PREDICTING TELECOM CUSTOMER CHURN: AN EXAMPLE OF IMPLEMENTING IN PYTHON

Ahmad M. A. Zamil, Prince Sattam bin Abdulaziz University
Nawras M. Nusairat, Applied Science Private University
T. G. Vasista, International School of Technology and Sciences Women's Engineering College
Marwan M. Shammot, College of Applied Studies & Community Service
Ahmad Yousef Areiqat, Business School, Al-Ahliyya Amman University

ABSTRACT

The telecom industry is facing fierce competition. Customer retention is becoming a real challenge. Telecom companies do not want their customers to leave them and look for other service providers. Thus addressing customer churn is becoming a problem. This paper determines the customer churn percentage for a given case of transaction data set as a secondary data. The objective of this research is to identify the probability of customer churn using predictive analytics technique using logistic regression model in order to assess the tendency of probability of customer churn. The result of model accuracy got is 0.8. Based on the existing telecom case study data, customer churn percentage is determined as shown in the graph in the main body of the paper and weighting factors of model function are computed using Python programming language and its libraries.

Keywords: Customer Churn, Logistic Regression, Machine Learning Applications, Predictive analytics, Python Solution for Customer Churn

INTRODUCTION

Customer satisfaction and customer loyalty are becoming important factors in modern retailing (Juhl, Kristensen & Ostergaard, 2002). The telecom industry is flooded with many service providers competing to have high customer base. But telecom companies are facing fierce competition now-a-days due to more telecom service providers as players coming in the market. Customer churn is becoming an ever-growing problem in the retailing sector including retail banking and Tele-communications. The business model for customer churn problem depends on the long-term relationships with customers in order to achieve profitability. Thus managing customer churn is becoming a top priority in service industries such as retail banking and telecommunications (Sweeney & Swait, 2008). Customer retention is one of the best strategies to increase the customers'/company's market share (Verhoef, 2003; Lipowski, 2018). Few statistical studies shown the impact of customer churn, for example the study conducted by studies shown the impact of customer churn, for example the study conducted by Kisioglu & Topcu (2011) showed that a 1% increase in customer retention strategies may decrease the churn rate by up to 5%. Further, customer retention is proved to be 5-6 times cheaper than acquiring new customers. For example, improving customer retention contributed in reducing churn rate from 20% to 10% annually saved about 25M GBP to the mobile operator orange as mentioned in Hassouna, Tarhini, Elyas & Trab (2015) in passing. Telecom companies do not want their customers to leave for other available service provider (Eria & Marikannan, 2019). Some of the factors that influence customer attrition from the existing service provider could be customer satisfaction, superior technology, cost of change and advertising etc. As per Hejazinia & Kazemi, (2014) in passing. However there could be other factors that influence churn rate. So customer

churn prediction is important in the telecom sector as telecom companies have to retain their valuable customers and wanted to enhance customer loyalty as a part of customer relationship management (Ullah, Raza, Malik, Imran, Islam & Kim, 2019).

Customer churn refers to "a customer leaving a service provider" (Wei & Chui, 2002). Customer churn is defined as the affinity of customer to finish the contact with a company. It is otherwise also called customer attrition (Hejazinia & Kazemi, 2014; Yang & Chiu, 2006). In simple terms, Churn is when a customer stops doing business or ends a relationship with a company. Churn rate is the rate at which customers are leaving a company during some period. If churn rate is higher than certain threshold value, it will have substantial and subtle effects (Malik, 2020) on the profitability of the company's business. In telecom service market, customer churn between different service providers is very common in search of best services features and rates (Ahmed & el-Sheikh, 2019). Mathematically churn rate can be computed as follows (Labhsetwar, 2020):

$$C\ (T) = TNCA\ (T)\ /TNCB\ (T) \times 100$$

Where,

C (T) represents churn percentage for a time frame T,

TNCA (T) represents the total number of customers after the time T,

TNCB (T) represents the total number of the customers before the time T.

