IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.12 ⁜ Vol. 10, Issue 2, February 2023

DOI: 10.17148/IARJSET.2023.10209

CUSTOMER CHURN ANALYSIS IN TELECOMMUNICATION

Abhishek Yadav[1], Gulshan Kumar S[2], Akash AS[3], Ajay S Kalburgi[4]

Department of Computer Science & Engineering, KSIT, Bengaluru, India[4-1]

Abstract: Customer acquisition and retention is a key concern forseveral industries and is particularly acute in fiercely competitive and fast-growth businesses. Retaining a loyal customer is far more important than acquiring a new one, thus making customer churn one of the critical concerns for big corporations. Finding factors triggering customer churn is vital to implement necessary remediation to pre-empt and cut back this churn. This research focuses on implementing machinelearning (ML) algorithms to identify potential churn customers, categorize them based on usage patterns,and visualize the analysis results. Results show that Decision Tree Classifier, Random Forest Algorithm, and Support Vector Machine have the best churn modeling performance, particularly for 80:20 dataset distribution with average AUC scores of 0.843, 0.787 and 0.735 respectively and low false negatives. The research demonstrates that ML algorithms can successfully predict potential customer churn and helpin devising customer retention programmers.

Keywords: Telecom Industry, Customer churn, programming, understanding.

I. INTRODUCTION

The Telecom industry is constantly evolving and innovating. Due to the ever-increasing competition among corporations, increased importance is being given to targeted promoting methods for customer churn management. Modern-day customersexpect the best services at affordable rates. In case they are not satisfied, they quickly switch to another telecom network. Companies must find innovative ways to predict potential customer churn in order to prosper in such a competitive market. Customer churn is defined as the proportion of customers who stopped using a particular company's products or services during a definite time frame.
Mathematically,

$$C(T) = \frac{A(T)}{B(T)} \times 100$$

Where,

C represents the churn % for a time frame T,
A(T) represents the total number of customers after time T,
B(T) represents the total number of customers before time T

Few statistical studies have shown the impact of customer churn on the industry. Research [1] showed that a 1% increase in customer retention strategies equipped with adequate increase in corresponding budget may decrease the churn rate by up to 5% and lead to vast increment in revenues earned.

Research [2] showed that cost of retaining a loyal customer is much lesser than that of acquiring a new one. Furthermore, once a company starts losing its customers, the rapid decline in revenues makes it financially infeasible to initiate new retention programs. Hence, predictive analysis of customer retention is an absolute necessity inall businesses, especially in the case of the Telecommunications Industry.

Machine learning algorithms provide the bestsolution to predictive analysis of customer churn. Companies can use such ML pipelines to initiate retention strategies on those customers who are classified as likely targets of churn.

In this research, multiple classification algorithms are employed and theirresults are compared to discover the most accurate algorithm for the prediction of customer churn in businesses. The various algorithms used include Logistic Regression, Gaussian Naive Bayes, Random Forest Classifier, Decision Tree Classifier, and Support Vector Machine. Also, multiple evaluation parametric are used to determine the best overall performing model for the prediction of churn at the earliest stages.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.12 ⁜ Vol. 10, Issue 2, February 2023

DOI: 10.17148/IARJSET.2023.10209

II. RELATED WORK

In this section, we review works that make use of gamification in education and how it affects learning and engagement in the e-learning environment by researchers. Customer churn or attrition refers to the percentage of customers who will discontinue with a company's service during a given timeframe.Churn rate is calculated by dividing the number of customers a company lost over a given period of time by the number of retained customer at the beginning of that time period. Churn prediction is a key predictor of the long term success or failure of aBusiness.

In this research, machine learning and deep learning techniques are explored in order to predict telecom customer churn. Ubiquitous techniques like Random Forest Classifiers and SVMs are compared with relatively newer architectures like XGBoost and Deep Neural networks to classify if a customer will churn or not. The efficiency of these models are further explored by passing them through a grid search. From this experiment, it could be inferred the Random Forest model works best for this particular use case with a prediction accuracy of 90.96% on the testing data before grid search.

In current days, the customers are getting more attracted towards the quality of service (QoS) provided by the organizations. However, the current era is evidencing higher competition in providing technologically advanced QoS to the customers. Nevertheless, efficient customer relationship management systems can be advantageous for the organization for gaining more customers, maintaining customer relationships and improve customer retention by adding more profit tothe organizational business. Furthermore, the machine learning models such as support vector machine algorithms can add more value to the customer retention strategies.

