Agent Based Modelling and Simulation: An Examination of Customer Retention in the UK Mobile Market

A thesis submitted for the degree of Doctor of Philosophy

By

Mohammed Bassam Hassouna

Department of Information Systems and Computing,

Brunel University

March 2012

ABSTRACT

Customer retention is an important issue for any business, especially in mature markets such as the UK mobile market where new customers can only be acquired from competitors. Different methods and techniques have been used to investigate customer retention including statistical methods and data mining. However, due to the increasing complexity of the mobile market, the effectiveness of these techniques is questionable.

This study proposes Agent-Based Modelling and Simulation (ABMS) as a novel approach to investigate customer retention. ABMS is an emerging means of simulating behaviour and examining behavioural consequences. In outline, agents represent customers and agent relationships represent processes of agent interaction. This study follows the design science paradigm to build and evaluate a generic, reusable, agent-based (CubSim) model to examine the factors affecting customer retention based on data extracted from a UK mobile operator. Based on these data, two data mining models are built to gain a better understanding of the problem domain and to identify the main limitations of data mining. This is followed by two interrelated development cycles: (1) Build the CubSim model, starting with modelling customer interaction with the market, including interaction with the service provider and other competing operators in the market; and (2) Extend the CubSim model by incorporating interaction among customers.

The key contribution of this study lies in using ABMS to identify and model the key factors that affect customer retention simultaneously and jointly. In this manner, the CubSim model is better suited to account for the dynamics of customer churn behaviour in the UK mobile market than all other existing models. Another important contribution of this study is that it provides an empirical, actionable insight on customer retention. In particular, and most interestingly, the experimental results show that applying a mixed customer retention strategy targeting both high value customers and customers with a large personal network outperforms the traditional customer retention strategies, which focuses only on the customer's value.

ABBREVIATIONS

ABMS: Agent-Based Modelling and Simulation

AUC: Area under the ROC Curve

BDI: Belief-Desire-Intention

CART: Classification and Regression Trees

CDR: Call Detail Record

CA: Cellular Automata

CHAID: Chi-Squared Automatic Interaction Detection

CAS: Complex Adaptive Systems

CLV: Customer Lifetime Value

CRISP-DM: Cross-Industry Standard Process-Data Mining

CRM: Customer Relation Management

CubSim: Customer Behaviour Simulation Model

DSR: Design Science Research

GUI: Graphical User Interface

KPI: Key performance indicator

MNOs: Mobile Network Operators

MVNOs: Virtual Mobile Operators

Ofcom: Office of Communications

PAYG: Pay-As-You-Go

PDF: Probability Density Function

ROC: Receiver Operating Characteristic

SNA: Social Network Analysis

UML: Unified Modelling Language

WOM: Word-of-Mouth

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEDICATIONS

I dedicate this work to God almighty, to whom all glory shall always be, for his grace and strength that helped me to accomplish this work. I also dedicate this work to my parents, for making me who I am today. To my dad, this work is dedicated to your memory. To my mum, I am very grateful to your prayers and unlimited support.

It is my pleasure to dedicate this work to my beloved wife, Mona, for her love, patience and understanding and to my precious daughter and son, Sondos and Bassam, the beauties of my life. In addition, I would like to express my gratitude to my grandmother, brothers and sisters for their encouragements and inspirations.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest gratitude to my principal supervisor, Professor Mark Lycett, for his valuable support, guidance, insight, professionalism and patience. This work would never have been accomplished without your help and support. Thanks must go to Dr. David Bell, my second supervisor, for his valuable advice and support. Also, I am very grateful to Dr. Sergio de Cesare for his kind support. Special thanks are extended to Professor Ray Paul for his kindness, empathy and inspiration, which motivated me to attain the best of my abilities.

This work would not be possible without the help of all the family, friends and colleagues who encouraged and supported me throughout my PhD journey. My warmest thanks go to Dr. Michael Anaman for his valuable insights and assistance. I would like to extend my thanks to my colleagues Mahir Arzoky and Firas Omar for their help and support.

Finally, I am very grateful to the Department of Information Systems and Computing (DISC) for funding my PhD study. I am also very thankful for the stimulating environment and the excellent resources that helped me to realise my potential.

DECLARATION

The following papers have been published as a result of the research conducted in this thesis.

Hassouna, M. and Arzoky, M. (2011) “Agent based Modeling and Simulation: Toward a New Model of Customer Retention in the Mobile Market”, *Proceedings of the 2011 Summer Computer Simulation Conference*, Netherlands, pp. 30-36.

Hassouna, M. and Omar, F. (2011) “Modelling Customer Churn in the Mobile Market - A Comparative Study of Decision Tree and Logistic Regression”, *Proceedings of the First Global Conference on Communication, Science & Information Engineering CCSIE 2011*, Middlesex University, UK, London.

Chapter 1: Introduction

1.1 Overview

This introductory chapter sets the scene for the whole thesis, which investigates customer retention in the UK mobile market. After highlighting the importance of customer retention as a field of study, the aim and objectives of the thesis are detailed. The research approach utilised to achieve the research aim and objectives is then explained, followed by an overview of the thesis chapters. To guide the reader through the thesis, a simple diagram is presented to illustrate graphically the structure of the thesis (see Figure 1.1).

This chapter is structured as follows: Section 1.2 provides background information on the research problem and discusses the motivation for this study. Section 1.3 defines the research aim and objectives based on the research problem and motivation. Section 1.4 describes the research approach adopted in the study. Finally, Section 1.5 presents the thesis structure and shows a diagram that summarises the contents of the thesis.

1.2 Background and Motivation

Customer retention is an important area of concern in the field of Customer Relationship Management (CRM). It mainly focuses on building and managing loyal, profitable and long-lasting customer relationships. Customer retention strategies are practised by different businesses not only to gain a competitive advantage but also to survive. Developing successful retention strategies is crucial for businesses in general, particularly mobile operators. Customer retention is an immediate issue in the mobile market because of the high customer churn rate (subscribers leaving the market or switching to other networks) and the high cost of acquiring new customers (Clarke, 2001; Seo et al., 2008).

The UK mobile market is one of the largest in Europe. According to Ofcom's latest report on the Communications Market, the total number of mobile subscribers in the UK market surpassed 81.1 million at the end of 2010, and the penetration rate was 130.1% - a figure expected to reach 150% by 2016 (Ofcom, 2011a, p.245; ITProPortal, 2011). The UK mobile market is mature and highly competitive because of the increased number of network operators and the intense competition between these operators. There are four mobile network operators (MNOs) offering mobile services in the UK mobile market. In addition, there are more than 100 virtual mobile operators (MVNOs), which buy airtime from MNOs and resell it, under different brands (Ofcom, 2011b, p.60).

McIlroy and Barnett (2000) revealed that attracting a new customer is five times more costly than keeping an existing one. Attracting new customers entails costs of advertising, setting up new accounts, educating customers and other costs not included in the case of retaining existing customers. Aydin and Ozer (2005) stated that improving customer retention and reducing churn rate from 20% to 10% brought about £25 million of annual savings to the Orange mobile operator.

Moreover, the UK mobile market has exhibited exponential growth for the past few years in spite of the economic conditions and downturns and entered the market saturation or mature stage (Ofcom, 2011a, p.245). As the UK mobile market reaches maturity and mobile services are increasingly offered on a subscription basis, the basis of competition has shifted from acquiring new subscribers to retaining existing customers and leveraging customer relationships by increasing their loyalty and decreasing churn. Although mobile operators are making concerted efforts to reduce churn rates (Clarke, 2001; Seo et al., 2008), they lose about 20% to 40% of their subscribers every year (Lee, Lee and Feick, 2001; Ahn et al., 2006; Seo et al., 2008). In the UK market, this rate increased from 33.4% in 2005 to 38.6% in 2007, and it is expected to continue increasing (Mobile Marketing Magazine, 2009).

As the mobile market becomes more competitive and customer expectations increase, understanding customer behaviour and needs is a key element of mobile operators' success. Customer churn analysis is one of the most pressing issues in the mobile market (Seo et al., 2008). Different methods and techniques have been used to investigate customer churn including statistical methods and data mining. Due to the increasing complexity of the mobile market, the appropriateness of using traditional statistical and data mining techniques to analyse the mobile market and customer retention are questionable (Twomey and Cadman, 2002).

Many of the published studies in customer churn in the mobile industry have only focussed on customers' characteristics and interactions with operators, while ignoring interactions with the mobile market and interactions among customers themselves. In addition, most of the research until now is predictive in nature and did not offer answers to the question of why a customer might churn. Accordingly, the importance of customer retention as a field of study in this thesis is supported by many factors, including:

- High customer churn rate
- Evolving customer needs and behaviour
- Globalisation and technological developments
- Increasing national and international competition
- Mature markets
- Inappropriateness of traditional retention analysis tools

In order to overcome the limitations of traditional customer retention analysis tools, this study employs Agent-Based Modelling and Simulation (ABMS) as proposed by Twomey and Cadman (2002) to capture the complexity of customer interactions in the UK mobile market. ABMS is a computational model for simulating the actions and interactions of autonomous individuals in a network, with a view to assessing their effects on the system as a whole. It involves creating artificial agents mimicking the attributes and behaviours of their real-world counterparts (Kyrylov et al., 2004). ABMS is composed of two main pillars: Modelling and simulation. Modelling, simply, is abstracting real-world phenomena into a representation or model, whilst simulation is a process of executing models over time to mimic and imitate real or proposed situations or systems. The potential of ABMS to overcome the limitations of traditional analysis tools is highlighted by several examples in the literature that model customer behaviour in different consumer markets (e.g. Brannon et al., 2000; Kuhn Jr et al., 2010; North et al., 2010).

1.3 Research Aim and Objectives

This study attempts to overcome the limitations of the traditional tool for retention analysis by exploiting the potential of ABMS to provide a holistic view for customer retention. Accordingly, the aim of this research is to answer the following question:

Can ABMS be used to provide insights into customer retention in the UK mobile market? If yes, how and what type of insights can it provide?

To fulfil the aim of this study, an illustrative case study of a mobile operator in the UK mobile market is used. In order to address the research question, the following research objectives were established.

Objective 1: Evaluate the current and most common analysis tools for customer retention to highlight their capabilities and limitations.

Objective 2: Develop an agent-based model for customer interaction with the mobile market, including interaction with the service provider and other competing operators in the market, to address the limitations of traditional retention analysis tools.

Objective 3: Extend the agent-based model to incorporate the interaction among customers in order to evaluate the influences of social networks on customer churn.

Objective 4: Perform a set of simulation experiments and analyse its results to establish the validity of the agent-based model.

Objective 5: Demonstrate the utility of the agent-based model by providing actionable insights into customer retention.

1.4 Research Approach

As this thesis seeks to generate knowledge and insights into customer retention by building simulations and artefacts, Design Science Research (DSR) paradigm is deemed appropriate (Purao, 2002). Therefore, Design Science Research (DSR) paradigm is employed in this study as a general methodological framework (Hevner et al., 2004). The design process in the DSR is different from other design activities that focus on building usable artefacts intended to accomplish goals. DSR is distinct from other design activities in that it involves creating, capturing and communicating the knowledge acquired throughout the design process (Vaishnavi and Kuechler, 2004).

To accomplish the aim and objectives of this study, four successive phases were incorporated into the research design and implementation process. These phases are: (1) Awareness of the problem; (2) solutions selection and suggestion; (3) development; and (4) evaluation. Phase 1 explores the business problem addressed in this study and the wider environment in which the problem is set. Phase 2 examines in more detail the traditional analysis tools that are commonly used in customer retention analysis. This phase aims at defining the limitations of the traditional analysis tools and suggesting ways to overcome these limitations. It involves one development cycle resulting in two data mining models (decision tree and logistic regression). The main development phase (phase 3) consists of two iterations: (1) Modelling customer interaction with the market; and (2) modelling interactions among customers. Phase 4 provides an evaluation and reflection of the research results and outcomes. To ensure continuous improvement of the artefacts designed in this study, an iterative/incremental development process was followed. In this process, the knowledge acquired through the design and implementation of each iteration is transferred to the next iteration. Chapter 3 provides further explanation of the phases and iterations that comprise this study.

1.5 Thesis Structure

This thesis is organised around seven chapters as follows:

Chapter 1 provides an introduction and an overview of this research. It highlights the importance of customer retention and underlines the need for this study. This chapter also presents the research aim and objectives. Further, it offers a brief overview of the research approach adopted to accomplish the aim and objectives of this study. Finally, this chapter concludes by outlining the thesis structure and providing a summary of each chapter.

Chapter 2 provides a literature review in which two research areas related to this study are discussed: Customer retention and modelling customer behaviour. This chapter includes four main sections. The first section introduces the concept of customer retention and highlights the effects of customer satisfaction on it. The second section narrows the focus to customer retention in the mobile market and explores the key drivers of customer churn in this market. In section three, a critical evaluation of literature on empirical studies of customer retention in the mobile market is presented. This section also underlines the main limitations in the reviewed literature. In addition, it identifies the commonly used techniques for customer retention analysis. The fourth section explores the potential of ABMS to overcome the limitations of previous studies by discussing some of its applications in different consumer markets.

Chapter 3 describes the research approach adopted to conduct this study. As noted before, the DSR paradigm is employed in this study as a general methodological framework. Before explaining the methodology employed in this study, a theoretical overview of the research paradigms in information system is presented. The research methods and techniques used in this study are then explained, along with a justification of why these techniques were selected. After that, the case study used as the main data source for this study is introduced. The chapter concludes by explaining the phases and tasks performed to fulfil the aim and objectives of this study.

Chapter 4 presents an empirical evaluation of two data mining tools commonly used to analyse customer retention: Decision tree and logistic regression. During the analysis and design in this phase, further insights are acquired into the problem domain. In addition, the specifications of the required solutions are defined. Before detailing the experiments performed in this chapter, a theoretical background on the decision tree and logistic regression modelling techniques is presented. In order to establish quality benchmarks for the different data mining models developed in this study, some of the performance evaluation metrics for these models are discussed and explained. Then, the modelling experiments and results are introduced along with the evaluation of the models and their limitations. The modelling experiments performed in this chapter were based on two data sets gained from the case study presented in Section 3.3.5. The empirical evaluation conducted in this chapter not only provides motivation for the following two chapters but also identifies the general requirements and specification for the ABMS model.

Chapter 5 proposes an agent-based conceptual model for the customer churn behaviour in the mobile market. This model is named “Customer Behaviour Simulation Model” (CubSim). In order to demonstrate the validity of the conceptual model, the design elements identified in the conceptual model are implemented and executed. This chapter focuses on implementing the design elements related to customers interaction with the mobile market. Additionally, the design and implantation process of the customer agent is also explained in this chapter, including agent attributes, agent architecture and behavioural rules. The last sections of this chapter are devoted to describing the verification, validation and calibration processes along with the simulation experiments that were performed on the CubSim model (Version 2). The calibration process is conducted based on the same data sets used to develop the data mining models in Chapter 4.

Chapter 6 expands on Chapter 5 by extending the CubSim model to incorporate the interaction among customers. Before discussing the empirical model extension process, a theoretical background on social networks is given. This is followed by a detailed explanation and demonstration of the extension process that includes modelling the social network, modelling the word-of-mouth flow and extending the agent behaviour to incorporate social influences. Similar to Chapter 5, the verification, validation and calibration processes performed in this chapter are outlined and discussed. After that, the results of the simulation experiments performed on the CubSim model (Version 4) are presented and interpreted. The simulation experiments are aimed at: (1) Establishing the validity of the model by performing a black-box validation experiment; and (2) establishing the utility of the model by demonstrating empirically how insights can be extracted from the CubSim model. Lastly, this chapter provides an evaluation of the model including its limitations and uses.

Chapter 7 presents the research conclusions and findings. These findings are divided into four logical parts, corresponding to the research phases. It also provides a brief dissection of how this research meets its defined objectives. Finally, it highlights the key contributions made by this study and discusses its limitations in order to draw up future research directions.

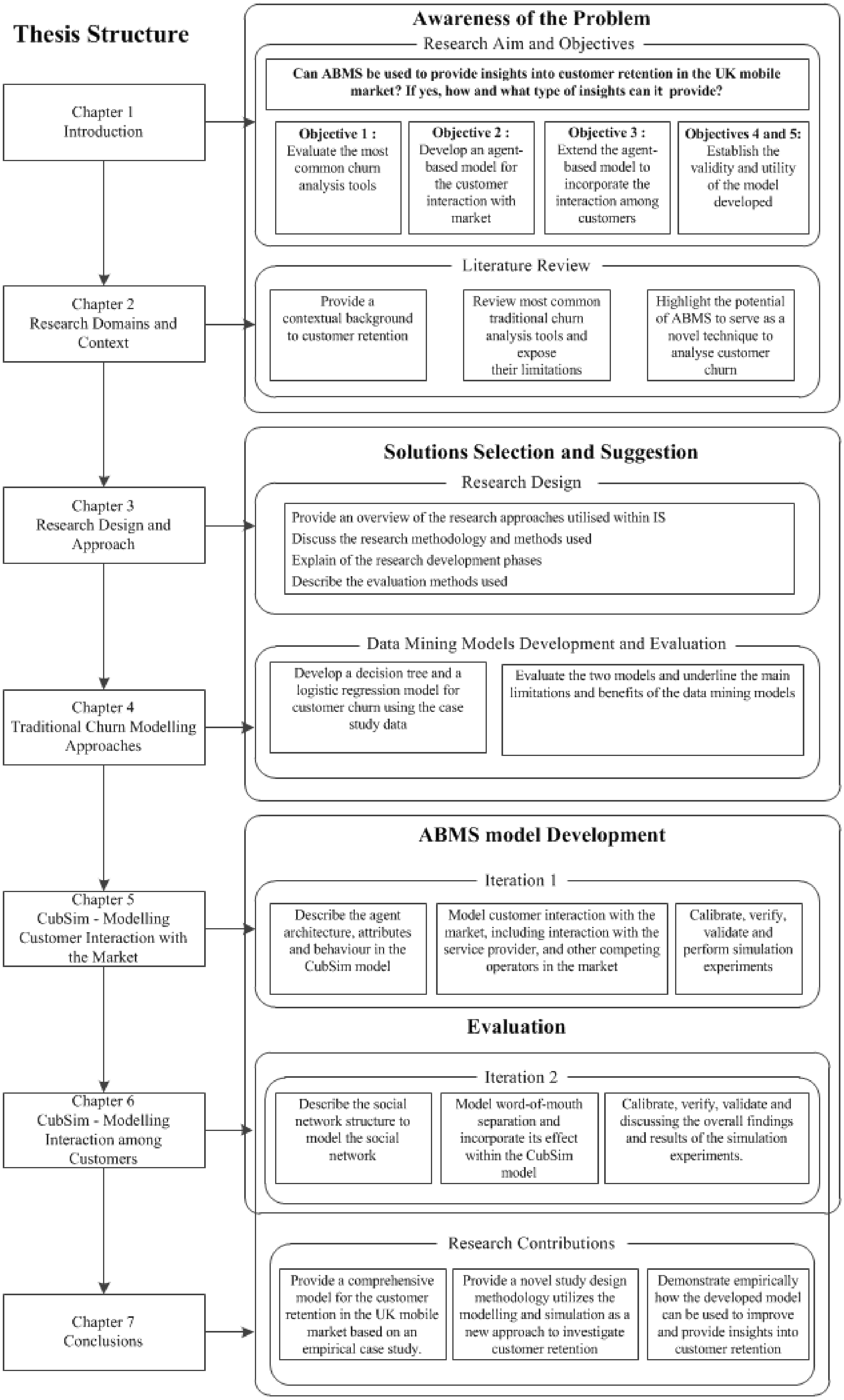


Figure 1.1: Overview of the Thesis

Chapter 2: Research Domains and Context

2.1 Overview

This chapter explores state-of-the-art of customer retention in the mobile market. The aims of this chapter are to: (1) Provide a contextual background to customer retention in the mobile market; (2) critically review traditional churn analysis techniques and expose their limitations; and (3) highlight the potential of ABMS to serve as a novel technique to analyse customer churn.

This chapter is structured as follows: Section 2.2 provides background information on CRM and customer retention. Section 2.3 narrows the focus to the customer retention in the mobile market. Section 2.4 reviews existing literature on modelling customer behaviour in the mobile market by using both traditional and ABMS techniques. Section 2.5 underlines some of the significant limitations of previous studies. Finally, Section 2.6 provides an overall summary of the chapter.

2.2 Customer Relationship Management

Customers are at the heart of today's business, especially in the service industries. Business prospects are dependent on satisfying customers and developing their relationships with the organisation. To build and manage loyal, profitable and long-lasting customer relationships businesses need to rationalise CRM strategies. Chen and Popovich (2003) defined CRM as a combination of people, technology and processes that seek to understand a company's customers. Furness (2001) saw it as a process geared towards increasing the value of the customers over their lifetime.

CRM is not a one-time effort; it is a systematic and continuous process. This process encompasses three core customer management processes: (1) Customer acquisition; (2) customer development; and (3) customer retention (Kamakura et al., 2005). In broad outline, acquisition is the process of obtaining more and profitable customers through different channels such as direct marketing. Once the customers have been acquired, customer development strategies are utilised to maximise the potential revenue from existing customers and new prospects. Cross-selling and upselling are some of the activities undertaken in customer development. After customer acquisition and development, customer retention is the key to long-term profitability and can be achieved by maintaining continuous and active relationships with customers.

Managing customer retention is a critical issue in the mobile market. Mobile operator companies lose about 20 % to 40% of their subscribers every year (Lee, et al., 2001; Ahn et al., 2006; Seo et al., 2008). Moreover, the mobile telecommunication industry faces combined difficulties such as increasing national and international competition, slower growth rates and mature markets. Furthermore, the changes that emerge because of deregulation, globalisation and technological developments also contribute to hinder the ability of business to retain their customers. The next section discusses customer retention in more detail.

2.2.1 Customer Retention

Customer retention simply means maintaining a continuous and active relationship with customers. Hoyer and Macinnis (2009, p.289) defined it as “the practice of working to satisfy customers with the intention of developing long-term relationships with them”. These relationships can be achieved by: (1) Subsequent purchases; (2) extending the customers’ contract with the service provider over a specified period; (3) the customers’ intention to make future purchases from the provider; or (4) inducing customers to refrain from terminating their contracts (Gerpott et al., 2001). For the purpose of this study, customer retention is defined as all organisational plans and actions to retain existing customers and potential customers by developing, maintaining and maximising mutually beneficial long-term relationships.

Business transformations that occurred because of deregulation, globalisation and technological developments have changed the business focus from being product-focussed to being customer-focused. In addition to the increased profit that customer retention strategies create, these strategies also minimise the cost of selling new products or services because it is easier to sell to existing customers than new ones (Aydin and Ozer, 2005). New customers are more difficult to find, and they often pay less than existing customers do, and are typically more price-sensitive (Stahl et al., 2003). Besides the obvious direct benefits that can be realised by adopting customer retention strategies, there are many indirect benefits. One example of these indirect benefits is attaining valuable customer feedback that can be used to improve business operations. Word-of-mouth (WOM) is another example of indirect benefits of the customer retention. It is one of the best forms of business advertising, which can be exploited to influence customers' attitudes and behaviours (Buckinx and Van den Poel, 2005).

Customer churn is an opposite construct to customer retention, and refers to the cases in which an operator's customers leave it for one of its competitors. The two terms will be used interchangeably in this study, as they are fundamentally two sides of the same coin. Customer churn can be classified into two types: Voluntary and non-voluntary churn. Although businesses have serious concerns about both types, much less attention has been paid to the latter, mainly because it is easier to spot than the voluntary churn. Fraud and customers with credit problems are some examples of non-voluntary churn. In fraud cases, businesses use real time fraud detection systems to evaluate customer transactions as they occur, in order to identify unauthorised activities. Credit problems are the easiest to be identified because they occur when customers fail to pay their bills for several months.

Voluntary churn occurs when the customers initiate termination of the relationship with the service provider. It can be divided further into two types: Incidental and deliberate churn. Incidental churn happens when the customers circumstances change in a way that prevents them from extending their relationship with the service provider. Financial circumstances and moving to new places provide example reasons of incidental churn. In contrast, deliberate churn occurs when customers decide to switch to a competing company, for reasons such as dissatisfaction with the service provided. This service encompasses many dimensions like technology and quality. Customers switching service provider due to seeking new technologies that their existing supplier does not provide is one example of technology-based churn. Another example of deliberate churn is that due to quality factors, like poor coverage.

Unlike incidental churn, which cannot be controlled or managed, deliberate churn is the problem that most churn solutions seek to identify and manage. Ahn et al. (2006) stated that understanding exactly why customers defect is a research gap. Therefore, this study focuses on the voluntary churn and more specifically deliberate churn, and why it happens. Relationship quality is a key determinant of customer retention (Ulaga and Eggert, 2006). Customer satisfaction is an important measure of the quality of the relationship between customers and business. The following section introduces customer satisfaction and investigates its effect on customer retention.

2.2.2 Customer Satisfaction and Customer Retention

Many researchers have linked customer retention with customer satisfaction and loyalty (e.g. Gerpott et al., 2001; Lee et al., 2001; Turel and Serenko, 2006). Customer satisfaction encompasses many dimensions, including products and services, customer services and service quality to name a few. Ranaweera and Prabhu (2003, p.337) defined customer satisfaction as “an evaluation of an emotion, reflecting the degree to which the customer believes the services provider evokes positive feelings”. Chen (2001) provided a more precise definition as “a post-consumption evaluation that a chosen alternative at least meets or exceeds expectations”. In simple words, customer satisfaction can be defined as the difference between what the customers think they should be getting and what they actually do get. The relationship between customer satisfaction and customer retention has been a matter of intense debate within the marketing literature (Lee et al., 2001). Three major views have been suggested concerning the relationship between customer satisfaction and customer retention:

- First, customer satisfaction is a key determinant of customer retention (Day, 1994). Aydin and Ozer (2005) supported this view by examining the relationships between the different dimensions of customer satisfaction and customer retention. They pointed out that satisfied customers stay loyal longer, are less price-sensitive and pay less attention to competitors’ advertising, and spread positive WOM.

- Second, customer satisfaction has no direct effect on customer retention (Tikkanen and Alajoutsijärvi, 2002). This view is reinforced by Reichheld (1993), who pointed out that sometimes satisfied customers leave their provider while unsatisfied customers do not. In the same line, Reichheld (1996) warned against what he called a ‘satisfaction trap’; research indicates that 60-80% of customers who appear to be satisfied still defect to competitors.

- Third, although customer satisfaction may positively influence customer retention, as some research indicates, customer satisfaction alone is not sufficient to explain customer retention (Clarke, 2001). Seo et al. (2008) maintained this view, and asserted that in competitive markets such as the mobile market, service providers need to look beyond mere basic satisfaction to employ ways of establishing, maintaining and strengthening ties of loyalty with customers. This study is consistent with this view, and examines it by investigating customer satisfaction along with other factors that might contribute to customer retention.

2.3 Mobile Market and Customer Churn

2.3.1 UK Mobile Market and Customer Churn

The UK mobile market is one of the largest in Europe. According to Ofcom's latest report on the Communications Market, the total number of mobile subscribers in the UK market surpassed 81.1 million at the end of 2010, and the penetration rate was 130.1% - a figure expected to reach 150% by 2016 (Ofcom, 2011a, p.245; ITProPortal, 2011). The UK mobile market is mature and highly competitive because of the increased number of network operators and the intense competition between these operators. There are four mobile network operators (MNOs) offering mobile services in the UK mobile market. In addition, there are more than 100 virtual mobile operators (MVNOs), which buy airtime from the MNOs and resell it under different brands (Ofcom, 2011b, p.60). Moreover, the UK mobile market has exhibited exponential growth for the past few years in spite of the economic conditions and downturns and entered the market saturation or mature stage (Ofcom, 2011a, p.245).

It has been shown that customer retention is a critical issue in the mobile market for reasons such as the high rate of churn and the high cost of acquiring new customers (Clarke, 2001; Seo et al., 2008). McIlroy and Barnett (2000) observed that attracting new customers is five times more costly than keeping existing ones. Attracting new customers entails the costs of advertising, setting up new accounts, educating customers and other costs not included in the case of retaining existence customers. Aydin and Ozer (2005) observed that improving customer retention and reducing churn rate from 20% to 10% brought about £25 million of annual savings to the Orange mobile operator. Moreover, a study by Seo et al. (2008) revealed that retaining customers is more profitable than increasing market share or decreasing costs.

As the UK mobile market reaches maturity and mobile services are increasingly offered on a subscription basis, the basis of competition has shifted from acquiring new subscribers to retaining existing ones and leveraging customer relationships by increasing their loyalty and decrease churn. Although mobile operators are making concerted efforts to reduce churn rates (Clarke, 2001; Seo et al., 2008), recent research shows that customer churn rates within the UK mobile market rose from 33.4% in 2005 to 38.6% in 2007, and they are expected to continue increasing (Mobile Marketing Magazine, 2009).

2.3.2 Churn Drivers in the Mobile Market

Ahn et al. (2006) identified four major groups of factors affecting customer retention: (1) Customer satisfaction; (2) switching costs; (3) service usage; and (4) customer-related variables. Added to these, social influence is another important factor, which directly affects customer retention (Birke and Swann, 2006; Dasgupta et al., 2008). According to Chu, Tsai and Ho (2007), price comes in the top of the contributors to the churn, followed by customer services and then by service quality and coverage (at the same level of importance). Mattison (2005, p.35) proposed a taxonomy for the major drivers of customer churn (see Figure 2.1). The highlighted blocks denote the areas of primary interest in this study.

Market research gives a more realistic view for the main contributors to churn. In a recent study carried out by Nokia Siemens Networks (2009) analysing customer experience in the mobile market in ten different countries, including the UK, it was revealed that there are seven key drivers for churn. These drivers are network coverage, cost of services, customer care services, offered rate packages, mobile service offering and (finally) billing. Figure 2.2 shows the major churn drivers in mature markets, which are closely related to the UK market, and the percentage of the customers who voted for each factor.

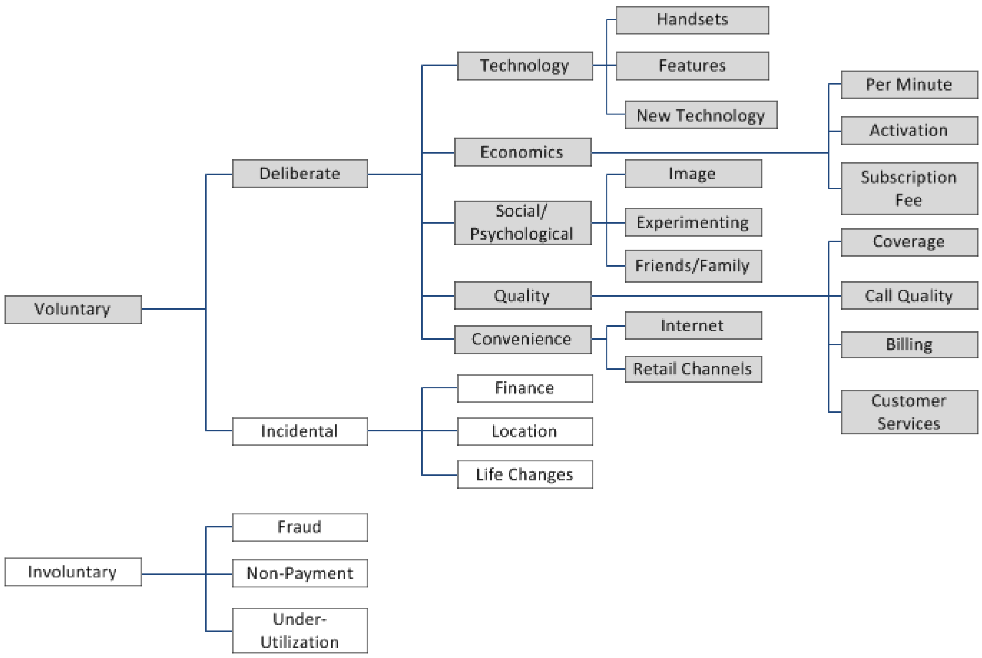


Figure 2.1: Churn Taxonomy (Compiled after Mattison, 2005)

The results from the market research are partially consistent with the literature. As previously reported in the literature, market research indicates that costs have the greatest influence on customers' decisions to churn. Contrary to the literature, network quality and coverage have minimal influence on customer decisions in mature markets. Network quality is already well established in mature markets, and there are no longer significant differences between market players. Another discrepancy with the literature is that contract structure has a great influence on customers' decision.

