Future Computing and Informatics Journal

Volume 6 | Issue 1 (2021) Article 1

2021

Review of Data Mining Techniques for Detecting Churners in the Telecommunication Industry

Mahmoud Ewieda
October 6 University, ewi111@hotmail.com

Mohamed Ismail Roushdy
Professor & Dean of Faculty of Computers & Information Technology, Future University in Egypt, Mohamed.Roushdy@fue.edu.eg

Essam Shaaban
Canadian International College School of business, Business technology department, Cairo, Egypt, essam.shaaban@cic-cairo.com

Recommended Citation

Ewieda, Mahmoud; Roushdy, Mohamed Ismail; and Shaaban, Essam (2021) "Review of Data Mining Techniques for Detecting Churners in the Telecommunication Industry," *Future Computing and Informatics Journal*: Vol. 6: Iss. 1, Article 1.
DOI: http://doi.org/10.54623/fue.fcij.6.1.1
Available at: https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/1

Review of Data Mining Techniques for Detecting Churners in the Telecommunication Industry

Cover Page Footnote

Dear editor, We thank you for considering our manuscript for publication in Future Computing and Informatics Journal. our manuscript "Review of Data Mining Techniques for Detecting Churners in the Telecommunication Industry " Sincerely, Best regards, Mahmoud Ewieda

Future Computing and Informatics Journal

Volume 6 | Issue 1 (2021) Article 1

2021

Review of Data Mining Techniques for Detecting Churners in the Telecommunication Industry

Mahmoud Ewieda
October 6 University, ewi111@hotmail.com

Mohamed Ismail Roushdy
Professor & Dean of Faculty of Computers & Information Technology, Future University in Egypt, Mohamed.Roushdy@fue.edu.eg

Essam Shaaban
Canadian International College School of business, Business technology department, Cairo, Egypt, essam.shaaban@cic-cairo.com

Recommended Citation

Ewieda, Mahmoud; Roushdy, Mohamed Ismail; and Shaaban, Essam (2021) "Review of Data Mining Techniques for Detecting Churners in the Telecommunication Industry," *Future Computing and Informatics Journal*: Vol. 6 : Iss. 1 , Article 1.
DOI: http://doi.org/10.54623/fue.fcij.6.1.1
Available at: https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/1

Review of Data Mining Techniques for Detecting Churners in the Telecommunication Industry

Cover Page Footnote

Dear editor, We thank you for considering our manuscript for publication in Future Computing and Informatics Journal. our manuscript "Review of Data Mining Techniques for Detecting Churners in the Telecommunication Industry " Sincerely, Best regards, Mahmoud Ewieda

https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/1
DOI: http://doi.org/10.54623/fue.fcij.6.1.1

Future Computing and Informatics Journal

Homepage: https://digitalcommons.aaru.edu.jo/fcij/

doi: http://Doi.org/10.54623/fue.fcij.6.1.1

Review of Data Mining Techniques for Detecting Churners in the Telecommunication Industry

Mahmoud Ewieda[1,a], Essam M Shaaban[2,b], Mohamed Roushdy[3,c]
[1] Business Information Systems, Helwan University, Helwan, Egypt
[2]Canadian International College School of business, Business technology department, Cairo, Egypt
[3]Faculty of Computers & Information Technology, Future University in Egypt, New Cairo, Egypt
[a] ewi111@hotmail.com, [b] essam.shaaban@cic-cairo.comb,[c] Mohamed.Roushdy@fue.edu.eg

ABSTRACT

The telecommunication sector has been developed rapidly and with large amounts of data obtained as a result of increasing in the number of subscribers, modern techniques, data-based applications, and services. As well as better awareness of customer requirements and excellent quality that meets their satisfaction. This satisfaction raises rivalry between firms to maintain the quality of their services and upgrade them. These data can be helpfully extracted for analysis and used for predicting churners. Researchers around the world have conducted important research to understand the uses of Data mining (DM) that can be used to predict customers' churn. This paper supplies a review of nearly 73 recent journalistic articles starting in 2003 to introduce the different DM techniques used in many customer-based churning models. It epitomizes the present literature in the field of communications by highlighting the impact of service quality on customer satisfaction, detecting churners in the telecoms industry, in addition to the sample size used, the churn variables used and the results of various DM technologies. Eventually, the most common techniques for predicting telecommunication churning such as classification, regression analysis, and clustering are included, thus presenting a roadmap for new researchers to build new churn management models.

Keywords: Quality of service, Churn prediction, Classification, Data mining and Prediction model.

1. INTRODUCTION

In light of the competitive climate prevailing these days in which customers receive unprecedented support, organizations are looking for Excellency through this competition and are working to reduce customer dropouts by making improvements to gain their satisfaction [1]. With the appropriate support solution, the problem is diagnosed, and then the service office can automate the repetitive problem fixing processes, and not reinventing solutions every time any of these problems arise. Customer service is a concept that is more comprehensive than just solving problems and should not wait for the customer to search for it himself; this is an imperfect role because the company only takes care of customers who have problems. Firms in multiple industries struggle to attract and retain customers [1]. Due to rapid techniques advancement and increasing rivalry, as offers and prices [2] customers have multiple choices to choose.Also, customer satisfaction is the measure of products and services that meet or exceed customer expectations provided by the company, and it has become a huge challenge for telecommunication operators. Telecommunication

Published by Arab Journals Platform, 2021

firms lose much revenue because of losing their current customers, and this operation is called "churn" [3].Sufian Albadawi et al. (2017) [4] obtain databases about DM and churn.

Churn in the telecommunication-based industry is the measure of customers who leave service during a certain term of time. Churning can be voluntary and involuntary. Voluntary occurs when the current customer departs the service provider [5]. Involuntary churn occurs when the service provider asks the customer to depart for reasons such as non-payment, and others. Voluntary churn can be divided into incidental and intentional deliberate [6, 7]. Incidental churns happen not because customers planned it but something occurred in their life like financial situation, alteration of their place, etc. Deliberate churn happens for technology-related reasons (customers want novel or better technique, price sensitivity, quality of service (QoS) factors, social or psychological factors, and amenities) [3]. Customer retention is the major target of Customer Relationship Management (CRM) that involves customer retention of whole measures the organizations take to guarantee customer loyalty and decrease churn [8, 9].

The importance of churn prediction is the ability to predict a specific customer at risk of disruption because acquiring new customers is more expensive than trying to win back a customer or customer you have known before [10]. The expression "Churn Management" is a procedure used by telecom firms to retain their lucrative subscribers such as proactive in retaining customers [11]. DM techniques in communications are set to work for CRM due to the rapid growth of enormous amounts of data, high speed of competition in the market, and a high rate of churn.DM techniques are the analysis of large amount of data to create rules, examples, models that can be used to guide decision-makers, to predict future behavior, and aim to extract information hidden in large data blocks. Modern technology has imposed itself strongly in the information age and provides all businesses and organizations [10]. DM techniques prediction states can be used in the classification and clustering of these techniques and it may be utilized to predict churners such as Logistic Regression (LR), Naive Bayesian (NB), k-Nearest Neighbor (KNN), Gradient Boosted Trees (GBT), Genetic Algorithms (GA), or Fuzzy Logic (FL). These techniques have a powerful effect on prediction. Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and Neural Networks (NN) [3, 12, 13].

