and Good Weather)



Initial-ROC

Post-ROC



DAQ

KPI Dashboard

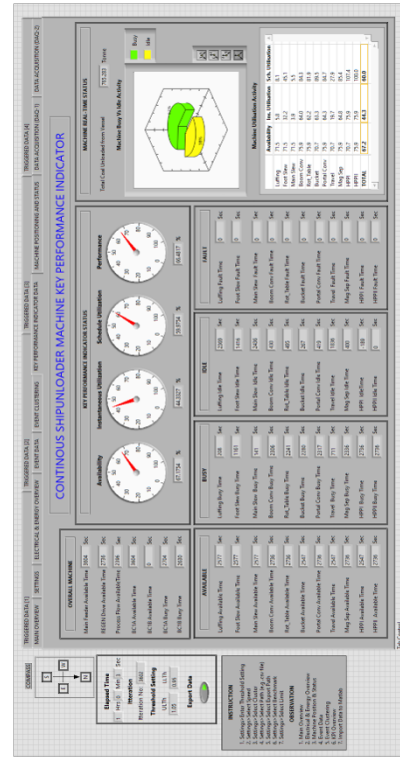


Figure 6.17: CSU Machine - Pattern 8 (Operating, Downtime and Bad Weather)

movement to park the CSU machine at its initial position. Last but not least, the portal conveyor current has a significant correlation with other motors, just like Pattern 6, which was discussed earlier. The reason for this is that the portal conveyor current is linear with the coal flow from the central chute. In principle, when the coal is dropped from the central chute to the portal conveyor, the loading current of the motor is dependent on the amount of coal. In addition to the effect of the dropping, the portal conveyor current will be fluctuating which were detected by the event modeller. On the other hand, the KPIs indicate that the performance was 31.93 % with the availability of 68.55 %. The instantaneous utilisation and Schedule Utilisation were at the lower percentage with 32.93 % and 48.00 % consecutively.

6.3.9.8 Pattern 8 - Operating, Downtime and Bad Weather

For the final pattern, which is the worst-case scenario for all patterns, the scenario is very unlikely to happen. Only 5 datasets are available from the population as shown in Table 6.7. Figure 6.17 shows the 4-quadrant result for Pattern 8 dated 26th October 2018 for the 1-hour period. In this example, the CSU machine operates at its steady-state before a fault is triggered. The ROC pattern in quadrant 4 shows the outcome of the fault. The chart shows that the Main Incomer Status only responds to the ED series. This indicates that the machine was completely shut down. The outcome from the motor status was only linked with the HPP Pumps (ED21 and ED22) and Main Slew (ED7).

6.4 Results and Discussion

Having discussed the dataset structure in the previous section, this segment would concentrate on experimental work, followed by results and discussion. Two major experiments were conducted using the system development implementation in Section 6.2. The first experiment is to evaluate which parameter influences machine failures within the CSU machine data set. The ultimate goal is to observe how Event Modeler and the translated KPIs are correlated with these parameters. The second experiment is to predict the pattern to which the generated predictive model dataset belongs. The intention is to validate the predictive model dataset using machine learning tools by comparing the classification accuracy of the three leading classifier models.

6.4.1 Experiment 1: Machine Failure Analysis with Key Performance Indicators

Harmonic filter failure can be influenced by various factors, such as faulty device, operator handling, and environmental factors in which the machine operates. It has also occasionally been affected by the combination of these factors. Based on the expert point of view of the CSU machine, the harmonic problem is associated with the occurrence of REGEN fault. In order to access the REGEN fault events, the abnormal dataset from Figure 6.7 was selected. The frequency of the REGEN failure was compared with other types of fault, as presented in Figure 6.18. The bar chart reveals that the highest number of REGEN faults was reported in October 2018, followed by August 2018, September 2018 and January 2019. Although the highest number of REGEN failures was recorded in October 2018, no report indicated that the REGEN drives had been replaced. However, the January 2019 reports indicated that the maintenance team had replaced the REGEN during this month due to a fault. The January dataset was therefore chosen to further explain the relationship of the variables using the EMDA technique.

6.4.1.1 Experiment Strategy

Building a solution capable of translating engineering data automatically into high-level management information is the principal motivation for this experiment. The KPIs shall be calculated based on the operation of the machine during the selected period. The method used in Chapter 5 is borrowed to run a fault dataset with a time interval of 5 minutes. The first 5 minutes are the events before the fault occurs, followed by the fault, and end with the last 5 minutes that accumulate up to 15 minutes of sampling time. The reason why this method was adopted is to prove that the event modeller could instantly solve a complex system without having to run the system for a longer period. It is easy to sample data in a simulation environment, as the user can decide when the fault may occur, but this is not the case with industrial data. The dataset was therefore analysed to determine which data was almost equivalent to the setting. The experiment strategy is summarised in Figure 6.19. The results are presented in two phases. First, the event modeller was used as a tool to reduce the dimensionality of the data being observed. Second, the relationship between the observed data and the KPIs has been compared. The results will be explained as follows.

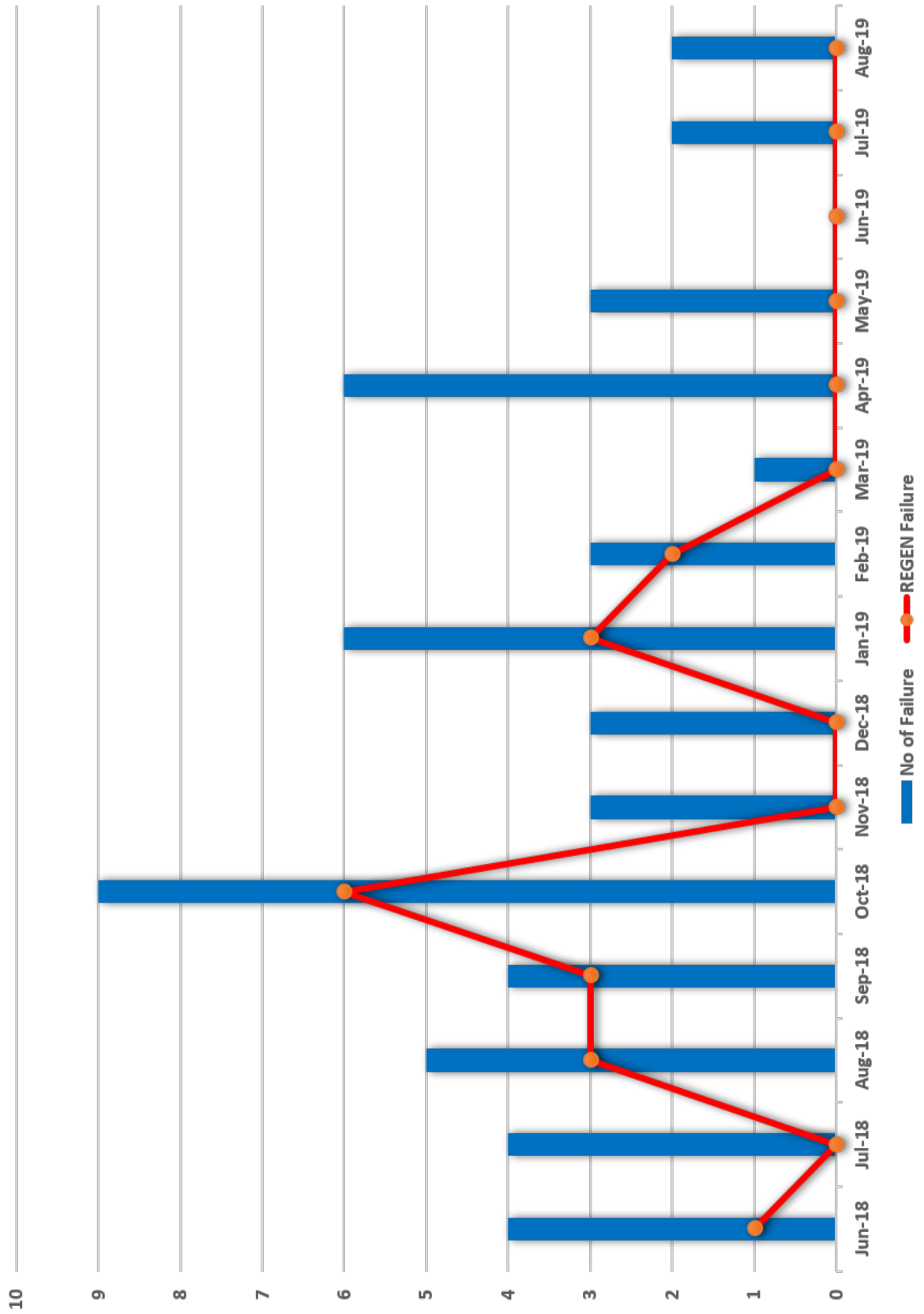


Figure 6.18: Tabulation of CSU Machine Failure vs REGEN Failure

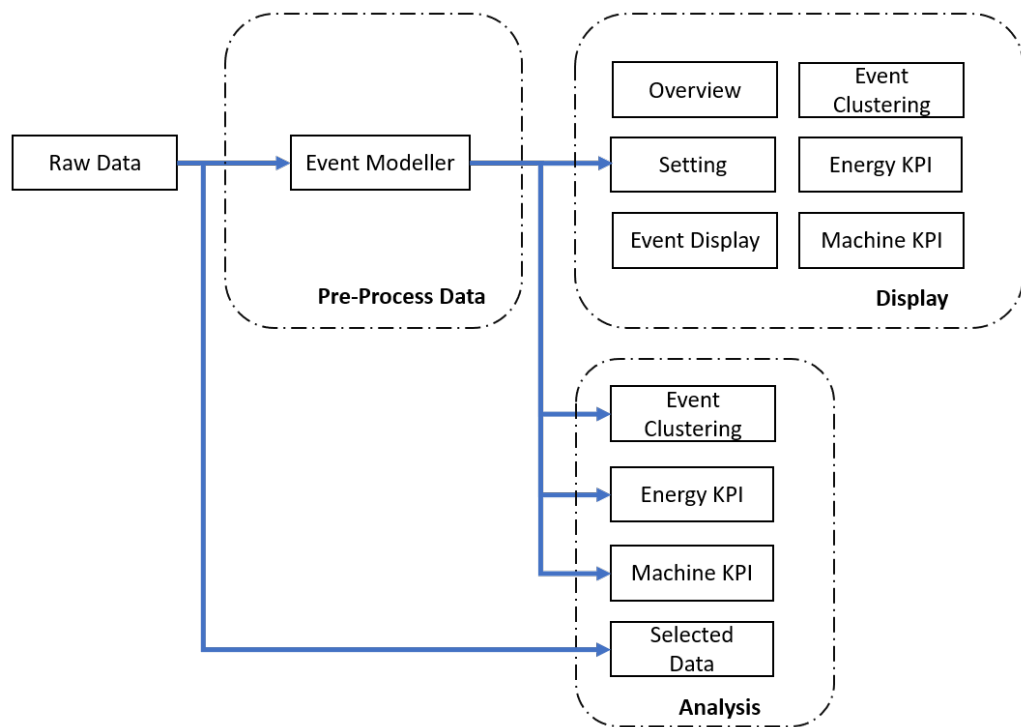


Figure 6.19: Experiment Strategy for KPIs Analysis

6.4.1.2 Relationship between the observed data against Event Modeller

In this example, a homogenous dataset representing Pattern 7 of the predictive model dataset population was sampled for 16 minutes. The dataset of 30th January 2019 was sampled between 12:47 pm and 13:03 pm. The dataset arrangement in Section 6.3 indicated that there were 96 TDs to be grouped against 24 EDs in the form of matrices to show which system parameters were highly correlated. It is noteworthy that the 96 X 24 matrices are too complex for a correlation analysis; therefore, the 96 TD was divided into 4 clusters, as shown in Table 6.3. For each TD cluster, 24 X 24 matrices were analysed using the event modelling technique. The sensitivity index of each input-output relationship has been measured and updated every second. The ROC pattern was observed in real-time and captured every minute. As a result, four-quadrant windows representing four key states were presented: Initial, Pre-Fault, Fault and Post Fault for each TD cluster. This analysis intends to examine how TDs influence the pattern of the ROC. It is necessary to clarify what happens during this 16-minute time stamp

before presenting the chart. Based on the report, the machine is running at its optimum level before it is fully phased out after 10 minutes. The reason why machine trips is unknown, which left the proposed method to be concluded further.

Figure 6.20 to Figure 6.23 shows the pattern of the ROC system state for all clusters. An important finding was that the first four EDs were the same for all clusters. They comprised of E-House Humidity (ED24), Portal Conveyor Current (ED18), Wind Speed (ED 12) and Portal Speed Drive (ED14). This indicates that one of these EDs would have been the cause of failure. The following observation was based on the ROC pattern. Similar findings were found in Cluster 2 and Cluster 3 patterns. Surprisingly, between the four quadrants, this pattern has not changed much. The only obvious is that it is the TDs are influenced by E-House Humidity (ED24). This suggests that the relationship between the TDs and the EDs for this quadrant is consistent. In other words, the fault may not be related to the TDs in those clusters. Moving to the ROC patterns for Cluster 1 and Cluster 4, major changes were observed after the fault in this quadrant. This suggested that the TDs could have originated from these clusters. Taking this into account, the highly correlated variables for both Cluster 1 and Cluster 4 were further examined. Equation 6.3 indicates the possible root cause of Cluster 1. If we overlap both relationships, the new equation for Cluster 1 is expressed in Equation 6.4:

$$\rho_{0.95} = (ED18) \times (TD6, TD7, TD11, TD8, TD9, TD10) + (ED24) \times (TD8, TD9, TD10, TD24, TD21, TD2, TD22, TD23, TD20, TD13, TD12) \quad (6.3)$$

$$\rho_{0.95} = (ED18 + ED24) \times (TD8, TD9, TD10) \quad (6.4)$$

On the other hand, the following equation 6.5 indicates the possible root cause of Cluster 4. If we overlap both relationships, the new equation for Cluster 4 is expressed in Equation 6.6.

$$\rho_{0.95} = (ED18) \times (TD15, TD18, TD16, TD17, TD20) + (ED24) \times (TD16, TD17, TD14, TD13, TD19, TD12) \quad (6.5)$$

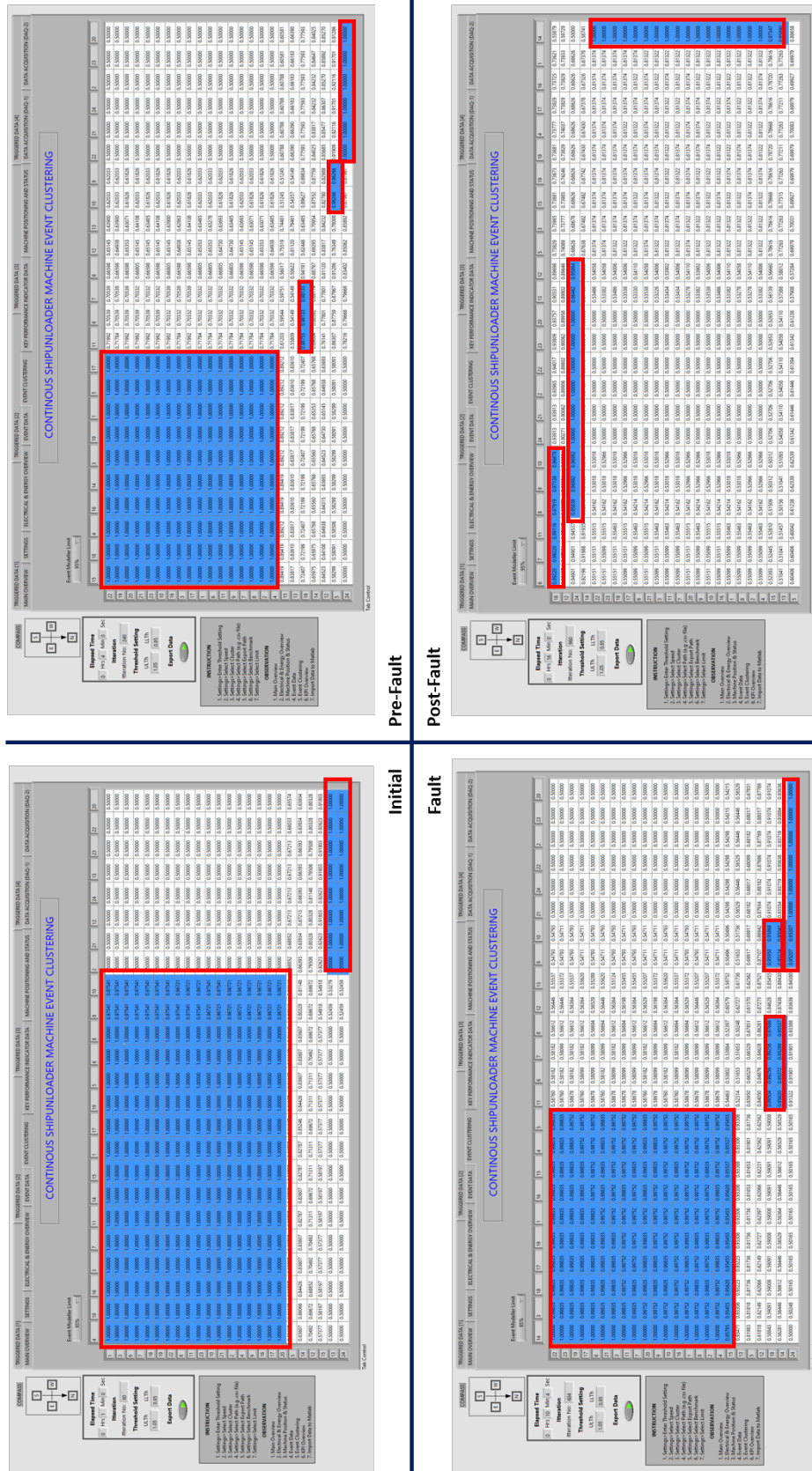


Figure 6.20: Rank Order Clustering - Cluster 1

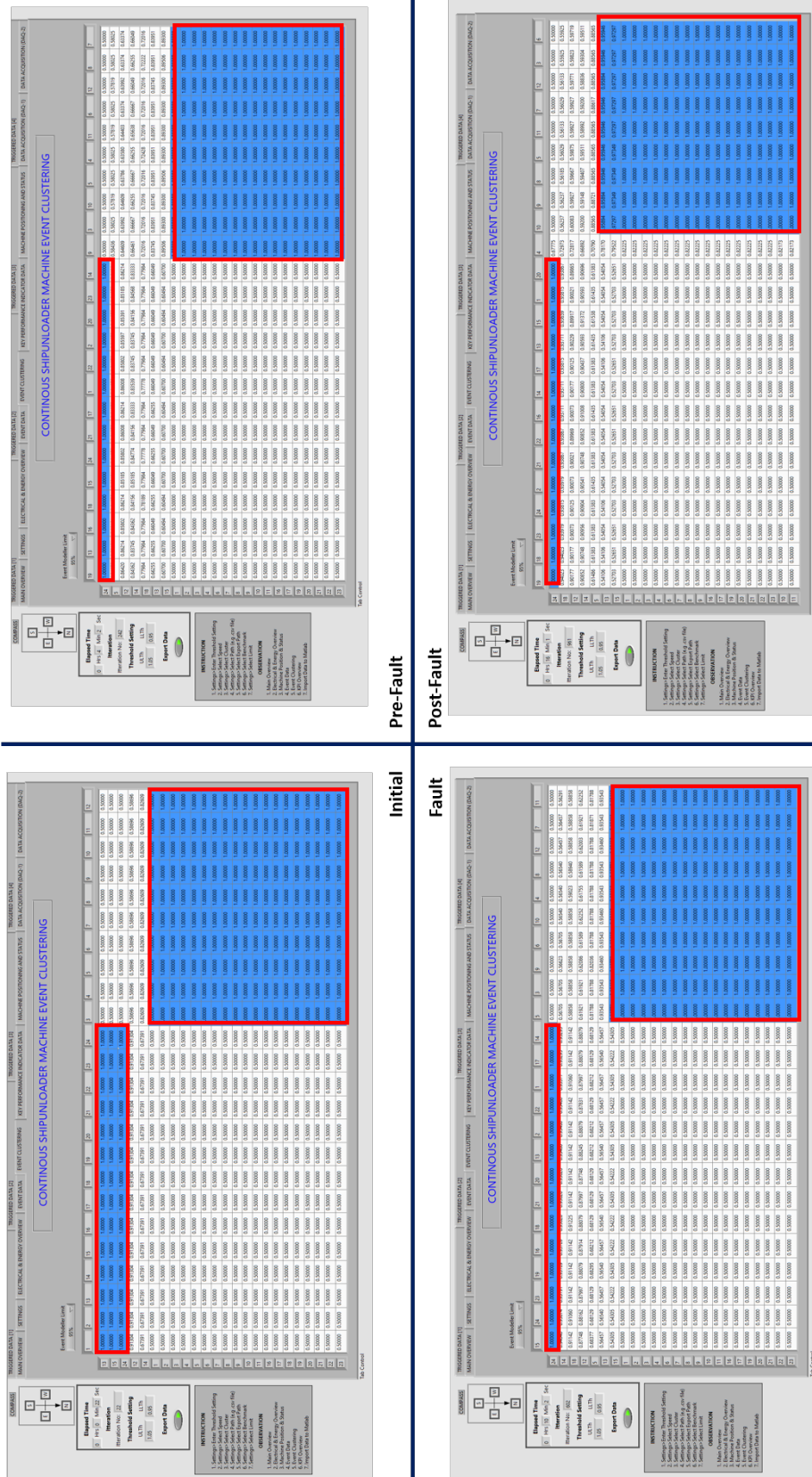
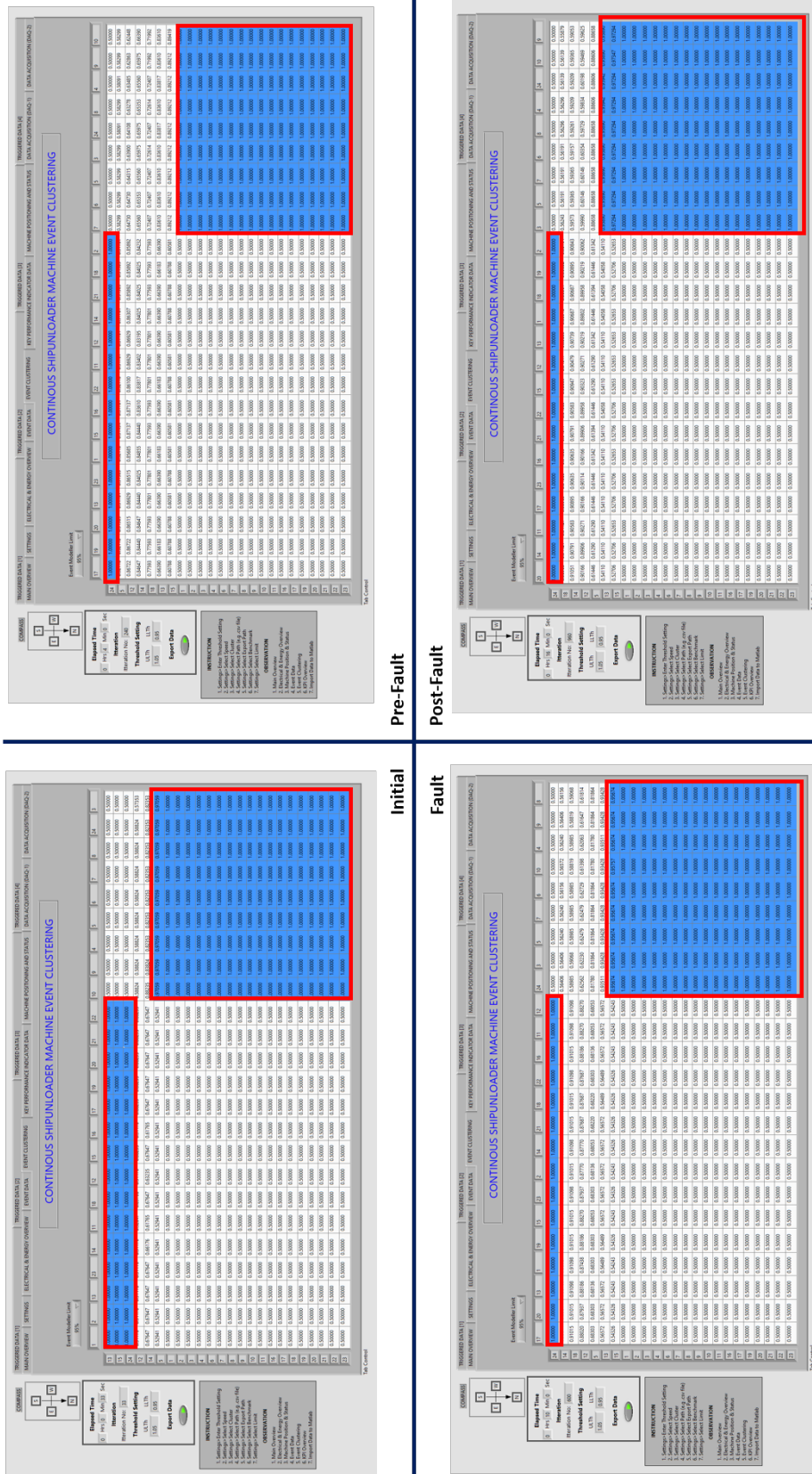