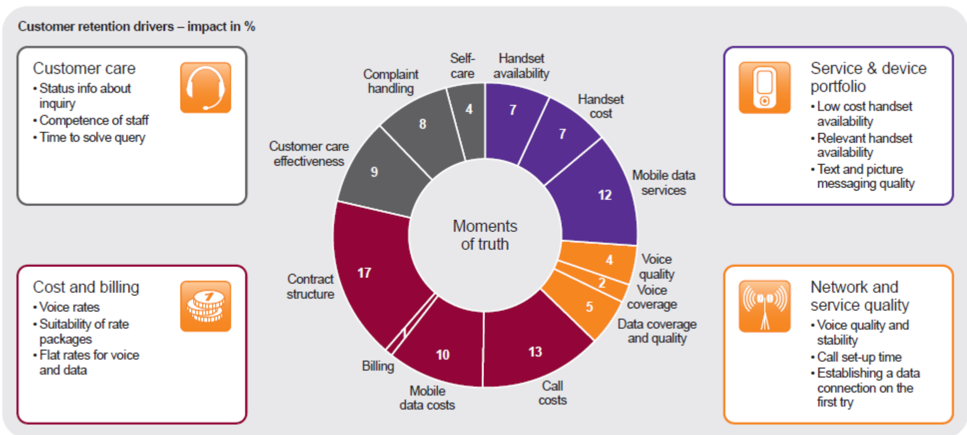


Figure 2.2: Churn Drivers in Mature Markets (Nokia Siemens Networks, 2009)

2.4 Customer Behaviour Modelling

The central principle of CRM is understanding and predicting customer behaviour. Customer behaviour modelling is the main process to produce this understanding. Kamakura et al. (2005) classified CRM models into two categories: Analytical and behavioural models. Based on this classification, analytical models involve using customers' data stored in the business data warehouse to develop longitudinal models that help the business to increase their profit from customer. Behavioural models utilise experiments and surveys to analyse psychological responses to the services interaction (Kamakura et al., 2005). Furness (2001) provided a more objective classification, classifying customer behaviour modelling into three categories: (a) Predictive modelling; (b) descriptive modelling; and (c) combined predictive and descriptive modelling. These categories are detailed below.

- Predictive modelling is mainly concerned with predicting how the customer will behave in the future by analysing their past behaviour. Predicting customers who are likely to churn is one example of the predictive modelling.
- Descriptive modelling attempts to answer "why" questions, whereas predictive modelling is often used to answer "who" questions. Customer clustering is one example of descriptive modelling, which is used to separate or segregate customers in different groups based on their behaviour.
- Combined descriptive and predictive, as its name indicates, integrate the two modelling approaches to provide more conclusive answers for "who" and "why" at the same time.

2.4.1 Traditional Modelling Approaches

Different methods and instruments have been used to investigate and analyse customer retention. Broadly speaking, these methods are divided into two categories: Statistical analysis methods, such as correlation analysis, and data discovery methods, such as data mining techniques. Data mining refers to the process of extracting knowledge from data. It employs a variety of tools and algorithms that originate in statistics, artificial intelligence (AI), and machine learning. It is usually used in the context of customer retention to predict future customers behaviour and trends by detecting patterns and relationships within raw data. Data mining methods dominate customer churn literature for two reasons: Their prediction results outperform the traditional statistical techniques (Ye, Liu and Li, 2008); and they are able to analyse much larger number of variables (Lemmens and Croux, 2006). Therefore, the focus in this study is on evaluating data mining in customer churn analysis.

Customer churn has been studied in different industries, including: Banking, insurance, and telecommunications (Coussement and Van den Poel, 2008). Customer churn studies also have been formulated and applied in different contexts, such as economics, behavioural and psychological (Kamakura et al., 2005). As a result, there is an extensive literature on customer retention, investigating different factors and influences using a wide variety of tools and techniques (Ahn et al., 2006). The following paragraphs shed light on some of the key issues reported in the customer churn literature. Table 2.1 provides a summary for some of the key studies in the churn analysis in the telecommunication industry. This includes the analysis tools, the modelling types and the main factors covered in these studies.

Gerpott et al. (2001) studied the German mobile market based on a sample consisting of 684 residential mobile users. They were mainly concerned with finding the relationship between the three constructs: Customer retention, loyalty and satisfaction. They showed that the three constructs are causally interlinked. In addition, they identified some other factors, which have significant effects on customer retention such as mobile service price, mobile service benefit perceptions and lack of number portability. The main limitation of this study is that it investigates customer behavioural intentions and attitudes and neglects the actual customer usage behaviours. In addition, this study employs traditional statistical techniques, which are now being questioned in the churn literature (Twomey and Cadman, 2002).

Kim and Yoon (2004) analysed customer churn in the Korean mobile market based on a survey of 973 mobile users. They conducted an empirical analysis to identify the determinants of churn. Their results indicated that churn probability is associated with the level of satisfaction. Furthermore, they identified other services attributes that might affect churn including call quality, tariff level, handsets, brand image and subscription duration. They concluded that the main cause of customer churn in the Korean mobile market is the desire to change handsets and dissatisfaction with specific service attributes, such as call quality or price level. This study shares the same limitation as Gerpott et al.'s (2001) study, namely the use of the questionnaire as a data collection tool. The concerns about potential biases in the questionnaire-based study have been reported widely in the literature (Liu, Guo and Lee, 2010). In another study of the same market conducted by Ahn et al. (2006), the limitation of using questionnaires to collect customer data was overcome by analysing actual transactions and real customers' data provided by a service provider. Ahn et al. (2006) investigated the key determinants of churn and reported service quality, customer usage and switching costs as the main determinants of churn.

Seo et al. (2008) analysed customer retention in the US mobile market based on a database of 31,769 customers and call log files for one of the top ten US mobile services providers. They used binary logistic regression modelling to analyse the behavioural and the demographical factors affecting customer retention. They investigated six variables: Service plan complexity, handset sophistication, length of association, connectivity quality, age and gender. As reported by Seo et al. (2008), one of the limitations of their study is that it was conducted before the introduction of local number portability. As noted earlier, the mobile market is changing rapidly and more recent and comprehensive research is needed.

Yu et al. (2005) overcame the limitations of survey-based studies by using Duke Teradata's 2003 churn modelling tournament data, which contains 100,000 customer records extracted from a major mobile operator in the US. They developed a model called 'Churn-Strategy Alignment Model' to evaluate churn based on 172 variables by using factor and reliability analysis. This model offers managers a new way to define customer retention strategies and helps them to understand why customers churn, rather than focusing only on who are going to churn. However, one major drawback of this model is that it fails to consider social influences.

Hwang et al. (2004) investigated the Korean mobile market and suggested a lifetime value model considering propensity to churn, past financial contribution and customer potential value at the same time. They used decision tree, logistic regression and neural network to develop and evaluate their model, which can be used to segment customers and develop customer retention strategies based on the customer lifetime value. Choosing the most profitable customers and retaining them while lessening or terminating relationships with less profitable customers is one of the most successful customer retention strategies to improve business profit (Kim et al., 2006). One important contribution of Hwang's et al. (2004) study is that it adds important findings to the empirical customer churn literature by considering customer lifetime value in the churn analysis. However, the study did not take into account other important factors that strongly affect customer churn, such as social influences and market characteristics.

Hung et al. (2006) used decision tree and neural networks to analyse customer churn in Taiwan based on data including customer demography data, billing, customer service interaction and call detail records data. Although the results of this study show a significant improvement (based on lift chart) from those in early studies, one major criticism of this work is that it did not take account of social influences along with other factors that may affect customer churn.

Lemmens and Croux (2006) applied bagging and stochastic gradient boosting (two data mining algorithms) to predict customer churn in a US mobile company. In their study, they use three groups of predictors: Behavioural, company interaction and customer demographics predictors. According to the study results, using the two algorithms performs comparably better than logistic regression. Another important finding of this study is that using ensemble classifiers produces superior performance over single classification models. In ensemble models, multiple classification models are combined into one classifier by using different methods, such as majority voting. Nevertheless, this study shares the same limitations as others in that it does not consider social influences.

In another major study, Neslin et al. (2006) investigated the performance of churn prediction models of different statistical and data mining tools. This study reported the results of a tournament in which researchers use the Duke Teradata's churn modelling data to build churn prediction models on that data. The results of this study demonstrate that logistic regression and decision tree models outperformed traditional statistical tools such as discriminant and explanatory models. The results of this study contribute to make churn prediction more accurate, but cannot provide answers for why a customer might churn (Hadden et al., 2007).

Lima et al. (2009) identified an important shortcoming of the existing prediction models, such as logistic regression and decision tree, which is that they lack transparency and comprehensibility. They suggested incorporating domain knowledge to improve the interpretability of the resulting models. Based on two telecom data sets they demonstrated how domain knowledge could be used to improve the interpretability of predictions models.

In summary, as shown in Table 2.1 most studies in the field of customer churn in the mobile industry have only focussed on customers' characteristics and interactions with the operator, while ignoring interactions with the mobile market and interactions among customers themselves. In addition, much of the research up to now have been predictive in nature, and has not offered answers to the question: "Why a customer might churn?"

Study	Method	Model type	Factors covered
Gerpott et al. (2001)	Explorative factor analysis and structural equations modelling	Descriptive	Customer behavioural intentions and attitudes
Kim and Yoon (2004)	Binomial logit modelling	Predictive	Call quality, tariff level, handsets, brand image, and subscription duration
Ahn et al. (2006)	Logistic regression	Predictive then descriptive	Service quality, usage, and switching costs
Seo et al. (2008)	Binary logistic regression modelling and two-level hierarchical linear model	Two level: predictive then descriptive	Service plan complexity, handset sophistication, length of association, connectivity quality, and demographic
Yu et al. (2005)	Factor and reliability analysis	Descriptive	Three groups of factors, covering product, customer solution and usage data
Hwang et al. (2004)	Decision tree, regression and neural networks	Predictive	Demographic, usage, billing and customer services
Kim et al. (2006)	Decision tree, regression and neural networks	Predictive	Demographic, usage, billing and customer services
Hung et al. (2006)	Decision tree and neural networks	Predictive	Demographic, usage, billing and customer services
Lemmens and Croux (2006)	Bagged and boosted decision tree and Logistic regression	Predictive	Three groups of factors, covering behavioural, company interaction and demographic
Neslin et al. (2006)	Neural networks, decision tree, logistic regression, and discriminate analysis	Predictive	Demographic, usage, billing and customer services
Lima et al. (2009)	Decision tree and logistic regression	Predictive	Demographic, usage, billing and customer services

Table 2.1: Literature Review of Key Studies on Churn in Telecommunications

2.4.2 Social Influences on Customer Retention

Social network influences have been identified as major contributing factors in customer retention (Dasgupta et al., 2008; Zhang et al., 2010). These influences are exercised via WOM. WOM refers to the spread of information, either positive or negative, between customers about a product or service. Keaveney (1995) found that 75% of defecting customers spread negative WOM to at least one other person. This section provides a brief discussion of relevant literature on social network influences on customer retention in the mobile market.

Detailed examination of the effect of social influences on customer retention by Dasgupta et al. (2008) showed that the probability of churn is significantly influenced by the number of friends who have churned. They developed a social network model for mobile customers based on information extracted from Call Detail Record (CDR) data. The social relationships in this model were based on the duration of voice calls and call frequency. The study of Dasgupta et al. (2008) looked at predicting customer churn by using decision tree algorithm on a data set containing three groups of attributes: (1) Usage attributes extracted from CRD data; (2) connectivity attributes extracted from the modelled social network; and (3) interconnectivity attributes extracted from the modelled social network. Similar to the study of Dasgupta et al. (2008), Zhang et al. (2010) used CDR data to extract an extensive set of attributes that describe the social network of mobile customers. Zhang's et al. (2010) study is distinguished from Dasgupta et al.'s (2008) in that the former incorporated the social network attributes along with the traditional attributes such as demographical and customer service-related attributes. Zhang et al. (2010) demonstrated empirically how the inclusion of social network attributes in churn prediction models could greatly improve churn prediction accuracy.

Although the two aforementioned studies contributed greatly to our understanding of the effect of social network on customer retention, they inherited the same limitations of traditional churn analysis techniques. The main limitation of such analysis techniques is that they do not reflect the whole complexity of customer churn in the telecommunication industry. Alternatively, ABMS is suggested as a new and novel technique to capture the underlying complexity in the telecommunication market (Bobek and Perko, 2006).

2.4.3 Agent Based Modelling Approach

Researchers and practitioners alike have been attempting to employ new techniques and approaches to investigate and analyse customer and market behaviours. Among these approaches is ABMS technique (Twomey and Cadman, 2002). ABMS provides insight into how the real systems behave under different scenarios and provides useful explanations and careful analysis for the relations between the different determinants in complex systems. ABMS is composed of two main pillars: Modelling and simulation. Modelling simply is abstracting real-world phenomena into a representation or model, whilst simulation is a process of executing models over time to mimic and imitate real or proposed situations or systems.

ABMS has linkages to many other fields, including social sciences, artificial life science, management science and complexity science (Ma and Nakamori, 2005; Macal and North, 2007). Moreover, ABMS benefits from the latest developments in the field of AI and Individual Based Modelling (IBM) in ecology. ABMS became widespread in the early 1990s, but preliminary work on developing models using the ABMS approach was undertaken by Thomas Schelling in 1978 (Macal and North, 2007). Schelling employed ABMS and cellular automata to investigate housing segregation patterns by modelling people and the socially relevant process, which represent interactions between people (Li et al., 2008).

ABMS has been widely used in many different science fields. As a result, many terms have been used to describe it. Some of these terms are too broad, and some of them are too specific to certain scientific fields. These terms are often used interchangeably in the literature: Multi-agent simulation (MAS) (Siebers and Aickelin, 2008), agent-based modelling (ABM) (Gorman et al., 2006), individual-based modelling (IBM) (Railsback, 2001), multi-agent-based simulation (MABS) (Edmonds, 2001), agent-based social simulation (ABSS) (Bryson, 2003), and agent-based modelling simulation (ABMS) (North and Macal, 2007). To avoid confusion, the term "ABMS" is used throughout this work.

ABMS can be used as a laboratory to experiment with various "what-if" scenarios (Twomey and Cadman, 2002). In addition, it can be used for exploratory, explanatory and predictive purposes (Siebers et al., 2007). Goldsman (2007) pointed out that simulations are widely used to analyse systems that are too complicated to be analysed by traditional analytic methods, such as standard probability and statistics. Ma and Nakamori (2005) considered ABMS to be a new way of thinking about the world, while Macal and North (2007) went further and considered ABMS to be a third way of doing science, in addition to traditional inductive and deductive methods. Gross and Strand's (2000) definition of ABMS was the most complete and appropriate for the purpose of this study. They defined ABMS as a:

> "Set of techniques in which relations and descriptions of global variable are replaced by an explicit representation of the microscopic features of the system, typically in the form of microscopic entities ('agents') that interact with each other and their environment according to (often very simple) rules in a discrete space–time" (Gross and Strand, 2000, p.27).

A few studies have investigated the use of ABMS to model the customer behaviour in the telecommunications market. Most of these studies have not reached the stage of development (Bobek and Perko, 2006). Baxter et al. (2003) employed ABMS to develop an intelligent CRM tool. In their model, they studied the influences of WOM on the adoption of products and services. Collings et al. (1999) used ABMS to model technology adoption in the telecommunication market. They use real field data to calibrate their model and then used it to investigate the effect of different structure, dynamics and distribution of social networks on adoption rates. They argued that ABMS provides a very useful tool to incorporate different type of interactions in one model.

Twomey and Cadman (2002) introduced the concept behind ABMS and highlighted the strengths, weaknesses and opportunities of using it in the telecoms and media markets. They also offered general guidelines for designing and developing ABMS models. To demonstrate the utility and the uses of ABMS in business applications, they provided some examples of ABMS models that have been developed and used for commercial purposes. Although this study contributes to limited literature of ABMS in the telecommunication market, it does not offer either an empirical or a theoretical model for customer behaviour.

Li, Lu and Zhou (2009) proposed an ABMS model of customer retention management. Their model considered interactions among firms and customers and interactions among customers themselves. The main limitation of this study is that it does not offer an empirical test for the proposed model. Moreover, it neglected some other important factors proven important in previous studies, such as customer satisfaction-related factors.

ABMS is relatively new to the customer modelling for business applications in general, and for the telecommunication industry in particular (Twomey and Cadman, 2002). It is still in its early development stages, and much of the research up to now has been theoretical in nature. However, there have also been some empirical ABMS studies, which attempted to implement, validate and experiment with ABMS models, rather than just proposed theoretical and conceptual models. Some of these studies are discussed below.

Said et al. (2002) developed an ABMS model to study the effect of customer interaction on customer behaviour. Their model focused on the interplay among interaction rules, socio-economic profiles to investigate customer purchase behaviour and adoption decision. In the model, customer behaviour is activated by different external stimuli such as negative and positive WOM, innovations diffusion, brand loyalty and promotions. Said et al. (2002) used a genetic algorithm to build a realistic customer population, which is compliant with a real market. The model developed in this study was built based on a comprehensive framework, which focused on elementary behavioural primitives. Although the model provides some useful insights into the customer behaviour in general, using it to investigate customer churn is not easily testable, especially in the telecommunications market, because of the special characteristics of this market.

Brannon et al. (2000) built an ABMS model for the apparel and textile market. The model incorporated external influences, such as promotion and social influences to study the customer purchase decisions. The model was validated by comparing its results with published results from diffusion innovation studies. This model shares the same limitations of other models in that it cannot replicate the special characteristics of the telecommunication market.

North et al. (2010) developed a large-scale agent-based model for consumer marketing. This model provided a holistic view of the consumer market and aimed to be used as virtual market learning lab and as a test bed to evaluate the effects of different strategies on the market. The model replicated the shopping behaviour of consumers and the business behaviour of companies in a simulated market. Although this model was applied successfully to solve real business problems, North et al. (2010) emphasised that the model cannot be universally adopted without being calibrated, verified and validated for a given usage.

Kuhn Jr et al. (2010) used an illustrative case study of an airline company to develop an ABMS of consumer airline market. In the model, the airline company was affected by federal regulation, internal business policies, competition with other companies and environmental factors, such as fuel costs. Regarding the model uses, the authors demonstrated how the model could be used as a decision support tool and as a driving force to support future research activities.

2.5 Limitations of Previous Studies

Drawing on existing literature and research findings, it can be argued that due to the differences between mobile markets and other markets, as well as the variations in the mobile market itself across countries, the results and the outcomes of such research cannot be generalised. Seo et al. (2008) supported this assertion, and observed that generalisability was one of the limitations of their study of the US mobile market. Therefore, there is a growing need to understand the distinctive characteristics of the UK mobile market, which can affect customer churn.

There have been no comprehensive studies on customer retention in the UK mobile market recently published. Grzybowski (2008) analysed some of the UK mobile regulatory variables and measures, such as switching costs, to investigate customers' choiceness of network operator. Birke and Swann (2006) examined the impact of social ties on consumer choice in network markets. In addition, they identified three important factors: Price, firm size and switching cost, which have a significant effect on customer retention. These two studies contribute to the collective understanding of factors affecting customer retention in the UK mobile market. In order to construct a comprehensive understanding for the customer retention, other important factors such as customer satisfaction factors and social influences need to be jointly considered.

Despite the large amounts of research done on customer retention, few studies offer a comprehensive understanding of the relationship between the different factors and customer retention. Ahn et al. (2006) pointed out that much of the literature about customer retention focused on finding a few specific factors like , for example, customer satisfaction and customer loyalty, and neglected other important factors such as social ties. Verbeke et al. (2010) underlined the same issue and asserted that much of the customer retention studies focused on predicting churn and paid less attention to the causes of churn. In addition, they indicated that the lack of comprehensibility is one of the major issues in customer retention models, and confirmed the need to build models that are more comprehensible.

As reported in the ABMS literature, the level and details of ABMS models of consumer markets range from very general to very specific. In addition, these models are built according to different theoretical and empirical approaches.

Using any one of these models in a different context to solve different problems is a challenging task and requires tailoring these models to suit the nature and the special characteristics of each market. Nonetheless, these models clearly show that ABMS is a viable tool to account for the complexity inherent in customer interactions. To summarise, Table 2.2 defines some of the significant limitations of the existing research that covering both traditional and agent-based customer modelling approaches.

Traditional Modelling Approaches	Agent-based Modelling Approache
• Models' generalisability is limited due to differences among national mobile markets • Much of the research focused on customers' characteristics and interactions with the operator, while ignoring interactions with the mobile market and interactions among customers themselves • Much of the research up to now has been predictive in nature, and does not offer answers for the question "why might a customer churn?" • There is no published research that provides a comprehensive view of customer retention in the UK mobile market	• There is limited empirical research on customer behaviour, specifically customer retention, in the mobile market • Existing ABMS models of consumer markets are either very general, or very specific, which limits their ability to capture the distinct characteristics of the mobile market

Table 2.2: Summary of the Key Limitations of the Existing Literature

2.6 Summary

This chapter provides the context and the grounding for this study. Customer retention is an important issue for any business, and it is more important in mature markets where new customers can only be acquired from competitors. Data mining provides very useful insights into customer churn and offers continuous and current knowledge of whole customer populations. Although the superiority of using data mining is well established in the customer churn analysis, its utility in capturing the complexity inherent in the mobile market and customer retention is questionable (Twomey and Cadman, 2002).

The literature review reported in this chapter reveals that decision tree and logistic regression were among the most commonly used data mining models. It also reveals that empirical customer retention studies suffer from two major limitations, namely that they: (1) Focus on analysing a few specific factors, like customer satisfaction and customer loyalty, neglecting other important factors such as social ties; and (2) focus on customers' characteristics and interactions with the operator, while ignoring interaction with the mobile market and interaction among customers themselves. To overcome these limitations, ABMS is proposed as a novel tool to model the customer churn behaviour in the UK mobile market. The potential of ABMS to overcome the limitations of traditional analysis methods is illustrated by discussing some of its empirical applications in different consumer markets. To demonstrate the utility of ABMS, a new empirical and agent based methodology to analyse customer retention will be introduced in subsequent chapters.

Chapter 3: Research Design and Approach

3.1 Overview

This chapter explains the research approach followed to investigate customer retention in the UK mobile market. Detailed description of the research methodology and methods used in designing and testing the proposed customer behaviour simulation (CubSim) model is also provided. In addition, this chapter provides an explanation of the research development phases along with the input and output of these phases. The Design Science Research (DSR) paradigm is employed in this study as a general methodological framework.

This chapter is structured as follows: Section 3.2 highlights the main research paradigms employed in Information Systems (IS) research. In addition, it provides the broad outline of the development research phases that characterise these paradigms. Section 3.3 presents data mining and ABMS as the main research methods used to conduct this study. Detailed explanation and justification of the use of these methods (used in this study) are also presented in this section. Section 3.4 illustrates the application of DSR to this study and demonstrates how the theoretical processes are transformed into four main practical phases. Finally, the key aspects of this chapter are summarised in Section 3.5.

3.2 Research Approaches and Paradigms in IS

Information systems is a multidisciplinary research field, because it involves a range of disciplines such as engineering, computer science, mathematics, management science and others (Baskerville and Myers, 2002). Accordingly, there is no single or standard approach to carry out information systems studies;

instead, there are a variety of research approaches, paradigms, techniques, methodologies and methods that can be employed. Multi-paradigm research is increasingly accepted in IS (Mingers, 2001). Terms such as paradigm, methodology, technique and method are open to multiple interpretations. In order to limit the different interpretations in the context of this study, specific interpretations of these terms have been adopted in this research. The next section is devoted to clarifying the specific interpretations adopted in this study. It must be admitted that the adopted interpretations are by no means correct in an absolute sense; rather, they seek to clarify how the terms are being defined and used in this study.

3.2.1 Definition of Terms

Research paradigm is defined as a set of beliefs or the underlying philosophical assumptions, which guide the activities conducted throughout the research process (Mingers, 2001). Hevner et al. (2004) classified research paradigms in IS into two complementary but distinct paradigms, behavioural science and design science. According to Hevner et al. (2004), behavioural science seeks to develop and justify theories that explain or predict organisational and human phenomena, whereas the design science seeks to create artefacts that aim to provide solutions to business problems. The following sections provide more details about the two paradigms and demonstrate how they were combined within this study.

Research methodology comprises the procedures that need to be carried out to conduct a research study. It normally encompasses components such as phases, activities, methods, techniques and tools. There is often confusion between the methodology and the methods. Mingers (2001) distinguished between three connotations of the term methodology: The first, the most general one, is denoting the study of methods; the second, the most specific one, is used when describing the methodology of a particular research study; the third is a generalisation of the second connotation. This generalisation includes all the principles and rules that used in a particular area of study. For this study, the second meaning is adopted, and the term methodology has been used to describe the procedures followed to conduct this study. Methodology is less prescriptive than a method, and it often consists of various methods or techniques (Mingers, 2001).

Methods and techniques are used, sometimes synonymously, to perform tasks inside the different processes of which a methodology consists. Mingers (2001) described methods and techniques as well-defined sequences of operations that if carried out proficiently yield predictable results. However, this study is concerned with combining multiple research method to strengthen the value and relevance of the research outcomes. For example, this study uses data mining and ABMS to gain deeper insights into the customer retention problem. In addition to combining different methods, different ideas of different paradigms also have been used in this study. The following sections discuss the two common IS paradigms, behaviour science and design science, adopted in this study.

3.2.2 The Behavioural Science Paradigm

Behaviour science originates from natural science and aims to describe, explain and predict phenomena (very often human behaviour). The outcomes of behavioural science research are likely to lead to the development or the evaluation of theories (Hevner et al., 2004). Similar to other research paradigms, behavioural science paradigm defines a set of activities to conduct research studies. Bukvova (2009) discussed five models of procedures normally applied to conduct research under the behavioural science paradigm. One of these models is that of Graziano and Raulin (2009, p.40), which provides a broad general view of the behavioural science research procedures (see Figure 3.1).

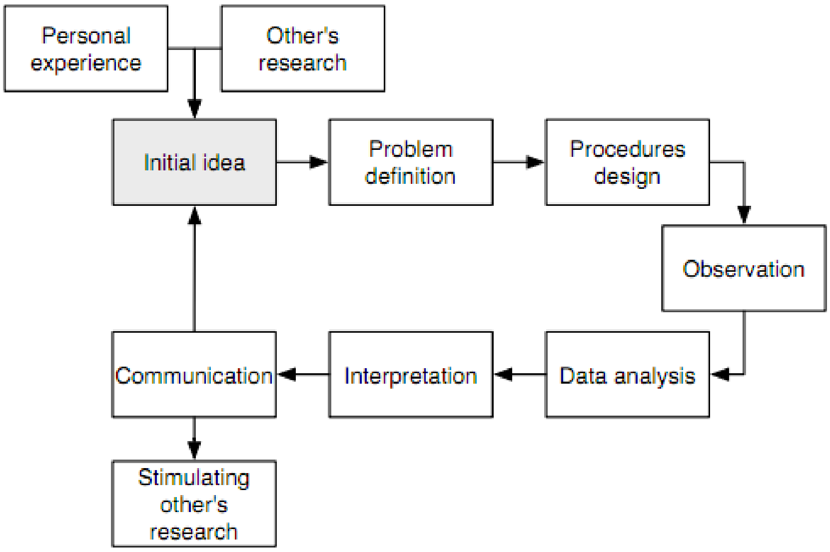


Figure 3.1: Behavioural Research Process (Source: Graziano and Raulin, 2009)

Graziano and Raulin's model shows the phases or groups of tasks that are normally carried out in behavioural since research. According to Graziano and Raulin's model, initial research ideas are generated based on personal experience or existing literature. To explore these ideas with the scientific research, they have to be clearly defined in the form of research questions. Then, the research procedures that should lead to the solution of the research questions are defined throughout the procedure-design phase. The outcomes of this phase determine the data-collection and data analysed methods. After the data collection has been carried out, the data is analysed and interpreted to answer the research questions. Then, finally, the research outcomes should be disseminated to the wider community, which in turn lead to stimulating and encouraging other research efforts. The following sections discuss the design research paradigm and then illustrate the cross-paradigmatic approach followed in this study.

3.2.3 The Design science paradigm

The DSR paradigm is inspired by Simon's view of the 'Science of the Artificial', wherein 'artificial' means a hand-made product or an artefact (Simon, 1996, p.123). March and Smith (1995, p.253) defined design science as 'an attempt to create things that serve human purpose'. DSR facilitates building and evaluating artefacts to address business needs and focuses mainly on building and evaluating IT artefacts that are described as purposeful, innovative and novel (Hevner et al., 2004). 'Purposeful' means that the produced artefacts should offer a 'utility' that address unsolved problems or offer a better solution that can enhance existing practices (Vaishnavi and Kuechler, 2004).

The design process in DSR is different from other design activities that focus on building usable artefacts intended to accomplish goals. DSR is distinct from other design activities in that it involves creating, capturing and communicating the knowledge acquired throughout the design process (Vaishnavi and Kuechler, 2004). DSR is also characterised by the iterative reconstruction of artefacts and assumes that knowledge emerges during the iterations effort (Vaishnavi and Kuechler, 2004). Clearly, the design process in DSR can be seen as a learning process, whereby understanding is enhanced from one iteration to another, which in turn helps to improve the artefacts' quality.

3.2.4 DSR Processes

DSR processes follow a systematic approach, structured in several phases. Vaishnavi and Kuechler (2004) categorised the DRS processes into five phases, as shown in Figure 3.2. These phases and their outputs are described and explained below.

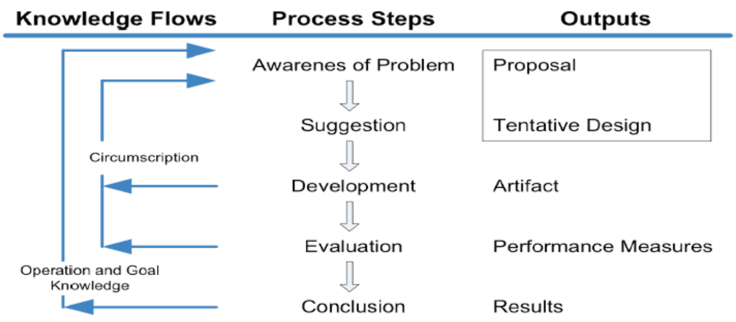


Figure 3.2: Design Science Research Phases (Source: Vaishnavi and Kuechler, 2004)

Awareness of problem: DSR starts with identifying and defining the problem under study. Different sources can be used in this phase ranging from the new developments in the industry to the literature. The research problem needs to be clearly articulated, and the research scope needs to be properly identified. The output of this phase is a proposal for a new research.

Suggestion: This phase involves exploring and evaluating potential solutions. During the initial analysis and design in this phase, further insights are acquired into the problem domain. In addition, the specifications of the required solutions are defined. A tentative design or a declarative representation of suggested solutions is the output of this phase.

Development: In this phase, the DRS artefacts are developed according to the suggested solutions resulted from the previous phase. The artefacts are the output of this phase and the core outcome of the whole DSR process. March and Smith (1955) classified artefacts into four categories: Construct, models, methods and instantiations (for definitions, see Table 3.2).

Evaluation: Once the artefacts are developed, these artefacts are analysed and evaluated against the specifications that have been set in the suggestion phase. If the outcomes of the development or the evaluation phases are not satisfactory, the design cycle goes back to the first phase (awareness of the problem), along with the new knowledge acquired during the previous phases. This loop may be repeated many times until the artefacts evaluation satisfies the solution requirements. Performance measurements are the output of this phase and meant to improve efficiency, effectiveness and the credibility of the artefacts.

Conclusion: This phase concludes the DSR cycle. The results of the whole DRS process are communicated to the wider audience in this phase. In addition to the final outcomes of the research study, research results encompass also the knowledge acquired during the design cycles, which can be used by the practitioners as guidelines on how to use the developed artefacts in similar situations.

3.3 Research Methods and Techniques

Because of the multidisciplinary nature of IS, there are diverse research methods and techniques that can be used to conduct IS research. For the purpose of this study, two research methods and techniques were used: Data mining and ABMS. The following sections introduce and discuss the two methods. Detailed explanation and justification of how these methods were applied in this study are also presented.

3.3.1 Data Mining Overview

Data mining is a cutting-edge tool used usually in the context of customer retention to predict future customer behaviours and trends by detecting patterns and relationships within raw data. Hung, Yen and Wang (2006, p.517) defined data mining as 'applying data analysis and discovery algorithms to detect patterns over data for prediction and description'. In the context of IS research, data mining can be seen as an analytical tool used to justify or evaluate theories and artefacts (see Figure 3.4).

Data mining is used to build models from data; these models can be classified according their functions into different type of data models. Ngai, Xiu and Chau (2009) classified data models into seven categories: Association, classification, clustering, forecasting, regression, sequence discovery and visualisation. Regression and decision tree are two models commonly employed by both academics and practitioners to study and model different factors that may affect customer retention (Hung, Yen and Wang, 2006). Regression and decision tree are discussed in details in Section 4.2. The next section introduces the general methodology followed to develop the data mining models in this study.

3.3.2 Data Mining Development Cycle

To meet research objective 1, two widely used data mining techniques for analysing customer churn, decision tree and logistic regression models, were employed. Detailed experiment results and model building procedures are provided in Chapter 4. The data mining development cycle follows the well-known data mining methodology CRISP-DM (Cross-Industry Standard Process-Data Mining) (Chapman et al., 2000). The CRISP-DM methodology provides complete and high-level instructions and procedures for applying data mining algorithms to solve real world problems. A CRISP-DM model consists of six phases: Business understanding, data understanding, data preparation, modelling, evaluation and deployment (see Figure 3.3). Business understanding entails understanding the business objectives and requirements and matching them accordingly with data mining applications. Data understanding includes data collection and familiarisation with the data. It also involves data exploration, data description and data quality verification. Data preparation is the process to transform the raw data into a suitable format for applying the data mining algorithms.