The most difficult problem faced by telecom industry is customer churn. Customer churn models aim to detect customers with a high probability to jump or leave the service provider. A database of customers who might churn allows the company to target those customers and start retention strategies that reduce the percentage of customer churning. In this study, I attempted to address the following: (1) Analysis of service quality with data to detect disruptions in the telecommunications industry (2). Knowing the variables that is affect the quality of service (3). Summarizing the most popular DM technologies used in industry-based forecasting models.

This study helps researchers avoid interfering with efforts and review a large part of references. The next sections provide a survey of the different literature on grasp the assortment of DM strategies utilized in building churn predicting models [11].

This paper targets to review research density through the year 2003 to 2020. The Research methodology is presented in Section 2. The review of the different mining techniques is mentioned in Section 3 and Section 4 gives the discussions. The conclusions and future scope of the paper are described in Section 5.

2. RESEARCH METHODOLOGY

Many researchers have extensive literature across several disciplines in the field of DM [14]. This paper reviews DM techniques for detecting churners in the telecommunication industry starting from 2003 to 2020.

2.1. Articles Source:

Papers are collected from the articles in the period between 2003 and 2020 from the standard sources as Elsevier, IEEE, Explorer Springer, Google Scholar, ACM Digital Library and etc.

2.2. Aspects of inclusion and exclusion:

- Articles submit application reports in the telecommunication industry only.
- Articles should be suggesting ideas or solutions to problems of telecoms customers.
- Articles should be about machine learning and its optimization algorithms.
- Articles must be highly effective.

Predicting the churn of telecommunications is frequently used in research for customer

https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/1
DOI: http://doi.org/10.54623/fue.fcij.6.1.1

satisfaction. Many machine learning models have communications datasets. This article contributes to a survey of the different machine learning techniques used between 2003 and 2020. Figure 1 show several standard articles published between 2000 and 2018. It has been observed that there is continuous development in the creation of flop prediction models through research, especially in the field of telecoms. This paper also reveals been used in various public and private public and private communications disruption data sets and key challenges in the telecommunications sector. Currently, hybrid groups are very popular due to their high predictability and great importance. Table 1 gives a concise overview of previous studies of telecoms industry-based churn prediction problems.

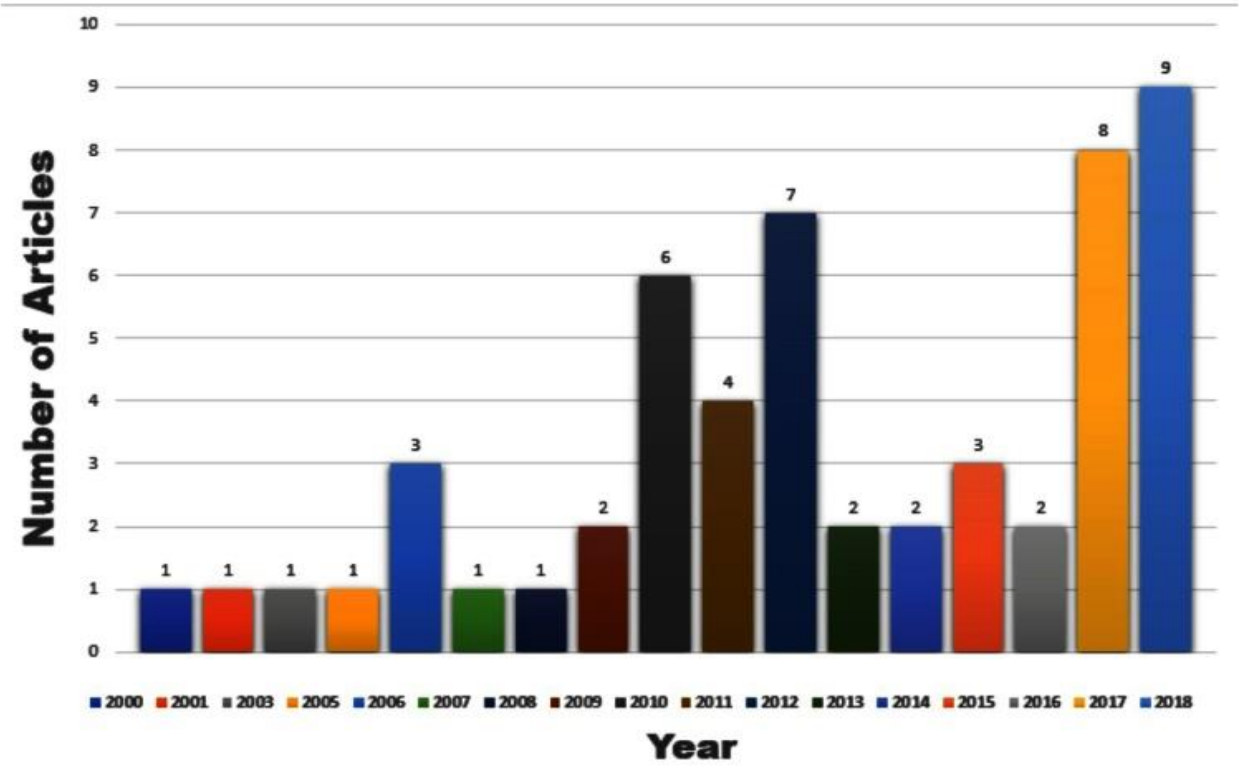

Figure1: The number of articles published annually in the prediction of the incidence of telecommunication disruption (2000-2018) [15].

Published by Arab Journals Platform, 2021

Table 1: Overview of previous research about churn prediction telecom-based industries