Figure 6.21: Rank Order Clustering - Cluster 2



Pre-Fault

Post-Fault

Initial

Fault

Figure 6.22: Rank Order Clustering - Cluster 3

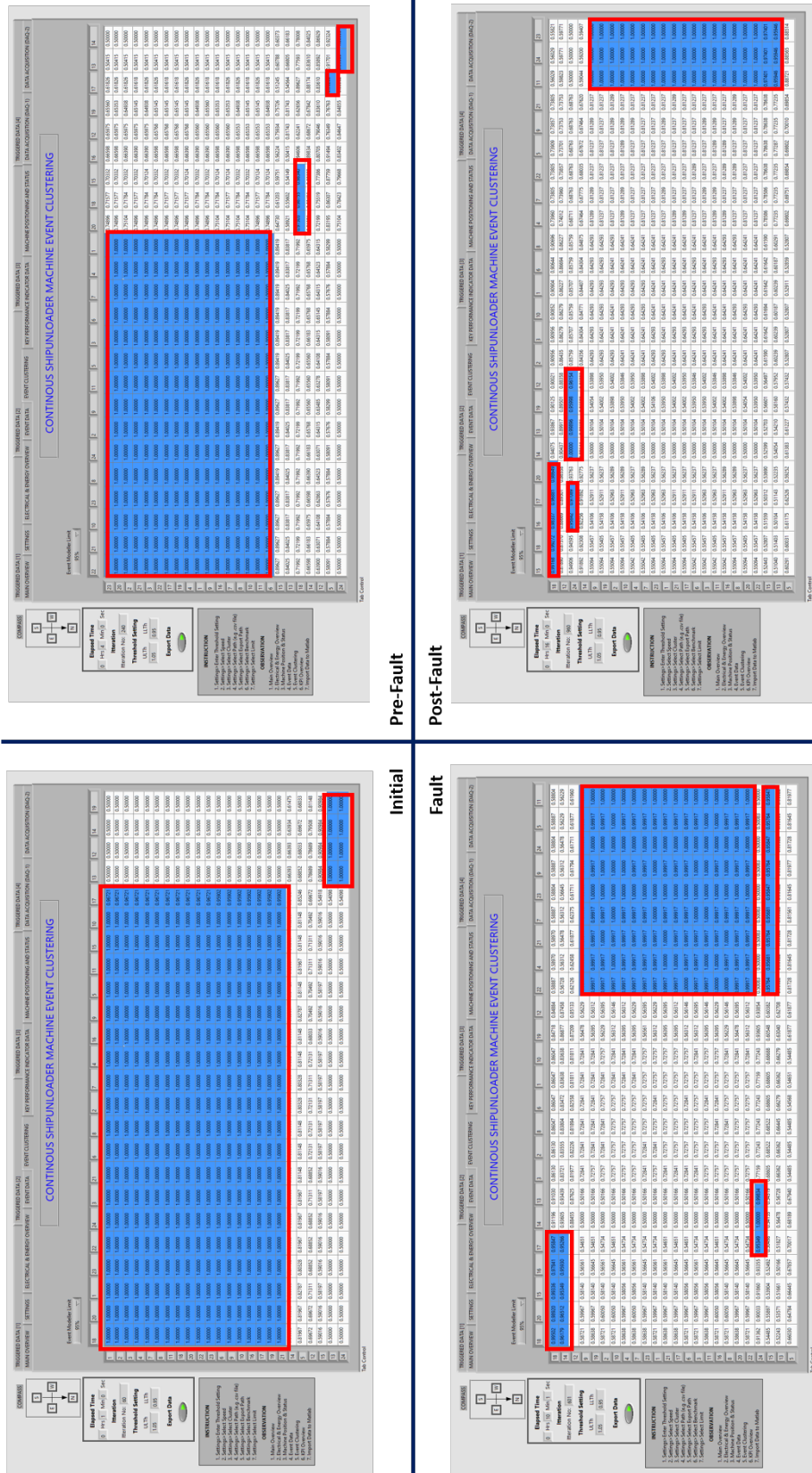


Figure 6.23: Rank Order Clustering - Cluster 4

$$\rho_{0.95} = (ED18 + ED24) \times (TD16, TD17) \quad (6.6)$$

Combining Equation 6.4 with Equation 6.6, the sensitivity index is expressed in Table 6.8. With that, Event Modeller has reduced the dimensional to a smaller scale that will help system engineer decide which relationship is liable to this fault. The following section will be discussed on the relationship between the observed data and the KPIs to validate the findings of the Event Modeller.

Table 6.8: Summary of Sensitivity Index

Description	TD8	TD9	TD10	TD16	TD17
ED18	0.97919	0.97138	0.96878	0.98337	0.96881
ED24	0.95838	0.96982	0.996982	0.95842	0.97089

6.4.1.3 Relationship between the observed data against Key Performance Indicators

The KPIs are the translated index of the machine operation as indicated at the beginning of the chapter. It provides operational information for the decision-making of both the machine and the system operator. Four main KPIs have been adapted in this experiment known as Availability (A), Instantaneous Utilisation (IU), Schedule Utilisation (SU) and Performance (P). Having these four KPIs in line with CSU machine data, the possible root cause of frequent harmonic failure could be determined. The following steps have been taken to analyse the data: (1) Select fault dataset from the Predictive Model Dataset. (2) Identify fault observed data location. (3) Distinguish observed data according to Environment Variables, Electrical Variables and Motor Variables. (4) Plot Environment Variables data against KPIs. (5) Plot Electrical Variables data against KPIs. (6) Plot Motor Variables data against KPIs.

The easiest way to identify fault events is by exploring the main incomer trip data. Theoretically, the entire electrical system will be tripped when there is a REGEN fault. Although it may conflict with other events in which the operator may trip the main incomer on purpose, the system could automatically distinguish between the actual fault and the simulated fault using the translated KPIs data. For example, when a fault occurred during the operation, the data of the KPIs will hold the value in percentage. This verifies that the fault that occurred

is genuine. If the machine is stopped or not operated, the value is either zero or hold to a specific value. In this example, the same dataset used in the previous experiment was chosen. This is to compare the relationship between the data observed using both techniques. The dataset dated 30th January 2019 was therefore sampled between 12:47 pm and 13:03 pm. The target data was then divided into three and plotted against the KPIs as follows:

Table 6.9: Summary of Electrical Variables

Description	ED1	ED2	ED4
Min	427.33 V	9.66 A	1.84 %
Max	427.47 V	242.93 A	2.19 %
Average	427.44 V	73.13 A	2.00 %

Figure 6.24 shows the electrical observed data against the KPIs when electrical faults occurred. This graph is quite revealing in several ways. First, the main fault has been observed. The chart reveals that the main fault occurred at 09:50. This is the turning point of all events. Next, the four KPIs are observed. Performance was initiated at 100% and remains for 1 minute. It was then exponentially reduced to 16% when the main fault occurred. This indicates that the unloading of coal has stopped immediately after 1 minute. On the other hand, both Instantaneous Utilisation and Schedule Utilisation started at 58% and gradually decreased to 35% at 04:24. At this point, Schedule Utilisation bounces back to 39% but Instantaneous Utilisation remains reduced to 30% until the main fault event occurs. The availability started at 95% and remained until 04:24 before it gradually decreased when the main fault occurred. These trends have shown that there is a relationship between the main fault and the KPIs. Table 6.10 summarised the KPI results.

Table 6.10: Summary of KPIs

Description	A	IU	SU	P
Min	64.37 %	25.01 %	34.82 %	11.52 %
Max	91.67 %	58.33 %	58.33 %	100.00%
Average	76.43 %	32.48 %	39.59 %	26.72 %

Moving to electrical variables, the main input voltage (ED1) remains at 427V for the entire time. Although the fault occurred, the main input voltage does not respond to this. This validates that the fault is not caused by power disturbances from the supply, such as transient, interruption, under-voltage or over-voltage. The Main Incomer Current (ED2) pattern was

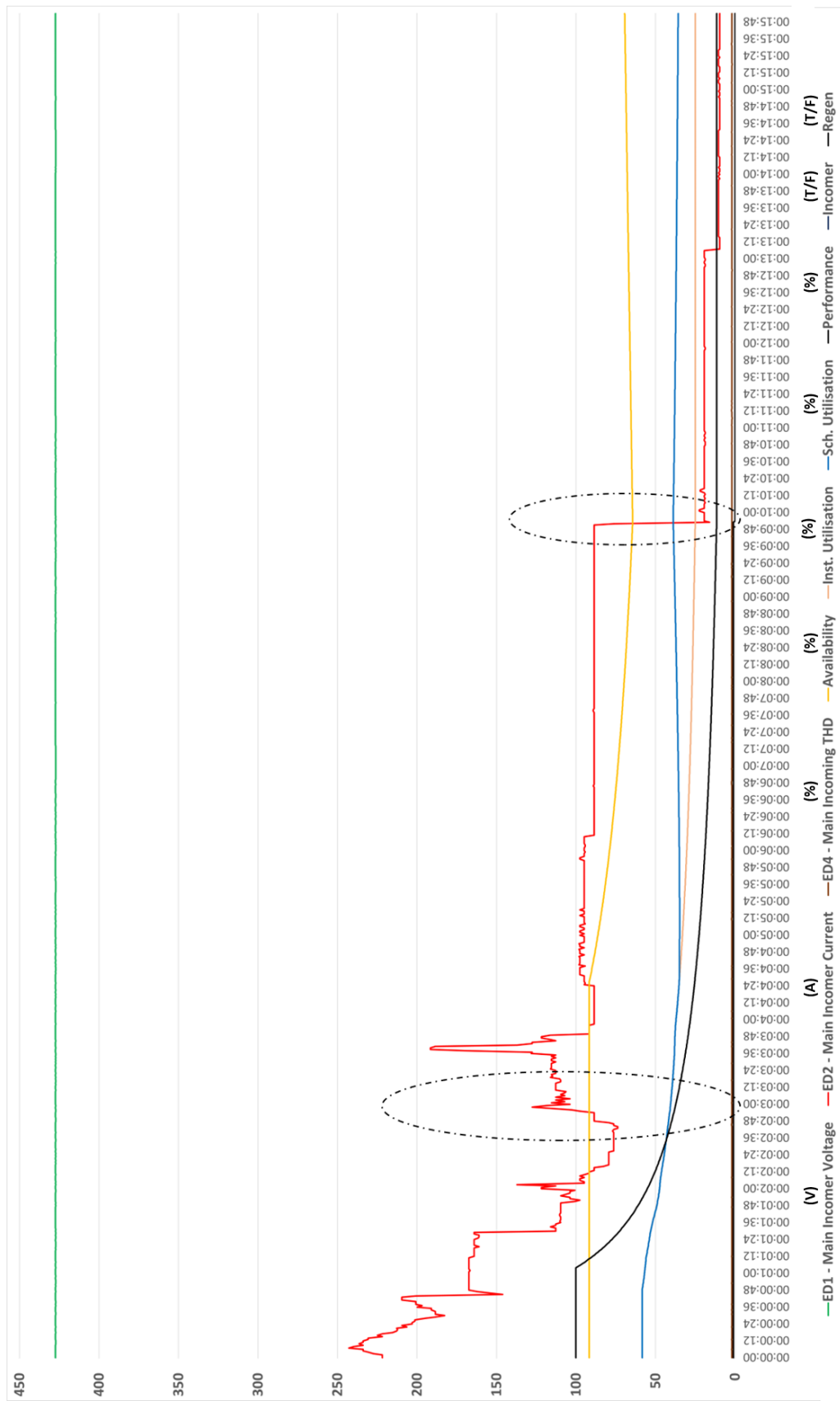


Figure 6.24: Electrical Variables vs KPIs

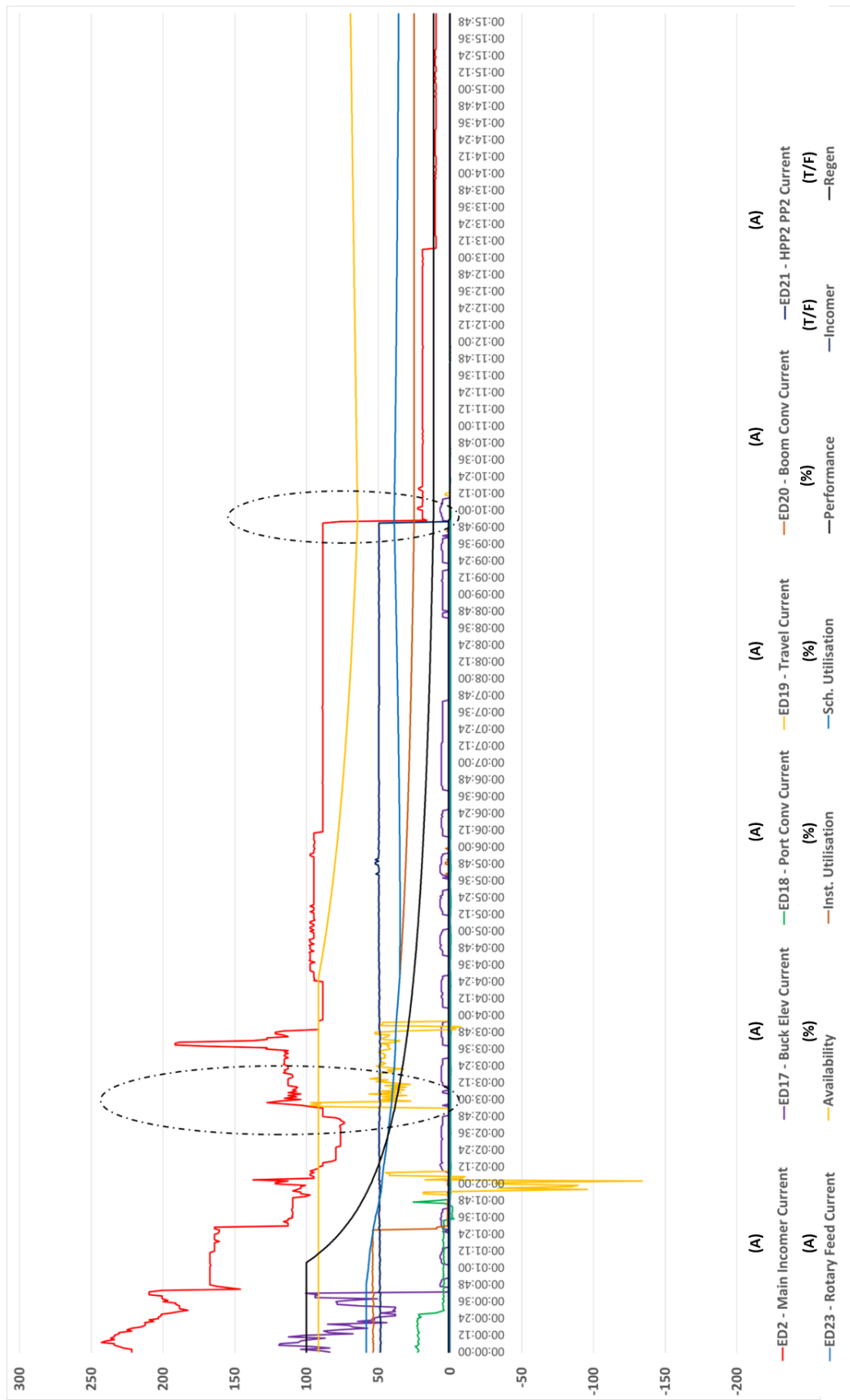


Figure 6.25: Motor Variables vs KPIs

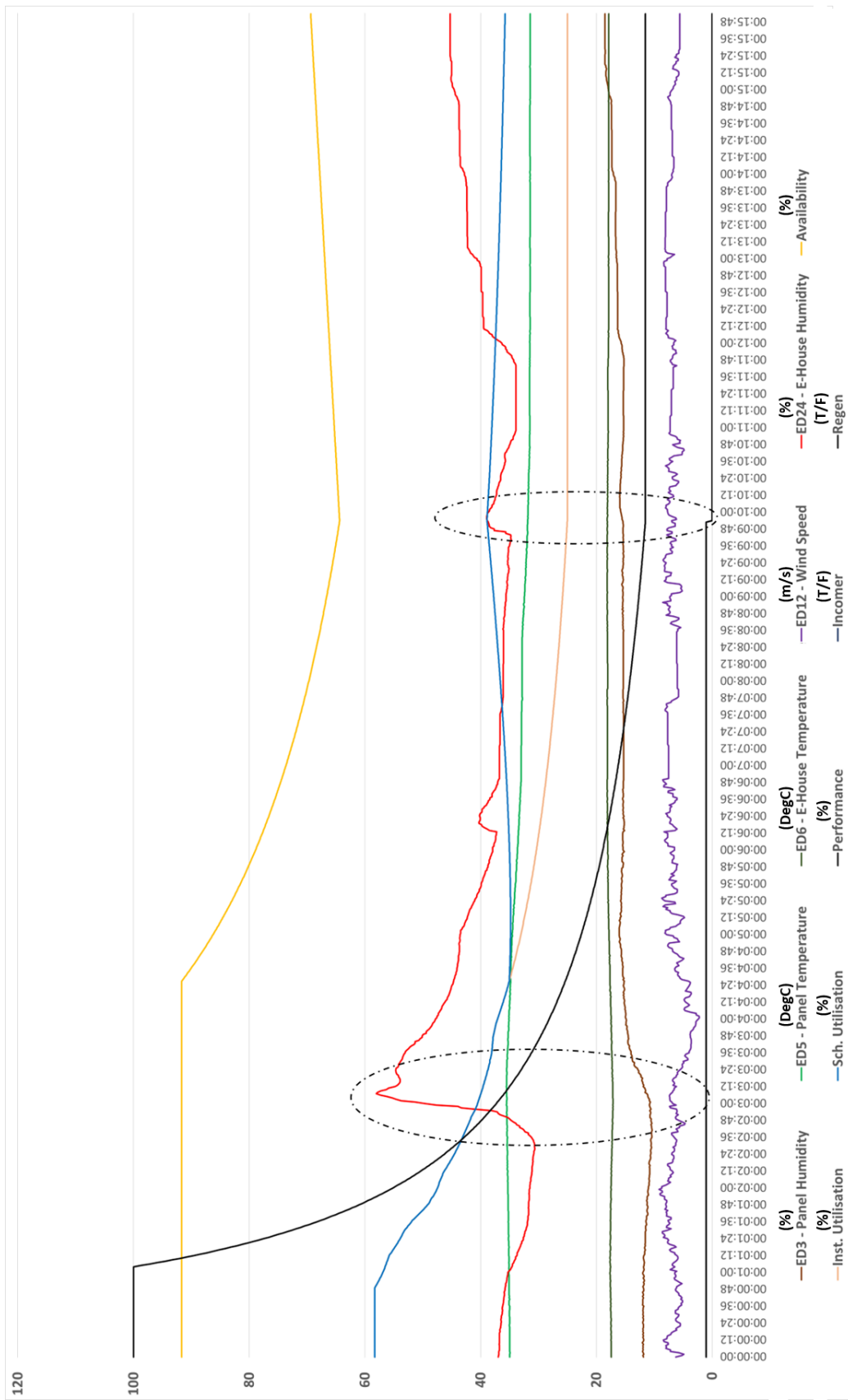


Figure 6.26: Environment Variables vs KPIs

then observed. Before the fault, ED2 shows the current drawing of the electrical distribution. The pattern is based on the loads of the motor during machine operation. However, when the fault occurred, trends show that ED2 is responding to the event. An instantaneous drop was discovered during this period. ED2 does not fall to zero because the energy was still drawn from the maintenance feeder for essential loads. The main incoming THD (ED4) has also been observed. As explained in Section 6.3.5, ED4 has been generated assuming that there will be no harmonic distortion during these events. The ED4 pattern remains at an average of 2.0% throughout the sampling time, indicating that the THD remains at IEEE 519 Standards. Table 6.9 summarised the electrical variables.

Table 6.11: Summary of Motor Variables

Description	ED17	ED18	ED19	ED20	ED21
Min	0.00 A	-2.96 A	-133.97 A	0.00 A	0.05 A
Max	119.31 A	25.67 A	97.87 A	54.41 A	52.05 A
Average	5.41 A	0.69 A	2.49 A	5.03 A	30.60 A

Meanwhile, to further explain the relationship between the motor variables and the KPIs, Figure 6.25 shows the current drawing of five main processes in the CSU machine. These include Bucket Elevator Current (ED17), Portal Conveyor Current (ED18), Travel Current (ED19), Boom Conveyor Current (ED20) and Hydraulic Pack Power (ED21). In order to validate the ED2 from the previous graph, this chart reveals that ED2 has a linear relationship with all motor variables within this graph. For example, when ED17 was drawn at the beginning of the sampling and all other motor loads were added during that period the ED2 measurement reached its peak value. Some attention is given to the small circle in the chart. It is clear that there was some instability happening at 01:48, which could be linked to the fault. ED18 had an impulsive signal and remained zero after that event. On the other hand, ED19 has also revealed some dramatic changes. Just after ED18 had an impulsive signal, ED19 experienced some negative distortion for 12 seconds, followed by an interruption for 48 seconds and a further positive distortion for 60 seconds completely before it stopped. Another attention has to be made at ED17. In the beginning, ED17 had drawn some fluctuation current and had spiked for a few seconds before it stopping at 00:40. The trend shows that ED17 is struggling to recover, but ends up stopping when the main fault has occurred. Otherwise, all other motor variables are operating in their normal state. Table 6.11 summarised the motor

variables.

Figure 6.26 displays the environment data observed against the KPIs when the same electrical faults have occurred. Five environmental data were reported, including Panel Humidity (ED3), Panel Temperature (ED5), E-House Temperature (ED6), Wind Speed (ED12) and E-House Humidity (ED24). Attention must be given to the humidity and temperature of the electrical room where the electrical distribution panel is located. The pattern has shown an instant increase of ED24 at 02:36. The humidity builds up and reaches its highest peak at 58% and slowly decreased with two spikes before the fault occurred. On the other hand, ED3 has also been increased but not as much as ED24. ED5 has an average temperature of 33.13 degrees Celsius, while ED6 has an average temperature of 17.82 degrees Celsius. Eventually, ED12 fluctuates within the normal range. Table 6.12 summarised the environment variables.