Types of Customer Churn

Hejazinia & Kazemi (2014) mentioned different types of customer churns as (i) Contractual Churn: When a customer is under a contract for a service and decides to cancel the service e.g. Cable TV, Software-as-a-Service (ii) Non-Contractual Churn: When a customer is not under a contract for a service and decides to cancel the service. E.g. Customer loyalty in retail stores (iii) Voluntary Churn: When a user voluntarily cancels a service e.g. Cellular connection. (iv) Involuntary Churn: When a churn occurs without any request of the customer e.g. Credit card expiration.

For example, Customer data contains all data related to customer's services and contract information (Ahmad, Jafar & Alijoumaa, 2019).

Customer churn is one of the major concerns of large companies. Hence it is important to predict potential customer churn. When churn prediction model is developed, it can assist telecom companies to predict customers who are more likely to churn. Therefore certain important and necessary actions are required to be taken to reduce the customer churn (Ahmad, Jafar & Aljoumaa, 2019). Information systems have become an integral part of every business process in an organization (Srivastava & Bagga, 2014). Strategic impact of business intelligence can be well understood (Tripathi, Bagga & Aggarwal, 2020). Business intelligence systems can assist in churn management in-line with the business strategy of the firm in terms of enhancing the strategic value of organization's value chain. Telecom industry is more prone to high churn rate. Therefore telecom customer satisfaction has a positive influence in retaining more customers (Singh & Samalia, 2014).

LITERATURE REVIEW

Though many approaches are followed to address the problem of predicting the customer churn in telecom companies (Adebiyi, Oyatoye & Amole, 2016), machine learning (Vafeiadis, Diamantaras, Sarigiannidis & Chatzisavvas, 2015) and data mining approach (Kolajo & Adeyemo, 2012; Rodpysh, 2013; Babu & Ananthanarayanan, 2014; Madan, Dave & Nijhawan,

2015) has been attracted as one of recently adopted approaches that focuses on comparing several strategies to predict churn. A comparative study of customer churn prediction in telecom industry using collection based classifiers is conducted by Mishra & Reddy (2017). Chaudhari & Potey (2018) research conducted on predicting the customer churn in telecom industry using Hybrid data mining techniques.

Ahmad, Jafar & Aljoumaa (2019); Jain, Khunteta & Srivastava (2020) research citations provides related work on this kind of problem in passing. Further research by Brandusoiu, Toderean & Beleiu (2016) worked on the prediction of customer churn in the pre-paid mobile telecommunications industry using a call records dataset that consists of 3333 customers with 21 attributes each (Tanneedi, 2016). Ullah, Raza, Malik, Imran, Islam & Kim (2019) worked on features influencing customer churn and their ranking in passing. Further the qualitative research by Sharma & Panigrahi (2011) has also provided citations on the telecom churn. Hashmi, Butt & Iqbal (2013) has provided many citations on the prediction of telecom churn using data mining techniques. Python based solutions of Prediction of telecom churn is worked previously by the following: (i) Pamina & Raja et al., (2019), (ii) Labhsetwar (2020) etc. Majority of those who had done is from the computer science perspective by comparing the accuracy of algorithm. From the business perspective very limited research has been done like Bhattarai, Shrestha & Sapkota (2019). Therefore this research has been focused to work on the evidently showing the effort of providing solution of telecom churn prediction using Python more from the retail management perspective.

RESEARCH OBJECTIVE

This paper determines the customer churn percentage for a given case of transaction dataset. The objective is to measure the impact of churn factors on customer loyalty and to discover ways to manage customer churn. It uses predictive analytics technique to develop churn prediction models that predict customer churn by assessing the tendency of probability to churn.

RELEVANT THEORIES

Broadly some potential underlying theories related to the work are: (a) Roger's Theory of innovation and (b) (Internet based) Davis's Technology Acceptance Model. As per the research study by Wahid (2007), the problem of innovation diffusion can be approached by many ways (Taylor &Todd, 1995) such as (i) Innovation diffusion theory, (ii) concerns-based adoption model, (iii) technology acceptance model and (iv) chocolate model (Ham, 2018).

However the following four theories are found more related to the current context of the research.