References	Dataset	Techniques	Records	Results
E. Shaaban et al. (2012)[6]	The data set was gained from the service provider	DT, NN, and SVM	5000 instances	DT 77.9% accuracy, NN 83.7% accuracy, and SVM 83.7% accuracy
W. Auet al. (2003) [16]	Malaysian subscriber	Data Mining by Evolutionary Learning (DMEL)	100,000 subscriber data and 251 subscriber-related variables.	DMEL effectively detected the rules and accurately predict churn
S.-Y. Hunget al. (2006) [17]	A data set owned by a Taiwanese telecommunication s company	DT(C5.0) and Back Propagation Network (BPN) techniques	160,000 subscribers beside 14,000 churners.	Both the DT and NN technologies can offer accuracy while BPN performs best than DT without a split.
K. Cao and P. Shao,(2008)[18]	Duke University (CRM)	SVM- Recursive Features Elimination (RFE)	65,000 customers and 171 attributes	Present SVM-RFE for attributes selection in the churn prediction that demonstrated the pathological predictive outcome
R. J. Jadhav and U. T. Pawar, (2011)[19]	Aggregated data from the internal customer database, private call log data from the company, and search survey.	NN- Back PropagationNetwor kalgorithm.	895 customers, of which 17.67% Churn & 82.33% non-churn.	A predicted customer at risk that will likely churn
I. Brandusoiu and G. Toderean, (2013) [20]	Communications Data Set, UCI warehouse, University of California, Irvine	Support Vector Machine (SVM)	3333 customers	88.56% accuracy
W. Verbekeet al. (2011) [21]	KDD Library, UCI Repository	C4.5 & repeated incremental pruning to produce an error RIPPER	5000 instances	ALBA appears with C4.5 and Ripper with highest accuracy whilst AntMiner + is a high performance
W. Verbekeet al. (2012) [22]	KDD Cup2009, UCI, and Duke	NN, Li SVM, rbfSVM, linLSSVM,rbfLSSV M, Ripper, Part, C4.5, CART, ADT, RF, LMT, Bag, Boost, RBFN, VP, Logit, KNN10, KNN100, BN, NB	It contains small data 2180 and bigger data 338,874	Excessive sampling does not improve the performance of classifiers as it does in predicting churn in the telecommunication sector, and a large group of classifiers has found similar performance.
K. Kim, C.-H. Jun, and J. Lee,(2014)[23]	Contains customer personal data & "CDR" data.	LR and Multilayer perceptron NN	the dataset included 89,412 instances and 9.7% churns	An effective approach was developed using the SPA method as a prevalence process.
G. Olle, (2014)[24]	Asian Mobile	LR &Voted perceptron	Includes 2000 subscriber & 23 variables with 534 churners.	Hybrid model to predicting churn that shows the most accurate outcome
A. Idris, A. Khan, and Y. Soo, (2013)[25]	Large orange datasets and Cell2Cell are publicly available to predict communication churns.	RotBoost (RB), RF, Rotation Forest, and Decorate (DEC) ensemble with mRMR	50,000 instances	RF+ rotation forest, and RB + DEC with better mRMR performance.
C. Kirui et al. (2013) [26]	European telecom dataset.	NB, BN, and C4.5	106405 instances 5.6% Churns	Decision tree performance achieved the best rate with high accuracy and the others at a higher positive rate

https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/1
DOI: http://doi.org/10.54623/fue.fcij.6.1.1

2.3. Datasets for predicting churn:

Predicting churn has been used in communications in both public and particular datasets. The private churn datasets used by the researchers, they have been compiled by several telecommunication operators. Most particular churn data sets are inaccessible due to property rights issues. Show the publicly available dataset used to predict the incidence of communications as shown in figure 2.

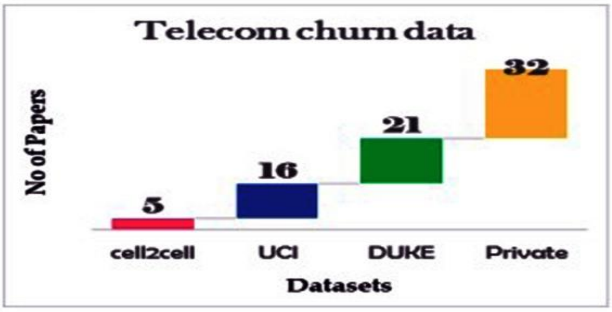

Figure2: Telecommunications dataset versus article count [15].

2.4. Detect churn

Here are several concepts and methods to detect customers who are about to switch to another operator. A good churn prediction system should not only detect at potential churners, but also provide a sufficiently long term forecast. When potential churners are identified, the marketing department usually contacts them and, if the customers are established as high churn risk, takes appropriate actions to prevent loss of customers. Through to analyze service quality, analysis data to detect churn in the telecommunications industry, and Detect of the variables most influencing data; as knowing by named “Feature Selection with the Boruta algorithm"[29].

2.4.1. Features selection

Features selection or dimension is the technique that is frequently used in machine learning to select a subset of features for a dataset to build a model. It is to select the most appropriate features for the problem. Feature selection gives a clearer understanding of the data by specifying the important features of data and its relationship with each other, which is the selection of the most influential variables in the data [27]. Algorithm and a method of selection select the characteristics that are able to classify by simplifying models to facilitate and avoid dimensions [28, 29].

2.4.2. Boruta Algorithm

Boruta is an encapsulated algorithm for the selection of relevant features, It can work with any classification that outputs the Variable Importance Measurement (VIM); Boruta uses RF. The method performs a descending search for pertinent features by comparing the significance of the original traits with the significance that can be randomly achieved, estimating them using their dilute copies, and phasing out the unrelated features to stabilize this test [30, 31].

2.5. Data mining techniques

This study presents review study of the most used algorithms and more effective for predicting customer churn.

2.5.1. Decision Tree

The decision tree is an exploratory model that appears in the form of a tree, and precisely each branch of the tree represents a taxonomic question, and its leaves represent parts of the database that belong to the classifications that have been built. The decisions tree a graphic style that helps decision-makers to clearly understand the available alternatives, errors, and expected results for each. Decision trees are easy to understand. Moreover, the decision tree can create models using data sets including numeric and categorical data. The advantage of this method is that it enables the decision-maker to know the impact of the decisions he makes at present on the alternatives he faces in the future.

There are 3 components of a decision tree:

A - Contract: It is of two types, a performance or disposition contract, and a chance or event contract.

B- Branches or Divisions: There are three types: performance branches, authentication branches, and end contract branches.

C- Returns and results.

The decision tree algorithm family contains the classic algorithms, as ID3, C4.5, and CART. Due to their ease of understanding, decision trees have been used in many areas such as customer classification and market conditions [32, 33].

Published by Arab Journals Platform, 2021

2.5.2. The neural network

The neural network is working in the same way as a human mind, just as the mind transmits processes and analyzes information, draws conclusions, and discovers patterns and predictions. Through it, can apply some of what the natural mind applies, although scientists are still discovering more about it today. It has a noteworthy ability to extract meaning from complex or imprecise data and functions as an input/output group. The neural network is a model for processing information that consists of a group of simple processing units called neurons.

Multilayer Perceptron Neural Networks (MLP) consists (usually three or more layers) neural networks consist of a group of artificial cells or nodes that are grouped in a matrix within layers connected, and that each connection between these nodes has a set of values called (Weights) that contribute to determining the values resulting from each processing element, based on the values entered for this element. Input units form a layer called the input layer and processing units form a layer that outputs the network's output, and between each of these layers, there are hidden layers that link each layer to the next layer. The network contains only one layer of input units, but it may contain more than one layer of processing, the outputs of each layer are inputs to the next layer except for the first layer that receives data as input from the external medium as shown in figure 3[34, 35].

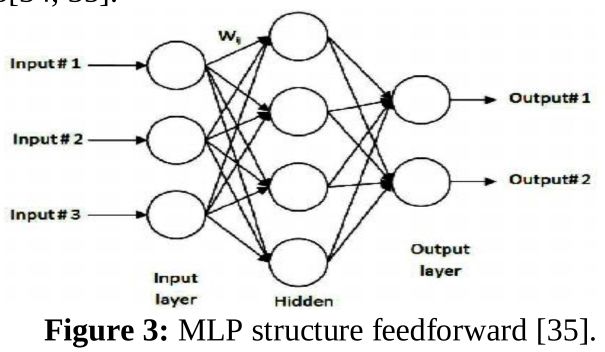

Figure 3: MLP structure feedforward [35].

