Data-driven Customer Behaviour Model Generation for Agent Based Exploration

Chidozie Mgbemena
Brunel University London
Kingston lane, Uxbridge, London
chidozie.mgbemena@brunel.ac.uk

David Bell
Brunel University London
Kingston lane, Uxbridge, London
david.bell@brunel.ac.uk

ABSTRACT
Customer retention is a critical concern for most mobile network operators because of the increasing competition in the mobile services sector. This concern has driven companies to exploit data as an avenue to better understand customer needs. Data mining techniques such as clustering and classification have been adopted to understand customer retention in the mobile services industry. However, the effectiveness of these techniques is debatable due to the increasing complexity of the mobile market itself. This study proposes an application of Agent-Based Modeling and Simulation (ABMS) as a novel approach to understanding customer retention. A dataset provided by a mobile network operator is utilized to automate decision trees and agent based models. The most popular churn modeling techniques were adopted in order to automate the development of models, from decision trees, and subsequently explore customer churn scenarios. ABMS is used to understand the behavior of customers and detect possible reasons why customers churned or stayed with their respective mobile network operators. Data analysis is able to identify that location and choice of mobile devices were determinants for the decision to churn or stay with their mobile network operator - with word of mouth as an important factor. Importantly, agent based simulation is able to explore further the determinants in the wider marketplace.

Author Keywords
Agent Based Modelling; CADET Approach; Decision Trees.

1. INTRODUCTION
Simulation models that describe agents have become a widely used tool for understanding phenomena. These models have been used across industries and often prove to be sufficient in providing insights to problems. A popular simulation model is the Agent-based model (ABM). The agent-based model allows researchers and practitioners to study how system level properties emerge from the adaptive behaviour of individuals and on the other hand how systems affect individuals[10].

Agent-based models consist of a number of entities with individual rules of behaviours. Entities in such models interact with one another and with their surrounding environment. Such interaction may influence the behaviour of the agents. Harnessing this information and understanding the influence of agents interaction with other agents and agent interaction with the environment can provide some useful insights to business problems, in this case customer churn.

There are various ways of describing agent based models. This paper presents the Customer Agent Decision Tree (CADET) approach, a novel approach to describing Agent-Based models. The CADET approach is a data-driven approach that provides the key drivers that collectively uncover the decision of individual agents in an environment. The CADET approach is not industry specific. Thus, it can be applied to different sectors. The CADET approach is applied in this research paper as an overarching structure to study customer retention in the mobile services industry.

2. CUSTOMER RETENTION IN THE MOBILE SERVICES INDUSTRY
Over the last decade, the number of mobile phone users have increased reaching an overwhelming number of 7 billion [24]. In developed countries, telecommunication companies have mobile penetration rates above 100% with no new customers [30]. Therefore, customer retention receives a growing amount of attention from telecommunication companies. The high penetration rate of above 100% in the mobile services sector has inspired many management scholars to devote in conducting more substantial research in customer retention. It has been shown in the literature that customer retention is profitable to a company because: (1) acquiring new customers cost five times more than retaining existing customers [17, 13]. (2) Existing customers generate higher profits, become less costly to serve, and may provide new referrals through providing positive word-of-mouth while dissatisfied customers might spread negative word of mouth [16, 9, 4] (3) Loosing customers may lead to opportunity costs because of reduced sales [15, 26] However, a little improvement in customer retention can lead to a significant increase in profit [7]. A number of studies in the area of customer retention have revealed that customer satisfaction is a strong predictor of customer retention [1].

Customer churn is a term widely used in the area of customer retention to describe customers who switch to a different mobile service provider or leave the market entirely.

There are two basic approaches that can be used to address customer churn namely, untargeted and targeted approaches [18]. Untargeted approaches rely on outstanding product and mass advertising to increase brand loyalty and retain customers while targeted approaches rely on identifying customers who are likely to churn and then providing them with either a direct incentive or a customised plan for them to stay [18]. Various types of information can also be used to predict customer churn, such as information on socio demographic data (e.g. sex, age, or post code) and call behaviour statistics (e.g. the number of international calls, billing information, or the number of calls to the customer helpdesk).

