COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING
e-ISSN: 2582-7464

Computational Intelligence and Machine Learning
Vol-2, Issue-2, October 2021

Prediction of Customer Churn in Telecom Industry: A Machine Learning Perspective

Lopamudra Hota[1], Prasant Kumar Dash jer[2*]

[1] M Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India.
[2] Department of Computer Science and Engineering, C. V. Raman Global University, Bhubaneswar, India.
*Corresponding Author Email: [2] prasant.oitburla@gmail.com

Abstract

The business world is becoming increasingly saturated in today's competitive environment. There is a great deal of competition in the telecommunications industry, especially due to various vibrant service providers. As a result, they have had difficulty retaining their existing customers. As attracting new customers is much more costly than retaining current ones, now is the time to ensure the telecom industry maintains value by retaining customers over acquiring new ones. Numerous machine learning and data mining methods have been proposed in the literature to predict customer churners using heterogeneous customer records over the past decade. This research gives a brief idea on the Customer Churn problem, and explores how various machine learning techniques can be used to predict customer churn via models such as XGBoost, GradientBoost, AdaBoost, ANN, Logistic Regression and Random Forest, and also compare the effectiveness of the models in term of accuracy.

Keywords

AdaBoost, Customer Churn, GBoost, Machine Learning, Prediction Model, Random Forest, XGBoost.

INTRODUCTION

It is much more expensive to acquire new customers than to retain existing ones. The cost of acquiring a new customer is six to seven times greater than retaining an existing customer [1]. Customers are regarded as the most significant assets in any industry or sector because they provide the majority of the profit. Companies today are putting a greater emphasis on convincing and retaining their existing customers. Consumers' churn can be reduced if the firm correctly predicts customer behavior, expands the link between consumer attrition, and has factors under its control. By determining the difference between churners and non-churners, you can predict churn [2]. Customer churn is perceived as a significant issue in service-based firms due to its direct effect on revenues. As a result, many businesses focus on reducing churn and identifying appropriate processes for doing so. Firms intend to keep their customers through spending and minimizing profits. The best way to retain consumers is to reduce the rate of churn, which refers to the phenomenon of a consumer switching from one service provider to another or ceasing to use a particular service over a certain period. A variety of reasons could be identified in the past if a firm analyzed its history of data and adopted machine learning technology, which can identify the consumers who are likely to churn [3]. Almost every firm now has data about its clients and about the behavior of their customers thanks to the development of data management. A major advantage of big data is the high quality and diversity of consumer data and the ability to provide a strategic benefit to the company. Data mining assists in identifying, identifying, and understanding the behavior of the consumer, thus optimizing business operations and enhancing customer management effectiveness [4].

Many factors can lead to a firm losing its customers. Cost, quality, and service quality all play a significant role in that. A huge outflow of consumers affects the valuation of any firm. Market reputations and stakeholder trust are destroyed mainly by most firms. However, it is also essential to determine the level of customer satisfaction to retain and attract customers. Identifying how satisfied consumers are can be a challenging task. As the base of consumers grows, it becomes more challenging. The value-added service is another cause of consumer churn. The telecom industry has introduced a new offer called Triple Play, which spans television, phone, and internet services. In addition to adding value, this offer helps retain consumers.

Furthermore, it maximizes the revenue allocated to each user directly by the company [5]. Telecom companies face a unique challenge in predicting churn. Telecom analytics is a type of business intelligence explicitly used to satisfy the demands of the telecom sector. Analytics in telecom is primarily focused on maximizing profits, minimizing costs, and decreasing fraud. The purpose of telecom analytics is to forecast, multidimensionally, and optimize. Most companies suffer from customer churn, affecting their revenues when a customer moves from one service provider to another in the telecom sector. To grow their revenue-generating base, Telco companies must both attract new customers and avoid terminations (churn). According to churn analysis, customers terminate their contracts for various reasons, including better price offers, more exciting packages, poor service experiences, and changes in their personal circumstances.