2.5.3. Support Vector Machine

A support vector machine is a supervised machine learning algorithm that can be used for classification or regression problems. However, it is mostly used in classification matters [36]. The SVM algorithm can plot each data element as a point in space with an n dimension (where n is the number of properties) with the value of each property that has been given coordinate value. Then, can perform the classification by finding the hyper-plane that characterizes the two classes well. A support vector machine is a directed learning algorithm under machine learning and is used for either classification or regression tasks. But supportvector machine is mostly used in classification. Support vector machine is based on the idea of creating a hyperplane that divides the dataset into two classes in the best way, as shown in the figure below [20, 27].

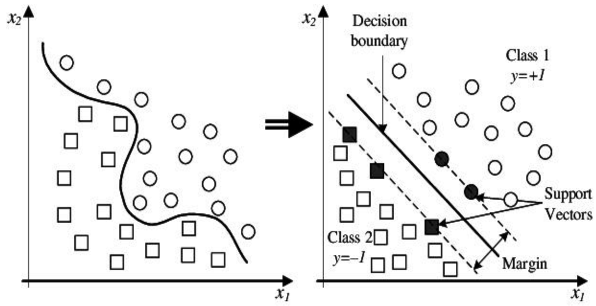

Figure 4: Precept of Support Vector Machines algorithm [20].

2.5.4. Random forest

The random forest is a versatile learning algorithm capable of performing regression and classification tasks. At the same time, it is also a method for reducing the dimensions of the data, used to deal with missing values, outliers, and other important steps in data exploration, and it has achieved good results. Besides, it also functions as an important method in group learning, as it shows its strength when combining several ineffective models into one effective model. In random forests, we will be creating a lot of decision trees, not just generating unique trees as in the CART model. When classifying and distinguishing a new object based on certain characteristics, each tree in a random forest will give its classification option and "vote" accordingly, and the total forest output will be the largest number of votes. Classification option; in the regression problem, the yield of the random forest will be the mean of all the outputs of the decision tree.

The decision trees were entered into the forest using random recurrence of data. The choice of the primary classifier is based on the variability and bias of the classifier, in general, classifiers with high contrast and low bias are preferred [37].

https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/1
DOI: http://doi.org/10.54623/fue.fcij.6.1.1

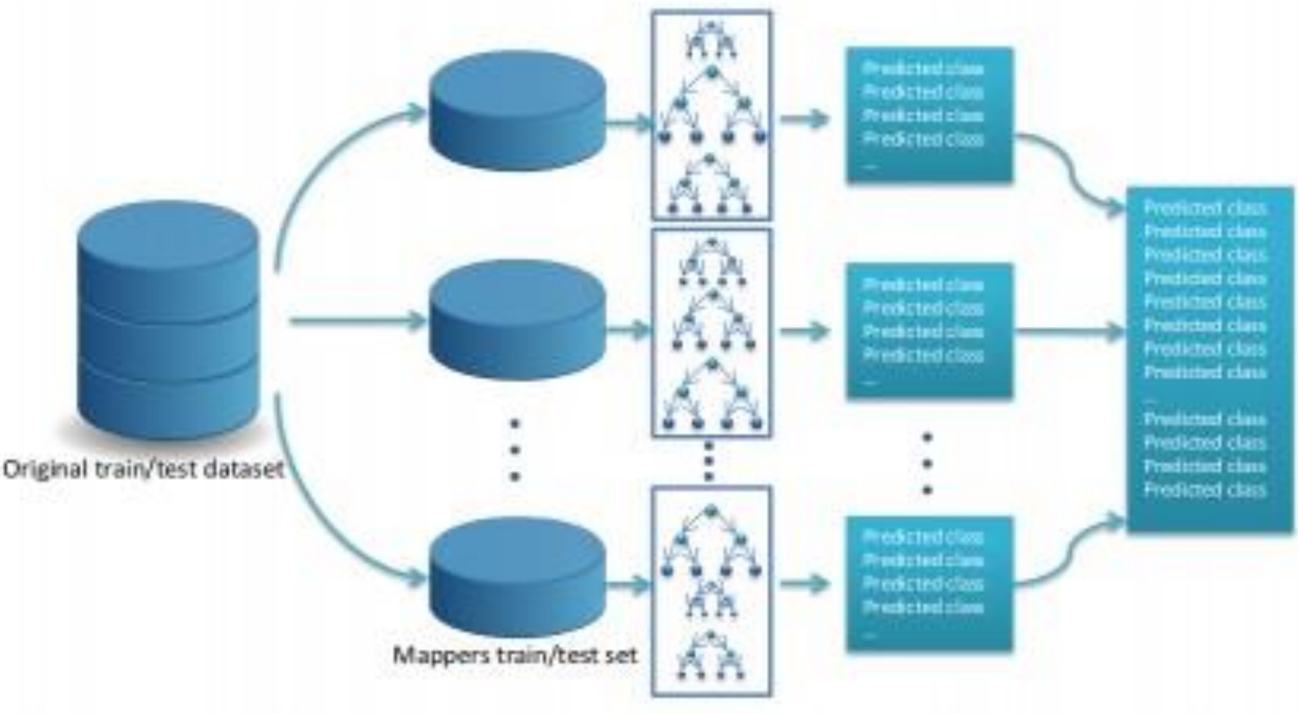

Figure 5: Random forest [37].

3. USING DATA MINING TO PREDICTING CHURN IN TELECOMS

The target of DM to infer knowledge from vast amounts of data, based on mathematical algorithms which are the basis of DM and is drawn from many sciences such as statistics, mathematics, logic, learning science, artificial intelligence and expert systems, pattern recognition science, and machine science[38,39]. Other sciences are considered smart and non-traditional sciences. Tools in DM are orange, WEKA, R project, mat lab, etc.

That is open-source software that is easy to use and helps researchers to complete their experiments and it is
software as orange, weka, R project, Matlab, etc. DM plays an important role in decision making and future predictions [40]. An important application of DM is amplification analysis in the telecom based industry [12,-41, 42, 43].

Agyei, Paul. (2013)[44] review how DM technologies are used in the telecommunication sector, how they affect how these technologies are used in the telecommunication sector, how they affect how these technologies are used in the telecommunication sector to improve their services and obtain more than predictable income, and classify customers and group them using appropriate algorithms. For this case and even knowing the link services are using algorithm through the weka tool [45]. Using the last details of all customers stored in the communication server, the authors also discussed the different techniques that are applied to analyze the service provided to customers and the importance of service quality. The results of every technology help companies to know their services and track errors where they are not far too late.