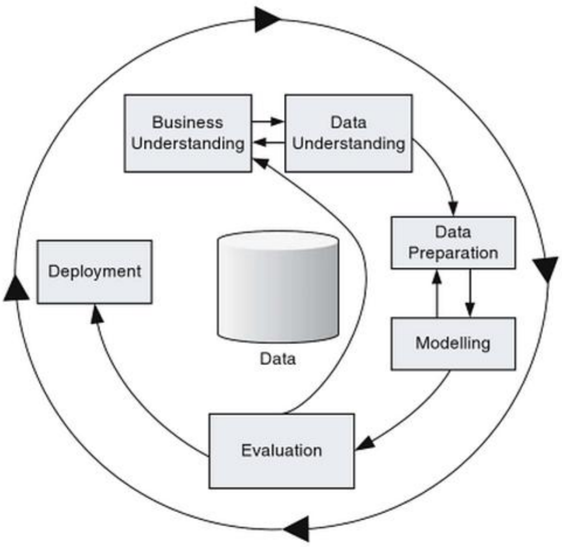


Figure 3.3: The CRISP-DM KD Process Model (Source: Chapman et al., 2000)

Data preparation includes data cleaning, data transformation and data reduction. Modelling is the process of selecting and applying suitable data mining algorithms for addressing business problem. In addition, modelling includes identifying and optimising data mining algorithms parameters and evaluating algorithms performance. Evaluation examines data mining algorithms results from a business objective perspective. Based on the evaluation outcomes the decision about adopting data mining models should be taken. Deployment is the last phase, wherein the discovered knowledge from the data mining process should be organised and presented in a format that the business can use. Deployment can be as straightforward as generating a report or as complex as integrating data mining solutions into daily business processes.

3.3.3 ABMS Overview

ABMS involves creating artificial agents mimicking attributes and behaviours of their real-world counterparts (Kyrylov et al., 2004). There is no consensus on a formal definition of the term agent, and some of the definitions are intentionally broad to provide an inclusive definition that better reflects its multidisciplinary nature. For the purpose of this study, the definition of Renovales and Orlando (2006) was adopted. They defined agents as autonomous, computational entities acting and behaving upon their environment and their own experience.

ABMS agents are different from other agents especially those found on shopping systems and web sites. Macal and North (2007) referred to these agents as mobile agents and described them as lightweight software proxies that travel over the web and execute some tasks for users and to some extent behave autonomously. Other examples of light agents are virtual tour guides and personal assistants. These agents are usually rule-based system with animated characters as frontend.

ABMS is a computational model for simulating the actions and interactions of autonomous individuals in a network, with a view to assessing their effects on the system as a whole. ABMS is composed of two main pillars: Modelling and simulation. Modelling, simply, is abstracting real-world phenomena into a representation or model, whilst simulation is a process of executing models over time to mimic and imitate real or proposed situations or systems. ABMS can be used in IS research to conduct simulation studies to justify or evaluate theories and artefacts. Figure 3.4 shows the simulation rule within the IS framework proposed by (Hevner et al., 2004).

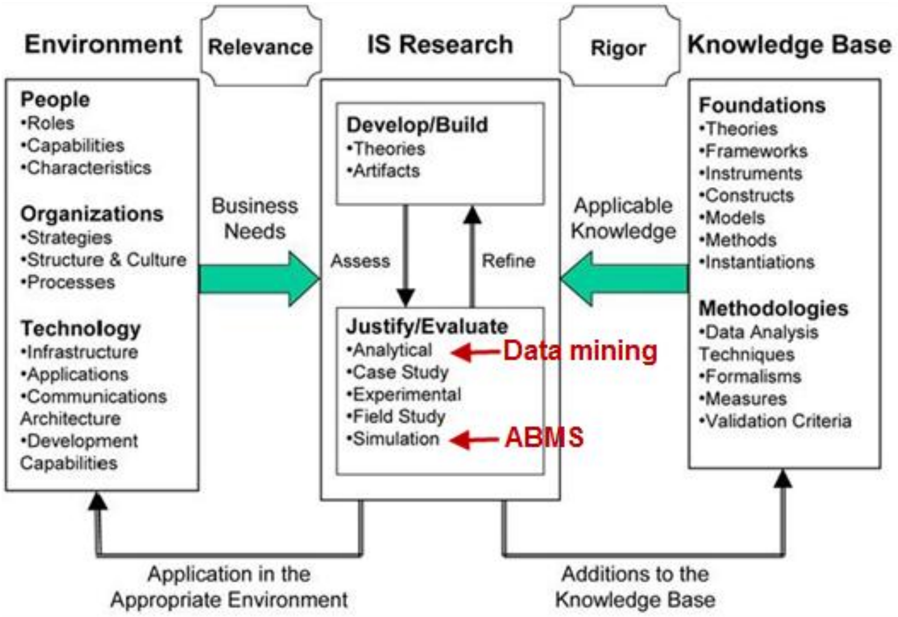


Figure 3.4: The IS Framework (Compiled after Hevner et al., 2004)

3.3.4 Why ABMS?

To present a new method or technique to overcome the limitations of the traditional methods to analyse and investigate churn phenomena, a demonstration of the feasibility and the validity of this method is needed. The following paragraphs shed light on the appropriateness of using ABMS to investigate and analyse customer behaviour and churn phenomena.

ABMS is usually used to model and analyse complex systems. In order to demonstrate the appropriateness of ABMS in the context of customer retention studies, a matching between the characteristics of customer behaviour and complex system is carried out. Using complexity in the technical context refers to the fact that the whole system behaviour cannot be determined by analysing the behaviour of each individual part separately (Gilbert, 2007). The UK mobile market exhibited unprecedented changes in many dimensions including technical, regulatory, legislative and demand issues. Additional complexity resulted from non-linear interactions between customers and mobile operators and between customers themselves.

Gilbert (2007) and Li et al. (2008) considered interactions within human societies to be a complex system. Gilbert (2007) claimed that the interactions among any system involving transitions of knowledge and materials affecting the behaviour of the recipients make this system complex. WOM is one example of how knowledge can affect the customer behaviour in the mobile market, and this support the argument that customer behaviour is a complex system. Twomey and Cadman (2002) expanded this more and identified six primary characteristics of complex systems: Non-linearity, self-organisation, heterogeneity, adaptation, feedback and emergent behaviour. Subsequently, the following paragraphs show matching between some of the characteristics of complex systems and customer behaviour.

Detailed examination of non-linearity in customer satisfaction and loyalty by Streukens and De Ruyter (2004) showed that the relationships in customer satisfaction and loyalty models are non-linear and asymmetric. Gilbert (2007) and Li et al. (2008) supported this argument and generalised the non-linearity description to encompass all of human society's interactions.

Emergent behaviour is another feature of customer behaviour whereby the churn phenomenon is indeed an emergent behaviour. Gilbert (2007) illustrated this by showing that the emergent behaviour occurs when the outcome of the collection of simple individuals work show unexpected behaviour that is significantly more complicated than what a single individual could do. A simple example of emergent behaviour is the flocking behaviour of birds (Sanchez and Lucas, 2002), where each bird in the flock is controlled by only three simple rules: (1) Collision avoidance; (2) velocity matching; and (3) cohesion with neighbours.

Gandon (2002) defined three main characters for adaptivity: (1) The capability of reacting flexibly to changes; (2) taking goal-directed initiatives; and (3) the ability of learning from own experience, environment and interactions with others. It is obvious that humans are adaptive by nature but are their behaviours adaptive. A longitudinal study of adaptivity in human behaviour undertaken by Mulder (1987) answered this question and maintained the adaptivity of human behaviour.

As a result of the development of complexity science, a new scientific field of Complex Adaptive Systems (CAS) has been established with a focus on how the properties of aggregations of individuals can be determined by the characteristics and behaviour of the individuals (Railsback, 2001). ABMS has strong and direct links with CAS (Macal and North, 2005). Nilsson and Darley (2006) demonstrated the suitability of ABMS to mode complex systems and pointed out that ABMS facilitates using CAS approaches to capture and model the behaviour of each of the participants within complex systems. In summary, customer behaviour has many similar characteristics to CAS, therefore; methods and techniques of analysing and investigating CAS, such as ABMS, are needed to examine customer behaviour.

In addition to the complexity issues related to analysing and investigating customer behaviour, the rapid development of software and computing capabilities facilitate building large-scale agent-based models that would not have been achievable just a few years ago. Moreover, recent developments in the data base systems and the availability of micro-data about customers at finer levels of granularity ease the process of developing agent-based models (Macal and North, 2005). In addition, these data are invaluable resources for verification and validation of ABMS.

3.3.5 Case Study Description

This study utilises three main data sources: (a) The literature; (b) customers transactional data provided by a UK mobile operator; and (c) Ofcom, the UK telecoms regulator. The customers of the mobile operator are selected as a case study to build the data mining and the ABMS models. Thomas (2011) defined case studies as a holistic analysis of persons, decisions, events, policies, periods, projects, organisations or other systems by one or multiple methods. Generalising the case study results into a more general concept is the ultimate goal of case studies. Thomas (2011) used the object-oriented approach to describe this relationship between the case study and the phenomena or the problem under investigation. He described the case study as an instance of a class representing the phenomena or the problem under study.

The mobile operator under consideration in this case study (Network 1 in the ABMS model) is one of the mobile operators in the UK mobile market. The services of this operator cover several countries and millions of customers. This study is conducted in the context of the UK business and focuses mainly on the communication services including voice and data.

In terms of customers, this study focuses on contract customers, as the mobile market exhibits a gradual shift of customers from pay-as-you-go (PAYG) to pay-monthly contracts, plus there tends to be more available data about customers in the case of contract customers. The number of contracts was 49% of total mobile connections in 2010, compared to 41% at the end of 2009 (Ofcom, 2011a, p.260). Contract customers have an on-going financial arrangement with mobile operators (typically of 12, 18 or 24 months).

They are locked-in during the contract period and cannot switch to other networks. Once the contract period ends, customers have the option either to renew their contract or to switch to other networks. Customers constantly seek to maximise their satisfaction, including all different satisfaction factors (for example cost, coverage and customer services). If customers had a bad experience or low satisfaction with current network, they would seek alternative mobile network.

Typically, the mobile operator seeks to maximise profit and minimise the customer churn rate. Customers within the mobile market are often segmented into different groups based on their business value. The High, Mid and Low value customers are some examples of customer classification based on business value. When a customer contract is due, operators usually assess customer value for the business. Based on this assessment, operators take some retention actions (like offering new handsets or matching offers of other networks) to retain high-value customers and increase profits.

3.4 Research Methodology

After introducing the paradigms and the methods adopted in this study, this section describes the broad outline of the development research phases that were conducted to accomplish this work. This study utilises ideas from different paradigms and employs two main research techniques, data mining and ABMS, to investigate the customer retention in the UK mobile market. Rather than being tied to particular paradigm boundaries, this study adopts a cross-paradigmatic approach bounded by the DSR paradigm. Behavioural science paradigm is utilised and used as a complementary to the DSR paradigm. The relationship and interplay between the two paradigms is illustrated in Figure 3.5.

The behavioural science paradigm is utilised to build the Customer Experience (CE) model. The CE model is developed to quantify customers experience by considering all factors that may influence customer satisfaction. The CE model adopts the components of the J.D. Power framework, which provide group of factors that have proven to be important for customer satisfaction in the UK mobile market (Anaman and Lycett, 2010). The model is validated by interviewing customers as they exit a customer service centre. This model is then implemented within the ABMS during design iteration 1 by following the DSR paradigm guidelines. In ABMS design iteration 2, theories and research outcomes resulting from behavioural science studies are employed to build the social network representation.

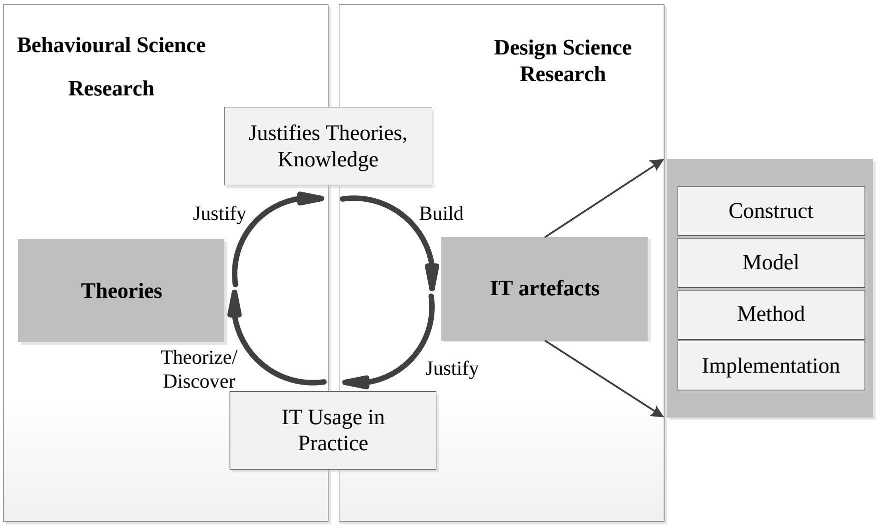


Figure 3.5: The Relationship Between DSR and Behavioural Science (Source: Niehaves and Becker, 2006)

DSR provides a theoretical framework that facilitates rigorous development of the artefacts developed in this study. The next section demonstrates how the theoretical DRS processes (introduced in Section 3.2.4) were transformed into four main practical stages: (1) Awareness of the problem; (2) solutions selection and suggestion; (3) development; and (4) evaluation. Phase two involves one development cycle resulted in two data mining models (decision tree and logistic regression). The main development phase (phase 3) consists of two iterations: (1) Modelling customer interaction with the market; and (2) modelling interactions among customers. Figure 3.6 outlines the research phases of this study. A description of each phase is provided in the next sections.

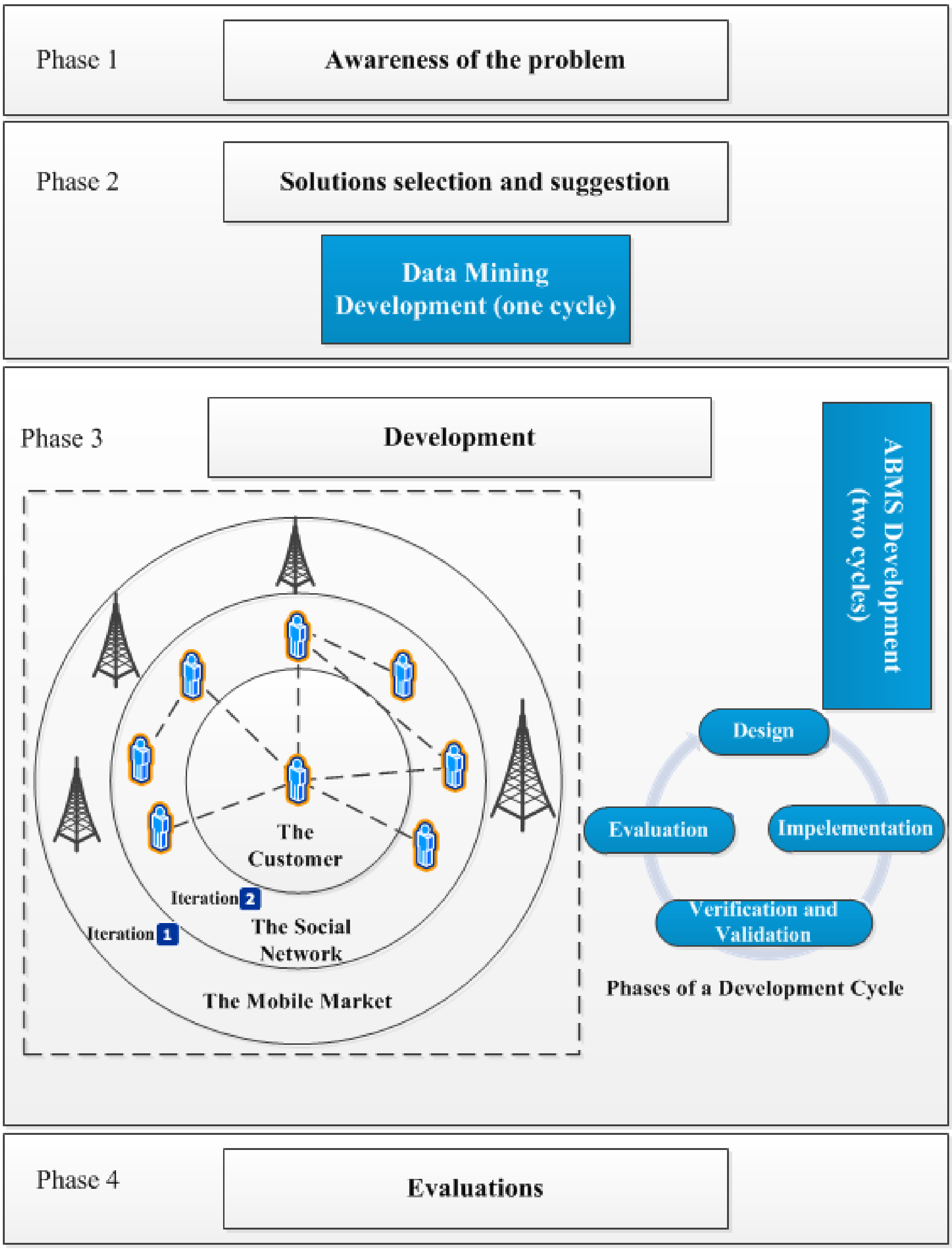


Figure 3.6: Outlines of the Research Phases

3.4.1 Awareness of the Problem

As the mobile telecommunication market becomes more competitive and the customers' expectations increases, understanding customer behaviour and needs is a key element of mobile operators' success. Customer churn analysis is one of the most pressing issues in the mobile market. Different methods and techniques have been used to investigate customer churn including statistical methods and data mining. As demonstrated in Chapter 2, traditional ways of studying and analysing customer retention are either too complex or inadequate. Customer retention contains many complicated feedback mechanisms, which are either poorly understood or extremely difficult to analyse. Moreover, customer retention encompasses a degree of fuzziness, where it represents a theoretical construct that cannot be observed directly. To get a better understanding of the customer retention, research studies need to examine all the factors affecting the customer retention simultaneously and jointly (Aydin and Ozer 2005). Therefore, new tools and methods are needed to improve and provide new insights into the customer churn problem. ABMS is proposed in this study as complementary to the traditional analysis tools.

3.4.2 Solutions, Selection and Suggestion

In this phase, the data mining development cycle is carried out to illustrate empirically how decision trees and logistic regression models are being used to investigate customer churn. The two data mining models are built based on two historical data sets related to the case study presented in Section 3.3.5. The data sets themselves are described in Section 4.4.1. The model building process follows the CRISP-DM methodology (see Section 3.3.2).

This phase focuses on leveraging insights gleaned from the data mining tools into the problem domain. In addition, the benefits and shortcomings of the decision trees and logistic regression models are identified in this phase. The limitations of these models motivate the author to propose ABMS as a new method to investigate customer retention. At the end of this phase and based on the outputs of the previous phase (awareness of the problem phase), the general requirements and specifications of the ABMS model are identified.

3.4.3 Development

This phase involves two design iterations to build the artefact proposed in the solutions selection and suggestion phase. During this phase, the ABMS model (artefact) is constructed in an iterative way through four processes: 1) Design; (2) implementation; (3) verification and validation; and (4) evaluation (see Figure 3.6, concerning the development phase).

The development of the ABMS model (CubSim model) consists of two iterations: (1) Modelling customer interaction with the market; and (2) modelling interactions among customers. The development processes in Figure 3.6 (Design Phase) are repeated in each iteration. To improve the rigor and the utility of the ABMS model, the development processes follow Rand and Rust's (2011) guidelines. Rand and Rust (2011) divided ABMS model development into four large-scale steps: (1) Decide if ABM is appropriate; (2) design the model; (3) construct the model; and (4) analyse the model. In addition, they provide a set of guidelines to maintain scientific rigor of the ABMS model. Table 3.1 lists Rand and Rust's (2011) guidelines, along with supporting evidence that demonstrates how these guidelines have been met in this study.

ABMS Development Iteration 1

In this iteration, a generic, reusable agent-based model (CubSim) is developed to examine the factors affecting customer retention in the UK mobile telecommunication market. The model development process involves creating artificial agents mimicking attributes and behaviours of their real-world counterparts. The focus in this iteration was on modelling customer interaction with the market. The decisions of the models' agents are based essentially on evaluating the satisfaction resulting from the interaction with their service provider. Based on the satisfaction level and interaction with other service provider in the market, agents take their decision on switching their network provider.

ABMS Development Iteration 2

The aim of this iteration is to extend the CubSim model by incorporating social networks. The extension process involves three main tasks: (1) Modelling the social network; (2) modelling the WOM flow; and (3) extending the agent behaviours to incorporate the social network effects. The decisions of the agents in this iteration are based on combining the satisfaction level and the social interaction with family and friends. Based on the amalgamation of interactions among all channels, including social network effect, agents take their final decision on switching their network provider.

3.4.4 Evaluation

In order to demonstrate the utility of DSR artefacts, an effective and rigorous evaluation process should be conducted. The evaluation process is very significant aspect of DSR because it establishes validity and reliability artefacts. In addition, it can generate knowledge that can lead to a deeper understanding of the problem domain and to the improvement of the quality of the artefacts themselves. Hevner et al. (2004) identified a set of evaluation methods to evaluate the quality and the effectiveness of artefacts (see Table 3.3).

To achieve the objective 4 of this phase, a systematic and comprehensive methodology to evaluate the study artefacts is applied. Different design evaluation methods are used in the three design cycles. In the first design cycle (data mining development), three evaluation metrics are used to evaluate the performance of the decision tree and the logistic regression models. These metrics are lift, area under curve and overall accuracy (see Section 4.3 for more details). In the second and third development cycles (ABMS development) a verification and validation process is performed. Similar to the first development cycle, a variety of different evaluation methods are used (e.g. statistical analysis, sensitivity analysis and black box testing) (see Section 5.6 and Section 6.5 for more detail).

<table>
<tr><th>Step</th><th colspan="2">Guideline focus</th><th>How these guidelines have been met</th></tr>
<tr><td>1</td><td colspan="2">Decide if ABM is appropriate</td><td>See Section 3.3.4 for discussion of this.</td></tr>
<tr><td rowspan="7">2</td><td rowspan="7">Design the Model</td><td>Scope of the Model</td><td>This study focuses on modelling contract customers within the UK mobile market based on a case study described in Section 3.3.5.</td></tr>
<tr><td>Agents</td><td>The CubSim model contains one agent type, customer agents, which represent mobile users.</td></tr>
<tr><td>Properties</td><td>Agent properties are described in Section 5.3.1.</td></tr>
<tr><td>Behaviours</td><td>The behaviours of the agents are based on a novel modelling approach combining a micro-economic utility model of customer satisfaction with a social interaction model. This approach provides balance between reactive decision and utility maximisation (see Section 5.3.3 for more details).</td></tr>
<tr><td>Environment</td><td>The agent's surrounding environment includes mobile operators and other agents in the market. The agent's social networks are built based on the social circles concept, which is described in section 6.3.1.</td></tr>
<tr><td>Input and Output</td><td>The input and output of the CubSim model are described in Section 5.4.3 and Section 5.6, respectively.</td></tr>
<tr><td>Time Step</td><td>Once the simulation session started, the agents start interacting with the surrounding environment. Agents take decisions at each simulation time step, which represents one month of real-world time.</td></tr>
<tr><td rowspan="4">3</td><td rowspan="4">Construct the Model</td><td>Initialisation</td><td>The initial customer pool size is set to 4000 agents. Section 5.6 illustrates how sensitivity analysis is carried out to identify the optimal number of agents. Other input parameters extracted from the data sources as explained in Section 5.5.</td></tr>
<tr><td>Adoption Decision</td><td>Agents make decisions at each simulation time step according to the mechanism described in Section 5.3.3.</td></tr>
<tr><td>Statistics Collection</td><td>The numbers of satisfied, unsatisfied and churned customers are recorded at each time step.</td></tr>
<tr><td>Repeat</td><td>Every simulation run comprised of 48 time steps, which represent a time window of four years.</td></tr>
<tr><td>4</td><td colspan="2">Analyse the model</td><td>Once the CubSim model has been verified and validated, it is used as an experimental tool to investigate different scenarios. Section 5.6 and Section 6.5 report some of the simulation experiments performed on the CubSim model.</td></tr>
</table>

Table 3.1: ABMS Development Guidelines (Source: Rand and Rust, 2011)

3.4.5 Application of the DSR Guidelines

Hevner et al. (2004) compiled a set of guidelines to provide rigor and relevance to DSR projects. As presented earlier, the development phase of this study incorporates Rand and Rust's (2011) guidelines to improve the rigor and the utility of the ABMS model (the main artefact of this study). More generally, this section illustrates the application of Hevner et al. (2004) guidelines to the whole research process in this study.

Guideline 1: Design an Artefact

Artefacts are the main outcomes of the DSR projects, and their roles are very essential in DSR. March and smith (1995) classified the DSR artefacts into four types: construct, model, method, and instantiation. Table 3.2 lists and explains the DSR artefacts types, along with matching the artefacts produced in this study with these types.

Artefact Type	Description	This Study Artefacts
Constructs	Constructs are conceptual vocabularies and symbols that provide a language to identify and share design problems and solutions.	None
Models	Models are design constructs to conceptualise and abstract problems and their solutions to improve understanding.	The CubSim model
Methods	Methods are defined processes that facilitate finding solutions. The methods can be formal (such as algorithms) or informal (such as focus groups), or a mix of both.	The procedures followed to develop the CubSim model are summarised in Table 3.1.
Instantiations	Instantiations are the implementations of constructs, model and methods in the form of working systems to provide a proof of the concept of the solution.	The Customer Behaviour Simulator built to examine and evaluate the CubSim model (see section 5.4.1).

Table 3.2: Design Research Artefacts Developed in this Study

Guideline 2: Problem Relevance

Section 1.2 provides information on the research problem and discusses the motivation for this study. Section 3.4.1 underlines problem relevance and highlights the need for new retention analysis tools.

Guideline 3: Design Evaluation

There are a variety of different research evaluation methods used in this study. Following the set of evaluation methods that identified by Hevner et al. (2004), Table 3.3 lists and describes these methods, and match the methods used in this study with these methods.

Guideline 4: Research Contributions

The design-science contributions of this research are the CubSim model and the design principles. The CubSim model demonstrates the feasibility of using ABMS to develop an artefact to model and study customer retention. The CubSim model is the first artefact to address customer retention in the mobile market; therefore, its development process is itself a contribution to design science. Furthermore, examples of using the model to provide insights into customer churn are provided. Using the model to generate new insights can be tested by future work. Detailed discussion of the contributions and values of this research is presented in Section 7.3, followed by an agenda for possible future work in Section 7.5.

Guideline 5: Research Rigor

To ensure the rigor of this study, two sets of best practice guidelines have been integrated into its design and research processes. These guidelines are employed in both the construction and evaluation of the designed artefacts. The first guidelines set focus on developing the ABMS model (See Table 3.1) (Rand and Rust, 2011). The second guidelines set cover the whole design and research process (Hevner et al., 2004). Section 3.4.5 demonstrates the application of these guidelines in the context of this study.

Guideline 6: Design as a Search Process

Utilising best practice guidelines in developing the artefacts not only provides structured and systematic approach to conduct this study, but also facilitates the search process for the best solution. In contrast to many other studies, this study seeks to model the problem rather than directly working with the problem itself. The modelling process is inherently iterative and developing the models is essentially a search process to produce the best representation for the problem. The iterative design process followed in this study contributes to improve the designed artefacts. In addition, it enhances the research-learning process.

Evaluation Method	Description	Method used in this study
1.Observational	Case Study: Study artefacts in depth in business environment	None
	Field Study: Monitor use of artefacts in multiple projects	None
2. Analytical	Static Analysis: Examine structure of artefacts for static qualities (e.g., complexity)	- One-way ANOVA statistical test is used to find the optimal number of agents for the CubSim model (see Section 5.6). - A Paired sample T-test is used to determine the effect of modelling the annual market growth on the CubSim performance measurements.
	Architecture Analysis: Study fit of artefacts into technical IS architecture	None
	Optimisation: Demonstrate inherent optimal properties of artefact or provide optimality bounds on artefacts behaviour	None
	Dynamic Analysis: Study artefacts in use for dynamic qualities (e.g., performance)	None
3. Experimental	Controlled Experiment: Study artefacts in controlled environment for qualities(e.g., usability)	None
	Simulation :Execute artefacts with artificial data	Simulation is used in all experiments performed on the CubSim model. Simulation experiments involve executing the CubSim model multiple times to mimic and imitate real scenarios.
4. Testing	Functional (Black Box) Testing: Execute artefacts interfaces to discover failures and identify defects	Black box testing is used to compare one of the performance measurement of the CubSim model (annual churn rate) with the annual churn rate released by the mobile operator (see Section 6.5).
	Structural (White Box) Testing: Perform coverage testing of some metric (e.g. execution paths) in the artefacts implementation	None
5. Descriptive	Informed Argument: Use information from the knowledge base (e.g., relevant research) to build a convincing argument for the artefacts utility	None
	Scenarios: Construct detailed scenarios around the artefacts to demonstrate its utility	This method is used to demonstrate the utility of the CubSim model when it used as what-if scenario tool. A new customer retention strategy considering customers social values in addition to their financial values is proposed and tested (Section 6.5).

Table 3.3: Design Research Evaluation Method Used in this Study

Guideline 7: Communication of Research

This study is targeted both technology-oriented and management-oriented audiences. To communicate the study findings to the technology-oriented audiences, detailed description of the artefacts and their design and implementation processes are provided. This detailed description enables practitioners to understand the processes by which the artefacts were developed and evaluated. In addition, this paves the way to further extension and evaluation of artefacts.

The ultimate goal of this study is not only offer decision makers a tool to better understand the customer churn problem, but also help them to explore different customer retention strategies and their probable consequences. To satisfy this goal, an emphasis placed on the importance of the customer churn problem and the novelty and effectiveness of the CubSim model as a retention analysis tool. Clear demonstration of how the CubSim model works and how it can be used is provided to enable decision makers to effectively apply this model.

3.5 Summary

This chapter describes the research approach adopted to conduct this study. It provides a road map for other researchers to follow when replicating the models developed in this study, and it offers a basis for a critical review of the research methodology. Due to the multidisciplinary nature of this study, a triangulation of multiple data sources is used to construct the artefacts developed. The research methodology adopted in this study is DSR, which involves the construction and evaluation of artefacts that resolve the customer retention problem. Data mining and ABMS are used as the main research tools in this study. The selection of these tools along with the DSR methodology adopted in this study is justified and clearly described.

Due to the central role of artefacts in DSR, the artefacts developed in this study are discussed and classified. In response to increasing calls for rigor and utility, this chapter concludes with illustrations of how this study has incorporated two sets of best practice guidelines to improve the utility of the ABMS model and to ensure the rigors and the validity of this study as a whole. With the DSR methodology in action, the research proceeds into the development of the data mining models.

Chapter 4: Traditional Churn Modelling Approaches

4.1 Overview

This chapter presents phase two of this study, which focuses on exploring and evaluating current customer retention analysis tools. During the analysis and design in this phase, further insights were acquired into the problem domain. In addition, the general specifications of the required solutions are defined in this chapter. Through empirical experimentation, this chapter compares the two widely used techniques of analysing customer churn: Decision tree and logistic regression. This is followed by a discussion of the limitations of these tools and the need for other tools and techniques to overcome the identified limitations.

This chapter is structured as follows: Section 4.2 provides background information on the decision tree and logistic regression modelling techniques. Section 4.3 describes some of the performance evaluation metrics for these techniques. Section 4.4 presents the modelling experiments and results along with the models' evaluation and limitations. Finally, Section 4.5 provides an overall summary of the chapter.

4.2 Churn Modelling Techniques

As noted in Section 2.4, different methods and techniques have been used to investigate and analyse customer churn. Data mining is a common method used in the context of customer churn to predict future customer behaviours and trends by detecting patterns and relationships within raw data. Regression analysis and decision tree are two of the most popular data mining techniques (Hung, Yen and

Wang, 2006). Because of the wide use of these two techniques in both academia and business (see Section 2.4.1 for relevant literature), they are selected for further empirical analysis. The next two sections discuss the decision tree and logistic regression methods.

4.2.1 Decision Tree

Decision trees are the most popular predictive models (Burez and Van den Poel, 2007). A decision tree is a tree-like graph representing the relationships between a set of variables. Decision tree models are used to solve classification and prediction problems where instances are classified into one of two classes, typically positive and negative, or churner and non-churner in the churn classification case. These models are represented and evaluated in a top-down manner. Developing decision trees involves two phases: Tree building and tree pruning.