The paper is organized as, Section 2 presents the review of literature related to customer churn based on machine learning, Section 3 states the churn problem in details.

Section 4, describes the proposed work comprising of the model design and result analysis; finally, section 5 depicts the conclusion of the work.

LITERATURE STUDY

Kassem et al. [6] identified the main factors influencing customer churn and identified customers likely to churn by analyzing social media. The results are analyzed using various machine learning algorithms such as Deep Learning, Logistic Regression, and Naïve Bayes. In [7], the study's main goal is to predict customer churn in telecom by using machine learning and big data platforms. Consumer churn can be predicted using machine learning methods. Consumer churn prediction using KNN and big data depicts the study results shows an accuracy rate of 0.80 percent for predicting consumer churn, and 1.01 percent for the area under the curve.

With specific reference to SyriaTel Telecom Company, Ahmad et al. [3] developed a mechanism for predicting the churn of consumers. Random Tree, Decision Tree, extreme gradient boosting algorithm, and GSM tree algorithm were chosen in this research. Selecting the features as well as adding the features of the mobile social network have had a significant influence on the success of the developed model as SyriaTel's area under the curve (AUC) value has reached 93.301 percent. In all measurements, the extreme gradient boosting algorithm achieved the best results. Almuqren et al. [8], offers a new approach to predicting churn and compares the telecom industry using social media mining. In this study, Arab Twitter mining was used to predict churn in Saudi Telecom companies for the first time. Based on various standard metrics to the ground-truth actual outcomes offered by a telecom company, the newly proposed method is proven to be effective.

Different techniques have been used to predict customer churn, including data mining, machine learning, and hybrid technologies. Churn predicting, and retention techniques help companies identify, predict, and prevent churn. Most of them used decision trees because it is a recognized method for determining customer churn, but it's a challenge to solving complex problems this way. The study shows that reducing the data improves the accuracy of the decision tree [9]. Customer prediction algorithms and historical analysis are sometimes used in data mining. In addition to discussion of regression trees, decision trees, neural networks, and some other data mining methods were examined in [10]. Our system is designed based on the data analysis and visualization of data collected from telecom department. The churn prediction and analysis of the machine learning models is done based on performance metrics such as precision, recall, f1-score and accuracy.

CHURN PROBLEM

When it comes to a business environment, customer attrition simply refers to customers switching services. Subscriber churn or customer churn is similar to attrition, when a customer switches from one service provider to another anonymously. In machine learning terms, churn prediction is a supervised (i.e. labeled) problem: Given a predetermined forecast horizon, one goal is to predict the number of subscribers that will churn over that time frame. Churn Prediction identifies churners in advance, before they leave the network. Therefore, the CRM department is able to prevent subscribers from churning in the future by implementing retention policies that attract and retain likely churners. Thus, the company would not suffer a potential loss. A mobile subscriber's past calls, along with his or her personal details and business information, are inputs into this problem. A list of churners is also provided for the training phase. When a model has been trained to the highest level of accuracy, it must be able to predict the churners from the real dataset which does not include any churn labels. The knowledge discovery process categorizes this problem as predictive modeling or data-mining. Figure 1 portray a model of churn prediction with four steps; 1) Preprocessing of customer data 2) Feature extraction for model design 3) Model design by classifiers and validation4) Computation of performance metrics for model comparison.