The number of service providers are beingincreased very rapidly in every business. In these days, there is no shortage of options for customers in the banking sector when choosing where to put their money. As a result, customer churn and engagement has become one of the top issues for most of the banks. In this paper, a method to predicts the customer churn in a Bank, using machine learning techniques, which is a branch of artificial intelligence is proposed.

The research promotes the exploration of the likelihood of churn by analyzing customer behavior. The KNN, SVM, Decision Tree, and Random Forest classifiers areused in this study. Also, some feature selection methodshave been done to find the more relevant features and toverify system performance. The experimentation was conducted on the churn modeling dataset from Kaggle. The results are compared to find an appropriate model with higher precision and predictability. As a result, theuse of the Random Forest model after oversampling is better compared to other models in terms of accuracy.

Customer churn analysis and prediction insector is an issue now a days because it's very important for telecommunication industries to analyze the behaviors of various customer to predict which customers are about to leave the subscription from telecom company. So data mining techniques and algorithm plays an important role for companies in today's commercial conditions because gaining a new customer's cost is more than retaining the existing ones.In this paper, we can focus on various machine learning techniques for predicting customer churn through which we can build the classification models such as Logistic Regression, SVM, Random Forest, and Gradient boosted tree and also compare the performance of these models.

III. GOALS AND OBJECTIVES

Churn analysis often reveals patterns that indicate common motivators for customers to leave you, such asprice sensitivity or poor product adoption. Improving customer experience comes with a constant understanding of customer expectations andmeeting their needs. Churn analysis involves analyzing historical customer data to make churn prediction possible. Churn analysis is beneficial at all times, but evenmore so in a downturn or recession.

Churn prediction means detecting which customers are likely to leave a service or to cancel a subscription to a service. It is a critical prediction for many businesses because acquiring new clients often costs more than retaining existing ones. Once you can identify those customers that are at risk of canceling, you should know exactly what marketing action to take for each individual customer to maximize the chances that the customer will remain.

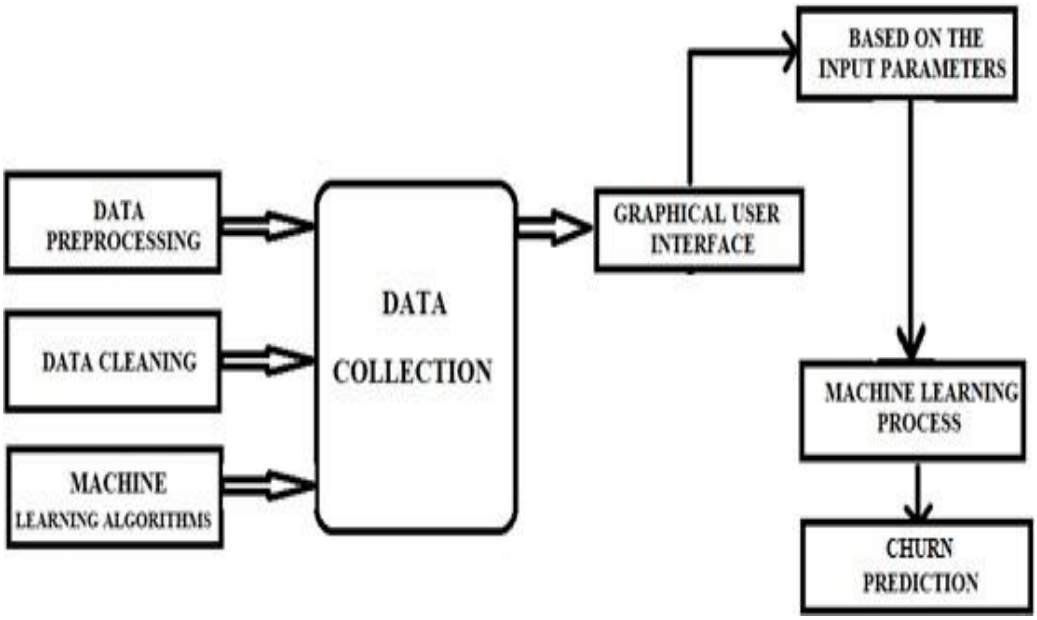
IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.12 ⁜ Vol. 10, Issue 2, February 2023