Table 6.12: Summary of Environment Variables

Description	ED3	ED5	ED6	ED12	ED24
Min	10.35 %	31.41 degC	17.12 degC	2.22 m/s	30.64 %
Max	18.52 %	35.54 degC	18.16 degC	9.19 m/s	58.06 %
Average	14.95 %	33.13 degC	17.82 degC	6.74 m/s	39.21 %

6.4.1.4 Discussion

The main outcome of the experiment is to show how the proposed technique can be used as a tool to reduce the dimensions of a highly complex system. In this case study, 96 TDs were analysed against 24 EDs in real-time to help system engineers visualise any changes to the system state. The EMDA implementation will group high correlation system parameters in the form of matrices and place them in mutually exclusive blocks. Having a huge matrix, however, will make things difficult. The matrix needs to be presented as a square for clear visibility. Thus, the 96 TDs were divided into 4 clusters, where each cluster has a relationship with 24 EDs. The example accessed a 16-minute dataset that belongs to a fault population. The result for each cluster was shown in a four-quadrant to show the pattern change. In this example, there is a major change in the pattern observed in Cluster 1 and Cluster 4. Further analysis on the two clusters suggested a new formula representing the state of the system. After the fault event, the new formula for Cluster 1 and Cluster 4 is expressed in Equation 6.4 and Equation 6.6 respectively. As a result of these equations, the main culprit of the fault

was identified. The ED variables were reduced from 24 EDs to 2 EDs (a decrease of 91.76%) while the TDs variable was reduced from 96TDs to 5TDs (a decrease of 94.79%). As listed in Table 6.8, the possible root causes of the fault are the combination of:

1. Portal Conveyor Current (ED18) with Rotary Feeding Table Motors (TD8), Bucket Elevator Motors (TD9), Bucket Elevator Brake Motors (TD10), Busy Rotary Table (TD16) and Busy Bucket Elevator (TD17) or
2. E-House Humidity (ED24) with Rotary Feeding Table Motors (TD8), Bucket Elevator Motors (TD9), Bucket Elevator Brake Motors (TD10), Busy Rotary Table (TD16) and Busy Bucket Elevator (TD17)

Further analysis with the KPIs will validate the existence of unknown events that could be linked to harmonic problem. In normal circumstances, electrical variables such as voltage, current and THD should have a linear relationship with the KPIs. As the KPIs responded to the fault, the electrical variables should also respond. However, the voltage and the harmonic measurement remain constant after the fault has occurred, as shown in Figure 6.3. Although current measurements have responded to this fault, trends show that the current value is not zero. It indicates that there was still a current drawing for the essential loads. Therefore, this eliminates the possibility of machine failure due to electrical variables.

Moving to the motor variables, as explained in the result, ED17, ED18 and ED19 showed some undiscovered events that could lead to fault. When comparing the performance trend to ED17, the performance measurement falls in line with the activity of the bucket elevator. Although the trends for ED17 fluctuate before it stops, this could be explained from the perspective of uneven coal inside the bucket. On the other hand, ED18 has experienced an impulsive signal that leads to an immediate stop. However, looking at the timing of this event, it took about 9 minutes for the fault to happen. The same goes for ED19, which experienced negative and positive distortions right after ED18. This can be claimed as a result of an event. However, when it comes to the type of operation, ED19 is a travel motor that operates in two directions. Although some distorted measurements were made, this could be explained by the number of motors that were operating. There were 10 motors in service working at different efficiency levels. Having said that, the motor variables is also excluded from being the root cause of the fault.

Turning to environment variables, the trends in Figure 6.26 have shown some indications that the humidity could be related to the fault. The rise in humidity up to 58.06% shows that the state of the system has been compromised. Although high humidity patterns started 7 minutes before the fault, the measurement continuously retained above 40 % after the fault. This type of phenomena takes a fixed amount of time to reach the fault level. To explain this phenomenon, we take an example of preheating an oven. When a person decides to bake a cake, the oven needs to be heated to the right temperature before putting the cake inside the oven. Although it takes a few seconds to turn the oven on, it may take a few minutes to get to the right temperature. This phenomenon is referred to a fixed time delay deterministic event. However, if the person warms up the oven, but forgets to close the oven door, it may take longer to get to the right temperature. This phenomenon is referred to a deterministic sequence of different input events. The heat produced in the oven was cooled by the amount of cold air present in the surroundings, which would delay the process of preheating the oven.

When we relate this to our case study, the power electronic component such as choke, rectifier, REGEN and control drives produce an excessive amount of heat during operation. To ensure that the heat is kept at certain limits, two units of air conditioners were designed to combat this heat. However, after considering the amount of heat generated in the room along with the hot and humid climate of Malaysia throughout the year, experts have reviewed and decided to install two additional air conditioners to cater for this amount of heat. As a result, when CSU machine is fully operational, the four-unit air conditioners are sufficient to cool down the electrical room. However, when the machine stops at intervals, the air conditioners are found oversized to the power electronic component. As a consequence, the humidity increases; the component becomes wet, resulting in machine failure. The humidity in the electrical switchgear room must be maintained at an elevated temperature relative to the ambient inside the room. Condensation is usually considered a problem only if the humidity in the switchgear room is 65% or greater [279]. Although the humidity in this example does not reach 65%, considering the machine has been running for a decade, the efficiency of the machine may decrease as humidity may be the contribution factor. This finding suggests that there is a relationship between the humidity and the fault.

The relationship between the observed data and the Event Modeller and the KPIs appears to be consistent. The findings of the two analyses suggested that there is a new unknown

parameter, which shows a reasonable correlation with the machine's performance. The unknown parameter is that the humidity of the operating environment has a significant impact on the occurrence of harmonic failure. This proved the hypothesis set in the underpinning research endeavour presented in this thesis. By controlling the operating environment's humidity, the event-based machine learning (EMDA) discussed in the following section could precisely classify the system state issue.

6.4.2 Experiment 2: Machine Failure Analysis with Predictive Model

Reviewing the purpose of this study, the objective is to deploy machine learning platforms that could predict the variables causing machine failure in real-time. As discussed earlier, machine failure may be due to a variety of reasons, such as climate change, device failure, system interruptions, or operator handling. Therefore, in order to decide on future events, a supervised learning model was used to analyse the relationship between these factors, thus predicting which pattern the system state represents in real-time. In supervised learning, training sets are labelled with the information of the system state determined by the predictive model dataset. In this experiment, three machine learning classifier models were chosen, namely, Decision Tree (DT), Multilayer Perceptron Neural Network (MLPNN) and Naive Bayes (NB), which will be discussed further in the following section.

6.4.2.1 Experiment Strategy

Building a solution that is capable of classifying pattern based on the predictive model is another key motivation of this thesis. The proposed EDMA technique are designed to provide information of the system state in real-time, which help system engineer to be alert with the future fault or unnoticed events which could harm the system. EMDA is a unique technique which is looking at the coincidence of the events with its triggered data. In order to ensure that machine learning is provided with sufficient machine information, the predictive model dataset in Section 6.3.9 has been taken and labelled according to the pattern. The data sampling for each pattern was then taken and stored in a new data set to run machine learning models. The same procedure has been used to validate this by using the raw dataset population in Section 6.3.2. The reason why we sampled the same population is to compare the accuracy of the classification for both methods. The experiment strategy is summarised in Figure 6.27. The

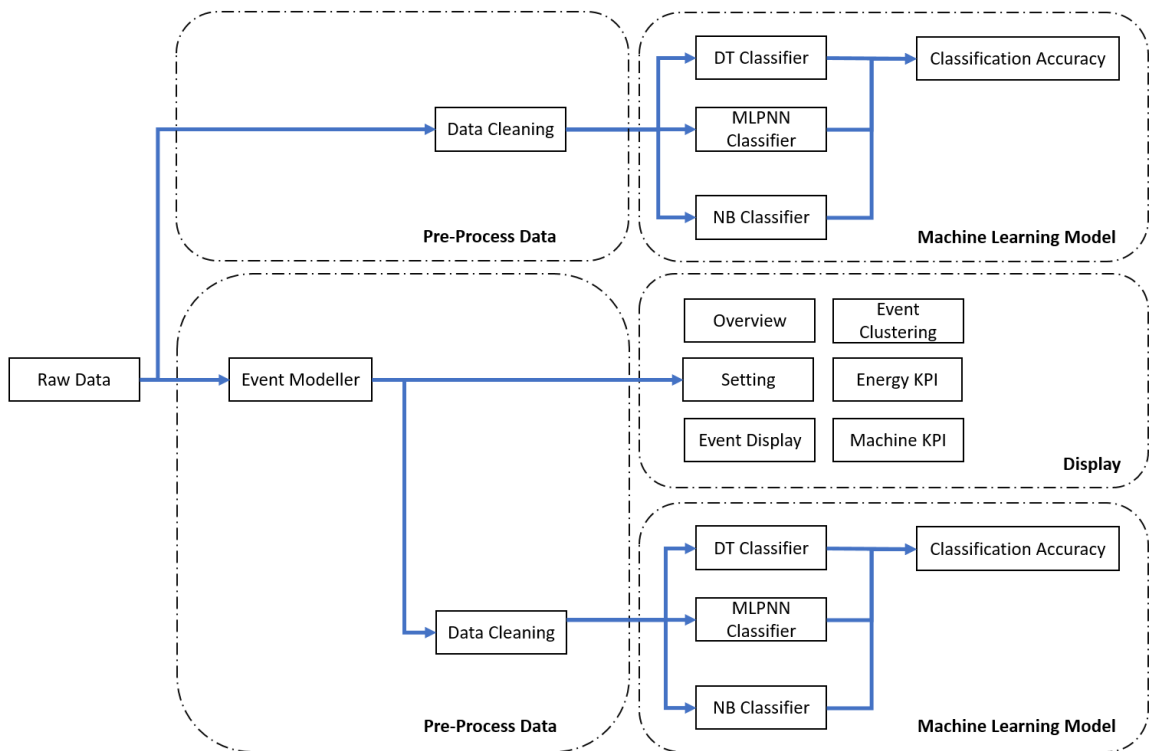


Figure 6.27: Experiment Strategy for Predictive Modelling Analysis

implementation of this experiment is achieved by using MATLAB R2019a Neural Network Toolbox [43] and by experimentally setting the parameters. Classification results for both experiments will be discussed in the following section.

6.4.2.2 Classification Results

The details of the experiment are given in Table 6.13. A total of 2400 data were obtained from the original raw dataset and predictive model dataset and divided into training data and testing data. The results of three different classifications were presented in Table 6.14. It has been observed that the majority of the classification accuracy for Decision Tree and MLPNN Classifiers has been 100%. This shows that the responses of the original dataset and the EMDA dataset are consistent. However, the results using the Naïve Bayes classifier show that the original data set is slightly less accurate but still showing a good indicator of high accuracy. Surprisingly, the EMDA dataset has achieved better accuracy than the original dataset. This has shown that EMDA is performing well, which makes it a good technique for machine learning tools.

Table 6.13: Experiment Details for CSU Machine with 15% Holdouts

Description	Data	Variables	Predictors	Train	Test
Original	2400	124	8	2040	360
EMDA	2400	580	8	2040	360

Table 6.14: Classification Result for CSU Machine with 15% Holdouts

Description	DT		MLPNN		Naïve Bayes	
	Original	EMDA	Original	EMDA	Original	EMDA
Sample 1	100.00 %	100.00 %	100.00 %	100.00 %	97.50 %	100.00 %
Sample 2	100.00 %	100.00 %	100.00 %	100.00 %	93.67 %	100.00 %
Sample 3	100.00 %	100.00 %	100.00 %	100.00 %	99.17 %	100.00 %
Sample 4	100.00 %	100.00 %	100.00 %	100.00 %	96.67 %	100.00 %
Sample 5	100.00 %	100.00 %	100.00 %	100.00 %	94.72 %	99.72 %
Sample 6	100.00 %	99.72 %	100.00 %	99.72 %	97.50 %	98.61 %
Sample 7	100.00 %	100.00 %	100.00 %	100.00 %	96.67 %	100.00 %
Sample 8	100.00 %	100.00 %	100.00 %	100.00 %	95.27 %	100.00 %
Sample 9	100.00 %	100.00 %	100.00 %	100.00 %	97.22 %	100.00 %
Sample 10	100.00 %	100.00 %	100.00 %	100.00 %	96.94 %	100.00 %

Hold-out is when you split your dataset into a 'train' and 'test' set. The training set is what the model is trained on, and the test set is used to see how well the model performs on unseen data. In the previous experiment, 85 % of the data was used for training, and the remaining 15 % of the data was used for testing. The experiment was repeated with a 30 % holdout to observe the consistency of the outcome. The details of the experiment are shown in Table 6.15. The comparison of results for the three classifiers is summarised in Table 6.16.

Experimental results have shown that both the EMDA and the original data show good accuracy. This suggests that the model has been classified according to the predictive model pattern. Although the results of this analysis sound convincing, they are not as good as we thought. Getting 100% accuracy is questionable but justifiable. Since this is a supervised learning, the labelling of data may not be as accurate as it is. The reason for this is that the labelling is based on a series of events. For example, when a fault occurs, some event must be triggered before it happens. As a result, not only the fault time stamp was labelled as an interruption, but the whole sequence of events. This leads to an incorrect time span of the data. The same applies to weather conditions such as rain and strong winds. Adverse weather conditions happen during a period of time. As a result, the whole period will be labelled as weather issues. In addition, the data in the study represents the operation of a

machine. There are variables that haven't changed as much as they do. For example, the voltage stays at 427.33V most of the time. When we train them in machine learning, there was no significant difference between the 8 patterns as they remained in the same state. The linearity of these data makes the classifier able to accurately predict most of the time. To overcome and comprehend these situations, further experiments can be carried out using other strategies, such as cross-validation or other methods of sampling industrial data in real-time.

Table 6.15: Experiment Details for CSU Machine with 30% Holdouts

Description	Data	Variables	Predictors	Train	Test
Original	2400	124	8	1680	720
EMDA	2400	580	8	1680	720

Table 6.16: Classification Result for CSU Machine with 30% Holdouts

Description	DT		MLPNN		Naïve Bayes	
	Original	EMDA	Original	EMDA	Original	EMDA
Sample 1	100.00 %	100.00 %	100.00 %	100.00 %	97.80 %	100.00 %
Sample 2	100.00 %	100.00 %	100.00 %	100.00 %	97.36 %	100.00 %
Sample 3	100.00 %	100.00 %	100.00 %	100.00 %	98.19 %	100.00 %
Sample 4	100.00 %	100.00 %	100.00 %	100.00 %	98.19 %	100.00 %
Sample 5	100.00 %	100.00 %	100.00 %	100.00 %	96.52 %	99.72 %
Sample 6	100.00 %	99.86 %	100.00 %	99.86 %	97.78 %	98.75 %
Sample 7	100.00 %	100.00 %	100.00 %	100.00 %	96.80 %	100.00 %
Sample 8	100.00 %	100.00 %	100.00 %	100.00 %	95.55 %	100.00 %
Sample 9	100.00 %	100.00 %	100.00 %	100.00 %	97.22 %	99.72 %
Sample 10	100.00 %	100.00 %	100.00 %	100.00 %	97.77 %	100.00 %

6.4.2.3 Computational Effort Results

Putting aside the batch labelling issue and static sensory data discussed in the previous section, the classification results for both EMDA and original data shows that all three ML classifiers achieved high accuracy. A major difference between these three ML classifiers resides in its computational effort. To ensure the setup run under the same experiment conditions, the analysis were running on a Personal Computer with Intel®Core™ i7-7820HQ CPU 2.90GHz and 16.00 GB memory RAM. In this experiment, the computational effort was measured using its time span and CPU utilisation. Table 6.17 presents the average time span for all experiment. The results reveal that MLPNN took the least computational time compares to DT and NB.

Table 6.17: Time Span Computational Effort Results

Description	DT		MLPNN		Naïve Bayes	
	Original	EMDA	Original	EMDA	Original	EMDA
Sample 1	5.38 s	4.71 s	3.85 s	4.00 s	7.77 s	8.04 s
Sample 2	5.03 s	4.58 s	3.73 s	3.83 s	7.52 s	7.89 s
Sample 3	5.02 s	4.54 s	3.94 s	3.96 s	7.73 s	8.02 s
Sample 4	5.59 s	4.69 s	3.86 s	3.95 s	7.76 s	8.17 s
Sample 5	5.22 s	4.92 s	3.79 s	4.01 s	7.80 s	8.26 s
Sample 6	5.12 s	6.08 s	3.80 s	4.18 s	7.54 s	8.01 s
Sample 7	5.29 s	4.53 s	3.85 s	3.96 s	7.56 s	8.07 s
Sample 8	5.23 s	4.87 s	3.90 s	3.92 s	7.66 s	7.37 s
Sample 9	5.15 s	4.43 s	3.64 s	3.55 s	7.85 s	7.22 s
Sample 10	4.93 s	4.39 s	3.44 s	3.53 s	7.59 s	7.99 s
Average	5.20 s	4.78 s	3.78 s	3.89 s	7.68 s	7.90 s

Table 6.18: CPU Utilisation Computational Effort Results

Description	DT		MLPNN		Naïve Bayes	
	Original	EMDA	Original	EMDA	Original	EMDA
Sample 1	67 %	67 %	41 %	40 %	78 %	84 %
Sample 2	66 %	59 %	40 %	37 %	62 %	71 %
Sample 3	68 %	63 %	39 %	35 %	59 %	69 %
Sample 4	70 %	66 %	36 %	34 %	69 %	78 %
Sample 5	66 %	61 %	38 %	36 %	74 %	75 %
Sample 6	66 %	69 %	35 %	41 %	69 %	75 %
Sample 7	69 %	65 %	40 %	40 %	71 %	80 %
Sample 8	69 %	64 %	41 %	37 %	76 %	81 %
Sample 9	69 %	60 %	35 %	34 %	77 %	80 %
Sample 10	74 %	61 %	39 %	35 %	74 %	77 %
Average	68 %	64 %	38 %	37 %	71 %	77 %

MLPNN spent an average of 3.78 seconds for original data and 3.89 seconds for EMDA data. Although it sounds original data classify faster by less than 0.1 seconds, knowing that EMDA data have high complexity compared to the original data, it is an achievement for MLPNN in classifying EMDA data. On the other hand, EMDA data in DT have a different perspective. It is notable that EMDA classified faster by 0.4 seconds compares to its original data. These findings may explain the relatively good structure of DT in branching the possible outcomes, which has an important implication for future work. Meanwhile, NB took the longest time, which almost doubles the time span for MLPNN.

Table 6.18 reveals the average CPU utilisation for all experiment. A similar outcome with the time span, MLPNN have the cheapest computational effort by only using 38% of the

CPU utilisation in classifying the model. On the contrary, DT and NB need almost double of the MLPNN effort. Comparing the original data to EMDA data, DT computes less effort in EMDA data compared to original data. This is consistent with the time span results discussed earlier, which shows that EMDA data are relevant to DT. Alternatively, while MLPNN has almost similar effort for both original and EMDA data, NB works best with its original data.

In conclusion, the results of this study indicate that MLPNN is the best classifier that works with original and EMDA data. With its low computational effort and reduced time span, MLPNN has shown a positive implication for the proposed EMDA technique. DT, on the other hand, are still relevant due to its success in handling EMDA data and acceptable time span, but needs to further research in optimising the CPU utilisation. Unfortunately, NB remains the least preference classifier due to its expensive computational effort.

6.5 Summary

Chapter 6 demonstrate the use of event based technique in solving the repetitive harmonic failure in one of the power plants in Malaysia. The aim of this work was to analyse the machine failure using the coincidence matrices of an input-output relationship with the latent knowledge (KPI) translated into the system. To find the root cause of this failure, the novel technique, which is namely EMDA, was examined on the real industry data. The system development implementation were developed using LabVIEW application, accessing 914 raw datasets with each dataset containing 120 sensors/seconds. To label the predictive model dataset, EMDA were used to convert the raw dataset into sensitivity index and KPIs before it was trained using 3 machine learning model known as Decision Tree, Multi-Layer Perceptron Neural Network and Naïve Bayes. The accuracy of the classifications and the computational effort were compared. Meanwhile, further analysis of the CSU machine against the REGEN failure was established. The observed data were compared with the KPIs and the Event Modeller ROC pattern. The results obtained suggests that humidity have a significant impact on the occurrence of harmonic failure. This proved the hypothesis underpinning this research which is *"A fault in a power system distribution will be influenced not only by internal events, but also by external events such as environment and climatic change"*. In terms of computational effort, the MLPNN classifier has found to perform better compared to DT and NB, which conclude the findings.

Chapter 7

Conclusion

7.1 Overview

Chapter 7 summarises the entire works carried out in this research. The aim of this chapter is to highlight the research outcomes and findings. Building on the findings, research limitations and proposals for future work are presented. This chapter comprises of five sections. Section 7.2 summarises all chapters of the thesis. Section 7.3 presents the contributions of the research, which will be of benefit to real-time industrial application. Limitations of the study are discussed in section 7.4. Finally, based on the limitation and findings, this thesis proposes future work in Section 7.5.

7.2 Summary of Dissertation

This thesis started with introducing the subject, defining the problem statement and establishing the research question in the domain problem: Power Quality (PQ). Managing PQ in machines using industrial data is very challenging. Industrial data are multi-dimensional data in nature and deterministic. The problem with analysing this data is that the machine events behave differently and requires expensive computational effort. The current practice embrace signal processing technique, which use various combinations of feature extraction, feature selection and classifier to locate or predict the PQ problem. However, this technique is solely based on the time domain and frequency domain of an electrical signal. It does not consider the external parameter, which has a significant influence on the PQ problem. For example, PQ

engineer used government meteorological websites or other weather data websites to obtain site temperature and humidity level separately from the target data. These websites provide offline data, and future data is based on prediction, which does not represent the accurate measurement at site. Since there was no research investigating the relationship between the PQ data with other external data in real-time, a novel machine learning architecture framework namely, Event Modeller Data Analytics was proposed in this research, to visualised the PQ system inputs and machine Key Performance Indicator (KPI) occurrences with predictive data analytic approach in real-time. As stated in Chapter 1, the research question for this study is defined as follows:

Table 7.1: Research Question

No	Research Question
RQ 1	What are the internal and environmental events within the power system distribution that affect harmonic filter performance? By internal events, we mean the dynamic and the interaction within plant machinery, and by environment event we mean the fluctuation in temperature, humidity, and pollution in the vicinity of the plant.
RQ 2	Is the bath tub theory of harmonic filters performance a reliable mechanism for predicting its performance at the three periods: (1) an “early failure” (burn-in) period, where the hazard function decreases over time, or (2) a “random failure” (useful life) period, where the hazard function is constant over time, or (3) a “wear-out” period, where the hazard function increases over time.

Chapter 2 has provided the literature review on the research areas. The domain PQ disturbance problems with its various techniques and methods to solve the problem have been discussed therein. The relevance of input variable selection and sensitivity analysis within this context have been associated with this research, to challenge the ways of feature extraction from the modern complex systems and its infrastructure to achieve time-critical reduction and low computational effort. The methods used in this research were identified beforehand based on the review from the literature.

Event modeller data analytics has been introduced in Chapter 3. This technique used event-based learning to connect the embedded system with its Key Performance Indicator (KPI) with machine learning tools for prediction purpose. It used a dynamic platform to produce a real-time critical accurate knowledge that represents the system state at a minimum cost and then further trained to make a prediction based on the pattern.

Chapter 4 discussed the main issue faced by our industrial partner in Malaysia. It was pointed out that the CSU machines, being responsible for transporting the coal from the vessel to the supercritical boiler of a power plant, is experiencing a major frequent harmonic filters failures. The exciting part is, results of the PQ assessment reported in this Section 4.4 have shown that all PQ parameters are within the tolerance value as per IEEE 519, 1992 standard and International Standard IEC 61000-3-4. The majority of the measured harmonic distortion at the PCC is relatively small, achieving below 2 % for voltage distortion and demand distortions. These findings suggest that the problem have other causes, potentially associated with other variables within the vicinity of the operating machines. Thus, the Event Modeller Data Analytic technique introduced in Chapter 3 was used to demonstrate this problem in Chapter 6. In order to collect accurate data from the industry whilst maintaining the client policy, an offline data collection setup has been made to access and capture real-time data from the machine. This chapter also explained the modification made to incorporate a limitation on the existing infrastructure to include additional external variables that may have an impact on the research.

Chapter 5 has introduced event modeller technique in the form of software in the loop architecture using synthetic data. This chapter has presented a real-time data simulation of Continuous Ship Unloader machine using three different environments which are pre-disturbance environment, k-disturbance environment and post-disturbance environment. The results have shown common findings during the transition between pre-disturbance to disturbance period, having the same average reduction trends of 8.2 % for Static TD and 15.94 % for Dynamic TD. Alternatively, it also shows a common finding during the recovery period from disturbance to post-disturbance with average increment trends of 3.218 % for Static TD and 6.023 % for Dynamic TD. With these outcomes, Dynamic TD has been chosen to be the proper method to capture the changes of input sensitivity which was used to demonstrate the problem in Chapter 6.