Tree building starts from the root node that represents a feature of the cases that need to be classified. Feature selection is based on evaluation of the information gain ratio of every feature. Following the same process of information gain evaluation, the lower level nodes are constructed by mimicking the divide and conquer strategy. Building a decision tree incorporates three key elements:

1- Identifying roles at the node for splitting data according to its value on one variable or feature.
2- Identifying a stopping rule for deciding when a sub-tree is created.
3- Identifying a class outcome for each terminal leaf node, for example, ‘Churn’ or ‘Non-churn’.

Decision trees usually become very large if not pruned to find the best tree. The pruning process is utilised not only to produce a smaller tree but also to guarantee a better generalisation. This process involves identifying and removing the branches that contain the largest estimated error rate and can be regarded as an experimentation process. The purpose of this process is to improve predictive accuracy and to reduce the decision tree complexity (Au, Chan and Yao, 2003). Once the model is built, the decision about a given case regarding to which of the two classes it belongs is established by moving from the root node down to all the leaves and interior nodes. The movement path is determined by the similarity calculation until a leaf node is reached, at which point a classification decision is made.

The representation of decision tree models is best explained with an example; Figure 4.1 shows a decision tree. The target variable is churn status (yes or no) and the input variables are Age, Gender and Tenure. To demonstrate a classification decision process, suppose that a customer has the following features: (Age = 32, Gender = M, Tenure = 12). The classification decision processes starts from the root by checking the age then moves down according to the outcomes from the internal nodes. The movement path through nodes is highlighted in Figure 4.1. Classification rules can be extracted easily from the decision tree. These rules are the paths to get from the root node to each leaf. The classification rule used in the previous case is given below.

If (Age >28, Gender= M, and Tenure <18), then Churn=Yes.

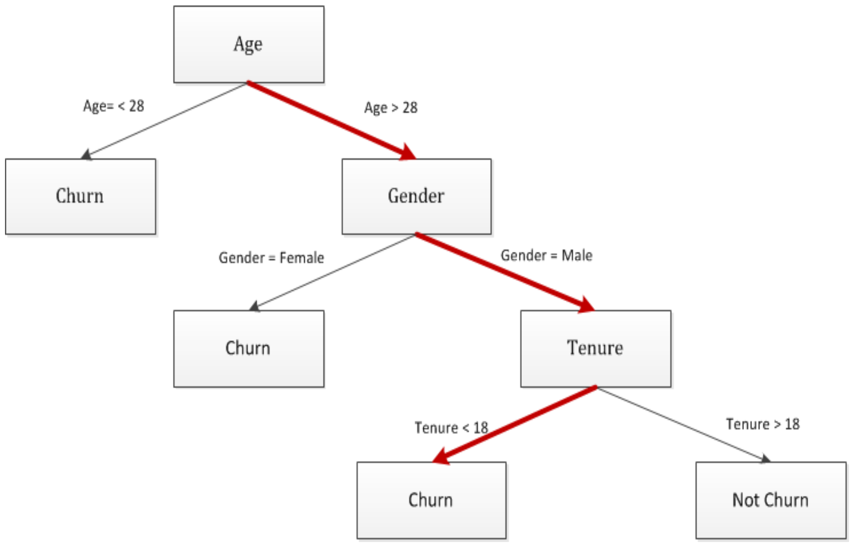


Figure 4.1: Classification Decision Processes Based on a Decision Tree

There are many algorithms for building a decision tree, including CART, C5.0, and CHAID. The CART (Classification and Regression Trees) algorithm is used to predict the values of both continuous and discrete variables. Using CART to predict continuous variables is called regression tree, while using it to predict discrete variables is called classification tree. In regression trees, a finite list of possible splits based on the different values that the variables take in the data set are identified as stopping rules. In classification trees, the stopping rule is satisfied when the final nodes are sufficiently pure and contain only one type of data (for example, 'Churn' or 'Non-churn'). Once the best split is found in regression trees or classification trees, CART searches again for each node below child nodes, until further splitting is stopped by a stopping rule, or splitting is impossible.

C5.0 is very similar to CART in that they both deal with continuous variables in much the same way. C5.0 performs all the possible tests, which can split a data set and selects the one that leads to the highest information gain. C5.0 differs from CART in that it deals with categorical variables in a different way. When C5.0 selects a categorical field's value as a splitter, the number of branches created will be equal to the number of values taken on by that variable (Bloemer et al., 2003).

Unlike CART, splitting rules in CHAID (Chi-Squared Automatic Interaction Detection) are based on the Chi-square test, to determine the best split at each node. CHAID searches at each node for the split point that has the smallest adjusted p-value (probability value). If the adjusted p-value is smaller than a predefined threshold, then the node is split. The splitting process continues until all the p-values of the split variables in the non-split nodes exceed the predefined threshold.

Decision trees have many apparent advantages. They are easy to understand and visualise. Moreover, the decision process can be simplified to a set of business rules as demonstrated earlier (Bakır et al., 2009). It is a nonparametric method; therefore, no prior assumptions about the data are needed. In addition, decision trees are able to process both numerical and categorical data. However, despite the apparent advantages of decision tree, some disadvantages are also present. Complex interactions among variables and attributes can affect the performance of decision trees. In addition, it becomes very complicated and difficult to visualise and interpret complex decision trees. Other reported disadvantages of decision trees are their lack of robustness and their over-sensitivity to training data sets (Burez and Van den Poel, 2007).

4.2.2 Logistic Regression

Logistic regression is one of applied statistics and data mining techniques that have been used extensively in customer churn studies (e.g. Ahn, Han and Lee, 2006; Burez and Van den Poel, 2007). Logistic regression utilises mathematically oriented techniques to analyse the influence of variables on the other variables. Logistic regression is used in customer churn analysis to predict the probability of customer churn occurrence. This prediction can be made by establishing a set of equations and relating input field values (factors affecting customer churn) with output field (probability of churn). The equations below present the standard mathematical formulas for a logistic regression model (Nisbet et al., 2009, p.250)

.

Where:

- y is the target variable for each individual j (customer in churn modelling). y is a binary class label (0 or 1)

- is a constant parameter
- is the weight given to the specific variable associated with each customer j (j=1,.... m)
- are the value of the predictor variables for each customer j, from which y is to be predicted.

To use logistic regression in customer churn analysis, customer data sets need to be analysed to establish the regression equation. Once the logistic regression equation is obtained, an evaluation process for each customer in the data set is performed. If the p-value for a customer is greater than a predefined threshold (e.g. 0.5), this customer is predicted to be at risk of churn. An example of predicting customer churn by using logistic regression is shown below.

A customer data set contains the following features: Target variable (Churn) and predator variables (Age, Gender and Tenure). Table 4.1 shows records of five customers including target and predator variables values, in addition to the p value, for each customer.

Age	Gender	Tenure	p	Churn
28	M	12	0.75	Yes
40	M	10	0.83	Yes
30	F	20	0.90	Yes
60	F	10	0.66	Yes
35	M	20	0.44	No

Table 4.1: Customer Records and their Probability of Churn

Regression analysis needs to be handled cautiously, because the results of such analysis can easily give misleading results, especially when carrying out causality and impact assessment (Cook and Weisberg, 1982). Logistic regression is a linear regression model to model linear systems. Under certain circumstances, logistic regression can be used to approximate and represent nonlinear systems (Cook and Weisberg, 1982). However, with the ever-increasing complexity of the mobile telecommunication market, the appropriateness of logistic regression needs to be evaluated (Twomey and Cadman, 2002).

4.3 Performance Evaluation Metrics

There are several customer churn modelling methods. Some of these methods were presented earlier in this chapter. A critical issue in using different churn modelling methods is how to assess efficiently the performance of these methods. Another crucial issue of model evaluation is how to benchmark and compare the relative performance among competing models. This section discusses and addresses some of the performance evaluation metrics and models comparison methods.

4.3.1 Classification Accuracy

One of the tools that can be used to measure the performance of binary classification models is the confusion matrix (also called contingency table). A confusion matrix is a visual representation of information about actual and predicted classifications produced by a classification model (Nisbet et al., 2009, p.292). Table 4.2 shows a confusion matrix for a binary classifier.

	Predicted classes		
Actual classes		**Class=Yes/+/ Churn**	**Class=No/-/No-churn**
	Class=Yes/+/ Churn	TP (true positive)	FN (false negative)
	Class=No/-/ No-churn	FP (false positive)	TN (true negative)

Table 4.2: A Confusion Matrix for a Binary Classifier

Different accuracy metrics can be derived from the confusion matrix. These metrics are classification accuracy, sensitivity and specificity. Classification accuracy (CA) is the percentage of the observations that were correctly classified. CA can be calculated from the confusion matrix by using the following equation (Vuk and Curk, 2006):

$$CA=$$

Classification accuracy alone can be a misleading indicator of classifier quality, especially in the case of imbalanced data. To illustrate this weakness, consider this example: A customer base contains 9,990 churn customers and 10 non-churn customers. If a model predicts that all 10,000 customers are at risk of churn, the classification accuracy will be 9,990/10,000 (99.9%). This high accuracy rate indicates that the model is very accurate in predicting customer churn, but this is not true because the model does not detect any non-churn customers. Correctly predicting churn cases is always more important than predicting non-churn cases because the cost of mis-predicting churn is substantially higher than the cost of mis-predicting non-churn. The cost of losing customers incorrectly predicted to be non-churners is more than the cost of the incentive efforts directed to customers incorrectly predicted as churners.

4.3.2 Sensitivity and Specificity

Sensitivity and specificity measurement tools can overcome some of the weakness of the accuracy metric. The definitions of sensitivity and specificity and how they can be calculated from the confusion matrix are given below (Verbeke et al., 2010).

Sensitivity is the proportion of actual positives that are correctly identified:

$$Sensitivity=$$

Specificity is the proportion of actual negatives that are correctly identified:

Specificity=

In the case of churn analysis, sensitivity is the percentage of churners that are correctly predicted, while specificity is the percentage of non-churners that are correctly predicted. Mobile operators favour models with high sensitivity over models with high specificity. This is because, as demonstrated earlier, the cost associated with the incorrect classification of churners is substantially higher than the cost associated with the incorrect classification of non-churners. Clearly, specificity should not be ignored totally. Typically, a trade-off between high specificity combined with reasonable sensitivity should be made. By doing so, mobile operators can effectively manage their marketing budget to achieve high customer retention.

4.3.3 Receiver Operating Characteristic Curve (ROC)

The ROC curve is a visual representation of the relationship between true positive rate and false positive rate or between sensitivity and (1 − specificity). It is plotted on the linear scale of x and y axis (Karahoca, Karahoca and Aydin, 2007). In churn classification, ROC represents the relationship between the percentage of churners predicted correctly as churners, and the percentage of non-churners predicted wrongly as churners. The ROC provides insight into relative trade-offs between true positive (benefits) and false positive (costs). The ROC curve is composed of points, each point corresponding to a prediction result or one instance of a confusion matrix. Figure 4.2 shows an example of a ROC curve.

The best performance model is that model in which the ROC curve passing through or close to (0, 1). In this case, the model sensitivity will be 100% (no false negatives) and the model specificity will be 100% (no false positive). Some of the models, such as logistic regression models, do not produce binary class decisions (i.e. churn or non-churn); instead, they produce a ranking or scoring. In this case, thresholds need to be used to produce a binary classifier. If the classifier output is greater than a threshold, the classification class will be churn, otherwise it will be non-churn. Varying the threshold produces different points in the ROC curve. The diagonal line in Figure 4.2 that divides the ROC space into two parts represents the ROC curve for a random predictor. ROC curves passing near this line correspond to random guessing classifiers (e.g. classifying by tossing a coin).

In general, models with ROC curves passing through the top left part of the ROC curve have better performance. The area under the ROC curve (AUC) is also used as a performance metric (Karahoca, Karahoca and Aydin, 2007). This area is equal to the probability that the predictive model will correctly distinguish between instances of churn and non-churn. The AUC value varies from 0.0 to 1.0, and models with greater AUC are usually considered having better performance. Models that perform better than random models should have an AUC value greater than 0.5, because the area under the ROC diagonal line is 0.5.

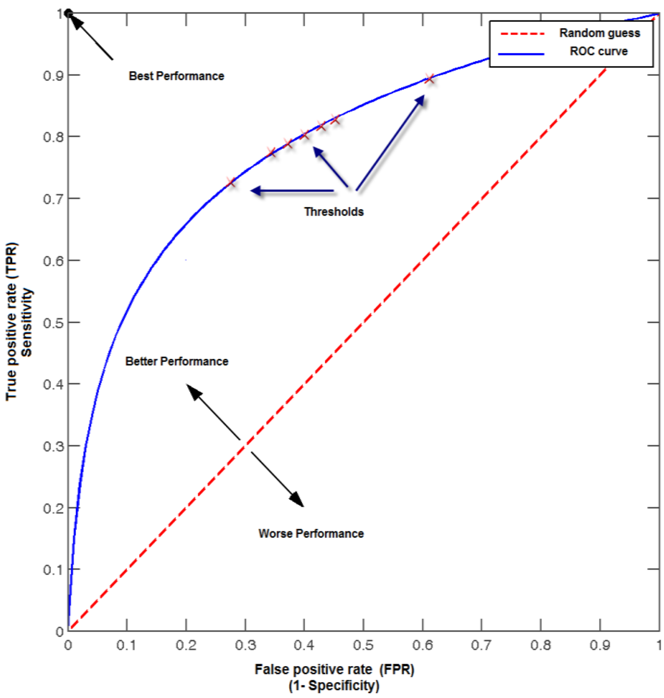


Figure 4.2: An Example of ROC Curve

4.3.4 Lift Chart

The Lift chart is similar to the ROC curve in that it provides means of evaluating models' performance and determining thresholds, which yield high true positive cases. In some cases, like predicting potential customers of a product, the total numbers of the classification instances are unknown and the true positive TP rate cannot be computed. In these cases, the ROC curve cannot be used for performance evaluation. Instead, the Lift chart can be used for measuring the models' accuracy (Vuk and Curk, 2006).

In customer churn, the Lift chart groups customers into deciles based on their predicted probability of churn. Each decile demonstrates the models' performance in predicting customer churn. Figure 4.3 shows an example of lift chart for a customer churn model. This lift chart shows that the top two deciles capture about 50% of churners, while the top five deciles captured about 90% of churners. Grouping customers according to their relative probability to churn makes Lift charts the preferred performance evaluation metric among marketers. Targeting customers in the top deciles rather than targeting all customers will keep the marketing cost at a minimum while simultaneously leading to high customer retention rate.

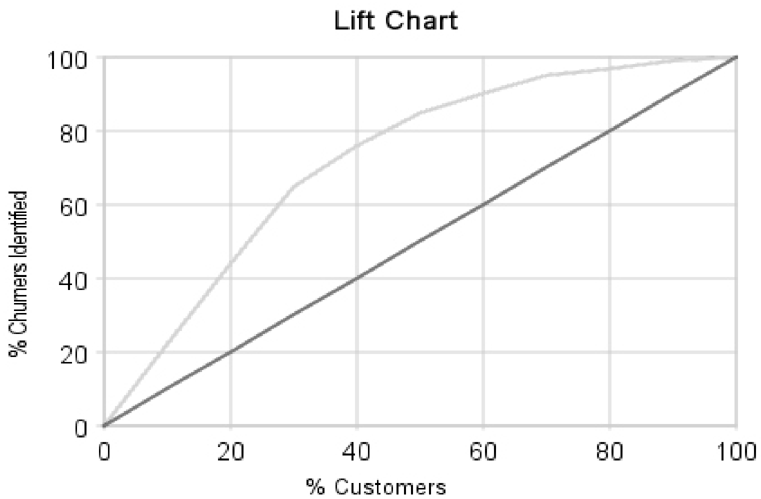


Figure 4.3: A Lift Chart of Customer Churn Model

4.4 Modelling Experiments and Results

This section reports churn analysing and predicting experiments by using two modelling methods: Logistic regression and decision tree. In these experiments, the well-known data mining methodology CRISP-DM was adopted to investigate customer churn in the telecommunications industry (Chapman et al., 2000). The CRISP-DM methodology provides complete and high-level instructions and procedures for applying data mining algorithms to solve real-world problems (see Section 3.3.2 for more details).

4.4.1 Data Sets Description

This section sheds light on activities involved in the data understanding phase. Two data sets were used to demonstrate how traditional churn analysis methods could be used to investigate churn. The data sets were extracted from a UK mobile telecommunication operator data warehouse. The analysis is based on two datasets of 15,519 and 19,919 customers with 23 variables. The dependant variable (output variable) is whether the customer churned or not. The predictor variables (input variables) are listed in Table 4.3. Some of the predictor variables are categorised into five groups: Demographics, cost, features/marketing, services usages, and customer services. Table 4.3 provides a brief description of some predictor variables. Both data sets contained 50% of customers who churned and 50% who stayed with the operator. The data sets contain contract customers on 12- and 18-month contracts. The data sets reflect the entire customer population base in all aspects.

The data sets cover customers' data from March 2008 through to December 2008, at which point the customers took the decision of churn. For the new or upgraded 12-month customers, the data were collected from month 4 to month 9. For those on 18 month contracts, the data were from month 9 to month 15. Figure 4.4 illustrates the time window of the data sets that were used in this study.

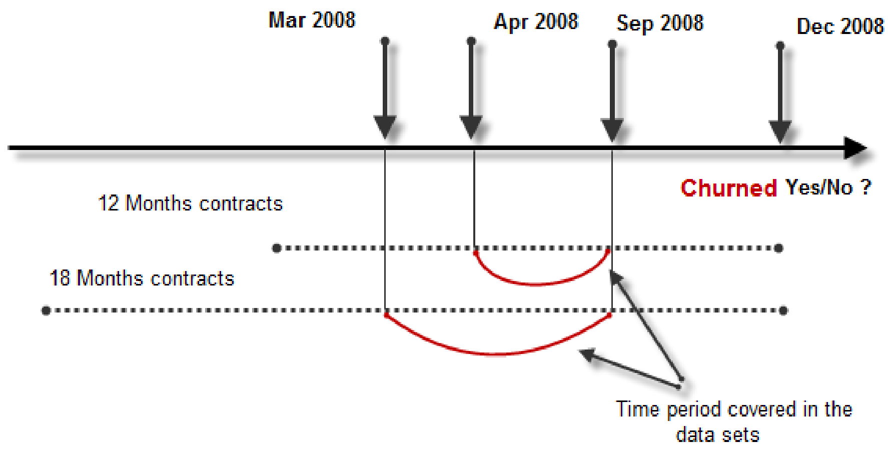


Figure 4.4: Time Window of Data Sets

Category	Variable Name	Description
Demographics	Lifestage_Segment	Subscribers' age stage and gender
	Gender	Subscribers' gender
	Post_Code	Post code in which subscribers live
Cost	Package_Cost	Cost of the package of services chosen by subscribers
	Contract_Length	Number of months of the contract
	Tenure	Number of months with the present mobile operator
Features/ Marketing	Tariff	The package of services chosen by subscribers
	Device_Desc	Handset model and manufacturer
	sales_channel	The first channel where the relationship with the customer was established
Usage Level	Q2_bytes	Data usages in the second quarter
	Q3_bytes	Data usages in the third quarter
	Q2_voice	Voice usages in the second quarter
	Q3_voice	Voice usages in the third quarter
Customer Services	No_of_Repairs	Number of times handset has been in for repair in a 12-month period
	Prob_Handset	Known issues with existing handset
	No_of_Complaints	Number of customer complaints regarding billing in a 12 month period

Table 4.3: Description of Data Sets' Variables

4.4.2 Data Preparation

The data preparation phase covers all activities to produce the final data sets for models building. The primary goal of data preparation is to enhance data quality to improve the performance of data analysis. Data preparation is not a one-time job, and each modelling algorithm has different requirements. Therefore, data preparation needs to be carried out in a more iterative manner until a conclusive outcome is reached.

In this study, data preparation activities include:

- Imputation of the missing values
- Discretisation of numerical variables
- Transformation from one set of discrete values to another
- Feature selection of the most informative variables
- New variable derivation

Imputation process involves replacing an incomplete observation with complete information based on an estimate of the true value of the unobserved variable. Discretisation and transformation are used to create new variables. Two new variables were created to measure voice and data usage change. The performance results of different modelling experiments with different feature lists demonstrate that models with selected features outperform models with full feature sets. As a result, despite the fact that the data sets used in this research have few variables, a feature selection process was carried out. Feature selection process as applied in this study involves three steps:

1. Feature screening by removing unimportant and unreliable variables. In this step, two variables (problem_handset and no_of_complaint) were removed from the feature list because they exhibited very little variation.
2. Feature ranking by sorting the remaining variables based on their importance and correlation with the dependant variable.

3. Feature selection by choosing the most important variables and filtering out all others.

4.4.3 Modelling

In this phase, various modelling techniques were applied, and their parameters were calibrated to optimal values. Two modelling techniques (logistic regression and decision tree) were selected for detailed analysis. Logistic regression and decision tree were chosen because of their widespread use both in academia and in industry. In addition, the two modelling techniques are relatively simple and have the ability to explain the relationship between input and output variables. IBM SPSS Modeller (formerly Clementine) was used to build and experiment with different modelling techniques.

4.4.4 Logistic Regression Analysis

A good model generally has 25% higher classification accuracy rate than the proportional by chance accuracy rate. The logistic regression analysis was initiated by computing the proportional by chance accuracy rate. This rate was computed by calculating the proportion of cases for each group (churn or non-churn). Based on the number of cases in each group in the classification table at Step 0, the proportions of the “non-churn” and the “churn” groups are shown in Table 4.4. Then, the proportional by chance accuracy rate was computed by squaring and summarising the proportion of cases in each group. The logistic regression accuracy rates, as shown in Table 4.4, are 25% higher than the proportional by chance accuracy rate. Therefore, the criteria for classification accuracy are satisfied, and the logistic regression model performs better than a random guess.

The presence of a relationship between the dependent variable (probability of churn) and combination of independent variables is verified by checking the statistical significance of the model chi-square at step 1, after the independent variables have been added to the analysis. The probabilities of the models' chi-squares for both data sets were less than or equal to the level of significance of 0.05. Thus, the null hypothesis that there is no difference between the model with only a constant and the model with independent variables was rejected. In addition, the existence of a relationship between the independent variables and the dependent variable was supported.

Data set	Proportion of churner	Proportion of non-churner	Proportional by chance	Logistic regression model accuracy rate
Dataset 1	0.5277	0.47220	0.5176	0.681
Dataset 2	0.499	0.5033	0.5023	0.642

Table 4.4: The Proportional by Chance Accuracy Rate

Multicollinearity is a concerning issue in logistic regression models (Nisbet et al., 2009, p.265). Multicollinearity is a result of strong correlations between independent variables. The existence of strong multicollinearity inflates the variances of the parameter estimates. Multicollinearity may also lead to wrong estimates for signs and the magnitude of the regression coefficient, which in turn can lead to incorrect conclusions about relationships between independent and dependent variables. Multicollinearity in logistic regression solution is detected by examining the standard errors for the B coefficients. A standard error larger than 2.0 indicates numerical problems, such as multicollinearity among the independent variables (Nisbet et al., 2009, p.265). None of the independent variables in this analysis had a standard error larger than 2.0. Therefore, multicollinearity is not a concern in this study. Since the analysis of the two data sets gave very similar results, we present the results of the first data set.

The built model for logistic regression is shown in Figure 4.5. The stream represents data operations that need to be performed on the raw data. Each operation is represented by an icon or node, and the nodes are linked together in a stream representing the flow of data through each operation. Table 4.5 shows the classification table for the best performing logistic regression model based on the classification cut-off of 0.5 and 95% confidence intervals (CI for exp (B)). The classification table is a table that records correct and incorrect estimates for the full model of all independents and constants.

Observed			Predicted		
			Flag		Correct Percentage
			No	Yes	
The Last step	Flag	No	5183	1294	80.0
		Yes	2175	2226	50.6
	Overall Percentage				68.1

Table 4.5: The Logistic Regression Classification Table

Table 4.6 lists the B parameter, Wald statistics, degree of freedom, significance levels and odd ratio. The Wald statistics and the significance level corresponding to it are used to test the statistical significance of each coefficient (B). If the Wald statistic is significant (i.e. less than 0.05) then the variable is significant in the model. Exp (B) values in Table 4.6 represent the results of a statistical test performed to assess the risk of customer churn. Exp (B) is the predicted change in odds for a unit increase in the corresponding independent variable. Odds ratio of less than 1 indicates a decrease in odds. Odds ratio above 1 indicates an increase in odds. Odds ratio of 1.0 indicates that the independent variable does not affect the dependant variable.

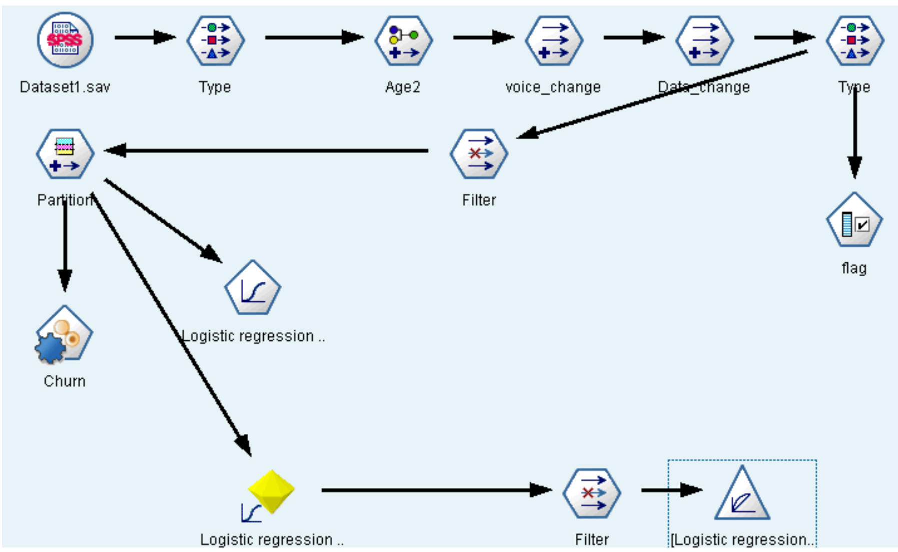


Figure 4.5: Data Stream for the Logistic Regression Model

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	CNTR_LENGTH	.084	.009	79.717	1	.000	1.087
	Gender	.050	.050	.972	1	.324	1.051
	no_of_repairs	-.220	.072	9.282	1	.002	.803
	Cost	-.002	.002	1.343	1	.247	.998
	billing_queries	-.057	.062	.830	1	.362	.945
	Tenure	-.026	.003	97.216	1	.000	.974
	Coverage	-.289	.052	30.309	1	.000	.749
	chang_data_usage	-.159	.048	10.781	1	.001	.853
	chang_voice_usage	-.943	.081	133.980	1	.000	.390
	bundle_avr_usage	-.106	.069	2.358	1	.125	.900
	Constant	.853	.247	11.946	1	.001	2.347

Table 4.6: Logistic Regression Analysis Results of the First Step

All of the variables with higher significance level (more than 0.05) are eliminated from the model. Then, the logistic regression model is applied progressively through several iterations to eliminate all non-significant variables. Table 4.7 shows all the significant variables, which were included in the final logistic model.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 6	CNTR_LENGTH	.085	.009	85.360	1	.000	1.089
	no_of_repairs	-.217	.072	9.088	1	.003	.805
	tenure	-.027	.003	105.587	1	.000	.974
	Coverage	-.292	.052	31.086	1	.000	.747
	chang_data_usage	-.162	.048	11.193	1	.001	.850
	chang_voice_usage	-.973	.079	152.903	1	.000	.378
	Constant	.735	.217	11.461	1	.001	2.085

Table 4.7: Logistic Regression Analysis Results of the Last Step

4.4.5 Decision Tree Analysis

A decision tree model was created to investigate customer churn behaviour (see Figure 4.6 for the decision tree data stream). Three decision tree algorithms

(CART, C5.0 and CHAID) were used to evaluate and compare their performance. The decision tree accuracy rates were 68.57 percent for CART model, 68.83 percent for CHAD model and 70.25 percent for C5 model. Since C5 has the best predictive performance among the three decision tree prediction models, its results are discussed in more detail. Figure 4.7 shows the decision tree constructed using C5 model. The most important independent variables (as shown by the top three levels of the decision tree) are CNTR_LENGTH, chang_voice_usage, and tenure. Ranking of the independent variables based on their importance is consistent with the logistic regression results.

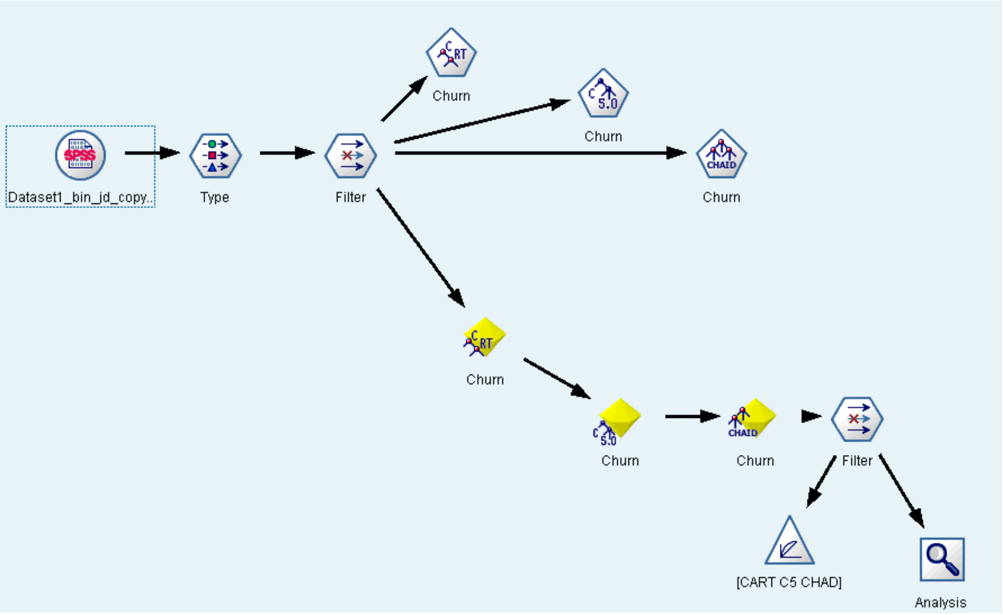


Figure 4.6: Data Stream for the Decision Tree Model

4.4.6 Models Evaluation and Discussion

The models that were presented in the last two sections were chosen based on their quality, from a data-analysis perspective. In the evaluation phase, a performance comparison of the different modelling techniques is carried out. At the end of this phase, a decision on which modelling technique to be adopted for deploying the final customer churn model should be taken. As described with regard to performance evaluation metrics (Section 4.3), there are many evaluation metrics that can be used to compare the performance of different modelling techniques. In this study, AUC, ROC curve, top-decile and overall accuracy are used as performance measurement metrics. After evaluation of the models, a discussion on the limitations of data mining in customer churn analysis is presented.

Model evaluation metrics

Figure 4.8 shows the ROC curve for three decision trees and two logistic regression models. Table 4.8 shows the three evaluation metrics adopted in this chapter. It becomes very clear from Figure 4.8 and Table 4.8 that decision tree in general and C5 in particular outperforms logistic regression model in this study. These results are consistent with some prior literature (e.g. Bakır et al., 2009). However, fair comparison with prior studies is difficult, since the data sets are different. Nonetheless, based on the evaluation results and the data sets used, this study suggests decision tree analysis as a potentially valuable tool for churn prediction.

The C5 model outperforms all other models, including the Logistic Reg. (2) model, which was developed by a data analytics team working for the mobile operator (see Table 4.8). The C5 model shows a lift value of 1.598 at 30 percentile. This value can be explained by that the C5 model can capture about 16% of the churners in the top 30% of a sorted list of churned customers.