The main factors that influence customer churn in the mobile services market are (1) Customer satisfaction, (2) Switching costs, (3) relationship quality and (4) Price [11, 27, 28]. Price is the most important factor for customer churn, followed by customer service, service quality and coverage quality [5]. However, social influence is another key driver to customer churn in the mobile services industry [25, 31]. This paper proposes an approach to ABM in order to explore social influence on customer churn behaviour. The next section provides a background of ABM.

3. AGENT BASED MODELLING

Agent based models (ABMs) are composed of agents and a structure for agent based interaction. The ABM technique can be applied to understand any aspect of phenomena. Agents can represent anything from number of patients in a hospital to consumers of a product or service. ABMs are often characterised by rules and these rules define the behaviour of agents in the system. These behaviours are often influenced by agent interactions with other agents in the system, making the outcome difficult to predict. In such cases, a balance may be difficult to reach, making the ability to study the underlying system and the dynamics of the behaviour imperative. ABMs are distinct from traditional modelling approaches where characteristics are often aggregated and manipulated [2].Traditional modelling techniques for modelling may be suitable for their informative purposes but they may not be able to provide adequate level of details in regards to the independent behaviours of agents. ABMs provide sufficient levels of detail in regards to the interdependent behaviors of consumers, retailers, and manufacturers. In addition, ABM is able to sufficiently represent interdependent systems even on a large scale i.e. incorporating a high number of factors, with each factor's level of detail and the behavioral complexity of those factors [23].

Agent based modelling has been applied to various fields including economics [8], health-care [6], and geography [12]. ABM also has linkages with other fields including social sciences, artificial life science, management science and complexity science [21, 20]. ABM benefits from the most recent development in the Artificial Intelligence (AI), and Individual Based Modelling (IBM) in ecology [21]. ABM has been applied to cellular automata to investigate housing segregation patterns by modelling people and the socially relevant process, which represent interactions between people [19]. In business, agent-based modelling has been applied to help decision makers understand underlying market structures and anticipate dynamics in the market place [21]. Simulations can help clarify uncertainty by identifying the causes of the market dynamics. Simulating an Agent based model is known as Agent based Modelling and Simulation (ABMS). The next section discusses social network analysis as a means to provide a background knowledge of how individuals are connected and the connection influences customer retention.

4. SOCIAL NETWORK ANALYSIS

Social network analysis (SNA) is an evolving scientific research area. SNA has developed to be a primary technique for describing the social structure and interaction between network representation. A social network is an inter-connection between nodes using various links. For the purpose of this study, nodes represent customers and the links represent the relationships between customers.
SNA typically focuses on static networks. Static networks are the mapping of relationships between discrete entities. These networks do not change their structure over time. Recently, researchers have focused on dynamic networks that are capable of representing the continuous transmission of information and influence [32]. Dynamic networks are a vital aspect of ABMS. Using dynamic network analysis involves understanding the agent rules that govern network structure and growth, and how networks and their relationships convey information [21]. SNA is an approach to anticipating and modelling society as different sets of people or groups linked to one another [29]. SNA is a method of enquiry that focuses on the relationships between subjects. This approach seeks to understand subjects by collecting information from different sources, analysing the information and visualising the results. SNA has proved to be useful in explaining some important phenomena such as investigating the spread of disease, understanding the Internet and explaining the small word effect to the spreading of information [22]. Sociologists and market researchers believe that the life of an individual depends on how that individual engages with the web of social connections. The social networks term is loosely used to refer to social and professional networking sites including but not restricted to Facebook, LinkedIn and Twitter. Each networking site represents an online community. An online community is an example of a social structure. The people (also known as the nodes) who sign up on these online communities alongside the relationship between them represent social networks. A social network structure consists of nodes and the interrelationship between the nodes. Due to the growing interest of social network analysis, researchers have investigated the principles of the network approach. These principles include [29];
1. Actors and their actions should be viewed as autonomous and independent units rather than as interdependent units.
2. The links between actors should be viewed as channels for transfer of material and non-material resources.
3. Social network models focusing on individuals view the structural environment as a network imposing certain constraints on individual actions.
4. Social network models conceptualise structure (social, economic, political, and so on) as long-lasting patterns of relations among actors.
[22] is an empirical research study of social structure and it introduces 'six degrees of separation' phenomenon while addressing the 'small world problem'. In this study, some participants were chosen at random and asked to deliver a letter