Chapter 6 comprises of the main experiment and findings of this research. The main focus of the chapter was to identify the root cause of the CSU machine abnormalities, which links to a harmonic failure problem. Dynamic platforms that integrate the developed Event Modeller technique with its Key Performance Indicator (KPI) had been examined. With this information, data analytics is developed to predict the homogenised system parameter that

represents the current state of the system in real-time. This chapter comprised of a set of an initial experiment to construct the predictive model dataset. The dataset is classified into 8 patterns, in which each pattern represent the truth table of the status of machine operation, abnormality state and weather conditions. To validate, each pattern is tested using the industrial data, to formulate the equivalent system state relationship that represents this pattern. Within the system development capacity, the event modeller clustered the high correlation parameter together, which help system engineer to identify which group of parameters have high influence at the current time. With the help of data analytics, the KPI parameters are used as predictors to determine which pattern is it belongs to. The accuracy of the prediction is good which prove that the proposed technique could be used in any complex data in the future. With regards to evaluating the external parameter that has the influence to the system parameter, which leads to CSU machine failure, the work in Chapter 6 has been introduced with the three modelling strategies in the experiments.

7.3 Contribution

The contribution of this research is two-fold. Firstly, this research offers software in the loop framework that connects a real-time embedded system data with data analytics. Secondly, it contributes knowledge to system engineers in formulating complex systems into an event-based environment.

7.3.1 Contribution as a System Engineering Tools

The outcomes of this research have contributed to some practical tools for system engineering in terms of the methods and strategies used in this research. The event modeller data analytics technique, which runs in real-time combines the cause-effect relationship between the input variables and performance measure to developed a pattern. With the help of various machine learning algorithms, these tools could potentially predict linear and nonlinear problems in real-time, enhancing quick decision making for system engineering problems. The first notable contribution in this field is that the Event Modeller Data Analytics technique is robust, workable with the dynamic and autonomous environment while being able to withstand real industry data which is non-linear and highly influenced with various factors. The

experiments in Chapter 5 are tested with highly volatile and random values upon the system is disturbed. The results show that this technique is capable to react to this system state instantly and substantially leads to better solutions in improving the system's performance. The second notable contribution is that this proposed technique could visualise the group of highly correlated variables in the form of occurrence matrices in real-time, proposing a mathematics equation that represents the current state of the system. The results demonstrated in Chapter 6, which used high complexity industrial data have to substantiate grouped the high correlation parameter together in a significantly reduced time, formulating a non-linear equation with its accurate weighting. The applicability of combining the event modeller with an analytic data system is the third notable contribution in this research. This platform which was trained using various machine learning tools is able to predict the future events accurately, making these tools relevant to machine learning and system engineering environment.

7.3.2 Contribution in Power Quality Disturbance

Discovery of system state pattern could draw interest from a power system perspective in diagnosing PQ disturbances. From the clustering arrangement suggested by event modeller technique, the pattern of the abnormalities which considers external parameter can be studied by power system engineer to predict PQ problems. The main objective of this research was to explore which external variables have a significant correlation with the event of harmonic failure. The environment variable such as temperature, humidity and wind speed were included in the system analysis, weighting the correlation relationship of these variables with system operation variables, to formulate new knowledge information. The end result of Event Modeller technique was in the form of a cluster of the input-output relationship. The clusters were studied for the significance of this input-output relationship to see the association between them in real-time. In normal operation conditions, the resultant cluster found by Event Modeller has the same agreement with the actual events provided by the daily operation report. This has proven the significance of the findings. With the information from the key performance indicator translated within the system operation, these calculated variables validate the input-output relationship for data analytics purposes. The outcome of the results has shown good accuracy, which makes EMDA technique relevant not only in solving PQ disturbance problem but also across other industrial controls and monitoring problems.

7.4 Limitation

This research produced a reliable system modelling for measuring non-linear industrial data. However, like any other research, a number of limitations existed in conducting the research. There are four limitations identified in this research. Please note that the scope and objectives of the research are within acceptable tolerance limits.

7.4.1 Total Harmonic Data

This research used primary data which were retrieved from the control system within the CSU machine. One main issue is that the Total Harmonic Distortion (THD) data is not available within the control system architecture. As explained in Chapter 4, various efforts have been made to capture the THD data from the Multi-function meter. However, none of the initiatives made a positive outcome. The initial plan was to add a new analogue output module to import the THD data in line with the existing Voltage and Current data. Based on the datasheet, the device is capable of providing up to 4 analogue output. As a result, binary data which represent the alarm status of THD setpoint is imported rather than the analogue value of the THD. Then, another method was adopted using a Modbus communication protocol. As the CSU machine operates throughout the year, there is a limitation on the maintenance window to perform the job. After a number of delays in installing the device, the Modbus communication does not respond to the data acquisition, which makes the THD data not available for this research. The implication of this missing data affects the main objectives of the study. After further review with the system expert, a sample of THD data has been taken offline. Hence, the THD data was simulated based on the sample data and International Standard IEEE 519, assuming the machines are operated within the tolerance limits.

7.4.2 Data Benchmarking

The scenario of events in the actual system varies according to the system characteristic and its system deterministic. An event could happen instantaneously, or it may have some delay before it reached to the target output. For example, the scenario of a room temperature does not change immediately when the heater or air conditioner was switched on. In comparison, a dark room will be immediately bright when the switch was turned on. Both of the scenarios

have made changes to the system state at a different pace. In order to detect this, a data benchmarking is required. In this research, data benchmarking was used to set the limits of typical operation characteristics. If an event occurred outside this limit, the event modeller registered it as a triggered event. As explained in Chapter 6, data benchmarking decision has been made based on the historical data and system expert point of view. A sample data has been selected randomly from the entire data population. The average and limits of the sampled data have been analysed and reviewed by the system expert to decide on this setting. However, these limits may and may not cover the actual scenario of events. The implication of this benchmark-setting may affect the performance of grouping the high correlated system parameter together. Consequently, the outcome of the results may not reflect the actual system state.

7.4.3 Data Quality

The data used in this research were gathered during the data collection period. As explained in Chapter 4, due to company policy which does not allow data to be available on the cloud, a data logger has been configured to store the target system parameter in the server. When the data is packet into a DAT file, the author took approximately 6 hours to convert a single file with no interruption; otherwise, it has to start all over again. Data screening found 3.72% of the data population was missing. The implication is that this missing data may represent some of the abnormality datasets which is the purpose of the study.

Besides that, there were also missing data points within the dataset. A further investigation found that whenever the main incomer trips, the entire control system were interrupted. Thus, there were no data stored on the server for system modelling. The implication is that the data were skipped for the duration of the interruption, which will make a significant impact on the quality of the data. In addition, some data were found to be frozen for a period of time. In this scenario, it holds the last value, which compromises the integrity of the data.

7.4.4 Reports

As the data collection has been made for the period of 15 months, additional supporting data is required for cross-reference with the actual data to interpret what is happening. Therefore, a daily report has been accessed to determine the operational activity for each day. Nvivo

software has been used to organise, store and sort the daily report in a single platform and further analysed using Microsoft excel for tabulation purpose. Nevertheless, there are evidence of unreported events or delay in reporting, which make the analysis very challenging. Based on the feedback from the system operator, the daily report is made at the end of their daily shift which sometimes is distorted from the original events.

7.5 Future Works

While the results are encouraging, note that this thesis is only a small step towards building an automatic industrial continuous intelligent system. The performance on system modelling can still be improved from three aspects: the variety of the industrial dataset, the method on how to retrieve the data and the selection of machine learning tools to predict the data.

7.5.1 Wide Variety of Industrial Dataset

In this study, a Software in the Loop application has been implemented to formulate and predict the occurrence of PQ disturbance problem in real-time. While the outcome of the study reveals that there was influence from the environment, which leads to a frequent harmonic filter failure, there are a wide variety of industrial data that can be explored. For example, implementing the Event Modeller Data Analytics (EMDA) in smart grid system will be beneficial to the power system engineer. As the smart grid system facilities are available in real-time, integrating the EMDA technique with the smart grid will not only monitor the system grid parameter, it could also observe the relationship to natural disaster events such as tsunami, earthquake, flood, and hurricane. As long as the relevant system parameter is included within the smart grid platform, the proposed technique will visualise and predict the occurrence of any of these events continuously.

7.5.2 Real-time Cloud Services

In this research, the CSU machine data were retrieved offline using a data logger. The reason is because of our industrial partner company policy, which does not allow the data to be made available live. While re-running the historical data as it runs in real-time offers good results, the real-time cloud services such as Google Cloud, Microsoft Azure and Amazon web services

could provide an excellent platform to analyse the data online. Private clouds services are also available for sensitive data. The main advantage of using this approach is that it offers a substantial reduction in computing time, and in most instances, this approach is cost-effective and scalable, especially for a multinational company owning various plant at different parts of the world. In that instance, all information on the running plant is shared in a dashboard which is accessible to the relevant party.

7.5.3 Other Machine Learning Algorithm

For proof of concept, the works in Chapter 6 embraced Neural Network, Naïve Bayes and Decision Tree algorithm for predictive modelling purposes. While the results on the performance are encouraging, they are by no means sufficient. There are various machine learning algorithms to be investigated in the future such as Linear Regression, Logistic Regression, k-nearest neighbours, Random Forest and Support Vector Machines, which promises better prediction accuracy.

Bibliography

- [1] E. Suhir, “Analytical bathtub curve with application to electron device reliability,” *Journal of Materials Science: Materials in Electronics*, vol. 26, no. 9, pp. 6633–6638, 2015.
- [2] K. R. Khalilpour, “Interconnected electricity and natural gas supply chains: The roles of power to gas and gas to power,” in *Polygeneration with Polystorage for Chemical and Energy Hubs*. Elsevier, 2019, pp. 133–155.
- [3] M. Al-Atabi, M. M. Shamel, E. Chung, T. Padmesh, and A. Al-Obaidi, “The use of industrial visits to enhance learning at engineering courses,” *Journal of Engineering Science and Technology*, pp. 1–7, 2013.
- [4] M. Panteli, D. N. Trakas, P. Mancarella, and N. D. Hatziargyriou, “Power systems resilience assessment: hardening and smart operational enhancement strategies,” *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1202–1213, 2017.
- [5] P. O. Glauner, A. Boechat, L. Dolberg, R. State, F. Bettinger, Y. Rangoni, and D. Duarte, “Large-scale detection of non-technical losses in imbalanced data sets,” *arXiv preprint arXiv:1602.08350*, 2016.
- [6] N. K. Sharma, P. K. Tiwari, and Y. R. Sood, “Solar energy in india: Strategies, policies, perspectives and future potential,” *Renewable and Sustainable Energy Reviews*, vol. 16, no. 1, pp. 933–941, 2012.
- [7] I. Statistics, “Co2 emissions from fuel combustion,” *International energy agency*, 2012.
- [8] A. Lind, E. Rosenberg, P. Seljom, K. Espegren, A. Fidje, and K. Lindberg, “Analysis of the eu renewable energy directive by a techno-economic optimisation model,” *Energy Policy*, vol. 60, pp. 364–377, 2013.

- [9] E. Department for Business and R. Reform., “Digest of united kingdom energy statistics, 2007,” 2017.
- [10] S. Tenaga, “Energy statistics handbook 2017,” *Putrajaya, Malaysia*, 2017.
- [11] M. Hannan, R. Begum, M. Abdolrasol, M. H. Lipu, A. Mohamed, and M. Rashid, “Review of baseline studies on energy policies and indicators in malaysia for future sustainable energy development,” *Renewable and Sustainable Energy Reviews*, vol. 94, pp. 551–564, 2018.
- [12] R. Kardooni, S. B. Yusoff, and F. B. Kari, “Renewable energy technology acceptance in peninsular malaysia,” *Energy Policy*, vol. 88, pp. 1–10, 2016.
- [13] M. Panteli and P. Mancarella, “Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies,” *Electric Power Systems Research*, vol. 127, pp. 259–270, 2015.
- [14] I.-Y. Chung, D.-J. Won, J.-M. Kim, S.-J. Ahn, and S.-I. Moon, “Development of a network-based power quality diagnosis system,” *Electric Power Systems Research*, vol. 77, no. 8, pp. 1086–1094, 2007.
- [15] M. S. Manikandan, S. Samantaray, and I. Kamwa, “Detection and classification of power quality disturbances using sparse signal decomposition on hybrid dictionaries,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 1, pp. 27–38, 2015.
- [16] M. H. Bollen and M. H. Bollen, *Understanding power quality problems: voltage sags and interruptions*. IEEE press New York, 2000, vol. 445.
- [17] J. Lee, H. Davari, J. Singh, and V. Pandhare, “Industrial artificial intelligence for industry 4.0-based manufacturing systems,” *Manufacturing letters*, vol. 18, pp. 20–23, 2018.
- [18] S. Khokhar, A. A. B. M. Zin, A. S. B. Mokhtar, and M. Pesaran, “A comprehensive overview on signal processing and artificial intelligence techniques applications in classification of power quality disturbances,” *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 1650–1663, 2015.

- [19] A. B. Wayne Thompson, Hui Li, “Artificial intelligence, machine learning, deep learning and beyond,” https://www.sas.com/en_us/insights/articles/big-data/artificial-intelligence-machine-learning-deep-learning-and-beyond.html#/, 2020, online; accessed 15 September 2020.
- [20] M. K. Saini and R. Kapoor, “Classification of power quality events—a review,” *International Journal of Electrical Power & Energy Systems*, vol. 43, no. 1, pp. 11–19, 2012.
- [21] M. Danishvar, A. Mousavi, and P. Sousa, “Eventclustering for improved real time input variable selection and data modelling,” in *Control Applications (CCA), 2014 IEEE Conference on*. IEEE, 2014, pp. 1801–1806.
- [22] J. Ypsilantis and H. Yee, “Machine learning of rules for a power system alarm processor,” in *1991 International Conference on Advances in Power System Control, Operation and Management, APSCOM-91*. IET, 1991, pp. 321–325.
- [23] L. Wehenkel, “Machine learning approaches to power-system security assessment,” *IEEE Expert*, vol. 12, no. 5, pp. 60–72, 1997.
- [24] T. Niimura, H. Ko, H. Xu, A. Moshref, and K. Morison, “Machine learning approach to power system dynamic security analysis,” in *IEEE PES Power Systems Conference and Exposition, 2004*. IEEE, 2004, pp. 1084–1088.
- [25] W. Pan, Y. Yuan, H. Sandberg, J. Gonçalves, and G.-B. Stan, “Online fault diagnosis for nonlinear power systems,” *Automatica*, vol. 55, pp. 27–36, 2015.
- [26] L. Zhang, L. Harnefors, and N. Johansson, “Power-system modeling-a transfer matrix approach,” *IET Conference Proceedings*, 2015.
- [27] S. Souag, F. Benhamida, Y. Salhi, A. Bendaoud, and F. Z. Gherbi, “Sensitivity factor for power system security analysis using labview,” in *2013 International Renewable and Sustainable Energy Conference (IRSEC)*. IEEE, 2013, pp. 385–390.
- [28] M. Kezunovic and Y. Liao, “A novel software implementation concept for power quality study,” *IEEE Transactions on power delivery*, vol. 17, no. 2, pp. 544–549, 2002.

- [29] A. C. Parsons, W. M. Grady, E. J. Powers, and J. C. Soward, "A direction finder for power quality disturbances based upon disturbance power and energy," in *8th International Conference on Harmonics and Quality of Power. Proceedings (Cat. No. 98EX227)*, vol. 2. IEEE, 1998, pp. 693–699.
- [30] Transmission, D. Committee *et al.*, "Ieee std c62. 41.1-2002 ieee guide on the surge environment in low-voltage (1 000 v or less) ac power circuits," 2002.
- [31] I. S. Association *et al.*, "Ieee recommended practice for monitoring electric power quality," *IEEE Std 1159-2019 (Revision of IEEE Std 1159-2009)*, pp. 1–98, 2019.
- [32] ———, "Ieee guide for identifying and improving voltage quality in power systems," *IEEE Std 1250-2018 (Revision of IEEE Std 1250-2011)*, pp. 1–0, 2018.
- [33] A. Standard, "C84. 1-2016," *American National Standards Institute: Rosslyn, VA, USA*, 2016.
- [34] I. S. Association *et al.*, "Ieee std 519-2014. recommended practice and requirements for harmonic control in electric power systems," *IEEE Power and Energy Society*, vol. 29, 2014.
- [35] I. Standard, "Iec 61000-3-3," *International Electrotechnical Commission*, 2017.
- [36] C. Li, W. Xu, B. Hughes, J. Gurney, and B. Neilson, "Virtual pq troubleshooter [power quality]," *IEEE Power and Energy Magazine*, vol. 1, no. 3, pp. 24–31, 2003.
- [37] S. Khokhar, A. M. Zin, A. Mokhtar, N. M. Ismail, and N. Zareen, "Automatic classification of power quality disturbances: a review," in *Research and Development (SCORED), 2013 IEEE Student Conference on*. IEEE, 2013, pp. 427–432.
- [38] Z. Moravej, S. Banihashemi, and M. Velayati, "Power quality events classification and recognition using a novel support vector algorithm," *Energy Conversion and Management*, vol. 50, no. 12, pp. 3071–3077, 2009.
- [39] J. Han, W.-K. Kim, J.-W. Lee, and C.-H. Kim, "Fault type classification in transmission line using stft," in *11th IET International Conference on Developments in Power Systems Protection (DPSP 2012)*. IET, 2012, pp. 1–5.

- [40] M. J. B. Reddy, R. K. Raghupathy, K. Venkatesh, and D. Mohanta, "Power quality analysis using discrete orthogonal s-transform (dost)," *Digital Signal Processing*, vol. 23, no. 2, pp. 616–626, 2013.
- [41] U. Singh and S. N. Singh, "Application of fractional fourier transform for classification of power quality disturbances," *IET Science, Measurement & Technology*, vol. 11, no. 1, pp. 67–76, 2017.
- [42] A. Augustine, R. Prakash, R. Xavier, and M. Parassery, "Review of signal processing techniques for detection of power quality events," *American Journal of Engineering and Applied Sciences*, vol. 9, no. 2, pp. 364–370, 2016.
- [43] MATLAB, *Global Optimization Toolbox: User's Guide (r2016b)*. Natick, Massachusetts: The MathWorks Inc., 2016.
- [44] S. Nath, A. Dey, and A. Chakrabarti, "Detection of power quality disturbances using wavelet transform," *World Academy of Science, Engineering and Technology*, vol. 49, pp. 869–873, 2009.
- [45] S. Santoso, E. J. Powers, W. M. Grady, and P. Hofmann, "Power quality assessment via wavelet transform analysis," *IEEE transactions on Power Delivery*, vol. 11, no. 2, pp. 924–930, 1996.
- [46] Y. Liao and J.-B. Lee, "A fuzzy-expert system for classifying power quality disturbances," *International journal of electrical power & energy systems*, vol. 26, no. 3, pp. 199–205, 2004.
- [47] L. Angrisani, P. Daponte, and M. D'Apuzzo, "Wavelet network-based detection and classification of transients," *IEEE transactions on instrumentation and measurement*, vol. 50, no. 5, pp. 1425–1435, 2001.
- [48] C.-H. Lin and C.-H. Wang, "Adaptive wavelet networks for power-quality detection and discrimination in a power system," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1106–1113, 2006.
- [49] S. Kaewarsa, K. Attakitmongcol, and T. Kulworawanichpong, "Recognition of power

- quality events by using multiwavelet-based neural networks,” *International Journal of Electrical Power & Energy Systems*, vol. 30, no. 4, pp. 254–260, 2008.
- [50] A. Gaouda, S. Kanoun, and M. Salama, “On-line disturbance classification using nearest neighbor rule,” *Electric power systems research*, vol. 57, no. 1, pp. 1–8, 2001.
- [51] J. Chung, E. J. Powers, W. M. Grady, and S. C. Bhatt, “Power disturbance classifier using a rule-based method and wavelet packet-based hidden markov model,” *IEEE transactions on power delivery*, vol. 17, no. 1, pp. 233–241, 2002.
- [52] M. Zhang, K. Li, and Y. Hu, “Classification of power quality disturbances using wavelet packet energy and multiclass support vector machine,” *COMPEL-The international journal for computation and mathematics in electrical and electronic engineering*, vol. 31, no. 2, pp. 424–442, 2012.
- [53] W. R. A. Ibrahim and M. M. Morcos, “Artificial intelligence and advanced mathematical tools for power quality applications: a survey,” *IEEE Transactions on Power Delivery*, vol. 17, no. 2, pp. 668–673, Apr 2002.
- [54] S. Meher, A. Pradhan, and G. Panda, “An integrated data compression scheme for power quality events using spline wavelet and neural network,” *Electric power systems research*, vol. 69, no. 2, pp. 213–220, 2004.
- [55] S. Khokhara, A. M. Zina, M. Bhayoa, A. Mokhtara, and N. Pakistan, “Automated recognition of single & hybrid power quality disturbances using wavelet transform based support vector machine,” *JURNAL TEKNOLOGI*, vol. 79, no. 1, pp. 97–105, 2017.
- [56] M. D. Borrás, J. C. Bravo, and J. C. Montaña, “Disturbance ratio for optimal multi-event classification in power distribution networks,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3117–3124, 2016.
- [57] N. Nashad, M. Islam, S. Alam, R. Rahat, M. T. Begum, and M. Alam, “A simplistic mathematical approach for detection and classification of power quality events,” in *Electrical, Computer and Communication Engineering (ECCE), International Conference on*. IEEE, 2017, pp. 698–703.

- [58] D. Guillén, G. Idárraga-Ospina, and C. Cortes, “A new adaptive mother wavelet for electromagnetic transient analysis,” *Journal of Electrical Engineering*, vol. 67, no. 1, pp. 48–55, 2016.
- [59] M. Sahani, S. Mishra, A. Ipsita, and B. Upadhyay, “Detection and classification of power quality event using wavelet transform and weighted extreme learning machine,” in *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on*. IEEE, 2016, pp. 1–6.
- [60] S. P. Puthenpurakel and P. Subadhra, “Identification and classification of microgrid disturbances in a hybrid distributed generation system using wavelet transform,” in *Next Generation Intelligent Systems (ICNGIS), International Conference on*. IEEE, 2016, pp. 1–5.
- [61] R. G. Stockwell, “A basis for efficient representation of the s-transform,” *Digital Signal Processing*, vol. 17, no. 1, pp. 371–393, 2007.
- [62] Y. Wang and J. Orchard, “Fast discrete orthonormal stockwell transform,” *SIAM Journal on Scientific Computing*, vol. 31, no. 5, pp. 4000–4012, 2009.
- [63] Y. Yan and H. Zhu, “The generalization of discrete stockwell transforms,” in *Signal Processing Conference, 2011 19th European*. IEEE, 2011, pp. 1209–1213.
- [64] M. Biswal and P. K. Dash, “Detection and characterization of multiple power quality disturbances with a fast s-transform and decision tree based classifier,” *Digital Signal Processing*, vol. 23, no. 4, pp. 1071–1083, 2013.
- [65] S. He, K. Li, and M. Zhang, “A real-time power quality disturbances classification using hybrid method based on s-transform and dynamics,” *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 9, pp. 2465–2475, 2013.
- [66] M. Chilukuri and P. Dash, “Multiresolution s-transform-based fuzzy recognition system for power quality events,” *IEEE Transactions on power delivery*, vol. 19, no. 1, pp. 323–330, 2004.
- [67] T. Nguyen and Y. Liao, “Power quality disturbance classification utilizing s-transform

- and binary feature matrix method,” *Electric Power Systems Research*, vol. 79, no. 4, pp. 569–575, 2009.
- [68] S. Mishra, C. Bhende, and B. Panigrahi, “Detection and classification of power quality disturbances using s-transform and probabilistic neural network,” *IEEE Transactions on power delivery*, vol. 23, no. 1, pp. 280–287, 2008.
- [69] H. S. Behera, P. K. Dash, and B. Biswal, “Power quality time series data mining using s-transform and fuzzy expert system,” *Applied Soft Computing*, vol. 10, no. 3, pp. 945–955, 2010.
- [70] A. Rodríguez, J. Aguado, F. Martín, J. López, F. Muñoz, and J. Ruiz, “Rule-based classification of power quality disturbances using s-transform,” *Electric power systems Research*, vol. 86, pp. 113–121, 2012.
- [71] P. R. Babu, P. Dash, S. Swain, and S. Sivanagaraju, “A new fast discrete s-transform and decision tree for the classification and monitoring of power quality disturbance waveforms,” *International Transactions on Electrical Energy Systems*, vol. 24, no. 9, pp. 1279–1300, 2014.
- [72] M. Biswal and P. K. Dash, “Measurement and classification of simultaneous power signal patterns with an s-transform variant and fuzzy decision tree,” *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 1819–1827, 2013.
- [73] R. Kumar, B. Singh, D. Shahani, A. Chandra, and K. Al-Haddad, “Recognition of power-quality disturbances using s-transform-based ann classifier and rule-based decision tree,” *IEEE Transactions on Industry Applications*, vol. 51, no. 2, pp. 1249–1258, 2015.
- [74] O. P. Mahela and A. G. Shaik, “Recognition of power quality disturbances using s-transform and fuzzy c-means clustering,” in *Cogeneration, Small Power Plants and District Energy (ICUE), International Conference on*. IEEE, 2016, pp. 1–6.
- [75] M. V. Reddy and R. Sodhi, “A rule-based s-transform and adaboost based approach for power quality assessment,” *Electric Power Systems Research*, vol. 134, pp. 66–79, 2016.
- [76] A. A. Abdoos, Z. Moravej, and M. Pazoki, “A hybrid method based on time frequency

- analysis and artificial intelligence for classification of power quality events,” *Journal of Intelligent & Fuzzy Systems*, vol. 28, no. 3, pp. 1183–1193, 2015.
- [77] S. Zhang, P. Li, L. Zhang, H. Li, W. Jiang, and Y. Hu, “Modified s transform and elm algorithms and their applications in power quality analysis,” *neurocomputing*, vol. 185, pp. 231–241, 2016.
- [78] Z. Lu, J. Smith, Q. Wu, and J. Fitch, “Empirical mode decomposition for power quality monitoring,” in *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES*. IEEE, 2005, pp. 1–5.
- [79] T. Jayasree, D. Devaraj, and R. Sukanesh, “Power quality disturbance classification using hilbert transform and rbf networks,” *Neurocomputing*, vol. 73, no. 7, pp. 1451–1456, 2010.
- [80] O. P. Mahela, A. G. Shaik, and N. Gupta, “A critical review of detection and classification of power quality events,” *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 495–505, 2015.
- [81] S. Shukla, S. Mishra, and B. Singh, “Empirical-mode decomposition with hilbert transform for power-quality assessment,” *IEEE transactions on power delivery*, vol. 24, no. 4, pp. 2159–2165, 2009.
- [82] F. Hafiz, A. H. Chowdhury, and C. Shahnaz, “An approach for classification of power quality disturbances based on hilbert huang transform and relevance vector machine,” in *Electrical & Computer Engineering (ICECE), 2012 7th International Conference on*. IEEE, 2012, pp. 201–204.
- [83] O. Ozgonenel, T. Yalcin, I. Guney, and U. Kurt, “A new classification for power quality events in distribution systems,” *Electric Power Systems Research*, vol. 95, pp. 192–199, 2013.
- [84] S. Dalai, D. Dey, B. Chatterjee, S. Chakravorti, and K. Bhattacharya, “Cross hilbert-huang transform based feature extraction method for multiple pq disturbance classification,” in *Condition Assessment Techniques in Electrical Systems (CATCON), 2013 IEEE 1st International Conference on*. IEEE, 2013, pp. 314–317.