Model	Lift (top 30%)	Area under curve	Overall accuracy
C5	1.598	0.763	70.25
CHAD	1.584	0.710	68.83
CART	1.409	0.603	68.57
Logistic Reg.	1.218	0.723	68.1
Logistic Reg.(2)	1.4	-	67.5

Table 4.8: Model Evaluation Metrics

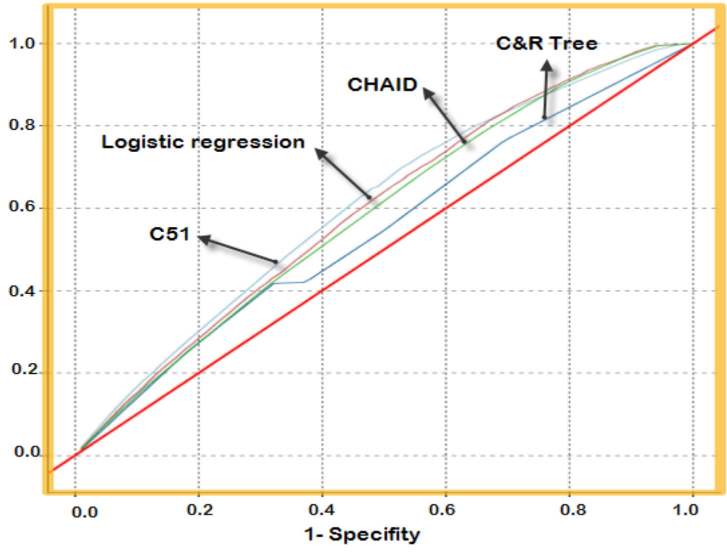


Figure 4.7: The ROC Curve of Decision Trees and Logistic Regression Models

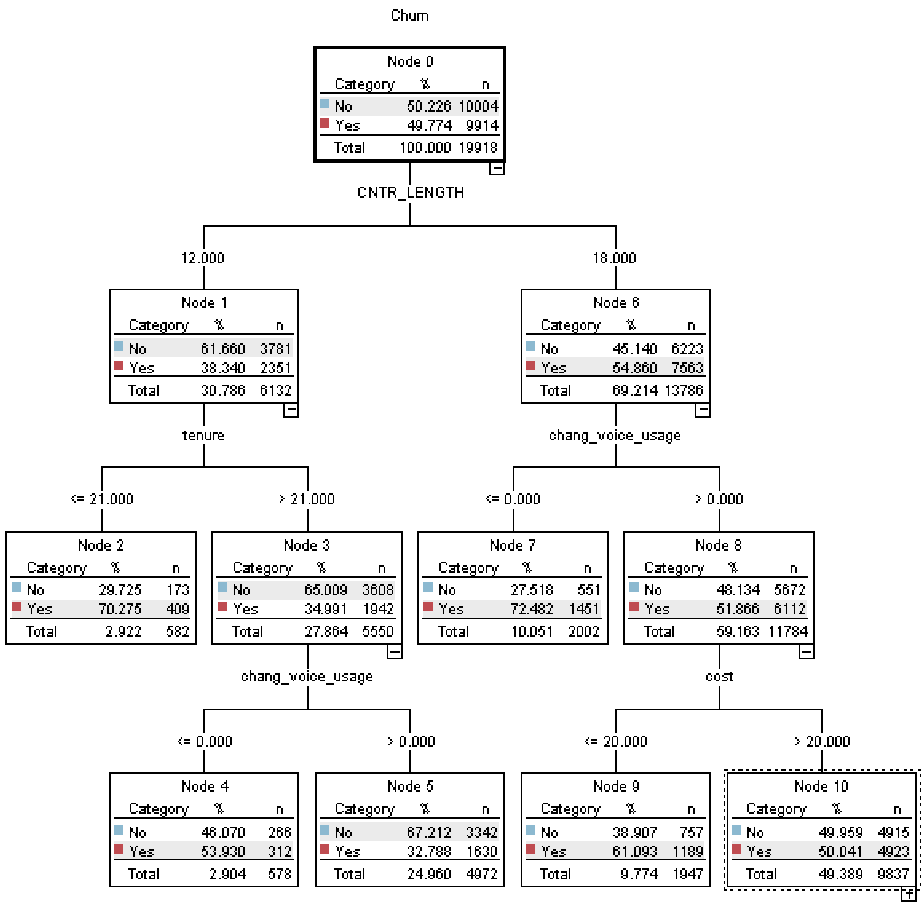


Figure 4.8: The C5 Decision Tree Model

Limitations of data mining

While data mining provides a significant advance in the type of customer churn analysis tools currently available, there are limitations to its use. Five limitations of data mining in customer churn analysis that are relevant to the aim of this study are presented in the following paragraphs. The first limitation is that although data mining is used to reveal patterns, it does not offer direct explanation of the value or significance of these patterns. To provide full explanation for these patterns, skilled data mining professionals are needed.

The second limitation is related to identifying causal relationships between variables. Data mining can identify links between customer behaviours and variables (e.g., demographic and usage variables), but in general it fails to identify the causal relationships between variables. However, decision trees can be used in some cases to infer causal relationships. The third limitation is that data mining models are dependent on the data quality and how recent the data is. The mobile market evolves very quickly, and everyday new standards and technologies are becoming available. Consequently, customer behaviour is changing, and the historical data become less relevant for future predictions. In other words, data mining models have a relatively short expiry date. Therefore, these models cannot be used for long-term planning.

The fourth limitation stems from the fact that churn is based on a multitude of influences, including complex interactions with other members of the population (friends, family). Since only limited data about customers are available, creating a comprehensive model to examine all the factors affecting customer churn simultaneously and jointly is impossible. For instance, there were no data on CDR; therefore, the available data cannot provide any insight into the effect of social networks on customer churn. The fifth limitation is that the level of the analysis in the data mining models is conducted at an aggregate level, which demolishes the ability of these models to capture the heterogeneity of the customers. To summarise, Table 4.9 lists the limitations of using data mining tools in customer churn analysis and describes the general requirements to overcome them.

No.	Description of the limitation	How to overcome it
1	Explanation of the value and the significance of the patterns and relationships necessitate skilled data mining professionals	Provide an easy and intuitive way to explain the value and the significance of relationships between variables
2	Low efficiency of identifying causal relationship	Enhance the ability of identifying causal relationships
3	Short expiry date and the dependency on data quality	Offer tools for long-term planning, and minimise the dependency on data quality
4	Neglecting customers' interactions	Include customer interaction in customer churn analysis
5	Cannot captures the heterogeneity of customers	Offer tools that take account of the heterogeneity of customers

Table 4.9: Data Mining Limitations in Customer Churn Analysis

4.5 Summary

Churn is a very significant and costly problem in today's mobile market. Mobile operators have adopted data mining techniques (e.g. regression and decision trees) for churn modelling. The results of the experiment of this chapter show that the decision tree model outperforms all the logistic regression models, including a model developed by a data analytics team working in the mobile operator. The decision tree model produces a lift value of 1.598 at 30 percentile compared with 1.4 at the same percentile produced by the data analytics team. Therefore, for this problem and for similar data sets, this study recommends decision tree for investigating customer churn. Although data mining provides very useful insights into customer churn, five main limitations of data mining in the customer retention analysis are reported in this chapter. To overcome these limitations, the rest of this study investigates the potential benefit of ABMS in customer churn analysis.

Chapter 5: CubSim - Customer Interaction with the Market

5.1 Overview

The previous chapter has identified some of the limitations of applying data mining in customer churn analysis. In order to overcome these limitations, ABMS is proposed as a novel approach to analyse and investigate customer retention. To demonstrate the applicability and utility of ABMS in the context of customer retention, this chapter presents a case study on modelling customer behaviour and introduces an agent-based Customer Behaviour Simulation Model (CubSim) to model customer behaviour in the UK mobile market. This chapter addresses iteration one of phase three of this study, which focuses on modelling customer interaction with the market.

This chapter is organised as follows: Section 5.2 introduces the concept of the CubSim model. Section 5.3 describes the agent architecture and attributes along with its behaviour. Section 5.4 presents the design process of the model and provides more details on the agent behavioural rules and agent interaction processes. Section 5.5 outlines the calibration, verification and validation procedures that were applied to the CubSim model. Section 5.6 explains the simulation experiments and discusses their results. Finally, the overall findings and results of this chapter are summarised in Section 5.7.

5.2 Conceptual Model Design

ABMS is a computational model for simulating the actions and interactions of autonomous individuals in a network, with a view to assessing their effects on the system as a whole. A generic, reusable, agent-based model (CubSim) is developed to examine the factors affecting customer retention in the UK mobile market. The model development process involves creating artificial agents mimicking the attributes and behaviours of their real-world counterparts. This approach allows an intuitive exploration of the dynamics and complexity of customer retention.

Drawing on a case study of customer experience (Anaman and Lycett, 2010), literature review and market data from Ofcom, the UK telecoms regulator, this research employs the ABMS paradigm to conceptualise and model the players, the actions and interactions within the mobile market. The research results of the case study that has been carried out with a sizeable UK mobile operator enable a better understanding of the problem domain. In addition, these results give insights into the dynamics of the whole system. Based on the system analysis, an innovative agent-based model has been developed to facilitate an understanding of the complex interactions between the different factors that influence customer retention.

The CubSim model uses the ABMS-based approach to understand the dynamic of customer retention. Contrary to many other studies (Ranaweera and Neely, 2003; Chu, Tsai and Ho, 2007), the CubSim model focuses systematically on customer retention via churn analysis. This model is an attempt to identify and understand the key factors that affect customer churn simultaneously and jointly. In addition, this model allows for a more natural description of the mobile market by including all relevant elements. A customer's social network, which has not been addressed adequately in the literature (Richter, Yom-Tov, and Slonim, 2010), is one of these fundamental elements.

Building the CubSim model starts through an iterative design process with a relatively simple model and progresses to more complex ones. The CubSim model is designed from the perspective of one operator competing with others. Agents in the CubSim model that represent customers take decisions based on evaluating satisfaction with the services provided by the mobile operator. The model also incorporates other external factors that have an impact on the agents' decisions, such as WOM.

The model consists of active, autonomous entities called agents and passive entities representing the surrounding environment. Agents in this model represent the population of customers. Each customer has a cognitive process dictating his decision to stay or to leave his mobile network. The model also takes into account different types of interactions that take place in the real market of which there are two types. The first one is the interaction between customers and mobile operators – e.g. with respect to issues like complaints and repair requests. The second one is the interaction between the customers themselves – e.g. recommendation and/or negative comment by family or friends.

Agents that represent customers are the core of the CubSim model. They are the main decision-making components in this model. Finding and representing realistic agent representation are what agent based modelling is all about. Remarkably, there is no consensus on how or what the best way of constructing an agent (Macal and North, 2005). However, there are many different agent architectures used for different applications.

Each agent has characteristics corresponding to the factors that affect customer churn. These factors are determined by a triangulation of data drawn from: (a) The literature; (b) an analysis of real customer data; and (c) industry-standard classification frameworks. From a model perspective, these factors are implemented as an agent's features. Each feature represents a single factor in the customer model. The next section presents a description of the agent architecture utilised in this research.

5.3 Agent Architecture

Agent architecture is the methodology of building agents. Maes (1991, p.115) defines agent architecture as follows:

> “Specifying how the agent can be decomposed into a set of component modules and how these modules should be made to interact. The total set of modules and their interactions has to provide an answer to the question of how the sensor data and the current internal state of the agent determine the actions and future internal state of the agent. An architecture encompasses techniques and algorithms that support this methodology”.

For simplicity, agent architecture can be defined as the map of the internal part of the agent including its data structure, the operations that will be performed on these data structures and the control flow between these data structures. Agent architectures are characterised by the nature of their decision making process. Müller (1999) highlights some of the main types of agent architecture including:

1. Reactive based architectures, in which the decision-making is achieved based on a very limited amount of information, and simple situation-action rules. Reactive agent based architecture makes its decisions directly based on sensory input without any symbolic representation of the world. This type of architecture focuses on producing robust behaviour rather than correct or optimal behaviour.

2. Deliberative based architectures, unlike the reactive agents, in which the decision-making process is based on planning by having an explicit world representation. The decision process takes into account all of the available sensory information and integrates them with the internal knowledge to create a plan of action. One of the famous deliberative architectures is the belief-desire-intention (BDI) architectures. Beliefs are the information that

the agent has about its surrounding environment, about itself and about other agents. Desire is simply goals that an agent wishes to achieve. Intentions are the drives that propel agents to achieve their goals.

3. Interaction based architectures or layered architectures, in which the decision-making is attained by explicitly reasoning about the environment via different software layers. This type of architecture has its origin in distributed artificial intelligence.

5.3.1 Agent Attributes

According to the problem definition and the model description, there are two main components in the CubSim model. The first component is the environment, which represents the mobile operators and mobile market. The second component is the customer agents, which represents mobile users. Table 5.1 and 5.2 list the attributes and the parameters of the customer agent and the environment component, respectively.

Name	Type	Values	Description
PricePerMin	Double	.0 - .99	Average price per minute
CustomerGrowthRate	Double	0 - 100 %	Annual customer growth rate
OurCust_services	Double	0 - 100 %	Customer service level of network 1
OurAggressiveness	Double	.0 - .99	Marketing aggressiveness level of network 1
OtherCust_services	Double	0 - 100 %	Customer service level of network 2
OtherAggressiveness	Double	.0 - .99	Marketing aggressiveness level of network 2
NoofAgents	Integer	4000	Initial number of agents

Table 5.1: Environment Parameters

Name	Type	Values	Description
Age	Double	15 - 100	Customer's age
Gender	Integer	1,2	Gender of the customer (1 for Male, 2 for Female)
Network	Integer	1,2	Customer's mobile network
Cntr_length	Double	12,24,36	Customer's contract length in months
Tenure	Double	0-150	Tenure is the time-period in which the customer is under a contract.
Cost	Double	15- 40	Cost of the customer bundle per month
Repair	Double	1-15	Number of times handset has been sent for repair in a 3 month period
Billing	Double	1-15	Number of customer complaints regarding billing in a 3 month period
Coverage	Double	1,0	Coverage strength at the customer post code (1 for good strength, 0 for bad strength)
Renew	Integer	0-50	Number of times contract has been renewed
Satisfaction	Double	0 - 100 %	Monthly customer's satisfaction level
SatisfactionAvg	Double	0 - 100 %	Average customer's satisfaction level of overall customer life cycle
VoiceRevenueV	Double	0-500	Monthly voice revenue realised by the customer
VASRevenueV	Double	0-500	Monthly non voice services revenue realised by the customer
CLV	Double	0-2000	Lifetime value of the customer
Value	Integer	0,1,2	Customer value group (2- High value, 1- Mid value, and 0- Low value)

Table 5.2: Agent Attributes

5.3.2 Agent Behaviour

Modelling customer behaviour is a challenging and complex task. This involves interdisciplinary processes and theory from sociology, psychology, economics, marketing and computing (Shah et al., 2005). In addition, the great heterogeneity of the customers' preferences and needs imposes a significant challenge to constructing a realistic model of customer behaviour. In this sense, this research seeks to elucidate the complexity of building a customer behaviour model by simplifying the modelling process in order to obtain operational marketing decision tools. Accordingly, this research proposes a novel modelling approach combining a micro-economic utility model of customer satisfaction with a social interaction model. In comparison with existing architecture reported in the literature, this research adapts a combined architecture based on a balance between reactive decision and utility maximisation.

The CubSim model is essentially based on evaluating customer satisfaction emanating from customer interaction with the service provider. Based on satisfaction level and social interaction with family and friends, agents make judgments about their relationship with their service provider. Lastly, based on the combination of interactions in all channels including interacting with other service providers in the market, agents take their final decision either to switch or not switch their network provider. Figure 5.1 shows the conceptual architecture of the CubSim model.

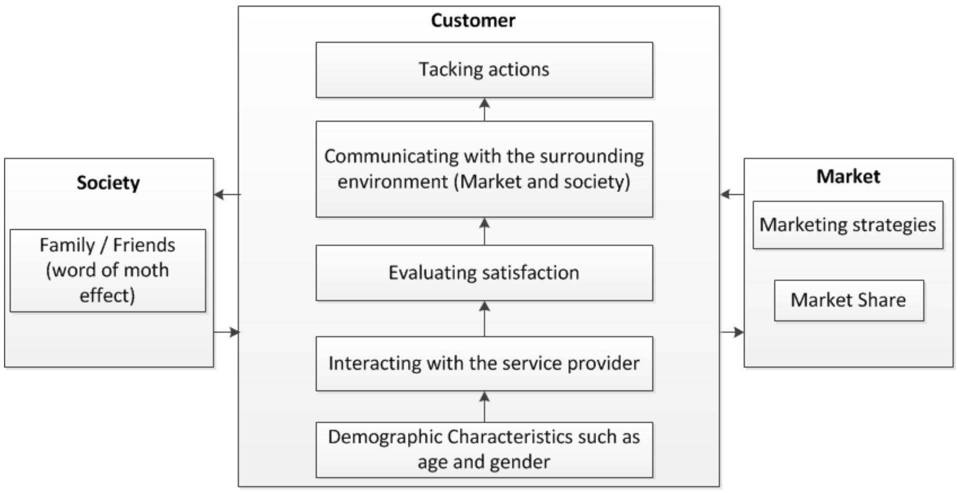


Figure 5.1: The Conceptual Architecture of the CubSim Model

5.3.3 State Diagram of the Customer Agent

Agents are characterised by two elements: Agent attributes and agent behaviour. Agent architecture and agent attributes were presented in the previous section. This section describes the behaviour of the customer agent. This behaviour is conceptualised in a state diagram that illustrates the different states a customer agent can be in. Possible transitions between the states and the events that initiate these transitions are also represented in the state diagram. Figure 5.2 shows the main state diagram of the customer agent. A narrative text description is followed to extend and clarify the agent state diagram. The description represents the point of view of customers when they make a decision of whether to switch or not to switch their network provider.

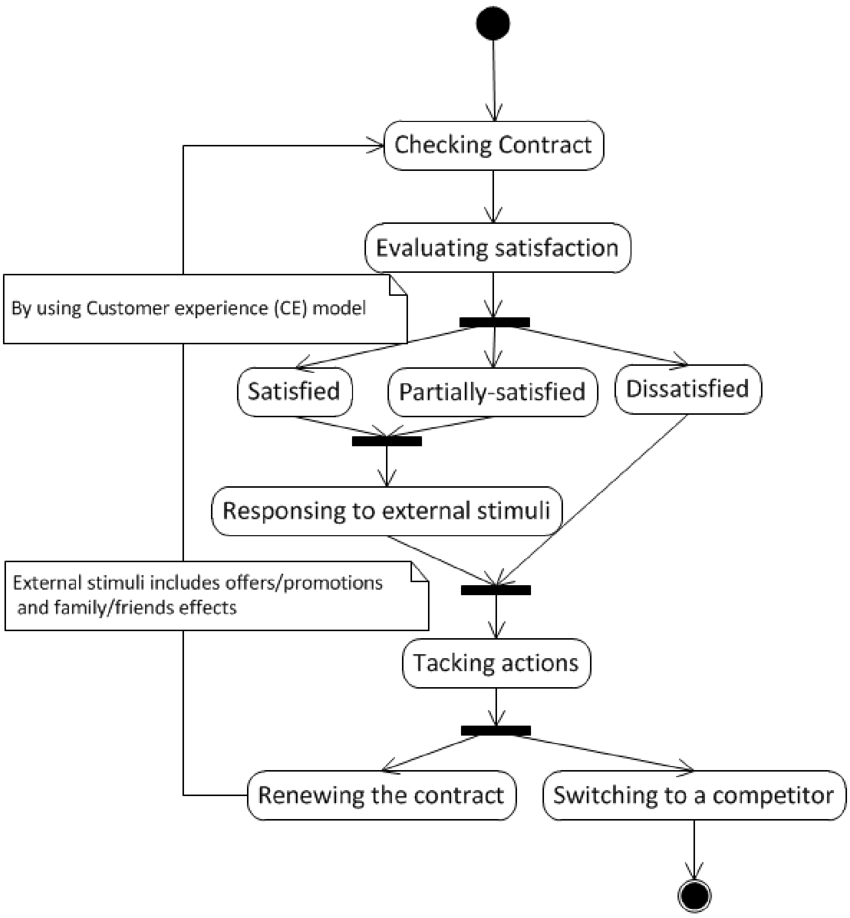


Figure 5.2: The Customer Agent State Diagram

1. A customer checks to find out whether his contract is due to expire. This task is repeated at each time-step, which corresponds to one month of real-world time.
2. The customer makes a judgment about his satisfaction by considering all aspects involved in his experience with the mobile operator. The satisfaction assessment and evaluation process are represented in the Customer Experience (CE) model. The next section illustrates the concept of the CE model.
3. The customer acknowledges his satisfaction level. There are three levels of satisfaction: Satisfied, partially-satisfied and dissatisfied. The level of satisfaction is assigned to each customer based on a predefined judgment trigger threshold.
4. The customer responses to various external stimuli and adjusts his decision accordingly. To simplify the modelling process, this research assumes that a satisfied customer does not look for alternatives and is not influenced by friends and family. Contrary to this, a dissatisfied customer checks for better alternatives from other network.
5. Finally, the customer takes final action either renewing or terminating his contract.

5.3.4 Customer Satisfaction Evaluation Mechanism

Customer satisfaction is an experience based assessment evaluated by customers. This assessment is carried out to find out the differences between the expectation and the reality of the experience of using mobile services (Gerpott, Rams and Schindler, 2001). To quantify and measure customer satisfaction, a CE model proposed in Anaman and Lycett (2010) is adopted. The CE model is an attempt to develop a shorthand method for evaluation of customer experience (inferring CE from data available in technical systems). The main purpose of this model is to create an easy-to-use and understandable method that would consider all key factors that may influence customer satisfaction. The primary outcome measure of this model is a score that represents the level of satisfaction of the customers. Prior to using the CE model within the CubSim model, the development process of CE is presented to illustrate the process of assessing customer satisfaction.

Developing the CE model involves two steps. The first step is identifying the general factors that may affect customer satisfaction. The CE model adopts the components of the J.D. Power framework, which provide group of factors that have proven to be important for customer satisfaction in the UK mobile market (Anaman and Lycett, 2010). The groups of factors are divided into six categories: Cost, handset, coverage, customer services, offerings & promotions and billing. Each category is further subdivided into experience items as showing in Table 5.3. The customer experience items are indicators derived from data collected from the mobile operator's data warehouse. The second step is assigning a rank to each experience item in order to denote the importance weight of that item. The weighting process was validated by interviewing customers as they exit a customer service centre. The adopted customer experience model is represented by a utility function. This utility function is the sum of the product of each experience item multiplied by its rank. The score scale is shown on a 0-to-100 point scale. The theoretical maximum of this score is 100 and the higher the score, the more satisfied the customer. A summary of the CE model is given in Table 5.3.

5.4 Model design and Implementation

5.4.1 Customer Behaviour Simulator

A multi-agent based simulator was built to examine and evaluate the CubSim model. This simulator is used to execute the CubSim model and to mimic and imitate the customer behaviour in the proposed mobile market. Initially, customers are allocated to each mobile operator according to the mobile operator market share. For simplification, two mobile networks are modelled in the proposed mobile market. The first one is called Network 1 and their customers denoted by ‘Our’ customers. This network represents the network under investigation in this study. The second network is called Network 2 and their customers denoted by ‘Other’ customers. This network represents the combination of the other networks in the mobile market.

Category	Weighting	Experience Item	Item Description	Weighting
Cost	18.18	Cost competitiveness	Telco customers’ cost per minute versus the cost per minute for the cheapest competitor	0.66
		Bundle efficiency	Percentage of bundle allocation used per month	0.34
Handset	9.09	Repairs	Number of times handset has been in for repair in a 12 month period	0.75
		Known issues	Known issues with existing handset	0.25
Coverage	23.38	Dropped calls	Percentage of dropped calls, based on totals number of calls made in that month	0.40
		Call Set-up failures	Percentage of call set up failures, based on the number of calls made	0.30
		Home coverage	Coverage rating at home post code	0.30
Customer Services	18.18	Complaint repetition	Percentage of customer complaints with the same reason code in a 12 month period	0.60
		Complaint volume	Number of customer complaints in a 12 month period	0.40
Offerings & Promotions	18.18	Decrease in voice usage	Percentage decrease in voice usage vs. previous month	0.45
		Decrease in data usage	Percentage decrease in data usage vs. previous month	0.45
		Decrease promotion usage	Percentage decrease in usage of latest promotional offer taken up	0.10
Billing	12.99	Billing complaints	Number of customer complaints regarding billing in a 12 month period	1.00

Table 5.3: Outline of the Customer Experience Model (Adapted from Anaman and Lycett, 2010)

Once the simulation session started, the agents start interacting with the surrounding environment including mobile operators and other agents in the market. Agents take decisions at each simulation time step, which represents one month of real-world time. During the simulation, the model users are able to change the mobile operators' policies and then predict the market share dynamics. In this way, users can compare different strategic scenarios.

The CubSim model and the simulator are implemented using a simulation/modelling toolkit. Utilisation of simulation/modelling toolkits greatly facilitates the development of agent-based models. These toolkits provide reliable templates for the design, implementation and visualisation of agent-based models. In this research, AnyLogic 6.6.0 university simulation software is adopted to implement the CubSim model. AnyLogic is a simulation software based on the object-oriented concept and unified modelling language (XJ Technologies, 2011). It provides a flexible visual development environment, which significantly speeds up the development process. In addition, it is based on native java environment, which enables users to integrate java code within their models to define complex structures and algorithms and to bypass the limitation of the graphical environment.

AnyLogic provides many built-in components to facilitate developing agent based modelling and simulation. State diagrams are an example of these components, which are used to describe behaviours graphically. Messages, which are used to transmit information between agents, are another example of these components. AnyLogic has been applied in different application problems including economic, business, social, biological and physics.

The Customer Behaviour Simulator (CBS) is a multi-agent system that simulates the action and the interactions of an agent population representing mobile customers. The simulator software is based on a multi-tiered architecture embracing three layers: Presentation, application and data layer. The presentation layer consists of user interface components. Through these components, users can interact with the model and run the simulation experiments by changing the simulation parameters or the simulation scenarios. The presentation layer also includes visualisation components, which allow the results of the simulation runs to be explored through diagrams and plots. The visualisation components play an important role in identifying and communicating important behaviours and can assist in the validation and the verification process. It also offers the means to follow the evolution of the simulation and to deliver insight into the emerging consumer behaviour.

The application layer contains the central component of the simulator, which is the simulation engine. Agents' rules and relationships between agents are implemented within this component. Data layer is a virtual layer, and its main function is to deal with input and output data. Figure 5.3 shows the CBS architecture.

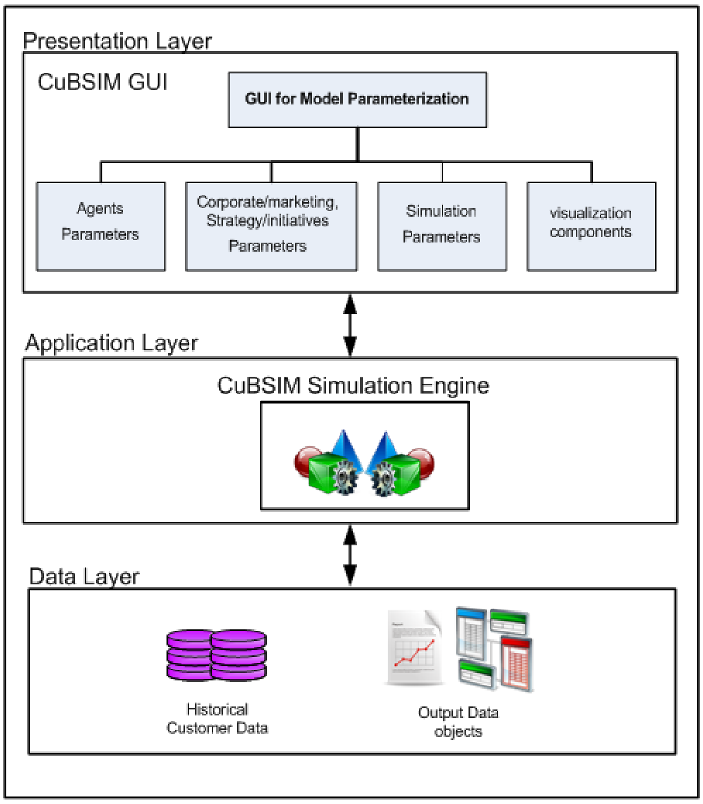


Figure 5.3: Customer Behaviour Simulator (CBS) Architecture

5.4.2 CubSim Module Implementation

After completing the conceptual design, a domain expert in customer retention management with 10 years of telecom experience was consulted to verify the key features that should be presented in the CubSim model. Then the implementation phase started to execute the design elements identified in the conceptual model design.

Two java classes were implemented in AnyLogic; one contains the customer agent implementation and another contains the environment implementation. The customer agent, as noted earlier, is the core of the CubSim model. Therefore, customer agent was modelled as an active agent, while the mobile operators were modelled as a part of the environment component. In the implementation phase, the conceptual state diagram in Figure 5.2 was turned into a design implementation.

Figure 5.4 shows a screenshot of the state diagram of the design implementation of the customer agent in CubSim V2. The state diagram in this figure defines the internal state and the corresponding state transition of the customer agent. The customer agent state diagram has two main states: ‘Our’ and ‘Other’. The Our state corresponds to that the agent is a Network1 customer. The Other state corresponds to that the agent is a Network 2 customer. The Our state is a composite state that depicts the internal state of any customer belongs to Network 1. At any point of time, the state of any customer can be ‘Checking’, ‘Satisfied’ or ‘Unsatisfied’.

The Checking state represents customers who locked into a contract and wait until their contracts end. When a customer contract finishes, the customer encounters a decision branch, the branch 1 in Figure 5.4, where the satisfaction assessment is carried out. Based on this assessment, the customer state will be changed into either Satisfied or Unsatisfied. If the customer agent is satisfied, the contract renewal process will be called and the customer agent state will back to the

Checking state. However, if the customer is unsatisfied, the customer encounters a decision branch, the branch 2 in Figure 5.4, where the decision between alternatives has to be taken. Based on the customer decision, customer state is either changed into the Checking state again or changed into the Other state.

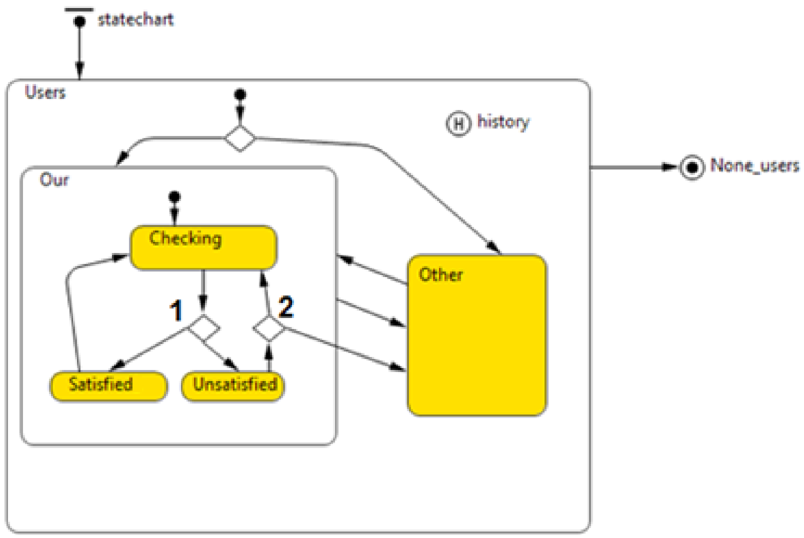


Figure 5.4: The State Diagram of the CubSim Design Implementation

5.4.3 Empirical Input Data

As demonstrated in the previous chapter, the level of analysis in the traditional customers churn analysis is implemented and reported at an aggregate level and does not consider the variability in customer populations. One of the strengths of agent-based modelling is its ability to model the variability and heterogeneity in the population of customers (Richetin et al., 2010).

To model systematic heterogeneity in customer characteristics, the probability density function (PDF) was used. This function describes the relative likelihood for a variable to occur at a given point (Berry and Linoff, 2004, p.133). Developing the PDF function, also known as empirical frequency distribution, was based on historical customer data extracted from the company's data warehouse. Table 5.4 shows some of the empirical frequency distributions used to populate information on customer agents' characteristics.

Name	Description
NetworkDistribution	An empirical probability distribution function used to populate the Network attribute of the customer agent based on the UK mobile market share.
AgeDistribution	An empirical probability distribution function used to populate the Age attribute of the customer agent based on real customer data.
GenderDistribution	An empirical probability distribution function used to populate the Gander attribute of the customer agent based on real customer data.
CntrDistribution	An empirical probability distribution function used to populate the Cntr_length (Customer's contract length) attribute of the customer agent based on real customer
CoverageDistribution	An empirical probability distribution function used to populate the Coverage attribute of the customer agent based on real customer data.
TenureDistribution	An empirical probability distribution function used to populate the Tenure attribute of the customer agent based on real customer data.
CostDistribution	An empirical probability distribution function used to populate the Cost attribute of the customer agent based on real customer data.
RepairDistribution	An empirical probability distribution function used to populate the Repair attribute of the customer agent based on real customer data.
BillingDistribution	An empirical probability distribution function used to populate the Billing attribute of the customer agent based on real customer data.

Table 5.4: Empirical Frequency Distributions

5.4.4 Main Features in the CubSim V1 and V2

The implementation processes were performed iteratively until a satisfactory representation of the system under study was achieved. This section documents some of the key features that were implemented in the CubSim model Version 1 and 2. The full documentation of the CubSim model is provided in Appendix A. The model code is available for download on the Google code platform at: http: http://code.google.com/p/cubsim/. A working version of the model can be viewed at http://www.cubsim.com/model/. The java virtual machine (JVM) is needed to run the model.

Market share (implemented in V1): To build a realistic representation for customer market share, Ofcom market data figures were used to build probability density function for customer market share. When the simulation starts, the simulator assigns customers to either Network 1 or Network 2 based on the "NetworkDistribution" frequency distributions. During the simulation experiments, model users can monitor the evolution of the customer market share. Figure 5.5 displays a screenshot of the time stack chart, which shows the market share of the two mobile networks. The horizontal axis represents time, and the vertical axis represents customer market share.