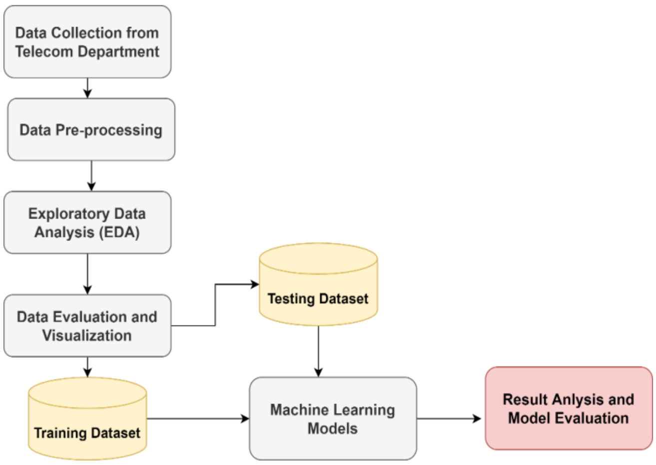

Figure 1. Flow Diagram of Proposed Work

PROPOSED MECHANISM

In this research, some of the addressed questions will be; analysis of the most important feature for customer churn, which type of customers are leaving more, and which machine learning model is the best one for result analysis and prediction. We explored classification techniques, compared their accuracy, as well as other metrics, precision, recall, f1-score, True/False Positive Rates. Data Preprocessing checks for missing values, correlated variables, and outliers; EDA for hypothesis generation; data scaling to improvise data accuracy; train and test dataset generation; training models for cross-validation and plotting data accuracy results from test data.

Dataset

IBM Telecom's Kaggle Dataset was used in this research paper. Several extremely important parameters for predictive churn analysis were included in the dataset, and the data is extremely large. 7043 instances of 21 attributes are contained in the dataset. Features include details about demographic information like gender, age, and dependents, services they have signed up for, contract information, payment methods, paperless billing, monthly charges, and a variable in which we anticipate which customers have left within the past month. Input data is in CSV format and visualized using various visual elements such as graphs, helping to identify trends, outliers, and patterns in the data. The analysis starts with data cleaning followed by graphical analysis, machine learning model, estimation and result analysis.

Methodology

Data Pre-processing

A data set consists of features and N rows. There are many formats used for values. Duplicate values and null values can lead to loss of accuracy in a dataset and dependent values. Various data sources have been used to collect data, so each uses a different format to represent a single value, such as whether someone represents Male/Female or M/F. In order to avoid noisy data, null values, and incorrect sizes, an image in 3-dimension should be reduced to a 2-dimension format by reducing it to 0 and 1. Images can be cleaned with OpenCV or Panda's tabular data. Making the data useful is paramount since generating unsatisfactory results or achieving less accurate results can be affected by unwanted or null values. Missing and incorrect values are prevalent in the data set. The entire dataset was analyzed and only the most useful features were listed. By listing features, the listing will be more accurate and contain only useful features. For a knowledge-based approach to data selection, feature selection is a crucial step. Out of the dataset here, we chose the features necessary for improving performance and helpful for decision-making, while the rest of the features have less importance.

Data Exploration

Explorative Data Analysis (EDA) provides a clear and better understanding of data patterns and potential hypothesis. The distribution of feature is an essential for trend analysis of dataset.

a. The gender distribution graph depicts that male and female distribution is nearly same as per figure 2.

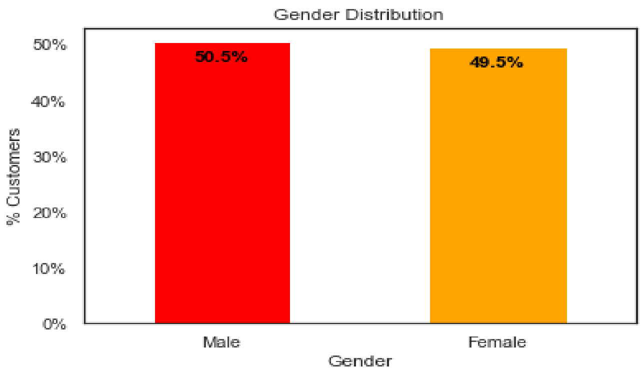

Figure 2. Gender Distribution

b. Most of the customers are youngsters rather than senior citizens as depicted in graph (figure 3).