Innovation Diffusion Theory–Churn

The research conducted by Libai, Muller & Peres (2009) combined the diffusion modeling with a customer relationship approach to investigate the influence of attrition on growth in service markets. The research of Lengyel, Clemente, Kertesz & Gonzalez (2018) cited that the Bass model has been used extensively to describe the diffusion process and forecasting the market size of new products. Complex contagion models provide the relationship between the diffusion of on-line behavior and on-line innovations. Further they cited that their research

studies shown that contagious mechanism of churn is similar to the diffusion. Thus they described the processes of spatial diffusion and churn by exploring the relationship of number of innovation adopters and churners. Brand switching can also be termed as churn and is becoming a major concern in many innovative industries (Peres, Mahajan & Muller, 2009).

Technology Acceptance Model–Churn Prediction

Research conducted by Galib, Ait Hammou & Steiger (2018) explored the prediction of consumer behavior as a part of extension of technology acceptance model. Johar & Awalluddin (2011) mentioned that Technology acceptance model in the form E-Commerce allows loyal customers to provide positive word-of-mouth and helps in customer retention. Technology Acceptance Model along with Theory of Reasoned Action, Theory of Planned Behavior model will better explain the acceptance of information technology and information system (Silva & Dias, 2007) for and the TAM factors will positively correlate with the usefulness of Decision Support Systems in Business domain (Dulcic, Pavlic & Silic, 2012). ML techniques adoption has its influence on software engineering applications (Rana, Staron, Hansson, Nilsson & Meding, 2014). However it is also important to assess the risk of using such technologies based on perceived risk theory (Hubert, Brock, Zhang, Koch & Riedi, 2019).

Commitment-Trust Theory and Customer Retention

Specific to customer retention, Morgan & Hunt's (1994) commitment-trust theory of relationship marketing would be very relevant as mentioned by Banerji (2016).

Rough Set Theory–Churn Prediction

Rough set theory is introduced by Pawlak (1982) and further Pawlak, Grzymala-Busse, Slowinski & Ziarko (1995). It is a mathematical modeling tool to deal with information uncertainty. It is an effective tool for financial analytics. Rough set theory in combination of neural networks help in predicting bank holding patterns (Ahn, Cho & Kim, 2000). Specific to customer churn, rough set theory of a mathematical model is used. It deals with class data classification problems to support decision support systems. The basic assumption of rough set theory is that every object of discourse can be associated with some information or data set (Qiasi, Roozbehani & Minaei-Bidgoli, 2013).

Rough set theory can be employed instead of regression analysis because it is better suited for categorical or discrete classification of output variables using logical if then rules. However multiple data analysis theories exist that deal with uncertain information like fuzzy set theory and evidence theory in addition Rough set theory (Radhakrishnapany, Wong, Tan, Chong, Tan, Aviso, Janiro & Chemmangattuvalappil, 2020).

RESEARCH METHODOLOGY

There are many ways of working for churn analysis. In machine learning aspect of data mining they are primarily categorized as (i) supervised learning methods (Khodabandehlou & Rahman, 2017) (ii) unsupervised learning methods (Patro, Kumar & Kumar, 2019). Azeem & Usman (2018) used supervised, semi-supervised and unsupervised learning algorithms to get

most important features in predicting churn and customer retention in the context prepaid customers in telecom industry.

In this research, initially, case study method approach is used to test the hypothesis as it helps in building the theory corresponds to the pragmatic considerations made in the case study such as understanding the factors influencing the customer churn (Iacono, Brown & Holtham, 2011). The transaction data set provides such factors and helps in understanding how many of such factors positively influence and also negatively influence the customer churn. Classification method of data mining and the use of logistic regression technique determine the weighting factors and thus rank the factors that influence customer churn. In this research, supervised learning method is used basically to classify the data based on what type of data it learns from given training data set. Data classification process has two steps. In the first step, model learning is known as training data set (Sarkar, Kayes & Watters, 2019). It is a pair of data consisting of an input and an output value. When the transaction data is becoming the input, probability becomes the output depicting whether the customer has churned or not (Pradeep, Sushmita, Swati & Akshay, 2017). In the second set of classification, model is used to test or predict the class labels for a given data set (Sarkar, Kayes & Watters, 2019).

Recent programming language called Python has the ability to provide these computations and statistical output through accessing the existing machine learning algorithms from the sci-kit libraries of Python (Pedregosa et al., 2012).