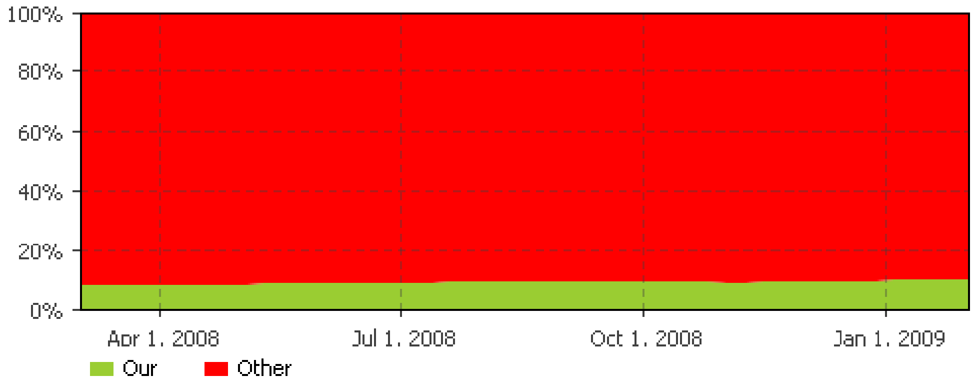


Figure 5.5: Time Stack Chart of the Market Share

Satisfaction assessment (implemented in V1): This feature was implemented based on the satisfaction evaluation mechanism that was explained earlier in the model design section. Two methods were written to implement the satisfaction assessment process. These methods are *Cal_satisfaction* and *Satisfied.* The *Cal_satisfaction* method is implemented to calculate the customer satisfaction score based on the CE model (see section 5.3.4). This score is passed to the *Satisfied* method, which returns a boolean value (true or false). If the customer satisfaction score exceeds a predefined threshold, the method returns true to indicate that this customer is a satisfied customer, and vice versa.

Renew customer contracts (implemented in V1): There are two cases where a customer contract can be renewed: (1) When the customer exits the Satisfied state; or (2) when the customer exists the Unsatisfied state and encounters the decision node 2 in Figure 5.4 and found that the other network is no better than his network. The customer agent attribute 'Renew' is used to indicate the number of times a customer contract has been renewed.

Contracts/due process (implemented in V1): Two methods in the customer agent class were written to implement this feature. The first method is the *Due* method and returns a boolean value (true or false). The second method is the *onContract* method and returns a double number equal to the number of months remaining until the customer contract expires. If the customer contract is not due, the customer will keep checking his contract for a period equal to the return value from the *onContract* method. If the customer contract is due, the customer agent will leave the Checking state.

Market growth (implemented in V2): To build more realistic representation of the mobile market, new customers coming to the market should be considered (Thomas, 2001). Based on Ofcom data, the *growthRate* method in the customer agent class takes one parameter that represents time in years and returns the annual customer growth rate. This method is invoked every year during the simulation run. The simulator then creates a number of new customer agents, which reflects the annual growth rate.

Customer lifetime value (implemented in V2): Customer Lifetime value (CLV) is typically defined as the total income earned from a customer during the lifetime of his relationship with the business. CLV is an important factor in analysing customer churn (Gupta et al. 2006). Therefore, in order to construct a realistic model of customer churn, CLV should be considered. The value of the CLV for each customer is stored as an agent attribute. This value is calculated by the average of the revenue derived from using voice and non-voice data. The revenue resulting from the voice data is calculated and returned by the *RPUVoice* method in the main class. The revenue resulting from the non-voice data is calculated and returned by the *RPUVAS* method in the main class. After calculating the total value of CLV, this value passes to the *customerValue* method in the main class. Then, the *customerValue* method returns an integer indicating the customer value category, of which there are three categories: High, Mid and Low.

Matching offers of other networks (implemented in V2): This is one of the key strategies that have been adopted by mobile operators to retain their valuable customers (Richter, Yom-Tov and Slonim, 2010). This feature is implemented within the *Switchtoother* method in the customer agent class. This method checks the customer value category. If the customer value category is High or Mid, the simulator will match Network 1 price with Network 2.

5.5 CubSim Model Calibration, Verification and Validation

Before the model is calibrated with data from the real system, a verification process is conducted. The purpose of the verification process is to confirm that the model design specifications are fulfilled by the model implementation (North and Macal, 2007). Three verification methods were performed: Design walkthroughs, code walkthroughs and model logging. The first two methods were performed by a modeller who has experience in the simulation software (AnyLogic). Through design walkthroughs, the modeller performed a visual check of the model components like state diagrams and transitions. Code walkthroughs involved checking the model code against the model design specification to identify errors. Finally, the model logging process was carried out by recording states and the values of some of the agents' variables over time to verify the correctness of the operation of the simulation. This verification process was performed continuously during model implementation until a satisfactory behaviour of the simulation model was achieved. In the calibration phase, model parameters were fitted to data from Ofcom and the mobile operator that was chosen in the case study. The data used for calibration include the following:

- Initial customer market share.
- Average minute price.
- Annual customer growth rate of the years covered in the simulation.
- Customer data from the chosen mobile network to build the empirical frequency distributions, those listed in Table 5.3.

Once the verification and calibration processes were finished, the validation process was carried out. The purpose of the validation was to be sure that the model represented and correctly reproduced the behaviour of the real system. A Black-box validation at a macro level was not applicable because there are only few cases relating to the real-world. Essentially, the analysis in this research was based on two historical data sets, and there was no data available at a macro level from the mobile network, which was chosen in the case-study.

As a result, sensitivity analysis was used to validate the CubSim model. The purpose of sensitivity analysis was to be sure that the simplifications that have been made in the model design did not diminish its credibility and utility for providing new insights to the problem domain. The next section presents and discusses the sensitivity test along with other experiments to establish the credibility and utility of the CubSim model.

5.6 Simulation Experiments and Analysis

This chapter reports the design and implementation of Version 1 and 2 of the CubSim model. The current implementation is focused on modelling customer interactions with the market. The interactions among customers are not represented in the current implementation. To realise the full potential of the CubSim model, the conceptual model design presented in Section 5.4.2 should be implemented fully. The incomplete implementation of the model limits the ability to conduct experiments on it and only a limited number of scenarios can be tested. However, this section reports the two experiments performed on the current implementation of the model. These two experiments are chosen to establish the credibility of the model and to identify the optimal number of agents that can be used in the next experiments to produce consistent, stable and reliable results.

Experiment 1: Customer Pool Size.

There are two objectives of this experiment: (1) To examine the sensitivity of the model outputs when the customer pool size is varied using four different pool sizes; and (2) to identify the optimal number of agents that produces reliable results. The pool sizes were varied from 1,000 to 7,000 customers in steps of 1000. In this experiment, the simulation runs were executed several times using the same parameters and different customer pool sizes. The number of satisfied customers, unsatisfied customers and churned customers were selected as comparative performance measurements. All the performance results were expected to be the same as the customer pool size increasing. The hypothesis for this experiment was as follows:

> Hypothesis 1: Varying the customer pool size will not affect the number of satisfied customers, unsatisfied customers and churned customers.

The model ran with customer pool size of less than 4,000, which produced unreliable and unstable results. The performance measurements started to become more stable at customer pool size equal to 4,000. Table 5.5 shows descriptive statistics results for Experiment 1(with customer pool size varying from 4,000 to 7,000). The results table lists the mean and standard deviation values for the three performance measurements at different customer pool sizes. The results are consistent with Hypothesis 1.

In order to validate rigorously Hypothesis 1, a One-way ANOVA test was performed. An ANOVA test was conducted to explore the impact of customer pool size on the chosen performance measurements. Levene's test was not significant ($p > .05$) for all the performance measurements. This indicated that there was indeed homogeneity of variances, and therefore, ($P < 0.05$) was considered statistically significant.

The ANOVA results in Table 5.6 show that there are no significant differences between the mean of the three performance measurements ($P>0.05$). This is sufficient to conclude that the Hypothesis 1 is true. Based on the results of this experiment, customer pool size of 4,000 is chosen to run all other experiments in this study.

Pool size	**4,000**		**5,000**		**6,000**		**7,000**	
Performance	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Satisfied customers	21.198	2.5233	21.493	2.5452	21.696	1.5805	22.229	1.6526
Unsatisfied customers	78.801	2.5233	78.506	2.5452	78.303	1.5805	77.771	1.6526
Churned customers	35.849	2.9773	37.494	1.750	37.749	2.4464	37.709	2.4446

Table 5.5: Descriptive Statistics Results of Experiment 1

		Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Satisfied	Between Groups	12.282	3	4.094	.902	.444
	Within Groups	376.727	83	4.539		
Unsatisfied	Between Groups	12.282	3	4.094	.902	.444
	Within Groups	376.727	83	4.539		
Churn	Between Groups	55.845	3	18.615	3.373	.022
	Within Groups	458.107	83	5.519		

Table 5.6: ANOVA Analysis Results of Experiment 1

Experiment 2: Effect of adding new customers

The aim of this experiment is to examine the effect of adding market growth representation to the model on the model performance measurements. In this experiment, the simulation experiments were executed several times using two implementations. The first one was with the representation of adding new customers to the market, which was discussed in Section 5.4.4, and the second one was without this representation. The number of satisfied customers, unsatisfied customers and churned customers were selected as comparative performance measurements. All the performance measurements were expected to be the same in the two model implementations. The hypothesis for this experiment was as follows:

> Hypothesis 2: There is no significant difference between the performance measurements of the two model implementations (with and without adding the new customers' representation).

Table 5.7 shows the results of Experiment 2. The results table shows the observation count (N), mean and standard deviation for three performance measurements. The results are grouped into three pairs. Each pair belongs to one performance measure, and each pair has two readings: One with the market growth representation and the other without. The results are not consistent with Hypothesis 2. In order to rigorously prove or (or disprove) this, a paired sample T-test was performed.

		Mean	**N**	**Std. Deviation**
Pair 1	Satisfied	21.9727	20	2.62782
	Satisfied_Without	20.1640	20	2.74893
Pair 2	Unsatisfied	78.0273	20	2.62782
	Unsatisfied_Without	79.8360	20	2.74893
Pair 3	Churn	32.6471	20	2.46771
	Churn_Without	5.5236	20	1.54795

Table 5.7: Descriptive Statistics Results for Experiment 2

The T-test results in Table 5.8 are significant for the three performance measurements. For the number of satisfied customers, there was a significant difference in the score with the new customers representation (M=21.9727, SD=2.62782) and without (M=20.1640, SD=2.74893); t (19) = 2.177, p = 0.042. For the number of unsatisfied customers, there was also a significant difference in the score with the new customers representation (M=78.0273, SD=2.62782) and without (M=79.8360, SD=2.74893); t (19) = -2.177, p = 0.042. Finally, the same pattern was also observed in the case of the number of churned customers. There was a significant difference in this score with the new customers' representation (M=32.6471, SD=2.46771) and without (M=5.5236, SD=1.54795); t (19) = 38.214, p = 0.00.

The results of this experiment suggest that considering the market growth in the model does have an effect on the three performance measurements. Specifically, the results of churned customers suggest that when the new customers representation was added, the number of churn customers increased. Table 5.8 shows T-test results for Experiment 2.

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1 - Satisfied customers	1.8086	3.7159	.8309	.0694	3.547	2.17	19	.042
Pair 2 - Unsatisfied customers	-1.8086	3.7159	.8309	-3.5477	-.0694	-2.17	19	.042
Pair 3 - Churned customers	27.1235	3.1742	.7097	25.6379	28.609	38.21	19	.000

Table 5.8: T-test Analysis Results for Experiment 2

5.7 Summary

This chapter proposes a novel method to gain insight into the dynamic and the complexity of customer retention and introduces an ABMS based model (CubSim) for customer churn behaviour in the UK mobile market. This chapter presents the first development iteration of the CubSim model, which focuses on customer interaction with the market, including interaction with the service provider and other competing operators in the market. This iteration starts with designing the architecture of the customer agent. In comparison with existing architecture reported in the literature, this study adapts a combined architecture based on a balance between reactive decision and utility maximisation. To model the interaction with the service provider, a micro-economic utility model of customer satisfaction is used. The main purpose of this model is to create an easy-to-use and understandable method that would consider all key factors that may influence customer satisfaction. The interaction with other competing operators in the market is represented by two variables: Customer services and the Marketing aggressive level. If a customer is unsatisfied with his service provider, he/she starts considering alternatives from the competing network based on these two variables.

The current implementation of the model focuses on customer interaction with the market. To realise the full potential of the CubSim model, the conceptual model design presented in this chapter need to be implemented fully. Therefore, the next steps for the CubSim model development include introducing social influences into the model and performing more validation experiments.

Chapter 6: CubSim - Interaction among Customers

6.1 Overview

The previous chapter introduced the conceptual design of the CubSim model along with the implementation of customers interaction with the market. Following on from that, this chapter addresses iteration two of phase three of this study, which focuses on modelling the interaction among customers. In addition, it demonstrates how the CubSim model can be used to generate insights into customer churn.

This chapter is organised as follows: Section 6.2 provides background information on social networks. Section 6.3 presents the CubSim extension process, which includes modelling the social network, modelling the word of mouth flow, and extending the agent behaviour to incorporate social influences. Section 6.4 outlines the calibration, verification, and validation procedures that were applied in this iteration. Section 6.5 explains the simulation experiment and discusses the results. Section 6.6 discusses and evaluates the CubSim model. Section 6.7 ends this chapter with a summary discussing the overall findings and results of the simulation experiments.

6.2 Social Networks

Social Network Analysis (SNA) is a growing methodological research area. SNA has emerged as a key technique for characterising social structure and interaction through network representations. SNA presents both a visual and mathematical analysis of the relationship between network entities. The social structure depicted in the network representations is called a 'social network', which encompasses nodes and links. In the context of this research, nodes are used to represent customers, while links are used to represent the relationships between customers.

Traditionally, SNA focuses on static networks or what is called 'symbolic' networks. Static networks are viewed as a mapping of the relationships between discrete entities. These types of networks do not change their structure over time. Recently, more attention has focused on dynamic networks that are capable of representing the continuous transmission of information and influence (Watts, 2004). Dynamic network analysis is an important aspect of ABMS. Utilising dynamic network analysis entails understanding the agent rules that govern network structure and growth, and how information and influences are transmitted through networks, and the type of relationships that are embodied in the network (Macal and North, 2005).

The social networks term has been used loosely to refer to online communities such as Facebook and Twitter. These communities are an example of social structure. Facebook and similar online communities are only tools to represent social networks however; the people (the nodes) who join these communities and the relationships between them are the social networks. A social network structure is formed of nodes that are connected by means of one or more links.

Vag (2007) identified four widely accepted principles of social networks:

1. The actors that represent nodes and their actions should be viewed as interdependent rather independent.
2. The links between nodes are channels for the transfer or flow of resources (either material or non-material).
3. Social network models represent the structural environment as a network imposing certain constraints on individual actions.

4. Network models conceptualise structure (social, political, economic, and so on) as lasting patterns of relations among actors.

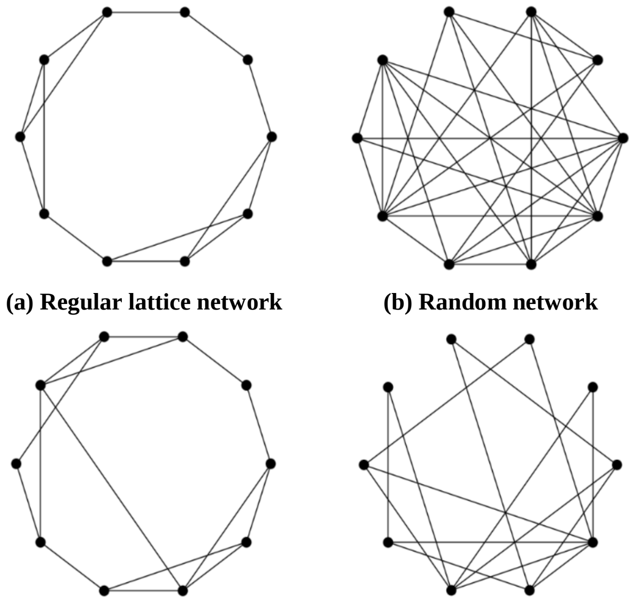
Milgram's (1967) study was one of the first empirical research studies of social structure, in which he introduced the ‘six degrees of separation’ phenomenon. Milgram distributed a number of letters to a sample of people in one city. All letters were addressed to the same address in another city. Milgram instructed all participants in his study to send the letters by passing them from person to person. When the letters finally reached their destination, Milgram found that the average number of recipients for a letter is six. Then he concluded that any two actors on the planet are separated by at most six degrees of separation.

Although Milgram’s study was not subjected to rigorous validation, his concept of a small world is now widely adopted in social network studies to provide an explanation of how information spreads. The small-world network has become one of the most popular social network models (Baxter, Collings and Adjali, 2003). The next sections discuss the characteristics of the small-world network along with other social network models.

6.2.1 Characteristics of Social Networks

Social network model representation is best explained with an example. The networks in Figure 6.1 represent four social network models of mobile customers. In these models there are ten nodes representing ten customers. The links between nodes represent the relationships between customers. These links are viewed as channels for transmitting influences. Social network models are used to visualise and model the patterns of interactions between nodes, of which there are four basic types of network reported in the literature: Regular lattice, random, scale-free and small-world.

Characteristics of any social network are defined by two sets of measurements: Nodes and network measurements. Hamill and Gilbert (2009) identified two measurements for nodes and three measurements for networks. Nodes are characterised by degree of connectivity and clustering coefficient, while networks are characterised by size, path and density. Table 6.1 lists nodes and network characteristics and their definitions.



(a) Regular lattice network (b) Random network

(c) Small-world networks (d) Scale-free networks

Figure 6.1: Examples of Social Network Models of Mobile Customers

Node characteristics	
Degree of connectivity	The number of links to or from the node. It is different from the degree of separation, which denotes path lengths.
Clustering coefficient/density	Measure of the degree to which nodes tend to cluster together. In other words, it is the extent to which one's friends are also friends of each other.
Network characteristics	
Size	The number of nodes and links in the network. If there are n nodes, the maximum possible number of undirected links is n (n-1)/2.
Path length	The distance between a pair of nodes measured by the number of links between the pair, provided that any node or link could only appear once in each path. The shortest path between two nodes is called a ‘degree of separation’.
Whole network density	The percentage of the actual number of links to the total possible.

Table 6.1: Social Network Nodes and Network Characteristics (Source: Hamill and Gilbert, 2009)

6.2.2 Social Network Models

To model social influences within ABMS, two important issues need to be addressed: (1) What is the most appropriate social network model that can sufficiently represent the social processes under study; and (2) what is the mechanism of passing information through the social network. This section discusses the main types of social network model and their characteristics.

Regular Lattice Networks

Regular lattice networks are the simplest types of network, in which each node is connected to all of its nearest neighbours. Figure 6.1 (a) shows a regular lattice network consisting of ten nodes. Regular lattice networks are characterised by low density, high local clustering and large average path length (Hamill and Gilbert, 2009). This type of network is generally used in cellular automata models. An example of ABMS models that employ this type of network is Wu's (1998) work, in which he uses a regular lattice network to investigate the complexities of urban developmental phenomena.

Random Networks

Random networks are formed by randomly linking nodes. In contrast to the regular lattice, random networks are characterised by a short path length and low local clustering. Although random networks have been used to model communication networks and the World Wide Web, it is unlikely that a random network structure can be found in empirical data about people’s interactions. An obvious reason for this is that people rarely interact with other people in a random manner. However, sometimes random networks are used as a test case in ABMS (Alam, Geller and Tsvetovat, 2011).

Scale-free networks

Scale-free networks are created by employing the dynamics of preferential attachment, in which it is more likely for a new node to attach to nodes with many links. The first serious discussion and analysis of scale-free networks emerged during the 1990s with Barabási and Albert's (1999) 'rich get richer' work. In that work, they represented the topology of a portion of the World Wide Web and found that some nodes, which they called "hubs", had many more links than others. This type of network is characterised by a short average path length, low clustering coefficient and low density. Scale-free structures do not generally apply to social networks (Hamill and Gilbert, 2009). Hence, it is very rare to find ABMS of social systems employing scale-free networks.

Small-world networks

The 'six degrees of separation' phenomenon is the foundation of small-world networks (Milgram, 1967). In this type of network, most nodes are not connected to each other and at the same time one can travel from one node to another using a relatively small number of links. Small-world networks have high clustering coefficients like regular lattice networks and a relatively short average path length like random networks. Small-world networks are employed within ABMS to model a wide variety of social processes. Some examples of these processes are the diffusion process of technology products (e.g. Garcia, 2005) and the spread of the SARS epidemic (e.g. Huang et al., 2004).

Social network structure has a significant effect on the local and global dynamics of ABMS (Garcia, 2005). Therefore, finding a network structure that better represents the real social network is very important. However, finding a network structure that fits well with sociological observations of real social networks remains an open problem. Detailed examination of network structures in ABMS by Hamill and Gilbert (2009) showed that none of the basic network structures represented in Section 6.2.2 could capture the important aspects of real social networks. Hamill and Gilbert (2009) provide a list of eight network structure requirements that match the basic characteristics of a social network. In addition, they compare the characteristics of the four common network structures against this list. Table 6.2 summarises the characteristics of the four basic network structures and compares them against the desirable characteristics of the real social networks.

Characteristic	Regular	Random	Small-world	Scale-free
Low density	✓	✓	✓	✓
Personal network size limited	✓	✓	✓	✗
Variation in size of personal network	✗	Limited	Limited	✓
Fat-tail (large skewness)	✗	✗	✗	✓
Assortative	✗	✗	✗	✗
High clustering	✓	✗	✓	✗
Communities	✗	✗	✗	✓
Short path lengths	✗	✓	✓	✗

Table 6.2: Characteristics of the Four Basic Social Network Models (Hamill and Gilbert, 2009)

6.3 Extending the CubSim model

This section discusses in detail the process of extending the CubSim model. This process involves three main tasks: (1) Modelling the social network; (2) modelling the WOM flow; and (3) extending the agent behaviours to incorporate social network influences.

6.3.1 Modelling the Social Network within CubSim

As noted, basic network structures cannot reflect the characteristics of the real social networks. The dominant approach to modelling social networks in the customer churn literature is extracting the network and nodes attributes from the Call Data Record (CDR) (see for example, Dasgupta et al., 2008).

Nevertheless, collecting data on social networks is a challenging task (Alam, Geller and Tsvetovat, 2011). For instance, Dasgupta et al. (2008) used about 60 gigabytes of raw CDR data to extract the attributes of the social network among mobile customers. CDR data are not available for this study; therefore, building the social network based on the CDR data is not applicable.

To overcome the limitations of the basic network structures and the non-existence of data about the social interactions, this study adopts Hamill and Gilbert's (2009) social network model to represent the social network within the CubSim model. The main advantage of this model compared to other models is that it is based on a minimum number of sociological assumptions and minimal data (Hamill and Gilbert, 2009). In addition, it is well suited to replicating the characteristics of the real social networks.

Hamill and Gilbert's model is based on a concept called 'social circles'. Simply, this concept is based on using circles to represent personal networks. Figure 6.2 shows two personal networks, A and B. To describe a personal network by using the social circles, one feature called 'social reach' is used. Social reach denotes the radius of the circle that represents the personal network. All persons within the circumference of a social circle are part of the personal network represented by this circle.

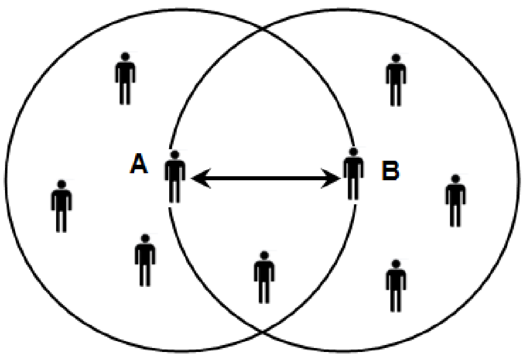


Figure 6.2: An Example of Personal Networks

The personal networks that comprise the whole social network in the CubSim model are built based on a simple single-reach model, where all customers in the model have a small, identical social reach. Any two agents are connected if the distance between them is less than nine, which denotes the value of the social reach. This approach has been chosen to replicate a network of close family and friends. The value nine was chosen based on the results of a recent study by Microsoft (2006), which found that the average size of personal networks based on stronger ties in the United Kingdom is around nine people.

Detailed implementation of the social network is presented in Section 6.3.4. Table 6.3 compares the characteristics of the social network within the CubSim model against the recommended characteristics listed in Table 6.2. In addition, Table 6.3 briefly explains and provides evidence of whether the recommended characteristics are presented in the CubSim model or not.

Recommended characteristic	**CubSim characteristics**
Low density	Yes, the whole network density is 0.032 and calculated by dividing the actual connections by the maximum possible connections.
Personal network size limited	Yes, the maximum number of links per agent is 13 (see Figure 6.3).
Variation in size of personal network	Yes, the size of the personal networks follows normal distribution (see Figure 6.3).
Fat-tail (large skewness)	No, the links distribution exhibits low skewness (-0.073), and that means a thin tail exists in the CubSim model.
Assortative	Yes, because those in densely populated regions have as many links as those to whom they are linked (see Figure 6.7).
High clustering	Yes, because all agents have the same social reach and they are located very close to each other, and therefore their circles are largely overlapped. Hence, these agents are highly clustered (see Figure 6.7).
Communities	Yes, the visualisation of the social network in Figure 6.7 shows groups of agents that are well connected within themselves but less connected to other groups.
Short path lengths	Yes, partially. See Figure 6.7, where it is shown that in many cases the number of links from one agent to another is relatively small.

Table 6.3: CubSim Social Network Characteristics

As illustrated in Table 6.3, the social network, which is implemented within the CubSim, satisfies most of the real-world social networks requirements identified by Hamill and Gilbert (2009). The next section discusses and explains the process of modelling the WOM flow.

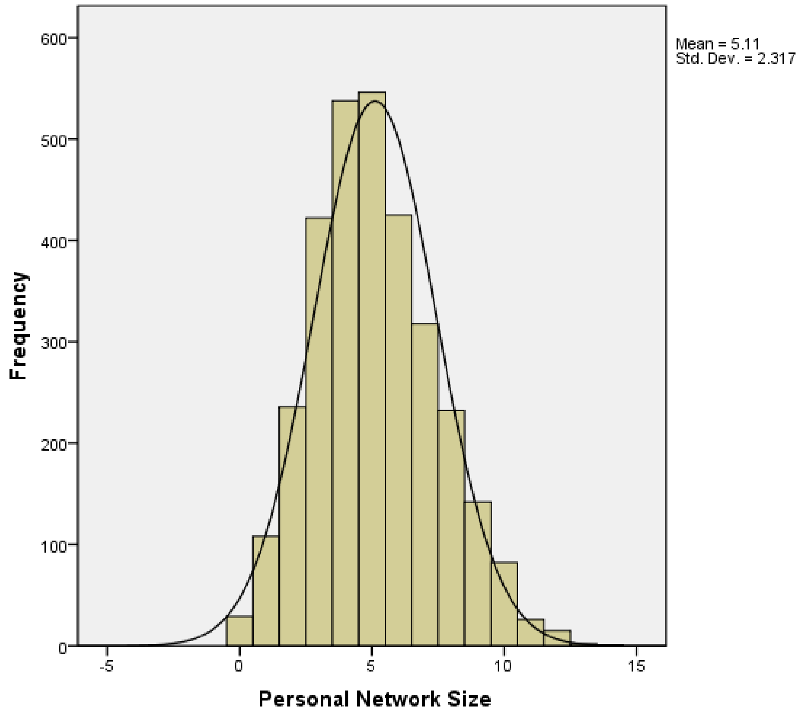


Figure 6.3: Histogram of the Personal Network Size

6.3.2 Modelling the WOM Flow within CubSim

A large and growing body of literature has investigated WOM in different domains and by different methodologies. This study is distinguished from previous studies in that the WOM is combined with other factors that have been shown to affect customer retention, more specifically customer satisfaction that was discussed in Chapter 5. Rather than focusing on few factors, this study presents a holistic view of customer retention. This approach not only contributes to increasing CubSim credibility, but also provides a better understanding of the WOM itself.

Allsop, Bassett and Hoskins (2007) support this approach and emphasise that traditional market research methods are no longer satisfactory to provide full understanding of the WOM. They suggest ABMS as an alternative method to investigate the WOM, combined with other related factors to provide solutions for business problems. The following paragraphs explain the approach that has been followed in this study to model the flow of WOM.

In contrast to the social network models, there is no dominant approach to modelling the WOM flow. Moreover, there is no consensus in the literature regarding the best way to model a consumer market with WOM interactions (Shi and Brooks, 2007). In remaining consistent with the work to this point, the model of WOM flow in this study is inspired by Reichheld's Net Promoter score (Reichheld, 1996). This scoring system classifies the company's customers into three categories: Promoters, passives and detractors. The classification process is carried out by asking the customers if they would recommend the company to a friend or colleague. Customer responses are recorded on a 0-to-10 point rating scale. The three categories and their corresponding scales are as follows:

- Promoters (score 9-10) are loyal, enthusiastic customers who will keep buying and referring others.
- Passives (score 7-8) are satisfied customers but unenthusiastic in referring others. Those customers are vulnerable to competitive offerings.
- Detractors (score 0-6) are unhappy customers who can damage the company brand by spreading negative word of mouth.

In this study, the customer classification process is not carried out by asking the customers if they are going to recommend the company to a friend or colleague. Rather, the customer category is assigned automatically, based on the customer satisfaction score as discussed in Section 5.4.4. If a customer satisfaction score reaches specific values, the customer category is changed to correspond with these values. Figure 6.4 shows the UML state diagram of the customer states, which represent the WOM flow. There are three states that correspond to the three customer categories: ‘Promoters’, ‘Passives’ and ‘Detractors’. If the customer state is changed from ‘Passive’ to ‘Promoter’, the customer starts spreading positive word of mouth about his mobile network. In the same way, if the customer state is changed from ‘Passive’ to ‘Detractor’, the customer starts spreading negative word of mouth about his mobile network. The next section describes how the WOM flow is amalgamated within the agent’s decision-making process.

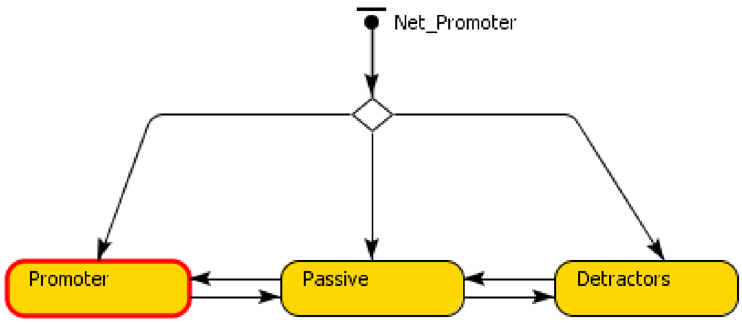


Figure 6.4: UML State Diagram of the WOM Flow

6.3.3 Extending the Agent Behaviour

In the previous development iteration (see Section 5.4), the agent’s decision-making process was based on the customer satisfaction evaluation and the trade-off between the two mobile networks that were modelled in the CubSim model. In this iteration, the social influences are introduced and the agent behavioural rules are modified to accommodate these influences. Extending the agent behaviours in this iteration involves two main tasks: (1) Implementing the WOM flow process that has been described in Section 6.3.2; and (2) modifying the agent’s decision-making process to incorporate social influences. As the first task has been discussed, concentration is now placed on decision-making.

To accommodate social influences, a new behavioural rule has been introduced. This rule is applicable in the case of the passive customers. The new rule is as follows:

If: Customer is Passive &&

Net2_Recommendations>Net1_Recommendations

Then: switch from the Network 1 to the Network 2

Where:

- *Net2_Recommendations* is the sum of the (positive and negative) WOM coming from customers within the personal network and belonging to Network2.
- *Net1_Recommendations* is the sum of the (positive and negative) WOM coming from customers within the personal network and belonging to Network1.

- The underlying assumption of this rule is that the passive customers are normally satisfied or partially satisfied customers, but they are vulnerable to recommendations. For this category of customers, if the recommendations coming from a competing network are greater than the recommendations from their current network they will switch their network.

After extending the agent behaviour, agents take their final decision on switching their network provider based on the amalgamation of interactions among all channels, including social network, interaction with the service provider and interaction with the competing operator. In order to simplify the modelling process, this research assumes that satisfied agents do not look for alternatives and are not influenced by social networks. Contrary to this, dissatisfied agents check for better alternatives from other network. Figure 5.2 shows the state diagram that illustrates how agents take their decisions.

6.3.4 Main Features in CubSim V3 and V4

The implementation processes of the CubSim model Versions 3 and 4 are built on top of the previous versions. These processes are carried out in an iterative manner to incorporate the new futures introduced in the third iteration of this study. This section reports some of the key features that are implemented in the CubSim model Versions 3 and 4. The full documentation of the CubSim model is provided in Appendix A. The model code is available for download on the Google code platform at: http: http://code.google.com/p/cubsim/. A working version of the model can be viewed at http://www.cubsim.com/model/. The java virtual machine (JVM) is needed to run the model.