- [85] R. Shilpa, S. S. Prabhu, and P. Puttaswamy, "Power quality disturbances monitoring using hilbert-huang transform and svm classifier," in *Emerging Research in Electronics, Computer Science and Technology (ICERECT), 2015 International Conference on*. IEEE, 2015, pp. 6–10.
- [86] Y. Huang, Y. Liu, and Z. Hong, "Detection and location of power quality disturbances based on mathematical morphology and hilbert-huang transform," in *Electronic Measurement & Instruments, 2009. ICEMI'09. 9th International Conference on*. IEEE, 2009, pp. 2–319.
- [87] W. Zhan, Z. Xiangjun, H. Xiaoxi, and H. Jingying, "The multi-disturbance complex power quality signal hht detection technique," in *Innovative Smart Grid Technologies-Asia (ISGT Asia), 2012 IEEE*. IEEE, 2012, pp. 1–5.
- [88] Y. Önal and Ü. Ç. Turhal, "The orthogonal hilbert-huang transform application in voltage flicker analysis," in *Power Engineering, Energy and Electrical Drives (POWERENG), 2013 Fourth International Conference on*. IEEE, 2013, pp. 700–704.
- [89] A. Foroughi, E. Mohammadi, and S. Esmaeili, "Application of hilbert–huang transform and support vector machine for detection and classification of voltage sag sources," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 22, no. 5, pp. 1116–1129, 2014.
- [90] T. Yalcin and M. Ozdemir, "Pattern recognition method for identifying smart grid power quality disturbance," in *Harmonics and Quality of Power (ICHQP), 2016 17th International Conference on*. IEEE, 2016, pp. 903–907.
- [91] S. Qian and D. Chen, "Discrete gabor transform," *IEEE transactions on signal processing*, vol. 41, no. 7, pp. 2429–2438, 1993.
- [92] D. Granados-Lieberman, R. Romero-Troncoso, R. Osornio-Rios, A. Garcia-Perez, and E. Cabal-Yepez, "Techniques and methodologies for power quality analysis and disturbances classification in power systems: a review," *IET Generation, Transmission & Distribution*, vol. 5, no. 4, pp. 519–529, 2011.

- [93] S.-H. Cho, G. Jang, and S.-H. Kwon, "Time-frequency analysis of power-quality disturbances via the gabor-wigner transform," *IEEE transactions on power delivery*, vol. 25, no. 1, pp. 494–499, 2010.
- [94] Z. Moravej, M. Pazoki, M. Niasati, and A. A. Abdoos, "A hybrid intelligence approach for power quality disturbances detection and classification," *International Transactions on Electrical Energy Systems*, vol. 23, no. 7, pp. 914–929, 2013.
- [95] T. A. Kawady, N. I. Elkalashy, A. E. Ibrahim, and A.-M. I. Taalab, "Arcing fault identification using combined gabor transform-neural network for transmission lines," *International Journal of Electrical Power & Energy Systems*, vol. 61, pp. 248–258, 2014.
- [96] S. Naderian and A. Salemnia, "Detection and classification of power-quality events using discrete gabor transform and support vector machine," in *Power Electronics, Drives Systems & Technologies Conference (PEDSTC), 2015 6th*. IEEE, 2015, pp. 544–549.
- [97] —, "An implementation of type-2 fuzzy kernel based support vector machine algorithm for power quality events classification," *International Transactions on Electrical Energy Systems*, vol. 27, no. 5, 2017.
- [98] —, "Method for classification of pq events based on discrete gabor transform with fir window and t2fk-based svm and its experimental verification," *IET Generation, Transmission & Distribution*, vol. 11, no. 1, pp. 133–141, 2017.
- [99] P. Dash and M. Chilukuri, "Hybrid s-transform and kalman filtering approach for detection and measurement of short duration disturbances in power networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 2, pp. 588–596, 2004.
- [100] A. A. Abdelsalam, A. A. Eldesouky, and A. A. Sallam, "Classification of power system disturbances using linear kalman filter and fuzzy-expert system," *International Journal of Electrical Power & Energy Systems*, vol. 43, no. 1, pp. 688–695, 2012.
- [101] E. Perez and J. Barros, "An extended kalman filtering approach for detection and analysis of voltage dips in power systems," *Electric Power Systems Research*, vol. 78, no. 4, pp. 618–625, 2008.

- [102] J. Reddy, P. K. Dash, R. Samantaray, and A. K. Moharana, "Fast tracking of power quality disturbance signals using an optimized unscented filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 12, pp. 3943–3952, 2009.
- [103] M. Wang and A. V. Mamishev, "Classification of power quality events using optimal time-frequency representations-part 1: theory," *IEEE Transactions on Power Delivery*, vol. 19, no. 3, pp. 1488–1495, 2004.
- [104] M. Wang, G. I. Rowe, and A. V. Mamishev, "Classification of power quality events using optimal time-frequency representations-part 2: application," *IEEE Transactions on Power Delivery*, vol. 19, no. 3, pp. 1496–1503, 2004.
- [105] S. Suja and J. Jerome, "Pattern recognition of power signal disturbances using s transform and tt transform," *International journal of electrical power & energy systems*, vol. 32, no. 1, pp. 37–53, 2010.
- [106] S. Jashfar, S. ESMAEILI, M. Z. JAHROMI, and M. Rahmanian, "Classification of power quality disturbances using s-transform and tt-transform based on the artificial neural network," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 21, no. 6, pp. 1528–1538, 2013.
- [107] G.-S. Hu, F.-F. Zhu, and Y.-J. Tu, "Power quality disturbance detection and classification using chirplet transforms," *Simulated Evolution and Learning*, pp. 34–41, 2006.
- [108] S. Ouyang and J. Wang, "A new morphology method for enhancing power quality monitoring system," *International journal of Electrical power & Energy systems*, vol. 29, no. 2, pp. 121–128, 2007.
- [109] C.-T. Hsieh, J.-M. Lin, and S.-J. Huang, "Slant transform applied to electric power quality detection with field programmable gate array design enhanced," *International Journal of Electrical Power & Energy Systems*, vol. 32, no. 5, pp. 428–432, 2010.
- [110] A. Subasi, A. S. Yilmaz, and K. Tufan, "Detection of generated and measured transient power quality events using teager energy operator," *Energy Conversion and Management*, vol. 52, no. 4, pp. 1959–1967, 2011.

- [111] J. J. G. de la Rosa, J. M. Sierra-Fernández, A. Agüera-Pérez, J. C. Palomares-Salas, and A. Moreno-Muñoz, “An application of the spectral kurtosis to characterize power quality events,” *International Journal of Electrical Power & Energy Systems*, vol. 49, pp. 386–398, 2013.
- [112] D. D. Ferreira, J. M. de Seixas, A. S. Cerqueira, and C. A. Duque, “Exploiting principal curves for power quality monitoring,” *Electric Power Systems Research*, vol. 100, pp. 1–6, 2013.
- [113] K. Manimala, K. Selvi, and R. Ahila, “Optimization techniques for improving power quality data mining using wavelet packet based support vector machine,” *Neurocomputing*, vol. 77, no. 1, pp. 36–47, 2012.
- [114] A. Wang, W. Yuan, J. Liu, Q. Wang, and Z. Yu, “A study of a multi-class classification algorithm of svm combined with art,” in *Natural Computation, 2007. ICNC 2007. Third International Conference on*, vol. 1. IEEE, 2007, pp. 59–63.
- [115] A. Y. Jaen-Cuellar, L. Morales-Velazquez, R. de Jesus Romero-Troncoso, D. Moriñigo-Sotelo, and R. A. Osornio-Rios, “Micro-genetic algorithms for detecting and classifying electric power disturbances,” *Neural Computing and Applications*, vol. 28, no. 1, pp. 379–392, 2017.
- [116] K.-E. Ko, S.-M. Park, J.-H. Park, and K.-B. Sim, “Training hmm structure and parameters with genetic algorithm and harmony search algorithm,” *Journal of Electrical Engineering and Technology*, vol. 7, no. 1, pp. 109–114, 2012.
- [117] R. Ahila, V. Sadasivam, and K. Manimala, “An integrated pso for parameter determination and feature selection of elm and its application in classification of power system disturbances,” *Applied Soft Computing*, vol. 32, pp. 23–37, 2015.
- [118] R. Hooshmand and A. Enshae, “Detection and classification of single and combined power quality disturbances using fuzzy systems oriented by particle swarm optimization algorithm,” *Electric Power Systems Research*, vol. 80, no. 12, pp. 1552–1561, 2010.
- [119] P. Jiang and X. Ma, “A hybrid forecasting approach applied in the electrical power

- system based on data preprocessing, optimization and artificial intelligence algorithms,” *Applied Mathematical Modelling*, vol. 40, no. 23, pp. 10 631–10 649, 2016.
- [120] B. Biswal, P. K. Dash, and B. K. Panigrahi, “Power quality disturbance classification using fuzzy c-means algorithm and adaptive particle swarm optimization,” *IEEE Transactions on Industrial Electronics*, vol. 56, no. 1, pp. 212–220, 2009.
- [121] B. Biswal, P. K. Dash, and S. Mishra, “A hybrid ant colony optimization technique for power signal pattern classification,” *Expert Systems with Applications*, vol. 38, no. 5, pp. 6368–6375, 2011.
- [122] S. Tavakoli, A. Mousavi, and S. Poslad, “Input variable selection in time-critical knowledge integration applications: A review, analysis, and recommendation paper,” *Advanced Engineering Informatics*, vol. 27, no. 4, pp. 519–536, 2013.
- [123] R. May, G. Dandy, and H. Maier, “Review of input variable selection methods for artificial neural networks,” in *Artificial neural networks-methodological advances and biomedical applications*. InTech, 2011.
- [124] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
- [125] M. Danishvar, A. Mousavi, P. Sousa, and R. Araujo, “Event-clustering for real-time data modeling,” in *Automation Science and Engineering (CASE), 2013 IEEE International Conference on*. IEEE, 2013, pp. 362–367.
- [126] S. Tavakoli, A. Mousavi, and P. Broomhead, “Event tracking for real-time unaware sensitivity analysis (eventtracker),” *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 2, pp. 348–359, 2013.
- [127] J. R. King, “Machine-component grouping in production flow analysis: an approach using a rank order clustering algorithm,” *International Journal of Production Research*, vol. 18, no. 2, pp. 213–232, 1980.
- [128] T. Ghosh and P. K. Dan, “Effective clustering method for group technology problems: a short communication,” *Journal of Science & Technology*, vol. 6, pp. 23–28, 2011.

- [129] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [130] F. Rosenblatt, “Principles of neurodynamics. perceptrons and the theory of brain mechanisms,” CORNELL AERONAUTICAL LAB INC BUFFALO NY, Tech. Rep., 1961.
- [131] M. Minsky and S. Papert, “Perceptrons cambridge,” *MA: MIT Press. zbMATH*, 1969.
- [132] P. J. Werbos, “Generalization of backpropagation with application to a recurrent gas market model,” *Neural networks*, vol. 1, no. 4, pp. 339–356, 1988.
- [133] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” in *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*. World Scientific, 1987, pp. 411–415.
- [134] D. Williams and G. Hinton, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–538, 1986.
- [135] M. Chester, *Neural networks: a tutorial*. Prentice-Hall, Inc., 1993.
- [136] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [137] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [138] Y. LeCun *et al.*, “Lenet-5, convolutional neural networks,” *URL: <http://yann.lecun.com/exdb/lenet>*, 2015.
- [139] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [140] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural networks*, vol. 61, pp. 85–117, 2015.
- [141] D. Devaraj, P. Radhika, V. Subasri, and R. Kanagavalli, “Power quality monitoring using wavelet transform and artificial neural networks,” in *Power Electronics, 2006. IICPE 2006. India International Conference on*. IEEE, 2006, pp. 425–430.

- [142] L. Saikia, S. Borah, and S. Pait, "Detection and classification of power quality disturbances using wavelet transform, fuzzy logic and neural network," in *India Conference (INDICON), 2010 Annual IEEE*. IEEE, 2010, pp. 1–5.
- [143] S. Alshahrani, M. Abbod, B. Alamri, and G. Taylor, "Evaluation and classification of power quality disturbances based on discrete wavelet transform and artificial neural networks," in *Power Engineering Conference (UPEC), 2015 50th International Universities*. IEEE, 2015, pp. 1–5.
- [144] Z. Yun, Z. Quan, S. Caixin, L. Shaolan, L. Yuming, and S. Yang, "Rbf neural network and anfis-based short-term load forecasting approach in real-time price environment," *IEEE Transactions on power systems*, vol. 23, no. 3, pp. 853–858, 2008.
- [145] E. Almaita and J. A. Asumadu, "On-line harmonic estimation in power system based on sequential training radial basis function neural network," in *Industrial Technology (ICIT), 2011 IEEE International Conference on*. IEEE, 2011, pp. 139–144.
- [146] A. M. Haidar, A. Mohamed, and A. Hussain, "Vulnerability assessment of a large sized power system using radial basis function neural network," in *Research and Development, 2007. SCORed 2007. 5th Student Conference on*. IEEE, 2007, pp. 1–6.
- [147] G. W. Chang, C.-I. Chen, and Y.-F. Teng, "Radial-basis-function-based neural network for harmonic detection," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 6, pp. 2171–2179, 2010.
- [148] T.-Y. Huang, C. J. Li, and T.-W. Hsu, "Structure and parameter learning algorithm of jordan type recurrent neural networks," in *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*. IEEE, 2007, pp. 1819–1824.
- [149] A. Thammano and P. Ruxpakawong, "Dynamic system identification using recurrent neural network with multi-valued connection weight," in *Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on*. IEEE, 2009, pp. 2077–2082.
- [150] X. Gao, X. Gao, and S. Ovaska, "A modified elman neural network model with application to dynamical systems identification," in *Systems, Man, and Cybernetics, 1996., IEEE International Conference on*, vol. 2. IEEE, 1996, pp. 1376–1381.

- [151] Z. J. Paracha, A. M. Mehdi, and A. Kalam, "Computational analysis of sag and swell in electrical power distribution network," in *Power Engineering Conference, 2009. AUPEC 2009. Australasian Universities*. IEEE, 2009, pp. 1–5.
- [152] J. Thongam, P. Bouchard, R. Beguenane, and I. Fofana, "Neural network based wind speed sensorless mppt controller for variable speed wind energy conversion systems," in *Electric Power and Energy Conference (EPEC), 2010 IEEE*. IEEE, 2010, pp. 1–6.
- [153] W. Zhao, L. Shang, and J. Sun, "Power quality disturbance classification based on time-frequency domain multi-feature and decision tree," *Protection and Control of Modern Power Systems*, vol. 4, no. 1, p. 27, 2019.
- [154] A. Milchevski and D. Taskovski, "Classification of power quality disturbances using wavelet transform and svm decision tree," in *11th International Conference on Electrical Power Quality and Utilisation*. IEEE, 2011, pp. 1–5.
- [155] R. K. Patnaik, K. Anjaiah, and R. K. Pattanaik, "A simple decision tree based event classification technique for multiple power quality disturbance signals," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, pp. 1315–1321, 2019.
- [156] M. R. Khan, S. Padhi, B. Sahu, and S. Behera, "Non stationary signal analysis and classification using fft transform and naive bayes classifier," in *2015 IEEE Power, Communication and Information Technology Conference (PCITC)*. IEEE, 2015, pp. 967–972.
- [157] F. Hafiz, A. Swain, C. Naik, and N. Patel, "Efficient feature selection of power quality events using two dimensional (2d) particle swarms," *Applied Soft Computing*, vol. 81, p. 105498, 2019.
- [158] M. H. Jopri, M. R. Ab Ghani, A. R. Abdullah, M. Manap, T. Sutikno, and J. Too, "K-nearest neighbor and naïve bayes based diagnostic analytic of harmonic source identification," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 6, pp. 2650–2657, 2020.
- [159] O. E. Adegbite and M. Okelola, "Training of the naïve bayes classifier for the detection of

- the power quality events (voltage dip, voltage swell and voltage interruption),” *European Journal of Electrical Engineering and Computer Science*, vol. 4, no. 4, 2020.
- [160] T. Kohonen, “The self-organizing map,” *Neurocomputing*, vol. 21, no. 1, pp. 1–6, 1998.
- [161] T. Babnik, R. Aggarwal, and P. Moore, “Data mining on a transformer partial discharge data using the self-organizing map,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 14, no. 2, 2007.
- [162] S. Malek, A. Salleh, and S. M. S. Ahmad, “Analysis of algal growth using kohonen self organizing feature map (som) and its prediction using rule based expert system,” in *Information Management and Engineering, 2009. ICIME’09. International Conference on*. IEEE, 2009, pp. 501–504.
- [163] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.
- [164] H. He and J. A. Starzyk, “A self-organizing learning array system for power quality classification based on wavelet transform,” *IEEE Transactions on Power Delivery*, vol. 21, no. 1, pp. 286–295, 2006.
- [165] P. Peilin and D. Guangbin, “Power quality detection and discrimination in distributed power system based on wavelet transform,” in *Control Conference, 2008. CCC 2008. 27th Chinese*. IEEE, 2008, pp. 635–638.
- [166] N. Huang, X. Liu, D. Xu, and J. Qi, “Power quality disturbance recognition based on s-transform and som neural network,” in *Image and Signal Processing, 2009. CISP’09. 2nd International Congress on*. IEEE, 2009, pp. 1–5.
- [167] E. Bentley, G. Putrus, S. McDonald, and P. Minns, “Power quality disturbance source identification using self-organising maps,” *IET generation, transmission & distribution*, vol. 4, no. 10, pp. 1188–1196, 2010.
- [168] Z. Zhemin, W. Tian, and L. Fenlan, “Fault diagnosis based on wavelet neural network,” in *Intelligent Computation Technology and Automation (ICICTA), 2012 Fifth International Conference on*. IEEE, 2012, pp. 482–485.
- [169] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, “A learning algorithm for boltzmann machines,” *Cognitive science*, vol. 9, no. 1, pp. 147–169, 1985.

- [170] J. Watada and H. R. M. Cisneros, "Meta-controlled boltzmann machine: Its convergence and application to power system," in *World Automation Congress (WAC), 2012*. IEEE, 2012, pp. 1–6.
- [171] H. Mori and S. Tsuzuki, "Determination of power system topological observability using the boltzmann machine," in *Circuits and Systems, 1990., IEEE International Symposium on*. IEEE, 1990, pp. 2938–2941.
- [172] U. M. Ali and K. Al Jebory, "Fieldbus scheduling using boltzmann machine," in *Systems Signals and Devices (SSD), 2010 7th International Multi-Conference on*. IEEE, 2010, pp. 1–10.
- [173] G. E. Hinton, "Learning multiple layers of representation," *Trends in cognitive sciences*, vol. 11, no. 10, pp. 428–434, 2007.
- [174] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer vision, graphics, and image processing*, vol. 37, no. 1, pp. 54–115, 1987.
- [175] Y. Pao, "Adaptive pattern recognition and neural networks," *U.S. Department of Energy Office of Scientific and Technical Information*, 1989.
- [176] S. Vasilic and M. Kezunovic, "Fuzzy art neural network algorithm for classifying the power system faults," *IEEE Transactions on power delivery*, vol. 20, no. 2, pp. 1306–1314, 2005.
- [177] X. Qingyang, M. Xianyao, and W. Ning, "Gas turbine fault diagnosis based on art2 neural network," in *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on*. IEEE, 2008, pp. 5244–5248.
- [178] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural networks*, vol. 4, no. 6, pp. 759–771, 1991.
- [179] I. Nikoofekr, M. Sarlak, and S. Shahrtash, "Detection and classification of high impedance faults in power distribution networks using art neural networks," in *Electrical Engineering (ICEE), 2013 21st Iranian Conference on*. IEEE, 2013, pp. 1–6.

- [180] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy layer-wise training of deep networks,” in *Advances in neural information processing systems*, 2007, pp. 153–160.
- [181] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [182] N. Anantrasirichai, I. D. Gilchrist, and D. R. Bull, “Visual salience and priority estimation for locomotion using a deep convolutional neural network,” in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1599–1603.
- [183] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” *Journal of Machine Learning Research*, vol. 12, no. Aug, pp. 2493–2537, 2011.
- [184] T. N. Sainath, A.-r. Mohamed, B. Kingsbury, and B. Ramabhadran, “Deep convolutional neural networks for lvcsr,” in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 8614–8618.
- [185] P. Binsha, S. Sachin Kumar, S. Athira, and K. Soman, “Power quality signal classification using convolutional neural network,” *International Journal of Control Theory and Applications*, vol. 9, no. 14, pp. 6405–6414, 2016.
- [186] Y. Tao, H. Chen, and C. Qiu, “Wind power prediction and pattern feature based on deep learning method,” in *Power and Energy Engineering Conference (APPEEC), 2014 IEEE PES Asia-Pacific*. IEEE, 2014, pp. 1–4.
- [187] K. Ueyoshi, T. Marukame, T. Asai, M. Motomura, and A. Schmid, “Memory-error tolerance of scalable and highly parallel architecture for restricted boltzmann machines in deep belief network,” in *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on*. Ieee, 2016, pp. 357–360.
- [188] H. J. Steinhauer, A. Karlsson, G. Mathiason, and T. Helldin, “Root-cause localization using restricted boltzmann machines,” in *Information Fusion (FUSION), 2016 19th International Conference on*. IEEE, 2016, pp. 248–255.