The limitation of this research is that Python programming is used to compute the churn rate and corresponding visualization is provided. However the steps of processing further to determine the weighting factors and thus ranking of factors that influence customer churn has not been pragmatically discussed.

Hypothesis Declaration

Null Hypothesis

H_0   *The telecom churnuse-case model that is built based on machine learning technique implemented in Python programming and its librarieswill perform better in terms of offering better decision support and decision making with visualization*

SOLUTION APPROACH

In this paper Python programming implementation is adopted to determine the percentage of churn for the given data set. Correspondingly the following attributes are becoming the part of telecom customer churn data set: Gender, Partner, Dependents, Phone Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device protection, Tech Support, Streaming TV, Streaming Movies, Contract, Paperless Billing, Payment method, Churn etc. Not all columns affect the customer churn (Elance, 2020).

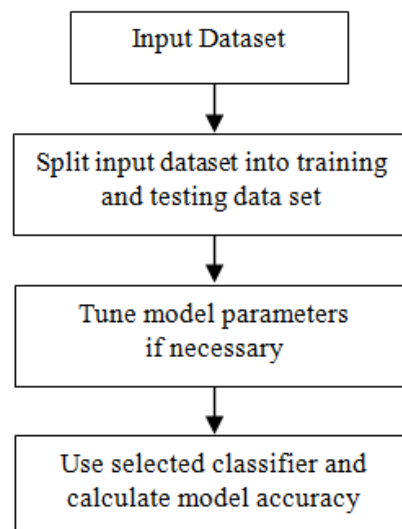## Role of Machine Learning Algorithm in Predicting Churn

Machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosted Machine Tree "GBM" and Extreme Gradient Boosting "XGBOOST" (Ahmad, Jafar & Alijoumaa, 2019) are used to predicting customer churnand also (Vafeiadis et al, 2015; Ullah, Raza, Malik, Imran, Islam & Kim, 2019). Machine learning algorithms such as decision tree algorithms process categorical data work better with numerical data. It uses nonparametric

method for this purpose and does not need prior assumption (Hassouna, Elyas & Trab, 2016). Therefore it is required to transform all the non-numeric and categorical data to numeric before applying the machine algorithm. One way of converting the categorical columns to numerical columns is to use one hot encoding technique so that in each column 1 is used to designate the categorical value for the current row and 0 otherwise.

For example, random forest algorithm is one of the powerful algorithms for classification problems that can be used. To train this algorithm, the fit method can be used to make predictions on the test set (Ahmad, Jafar & Aljoumaa, 2019).

Application of Classification Method in Data Mining for Telecom Churn Prediction

Various classification techniques of data mining include: decision tree, neural networks, logistic regression, cluster analysis, genetic algorithm, Markov model, Naïve Bayes, k-nearest-neighbor, Bayesian belief network, association rule, support vector machine, bagging, CART, CHAID, K-Means, Fuzzy C Means, Influence diffusion model, Partial Least Squares and structural model etc. (Hashmi, Butt & Iqbal, 2013).

FIGURE 1
PROCESS OF CLASSIFICATION DATA VALIDATION

Literature review of churn prediction models usually focus on predictive accuracy, comprehensibility and efficiency of customer retention with justifiability for marketing campaigns (Verbeke, Martens, Mues & Baesens (2011).

In Python, 'sci-kit-learn' package is used to import train_test_split libraries that help to split the data set into training data set and testing data set (bitdegree.org, 2019).

LOGISTIC REGRESSION

Previously analysis of customers churns prediction in telecom industry using decision trees and logistic regression is done by Dalvi, Khandge, Deomore, Banker & Kande, 2016; Mandak & Hanclova, 2019; Jain, Khunteta & Srivastava, 2020). Logistic regression is one of the most important statistical methods used in data mining for conduction data analysis and for

