PREDICTING CUSTOMERS' CHURN USING DATA MINING TECHNIQUES (MACHINE LEARNING APPROACH)

ABDUL AHMED ABDUL

OCTOBER, 2021

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY, BAYERO UNIVERSITY KANO

THESIS ADIVISOR
MR. ISMAIL ABU ZIBIRI

COORDINATOR
DR. BELLO BICHI

HEAD OF DEPARTMENT
DR. HABIBA KAKUDI

Submitted in partial fulfilment of the requirements for the degree of Masters of Computer Applications

ABSTRACT

Customer churn is a major problem and one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecommunication field, companies are seeking to develop means to predict potential customer to churn. Customers are the most important assets in any industry since they are considered as the main profit source. Churners are persons who quit a company's service for some reasons. Companies should be able to predict the behaviour of customer correctly in order to reduce customer churn rate. Customer churn has emerged as one of the major issues in every Industry. Researches indicates that it is more expensive to gain a new customer than to retain an existing one. In order to retain existing customers, service providers need to know the reasons of churn, which can be realized through the knowledge extracted from the data. To prevent the customer churn, a churn prediction model is developed in this project which can assist companies to predict customers who are most likely to churn. It uses machine learning techniques such as Logistic Regression, Decision Trees, and Random Forest to identify the primary determinants of customer churn along with the algorithm fit for such predictions. The dataset contains demographic details of customers, their data usage and the type of service they receive from the company. It comprises of churn data of 3333 customers spread over 11 attributes obtained from Global AI. Further on this investigation, the usage of the above-mentioned algorithms is described for predicting customer churn. Also, it has been found that the significance of data services is constantly increasing for the subscribers. These findings will be useful to both service providers and researchers to comprehend the service quality parameters considering mobile subscribers view point. This can help to re-look the existing benchmarks and increase customer satisfaction. However, the best results were obtained by applying Random Forest algorithm.

Background of study: Emphasizing the higher costs associated with attracting new customers compared with retaining existing customers, and the fact that long-term customers tend to produce more profits, (Verbeke et al. 2011) assert that customer retention increases with profitability. Many competitive organizations have realized that a key strategy for survival within the industry is to retain existing customers. (Tsai and Chen. 2010) argued that *"*this leads to the importance of churn management.*"* Customer churn represents a basic problem within the competitive atmosphere of telecommunication industry. The telecommunications sector has become one of the main industries in developed and developing countries. The technical progress and the increasing number of operators raised the level of competition (Gerpott TJ, Rams W, Schindler A, 2001). Companies are working hard to survive in this competitive market depending on multiple strategies. Tree main strategies have been proposed to generate more revenues (Wei CP, Chiu IT, 2002):

- acquire new customers,
- upsell the existing customers, and
- increase the retention period of customers.

However, comparing these strategies taking the value of return on investment (RoI) of each into account has shown that the third strategy is the most profitable strategy (Wei CP, Chiu IT, 2002), proves that retaining an existing customer cost much lower than acquiring a new one (Qureshii SA, et al. 2013), in addition to being considered much easier than the upselling strategy (Ascarza E, 2016). To apply the third strategy, companies have to decrease the potential of customer's churn, known as "the customer movement from one provider to another" (Bott. 2014). The development and digitalization of the world has led to new ways of doing business and companies all over the globe have been forced to adapt (R. Shkurti and A. Muca, 2014). Subscription based services are one of the outcomes of the explosive digitalization that has taken the world by storm and with this comes both possibilities and

challenges that require modern day solutions (C. Blank and T. Hermansson, 2018). Digitalization has not only changed the way business is conducted but the abundance of information available has also led to consumers facing a higher supply of subscription-based services. This can be viewed as a challenge for companies since retaining customers can potentially become more difficult. Digitalization within companies can lead to a decrease in labour costs, an increase in efficiency and a better overview of the company's operations within the organization (R. Shkurti and A. Muca, 2014). All of this is essential for staying competitive, and to gain an edge over other companies.