to a target person using only a chain of friends and acquaintances. There was a starting person and a target person in different states in America. The starting person was advised to send the letter across to his friend or acquaintance who is likely to know the target person. The procedure started in Nebraska and the destination of the letter was Boston, Massachusetts. The process will continue until the message gets to the target person. When the letters arrived in Boston, Milgram discovered that it took an average number of six steps for the letter to get to the destination. Hence, Milgram labelled the phenomena six degrees of separation. This study concludes by describing how people of a population are connected. Although [22] study was not subject to an evaluation process, his concept of the small world is widely adopted in social networks research to provide an explanation about how information spreads in the real world. The small-world network has become one of the most widely-used social network models [2]. The CADET approach will be utilised with a network of customers to demonstrate the influence of social networks on customer retention. The next section sheds light on the CADET approach and its link to SNA.

5. THE CADET APPROACH

A number of approaches can be used to describe an agent based modelling process. Researchers have extended agent oriented methodologies in two areas; object oriented (OO) methodologies and knowledge engineering methodologies[14]. The three common views in OO technologies for describing ABMs are static for describing the structure of objects; dynamic for describing object interaction and functional for describing the data flow of the methods of the objects [14]. The flow chat is another popular example for representing the process flow of an agent based model [10].
We present the CADET approach, a novel approach for agent based modelling using results derived from decision tree analysis. The agent attributes are derived from the decision tree flow process. The CADET approach can be applied to various domains and industries. To further explain the structure and functionality of the CADET approach, we present a scenario below;

A Mercedes Benz car dealership provides a number of products and services. In order to understand customer purchase behaviour, a decision tree analysis is conducted. The analysis is carried out to improve understanding on customer purchase behaviour. A decision tree approach is chosen to understand the trend and the pattern for the process of customer decision. The analysis shows that customers who have certain attributes purchase Mercedes Benz cars from the Mercedes Benz car dealership. The variables used for this analysis are age, height, salary, credit score, previous Mercedes Benz purchase, number of services/repairs and satisfaction level. The target variable is either to purchase a Mercedes Benz car or to go purchase another car brand.

The decision tree analysis shows that customers that are over 30 years old, are often interested in purchasing a Mercedes Benz while customers who are less than 30 years old often buy other car brands. The next variable on the decision tree is height. Customers with height less than 170cm often buy