binary classification response from data sets (Maalouf, 2009). Logistic regression is a generalized class concept of linear regression. It is used for estimating not only binary class dependent variables but also for multi-class dependent variables. It is a statistical model that uses sigmoid function to model binary dependent variables (Horvat, Havas & Srpak, 2020). Logistic regression is a classical model in probability and statistics used for estimating conditional probabilities (Foster, Kale, Luo, Mohri & Sridharan, 2018). Logistic regression technique of a supervised learning classification algorithm is used to predict the probability of a target variable (Sarkar, Kayes & Watters, 2019). The target or dependent variable is dichotomous or binary in nature. The data coding follows as either 1, which denotes success/yes or 0, which denotes for failure/no. Mathematical representation of a logistic regression model for prediction is P(Y=1) as a function of X. Logistic regression model can have following types: (i) Binary model – in which dependent variable will have only two possible types *i.e.* either 1 or 0. (ii) Multinomial model – in which dependent variable can have three or more possible unordered values having no quantitative significance for e.g. A, B or C types. (iii) Ordinal – in which dependent variable can have three or more possible but ordered types having quantitative significance for example poor, good, very good and excellent can have the scores 0, 1, 2, 3 (Elance, 2020).

The process of logistic regression in computing customer churn has the following steps: (i) Read the transaction data set, (ii) Computes the Churn percentage, (iii) does data pre-processing, (iv) Based on applying the classification algorithm, it splits the data set into training and testing data set, and find the accuracy score (v) It computes the weight of variables that affects the churn value. Thus it helps in handling only those variables in preventing the customer churn. For the given set of transaction data set called 'Telco-Customer-Churn.csv', the output of the accuracy of the model is 80.84% and weights of the variables are shown Table 1 in Appendix-I.

Once the accuracy level is acceptable, accuracy and confusion matrix are used as parameters to judge how accurately this model is behaving. If the value of accuracy is high, it suggests that the model is a better fit. Similarly confusion matrix shows a matrix of true positive and true negative values as well as false positive and false negative values. If there is higher percentage of true values as compared to false values it indicates that the model is a better model. In the next step, how each of the fields or variables affects the churn value is judged. These specific variables will have greater impact on the churn. Therefore those variables that impact greater will be handled to prevent the customer churn. Weight of each variable is computed by setting the coefficients in the classifier to zero (Elance, 2020). These variables give the direction of the impact of explanatory variables that can be assessed by the plus or minus sign before the estimated regression coefficients. Table 1 of Appendix I shows the division of variables into positive and negative impact on the probability of churn (Mandak, 2018).

The predict method simply substitutes these values of the weights into the logistic model equation and returns the result. This returned value is the required probability.

CONCLUSION

Presently telecom market is facing severe competition. Customer churn prediction has become an important issue of customer relationship management to retain valuable customers. Therefore by performing research, key factors of churn to retain customers and their influence on churn will be well understood. Thus CRM problems can be solved by the decision makers of the company (Ullah, Raza, Malik, Imran, Islam & Kim, 2019). The research conducted by Sweeney

& Swait (2008) indicated that brand management and word of mouth can influence by reducing the customer churn. According to Anderson, Jolly & Fairhurst (2007), increasing customer acquisition and customer retention through customer loyalty programs, using data mining tools for making data analysis in order to elicit insights to better understand customers are important to implement a more customer-centric business strategy.

In this study, a customer churn model is provided for the existing secondary data based case study from Kaggle.com and Github web sites to work on the metrics evaluation. The results show that the use-case model built based on machine learning technique is performing better.

In this case study, logistic regression model is implemented using Python libraries with Python programming.

Proper churn management can save a huge amount of money to company. Thus the economic value of customer retention can be summarized as: (i) satisfied customers can bring new customers (ii) long-term customers are usually do not get influenced much with competitors (iii) long-term customers tend to buy more (iv) company can focus on satisfying existing customer's needs (v) lost customers share negative experience and thus will have negative influence on the image of the company (Van Den Poel & Lariviere, 2004). Thus customer retention as a function of *i.e.* f{Price, service quality, customer satisfaction, brand image} could lead to better customer loyalty (Martey, 2014).

FUTURE RESEARCH

As a future research one can experiment by comparing different algorithms and their model accuracy as well as to experiment with the use of Logit Boost technique to find and compare the accuracy.

REFERENCES

Adebiyi, S.O., Oyatoye, E.O. & Amole, B.B. (2016). Improved customer churn and retention decision management using operations research approach.