Gürsoy, Şimşek&Tuğba, Umman (2010) [46] identify customers who want to run a lot of specific campaigns using the customer data of the main telecom company in Turkey to determine the causes of customer churn. LR and DT analysis are some of the classification techniques. In this case, the DM is required to understand business needs, define a communication model, use resources effectively, and improve service quality. The telecom industry is now a mature market and it understands the importance of CRM, using DM technologies to demonstrate models for forecasting customer behavior, improving CRM, developing

Published by Arab Journals Platform, 2021

various campaigns, and marketing planning procedures to retain loyalty [47].

Sidra Ansar & SamreenLodhi (2015) [48] focuses on finding the impact of service quality on consumer satisfaction in the telecommunications sector in Pakistan. U fone, Warid, Telenor, Zong, and Mobilink took this research and design, aiming to investigate the impact of service quality on consumer satisfaction in the telecommunication sectors,
The result shows that service pricing is not able to manage customer needs because of rapid changes in technology, it will give relatively low communication cost and increasing in competition.Service rates are not necessary for developing customer loyalty because it is not required that high level of customer satisfaction may lead to increase customer loyalty [49].

F., Sahar. (2018)[28] Used a method for building a customer churn prediction model. The most communal techniques are applied to build customer churn forecast models based on identifying the determinants of customer satisfaction. In general, decision trees have the highest accuracy for building customer momentum forecast models [50]. The second contribution to this study is the method used to uncover the determinants of customer satisfaction, customer service, quality of service, quality of coverage, and prices were found to be major determinants of customer satisfaction in the telecom services industry. Using four widely used techniques for customer retention analysis: Classification and Regression Trees (CART), RF, RL, and SVM. The theoretical background of selected customer retention analysis techniques is also offered [51].

Deyu, Lulu (2014) [52] did many experiments with Weka. This is done to find out how many groups logically split Key Performance Indicator (KPI) data by applying the simple k-mean algorithm and it was the best classification model comparing J48 decision tree, Naïve Bayes, as well as multi-layer perception classifiers. It has been found that the cluster model is worth dividing KPI data into five groups based on their natural proximity. Typical data on triple stamps such as high traffic hours for 24 hours, one month, one week, day and night, is indicating that QoS KPIs are taken from the ethio live network. In telecommunication context, these tasks include CRM, and fraud detection, network failure detection, and isolation, are performed on the large and complex data generated by the network as this study does. The authors' study attempts to analyze KPI data that indicate general QoS in a mobile network by applying different data extraction techniques. On the other hand, various experiments are performed to determine the best classification model.
Thus, Multilayer Perceptron (MLP)-based classification model has the best accuracy for J48 and Naïve Bayes classification algorithms. However, other clustering and classification algorithms may reveal better accuracy.

Babu, P. S., N. R. Ananthanarayanan and R. Jadhav (2016) [12] reviewed the most applicable cases of churning classification. The accuracy of the selected samples was estimated in the customer's carrier database. Depending on the results of this study, group learning techniques that give better resolution to the RF and Adaboost models are recommended. Although the study can be expanded by using hybrid and deep learning models, other performance measurements can be used to evaluate performance. Model timing scales can also be a major indicator of performance. Models can also evaluate different datasets from various fields. In this paper, the authors attempt to compare and analyze the performance of different machine learning techniques used in the prediction question. Ten different techniques were selected in this study. Models were used in a communications dataset containing 3333 records. The results appear to outperform all other technologies with the same accurate identification techniques including DT CART, k-neighbors closest, SVM, logistic regression, RF, AdaBoosting trees, Stochastic Gradient Boosting, NB, and NN.

Oralhan, Burcu & Uyar, Kumru & ORALHAN, Zeki (2016) [41] reviewed the effect of six elements on customer churn expectations through DM methods. After-sales service software database is a data source. Data source variables are customer type, usage type, the reason for the change, subscriber period, and tariff. The data is scanned by a data extraction program. The data is compared with 8 classification algorithms and grouped simply by K. It will identify the variables that are most effective in customer churn

https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/1
DOI: http://doi.org/10.54623/fue.fcij.6.1.1

prediction. In this paper, use WEKA data mining algorithms for different classifications and clustering, to predict churning. This study examined the determinants of customer churn in the Turkish telecom industry services market using a sample of 498,357 actual and churning customer data.
Efficiency and performance are compared to Naive Bayes Network, Bayes Network from Bayes Classifier, Multilayer PerceptronJobs, Joint Reserve Intelligence Program (JRip), pruning rule-based classification tree (PART), one rule (OneR) from Rules Classifier, Random Forest, and J48 from Trees. This study aims to identify and explain the relevant factors in the modeling of both active and churn prediction [53]. On the other hand, this study aims to pool application data with k data extraction, which means a combined, classified and grouped technique used for prediction of occurrence in data extraction.

Anuj Sharma, Prabin Kumar Panigrahi (2013) [54] describes a methodology of neural Networks to predict churners' cellphone wireless service subscriptions. The experimental results indicate that the neural network-based method can expect the damper accuracy to be more than 92%. It has been observed that midsize NNs perform better predictions of customers churn when experimenting with different neural network structures. This means that he intends to demonstrate how to apply data mining techniques to help manage communications. Since the pre-defined data Processing stage in the DM is an important point for the performance of the final prediction model, the phase may include dimensional reduction or feature selection as well as data reduction. Second, along with NN, other popular prediction techniques can be applied together, such as SVM, GA, etc. to develop hybrid models.

HomaMeghyasi and Abas Rad (2020) [55] described one of the most effective methods for predicting flustered customers by using data mining and an Artificial Neural Network (ANN) which are more efficient than other methods available in forecasting due to their diversity in structure and different training algorithms [34].In this paper, a hybrid approach based on a genetic algorithm and a standard synthetic neural network is presented to predict which customers are likely to be moved in telecom operators. The purpose of implementing the genetic algorithm is to obtain the optimum unit structure in standard ANN so that it can provide the highest accuracy in failure prediction. Examination of the results of the proposed method and its comparison with other methods shows that the proposed method is superior to the other methods implemented with an accuracy of approximately 95.5% for the US telecommunications data (which was used for evaluation) and 88.1% for the Iran cell dataset. The reason for this result is the use ofenaticalgorithm to provide the optimal structure of the ANNs. As well as the use of the Feedforwardis ANN in which the connections between the nodes do not form a cycle.

Hossein Abbasimehr, Mostafa Setak, and M J Tarokh (2011) [56] presented the application of an Artificial Neural Network System (ANFIS) in the context of confusion prediction. In particular, they compared ANFIS as an ambiguous neuronal classifier with clear and functional classifiers including C4.5 and RIPPER. Predictions use outcome software using high output ANFIS application in the context of flop prediction. In particular, in this paper, they construct two ANFIS models including ANFIS-Subtractive Fuzzy Inference System (FIS) and ANFIS-FCM (FIS-based Fuzzy C-means (FCM)). The results showed that both ANFIS presentation and ANFIS-FCM modeling had acceptable performance in terms of accuracy, quality, and sensitivity. Besides, the algorithms used in the paper are KNN, RF, and GP with AdaBoost. This research uniquely uses the AdaBoost-based approach to predict motion modeling. The GP improvement procedure is called for by incorporating the Gadabouts optimization method (Gadabouts algorithm is applied to elicitation raw feature data) for multiplying program development in each semester and the latest predictions are made based on the weighted number of GP program outputs [57]. The momentum of the proposed prediction showed strong predictive capabilities of two churners for two standard communication dataset, as the high accuracy of the mixing prediction was reported at a resolution of0.89 in the cell dataset.