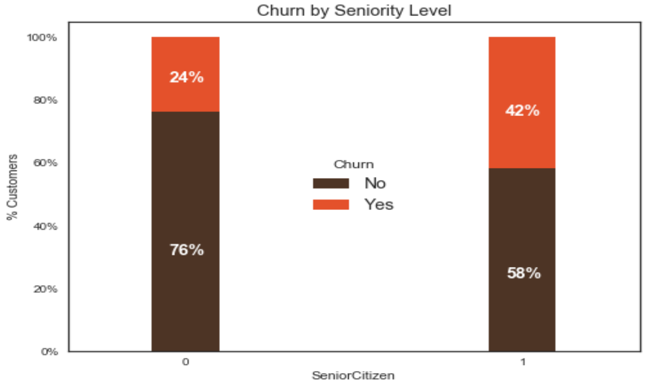

Figure 3. Customer Churn based on Seniority Level

c. Figure 4 depicts 48 percentage of the customers partner dependent, while 30 percentage have dependents. Fascinatingly, only about half of the customers who have a partner have a dependent, while the other half do not. Furthermore, a majority (80%) of the customers without a partner do not have dependents.

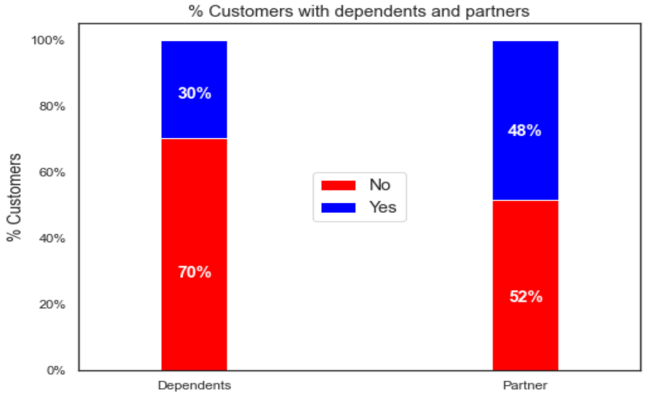

Figure 4. Customer Churn based on Dependents and Partners

d. Customer Tenure and Account Information: From the histogram as in figure 5, we can see that many customers have been with the telecom company just for a month, while many others have been with the company for about 72 months. Different contracts may apply to different customers. Due to this, depending on the contract they are in, it might be easier or harder for customers to stay or leave the telecom company.

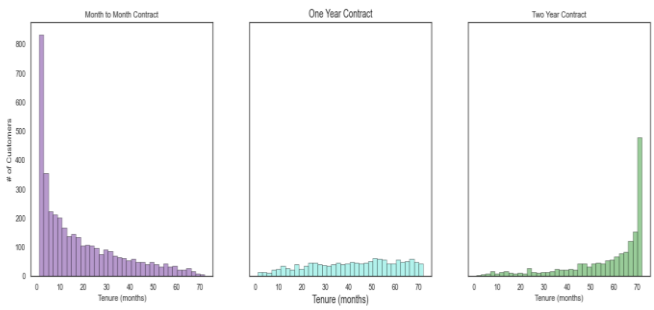

Figure 5. Customer Churn based on tenure and year of contract

Contrary to popular belief, the typical monthly contract usually lasts between one and two months, while the typical two-year contract lasts around seventy months. People who sign longer contracts and stick with the company longer show customer loyalty to the company.

e. Customer Distribution:

The graph depicts the distribution of services used by customers. From the graph of the relation between monthly and total charges, the total charges are directly proportional to customers' monthly bills. Finally, the rate of churn is depicted in figure6.