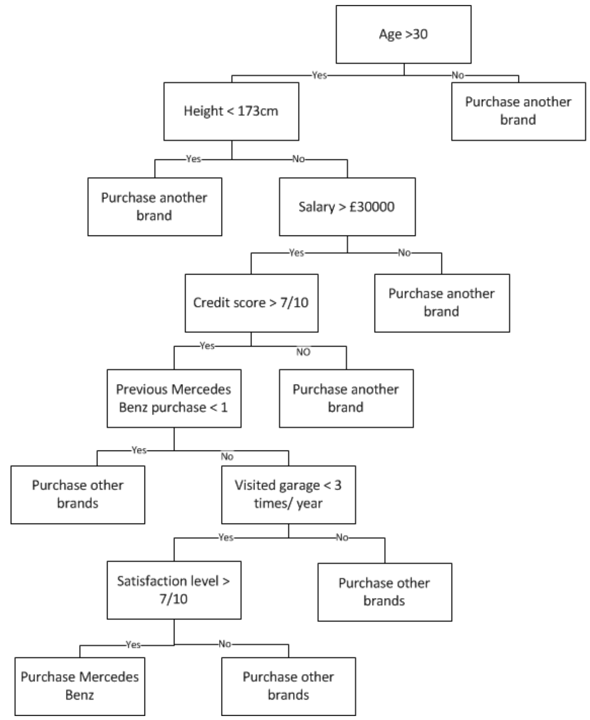

Figure 1. An example of describing an ABM using the CADET approach

other brands while taller customers are often interested in purchasing Mercedes Benz. The next variable on the decision tree is salary. If salary is greater than 30,000, move to the next step, else consider purchasing other brands. The next step is credit score. If credit score is greater than 7, move to the next step otherwise, consider purchasing other brands. This process is followed until the end of the tree. Figure 1 displays the decision tree analysis process described above. The TEA-SIM tool is used to validate the CADET approach. Hence, it is briefly described below.

5.1 TEA SIM Tool

Over the years, companies have spent a large amount of money on products and services that enable them manage and understand their customer behaviour more effectively.

Advertisement and word of mouth can be a powerful tools for customer retention. However, customers are often skeptical about advertisement and may turn to their family or friends within their network to seek advice before deciding whether to purchase or repurchase a product or service. Some companies have manipulated word of mouth by running schemes that offer customers benefits for expressing positivity about their product or service. As positive word of mouth is a great tool for marketing, negative word of mouth can also be a damaging tool for businesses. Dissatisfied customers of a product or service may share their experience about their dissatisfaction with members of their social network which can include family and friends.

The TEA-SIM tool is a data-driven agent based simulation platform [3] built by incorporating a cognitive process for understanding how the members of a small-world network [22]

Name	Description
Contract length	Length of contract
Gender	Customer's gender
Sales channel	Company that delivered contract
Post code	Post code in which customer lives
County Name	The name of county where customer lives
Region	The region where customer lives
Devices	Name and model of device used by customer
Tenure	Number of months with the present mobile service provider
Life stage segment	Customer age
Number of complaints	Number of complaints throughout contact period
Q2_bytes	2nd Quarter Data Usage
Q3_bytes	Third Quarter Data Usage
Q2_Voice	2nd Quarter Voice Usage
Q3_Voice	3rd Quarter Voice Usage
No_of_Repairs	Number of times customer phone has been repaired
Prob_Handset	Number of Times Customer has reported problems with handset

Table 1. Dataset description

make decisions. In addition, the TEA-SIM tool is a decision support tool that can be adopted by various industries to model various entities such as customers, products and services. It also provides a medium for companies to see the interaction process between agents and how the interactions influence agent decisions. The result derived from this process can be used to provide information to organisations so that they can strengthen their customer relationship management (CRM) strategies by exploring the effect of word of mouth to improve customer retention.

The dynamic nature of the TEA-SIM tool provides a unique approach to understanding customer behaviour. It can be used to observe the pattern of customer interaction and perform further analytics to explore the possibilities of incorporating those patterns into the marketing strategies in order to increase revenues. The TEA-SIM tool also works as a generic model that captures the key drivers behind customer change of behaviour and it can also work well in a consultancy environment. It is a cross-industry tool that is not unique to any industry. Experimenting with the TEA-SIM tool proves that it can improve understanding of CRM issues. The TEA-SIM tool is not a precise prediction tool. As a result, agent-based modelling is used to provide insight into the behaviour of a population of customers.