Ahmad, A.K., Jafar, A. & Aljoumaa, K. (2019). 'Customer churn prediction in telecom using machine learning in big data platform'. *Journal of Big data, 6*(28).

Ahn, B.S., Cho, S.S. & Kim, C.Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with applications, 18*, 65-74.

Ahmed, H., & El-Sheikh, K. (2019). The impact of Customer Churn Factors (CCF) on customer's Loyalty: The case of telecommunication service providers in Egypt. *International Journal of Customer Relationship Marketing and Management,10*(1).

Anderson, J.L., Jolly, L.D. & Fairhurst, A.E. (2007). Customer relationship management in retailing: A content analysis of retail trade journals, *Journal of Retailing and Consumer services*, *14* (6), 394-399.

Andrews, R, Zacharias, R., Antony, S. & James, M.M. (2019). Churn prediction in telecom sector using machine learning. *International Journal of Information Systems and computer sciences*, *8*(2), 132-134.

Azaeem, M. & Usman, M. (2018). A fuzzy based churn prediction and retention model for prepaid customers in telecom industry. *International Journal of computational intelligence systems, 11*(1), 66-78.

Banerji, D. (2016). *Which is the most suitable theory for customer retention in service industry?*

Bhattarai, A., Shrestha, E. & Sapkota, R. (2019). Customer churn prediction for imbalanced class distribution of data in business sector. *Journal of Advanced college of engineering and management*, *5*, 101-110.

Bitdegree.org (2019). *Splitting datasets with the sklearn train_test split function*.

Brandusoiu, I., Toderean, G., & Beleiu, H.G. (2016). Methods for churn prediction in the pre-paid mobile telecommunications industry. *In proceedings of International Conference on communications*.

Choudhari, A.S., & Potey, M. (2018). Predictive to prescriptive analysis for customer churn in telecom industry using Hybrid data mining techniques. *In proceedings of fourth international conference on computing communication control and automation (ICCUBEA)*, Pune, India, 1-6.

Dalvi, P., Khandge, S.K., Deomore, A., Banker, A., & Kanade, V.A. (2016). Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. *In Proceedings of IEEE Symposium on Colossal Data Analysis and Networking*, March 18-19, Indore, Madhya Pradesh.

Dulicic, Z., Pavlic, D., & Silic, I. (2012). Evaluating the intended use of Decision Support System (DSS) by applying Technology Acceptance Model (TAM) in business organizations in Crotia. *In Proceedings of 8th International Strategic Management Conference. Procedia. Social and Behavioral Science, 58,* 1565-1575.

Elance, P. (2020). *Predicting Customer Churn in Python*.

Eria, K., & Marikannan, B.P. (2019). 'Significance-based feature extraction for customer churns prediction data in the telecom sector'. *Journal of computational and theoretical nanoscience, 16*, 1-4.

Foster, D.J., Kale, S., Luo, H., Mohri, M., & Sridharan, K. (2018). Logistic regression: The importance of being improper. *In Proceedings of Machine Learning Research at 31st Annual conference on Learning Theory*, 75, 1-42.

Galib, M.H., Ait H.K., & Steiger, J. (2018). Predicting consumer behavior: An extension of technology acceptance model. *International Journal of Marketing Studies, 10*(3), 73.

Ham, M. (2018). Chapter 1. Theories of innovation adoption and real-world, Correia, A. (ed). *Driving educational change: Innovations in action*. The Ohio State University: Press Books.

Hashmi, N., Butt, N.A., & Iqbal, M. (2013). Customer churn prediction in telecommunications: A decade review and classification. *International Journal of Computer Science Issues, 10*(1).

Hassouna, M., Tarhini, A., Elyas, T., & Trab, M.S.A. (2015). Customer churn in mobile markets: A comparison techniques. *International Business Research, 8*(6), 224-237.

Hejazinia, R., & Kazemi, M. (2014). Prioritizing factors influencing customer churn. *Interdisciplinary Journal of Contemporary Research in Business, 5(12)*:227-236.

Hubert, M., Blut, M., Brock C., Zhang, R.W., Koch, V., & Riedl, R. (2019). The influence of acceptance and adoption drivers on smart home usage. *European Journal of Marketing, 53* (6).