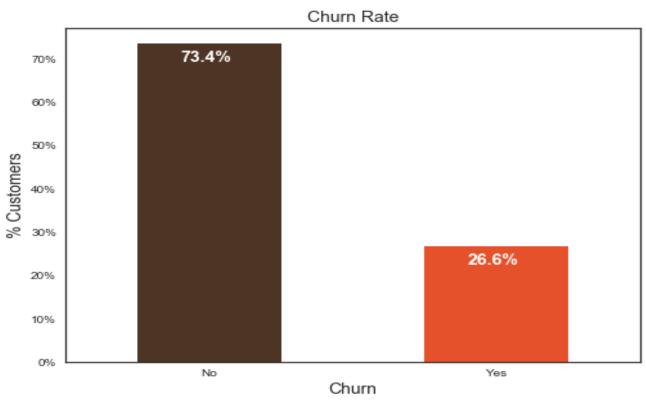

Figure 6. Customer Churn Rate Visualization

f. Churn by Monthly Charges vs Total Charges: customer's churn directly proportional to monthly charges; and churn rate is inversely proportional to total charges depicted in figure

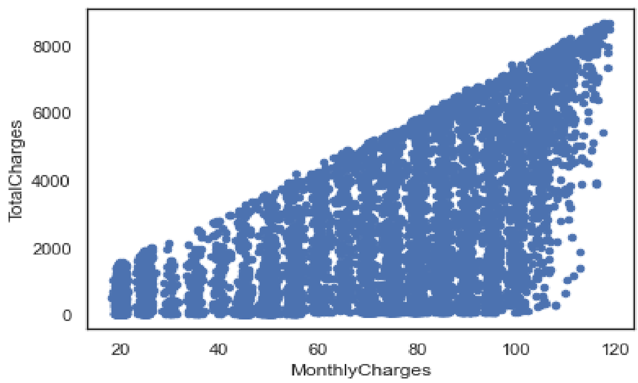

Figure 7. Monthly vs Total Charges

Model Design and Analysis

Machine Learning Models

ML (machine learning) is a form of artificial intelligence, in which software applications make better predictions without being explicitly programmed. Here, historical data are trained to predict the test result or output [11]. There are typically three types of ML algorithms Supervised, Un-supervised and Reinforcement Learning algorithms.

Logistic regression:

Logistic regression is a supervised learning approach for predicting a target variable's probability. Since the variable of interest is dichotomous, there are only two possibilities. In other words, the dependent variable comprises binary data that can either be coded as 0 (for failure) or as 1 (for success). As a function of X, logistic regression predicts the value of P(Y=1). This is one of the simplest ML algorithms that can be applied to a variety of classification problems such as diabetes prediction, cancer detection, fraud detection, spam detection and many more.

Random Forest:

The Random Forest algorithm is both a classification and regression learning algorithm that is used for supervised learning. This method is mainly used to solve classification problems. Forests can be thought of as a forest of trees, and a forest that is more robust has more trees. In the same way, random forests create decision trees using data samples and then obtain their predictions. Ultimately, they vote on which solution is the best. By averaging the result, it allows us to reduce over-fitting to a minimum.

AdaBoost:

It is short for Adaptive Boosting - is one of many Ensemble Methods used to improve neural networks. The Adaptive Boosting method assigns higher weights to incorrectly classified instances since the weights are re-assigned. A boost is used in supervised learning to reduce bias and variance. The system is based on the principle of sequential growth of learners. In all cases, except the first, the subsequent learners are grown from the previous ones. To put it simply, weak learners are transformed into strong ones.

GradientBoost:

Gradient Boosting Machines combine predictions from multiple decision trees into a final prediction. It is important to remember that gradient boosting machines only use weak learners. To select the best split in every decision tree, different nodes take into account different factors. In other words, each tree is different, and thus it can capture different signals from the data. Additionally, each new tree corrects previous errors. As a result, every subsequent decision tree builds upon the mistakes of the previous trees. An algorithm for gradient boosting machines builds trees sequentially in this manner.

XGBoost:

It is a gradient boosting-based ensemble Machine Learning algorithm based on decision trees. In unstructured data prediction problems (images, text, etc.); it is suitable for solving regression and classification problems, ranking problems, and user-defined prediction problems.