5.2 Applying the CADET Approach to a Real-World Dataset

Models are built for the purpose of mimicking real-world events. The CADET approach for conducting agent based modelling is a novel approach for building agent based models using decision trees. In this study, the decision tree analysis is used to describe customers that either remain with their mobile network operator (MNO) after their contract period and customers that leave their MNO after their contract period. The CADET approach shows that customers decision to stay or leave their MNO is driven by a set of subjective attributes.The attributes are the individual nodes on the decision tree. Mobile network customers may have attributes such as mobile phone type, customer location, price plan and data usage. The decision to churn or stay is composed of a number of phases represented on the decision tree. The final node of the decision tree is customer final decision to stay with or leave a MNO. Decision trees consist of dependent and independent variables. The dependent variables are the nodes that make up the final node. These variables influence customers final decision. The primary purpose of applying the CADET approach for ABMS using the TEA-SIM application is to understand how much influence a customer's environment, family, and friends within their network have on customer retention. In addition, the CADET approach utilised with the TEA-SIM model provides information on the possible decisions a customer might make when they interact with other customers within their network or environment. We believe that if a customer meets another customer within the environment, and they share the same MNO, they may have a conversation on their overall customer experience and that conversation may influence a customers decision to renew their contract. The CADET approach to agent based modelling seeks to provide an understanding on how customer variables along with customer network can influence customer retention.

5.3 Dataset Description

To demonstrate the usability of the CADET approach, we collected a dataset from a UK telecommunications company. The dataset comprises of 19,919 observations and 16 variables. The dependent variable (output variable) for this dataset is whether the customer churned or stayed with their