Monthly churn rate calculation (implemented in V3): To provide more intuitive and relevant results, a new output measure was introduced in Version 3. This measure is the monthly churn rate, which is often reported by the mobile operators as a key performance indicator (KPI). To implement this feature, the method *MonthlyChurn* was added to the main class. This method takes an integer parameter *t* representing the month, and returns a double value corresponding to the churn rate in that month. To display the monthly churn rate, a time plot diagram (AnyLogic component) was used. Figure 6.5 shows a screen shot for the time plot of the monthly churn rate, where the horizontal axis represents month numbers, and the vertical axis denotes the churn rate.

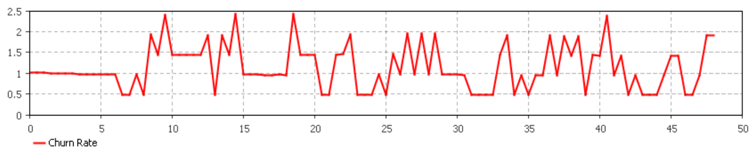


Figure 6.5: A Time Plot of the Monthly Churn Rate

Social network representation (implemented in V3): A distance-based network type (AnyLogic network type) was employed to implement the social network structure described in Section 6.3.1. This type of network connects two agents if the distance between them is less than a predefined maximum value. In this study, this value was set at nine. Figure 6.6 shows the social network representation, where each circle represents a customer. To distinguish the customers of the two mobile networks, two different colours are used to fill the circles. The blue colour denotes Network 1 customers, and the red colour denotes Network 2 customers.

WOM flow (implemented in V4): A new UML state diagram was added to the customer agent in the CubSim model to implement this feature. Figure 6.4 shows the state diagram that encompasses the states and the state transitions. As explained in Section 6.3.2, when an agent state changes from one state to another, the agent starts spreading WOM (positive or negative) to its personal network.

Quantifying social network influences (implemented in V4): Three variables *Net1_Recommendations, Net2_Recommendations and myLinks* and one method *womFunction* were added to the customer agent class to implement this feature. *Net1_Recommendations* and *Net1_Recommendations* variables are used to accumulate the recommendations coming from Network1 and Network2 respectively. Both negative and positive recommendations were considered; the positive recommendations were added and the negative recommendations were subtracted. The method *womFunction* was written to implement the rule that is described in Section 6.3.3.

Enhancing the GUI of the CubSim simulator (implemented in V4): In order to be able to use the CubSim model to test real business scenarios, a set of business graphics components and control elements were added to the CubSim GUI. In the implementation of Version 4, two groups of sliders were added and linked with two parameters: The Customer services and the Marketing aggressive level. By changing the slider position, the user can test the effect of changing the two parameters on customer churn. As shown in Figure 6.6, three different types of chart are used to present both simulation output data and simulation runtime data. The pie chart is used to display at runtime the percentage of churned and retained customers. The time stack chart in Figure 6.6 is used to show the history of contribution of the market shares of the two networks' customers. The time plot diagram in Figure 6.6 was used to display the monthly churn rate. In addition to these charts, an animated graph is used to illustrate the dynamic behaviour of the social network (see Figure 6.7).

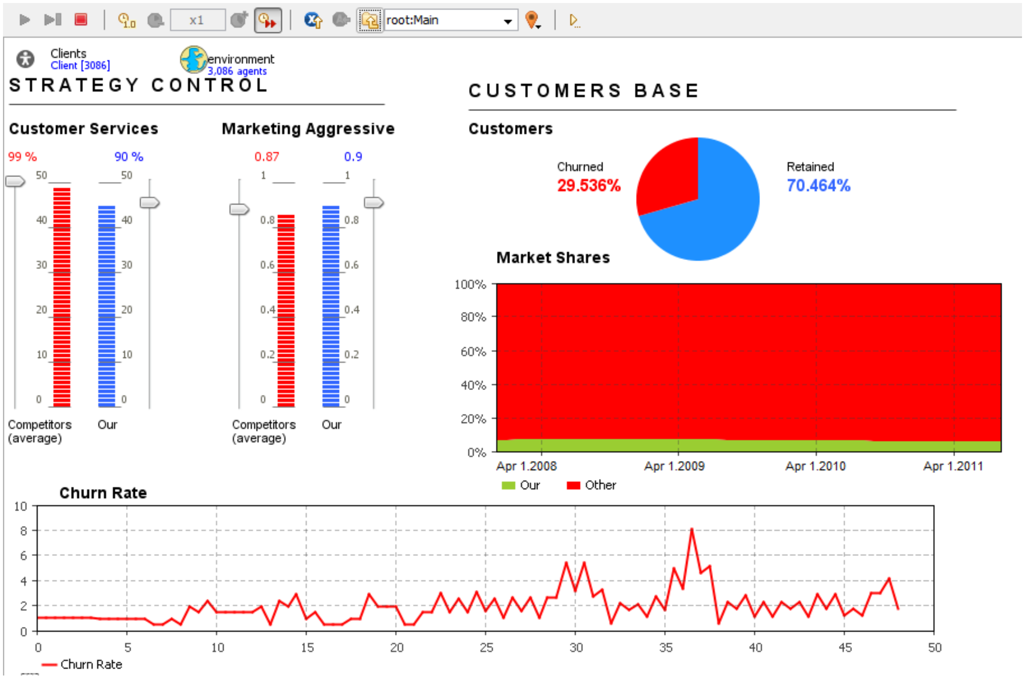


Figure 6.6: The GUI of the CubSim Model

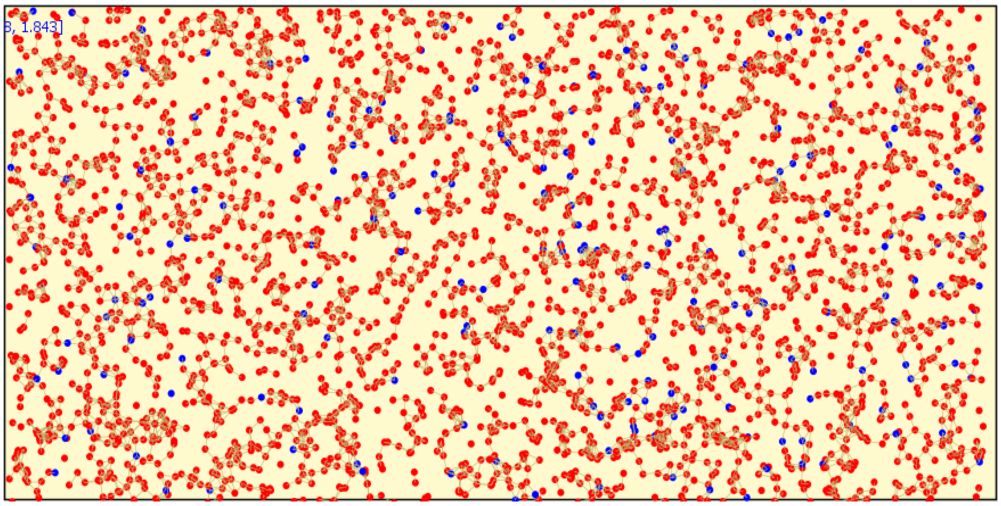


Figure 6.7: A Visualisation of the Social Network Representation in the CubSim Model

6.4 Verification and Validation of the Extended CubSim Model

Similar to the last chapter, three verification methods were used to examine the efficacy of the model: Design walkthroughs, code walkthroughs and model logging. Figure 6.8 shows a screenshot for a model-logging run where the states and the customer agent number were recorded. The aim of this logging run was to verify the correctness of the simulation operation after extending the CubSim model. Based on the CubSim conceptual model introduced in Section 5.3.2, it is expected that some customer agents will churn, although they are satisfied. As shown in Figure 6.8, there are a number of agents who have churned because of the social network effect, not because of dissatisfaction with the service provided.

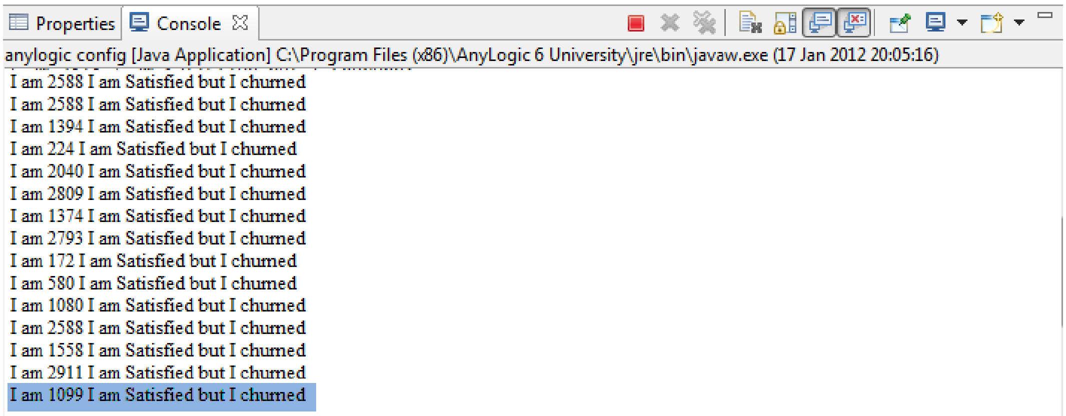


Figure 6.8: A Screenshot for a Model Logging Run

By checking the variables of the highlighted agent, it was possible to verify the logic and correctness of the agent's behavioural rule presented in Section 6.3.3. As shown in Figure 6.9, the agent number 1099 was satisfied because the satisfaction index was greater than the satisfaction threshold, which was set at 60%. The agent had received some negative WOM and the agent state was 'Passive'. The churn flag of this agent showed that he had switched his network.

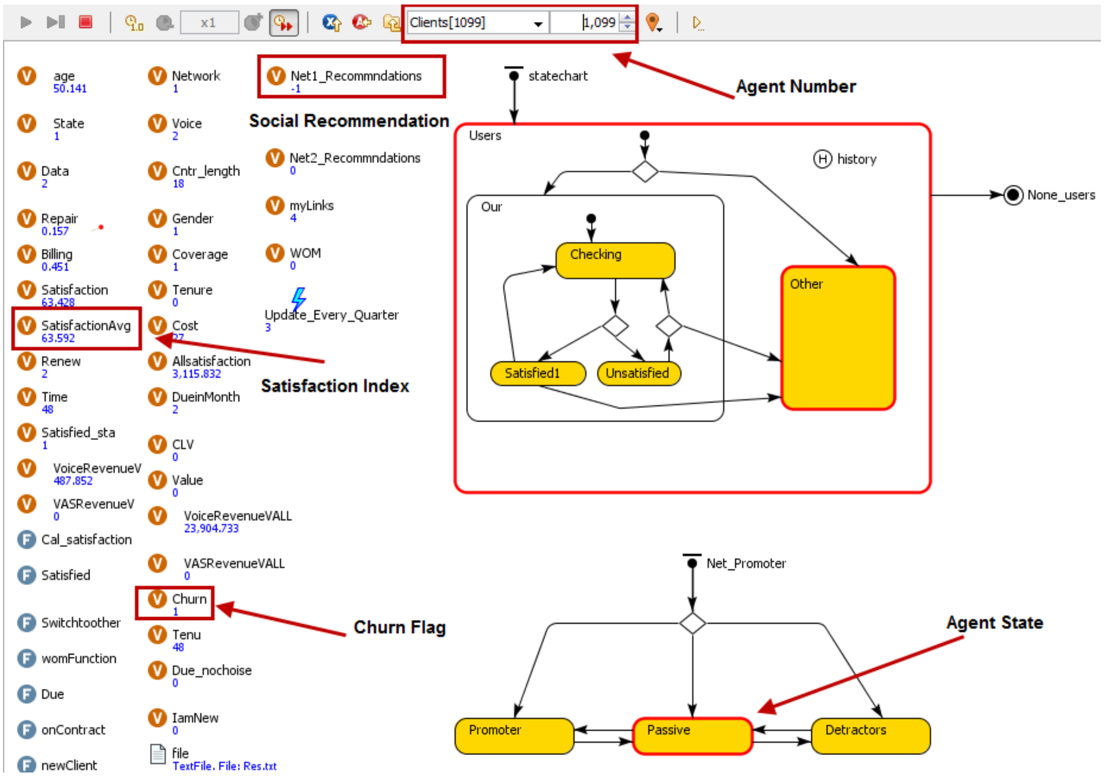


Figure 6.9: The Agent's States and Attributes

The verification processes were performed continuously during model implementation until a satisfactory behaviour of the simulation model was achieved. Prior to performing the validation, the CubSim was calibrated by using the same data used in iteration 1 (see Section 5.5), along with two additional parameters: The number of agents and the social reach diameter. Based on the results of Experiment 1 in Section 5.6, the number of agents was set at 4,000. The social reach diameter or the average size of the personal networks was set to nine as suggested in current literature (see Section 6.3.1). After introducing the monthly churn rate measurement in Version 3 of the CubSim model (see Section 6.3.4), it becomes possible to perform black-box validation at macro-level by comparing the monthly churn rate produced by the CubSim model with the rates reported by the mobile operator.

Experiment 3: Black-Box Validation

There are two objectives of this experiment: (1) To perform a black-box validation at macro-level by comparing annual churn rates predicted by the CubSim model with the rates reported by the mobile operator; and (2) to examine the difference between performance of the CubSim model with and without the social network representation. The mobile operator provided only an average combined monthly churn rate. To perform a black-box validation at macro-level, the performance measurement (monthly churn rate) was converted to a compound annual rate. Then, the simulation runs were repeated several times to average the results. Each simulation run began with a different starting value for the random number seed. By changing the random number seed from run-to-run, the CubSim model produced a range of customer churn values. These values were averaged to produce the means reported in Table 6.4, which shows descriptive statistic results for this experiment.

The results table lists the mean and standard deviation values for the annual churn rates predicted by two versions of the CubSim model: With and without the social network representation. Actual churn rates reported by the mobile operator are also listed in Table 6.4 (see Hutchison Whampoa Limited, 2011).

Year	Without the Social Network				With the Social Network				Reported Average Combined Monthly Churn Rate
	Min.	Max.	SD	Mean	Min.	Max.	SD	Mean	
2008	.483	2.404	.427	1.128	.483	2.404	.427	1.461	1.6
2009	.474	2.415	.603	1.203	.474	2.941	.828	1.582	1.9
2010	.476	1.961	.506	1.027	.980	5.076	1.19	2.445	2.1
2011	.469	2.381	.589	1.155	.457	8.081	1.73	2.229	-

Table 6.4: Descriptive Statistic Results for Experiment 3

Unlike Experiments 1 and 2, statistical tests are not practical here because of the small sample size. Nonetheless, the experiment results show that the average combined monthly churn rates of the extended CubSim model (with the social network) are much closer to the values reported by the mobile operator. These results lead to the conclusion that accounting for the social network influences in the customer behaviour modelling contributes to a more realistic and accurate representation of customers behaviour.

6.5 Simulation Experiments and Analysis

After verification and validation processes, the CubSim model can be used now to generate insights into customer churn behaviour. This section demonstrates the practical use of the model as a decision support tool to investigate various questions relating to possible changes in customer retention strategies. Experiment 4 compares three customer retention strategies: (1) Retaining high value customers only; (2) retaining social leaders only; (3) retaining high value customers and social leaders.

An effective retention strategy focuses on retaining the right customers, not every customer. Resources allocated to retention efforts are normally limited. The strategy of blanket customer retention is inapplicable and can damage the overall value of the customer base if it is applied. Therefore, the core of any retention strategy is to identify which customers are most valuable and why. Retaining high value customers is a dominant customer retention strategy in the mobile market (Richter, Yom-Tov and Slonim, 2010). In that light, Experiment 4 illustrates how the CubSim model is used to compare a new retention strategy based on targeting both high value customers and social leaders at the same time. The premises of this strategy are that: (1) Keeping high value customers is proven a highly effective customer retention strategy; and (2) retaining customers with large social networks can drive positive WOM and, as a result, increase customer retention.

Experiment 4:

There are two objectives of this experiment: (1) To demonstrate a practical use of the CubSim model; (2) to compare the performance of the three customer retention strategies discussed in this section. The performance measurements used in the previous experiments, mainly the number of churned customers, cannot measure the benefit of the different retention strategies in an absolute sense. For instance, the total number of churned customers does not infer any financial implications.

To understand customer churn and its financial implications, a new performance measurement for customer retention is proposed, since losing high value customers is more significant than losing low value customers. Therefore, the new measurement is developed to quantify customer churn in different ways by considering the customers' value. The new comparative performance measurement is stored in a variable named *CLV_Churn_KPI*. The value of this variable is calculated by dividing the summation of the customer values of churned customers by the number of churned customers. The process of calculating customer values is explained in Section 5.4.4. The greater the value of the *CLV_Churn_KPI*, the less favourable the evaluation, and the less effective retention strategies are.

The simulation runs of this experiment were performed in three parts; each part testing one customer retention strategy. The three experiment parts were then executed several times to average the results. During the simulation runs, all the simulation parameters and configurations were fixed to be sure all the performance differences between the three customer retention strategies resulted from changing the retention strategy and not from other effects.

Table 6.5 shows the descriptive statistic results for this experiment. The results table lists the mean and standard deviation values for the variable *CLV_Churn_KPI* grouped by the three customer retention strategies that have been tested. A description of the simulation configuration used is also provided in Table 6.5. The variable *CLV* is an agent attribute that denotes the customer value level, in which there are three levels: High, mid and low, which are represented by the values of 2, 1 and 0 respectively. The variable *Mylinks* denotes the personal network size. The value of this variable was set to equal the mean of the personal sizes. Figure 6.10 shows graphically the effect of the three retention strategies on the comparative performance measurement selected in this experiment.

No	Retention Strategy	Mean	SD	Simulation Configuration
1	Retaining high value customers only	251.012	1.642	CLV>=1
2	Retaining social leaders only	1553.849	2.860	Mylinks >= 5
3	Retaining high value customers and social leaders	185.6225	1.623	CLV>=1 or Mylinks>=5

Table 6.5: Descriptive Statistic Results of Experiment 4

As the experimental results show, applying mixed customer retention strategies targeting both high value customers and customers with a large personal network size outperforms the traditional customer retention strategies, which focuses only on the customers' value.

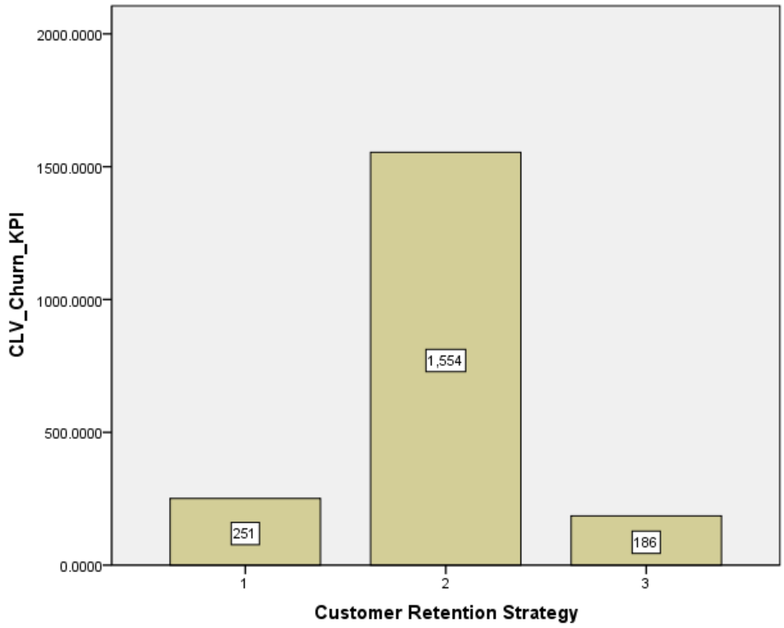


Figure 6.10: The Effect of the New Customer Retention Strategy on the Performance Measurement

6.6 CubSim Model evaluation and discussion

ABMS is a relatively new tool for modelling human behaviour; and, more recently, has been proposed as an effective decision support tool in the field of CRM. The CubSim model presented in this study is an attempt to demonstrate empirically the benefits of using ABMS in customer retention analysis. The CubSim model is neither intended to predict the customer churn of the mobile users under study nor the customer market share in the mobile market. Rather, it is developed to be an exploratory environment to improve and provide insights into the complexity of customer churn. The ultimate goal of the CubSim model is to provide better decision support for the telecommunication manager and to offer a test bed tool for analysing and studying complex customer behaviour by considering the key factors that may affect customer churn.

As explained in Section 3.4.4, two sets of best practice design and implementation guidelines have been integrated into the CubSim model development process to ensure the rigor of the model. The development of the CubSim model follows an incremental and iterative development approach that spans over two iterations to produce four consecutive versions of the model. This approach is used not only to improve the explanatory power of the CubSim model, but also to enhance the research-learning process. The KIDS approach (keep it descriptive, stupid) of Edmonds and Moss (2005) is followed to keep the CubSim model as descriptive as possible. Increasing the complexity (scope and level of details) of the CubSim model requires more data at both micro and macro-level. Since there is only limited data available on the case study used to build the CubSim model, a number of assumptions have been made to build the model. All of these assumptions are based on theoretically plausible foundations reported in the literature.

To demonstrate the utility of the model, four experiments were carried out. Experiment 1 identifies the optimal number of agents to run and experiment with the model. This experiment is very important in reducing the cost (in terms of computational time) needed to construct and run the model. Experiment 2 demonstrates the importance of considering customer acquisition in customer churn analysis, which is in agreement with the results reported in the literature (e.g. Chuan and Yun, 2008). Experiment 3 is performed to establish the validity and the reliability of the model by conducting black-box validation as explained earlier. In addition, the results of this experiment show the importance of WOM on customer retention, which is consistent with the findings of customer churn literature (e.g. Richter, Yom-Tov and Slonim, 2010).

Starting from Experiment 3 outcomes, Experiment 4 demonstrates empirically how insights can be extracted from the CubSim model. The most important finding of this experiment is that targeting high value customers in addition to the customers with large social network outperforms the traditional customer retention strategy, which focuses only on the customers' value. In addition to the highlighted benefits of using the CubSim model, other uses of the model are also possible. Some of these uses are: (1) Make explicit models of causality to stimulate discussion and critical comments; (2) use as a 'what if' scenarios tool to test different strategies; and (3) direct customer churn research and data collection processes.

6.7 Summary

This chapter expands on the previous one and extends the CubSim model to represent interaction among customers. To replicate the characteristics of the real social networks, a network structure based on a minimum number of sociological assumptions and minimal data is used. After building the social network, the WOM flow within the CubSim model is implemented. The WOM flow implementation is inspired by Reichheld's Net Promoter score (Reichheld, 1996), in which the company's customers are classified into three categories: Promoters, passives and detractors. These categories govern the WOM flow and form the social interaction dynamics. Finally, the agent's decision-making process is modified to incorporate social influences. By then, the agents response and adjust their decisions based on two external stimuli and one internal stimulus. The external stimuli are offers from competing networks and WOM from social networks while the internal stimulus is the satisfaction level with the service provider.

In order to simplify the modelling process, this research assumes that satisfied customers do not look for alternatives and are not influenced by friends and family. Contrary to this, dissatisfied customers check for better alternatives from other network. To establish the validity of the model, a black-box validation experiment is performed. After establishing the validity of the model, Experiment 4 is conducted to establish the utility of the model by demonstrating empirically how insights can be extracted from the CubSim model. Drawing on the experimental results, the author is convinced that the CubSim model is able to improve and provide insights into the customer churn problem. In particular, and most interestingly, the experimental results show that applying a mixed customer retention strategy targeting both high value customers and customers with a large personal network size outperforms the traditional customer retention strategies, which focuses only on the customer's value.

Chapter 7: Conclusion

7.1 Overview

This chapter presents the research conclusions and findings. It highlights the key contributions made by this study and discusses its limitations to draw up future research directions. The remainder of this chapter is structured as follows: Section 7.2 provides an overview of the thesis by showing the main theme and the rationale of each chapter. Section 7.3 identifies the research contributions. Section 7.4 discusses how this study achieves its defined objectives. Section 7.5 suggests some future research avenues that will provide further development to research on customer retention. Finally, concluding remarks are given in Section 7.6.

7.2 Research Overview and Findings

This research aimed at developing a generic, reusable, agent-based model to examine the factors affecting customer retention in the UK mobile market by focusing on churn behaviour analysis. The work presented in this thesis falls into four logical parts, corresponding to the four research phases: (1) Awareness of the problem; (2) solutions selection and suggestion; (3) development; and (4) evaluation.

Part 1: This part included Chapters 1 and 2, which covered the phase of problem awareness. After establishing the importance and severity of customer churn in the UK mobile market, traditional analysis tools that are commonly used in customer retention analysis were explored and their limitations were exposed. This part also examined the potential of ABMS to overcome the limitations of traditional tools.

Customer retention is an important issue for any business, and it is more important in mature markets where new customers can only be acquired from competitors. Acquiring new customers in mature markets is not only difficult; it is also expensive in terms of the costs of advertising, setting up new accounts, educating new customers and other costs that are not included when retaining existing ones.

Compared with the traditional research market survey, the superiority of using data mining is well established. Research market survey is often carried out through voluntary questionnaires or interviews. This type of research suffers from a number of limitations, such as relatively high cost, limited access to the population of interest and self-reporting of data during interviews. Data mining, in contrast, can provide continuous and current knowledge of whole customer populations. Although data mining provides very useful insights into customer churn, its utility in capturing the complexity inherent in the mobile market and customer retention is questionable (Twomey and Cadman, 2002).

The literature review reported in this part revealed that decision tree and logistic regression were among the most commonly used data mining models. Empirical customer retention studies suffer from two major limitations, namely that they: (1) Focused on analysing a few specific factors, like customer satisfaction and customer loyalty, neglecting other important factors such as social ties; and (2) focused on customers' characteristics and interactions with the operator, while ignoring interaction with the mobile market and interaction among customers themselves. Other limitations are reported in Table 2.2.

To overcome the limitations identified in the previous studies, this part explored the potential of ABMS to provide a holistic view of customer behaviour in the UK mobile market. An examination of ABMS literature revealed that ABMS was utilised by leading market analysts to gain deeper insight in the market and customer dynamics in different industries (e.g., North et al., 2010). Although ABMS studies are promising, the utilisation of ABMS for business applications is still in its infancy, and existing ABMS models of consumer markets are either very general, or very specific, which limits their ability to capture the distinct characteristics of the mobile market. Therefore, the need for further ABMS studies of the customer behaviour in the mobile market has become paramount. Consequently, the aim of this study was to develop a generic, reusable, agent-based model to examine the factors affecting customer retention in the UK mobile market.

Part 2: This part included Chapters 3 and 4, which covered the selection of solutions and suggestion phase. The importance of this part of the thesis lies on two aspects. First, it identified the general requirements and specifications of the ABMS model that needs to be built. Second, it provided a road map for other researchers to follow when replicating the models developed in this study, and it provided a basis for a critical review of the research methodology. Due to the multidisciplinary nature of this study, a triangulation of multiple data sources was used to construct the artefacts developed. The main data sources utilised in this study were: (a) Previous literature; (b) customers' transactional data provided by a UK mobile operator; and (c) Ofcom, the UK telecoms regulator. The customer transactional data were used to compile the illustrative case study presented in Section 3.3.5. This case study was used to develop decision trees and logistic regression models for modelling customer churn. The case study along with data from the other resources was used to build the CubSim model.

In addition, to identify the limitations of data mining tools in churn analysis, Chapter 4 demonstrated the capabilities and benefits of data mining tools in churn analysis by empirically developing two data mining models and analysing their results. These results revealed that the C5 decision tree model outperformed all the logistic regression models, including a model developed by a data analytics team working for the mobile operator. The C5 decision tree model produced a lift value of 1.598 at 30 percentile compared with 1.4 at the same percentile produced by the data analytics team (see Table 4.8).

Part 3: This part was the most important part of the thesis, which included Chapter 5 and 6 and covered the development phase. In this part, three main challenges of developing ABMS models were addressed: (1) Identifying an appropriate agent architecture and behaviour; (2) identifying a social network structure that is able to represent the characteristics of the social network under study; and (3) identifying a suitable approach to model the flow of WOM.

Chapter 5 focused on customer interaction with the market, including interaction with the service provider, and other competing operators in the market. This chapter presented the first development iteration of the CubSim model. This iteration started with designing the architecture of the customer agent. In comparison with existing architecture reported in the literature, this study adapted a combined architecture based on a balance between reactive decision and utility maximisation (see Section 5.4.2). To model the interaction with the service provider, a micro-economic utility model of the customer satisfaction was used. The main purpose of this model was to create an easy-to-use and understandable method that would consider all key factors that may influence customer satisfaction. The interaction with other competing operators in the market was represented by two variables: Customer services and Marketing aggressive level. If a customer was unsatisfied with his service provider, he/she would start considering alternatives from the competing network based on these two variables.

Chapter 6 focused on interaction among customers. This chapter presented the second development iteration of the CubSim model. To replicate the characteristics of the real social networks and the social influences within the CubSim model, a network structure based on a minimum number of sociological assumptions and minimal data was used. Based on this structure, the WOM flow within the CubSim model was implemented. This implementation was inspired by Reichheld’s Net Promoter score (Reichheld, 1996), in which the company’s customers are classified into three categories: Promoters, passives and detractors.

These categories governed the WOM flow and formed the social interaction dynamics. At the end of the second development iteration of the CubSim model, the conceptual model presented in Figure 5.1 was fully implemented. By then, the influences stemming from the interaction with the market and with other customers were incorporated into the agent behaviour, which dictated the agent decision to stay or to leave the mobile network.

Part 4: This part included parts from Chapters 5, 6 and 7 and focused primarily on the evaluation the CubSim model. The CubSim model was neither intended to predict the customer churn of the mobile users under study nor the customer market share in the mobile market. Rather, it was developed to be an exploratory environment to improve and provide insights into the complexity of customer churn. The ultimate goal of the CubSim model was to provide better decision support for telecommunications managers and to offer a test-bed tool for analysing and studying complex customer behaviour by considering the key factors that might affect customer churn. In light of these considerations, four simulation experiments were performed on the CubSim model to establish the validity and utility of the model.

The CubSim model was developed on a typical desktop PC to make it available for the end users at no extra cost. In order to reduce the model running time, **Experiment 1** identified the optimal number of agents to run and experiment with the model. In this experiment, sensitivity analysis was used to evaluate the sensitivity of the model outputs when the customer pool size was varied.

This experiment is important not only to identify the optimal number of agents that produced reliable results but also to demonstrate the output precision of the model. The output precision of the model denotes to the closeness of the match between the various results produced by the model (North and Macal, 2007, p.12). The CubSim model is a stochastic model as it includes stochastic elements represented in the probability density function, which were used to model the heterogeneity in the customer characteristics (see Section 5.4.3). Stochastic means that the model can produce different outputs given the same inputs (North and Macal, 2007, p.12). Table 5.5 demonstrates the output precision of the model by comparing the model results using different customer pool size.

Experiment 2 demonstrated that customer acquisition has an effect on customer retention, which is in agreement with the results reported in the literature (e.g. Chuan and Yun, 2008). Although the results of this experiment showed differences between the two model implementations, with and without the market growth, these differences did not evidence the superiority of one implementation over another.

Experiment 3 was performed to establish the validity and reliability of the model by conducting black-box validation. In this validation test, the annual churn rates produced by the model were compared with the rates reported by the mobile operator. The experiment results showed that the model with the WOM implementation tended to move in the same direction as the corresponding rates reported by the mobile operator. These results can be used to express the output accuracies (North and Macal, 2007). In addition, these results demonstrated that considering WOM jointly with other factors is important to develop a better understanding of customer retention, which is consistent with the findings of customer churn literature (e.g. Richter, Yom-Tov and Slim, 2010).

Experiment 4 demonstrated empirically how insights can be extracted from the CubSim model. The effect of social leaders on customer churn has proven significant as discussed in Section 2.4.2. The results of this experiment not only offered further evidence on the effect of social leaders on customer retention but also provided new insights for further exploration of the significance of social influences in customer retention. The experiment results showed that applying mixed customer retention strategies targeting both high-value customers and customers with large personal networks outperforms traditional customer retention strategies, which focus only on customers' value. Another interesting result of the experiment was the introduction of new measurements for customer churn, as the total number of churned customers does not infer any financial implications. The proposed measurement quantifies the customer churn in a different way by considering customers' value.

7.3 Research Contributions and Value

The contributions of this research covered three areas: (1) The developed artefacts; (2) the design construction knowledge (i.e. foundations); and (3) the design evaluation knowledge (i.e. methodologies). The main contribution of this study was the CubSim model, which empirically demonstrated the feasibility of using ABMS to develop an artefact to investigate customer retention. Unlike previous studies, which either focused on a few factors or provided a general model for customer behaviour, this study presented a comprehensive model for customer churn based on an empirical case study.

The developed artefacts

The CubSim model, as far as the author is aware, is the first artefact to address customer retention in the mobile market; therefore, its development process is itself a contribution to design science. The CubSim model investigated the key factors that affect customer retention simultaneously and jointly. In this manner, the CubSim model is better suited to account for the dynamics of customer churn behaviour in the UK mobile market than all other existing models.