- [189] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [190] N. Lu, T. Li, X. Ren, and H. Miao, “A deep learning scheme for motor imagery classification based on restricted boltzmann machines,” *IEEE transactions on neural systems and rehabilitation engineering*, vol. 25, no. 6, pp. 566–576, 2017.
- [191] Q. Yang, H. Wang, T. Li, and Y. Yang, “Deep belief networks oriented clustering,” in *Intelligent Systems and Knowledge Engineering (ISKE), 2015 10th International Conference on*. IEEE, 2015, pp. 58–65.
- [192] Z. Chen, J. Wang, H. He, and X. Huang, “A fast deep learning system using gpu,” in *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 1552–1555.
- [193] C. Zhang, J. H. Sun, and K. C. Tan, “Deep belief networks ensemble with multi-objective optimization for failure diagnosis,” in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 32–37.
- [194] G. Han and K. Sohn, “Clustering the seoul metropolitan area by travel patterns based on a deep belief network,” in *Big Data and Smart City (ICBDSC), 2016 3rd MEC International Conference on*. IEEE, 2016, pp. 1–6.
- [195] L. A. Zadeh, “Fuzzy sets,” in *Fuzzy Sets, Fuzzy Logic, And Fuzzy Systems: Selected Papers by Lotfi A Zadeh*. World Scientific, 1996, pp. 394–432.
- [196] J. Lukasiewicz and A. Tarski, “Investigations into the sentential calculus,” *Logic, Semantics, Metamathematics*, pp. 38–59, 1956.
- [197] S. Hasheminejad, S. Esmaili, and S. Jazebi, “Power quality disturbance classification using s-transform and hidden markov model,” *Electric Power Components and Systems*, vol. 40, no. 10, pp. 1160–1182, 2012.
- [198] H. Dehghani, B. Vahidi, R. Naghizadeh, and S. H. Hosseinian, “Power quality disturbance classification using a statistical and wavelet-based hidden markov model with dempster–shafer algorithm,” *International Journal of Electrical Power & Energy Systems*, vol. 47, pp. 368–377, 2013.

- [199] M. Ahmadlou and H. Adeli, “Enhanced probabilistic neural network with local decision circles: A robust classifier,” *Integrated Computer-Aided Engineering*, vol. 17, no. 3, pp. 197–210, 2010.
- [200] N. Huang, D. Xu, X. Liu, and L. Lin, “Power quality disturbances classification based on s-transform and probabilistic neural network,” *Neurocomputing*, vol. 98, pp. 12–23, 2012.
- [201] M. Manjula, S. Mishra, and A. Sarma, “Empirical mode decomposition with hilbert transform for classification of voltage sag causes using probabilistic neural network,” *International Journal of Electrical Power & Energy Systems*, vol. 44, no. 1, pp. 597–603, 2013.
- [202] K. Manimala, I. G. David, and K. Selvi, “A novel data selection technique using fuzzy c-means clustering to enhance svm-based power quality classification,” *Soft Computing*, vol. 19, no. 11, pp. 3123–3144, 2015.
- [203] M. Ijaz, M. Shafullah, and M. Abido, “Classification of power quality disturbances using wavelet transform and optimized ann,” in *Intelligent System Application to Power Systems (ISAP), 2015 18th International Conference on*. IEEE, 2015, pp. 1–6.
- [204] R. Ak and R. Bhinge, “Data analytics and uncertainty quantification for energy prediction in manufacturing,” in *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, 2015, pp. 2782–2784.
- [205] J. Man, Z. Zhang, and Q. Zhou, “Data-driven predictive analytics of unexpected wind turbine shut-downs,” *IET Renewable Power Generation*, vol. 12, no. 15, pp. 1833–1842, 2018.
- [206] K. P. Hock, K. Radjabli, D. McGuinness, and M. Boddeti, “Predictive analysis in energy management system,” in *2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*. IEEE, 2016, pp. 1–4.
- [207] Predictive Analytics Today, “What is Predictive Analytics ?” <https://www.predictiveanalyticstoday.com/what-is-predictive-analytics/>, 2019, online; accessed 5 May 2019.

- [208] W. W. Eckerson, “Predictive analytics,” *Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report*, vol. 1, pp. 1–36, 2007.
- [209] P. Ongsulee, V. Chotchaung, E. Bamrunsi, and T. Rodcheewit, “Big data, predictive analytics and machine learning,” in *2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE)*. IEEE, 2018, pp. 1–6.
- [210] D. Bañeres and M. Serra, “Predictive analytics: another vision of the learning process,” in *Software Data Engineering for Network eLearning Environments*. Springer, 2018, pp. 1–25.
- [211] N. Mishra and S. Silakari, “Predictive analytics: A survey, trends, applications, opportunities & challenges,” *International Journal of Computer Science and Information Technologies*, vol. 3, no. 3, pp. 4434–4438, 2012.
- [212] P. O’donovan, K. Leahy, K. Bruton, and D. T. O’Sullivan, “Big data in manufacturing: a systematic mapping study,” *Journal of Big Data*, vol. 2, no. 1, p. 20, 2015.
- [213] T. Taneja, A. Jatain, and S. B. Bajaj, “Predictive analytics on iot,” in *2017 International Conference on Computing, Communication and Automation (ICCCA)*. IEEE, 2017, pp. 1312–1317.
- [214] Y. Zhang, G. Zhang, J. Wang, S. Sun, S. Si, and T. Yang, “Real-time information capturing and integration framework of the internet of manufacturing things,” *International Journal of Computer Integrated Manufacturing*, vol. 28, no. 8, pp. 811–822, 2015.
- [215] J. Man and Q. Zhou, “Remaining useful life prediction for hard failures using joint model with extended hazard,” *Quality and Reliability Engineering International*, vol. 34, no. 5, pp. 748–758, 2018.
- [216] J. Harding, M. Shahbaz, A. Kusiak *et al.*, “Data mining in manufacturing: a review,” *Journal of Manufacturing Science and Engineering*, vol. 128, no. 4, pp. 969–976, 2006.
- [217] D. Lechevalier, A. Narayanan, and S. Rachuri, “Towards a domain-specific framework for predictive analytics in manufacturing,” in *2014 IEEE International Conference on Big Data (Big Data)*. IEEE, 2014, pp. 987–995.

- [218] A. Y. Dorogov, “Technologies of predictive analytics for big data,” in *2015 XVIII International Conference on Soft Computing and Measurements (SCM)*. IEEE, 2015, pp. 182–183.
- [219] S. Bagga and A. Sharma, “Big data and its challenges: A review,” in *2018 4th International Conference on Computing Sciences (ICCS)*. IEEE, 2018, pp. 183–187.
- [220] M. Beyer, “Gartner says solving ‘big data’ challenge involves more than just managing volumes of data,” *Gartner. Archived from the original on*, vol. 10, 2011.
- [221] A. Abouzeid, K. Bajda-Pawlikowski, D. Abadi, A. Silberschatz, and A. Rasin, “Hadoopdb: an architectural hybrid of mapreduce and dbms technologies for analytical workloads,” *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 922–933, 2009.
- [222] J. B. Rothnie Jr, P. A. Bernstein, S. Fox, N. Goodman, M. Hammer, T. A. Landers, C. Reeve, D. W. Shipman, and E. Wong, “Introduction to a system for distributed databases (sdd-1),” *ACM Transactions on Database Systems (TODS)*, vol. 5, no. 1, pp. 1–17, 1980.
- [223] M. Prakash, G. Padmapriy, and M. V. Kumar, “A review on machine learning big data using r,” in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, 2018, pp. 1873–1877.
- [224] L. R. Sebastian, S. Babu, and J. J. Kizhakkethottam, “Challenges with big data mining: A review,” in *2015 International Conference on Soft-Computing and Networks Security (ICSNS)*. IEEE, 2015, pp. 1–4.
- [225] A. K. Manekar and G. Pradeepini, “Cloud based big data analytics a review,” in *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE, 2015, pp. 785–788.
- [226] T. Dillon, C. Wu, and E. Chang, “Cloud computing: issues and challenges,” in *2010 24th IEEE international conference on advanced information networking and applications*. Ieee, 2010, pp. 27–33.

- [227] H. Yang, M. Park, M. Cho, M. Song, and S. Kim, “A system architecture for manufacturing process analysis based on big data and process mining techniques,” in *2014 IEEE International Conference on Big Data (Big Data)*. IEEE, 2014, pp. 1024–1029.
- [228] A. Saltelli, S. Tarantola, and K.-S. Chan, “A quantitative model-independent method for global sensitivity analysis of model output,” *Technometrics*, vol. 41, no. 1, pp. 39–56, 1999.
- [229] H. Cloke, F. Pappenberger, and J.-P. Renaud, “Multi-method global sensitivity analysis (mmgsa) for modelling floodplain hydrological processes,” *Hydrological Processes: An International Journal*, vol. 22, no. 11, pp. 1660–1674, 2008.
- [230] E. Borgonovo, “A new uncertainty importance measure,” *Reliability Engineering & System Safety*, vol. 92, no. 6, pp. 771–784, 2007.
- [231] A. Saltelli, “Sensitivity analysis for importance assessment,” *Risk analysis*, vol. 22, no. 3, pp. 579–590, 2002.
- [232] E. Fock, “Global sensitivity analysis approach for input selection and system identification purposes—a new framework for feedforward neural networks,” *IEEE transactions on neural networks and learning systems*, vol. 25, no. 8, pp. 1484–1495, 2013.
- [233] M. Danishvar, A. Mousavi, and P. Broomhead, “Eventic: A real-time unbiased event-based learning technique for complex systems,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018.
- [234] B. Krzykacz-Hausmann, “An approximate sensitivity analysis of results from complex computer models in the presence of epistemic and aleatory uncertainties,” *Reliability Engineering & System Safety*, vol. 91, no. 10, pp. 1210–1218, 2006.
- [235] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, “Sensitivity analysis in practice: a guide to assessing scientific models,” *Chichester, England*, 2004.
- [236] S. Tavakoli, A. Mousavi, and A. Komashie, “A generic framework for real-time discrete event simulation (des) modelling,” in *Proceedings of the 40th Conference on Winter Simulation*. Winter Simulation Conference, 2008, pp. 1931–1938.

- [237] Z. Huang, M. Li, A. Mousavi, M. Danishva, and Z. Wang, “Egep: An event tracker enhanced gene expression programming for data driven system engineering problems,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 2, pp. 117–126, 2019.
- [238] M. R. Kumar and G. P. Rao, “Design and implementation of 32 bit high level wallace tree multiplier,” *Int. Journ. of Technical Research and Application*, vol. 1, pp. 86–90, 2013.
- [239] M. Danishvar, “Modelling and design of the eco-system of causality for real-time systems,” Ph.D. dissertation, Brunel University London, 2015.
- [240] M. P. Groover, *Automation, production systems, and computer-integrated manufacturing*. Pearson Education India, 2016.
- [241] A. Mousavi and H. Siervo, “Automatic translation of plant data into management performance metrics: a case for real-time and predictive production control,” *International Journal of Production Research*, vol. 55, no. 17, pp. 4862–4877, 2017.
- [242] S. X. Ding, S. Yin, K. Peng, H. Hao, and B. Shen, “A novel scheme for key performance indicator prediction and diagnosis with application to an industrial hot strip mill,” *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2239–2247, 2012.
- [243] K. Ragnathan, S. Ravindranathan, and S. Naveen, “Statistical kpis in hmi panels,” in *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*. IEEE, 2015, pp. 838–843.
- [244] N. Kang, C. Zhao, J. Li, and J. A. Horst, “Analysis of key operation performance data in manufacturing systems,” in *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, 2015, pp. 2767–2770.
- [245] A. De Ron and J. Rooda, “Equipment effectiveness: Oee revisited,” *IEEE transactions on semiconductor manufacturing*, vol. 18, no. 1, pp. 190–196, 2005.
- [246] W. Kelton, R. Sadowski, and N. Zupick, “Simulation with arena, 6: e uppl,” 2015.

- [247] G. UK, “Greenhouse gas reporting: Conversion factors 2017,” *GOV. UK [Online]*. Available: <https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2017>., 2017.
- [248] S.-j. Wu, N. Gebraeel, M. A. Lawley, and Y. Yih, “A neural network integrated decision support system for condition-based optimal predictive maintenance policy,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 37, no. 2, pp. 226–236, 2007.
- [249] E. Mohamed, A. Abdelaziz, and A. Mostafa, “A neural network-based scheme for fault diagnosis of power transformers,” *Electric Power Systems Research*, vol. 75, no. 1, pp. 29–39, 2005.
- [250] J. Li and J. Shi, “Knowledge discovery from observational data for process control using causal bayesian networks,” *IIE transactions*, vol. 39, no. 6, pp. 681–690, 2007.
- [251] H. Maki and Y. Teranishi, “Development of automated data mining system for quality control in manufacturing,” in *International Conference on Data Warehousing and Knowledge Discovery*. Springer, 2001, pp. 93–100.
- [252] N. Bhargava, G. Sharma, R. Bhargava, and M. Mathuria, “Decision tree analysis on j48 algorithm for data mining,” *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 6, 2013.
- [253] S. S. Haykin *et al.*, “Neural networks and learning machines/simon haykin.” 2009.
- [254] J. Quinlan, “The morgan kaufmann series in machine learning,” *San Mateo*, 1993.
- [255] M. J. Berry and G. S. Linoff, *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 2004.
- [256] Q. Wang, G. M. Garrity, J. M. Tiedje, and J. R. Cole, “Naive bayesian classifier for rapid assignment of rna sequences into the new bacterial taxonomy,” *Appl. Environ. Microbiol.*, vol. 73, no. 16, pp. 5261–5267, 2007.
- [257] I. Androutsopoulos, J. Koutsias, K. V. Chandrinou, G. Paliouras, and C. D. Spyropoulos, “An evaluation of naive bayesian anti-spam filtering,” *arXiv preprint cs/0006013*, 2000.

- [258] H. Zhang, Y.-L. Kang, Y.-Y. Zhu, K.-X. Zhao, J.-Y. Liang, L. Ding, T.-G. Zhang, and J. Zhang, “Novel naïve bayes classification models for predicting the chemical ames mutagenicity,” *Toxicology in Vitro*, vol. 41, pp. 56–63, 2017.
- [259] A. Krichene, “Using a naive bayesian classifier methodology for loan risk assessment,” *Journal of Economics, Finance and Administrative Science*, 2017.
- [260] K. M. Leung, “Naive bayesian classifier,” *Polytechnic University Department of Computer Science/Finance and Risk Engineering*, vol. 2007, pp. 123–156, 2007.
- [261] I. S. Association *et al.*, “Ieee recommended practices and requirements for harmonic control in electrical power systems,” *IEEE Std 519-1992*, pp. 1–112, April 1993.
- [262] S. K. Jain and S. Singh, “Harmonics estimation in emerging power system: Key issues and challenges,” *Electric Power Systems Research*, vol. 81, no. 9, pp. 1754 – 1766, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378779611001088>
- [263] M. Rausand and H. Arnljot, *System reliability theory: models, statistical methods, and applications*. John Wiley & Sons, 2004, vol. 396.
- [264] I. . W. Group *et al.*, “Ieee recommended practices and requirements for harmonic control in electrical power systems,” *IEEE STD*, pp. 519–1992, 1992.
- [265] B. IEC, “61000-3-4,” *Electromagnetic compatibility (EMC)-part*, pp. 3–4, 1998.
- [266] S. Radack, “The system development life cycle (sdlc),” National Institute of Standards and Technology, Tech. Rep., 2009.
- [267] S. D. Eppinger, “A planning method for integration of large-scale engineering systems,” in *International Conference on Engineering Design*, vol. 97, 1997, pp. 1–6.
- [268] Y. Zhu, *Multivariable system identification for process control*. Elsevier, 2001.
- [269] N. Karthik and V. Ananthanarayana, “Sensor data modeling for data trustworthiness,” in *2017 IEEE Trustcom/BigDataSE/ICSS*. IEEE, 2017, pp. 909–916.
- [270] X. Dai, C. Ke, Q. Quan, and K.-Y. Cai, “Simulation credibility assessment methodology with fpga-based hardware-in-the-loop platform,” *arXiv preprint arXiv:1907.03981*, 2019.

- [271] X. Zhang, A. Nassehi, M. Safaieh, and S. T. Newman, “Process comprehension for shopfloor manufacturing knowledge reuse,” *International Journal of Production Research*, vol. 51, no. 23-24, pp. 7405–7419, 2013.
- [272] J. K. Wong, H. Li, and S. Wang, “Intelligent building research: a review,” *Automation in construction*, vol. 14, no. 1, pp. 143–159, 2005.
- [273] En.climate-data.org, “Climate Malaysia: Average temperature, weather by month & weather for Malaysia,” <https://en.climate-data.org/asia/malaysia-25/>, 2019, online; accessed 11 November 2019.
- [274] B. de Bruijn, T. A. Nguyen, D. Bucur, and K. Tei, “Benchmark datasets for fault detection and classification in sensor data.” in *SENSORNETS*, 2016, pp. 185–195.
- [275] R. H. Tan and V. K. Ramachandaramurthy, “Numerical model framework of power quality events,” *European Journal of Scientific Research*, vol. 43, no. 1, pp. 30–47, 2010.
- [276] N. Viswanadham and Y. Narahari, “Performance modelling of automated manufacturing systems, 1992.”
- [277] J. Banks, J. S. Carson, B. L. Nelson, D. M. Nicol *et al.*, *Discrete-event system simulation*. Prentice hall Upper Saddle River, NJ, 1996, vol. 3.
- [278] A. Mousavi, A. Komashie, and S. Tavakoli, “Simulation-based real-time performance monitoring (simmon): A platform for manufacturing and healthcare systems,” in *Proceedings of the Winter Simulation Conference*. Winter Simulation Conference, 2011, pp. 600–611.
- [279] J. E. Bowen, M. C. Moore, and M. W. Wactor, “Application of switchgear in unusual environments,” in *2010 Record of Conference Papers Industry Applications Society 57th Annual Petroleum and Chemical Industry Conference (PCIC)*. IEEE, 2010, pp. 1–10.