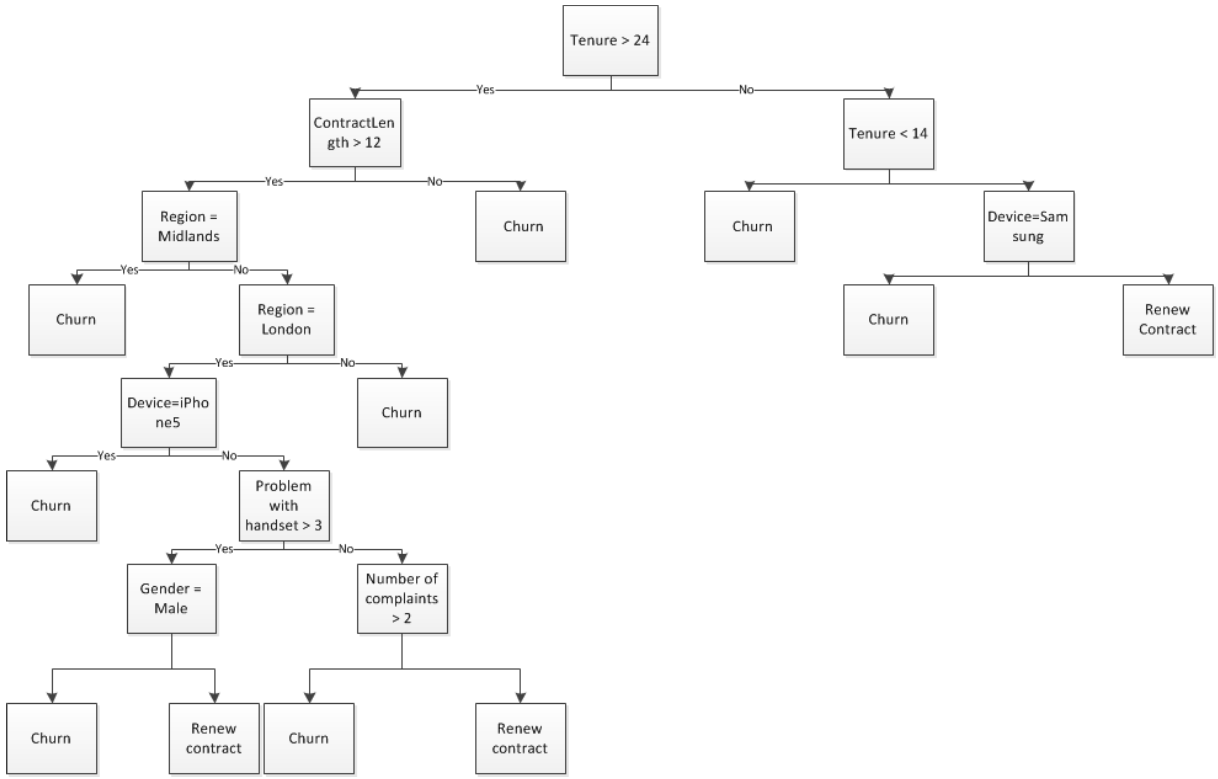

Figure 2. Decision tree analysis

```
<?php

class customers_ChurnCustomers extends Agent {
    function step($step) {
            if (!$this->churn && $this->anyNeighbour (1, 2, array("churn"=>false),
array("churn" => true, "_img" => "img/ChurnCustomer.jpeg"))) {
                    $this->churn = true;
                    $this->_img = "img/ChurnCustomer.jpeg";
            }
            $this->move(1);
            $this->morph(2);

    }
}

class customers_StayCustomers extends Agent {
    function step($step) {
        $this->move(1);
            }
        }
```

Figure 3. Stepper function for dataset

mobile Network operator after their contract. The predictor variables (input variables) are customers' data (such as type of device, price plan and region). The dataset contains 50% of customers who churned and 50% of customers who stayed with the MNO until the end of their contract. The dataset is based on a 24 month contract. Some customers stayed with the MNO after the end of the 24 month period (i.e they renewed their contract) while other customers left the MNO after the 24 month contract. The dataset comprises of different data types as a means to represent the entire customer base. Table 1 presents a tabular description of the dataset.

5.4 Model Structure & Experiment

To apply the CADET approach with the TEA-SIM tool, the CART decision tree algorithm is run on the dataset described above and the result of the decision tree is visualised. From the decision tree (see figure 2), we can see the flow of the decision process for customers who have decided to stay with their MNO or move to a different MNO. The top of decision tree analysis shows that a customer's tenure is either greater than 24 or not. If tenure is greater than 24 then move to the next variable on the left. However, if tenure is not greater than 24, then move to the next variable on the right-side of the tree. This process continues until the end of the tree. The end of the tree displays the end result of the process i.e. either customer churned or renewed their contract at the end of the 24-month contract period. Figure 2 displays the decision tree analysis diagram.

To simulate this process, agents are fed into the Tea-Sim model with the attributes used for the decision tree analysis. Figure 3 presents the description of agent behaviour specification. Also, it shows how each agent transitions from one phase to another i.e how stay customers transition to be churn customers and vice versa (stay customers are represented with a smiley face, while churn customers are represented with an angry crying face (see figure 4)). This is based on the belief that customers interact with one another in a given environment and customer interaction with other customers may influence their decision to stay with their MNO or to leave their MNO. Figure 4 displays the process in which the agents interact on the TEA-SIM model. Agents move from one grid element to another, if an agent meets another agent of another type, it is likely to morph into the other type of agent as described in the stepper function displayed in figure 3. This