As information technology is a growing trend, the available amount of data and information has increased significantly during the past years. This rapid growth has enabled storage and processing of great amounts of data while increasing the necessity of automatically finding valuable information and creating knowledge (D. Buo and M. Kjellander, 2014). With meaningful information extracted from the stored data, firms can make appropriate decisions in order to grow the business. With this growth, the use of data mining techniques and machine learning has increased, due to its ability of handling and analysing great amounts of data (K. Mishra and R. Rani, 2017). The digitalization has also brought forward an ongoing trend to improve current data processing activities as a part of customer relationship management (CRM) strategies. The idea of knowledge management and customer relationship management has lately obtained more attention in the subscription-based business model and the concepts focus on distribution of resources to activities that are customer-centric, to be able to increase competitive advantages (H. Gebert, 2003 and M. Sergue, 2020). Customer knowledge is the knowledge that businesses can obtain through interaction with their customers (F. Khodakarami and Y. Chan, 2014). Customer relationship management systems are systems that support interaction between businesses and customers with the objective of collecting, storing and analysing data to get an overview of their customers (F. Khodakarami and Y. Chan,

2014). Such systems have evolved through the past years and by using technology and different data analysis tools, enterprises can find patterns in customer's behaviour, which would be almost impossible to discover manually (D. Buo and M. Kjellander, 2014). Such patterns could vary from a customer's purchasing behaviour or patterns related to customer churning. In a subscription-based business model, a fundamental part of success is to minimize the rate of customers ending their subscriptions, in other words, to minimize churn (D. Buo and M. Kjellander, 2014).

Customer churning refers to the action of when a customer chooses to abandon their service provider (D. L. Garcia, et. al 2017). The term is relatively new and has gained more relevance with the emergence of online services. Firms across the globe recognize customer churning as a great loss since they have already invested in attracting these customers. This is one of the major reasons that customer retention is beneficial for a firm. Customers can churn for many reasons and it is hard to pinpoint a general reason for churning. The availability of information has given consumers a bargaining power, and nowadays customers can easily find the service provider, which provides the same product with a more satisfying deal (N. Gordini and V. Veglio, 2017). To manage this, firms invest in customer churn prediction, which means that companies try to predict which of their customers will churn, so that they can apply preventative measures. These preventive measures could differ depending on the reason a customer might churn, and could be for example, offering a lower price or including an extra service. As mentioned earlier, analysing customer behaviour serves as the basis for predicting customers who might churn, which is important for many reasons. One reason is that for companies who rely on subscription-based income, it can make a big difference on whether they can keep a steady income level or if they need to make changes to their services to keep customers. Another reason is that, compared to retaining customers, attracting new ones is costlier and firms can save money by retaining their existing customer base (I. Ullah, 2019).

More so, with new companies continuously emerging in the telecommunications industry, and new customers to those companies sparse, markets have matured to a point where they have saturated, so possible new customer acquisition is scarce. Because of this, companies are beginning to recognise that their most valuable assets are their existing customers. Holding onto existing customers is not an undemanding procedure, with further complications arising from governing bodies such as NCC. These authorities have emerged to act in the customers' best interests, and the promotion of healthy competition amongst service providers to ensure that the customer has a choice, and services are not monopolised. A further challenge that is commonly being recognised regarding the acquisition of new customers is lack of data, making targeted sales campaigns difficult to deploy.

In today's business world companies are recognising that customer value and increased revenue is more likely to come from their existing customer base than from new customer acquisition. The reasoning behind this is that companies know their existing customers, already have a relationship with them, and amplitude of data on them. In common recognition of this problem, industry has seen an emergence of customer relationship management (CRM) products. Software companies have realised that CRM has become a receptive topic within industry and therefore an area of potential wealth and opportunity for the development and promotion of products that promise to not only boost customer retention, but also for increasing selling opportunities (Hadden J., et al 2007).