Appendix A

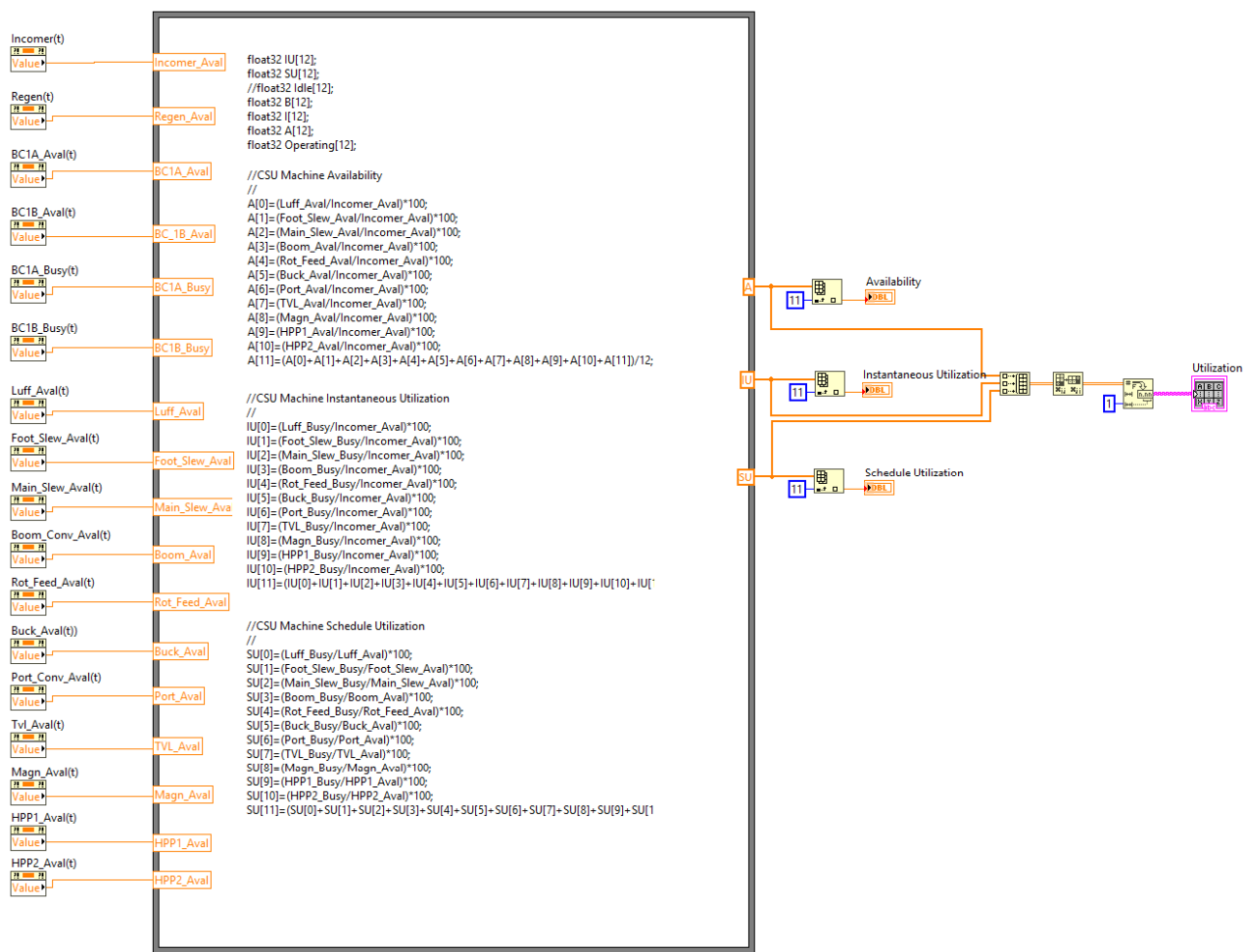


Figure A.1: Snippets of CSU Machine KPI

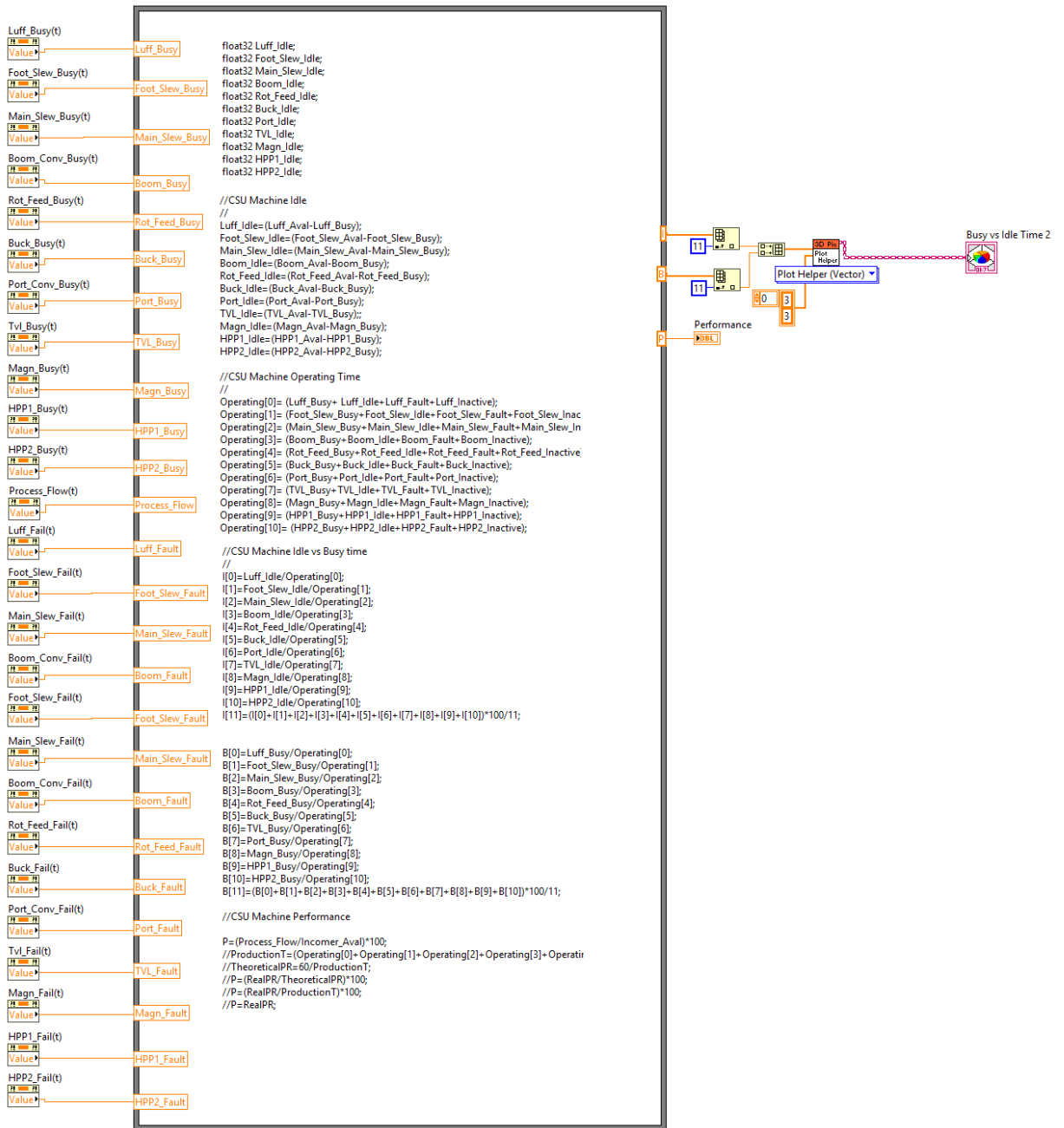


Figure A.2: Snippets of CSU Machine KPI

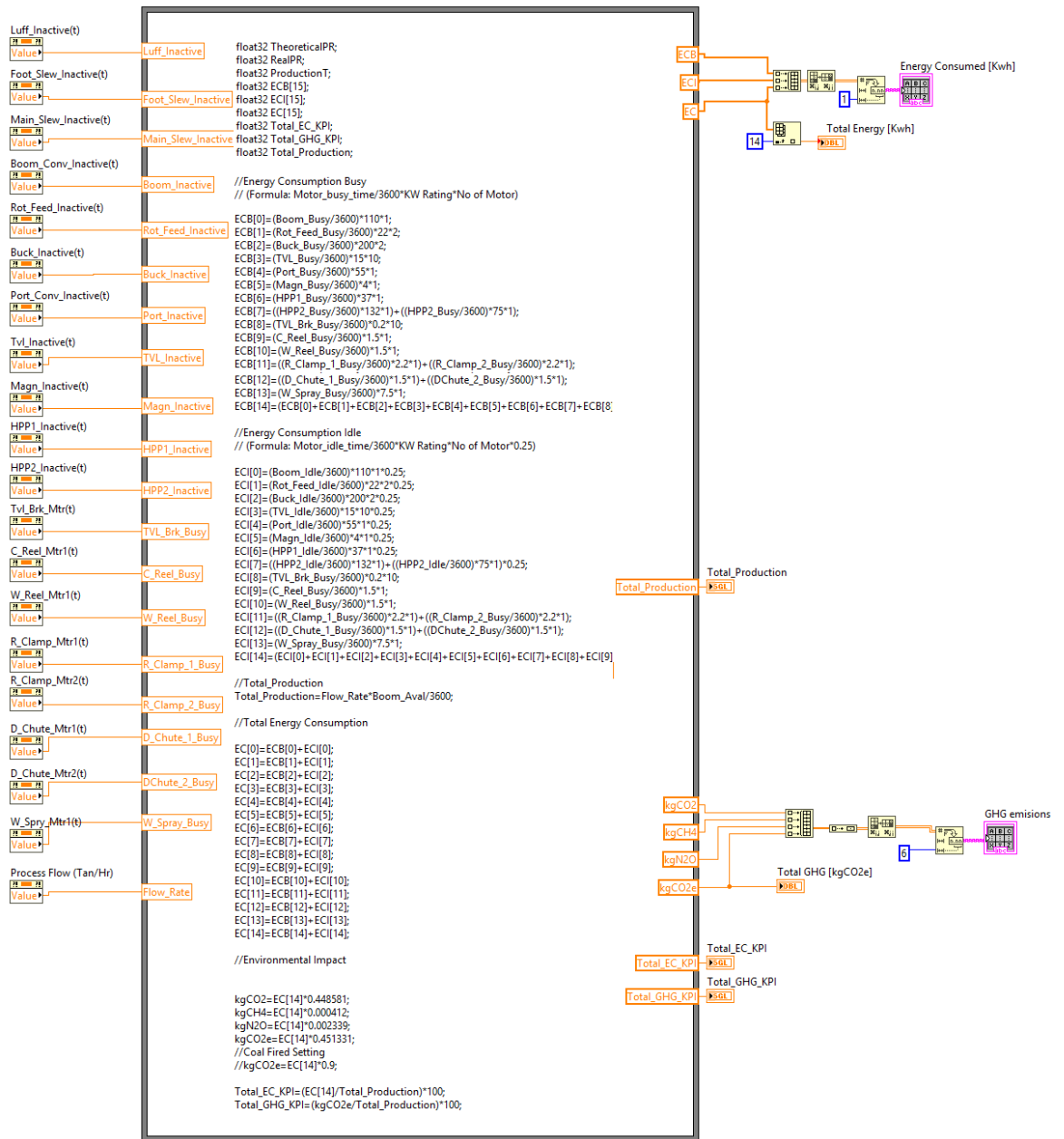


Figure A.3: Snippets of Energy Efficiency KPI

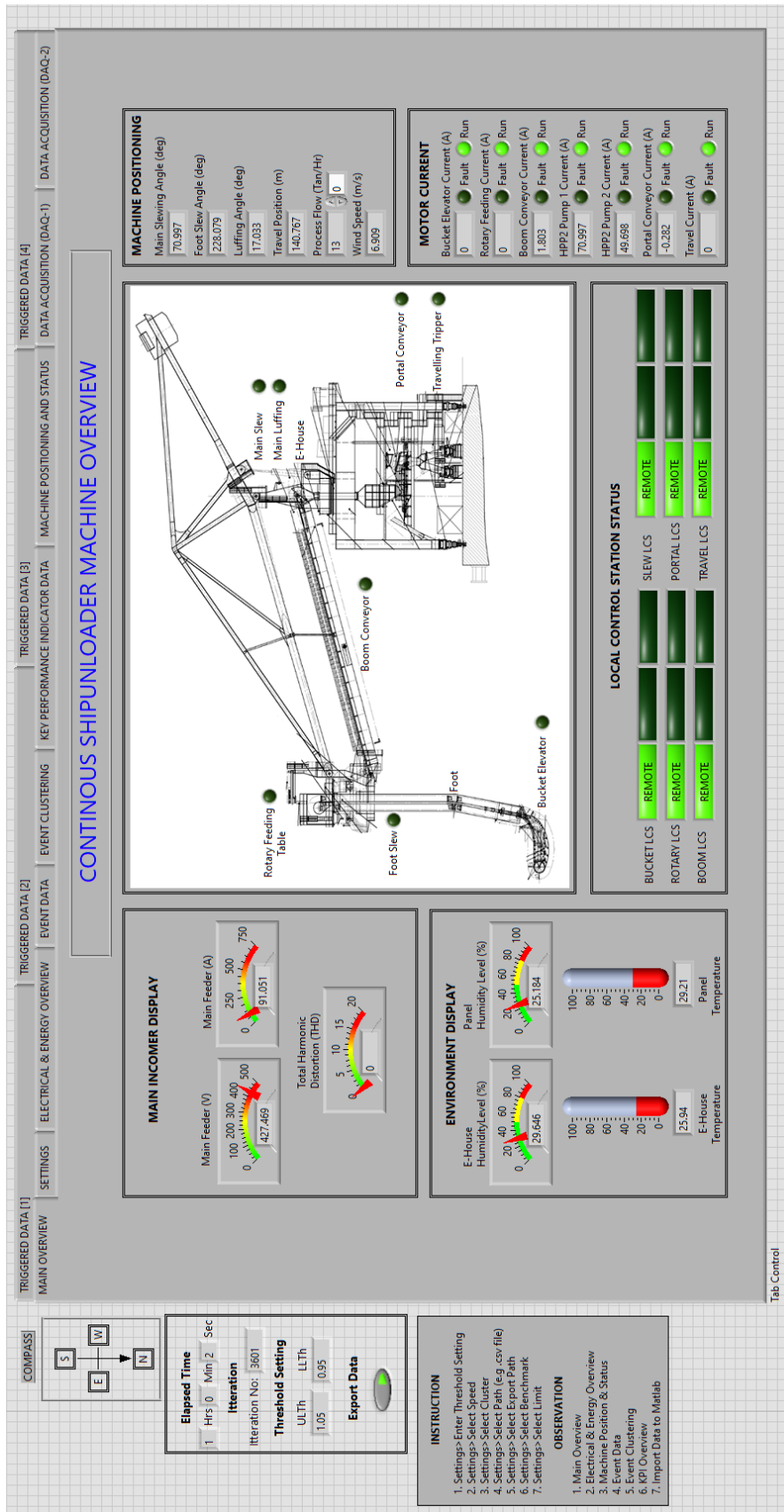


Figure A.4: CSU Machine Event Modeller - Main Overview

TRIGGERED DATA [1]
TRIGGERED DATA [2]
TRIGGERED DATA [4]

MAIN OVERVIEW
SETTINGS
ELECTRICAL & ENERGY OVERVIEW
EVENT DATA
EVENT CLUSTERING
KEY PERFORMANCE INDICATOR DATA
MACHINE POSITIONING AND STATUS
DATA ACQUISITION (DAQ-1)
DATA ACQUISITION (DAQ-2)

CONTINUOUS SHIPUNLOADER MACHINE EVENT MODELLER SETTINGS

Elapsed Time

1 Hrs 0 Min 2 Sec

Iteration

Iteration No: 3601

Threshold Setting

ULTh 0.95

LLTh 1.05

Export Data

1.0 Enter Threshold Setting:
(e.g. 0.05)

0.05

2.0 Select Cluster

Cluster 1

3.0 Select Speed:

Speed (ms) 1000

Algo Delay 1000

4.0 Select Path (e.g. .csv file):

Import PLC Data File Source

C:\Futura\...\6.Data Selected\2018 11 02 0000 (Float).csv

5.0 Select Export Path:

Export Data to MATLAB

C:\...\6.Database Export\20181102_Data_Export_001.csv

6.0 Select Benchmark

Select all Benchmark Setting for ED Variables. This setting could be adjusted based on System Expert.

7.0 Select Limit

Select all Limit Setting for ED Variables. This setting could be adjusted based on System Expert.

INSTRUCTION

1. Settings->Enter Threshold Setting
2. Settings->Select Speed
3. Settings->Select Cluster
4. Settings->Select Path (e.g. .csv file)
5. Settings->Select Export Path
6. Settings->Select Benchmark
7. Settings->Select Limit

OBSERVATION

1. Main Overview
2. Electrical & Energy Overview
3. Machine Position & Status
4. Event Data
5. Event Clustering
6. KPI Overview
7. Import Data to Matlab

ED	ED 1: MAIN INCOMER VOLTAGE (V)	ED 2: MAIN INCOMER CURRENT (A)	ED 3: PANEL HUMIDITY (%)	ED 4: MAIN INCOMING THD (%)	ED 5: PANEL TEMPERATURE (deg C)	ED 6: E-HOUSE TEMPERATURE (deg C)	ED 7: MAIN SLEW ANGLE (deg)	ED 8: TRAVEL POSITION (m)	ED 9: FOOT SLEW ANGLE (deg)	ED 10: LUFFING ANGLE (deg)	ED 11: PROCESS FLOW (Tonn/hr)	ED 12: WIND SPEED (m/s)	ED 13: ROTARY SPEED DRIVE	ED 14: PORTAL SPEED DRIVE	ED 15: BUCKET SPEED DRIVE	ED 16: TRAVEL SPEED DRIVE	ED 17: BUCKET ELEVATOR CURRENT (A)	ED 18: PORTAL CONVEYOR CURRENT (A)	ED 19: TRAVEL CURRENT (A)	ED 20: BOOM CONVEYOR CURRENT (A)	ED 21: HPP2 PUMP 2 CURRENT (A)	ED 22: HPP2 PUMP 1 CURRENT (A)	ED 23: ROTARY FEEDING CURRENT (A)	ED 24: E-HOUSE HUMIDITY LEVEL (%)																				
427	427.469	0	91.051	0	25.184	0	29.208	0	25.936	1	70.997	0	140.767	0	228.079	0	17.033	0	13	0	6.909	1	0.051	1	-0.051	1	0.102	1	-0.102	0	0	0	-0.282	1	1.803	0	48.698	0	70.997	0	0	0	29.646	0

BENCHMARK	LIMIT +/-	READING CHANGES	READING CHANGES
427	1	0	0
300	300	0	0
20	10	0	0
1	1	0	0
30	5	0	0
20	5	1	0
80	60	0	0
150	50	0	0
180	180	0	0
5	15	0	0
500	500	0	0
3	3	0	0
25	15	1	0
25	15	1	0
25	15	1	0
0	15	0	0
300	300	0	0
15	15	1	0
0	200	0	0
40	40	0	0
40	20	0	0
100	50	0	0
0	10	0	0
30	10	0	0

TD	TD 1: Main Incomer Status	TD 2: Maintenance Feeder	TD 3: REGEN 1 Enable	TD 4: REGEN 2 Enable	TD 5: REGEN 3 Enable	TD 6: Boom_MTR1	TD 7: Boom_Brake_MTR1	TD 8: Rotary_MTR1 & MTR2	TD 9: Bucket_Elevator_MTR1 & MTR2	TD 10: Bucket_Elevator_Brake_MTR1& MTR2	TD 11: Portal_MTR1	TD 12: Travel_MTR1 - MTR 10	TD 13: Travel_Brake_MTR1 - MTR 10	TD 14: Magnetic_Separator_MTR1	TD 15: HPP1_Foot_MTR1	TD 16: HPP1_MTR1	TD 17: HPP1_MTR2	TD 18: Cable_Reel_MTR1	TD 19: Water_hose_Reel_MTR1	TD 20: Rail Clamp_MTR1	TD 21: Rail Clamp_MTR2	TD 22: Diverter_Chute_MTR1	TD 23: Diverter_Chute_MTR2	TD 24: Water_Spray_MTR1	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tab Control