DOI: 10.17148/IARJSET.2023.10209

IV. METHODOLOGY



IMPORTING LIBRARIES

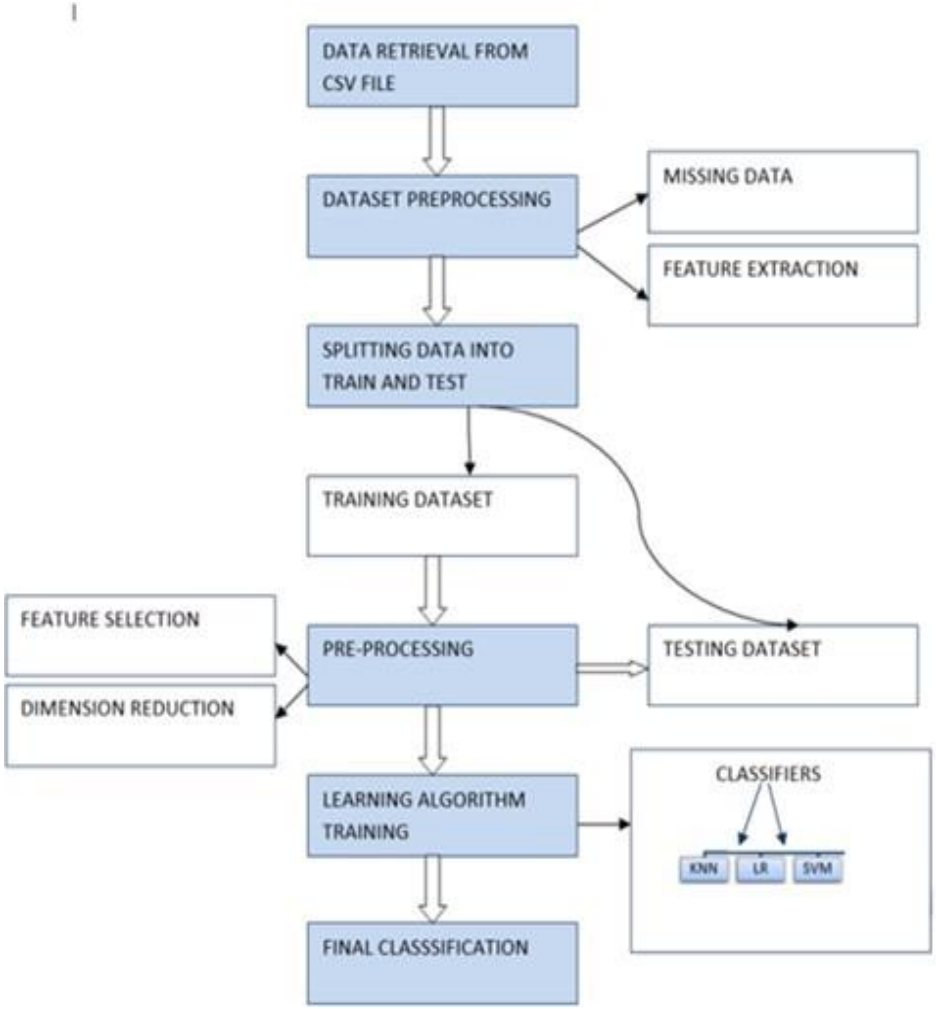
- ‘numpy’ is a Python module for scientific computing. This library will beutilized throughout the project and is imported as 'np’.
- ‘pandas’ is used to manipulate and analyse data. pandas is a BSD-licensed open-source library with basic data structuresand data analysis skills as pd, the pandas package is imported.
- matplotlib. pyplot Matplotlib offers acollection of command-style functions that allow it to operate similarly to MATLAB. It has the form of plt.
- ‘seaborn’ is a Python data visualization package for appealing and useful statistical visuals based on matplotlib. Data Pre-processing
- Data-Gathering – gathering the raw dataset from organized dataset providers ie.Kaggle.com or ucirepositry.com.Pre-processing is the term for the adjustments we make to our data before sending it to the algorithm, as seen in figure
- Data Preprocessing is a method for transforming messy data into a tidy collection.
- To put it another way, when data is collected from several sources, it is done so in araw form that prevents analysis. Performing a NaN check. Checking for NaN is critical during data pre-processing. We were only able to find a few NaNs on this try. Changing the value of NaN.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.12 ⁜ Vol. 10, Issue 2, February 2023