Also, in the past years, companies have been able to store and process huge amounts of data while realizing that being customer-centric was becoming a main requirement to stand out of the competition. Indeed, due to saturated markets, focusing on Customer Relationship Management (CRM) in order to retain existing customer base is not optional anymore, but an absolute necessity for competitive survival. In its research for Bain and Company, Frederick Reichheld stated that the cost of

acquiring a new customer could be higher than that of retaining a customer by as much as 70%, and that increasing customer retention rates by a mere 5% could increase profits by 25% to 95%. More generally, data-driven decision making is way for businesses to make sure their next move will benefit both them and their customers. Almost every company, especially in the Tech ecosystem has now put into place a tracking process, to gather data related to their customers' behaviour. The data to track varies along with the specific business model of each company and the problem service they aim to address. By analysing how, when and why customers behave a certain way, it is possible to predict their next steps and have time to work on fixing issues beforehand. This prediction and quantification of the risk of losing customers can be done globally or individually and is mainly used in areas where the product or service is marketed on a subscription basis. The prediction of churn is generally done by studying consumer behaviour or by observing individual behaviour that indicates a risk of attrition. It involves the use of modelling and machine learning techniques that can sometimes use a considerable amount of data.

Research Objectives and Goals: As aforementioned, the digitalization and the access to information is in large proportions, companies are under constant threats of customer churning, in particular subscription-based service providers that rely on income. Hence, it is important to study churn prediction and to support companies in this problem. The purpose of this study is to contribute knowledge in the field of churn prediction in a subscription-based service context and to develop a customer profiling methodology for predicting churn in advance. In this study we employ different machine learning techniques, in order to investigate the effects on the performance of churn prediction models in a subscription-based service context. More Over, the goal is to propose and evaluate different approaches and techniques of how

machine learning can be used to predict churn; hence creating value for businesses offering subscription-based services. The objectives of this study are identified as follows:

- Identify and select appropriate mathematical (i.e., statistical and data mining) techniques for developing predictive models.
- compare different types of algorithms for the built model, and
- study the effects of balancing the training dataset for the considered model

Research Design: As aforementioned, we decided to use the machine learning process proposed by Marsland (2014) as shown below. The proposed machine learning process includes data collection and preparation, feature selection, algorithm choice, parameter and model selection, training and evaluation. We do not consider the proposed machine learning process as a linear process; there is a lot of back and forth between the steps in other to create the most optimal model, in our case a churn prediction model.

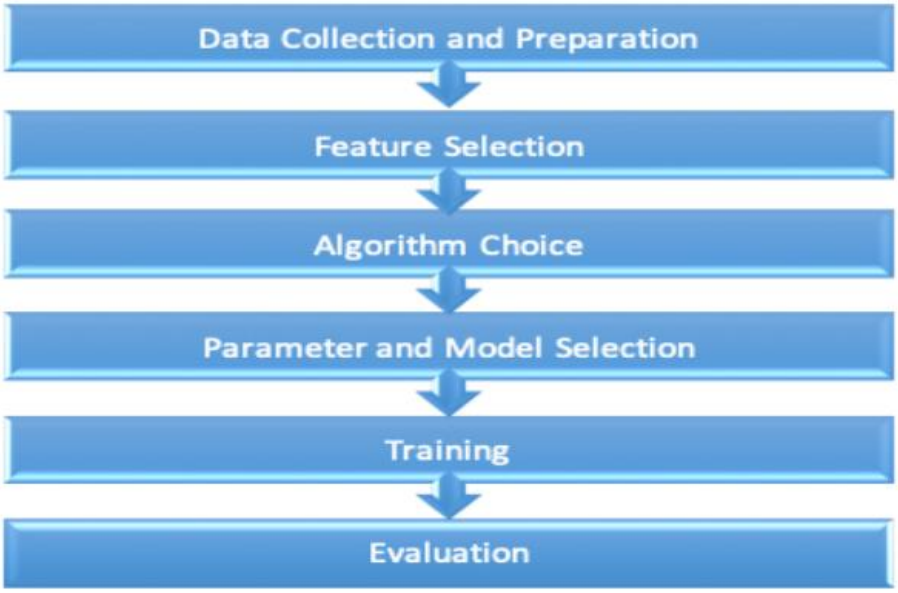

Figure 1: Machine learning process proposed by Marsland (2014)