Published by Arab Journals Platform, 2021

4. DISCUSSIONS

Table 2 displays different DM techniques (DT, NN, SVM, RF clustering, association, classification, and others) that are used for predicting customer churn from 2003 and 2020 in the telecommunication-based industry [43,59].
According to [60, 61, 62], the most used and communal techniques reviewed in the academic literature as follows:

- Neural Network: Try to discover natural combinations of observations in the data, whichits information process using a communication agent.
- The random forest has a powerful prediction in classifying the dataset to investigate the relationships between variables [64].
- The support vector machine is mostly used in classifying problems [65]. Classification can be done by finding the superlative formula that distinguishes the two categories well [61].
- Decision trees: These are the most popular prediction models used by researchers [66].
- Cluster analysis: It tries to detect the natural gatherings of observations in the data [52].
- Regression Analysis: Regression analysis is the following common method used to examine relationships. Regression analysis is performed to evaluate the effect of several explanatory variables on the dependent variable [58].

The different mining techniques used to churn have been classified in Table 2. The most common data mining techniques are decision tree, regression, neural network, and clustering as shown in figure 6. Nevertheless, there is no clear common unanimity on the prediction method that will be used in the data collected. [68], [69].
Several authors have reviewed different studies of churn prediction models based on DM concepts that use machine learning and the most commonly used algorithms. [70, 71] Including the most important churn prediction techniques, that has been updated in recent years. The primary goal in churn in the telecommunication industry is to accurately assess the customer's survival and the problems that the customer is exposed to, to gain full knowledge of the churn in the customer's service life. The higher the quality is, the more customer retention, and proactive measures will be as a big factor in retaining customers [72, 73].

Table 2: Overview of contracts for research techniques in the industry used

References	algorithm
Sahar F. (2018) [28]	RF, RL, RF.DT, Adaboost, k-nearest Neighbor, ANN, LDA, Naïve Bayesian.
Oralhan, Burcu et al. (2016) [41]	Naive Bayes Network, Bayes Network from Bayes Classifier, Multilayer Perceptron Jobs, JRip, PART, OneR from Rules Classifier, Random Forest and J48
Agyei, Paul. (2013) [44]	Apriori algorithm
Gürsoy, Şimşek&Tuğba, Umman. (2010) [46]	Classification techniques LR and DT
Sidra Ansar, &SamreenLodhi (2015) [48]	Classification techniques
Deyu, Lulu. (2014)[52]	K-mean classification J48 decision tree, Naïve Bayes, multi-layer perception classifiers.
HomaMeghyasi and Abas Rad (2020) [55]	Artificial Neural Network (ANN)
Anuj Sharma,&Prabin Kumar Panigrahi(2013) [54]	(NN) to predicting churners
HosseinAbbasimehr, et al. (2011)[56]	RF.KNN, FIS, GP, ANFIS-FCM (FIS-based C-Fuzzy (FCM)).
Rodan, Ali et al. (2014) [58]	Classification and regression trees (CART), RF, RL, and

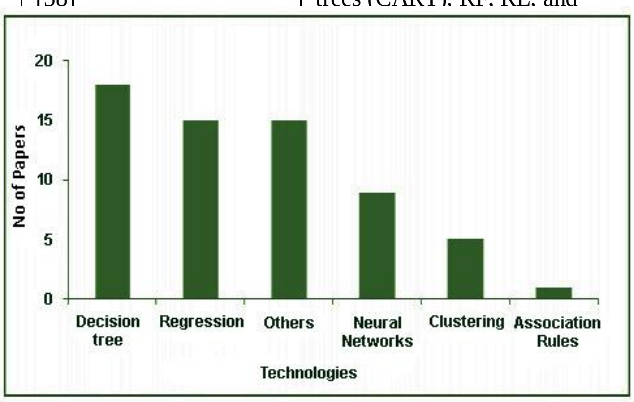

Figure 6: Research papers to predict churn [67].

https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/1
DOI: http://doi.org/10.54623/fue.fcij.6.1.1

5. CONCLUSIONS AND FUTURE WORK

The survey results show that the different DM techniques utilized to predict customer churn between 2003 and 2020 in the field of telecommunication. It identifies existing trends, intervals utilized, and challenges of DM techniques in the telecommunications area. The different mining techniques have been categorized for churn. The most common DM techniques are DT, NN, SVM, and RF. But it is the most effective and impactful in churn predicting of flocculation and has the highest accuracy in detecting churning. Whereas there is no clear general unanimity on the prediction method that will be utilized in the data that has been compiled. Moreover, most of the current studies participating in the survey use little sample data of customers' records. It undermines the accuracy and truth of analysis outcomes. It means that an experimental study with a huge dataset with additional dimensions raises the accuracy of the outcome. It is proposed that the future evolution of DM technologies could become more attitude to problems, plan proactively, and be specific to anticipating the desired "churner type". Furthermore, the crossbred models can be presented and compared with the current models. It can help design diversified dimensional customer datasets and innovate new technologies to manage various datasets. Decision-making depends on analyzing from DM technologies, can make churn predict more accurate and worthy insights for industry-based on telecommunications.

References

[1] Qadeer, Sara, "Service Quality & Customer Satisfaction: A case study in Banking Sector," pp.1-101, 2014.

[2] Jain S., Sharma N.K., Gupta S., Doohan N., "Business Strategy Prediction System for Market Basket Analysis" In: Kapur P., Kumar U., Verma A. (eds) Quality, IT and Business Operations. Springer Proceedings in Business and Economics. Springer, Singapore, pp.93-106, 2017.

[3] Alabar, T. Timothy, Ode Egena, Gbande, I. Richard, "Service Quality and Customer Satisfaction in Nigerian Mobile Telephony," British Journal of Marketing Studies, Vol.5, No.1, pp.37-49, 2017.

[4] SufianAlbadawi, Khalid Latif, Muhammad MoazamFraz and FatinKharbat, "Telecom Churn Prediction Model Using Data Mining Techniques," Bahria University Journal of Information & Communication Technologies Vol. 10, Special Issue, pp.8-14,2017.

[5] Amoo A. O, Akinyemi B. O, Awoyelu I. O, Adagunodo E. R., "Modeling& Simulation of a Predictive Customer Churn Model for Telecommunication Industry," Journal of Emerging Trends in Computing and Information Sciences, Vol. 6, No.11, pp.630-641, 2015.

[6] E. Shaaban, Y. Helmy, A. Khedr, and M. Nasr, "A proposed churn prediction model," International Journal of Engineering Research and Applications, Vol. 2, No. 4, pp. 693 – 697, 2012.