Moreover, the CubSim model provided an empirical, actionable insight on customer retention. In particular, and most interestingly, the experimental results showed that applying a mixed customer retention strategy targeting both high value customers and customers with a large personal network outperformed the traditional customer retention strategies, which focused only on the customer's value. In comprising with the existence customer retention models, the CubSim model has the following advantages:

- Provide an easy and intuitive way to explain the value and the significance of relationships between variables.
- Enhance the ability of identifying causal relationships.
- Offer tools for long-term planning, and minimise the dependency on data quality.
- Include customer interaction in customer churn analysis.
- Offer tools that take account of the heterogeneity of customers.

The C5 decision tree model developed in Chapter 4 was another contribution under the artefacts category. This model outperformed all other models developed by a data analytics team working for the mobile operator and produced a lift value of 1.598 at 30 percentile compared with 1.4 at the same percentile produced by the data analytics team.

The design construction knowledge

In terms of foundations, this study employed a novel design methodology that utilised modelling and simulation as a new approach to investigate customer retention. This approach contributed significantly to improve problem understanding and solution development. This study contributed to the design construction knowledge by providing detailed descriptions of the artefact development processes. These processes followed sets of best practice guidelines to ensure the rigour and credibility of this study. Moreover, this study employed multiple research methods and tools to strengthen the value and relevance of the research outcomes.

This study also provided novel empirical support for one of the theories that explain causality relationship between customer satisfaction and customer retention (see Section 2.2.2). The theory that was supported is that customer satisfaction alone is not sufficient to explain customer retention and other factors that may affect customer retention need to be considered simultaneously and jointly. This study demonstrated that accounting for the social interaction among customers is of fundamental importance not only to produce more accurate customer retention model but also to gain a better understanding of customer churn.

Another contribution in terms of foundations was the design and implementation of the social network structure within the CubSim model which was based on a minimum number of sociological assumptions and minimal data. This design represented a successful attempt to replicate the characteristics of real social networks. The validity of this design was verified and confirmed by simulation experiments. Modelling the WOM flow in the social network was another foundation-related contribution. This study proposed an original approach to model the WOM flow inspired by Reichheld's Net Promoter score. The originality of this approach was in the integration of customer satisfaction into the WOM process.

The design evaluation knowledge

The contribution of this study to design evaluation knowledge was represented by employing a different mix of evaluation methods to cope with the limitations of some of these methods (see Table 3.3). In addition, using multiple evaluation methods contributed to improving the validity and reliability of the study outcomes. The primary value and benefits of this study to both practitioners and academics are as follows:

Practitioners will be provided with a generic, usable, comprehensive model of customer churn behaviour in the UK mobile market. This model provides practitioners with a valuable, systematic tool to: (1) Make explicit models of causality to stimulate discussion and critical comments; (2) use as a 'what if' scenarios tool to test different strategies; and (3) generate new insights into customer retention by evaluating the effects of different strategies on the market.

Academics will benefit from the blending and cross-fertilisation of different disciplines of social science, management science and information system. They will also benefit from the emergent knowledge from the design and implementation of this study, which can open up a new area of original research.

7.4 Meeting the Research Objectives

The research objectives were formulated at the start of this study and presented in Section 1.3. These objectives are revisited and discussed below to demonstrate how they have been achieved.

Objective 1: *Evaluate the current and most common analysis tools for customer retention to highlight their capabilities and limitations.* This objective was achieved in Chapters 2 and 4. In Chapter 2, the literature review identified decision tree and logistic regression as the most common tools used in customer retention analysis. Through empirical experimentation, Chapter 4 identified the limitations and benefits of decision tree and logistic regression in customer retention analysis.

Objective 2: *Develop an agent-based model for customer interaction with the mobile market, including interaction with the service provider and other competing operators in the market, to address the limitations of traditional retention analysis tools*. This objective was achieved in Chapter 5, in which a generic, reusable agent-based model (CubSim) was developed to examine the factors affecting customer retention in the UK mobile market. The model development process involved creating artificial agents mimicking the attributes and behaviours of their real-world counterparts. The focus in this chapter was on modelling customer interaction with the market. The decisions of the agents were based essentially on evaluating the satisfaction resulting from the interaction with their service provider and comparing the service provided by their network with the service offered by the competing network.

Objective 3: *Extend the agent-based model to incorporate the interaction among customers in order to evaluate the influences of social networks on customer churn*. This objective was achieved in Chapter 6, in which the CubSim model was extended to incorporate social networks. The extension process involves three main tasks: (1) Modelling the social network; (2) modelling the WOM flow; and (3) extending the agent behaviours to incorporate the social network effects. The decisions of the agents in this iteration are based on combining the satisfaction level and the social interaction with family and close friends. Based on the amalgamation of interactions among all channels, including social network, interaction with the service provider and interaction with the competing operator, agents take their final decision on switching their network provider.

Objective 4: *Perform a set of simulation experiments and analyse its results to establish the validity of the agent-based model*. This objective was achieved in Chapters 5 and 6. A sensitivity analysis was performed in Experiment 1 in Chapter 5 to ensure that the simplifications made in the model design do not diminish its credibility. In Experiment 3, a black-box validation experiment was carried out. In this validation test, the annual churn rates produced by the model are compared with the rates reported by the mobile operator.

Objective 5: *Demonstrate the utility of the agent-based model by providing actionable insights into customer retention*. This objective was achieved in Chapters 5 and 6. Experiment 2 in Chapter 5 investigated the effect of customer acquisition on customer retention. The results of this experiment confirmed the effect of acquisition on customer retention, which is in agreement with the literature. Experiment 4 in Chapter 6 examined the effect of imposing a new customer retention strategy on customer churn. The result of this experiment suggested that targeting both high-value customers and social leaders outperforms the traditional customer retention strategies, which focus only on customers' value.

7.5 Research Limitations and Future Works

The lack of adequate data is one of the glaring problems of ABMS (Twomey and Cadman, 2002). This study clearly experiences this limitation, which comes with two consequences: First, lowering the fidelity of the problem representation, and second, limiting the validation options. As noted earlier, a number of theoretically plausible assumptions have been made to overcome the limitation with regard to the fidelity level. Concerning the validation options, Gilbert (2004) stressed that, to validate a model completely, it is necessary to validate both the micro- and macro-level of that model. Historical longitudinal data on customers are not available for this study. Therefore, the CubSim model does not track individuals and it is not possible to validate the model at the micro-level by extracting longitudinal trajectories form the model and comparing it with the historical data. The CubSim model is validated, partially, at the macro level by comparing the compound annual rates produced by the model and the rates reported by the mobile operator, as demonstrated in Experiment 3. In short, it is important to stress that there is no perfect model. As North and Macal (2007) emphasise, no model using ABMS or any other computational modelling approaches will ever be fully validated, but a high degree statistical certainty is still desired. North et al. (2010) cited the famous words of George Box: ‘‘All models are wrong, but some models are useful’’; this is certainly applicable to the CubSim model.

Selecting an appropriate process or mechanism to represent the agent behaviour is a challenging task (Twomey and Cadman, 2002). In addressing this challenge, a trade-off between the fidelity of the model and the descriptive power of the model has been made. This trade-off directs the model development process to be more explicit about the model assumptions to extract the essence of the customer churn problem. Accordingly, this imposes another limitation to the CubSim model, which is that it provides a general outline of customer behaviour and pays less attention to the underlying cognitive processes of customer churn. Unsurprisingly, the trade-off between the fidelity and the descriptiveness power comes at a price, as achieving one leads to doing less well at another. Although there are some limitations attached to the CubSim model, the model results are highly promising, demonstrating the potential of ABMS for both understanding processes behind customer churn and suggesting new strategies for customer retention.

Another limitation of the CubSim model is that it has only two mobile operators in the artificial market, which cannot fully capture the phenomena of market with competition. However, because the aim of this model is to find main rules and patterns between the mobile operator and customer in the real market based on the case study presented in Section 3.3.5, this study makes several assumptions to simplify the model and to avoid making additional assumptions regarding unavailable data. Nonetheless, although the CubSim model was used in the specific case study as a source of data, the author believes that the general model could be adapted to other mobile operators.

To improve on the limitations of this study, future research in the short term should focus on performing more validation experiments by using new data. The promising outcomes of the CubSim model provide a demonstrator to persuade mobile operators to provide more data for further model validation and testing. In addition to continuing to validate the CubSim model, further experiment with different customer retention strategies will be carried out. In the long term, future research will consider increasing the fidelity of the CubSim model by extending the model to incorporate other factors that may affect customer retention including cognitive functions. In addition, future research can be directed toward enhancing the capabilities of the agents of the CubSim model by implementing new features such as co-operation and negotiation, which in turn will lead to modifying the agent architecture and behaviours.

7.6 Concluding Remarks

This research has investigated the use of ABMS for analysing customer retention. A comprehensive model that can directly account for the different interactions among customers and mobile market is developed. Although the model has some shortcomings, the model has been proven useful in the customer retention analysis. The objective of ABMS in general is to explore and analyse the patterns and properties that emerge through the interaction of agents. ABMS is not best suited to build a descriptively accurate or predictive model of customer behaviour. In contrast, data mining has better prediction accuracy than all other tools. To this end, ABMS should not be regarded as a replacement for data mining, however, but perhaps as a necessary complement to it. Consequently, both ABMS and data mining should be part of the methodological tool kit of researchers and practitioners of customer behaviour.

References

Ahn, J.H., Han, S.P. and Lee, Y.S. (2006) "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry", *Telecommunications Policy,* vol. 30, no. 10-11, pp. 552-568.

Alam, S.J., Geller, A. and Tsvetovat, M. (2011) "Networks in Agent-based Social Simulation", in *Agent-Based Models of Geographical Systems* Springer Netherlands, pp. 199-216.

Allsop, D.T., Bassett, B.R. and Hoskins, J.A. (2007) "Word-of-mouth research: principles and applications", *Journal of Advertising Research,* vol. 47, no. 4, pp. 398-411.

Anaman M, Lycett M. (2010) “Toward a model of customer experience: An Action Research study within a mobile telecommunications company”, *12th International Conference on Enterprise information Systems,* 8-12 June 2010, Funchal, Madeira – Portugal.

Au, W.H., Chan, K.C.C. and Yao, X. (2003) "A novel evolutionary data mining algorithm with applications to churn prediction", *Evolutionary Computation, IEEE Transactions on,* vol. 7, no. 6, pp. 532-545.

Aydin, S. and Ozer, G. (2005) "The analysis of antecedents of customer loyalty in the Turkish mobile telecommunication market", *European Journal of Marketing,* vol. 39, no. 7, pp. 910-925.

Bakır, B., Batmaz, İ., Güntürkün, F., İpekçi, İ., Köksal, G. and Özdemirel, N. (2009) "Defect Cause Modeling with Decision Tree and Regression Analysis", *Relation,* vol. 10, no. 1.133, p. 8969.

Barabási, A.L. and Albert, R. (1999) "Emergence of scaling in random networks", *Science,* vol. 286, no. 5439, pp. 509-605.

Baskerville, R.L. and Myers, M.D. (2002) "Information systems as a reference discipline", *Mis Quarterly,* Vol.26, no. 1, pp. 1-14.

Baxter, N., Collings, D. and Adjali, I. (2003) "Agent-Based Modelling—Intelligent Customer Relationship Management", *BT Technology Journal,* vol. 21, no. 2, pp. 126-132.

Berry, M.J.A. and Linoff, G. (2004) *Data mining techniques: for marketing, sales, and customer relationship management,* Wiley Computer Publishing.

Birke, D. and Swann, G.M.P. (2006) "Network effects and the choice of mobile phone operator", *Journal of Evolutionary Economics,* vol. 16, no. 1, pp. 65-84.

Bloemer, J.M.M., Brijs, T., Vanhoof, K. and Swinnen, G. (2003) "Comparing complete and partial classification for identifying customers at risk", *International Journal of Research in Marketing,* vol. 20, no. 2, pp. 117-131.

Bobek, S. and Perko, I. (2006) "Intelligent Agent based Business Intelligence", *University of Maribor, Faculty of Economics and Business, Razlagova ulica14, SI-2000 Maribor,* Slovenia.

Brannon, E., Ulrich, P., Anderson, L., Presley, A., Thommesen, S. and Missam, M. (2000) "Agent-based simulation of the consumer’s apparel purchase decision", *National textile center annual report I98-A09-A1, November,* vol. 98.

Bryson, J.J. (2003) "Action selection and individuation in agent based modelling", *In Proceedings of the Argonne National Laboratories Agent Conference*, pp. 317–330.

Buckinx, W. and Van den Poel, D. (2005) "Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting", *European Journal of Operational Research,* vol. 164, no. 1, pp. 252-268.

Bukvova, H. (2009) "Research as a Process: A Comparison between Different Research Approaches", *Sprouts: Working Papers on Information Systems,* vol. 9, no. 29.

Burez, J. and Van den Poel, D. (2007) "CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services", *Expert Systems with Applications,* vol. 32, no. 2, pp. 277-288.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000) "CRISP-DM 1.0 Step-by-step data mining guide". Available at: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=YTW03084USEN (Accessed:02/04 2012).

Chen, I.J. and Popovich, K. (2003) "Understanding customer relationship management (CRM): People, process and technology", *Business Process Management Journal,* vol. 9, no. 5, pp. 672-688.

Chen, T.Y. (2001) "Using competence sets to analyze the consumer decision problem", *European Journal of Operational Research,* vol. 128, no. 1, pp. 98-118.

Chu, B., Tsai, M. and Ho, C. (2007) "Toward a hybrid data mining model for customer retention", *Knowledge-Based Systems,* vol. 20, no. 8, pp. 703-718.

Chuan, F. and Yun, C. (2008) "The Research on Customer Lifetime Value in Multi-agent Artificial Market", *Fifth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD'08*.

Clarke, K. (2001) "What price on loyalty when a brand switch is just a click away?", *Qualitative Market Research: An International Journal*, vol. 4, no. 3, pp. 160-168.

Collings, D., Reeder, A., Adjali, I., Crocker, P. and Lyons, M. (1999) "Agent Based Customer Modelling", *Computing in Economics and Finance 1999*.

Cook, R.D. and Weisberg, S. (1982) "Criticism and influence analysis in regression", *Sociological methodology,* vol. 13, pp. 313-361.

Coussement, K. and Van den Poel, D. (2008) "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques", *Expert Systems with Applications,* vol. 34, no. 1, pp. 313-327.

Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A.A. and Joshi, A. (2008) "Social ties and their relevance to churn in mobile telecom networks", *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*, ACM, pp. 668-678.

Day, G.S. (1994) "The capabilities of market-driven organizations", *Journal of Marketing,* vol. 58, no. 4, pp. 37-34.

Edmonds, B. (2001) "The use of models-making MABS more informative", *Multi-agent-based simulation: second international workshop, MABS 2000, Boston, MA, USA, July: revised and additional papers,* Springer Verlag, p. 15.

Edmonds, B. and Moss, S. (2005) "From KISS to KIDS–an ‘anti-simplistic’ modelling approach", in: *Multi-Agent and Multi-Agent-Based Simulation,* pp. 130-144.

Furness, P. (2001) "Techniques for customer modelling in CRM", *Journal of Financial Services Marketing,* vol. 5, no. 4, pp. 293-307.

Gandon, F. (2002) "Distributed Artificial Intelligence and Knowledge Management: ontologies and multi-agent systems for a corporate semantic web", *Doctoral school of sciences and technologies of information and communication*, University of Nice Sophia Antipolis.

Garcia, R. (2005) "Uses of agent-based modeling in innovation/new product development research", *Journal of Product Innovation Management,* vol. 22, no. 5, pp. 380-390.

Gerpott, T.J., Rams, W. and Schindler, A. (2001) "Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market", *Telecommunications Policy,* vol. 25, no. 4, pp. 249-269.

Gilbert, N. (2004). Open problems in using agent-based models in industrial and labor dynamics. In R. Leombruni & M. Richiardi (Eds.), *Industry and labour dynamics: The agent-based computational approach* (pp. 401–405). Singapore: World Scientific.

Gilbert, N. (2007) "Computational social science: Agent-based social simulation", *Multi-Agent Models and Simulation for Social and Human Sciences,* Oxford, The Bardwell-Press.

Goldsman, D. (2007) "Introduction to simulation", *Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come,* IEEE Press Piscataway, NJ, USA, , pp. 26-35.

Gorman, D.M., Mezic, J., Mezic, I. and Gruenewald, P.J. (2006) "Agent-based modeling of drinking behavior: a preliminary model and potential applications to theory and practice", *American Journal of Public Health,* vol. 96, no. 11, pp. 2055.

Graziano, A.M. and Raulin, M.L. (2009) *Research methods: A Process of Inquiry*, (7th ed.). Boston, MS: Allyn & Bacon.

Gross, D. and Strand, R. (2000) "Can agent-based models assist decisions on large-scale practical problems? A philosophical analysis", *Complexity*, vol. 5, no. 6, pp. 26–33.

Grzybowski, L. (2008) "Estimating Switching Costs in Mobile Telephony in the UK", *Journal of Industry, Competition and Trade*, vol. 8, no. 2, pp. 113-132.

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N. and Sriram, S. (2006) "Modeling customer lifetime value", *Journal of Service Research*, vol. 9, no. 2, pp. 139-155.

Hadden, J., Tiwari, A., Roy, R. and Ruta, D. (2007) "Computer assisted customer churn management: State-of-the-art and future trends", *Computers & Operations Research*, vol. 34, no. 10, pp. 2902-2917.

Hamill, L. and Gilbert, N. (2009) "Social circles: A simple structure for agent-based social network models", *Journal of Artificial Societies and Social Simulation*, vol. 12, no. 2, pp. 3-18.

Hevner, A.R., March, S.T. & Park, J. 2004, "Design Science in Information Systems Research", MIS Quarterly, vol. 28, no. 1, pp. 75-105.

Hoyer and Macinnis (2009) *Consumer Behavior*, 5th edn, South-Western.

Huang, C.Y., Sun, C.T., Hsieh, J.L. and Lin, H. (2004) "Simulating SARS: Small-world epidemiological modeling and public health policy assessments", *Journal of Artificial Societies and Social Simulation*, vol. 7, no. 4, pp. 100-131.

Hung, S., Yen, D.C. and Wang, H. (2006) "Applying data mining to telecom churn management", *Expert Systems with Applications*, vol. 31, no. 3, pp. 515-524.

Hutchison Whampoa Limited (2011) *Telecommunication Annual Reports 2007-2010*. Available at: http://www.hutchison-whampoa.com/eng/investor/annual/annual.htm (Accessed: 01/22 2012).

Hwang, H., Jung, T. and Suh, E. (2004) "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry", *Expert Systems with Applications,* vol. 26, no. 2, pp. 181-188.

ITProPortal (2011) *Mobile Penetration in UK to Exceed 150% by 2016 Claims Research*. Available at: http://www.itproportal.com/2011/11/24/mobile-penetration-uk-exceed-150-2016-claims-research/ (Accessed: 02/04 2012).

Kamakura, W., Mela, C.F., Ansari, A., Bodapati, A., Fader, P., Iyengar, R., Naik, P., Neslin, S., Sun, B. and Verhoef, P.C. (2005) "Choice models and customer relationship management", *Marketing Letters*, vol. 16, no. 3, pp. 279-291.

Karahoca, A., Karahoca, D. and Aydin, N. (2007) "GSM Churn Management Using an Adaptive Neuro-Fuzzy Inference System", *The 2007 International Conference on Intelligent Pervasive Computing, IPC*, pp. 323-330.

Keaveney, S.M. (1995) "Customer switching behavior in service industries: An exploratory study", *The Journal of Marketing*, vol.59, April, pp. 71-82.

Kim, H.S. and Yoon, C.H. (2004) "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market", *Telecommunications Policy,* vol. 28, no. 9-10, pp. 751-765.

Kim, S.Y., Jung, T.S., Suh, E.H. and Hwang, H.S. (2006) "Customer segmentation and strategy development based on customer lifetime value: A case study", *Expert Systems with Applications,* vol. 31, no. 1, pp. 101-107.

Kuhn Jr, J.R., Courtney, J.F., Morris, B. and Tatara, E.R. (2010) "Agent-based analysis and simulation of the consumer airline market share for Frontier Airlines", *Knowledge-Based Systems,* vol. 23, no. 8, pp. 875-882.

Kyrylov, V., Kyrylova, T. and Kanwal, A. (2004) "Agent-Based Simulation of the Telecommunications Services Market in a Metropolitan Area", *SIMULATION SERIES,* vol. 36, no. 2, pp. 62-70.

Lee, J., Lee, J. and Feick, L. (2001) "The impact of switching costs on the customer satisfaction-loyalty link: mobile phone service in France", *Journal of Services Marketing,* vol. 15, no. 1, pp. 35-48.

Lemmens, A. and Croux, C. (2006) "Bagging and boosting classification trees to predict churn", *Journal of Marketing Research,* vol. 43, no. 2, pp. 276-286.

Li, E., Lu, W. and Zhou, Y. (2009) "A Multi-agent Simulation Paradigm of Customer Retention Management", *2009 WRI World Congress on Computer Science and Information Engineering,* IEEE, pp. 581-589.

Li, X., Mao, W., Zeng, D. and Wang, F.Y. (2008) "Agent-Based Social Simulation and Modeling in Social Computing", *Lecture Notes in Computer Science,* vol. 5075, pp. 401-412.

Lima, E., Mues, C. and Baesens, B. (2009) "Domain knowledge integration in data mining using decision tables: case studies in churn prediction", *Journal of the Operational Research Society,* vol. 60, no. 8, pp. 1096-1106.

Liu, C.T., Guo, Y.M. and Lee, C.H. (2010) "The effects of relationship quality and switching barriers on customer loyalty", *International Journal of Information Management,* vol. 31, no.1, pp,71-79.

Ma, T. and Nakamori, Y. (2005) "Agent-based modeling on technological innovation as an evolutionary process", *European Journal of Operational Research,* vol. 166, no. 3, pp. 741-755.

Macal, C.M. and North, M.J. (2007) "Agent-based modeling and simulation: desktop ABMS", *Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come,* IEEE Press, pp. 95-102.

Macal, C.M. and North, M.J. (2005) "Tutorial on agent-based modeling and simulation", *Proceedings of the 37th conference on Winter simulation*Winter Simulation Conference, pp. 2-8.

Maes, P. (1991) "The agent network architecture (ANA)", *SIGART Bulletin,* vol. 2, no. 4, pp. 115-120.

March, S.T. and Smith, G.F. (1995) "Design and natural science research on information technology", *Decision Support Systems,* vol. 15, no. 4, pp. 251-266.

Mattison, R. (2005) *The Telco Churn Management Handbook,* XiT Press, Oakwood Hills, Illinois.

McIlroy, A. and Barnett, S. (2000) "Building customer relationships: do discount cards work?", *Managing Service Quality,* vol. 10, no. 6, pp. 347-355.

Microsoft (2006) *Windows Live: The Anatomy of Friendship*. Available at: http://www.darrenstraight.com/blog/downloads/2006/11/Buddy%20Barometer%20FINAL.pdf (Accessed: 01/20 2012).

Milgram, S. (1967) "The small world problem", *Psychology today,* vol. 2, no. 1, pp. 60-67.

Mingers, J. (2001) "Combining IS research methods: towards a pluralist methodology", *Information systems research,* vol. 12, no. 3, pp. 240-259.

Mobile Marketing Magazine (2009) *Study Says Mobile Churn is Approaching 40%*. Available at: http://www.mobilemarketingmagazine.co.uk/2008/01/ study-says-mobi.html (Accessed: 01/27 2012).

Mulder, M.B. (1987) "Adaptation and evolutionary approaches to anthropology", *Man*, pp. 25-41.

Müller, J. (1999) "The right agent (architecture) to do the right thing", *Intelligent Agents V: Agents Theories, Architectures, and Languages*, pp. 211-225.

Neslin, S.A., Gupta, S., Kamakura, W., Lu, J. and Mason, C.H. (2006) "Defection detection: Measuring and understanding the predictive accuracy of customer churn models", *Journal of Marketing Research*, vol. 43, no. 2, pp. 204-211.

Ngai, E., Xiu, L. and Chau, D. (2009) "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Applications*, vol. 36, no. 2P2, pp. 2592-2602.

Niehaves, B.; Becker, J.: "Design Science Perspectives on ITConsulting", in: Lehner, F.; Nösekabel, H.; Kleinschmidt, P. (Hrsg.): Tagungsband 1 der Multikonferenz Wirtschaftsinformatik in Passau 2006.Berlin, 2006 , pp.7-17.

Nilsson, F. and Darley, V. (2006) "On complex adaptive systems and agent-based modelling for improving decision-making in manufacturing and logistics settings: Experiences from a packaging company", *International Journal of Operations & Production Management*, vol. 26, no. 12, pp. 1351-1373.

Nisbet, R., Elder, J., Elder, J.F. and Miner, G. (2009) *Handbook of statistical analysis and data mining applications*, Academic Press.

Nokia Siemens Networks (2009) *How to generate customer loyalty in mobile markets, Global Summary Report*. Available at: http://www.nokiasiemensnet works.com/sites/default/files/document/AR_Mobile_A4_0107.pdf (Accessed: 01/27 2012).

North, M.J. and Macal, C.M. (2007) *Managing business complexity: discovering strategic solutions with agent-based modeling and simulation,* Oxford University Press, USA.

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. (2010) "Multiscale agent-based consumer market modeling", *Complexity,* vol. 15, no. 5, pp. 37-47.

Office of Communications (Ofcom) (2011b) *The Consumer Experience 2011*. Available at: http://stakeholders.ofcom.org.uk/binaries/research/consumer-experience/tce-11/research_report_of511a.pdf (Accessed: 02/05 2012).

Office of Communications (Ofcom) (2011a) *Communications market report 2011*. Available at: http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr11/UK_CMR_2011_FINAL.pdf (Accessed: 02/04 2012).

Purao, S. (2002) "Design research in the technology of information systems: Truth or dare", *Pennsylvania State University,* Available at: http://purao.ist.psu.edu/working-papers/dare-purao.pdf (Accessed: 02/05 2012).

Railsback, S.F. (2001) "Concepts from complex adaptive systems as a framework for individual-based modelling", *Ecological Modelling,* vol. 139, no. 1, pp. 47-62.

Ranaweera, C. and Prabhu, J. (2003) "The influence of satisfaction, trust and switching barriers on customer retention in a continuous purchasing setting", *International Journal of Service Industry Management,* vol. 14, no. 4, pp. 374-395.

Rand, W. and Rust, R.T. (2011) "Agent-based modeling in marketing: Guidelines for Rigor", *International Journal of Research in Marketing*, vol. 28, no. 3, pp. 181-193.

Reichheld, F.F. (1996) "Learning from customer defections", *Harvard business review*, vol. 74, pp. 56-70.

Reichheld, F.F. (1993) "Loyalty-based management", *Harvard business review*, vol. 71, pp. 64-64.

Renovales, R.D. and Orlando, F., 2006. Agent-Based Modeling and Simulation: A Survey. ESI6532, pp.1-4. Available at: http://pegasus.cc.ucf.edu/~wangyan/esi6532/papers/ESI6532_RaquelRenovales_paper.pdf (Accessed: 01/27 2012).

Richetin, J., Sengupta, A., Perugini, M., Adjali, I., Hurling, R., Greetham, D. and Spence, M. (2010) "A micro-level simulation for the prediction of intention and behavior", *Cognitive Systems Research*, vol. 11, no. 2, pp. 181-193.

Richter, Y., Yom-Tov, E. and Slonim, N. (2010) "Predicting customer churn in mobile networks through analysis of social groups", *Proceedings of the 2010 SIAM International Conference on Data Mining (SDM 2010)*.

Rudowsky, I. (2004) "Intelligent Agents", *Communications of the Association for Information Systems,*vol.14, no.1,pp. 275-290.

Said, L.B., Bouron, T. and Drogoul, A. (2002) "Agent-based interaction analysis of consumer behavior", *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*ACM New York, NY, USA, pp. 184-190.

Sanchez, S.M. and Lucas, T.W. (2002) "Exploring the world of agent-based simulations: simple models, complex analyses: exploring the world of agent-based simulations: simple models, complex analyses", *Proceedings of the 34th conference on Winter simulation: exploring new frontiers* Winter Simulation Conference, pp. 116-122.

Seo, D., Ranganathan, C. and Babad, Y. (2008) "Two-level model of customer retention in the US mobile telecommunications service market", *Telecommunications Policy,* vol. 32, no. 3-4, pp. 182-196.

Shah, S., Roy, R., and Tiwari, A. "Technology selection for human behaviour modelling in contact centres", In Decision Engineering Report Series, Cranfield University Press, UK, 2005. Available at: http://hdl.handle.net/1826/1212 (Accessed: 02/04 2012).

Shi, D.F. and Brooks, R.J. (2007) "The range of predictions for calibrated agent-based simulation models", *Simulation Conference, 2007 Winter* IEEE, pp. 1198-1205.

Siebers PO and Aickelin U (2008) 'Introduction to Multi-Agent Simulation'. In: Adam F and Humphreys P (Eds.). Encyclopedia of Decision Making and Decision Support Technologies, Pennsylvania: Idea Group Publishing, pp 554-564.

Siebers, P.O., Aickelin, U., Celia, H. and Clegg, C. (2007) "A multi-agent simulation of retail management practices", *Proceedings of the 2007 summer computer simulation conference*Society for Computer Simulation International San Diego, CA, USA, pp. 959-965.

Simon, H.A. (1996) *The sciences of the artificial,* the MIT Press.

Stahl, H.K., Matzler, K. and Hinterhuber, H.H. (2003) "Linking customer lifetime value with shareholder value", *Industrial Marketing Management,* vol. 32, no. 4, pp. 267-279.

Streukens, S. and De Ruyter, K. (2004) "Reconsidering nonlinearity and asymmetry in customer satisfaction and loyalty models: An empirical study in three retail service settings", *Marketing Letters,* vol. 15, no. 2, pp. 99-111.

Thomas, G. (2011) "A Typology for the Case Study in Social Science Following a Review of Definition, Discourse, and Structure", *Qualitative Inquiry,* vol. 17, no. 6, pp. 511.

Thomas, J.S. (2001) "A methodology for linking customer acquisition to customer retention", *Journal of Marketing Research,* vol. 38, no. 2, pp. 262-268.

Tikkanen, H. and Alajoutsijärvi, K. (2002) "Customer satisfaction in industrial markets: opening up the concept", *Journal of Business & Industrial Marketing,* vol. 17, no. 1, pp. 25-42.

Turel, O. and Serenko, A. (2006) "Satisfaction with mobile services in Canada: An empirical investigation", *Telecommunications Policy,* vol. 30, no. 5-6, pp. 314-331.

Twomey, P. and Cadman, R. (2002) ‘Agent-Based Modeling of Customer Behavior in the Telecoms and Media Markets’, *Journal of Policy, Regulation and Strategy for Telecommunications,* vol 4, no 1, pp. 56-63.

Ulaga, W. and Eggert, A. (2006) "Relationship value and relationship quality: Broadening the nomological network of business-to-business relationships", *European Journal of Marketing,* vol. 40, no. 3/4, pp. 311-327.

Vag, A. (2007) "Simulating changing consumer preferences: A dynamic conjoint model", *Journal of Business Research,* vol. 60, no. 8, pp. 904-911.

Vaishnavi, V. & Kuechler, W. 2004, Design Research in Information Systems [Homepage of Association for Information Systems],Available at: http://desrist.org/design-research-in-information-sys tems (Accessed: 02/04 2012).

Verbeke, W., Martens, D., Mues, C. and Baesens, B. (2010) "Building comprehensible customer churn prediction models with advanced rule induction techniques", Expert Systems with Applications, vol. 38, no. 3, pp. 2354–2364.

Vuk, M. and Curk, T. (2006) "ROC curve, lift chart and calibration plot", *Metodoloski zvezki*, vol. 3, no. 1, pp. 89-108.

Watts, D.J. (2004) "The" new" science of networks", *Annual review of sociology*, vol.30, pp. 243-270.

Wu, F. (1998) "An experiment on the generic polycentricity of urban growth in a cellular automatic city", *Environment and Planning B: Planning and Design*, vol. 25, no. 5, pp. 731-752.

XJ Techologies, Anylogic 6.6.0. Available at: http://www.xjtek.com/ (Accessed: 02/04 2012).

Ye, Y., Liu, S. and Li, J. (2008) "A Multiclass Machine Learning Approach to Credit Rating Prediction", *Proceedings of the 2008 International Symposiums on Information Processing-Volume*, IEEE Computer Society, pp. 57-65.

Yu, W., Jutla, D.N. and Sivakumar, S.C. (2005) "A churn-strategy alignment model for managers in mobile telecom", *Communication Networks and Services Research Conference, Proceedings of the 3rd Annual*, pp. 48-56.

Zhang, X., Liu, Z., Yang, X., Shi, W. and Wang, Q. (2010) "Predicting customer churn by integrating the effect of the customer contact network", *Conference on Service Operations and Logistics and Informatics (SOLI)*, IEEE, pp. 392-340.

Appendix A: CubSim Model Documentation