Horvat, T., Havas, L., & Srpak, D. (2020). The impact of selecting a validation method in machine learning on predicting basketball game outcomes. *Symmetry, 12*(431), 1-15.

Iacono, J.C., Brown, A., & Holtham, C. (2011). The use of case study method in theory testing: The example of emarket places. *Journal of Business Research Methods*, 9(1), 57-65.

Jain, H., Khunteta, A., & Srivastava, S. (2020). Churn Prediction in Telecommunication using Logistic Regression and Logit Boost. *In Proceedings of International Conference on computational intelligence and Data Science(ICCIDS2019). Procedia computer Science,* 167, 101-112.

Johar, M.G.M., Awalluddin, J.A.A. (2011). The role of technology acceptance model in explaining effect on e-commerce application system. *International journal of managing information technology*, 3(3), 1-14.

Juhl, H.J., Kristensen, K., & Ostergaard, P. (2002). 'Customer satisfaction in European food retailing'. *Journal of Retailing and consumer services*, 9, 327-334.

Khodabandehlou, S. & Rahman, M.Z. (2017). Techniques for customer churn prediction based on analysis of customer behavior. *Journal of Systems and Information Technology, 19*(1/2), 65-93.

Kolajo, T., & Adeyemo, A.B. (2012). Data mining technique for predicting telecommunications industry customer churn using both descriptive and predictive algorithms. *Computing Information Systems & Development Informatics Journal, 3*(2), 27-34.

Kisioglu, P., & Topcu, I.Y. (2011). Applying Bayesian belief network approach to customer churn analysis: A case study on the Telecom industry of Turkey. *Expert Systems with Applications, 38*(6), 7151-7157.

Labhsetwar, S.R. (2020). Predictive analysis of customer churns in Telecom industry using supervised learning.

Lengyel, B., Di Clemente, R., Kertesz, J., & Gonzalez, M.C. (2018). *Spatial diffusion and churn of social media*.

Libai, B., Muller, E., & Peres, R. (2009). The diffusion of services. *Journal of marketing research, 46*(2), 163-175.

Lopwski, M. (2018). 'Customer Churn as a purchasing journey stage'. *Proceedings of 32nd IBIMA Conference. November 15-16, Seville, Spain*.

Madan, M., Dave, M., & Nijhawan, V.K. (2015). A Review on: Data mining for telecom customer churn management. *International Journal of Advanced Research in Computer Science and Software engineering, 5*(9).

Maalouf, M. (2009). 'Logistic regression in data analysis: An Overview'. *International Journal of Data Analysis Techniques and Strategies,* 1-20.

Malik, U. (2020). *Customer Churn Prediction with Python*.

Mandak, I.J. (2018). Proposal and implementation of churn prediction system for telecommunication company. A PhD dissertation submitted VSB-Technical University of Ostrava, Czech Republic, EU.

Mandak, J., & Hanclova, J. (2019). Use of logistic regression for understanding and prediction of customer churn in telecommunications. *Statistika, 99*(2), 129-149.

Martey, E.M. (2014). The relationship between customer retention and customer loyalty in the restaurant industry in Ghana, 1(8), 51-66.

Mishra, A., & Reddy, U.S. (2017). A comparative study of customer churn prediction in telecom industry using ensemble based classifiers. *In Proceedings of the International conference on Inventive computing and informatics*, 721-725.

Morgan, R.M. & Hunt, S.D. (1994). 'Theory of relationship marketing'. *Journal of Marketing, 58*(3), 20-38.

Pamina, J., & Raja J.B. (2019). An effective classifier for predicting churn in telecommunication. *Journal of Advanced Research in Dynamical & Control Systems*, *11*(S1), 221-229.

Patro, P., Kumar, K., & Kumar, G.S. (2019). Various classifiers performance based machine learning methods. *International Journal of Recent Technology and Engineering*, *8*(S3), 305-310.

Pedregosa, et al. (2012). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Peres, R., Mahajan, V., & Muller, E. (2009). *Innovation diffusion and new product growth models: A critical review and research directions*.

Pawlak, Z. (1982). Roughsets. *International Journal of Information and Computer sciences*, *11*, 341-356.