Conclusion: Churn prediction is one of the most effective strategies used in telecom sector to retain existing customers. It leads directly to improved cost allocation in customer relationship management activities, retaining revenue and profits in future. It also has several positive

indirect impacts such as increasing customer's loyalty, lowering customer's sensitivity to competitors marketing activities, and helps to build positive image through satisfied customers. The results predicted by the Random Forest algorithm were the most efficient with an accuracy of 92.1%. Therefore, companies that want to prevent customer churn should utilize this algorithm thereby giving them more flexibility and providing additional services such as device protection and multiple phone lines proves to be of little value to customer attrition.

More so, focusing on enhancing the experience of loyal customers who have stayed with the company for long will prove worthwhile, ensuring their retention. The ability to identify customers that aren't happy with provided solutions allows businesses to learn about product or pricing plan weak points, operation issues, as well as customer preferences and expectations to proactively reduce reasons for churn.

REFERENCES

Ascarza E, Iyengar R, and Schleicher (2016) M. The perils of proactive churn prevention using plan recommendations: evidence from a feld experiment. J Market Res; 53(1):46–60

Bott. (2014) Predicting customer churn in telecom industry using multilayer perceptron neural networks: modelling and analysis. Igarss;11(1):1–5

C. Blank and T. Hermansson, (2018)"A Machine Learning approach to churn prediction in a subscription-based service," KTH, Stockholm.

D. Buo and M. Kjellander, (2014)"Predicting Customer Churn at a Swedish CRM-system Company," Linkoping's Universitet, Linkoping.

D. L. Garcia, A. Nebot and A. Vellido, (2017)"Intelligent data analysis approaches to churn as a business problem: a survey," Knowledge and Information Systems, vol. 51, no. 3, pp. 1-56.

F. Khodakarami and Y. Chan, (2014) "Exploring the role of customer relationship management (CRM) systems in customer knowledge creation," Information & Management, vol. 51, pp. 27-42.

Gerpott TJ, Rams W, Schindler A. (2001) Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. Telecommunication Policy. 25:249–69

H. Gebert, M. Geib, L. Kolbe and W. Brenner, (2003) "Knowledge-enabled customer relationship management: integrating customer relationship management and knowledge management concepts," Journal of Knowledge Management, vol. 7, no. 5, pp. 107-123.

Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer-assisted customer churn management: State-of-the-art and future trends. Computers and Operations Research 34(10):2902–2917

I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, (2019) "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector," IEEE Access, pp. 60134-60149.

K. Mishra and R. Rani, (2017)"An inclusive survey on machine learning for CRM: a paradigm shift," in 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India.

M. Sergue, (2020)"Customer Churn Analysis and Prediction using Machine Learning for a B2B SaaS company," KTH, Stockholm.

N. Gordini and V. Veglio, (2017) "Customer's churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry," Industrial Marketing Management, vol. 62, pp. 100-107.

Qureshii SA, Rehman AS, Qamar AM, Kamal A, Rehman A. (2013) Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management. p.131–6

R. Shkurti and A. Muca (2014) "An Analysis of Cloud Computing and Its Role in Accounting Industry in Albania," Romanian-American University, vol. 8, no. 2, pp. 219-229.

S. Marsland, (2014) Machine Learning: An Algorithmic Perspective, 2nd edition ed., Chapman and Hall/CRC.

Tsai C-F, Chen M-Y (2010) Variable selection by association rules for customer churn prediction of multimedia on demand. Expert Syst Appl 37:2006–2015.

Verbeke W, Dejaeger K, Martens D, Hur J, and Baesens B (2011) New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. European Journal of Operational Research 218:211–219

Wei CP, Chiu LT (2002) Turning telecommunications call details to churn prediction: a data mining approach. Expert Systems with Applications 23(2):103–12