Figure A.5: CSU Machine Event Modeller - Main Setting

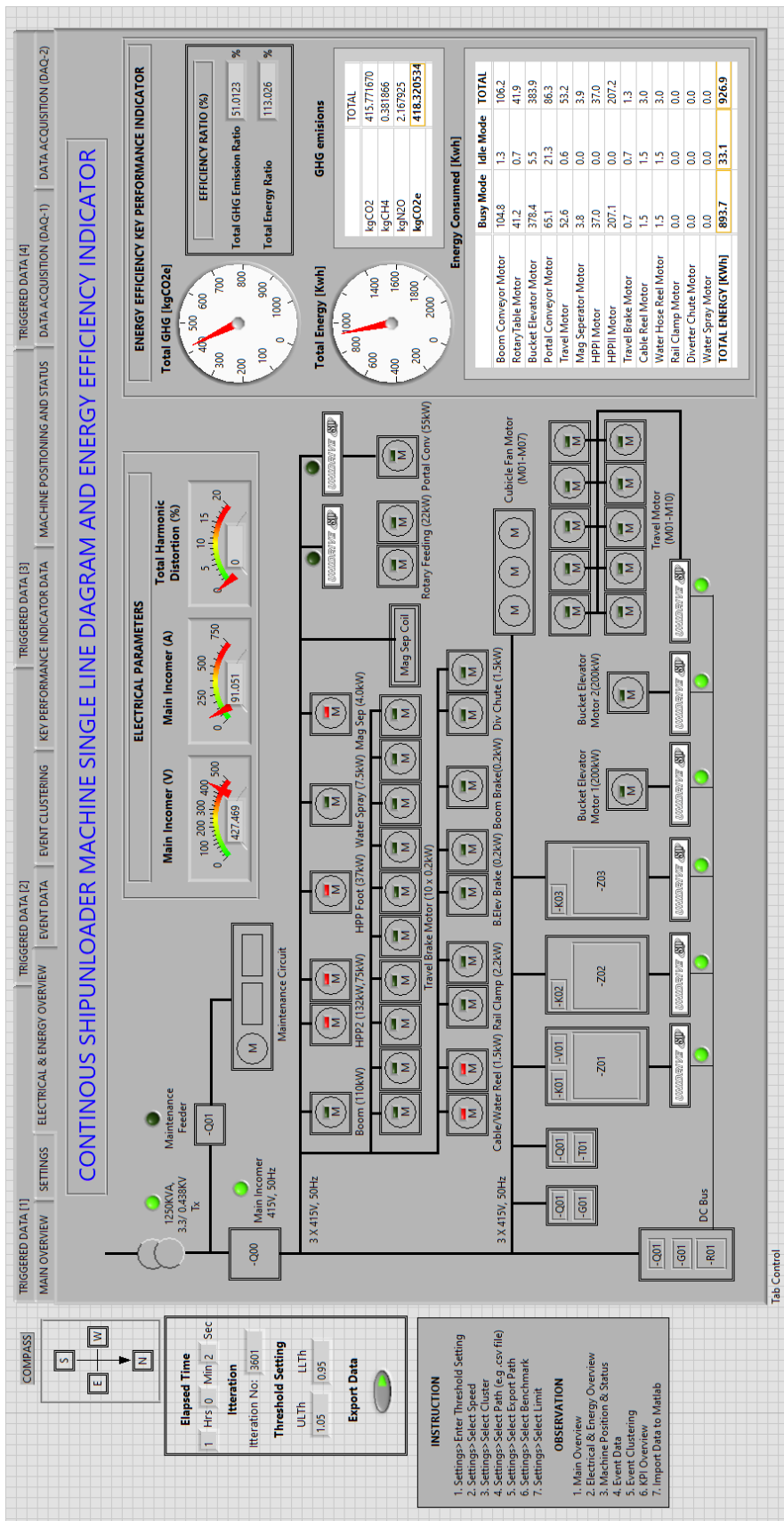


Figure A.6: CSU Machine Event Modeller - SLD and Energy Efficiency

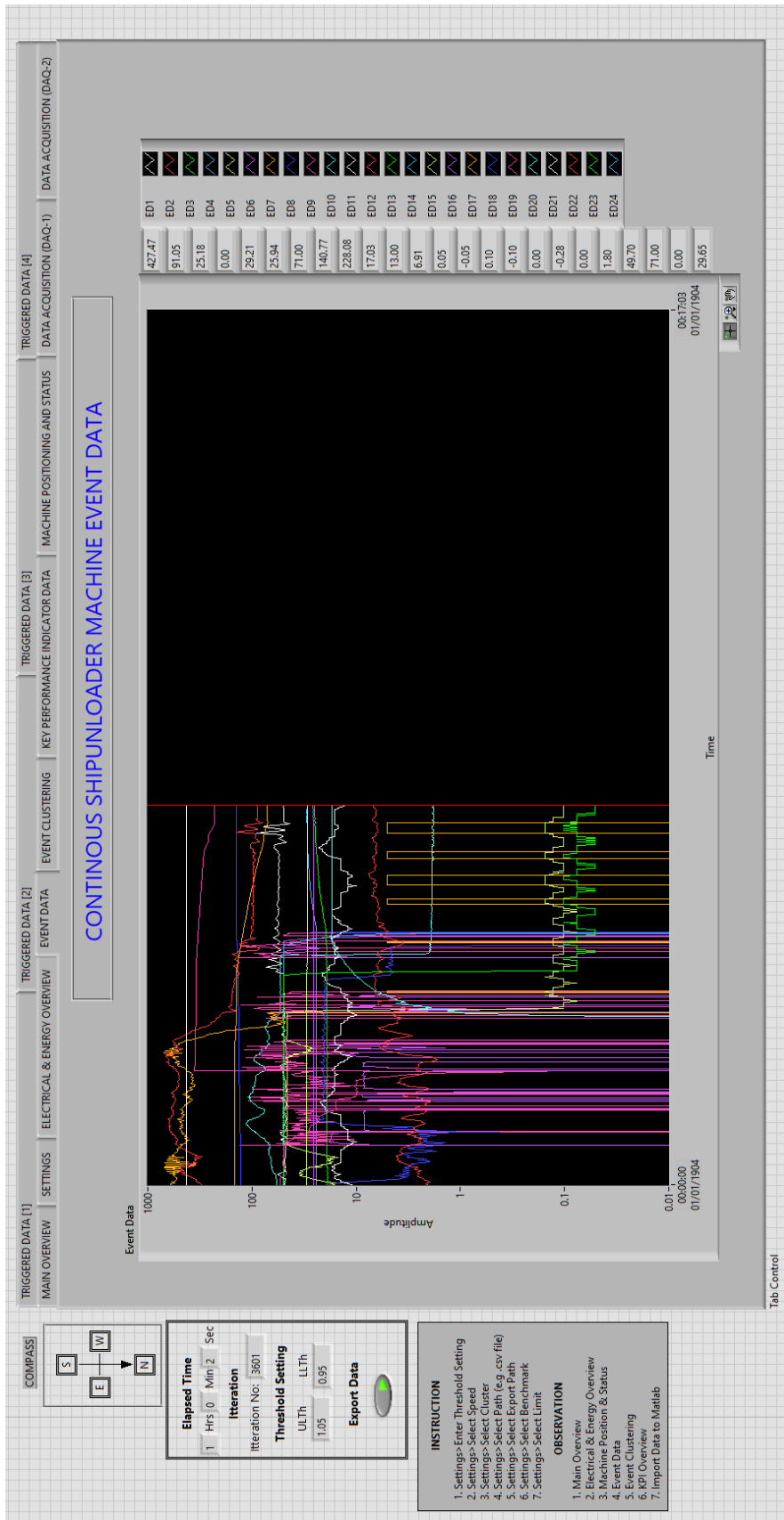


Figure A.7: CSU Machine Event Modeller - Event Data Trending

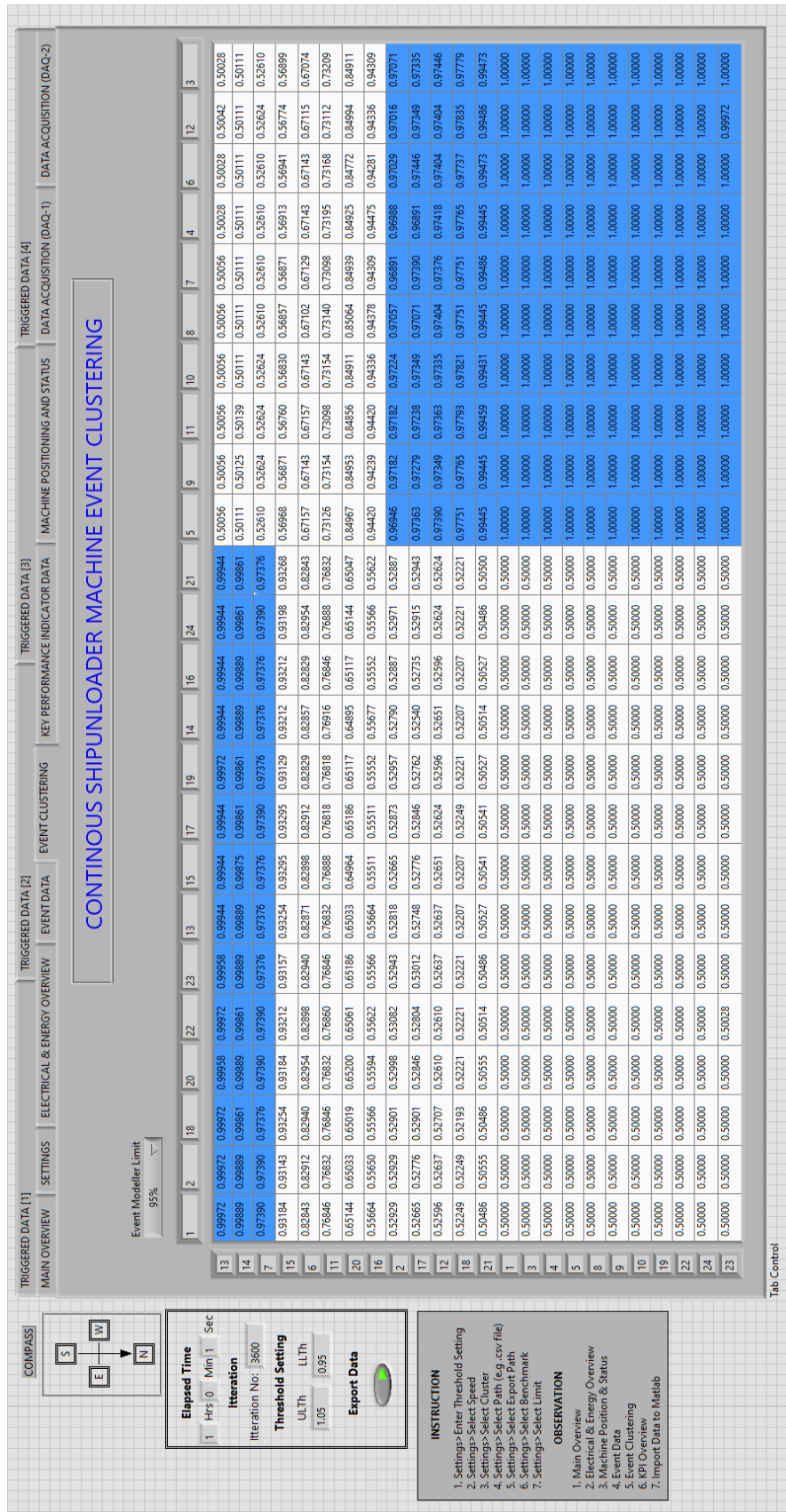


Figure A.8: CSU Machine Event Modeller - Event Clustering

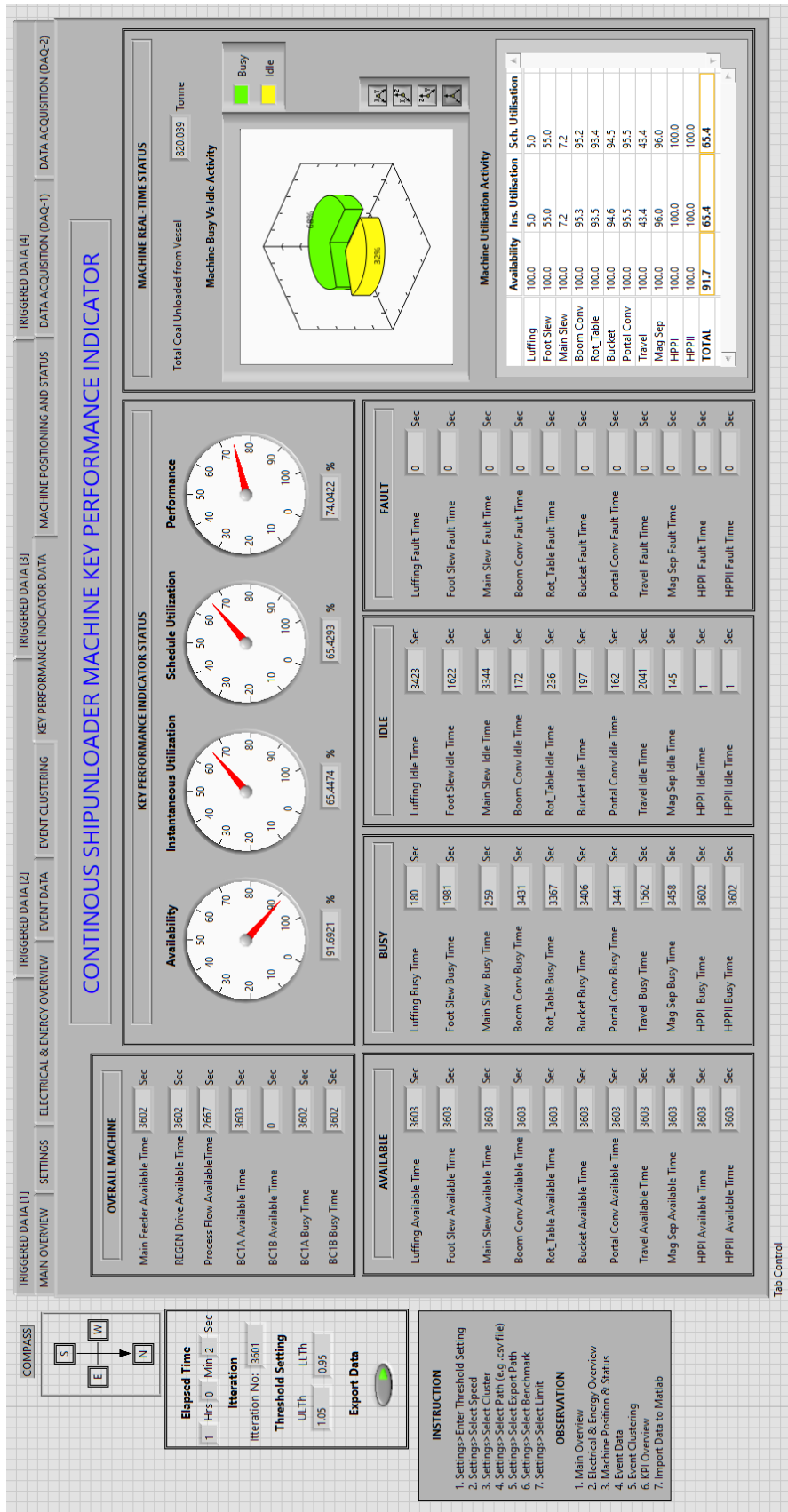


Figure A.9: CSU Machine Event Modeller - Key Performance Indicator

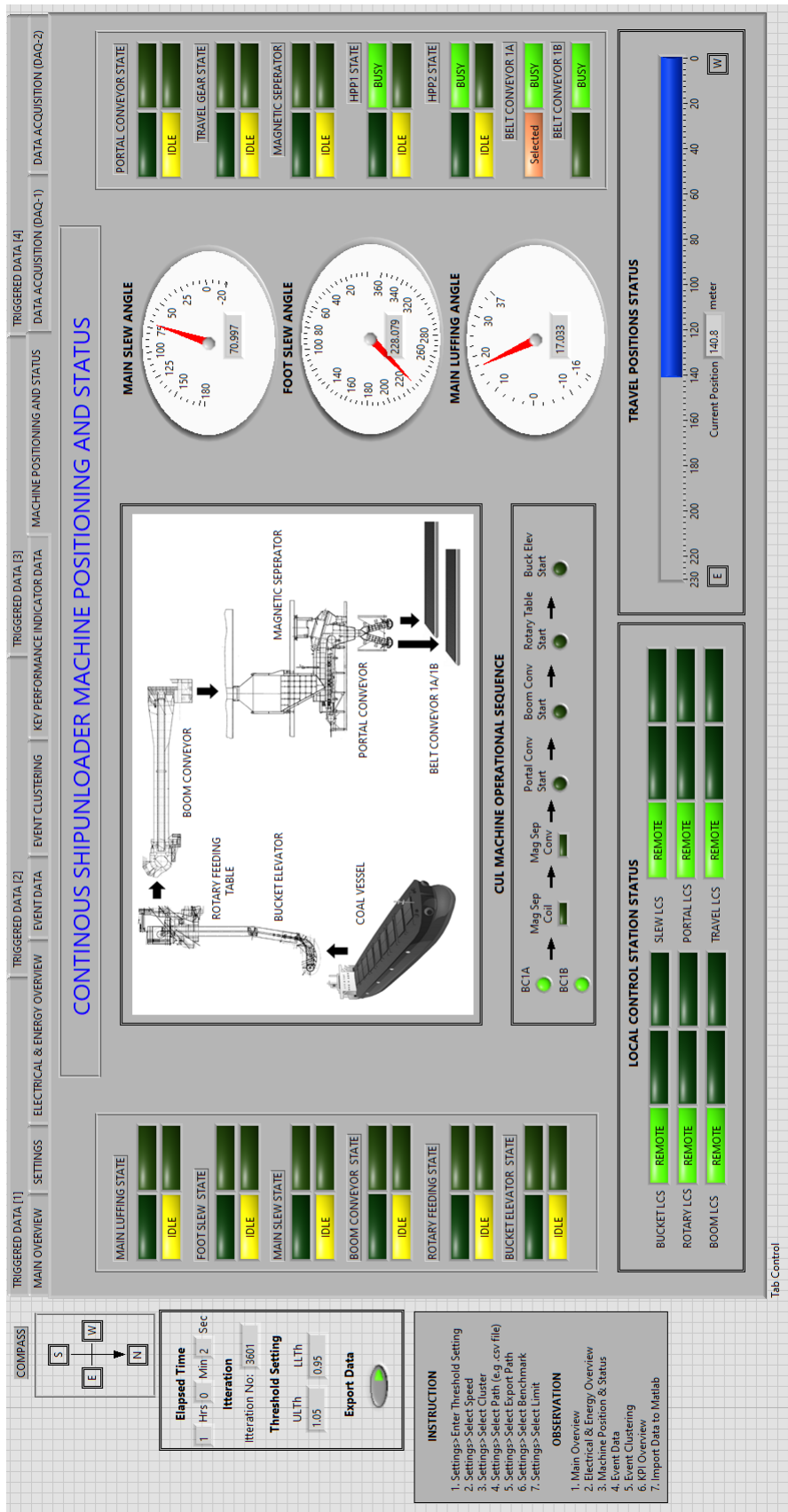


Figure A.10: CSU Machine Event Modeller - Machine Positioning & Status

Appendix B

Table B.2: Test Plan for Event Modeller Development

Test ID	Functionality Description	Test Input Data	Expected Output	Actual Output	Status
1	Run with Default Setting	N/A	Program run simultaneous data from subVI with invalid output values.	The output value indicates binary '1' for all values.	Fail
2	Run with TD Type = Boolean	Toggle Switch	TD Status bit is display Green LED	TD Status bit is display Green LED	Pass
3	Run with TD Type = Analogue	Toggle Switch	TD Status bit did Not appears	TD Status bit did Not appears	Pass
4	Run with Threshold Setting - 0.05	0.05	UL and LL denotes as 1.05 and 0.95	UL and LL denotes as 1.05 and 0.95	Pass
5	Run with Threshold Setting - 0.10	0.10	UL and LL denotes as 1.10 and 0.90	UL and LL denotes as 1.15 and 0.90	Pass
6	Run with Threshold Setting - 0.15	0.15	UL and LL denotes as 1.15 and 0.85	UL and LL denotes as 1.15 and 0.85	Pass
7	Run with Event Modeller Limit - 0.90	0.90	The pattern shows correlation with 90% confidence	The pattern shows correlation with 90% confidence	Pass
8	Run with Event Modeller Limit - 0.80	0.80	The pattern shows correlation with 80% confidence	The pattern shows correlation with 80% confidence	Pass
9	Run with Event Modeller Limit - 0.70	0.70	The pattern shows correlation with 70% confidence	The pattern shows correlation with 70% confidence	Pass
10	Run with Voltage Disturbance button ON	Toggle Button	Program run simultaneous data from subVI with Voltage Disturbance	The output wave-form for Voltage is fluctuating	Pass
11	Run with Humidity Disturbance button ON	Toggle Button	Program run simultaneous data from subVI with Humidity Disturbance.	The output wave-form for Humidity is fluctuating	Pass
12	User select all required Settings	N/A	Program run simultaneous data from subVI with good output values.	The output shows good correlation matrix	Pass

Table B.3: Summary of System Operator Daily Log Sheet for CSU Machine

Date	No	Shift	CUL Status	Ship Name	Total Unloaded	Abnormality	Weather	Abnormality Remarks	Weather Remarks
01-Oct-18	6324	Day	Missing	Missing	Missing	Missing	Missing	Missing	Missing
01-Oct-18	6325	Night	Healthy	N/A	N/A	No	Yes	N/A	Raining - 02:05AM - 03:55 AM
02-Oct-18	6326	Day	Healthy	N/A	N/A	No	No	N/A	N/A
02-Oct-18	6327	Night	Healthy	N/A	N/A	No	No	N/A	N/A
03-Oct-18	6328	Day	Healthy	N/A	N/A	No	No	N/A	N/A
03-Oct-18	6329	Night	Healthy	N/A	N/A	No	No	N/A	N/A
04-Oct-18	6330	Day	Healthy	N/A	N/A	No	No	N/A	N/A
04-Oct-18	6331	Night	Healthy	N/A	N/A	No	No	N/A	N/A
05-Oct-18	6332	Day	In-Service	MV POS LOGISTIC 2	600 MT	No	No	N/A	N/A
05-Oct-18	6333	Night	In-Service	MV POS LOGISTIC 2	11618 MT	No	No	N/A	N/A
06-Oct-18	6334	Day	In-Service	MV POS LOGISTIC 2	18300 MT	No	No	N/A	N/A
06-Oct-18	6335	Night	In-Service	MV POS LOGISTIC 2	29854 MT	No	No	N/A	N/A
07-Oct-18	6336	Day	In-Service	MV POS LOGISTIC 2	37065 MT	No	No	N/A	N/A
07-Oct-18	6337	Night	In-Service	MV POS LOGISTIC 2	37065 MT	No	No	N/A	N/A
08-Oct-18	6338	Day	In-Service	MV POS LOGISTIC 2	46300 MT	No	No	N/A	N/A
08-Oct-18	6339	Night	In-Service	MV POS LOGISTIC 2	46500 MT	No	Yes	N/A	Raining - 01:00AM - 06:40AM
09-Oct-18	6340	Day	In-Service	MV POS LOGISTIC 2	60081 MT	No	No	N/A	N/A
09-Oct-18	6341	Night	In-Service	MV POS LOGISTIC 2	60081 MT	No	No	N/A	N/A
10-Oct-18	6342	Day	In-Service	MV POS LOGISTIC 2	60091 MT	No	No	N/A	N/A
10-Oct-18	6343	Night	In-Service	MV POS LOGISTIC 2	66000 MT	No	Yes	N/A	Raining - 19:01PM - 06:40AM
11-Oct-18	6344	Day	In-Service	MV POS LOGISTIC 2	68903 MT	No	No	N/A	N/A
11-Oct-18	6345	Night	In-Service	MV POS LOGISTIC 2	68903 MT	No	Yes	N/A	Raining - 19:01PM - 01:09AM
12-Oct-18	6346	Day	In-Service	MV POS LOGISTIC 2	69543 MT	No	Yes	N/A	Raining - 16:24PM - 17:55PM
12-Oct-18	6347	Night	In-Service	MV POS LOGISTIC 2	72090 MT	No	Yes	N/A	Raining - 05:39AM
13-Oct-18	6348	Day	In-Service	MV POS LOGISTIC 2	73100 MT	Yes	Yes	Reset regent fault at CUL 2	Raining - 07:00AM
13-Oct-18	6349	Night	In-Service	MV POS LOGISTIC 2	76901 MT	No	Yes	N/A	Raining - 06:00AM
14-Oct-18	6350	Day	In-Service	MV TASIK SAKURA	3500 MT	No	No	N/A	N/A
14-Oct-18	6352	Night	In-Service	MV TASIK SAKURA	8969 MT	No	No	N/A	N/A
15-Oct-18	6353	Day	In-Service	MV TASIK SAKURA	12700 MT	No	Yes	N/A	Raining - 07:01AM
15-Oct-18	6354	Night	In-Service	MV TASIK SAKURA	17964 MT	No	Yes	Reset regent fault at CUL 1	Raining - 03:29AM
16-Oct-18	6355	Day	In-Service	MV TASIK SAKURA	20501 MT	No	Yes	N/A	Raining - 07:01AM
16-Oct-18	6356	Night	In-Service	MV TASIK SAKURA	26446 MT	No	No	N/A	N/A
17-Oct-18	6357	Day	Missing	Missing	Missing	Missing	Missing	Missing	Missing
17-Oct-18	6358	Night	In-Service	MV TASIK SAKURA	35600 MT	Yes	Yes	*CUL 2 regen fault alarm	Raining - 20:51 PM
18-Oct-18	6359	Day	In-Service	MV TASIK SAKURA	44300 MT	No	No	N/A	N/A
18-Oct-18	6360	Night	In-Service	MV TASIK SAKURA	54300 MT	No	No	N/A	N/A
19-Oct-18	6361	Day	In-Service	MV TASIK SAKURA	62889 MT	No	No	N/A	N/A
19-Oct-18	6362	Night	In-Service	MV TASIK SAKURA	62889 MT	Yes	No	Alarm foreign body protection actuated (03:45AM) - CUL 2 N/A	N/A
20-Oct-18	6363	Day	In-Service	MV TASIK SAKURA	76500 MT	No	No	N/A	N/A
20-Oct-18	6364	Night	Healthy	N/A	N/A	No	No	N/A	N/A
21-Oct-18	6365	Day	Healthy	N/A	N/A	No	No	N/A	N/A
21-Oct-18	6366	Night	Healthy	N/A	N/A	No	No	N/A	N/A
22-Oct-18	6367	Day	Healthy	N/A	N/A	No	No	N/A	N/A
22-Oct-18	6368	Night	In-Service	MV OCEAN OCEANUS	3295 MT	No	Yes	N/A	Raining - 02:40AM, 04:28AM
23-Oct-18	6369	Day	Missing	Missing	Missing	Missing	Missing	Missing	Missing
23-Oct-18	6370	Night	In-Service	MV OCEAN OCEANUS	34672 MT	No	No	N/A	N/A
24-Oct-18	6371	Day	In-Service	MV OCEAN OCEANUS	47151 MT	No	No	N/A	N/A
24-Oct-18	6372	Night	In-Service	MV OCEAN OCEANUS	65457 MT	Yes	No	CUL 2 HPPP hose leak.	N/A
25-Oct-18	6373	Day	Healthy	MV OCEAN OCEANUS	78238 MT	No	No	N/A	N/A
25-Oct-18	6374	Night	Healthy	N/A	N/A	No	No	N/A	N/A
26-Oct-18	6375	Day	In-Service	MV INCEPTION	3880 MT	No	No	N/A	N/A
26-Oct-18	6377	Night	In-Service	MV INCEPTION	12700 MT	Yes	Yes	CUL rotary feeding table underspeed	Raining - 19:01PM, 22:40PM
27-Oct-18	6378	Day	In-Service	MV INCEPTION	29100 MT	Yes	No	CUL 2 bucket touching with foreign body. dismantle foreign body to repair	N/A
27-Oct-18	6379	Night	In-Service	MV INCEPTION	36537 MT	No	Yes	N/A	Raining - 19:20 PM
28-Oct-18	6380	Day	In-Service	MV INCEPTION	50500 MT	Yes	No	Reset regent fault at CUL 2	N/A
28-Oct-18	6381	Night	In-Service	MV INCEPTION	64039 MT	No	No	N/A	N/A
29-Oct-18	6382	Day	In-Service	MV INCEPTION	77714 MT	Yes	No	Reset regent fault at CUL 2	N/A
29-Oct-18	6384	Night	Healthy	N/A	N/A	No	No	N/A	N/A
30-Oct-18	6385	Day	Healthy	N/A	N/A	No	No	Reset regent fault at CUL 1	N/A
30-Oct-18	6386	Night	Healthy	N/A	N/A	No	No	N/A	N/A
31-Oct-18	6387	Day	Healthy	N/A	N/A	Yes	No	Reset regent fault at CUL 2	N/A
31-Oct-18	6388	Night	Healthy	N/A	N/A	No	No	N/A	N/A

Table B.4: Conversion factor for Coal Combustion

Fuel	Unit	kg CO2e	kg CO2	kg CH4	kg N2O
Coal (industrial)	tonnes	2,437.49	2,414.19	6.66	16.64
	kWh (Net CV)	0.34149	0.33823	0.00093	0.00233
	kWh (Gross CV)	0.32442	0.32132	0.00089	0.00221
Coal (electricity generation)	tonnes	2,244.63	2,231.40	0.64	12.59
	kWh (Net CV)	0.32490	0.32298	0.00009	0.00182
	kWh (Gross CV)	0.30865	0.30683	0.00009	0.00173
Coal (domestic)	tonnes	2,861.96	2,632.92	191.93	37.11
	kWh (Net CV)	0.36008	0.33127	0.02415	0.00467
	kWh (Gross CV)	0.34208	0.31470	0.02294	0.00444
Coking coal	tonnes	3,126.81	3,069.93	40.12	16.77
	kWh (Net CV)	0.37224	0.36547	0.00478	0.00200
	kWh (Gross CV)	0.35363	0.34719	0.00454	0.00190
Petroleum coke	tonnes	3,385.40	3,375.14	2.94	7.32
	kWh (Net CV)	0.35875	0.35766	0.00031	0.00078
	kWh (Gross CV)	0.34081	0.33978	0.00030	0.00074
Coal (home produced coal only)	tonnes	2,244.63	2,231.40	0.64	12.59
	kWh (Net CV)	0.33861	0.33661	0.00010	0.00190
	kWh (Gross CV)	0.32168	0.31978	0.00009	0.00180

Table B.5: 96 Input Variable List

No	Description	No	Description	No	Description	No	Description
TD1	Main Incomer Status	TD25	Rotary Drive Enable	TD49	Inactive Luff	TD73	Idle Luff
TD2	Maintenance Feeder	TD26	Portal Drive Enable	TD50	Inactive Foot Slew	TD74	Idle Foot Slew
TD3	REGEN 1 Enable	TD27	Bucket Drive Enable	TD51	Inactive Slew	TD75	Idle Slew
TD4	REGEN 2 Enable	TD28	Travel Drive Enable	TD52	Inactive Boom	TD76	Idle Boom
TD5	REGEN 3 Enable	TD29	Rotary Feed State	TD53	Inactive Rotary	TD77	Idle Rotary
TD6	Boom Motor	TD30	Portal State	TD54	Inactive Buck	TD78	Idle Buck
TD7	Boom Brake Motor	TD31	Slew Remote LCS	TD55	Inactive Portal	TD79	Idle Portal
TD8	Rotary Motors	TD32	Boom Remote LCS	TD56	Inactive Travel	TD80	Idle Travel
TD9	Buck Motors	TD33	Rotary Remote LCS	TD57	Inactive Mag Separator	TD81	Idle Mag Separator
TD10	Buck Brake Motors	TD34	Buck Remote LCS	TD58	Inactive HPPI	TD82	Idle HPPI
TD11	Portal Motor	TD35	Portal Remote LCS	TD59	Inactive HPPII	TD83	Idle HPPII
TD12	Travel Motors	TD36	Travel Remote LCS	TD60	Failure Luff	TD84	Busy Luff
TD13	Travel Brake Motors	TD37	Slew Local LCS	TD61	Failure Foot Slew	TD85	Busy Foot Slew
TD14	Mag Separator Motor	TD38	Boom Local LCS	TD62	Failure Slew	TD86	Busy Slew
TD15	HPPI Motor	TD39	Rotary Local LCS	TD63	Failure Boom	TD87	Busy Boom
TD16	HPPII Motor 1	TD40	Buck Local LCS	TD64	Failure Rotary	TD88	Busy Rotary
TD17	HPPII Motor 2	TD41	Portal Local LCS	TD65	Failure Buck	TD89	Busy Buck
TD18	Cable Reel Motor	TD42	Travel Local LCS	TD66	Failure Portal	TD90	Busy Portal
TD19	Water Hose Reel Motor	TD43	Slew Disable LCS	TD67	Failure Travel	TD91	Busy Travel
TD20	Rail Clamp Motor 1	TD44	Boom Disable LCS	TD68	Failure Mag Separator	TD92	Busy Mag Separator
TD21	Rail Clamp Motor 2	TD45	Rotary Disable LCS	TD69	Failure HPPI	TD93	Busy HPPI
TD22	Diverter Chute Motor 1	TD46	Buck Disable LCS	TD70	Failure HPPII	TD94	Busy HPPII
TD23	Diverter Chute Motor 2	TD47	Portal Disable LCS	TD71	Selected BC1A	TD95	Busy BC1A
TD24	Water Spray Motor	TD48	Travel Disable LCS	TD72	Selected BC1B	TD96	Busy BC1B

Table B.6: 24 Output Variable List

No	Description	Unit
ED1	Main Incomer Voltage	V
ED2	Main Incomer Current	A
ED3	Panel Humidity	%
ED4	Main Incoming THD	%
ED5	Panel Temperature	degC
ED6	E-House Temperature	degC
ED7	Main Slewing Angle	deg
ED8	Travel Position	m
ED9	Foot Slew Angle	deg
ED10	Luffing Angle	deg
ED11	Process Flow	Tan/hr
ED12	Wind Speed	m/s
ED13	Rotary Drive Speed	r/min
ED14	Portal Drive Speed	r/min
ED15	Bucket Drive Speed	r/min
ED16	Travel Drive Speed	r/min
ED17	Bucket Elevator Current	A
ED18	Portal Conveyor Current	A
ED19	Travel Current	A
ED20	Boom Conveyor Current	A
ED21	HPP2 Pump 2 Current	A
ED22	HPP2 Pump 1 Current	A
ED23	Rotary Feeding Current	A
ED24	E-House Humidity	%