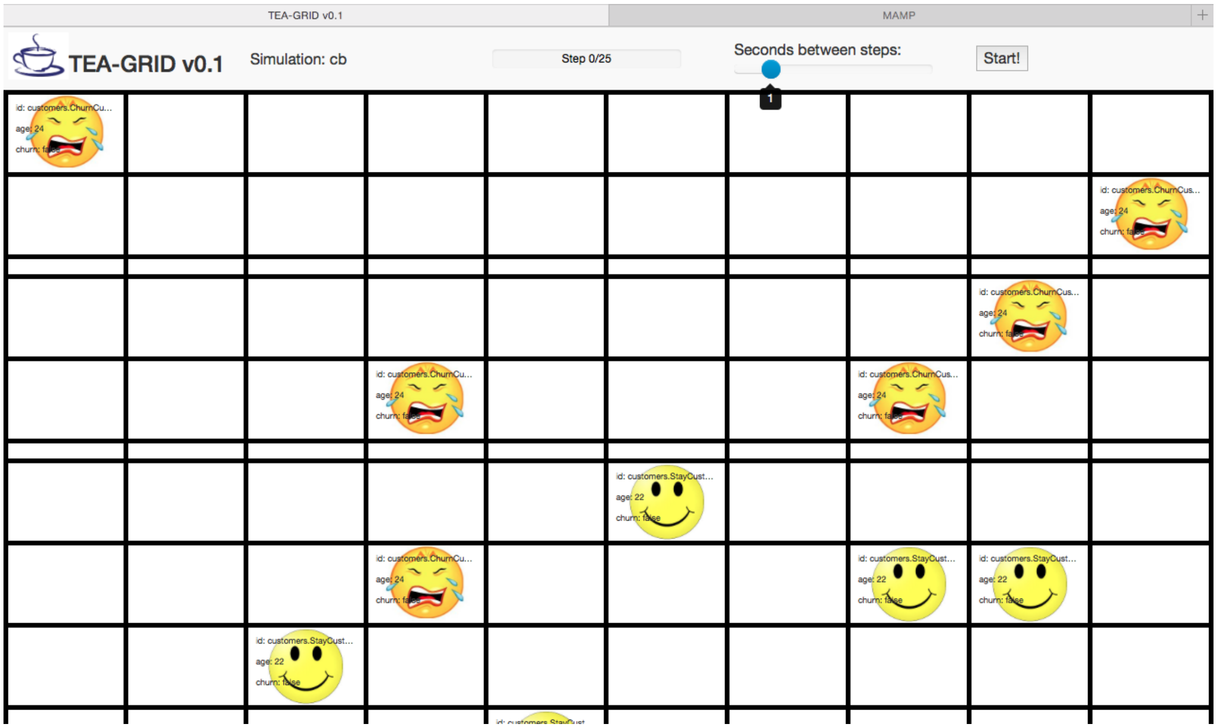

Figure 4. Agent interaction process

indicates that if a churn customer meets a stay customer and they have a conversation, the churn customer may decide to renew their contract and vice versa.

6. CONCLUSION

Numerous factors can influence the decision for customers to purchase or adopt a product or service. However, customers are likely to trust the word of mouth of someone in their network. This paper introduced a novel approach to agent-based modelling with the use of a novel tool, the TEA-SIM tool.

The TEA-SIM tool provides a generic approach for companies across industries to better understand their customers. The tool can help decision makers in organisations work on better strategies to understand customer behaviour and enhance customer retention, and subsequently improve CRM.

This paper also presents a real-life example of implementing the CADET approach using the TEA-SIM tool. The CADET approach was used to effectively provide an Agent-Based model using the results derived from decision trees. In this study, the CADET approach was applied to a dataset from a telecommunications company. However, the concept of the CADET approach can be extended and implemented in other industries such as health-care, manufacturing and financial services.

REFERENCES

1. Baumann, C., Elliott, G., and Burton, S. Modeling customer satisfaction and loyalty: survey data versus data mining. *Journal of services marketing 26*, 3 (2012), 148–157.
2. Baxter, N., Collings, D., and Adjali, I. Agent-based modelling– intelligent customer relationship management. *BT Technology Journal 21*, 2 (2003), 126–132.
3. Bell, D., Kashefi, A., Saleh, N., and Turchi, T. A data-driven agent based simulation platform for early health economics device evaluation. *to appear in Proceedings of the SpringSim'16* (2016).
4. Blodgett, J. G., Wakefield, K. L., and Barnes, J. H. The effects of customer service on consumer complaining behavior. *Journal of Services Marketing 9*, 4 (1995), 31–42.
5. Chu, B.-H., Tsai, M.-S., and Ho, C.-S. Toward a hybrid data mining model for customer retention. *Knowledge-Based Systems 20*, 8 (2007), 703–718.
6. Epstein, J. M. Modelling to contain pandemics. *Nature 460*, 7256 (2009), 687–687.
7. Epstein, M. J., and Westbrook, R. A. Linking actions to profits in strategic decision making. *Image* (2012).
8. Farmer, J. D., and Foley, D. The economy needs agent-based modelling. *Nature 460*, 7256 (2009), 685–686.
9. Ganesh, J., Arnold, M. J., and Reynolds, K. E. Understanding the customer base of service providers: an examination of the differences between switchers and stayers. *Journal of marketing 64*, 3 (2000), 65–87.
10. Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz,

S. K., Huse, G., et al. A standard protocol for describing individual-based and agent-based models. *Ecological modelling 198*, 1 (2006), 115–126.