DOI: 10.17148/IARJSET.2023.10209



This may be accomplished by: removing the whole column having a largenumber of NaN values, Method of forward fillna, the Method of backward fillna,Using the mean technique data analysis is the process of dissecting, sanitizing, modifying, and modeling data with the aim of revealing relevant information, guiding deductions, and assisting in decision making. Data analysis has many different components and steps, including a wide variety of methods with different names that are applied in a number of business, scientific, and social science fields. Because it helps businesses to operate more efficiently and makemore scientific judgments, data analysis is essential in today's business environment. Feature Extraction Feature extraction is the process of converting raw data into numerical traits that may be used while keeping the specifics of the original data set.Compared to just applying machine learning to raw data, it produces superior outcomes. As a consequence, when training adataset, it is possible to quantify how much each feature lowers impurity. The greater an attribute's ability to eliminate impurity, the more significant it is. In random forests, the impurity decrease from each feature may be averaged across datasets to determine the variable's final significance. Train and Test dataset.It's time to fit the first machine learning model into your data once you've cleaned it up, visualized it, and learned more about it. Creating two sets of data: one for training and one for testing.

Training Dataset: A portion of thedata was used to fit the model. The test dataset is used to objectivelyassess the final model's fit to the training dataset. Prediction and Accuracy Stated machine learning algorithms are taught to forecast the customer's smart phone decision.The ability to forecast the customer's choice of smartphone is critical in helping smartphone makers improve their standards by observing what characteristics are important to customers when choosing a smartphone. Simply put, accuracy refers to how well yourmachine learning model predicts the proper class .

V. CONCLUSION

The significance of churn projection would assist many firms, ordinarily in telecom firms, to have a worthwhile profit also reap top gain taking. Customers in the churn projection. As the principal trouble in the telecom firm, so because of this, corporations are attempting to hold the already present ones from going away.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.12 ⁜ Vol. 10, Issue 2, February 2023

DOI: 10.17148/IARJSET.2023.10209

Alternatively than obtaining a new customer. Three tree-based algorithms Were chosen due to the fact of their applicability and variety in this kind of application.By using Random Forest, xgBoost, we will get an extra accurate evaluation of another algorithm. Here the use of the dataset of clients about their's carrier format Checks their values of them and has a unique projection, which would assist to perceive the users who will be going to migrate to different corporation facilities.

The telecom firm Can have a clear point of view and can grant them some exciting gives to remain in that service. The Obtained consequences exhibit proposes churn mannequin produces higher effects and performed Better. the way of the usage of laptops getting to know techniques. Random Forest produced higher accuracy among quite a number of strategies

REFERENCES

[1]. S. Babu, D. N. Ananthanarayanan, and V. Ramesh, ''A survey on factorsimpacting churn in telecommunication using datamininig techniques,'' Int.J. Eng. Res. Technol., vol. 3, no. 3, pp. 1745–1748, Mar. 2014

[2]. C. Geppert, ''Customer churn management: Retaining high-margin customers with customer relationship management techniques,'' KPMG &Associates Yarhands Dissou Arthur/Kwaku Ahenkrah/David Asamoah,2002

[3]. W. Verbeke, D. Martens, C. Mues, and B. Baesens, ''Building comprehensible customer churn prediction models with advanced rule inductiontechniques,'' Expert Syst. Appl., vol. 38, no. 3, pp. 2354–2364, Mar. 2011.

[4]. Y. Huang, B. Huang, and M.-T. Kechadi, ''A rule-based method for customer churn prediction in telecommunication services,'' in Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining. Berlin, Germany: Springer2011, pp. 411–422

[5]. A. Idris and A. Khan, ''Customer churn prediction for telecommunication: Employing various various features selection techniques and tree based ensemble classifiers,'' in Proc. 15th Int. Multitopic Conf., Dec. 2012.

[6]. M. Kaur, K. Singh, and N. Sharma, ''Data mining as a tool to predict the churn behaviour among Indian bank customers,'' Int. J. Recent Innov. Trends Comput. Commun., vol. 1, no. 9, pp. 720–725, Sep. 2013.

[7]. V. L. Miguéis, D. van den Poel, A. S. Camanho, and J. F. e Cunha,''Modeling partial customer churn: On the value of first product-category purchase sequences,'' Expert Syst. Appl., vol. 12, no. 12, pp. 11250–11256,Sep. 2012.

[8]. D. Manzano-Machob, ''The architecture of a churn prediction system based on stream mining,'' in Proc. Artif. Intell. Res. Develop., 16th Int.Conf. Catalan Assoc. Artif Intell., vol. 256, Oct. 2013, p. 157.

[9]. F. F. Reichheld and W. E. Sasser, Jr., ''Zero defections: Quality comes toservices,'' Harvard Bus. Rev., vol. 68, no. 5, pp. 105–111, 1990.