ANN:

Artificial Neural Network (ANN) can be considered the core element of Deep Learning. In addition to their versatility, adaptability, and scalability, ANNs are also suitable for handling large datasets and highly complex Machine Learning problems, like image classification, speech recognition, or video recommendation. In ANN algorithms, the aim is to create the most minimal error function possible by selecting the optimal weights and bias terms [12]. thought of as the most sophisticated version of Machine Learning.

Result Analysis

Performance Metrics

Co-relation Matrix: A correlation is a description of how variables are related to one another. Feature variables such as these can be used as input for forecasting our target variable. A correlation is a statistical technique that determines how one variable moves/changes in relation to another. A correlation matrix is a table presenting multiple variables and their 'correlations'. Rows and columns in this matrix represent variables, and each value in the matrix represents a correlation coefficient between variables depicted in figure 7.

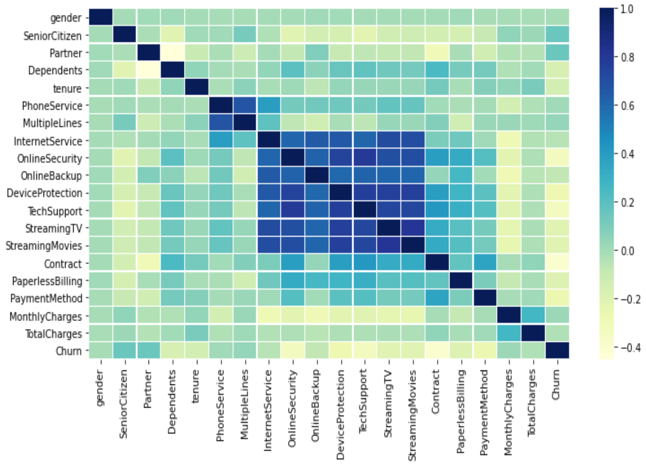

Figure 8. Confusion Matrix

Confusion Matrix shows how many True Positives/True Negatives and False Positives/False Negatives there are in a prediction.

TP: Number of customers who will actually default is also predicted as defaulting

TN: Number of customers not expected to default is also reported as non-default

FP: Number of customers who are predicted to default but won't actually default

FN: Number of customers predicted to default but actually defaulting

A telecom company needs to understand which customers will default. Therefore, we should keep the number of False Positives (FP's) as low as possible, as this will predict that the riskier customers will not be too risky. All of these factors should be considered as we assess every classification model.

'Accuracy' is the easiest performance metric to grasp and is simply the ratio of rightly predicted observations to the total observations. It is easy to assume that the best model is the one with high accuracy. It is true that accuracy is an important measure, but only if you have symmetric datasets with relatively equal values of false positives and false negatives. As a result, you have to evaluate the model's performance by looking at other parameters.

Accuracy = TP+TN/TP+FP+FN+TN

Precision is the percent of positively predicted observations among all predicted positive observations. This metric answers the question, how many of all passengers that were labeled as survivors actually survived? Precision is related to a low false positive rate.

Precision = TP/TP+FP

Sensitivity (Recall) - Recall is measured by how many of the correctly predicted positive observations have actually occurred in the class - yes. How many passengers did we label from all of those that truly survived? is the question that can be answered here.

Recall = TP/TP+FN

An F1 score is calculated by summing Precision and Recall. This score considers both false positives and false negatives. Although F1 is not intuitively as easy to understand as accuracy, it is usually more useful than accuracy if your classes are unevenly distributed. True positives and false negatives have similar costs when it comes to accuracy. Precision and recall should be considered along with the cost of false positives and false negatives.

F1-Score = 2 TP / (2TP + FP + FN)

Model Comparison

Machine learning algorithms were used on the dataset to perform several experiments on the proposed churn model. The results were observed pertaining to precision, recall, f1-score, and accuracy values. Table 1 presents the details of the performance results of all the models based on the metrics. As per the result, XGBoost model has the highest accuracy value of 81.14%, followed by RF, AdaBoost, GBM, LR and ANN. But, for the rest of the metrics like precision, f1-score and recall ANN has a higher value followed by three of the boosting models (XGBoost, AdaBoost, GBM) with approximately the same result, then LR and RF.