[7] Mamčenko, J. &Gasimov, J., "Customer churn prediction in mobile operator using combined mode," (ICEIS) Proceedings of the 16th International Conference on Enterprise Information Systems.Vol.1, pp.233-240, 2014.

[8] K. Dahiya and S. Bhatia, "Customer churn analysis in telecom industry," 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida , pp. 1-6, 2015.

[9] A. Ahmed and D. M. Linen, "A review and analysis of churn prediction methods for customer retention in telecom industries," *4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, pp. 1-7, 2017.

[10] Manpreet Kaur, Dr. PrernaMahaja, "Churn Prediction in Telecom Industry Using R," International Journal of Engineering and Technical Research (IJETR), Volume-3, Issue-5, pp.46-53, 2015.

[11] Golchia Jenabi, SeyedAbolghasamMirrosh, "Using Data Mining Techniques for Improving Customer Relationship management," European Online Journal of Natural and Social Sciences; Vol.2, No.3 Special Issue on Accounting and Management, pp.3143- 3149, 2013.

[12] Babu, P. S., N. R.Ananthanarayanan and R. Jadhav, "A Review on Customer Churn Prediction in Telecommunication Using Data

Mining Techniques,"International Journal of Scientific Engineering and Research (IJSER), Volume 4 Issue 1, pp .35-40, 2016.

[13] Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K., "Customer churn prediction in the telecommunication sector using a rough set approach," Neurocomputing, 237, pp. 242–254, 2017.

[14] G. Wang and N. Wu, "A Comparative Study on Contract Recommendation Model: Using Macao Mobile Phone Datasets," in IEEE Access, vol. 8, pp. 39747-39757, 2020.

[15] J. Pamina, T. DhiliphanRajkumar, S. Kiruthika, T. Suganya, Femila.F, "Exploring Hybrid and Ensemble Models for Customer Churn Prediction in Telecom Sector," International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-2, pp 299-309, 2019.

[16] W. Au, K. C. C. Chan, and X. Yao, "A Novel Evolutionary Data Mining Algorithm With Applications to Churn Prediction," IEEE Trans. Evol. Comput, Vol. 7, No. 6, 2003, pp. 532-545.

[17] S.-Y. Hung, D. C. Yen, and H.-Y. Wang, "Applying data mining to telecom churn management," Expert Syst. Appl., Vol. 31, No. 3, pp. 515-524, 2006.

[18] K. Cao and P. Shao, "Customer Churn Prediction Based on SVM-RFE, International Seminar on Business and Information Management," Wuhan, Vol. 1, pp. 306-309, 2008.

[19] R. J. Jadhav and U. T. Pawar, "Churn Prediction in Telecommunication Using Data Mining Technology," Int. J. Adv. Comput. Sci. Appl., Vol. 2, No. 2, pp. 17-19, 2011.

[20] Ionut Brandusoiu & G. Toderean, "Churn Prediction in the Telecommunications Sector Using Support Vector Machines," Annals of the Oradea University. Fascicle of Management and Technological Engineering. XXII (XII), pp. 19-22, 2013.

[21] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," Expert Syst. Appl., Vol. 38, No. 3, pp. 2354-2364, 2011.

[22] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," Eur. J. Oper. Res., Vol. 218, No. 1, pp. 211-229, 2012.

[23] K. Kim, C.-H. Jun, and J. Lee, "Improved churn prediction in telecommunication industry by analyzing a large network," Expert Syst. Appl., Vol. 41, No. 15, pp. 6575-6584, 2014.

[24] G. Olle, "A Hybrid Churn Prediction Model in Mobile Telecommunication Industry," Int. J. e-Education, e-Business, e-Management e-Learning, Vol. 4, No. 1, pp. 55-62, 2014.

[25] A. Idris, A. Khan, and Y. Soo, "Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification," in Springer Science Business Media New York, 39 (3), pp. 659-672, 2013.

[26] C. Kirui, L. Hong, W. Cheruiyot, and H. Kirui, "Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining,"IJCSI Int. J. Comput. Sci. Issues, Vol. 10, No. 2, pp. 165-172, 2013.

[27] Abdulrahman, S. A., Khalifa, W., Roushdy, M., & Salem, A.-B. M., "Comparative study for 8 computational intelligence algorithms for human identification. Computer Science Review," 36, 100237, pp.1-11, 2020.

[28] F.Sahar, "Machine-Learning Techniques for Customer Retention: A Comparative Study," International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 9, pp.273-281, 2018.

[29] A. Idris, A. Khan and Y. S. Lee, "Genetic Programming and Adaboosting based churn prediction for Telecom,"*IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Seoul, pp. 1328-1332, 2012.

[30] Miron B. Kursa, Aleksander Jankowski, Witold R. Rudnicki, "Boruta - A System for Feature Selection," Fundamental Informaticae volume101, pp. 271–285, 2010.

[31] Gangadharan, Sandhya, "Exploring Effective Feature Selection Methods for Telecom Churn Prediction," (ijitee) International Journal of Innovative Technology and Exploring Engineering.C8121.019320, pp.632-635, 2020.

https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/1
DOI: http://doi.org/10.54623/fue.fcij.6.1.1