11. Hanif, M., Hafeez, S., and Riaz, A. Factors affecting customer satisfaction. *International Research Journal of Finance and Economics 60* (2010), 44–52.
12. Heppenstall, A. J., Crooks, A. T., See, L. M., and Batty, M. *Agent-based models of geographical systems*. Springer Science & Business Media, 2011.
13. Hughes, A. M. *Strategic database marketing*. McGraw-Hill New York, 2006.
14. Iglesias, C. A., Garijo, M., and González, J. C. A survey of agent-oriented methodologies. In *Intelligent Agents V: Agents Theories, Architectures, and Languages*. Springer, 1999, 317–330.
15. Itzkowitz, J. Customers and cash: How relationships affect suppliers' cash holdings. *Journal of Corporate Finance 19* (2013), 159–180.
16. Jones, M. A., Taylor, V. A., and Reynolds, K. E. The effect of requests for positive evaluations on customer satisfaction ratings. *Psychology & Marketing 31*, 3 (2014), 161–170.
17. Keramati, A., and Ardabili, S. M. Churn analysis for an iranian mobile operator. *Telecommunications Policy 35*, 4 (2011), 344–356.
18. Khan, A. A., Jamwal, S., and Sepehri, M. Applying data mining to customer churn prediction in an internet service provider. *International Journal of Computer Applications 9*, 7 (2010), 8–14.
19. Li, X., Mao, W., Zeng, D., and Wang, F.-Y. Agent-based social simulation and modeling in social computing. In *Intelligence and Security Informatics*. Springer, 2008, 401–412.
20. Ma, T., and Nakamori, Y. Agent-based modeling on technological innovation as an evolutionary process. *European Journal of Operational Research 166*, 3 (2005), 741–755.
21. Macal, C. M., and North, M. J. Tutorial on agent-based modelling and simulation. *Journal of simulation 4*, 3 (2010), 151–162.
22. Milgram, S. The small world problem. *Psychology today 2*, 1 (1967), 60–67.
23. North, M. J., Macal, C. M., Aubin, J. S., Thimmapuram, P., Bragen, M., Hahn, J., Karr, J., Brigham, N., Lacy, M. E., and Hampton, D. Multiscale agent-based consumer market modeling. *Complexity 15*, 5 (2010), 37–47.
24. Ozcan, A. Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools. *Lab on a chip 14*, 17 (2014), 3187–3194.
25. Phadke, C., Uzunalioglu, H., Mendiratta, V. B., Kushnir, D., and Doran, D. Prediction of subscriber churn using social network analysis. *Bell Labs Technical Journal 17*, 4 (2013), 63–75.
26. Rust, R. T., and Zahorik, A. J. Customer satisfaction, customer retention, and market share. *Journal of retailing 69*, 2 (1993), 193–215.
27. Silva, K., and Yapa, S. Customer retention: with special reference to telecommunication industry in sri lanka.
28. Svendsen, G. B., and Prebensen, N. K. The effect of brand on churn in the telecommunications sector. *European Journal of Marketing 47*, 8 (2013), 1177–1189.
29. Vag, A. Simulating changing consumer preferences: a dynamic conjoint model. *Journal of business research 60*, 8 (2007), 904–911.
30. Verbeke, W., Dejaeger, K., Martens, D., Hur, J., and Baesens, B. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research 218*, 1 (2012), 211–229.
31. Verbeke, W., Martens, D., and Baesens, B. Social network analysis for customer churn prediction. *Applied Soft Computing 14* (2014), 431–446.
32. Watts, D. J. The" new" science of networks. *Annual review of sociology* (2004), 243–270.