Pawlak, Z., Grzymala-Busse, J., Slowinski, R., & Ziarko, W. (1995). Roughsets. *Communications of ACM, 38(11), 89-95.*

Pradeep, B., Sushmita, V.R., Swati, M.P., & Akshay, H. (2017). Analysis of customer churn prediction in logistic industry using machine learning. *International Journal of Scientific and Research Publications, 7*(11), 401-403.

Qiasi, R., Roozbehani, Z., & Minaei-Bidgoli, B. (2013). Predict customer churn by using rough set theory and neural network. *Proceedings of 9th International Industrial Engineering Conference*, Jan 20-21, Tehran, Iran.

Radhakrishnapany, K.T., Wong, C.Y., Tan, F.K., Chong, J.W., Tan, R.R., Aviso, K., & Janairo J.I.B. (2020). Design of fragrant molecules through the incorporation of rough sets into computer-aided molecular design. *Molecular Systems Design & Engineering*, 1-43.

Rana, R., Staron, M., Hansson, J., Nilsson, M., & Meding, W. (2014). A framework for adoption of machine learning in industry for software defect prediction. *In Proceedings of 9th International Conference on Software Engineering and Applications*.

Rodpysh, K.V. (2013). Applying data mining to customer churn analysis: A case study on the insurance industry of Iran. *Open Journal of Artificial Intelligence, 1*(1), 8-12.

Sarkar, I.H., Kayes, A.S.M. & Watters, P. (2019). Effectiveness analysis of machine learning classification model for predicting personalized context-aware smartphone usage. *Journal of Big data, 6*(57), 1-28.

Sharma, A., & Panigrahi, P.K. (2011). A Neural network based approach for predicting customer churn in cellular network services. *International Journal of Computer Applications, 27*(11), 56-89.

Singh, H., & Samalia, H.V. (2013). A business intelligence perspective for churn management. *Procedia –Social and Behavioural Science*, *109*, 51-56.

Silva, P.M., & Dias, G.A. (2007). Theories about technology acceptance: Why the users accept or reject the information technology? *Brazilian Journal of Information Science, 1*(2), 69-86

Srivastava, S., & Bagga, T. (2014). A comparative study on the usage of HRIS in the IT/ITES, Services and manufacturing sectors in the Indian scenario. *Prabandhan: Indian Journal of Management, 7*(6).

Sweeney, J., & Swait, J. (2008). The effects of brand credibility on customer loyalty. *Journal of Retailing and consumer services*, *15*, 179-193.

Tanneedi, N.N.P.P. (2016). *Customer churn prediction using big data analytics*. Thesis no. MSEE-2016:37, submitted to Faculty of computing, Blekinge Institute of Technology, Sweden.

Taylor, S., & Todd, P. (1995). Understanding information technology usage: A test of competing model. *Information Systems Research*, *6*(2,), 144-176.

Tripathi, A., Bagga, T., & Aggarwal, R. (2020). Strategic impact of business intelligence: A review of literature. *Prabandhan: Indian Journal of Management*, *13*(3), 35.

Tutorialpoint.com (2020). *Machine Learning – Logistic Regression*.

Ullah, I., Raza, B., Malik, A.K., Imran, M., Islam, S., & Kim, S.W. (2019). *A Churn Prediction model using Random Forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector*.

Silva, P.M., & Dias, G.A. (2007). Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*, 67, 30-45.

Vafeiadis, T., Diamantras, K.I., Sargiannidism, G., & Chatzisavvas, K.Ch. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory, 55*, 1-9.

Verbeke, W., Martens, D., Mues, C. Baesebs, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert systems with Applications, 38*, 2354-2364.

Wahid, F. (2007). Using the technology adoption model to analyze internet adoption and use among men and women in Indonesia. *The Electronic Journal on Information systems in Developing Countries, 32*(6), 1-8.

Wei, C.P., & Chiu, I.T. (2002). Turning telecommunications to churn prediction: A data mining approach. *Expert Systems with Applications, 23*, 103-112.

Yang, L.S., & Chiu, C. (2006). *Knowledge Proceedings of the 10th WSEAS international conference on applied mathematics*, Dallas, Texas.