[32] Nie, G., Rowe, W.; Zhang, L.; Tian, Y.; Shi, Y., "Credit card churn forecasting by logistic regression and decision tree," Expert Systems with Applications: An International Journal Vol.38, pp. 15273–15285, 2011.
[33] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," in: *IEEEAccess*, Vol. 7, pp. 60134-60149, 2019.
[34] Bharati, M. &Ramageri, "data mining technique applications," Indian Journal of Computer Science and Engineering. Vol. 1 No. 4, pp.301-305, 2010.
[35] Adwan, Omar &Faris, Hossam&Jaradat, Khalid &Harfoushi, Osama &Ghatasheh, Nazeeh, "Predicting Customer Churn in Telecom Industry using Multilayer Preceptron Neural Networks: Modeling and Analysis," Life Science Journal. 11, pp. 75-81, 2014.
[36] A. Naghibi, K. Ahmadi, A. Daneshi, "Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping," Water Resources Management ,31 (9) ,pp. 2761-2775,2017.
[37] Jayaswal, Pretam& Prasad, Bakshi&Tomar, Divya&Agarwal, Sonali. , "An Ensemble Approach for Efficient Churn Prediction in Telecom Industry," International Journal of Database Theory and Application. 9, pp. 211-232, 2016.
[38] Giuseppe Ciaburro Balaji Venkateswaran, "Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles," pp.1-272, 2017.
[39] C. Zhao, Y. Wu and H. Gao, "Study on Knowledge Acquisition of the Telecom Customers' Consuming Behaviour Based on Data Mining," 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, Dalian, pp. 1-5, 2008.
[40] Layla Safwat Jamil, "Data Analysis Based On Data Mining Algorithms Using Weka Workbench,"International Journal Of Engineering Sciences & Research Technology, 5(8), pp. 262-267, 2016.
[41] Oralhan, Burcu&Uyar, Kumru& ORALHAN, Zeki. , "Customer Satisfaction Using Data Mining Approach," (I J I S) and Applications in Engineering, pp. 63- 66, 2016.
[42] http://en.wikipedia.org/wiki/Weka_(machine_learning)
[43] Krotov, Vlad, "A Quick Introduction to R and RStudio," pp.1-15, 2017.
[44] Paul MensahAgyei, James M. Kilika, "The Relationship between Service Quality and Customer Loyalty in the Kenyan Mobile Telecommunication Service Industry," European Journal of Business and Management, Vol.5, No.23, pp.26-36, 2013.
[45] Jagtap, Sudhir B. and G. Kodge B., "Census Data Mining and Data Analysis using WEKA," *ArXiv* abs/1310.4647, (ICETSTM), pp 35-40, 2013.
[46] Gursoy, Şimşek&Tugba, Umman, "Customer churn analysis in telecommunication sector," *Istanbul University Journal of the School of Business*, Vol: 39, No: 1, pp.35-49, 2010.
[47] Abu Ellaban, Mahmoud Ayesh, "The Role Of Data Mining Technology In Building Marketing And Customer Relationship Management (CRM) For Telecommunication Industry Case Study: Jawwal Mobile Operator - Gaza Strip," pp.1-134, 2013.
[48] Sidra Ansar&SamreenLodhi, "The impact of service quality on customer satisfaction in telecom sector of Pakistan, An empirical study of Pakistan," International Journal of Scientific & Engineering Research, Volume 6, Issue 10, pp.1639-1645, 2015.
[49] Kayaalp, Fatih, "Review of Customer Churn Analysis Studies in Telecommunications Industry," Karaelmas Science and Engineering Journal. Vol.7, pp. 696-705, 2017.
[50] Hassouna, M., Tarhini, A., Elyas, T., &AbouTrab, M. S., "Customer Churn in Mobile Markets: A Comparison of Techniques," International Business Research, 8(6), pp. 224-237, 2015.
[51] T. Bujlow, T. Riaz and J. M. Pedersen, "A Method For Classification Of Network Traffic Based On C5.0 Machine Learning

Algorithm,"International Conferenceon Computing, Networking and Communications *(ICNC)*, Maui, HI, pp. 237-241, 2012.
[52] Deyu, Lulu, "Data Mining Approach to Analyze Mobile Telecommunications Network Quality of Service: The Case of Ethio-Telecom," pp.1-117, 2014.
[53] Freddie MathewsKau, Hlaudi DanielMasethe and Craven KlaasLepota, "Service Provider Churn Prediction for Telecoms Company using Data Analytics," (WCECS) Proceedings of the World Congress on Engineering and Computer Science, Vol. I, pp.1-4, 2017.
[54] Sharma, Anuj &Panigrahi, "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services," International Journal of Computer Applications. Volume: 27 Issue: 11, pp. 26 – 31, 2013.
[55] HomaMeghyasi and AbasRad, "Customer Churn Prediction in Irancell Company Using Data Mining Methods," EasyChair Preprint No. 2422, pp.1-6, 2020.
[56] HosseinAbbasimehr and MostafaSetak and M. J. Tarokh, "ANeuro-Fuzzy Classifier for Customer Churn Prediction," International Journal of Computer Applications, Vol.19 (8), pp.35-41, 2011.
[57] Ahmad, A.K., Jafar, A. &Aljoumaa, K., "Customer churn prediction in telecom using machine learning in big data platform," Journal of Big Data,6(1), pp.1-24, 2019.
[58] Rodan, Ali &Faris, Hossam& Al-sakran, Jamal & Al-Kadi, Omar, "A Support Vector Machine Approach for Churn Prediction in Telecom Industry," International journal on information, Vol. 17(8), pp.3962-3971, 2014.
[59] Lee, S., Bongsik Shin and H. Lee, "Investigation of Factors That Have Impacts on Usage Increase and Decrease of Mobile Data Service," Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS), pp.82-82, 2008.
[60] Hashmi, N.; Butt, N. A.; Iqbal, M., "Customer Churn Prediction in Telecommunication: A Decade Review and Classification," (IJCSI) International Journal of Computer Science Issues, (10), pp. 271-282, 2013.
[61] Ammar A.Q. Ahmed, D. Maheswari, "Churn prediction on huge telecom data using hybrid firefly based classification," Egyptian Informatics Journal, Vol. 18, No. 3, pp.215-220, 2017.
[62] Oghojafor, Benjamin &Bakarea, Rasaki&Omoera, Charles & Adeleke, I.A., "Discriminant Analysis of Factors Affecting Telecoms Customer Churn," International Journal of Business Administration, vol. 3, No. 2, pp.59-67, 2012.
[63] Adarsh Mishra. Mohan RaoMamdikar Asst. Prof DiptiVerma, "A Review on Customer Churn Detection Techniques Using Map Reduce," International Journal of Advance Research in Computer Science and Management Studies, Volume 6, Issue 7, pp.1-6, 2018.
[64] P.K.D.N.M. Alwis, B.T.G.S. Kumara and H.A.C.S. Hapuarachchi, "Customer Churn Analysis and Prediction in Telecommunication for Decision Making,"International Conference On Business Innovation (ICOBI) At Colombo, pp.40-45, 2018.
[65] I. M. M. Mitkees, S. M. Badr and A. I. B. ElSeddawy, "Customer churn prediction model using data mining" techniques, 13th International Computer Engineering Conference (ICENCO), Cairo, pp. 262-268, 2017.
[66] L. Bin, S. Peiji and L. Juan, "Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service,"*International Conference on Service Systems and Service Management*, Chengdu, pp. 1-5, 2007.
[67] Mahajan, Vishal &Misra, Richa&Mahajan, Renuka, "Review of Data Mining Techniques for Churn Prediction in Telecom," Journal of Information and Organizational Sciences, 39,pp 183-197,2015.
[68] Mgbemena, Chidozie, "A data-driven framework for investigating customer retention," Brunel University London. Book. (2016)
[69] B. Mishachandar, Kakelli Anil Kumar, "Predicting customer churn using targeted proactive retention," International Journal of

https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/1
DOI: http://doi.org/10.54623/fue.fcij.6.1.1

Engineering & Technology, Vol. 7, No (2.27), pp. 69-76, 2018.
[70] M. Kaur and S. Kang, “Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining,” Procedia Computer Science, Vol. 85, pp.78-85, 2016.
[71] P. Y. Niranjan, H. H. Kenchannavar and R. Medar, “Data mining techniques in telecom sector,” 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, (2017), pp. 2727-2731.`
[72] Nadeem Ahmad Naz, Umar Shoaib and M. ShahzadSarfraz, “A Review On Customer Churn Prediction Data Mining Modeling Techniques,” Indian Journal of Science and Technology, 11(27), pp. 1-7, 2018.
[73] V. Umayaparvathi, K. Iyakutti, “Applications of Data Mining Techniques in Telecom Churn Prediction,” International Journal of Computer Applications (0975 – 8887) Vol. 42, No.20, pp.5-9, 2012.