Table 1. Model Analysis

ML Models	Accuracy	Precision	Recall	F1-Score
GradientBoost	80.41%	0.66	0.58	0.59
AdaBoost	80.59%	0.67	0.57	0.58
XGBoost	82.20%	0.66	0.45	0.56
ANN	79.98%	0.89	0.84	0.85
LR	80.29%	0.68	0.56	0.60
RF	81.10%	0.66	0.49	0.56

CONCLUSION

As the telecommunications industry continues to grow, the problem of customer churn has significant growth as well. Retaining customers is a critical challenge in the telecommunications industry since it reduces customer churn through increased customer satisfaction. The use of predictive analytics can help tackle this threat by identifying vulnerable customers and implementing customer-centric retention measures. The proposed prediction analysis can be solved with the help of machine learning models. The paper describes the problem of churn and the importance of preventing the same. The paper investigated the realm of machine learning methods and applied them to the dataset. By modeling and testing, ANN and XGBoost models were found to out-perform other models in terms of precision, recall, f1-score and accuracy. We can extend the same for various other datasets from the telecom department and implement methodologies to achieve better results in future work. Big-data analytics with machine learning approach can also be implemented for the datasets.

REFERENCES

[1] Suguna, R., M. Shyamala Devi, and Rincy Merlin Mathew. "Customer churn predictive analysis by component minimization using machine learning." International journal of innovative technology and exploring engineering 8.8 (2019): 3229-3233.

[2] Vafeiadis, Thanasis, et al. "A comparison of machine learning techniques for customer churn prediction." Simulation Modelling Practice and Theory 55 (2015): 1-9.

[3] Ahmad, Abdelrahim Kasem, Assef Jafar, and Kadan Aljoumaa. "Customer churn prediction in telecom using machine learning in big data platform." Journal of Big Data 6.1 (2019): 1-24.

[4] Ahn, Hyunchul, et al. "Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques." Expert Systems with Applications 38.5 (2011): 5005-5012.

[5] Kumar.N and Naik.C. “Comparative analysis of machine learning algorithms for their effectiveness in churn prediction in the Telecom industry”. International research journal of engineering and technology, (2017), vol 4, iss 8, pp: 485-489.

[6] Abou el Kassem, Essam, et al. "Customer Churn Prediction Model and Identifying Features to Increase Customer Retention based on User Generated Content", (IJACSA) International Journal of Advanced Computer Science and Applications, (2020), 11. 5.

[7] M. A. Hassonah, A. Rodan, A. Al-Tamimi and J. Alsakran, "Churn Prediction: A Comparative Study Using KNN and Decision Trees," 2019 Sixth HCT Information Technology Trends (ITT), (2019), 182-186, doi: 10.1109/ITT48889.2019.9075077.

[8] Almuqren, L.A.; Moh’d Qasem, M.; Cristea, A.I. Using Deep Learning Networks to Predict Telecom Company Customer Satisfaction Based on Arabic Tweets; ISD: Tolerance, France, (2019).

[9] Mahajan, Vishal, and Renuka Mahajan. "Variable Selection of Customers for Churn Analysis in Telecommunication Industry." International Journal of Virtual Communities and Social Networking (IJVCSN) 10.1 (2018): 17-32..

[10] Hudaib, Amjad, et al. "Hybrid data mining models for predicting customer churn." International Journal of Communications, Network and System Sciences 8.05 (2015): 91-96.

[11] Labhsetwar, Shreyas Rajesh. "Predictive Analysis of Customer Churn in Telecom Industry using Supervised Learning." ICTACT Journal on Soft Computing 10.2 (2020): 2054-2060..

[12] Hota, L., Dash, P. “Comparative Analysis of Stock Price Prediction by ANN and RF Model”, Computational Intelligence and Machine Learning, 2.1, (2020), 1-9.