

Case Study on Customer Churn Predication in Telecom

RENUKA KURLE¹, KRANTI JADHAV², NUTAN RANE³, PRAJAKTA RANE⁴

^{1, 2, 3, 4} Computer Engg. Department, SSPM's College of Engineering, Kankavli

Abstract- *The Telecommunication Sector has risen to become one of the world's fastest-growing industries. It's an organization where the customer comes first, and as a result, client satisfaction is important to the success of businesses in this area. Because of this industry's global nature, consumers now have a plethora of options when it comes to receiving services. Consumer's decision to use a certain service provider is influenced by the pricing, flexibility, and customizability of the service. To address these demands, telecom companies work hard to establish policies and services that will entice customers and assist them acquire market position, however in current world, customer churn is a problem in telecom business, so it is vital for telecommunication companies to monitor the behaviors of different customers in order to predict which customers are going to terminate their subscriptions. Customers who are changing their service from current ones are termed as churners. Their could be many reasons for the churning. Researchers have been interested in predicting telecom churners, and several have worked on various algorithms to forecast telecom customer churn. Churn prediction is a crucial determinant of an organization's ultimate success or failure. As a result, There is an ever-increasing demand to forecast possible churners before they actually leave a service so that retention measures may be tailored to them and the company can grow by maximizing overall income.*

Indexed Terms- *Churn, Exploratory data analysis, univariate, biivariate, Random Forest, Decision tree*

I. INTRODUCTION

There are numerous Telecom organizations. Companies are now offering a variety of schemes and services in hopes of attracting clients to transfer from competitors' services to their own. This is because customers may be dissatisfied with existing programmes and services. As a result, the major

objective of those businesses is to keep their current consumers. As a result, such organisations must know ahead of time whose customers may migrate from their service to a rival organisation's service.

The Churn prediction's purpose is to forecast which customers will stop using a product or service. So that their unique organization's retention strategy can be tailored to them.

In Telecommunication sector, getting new consumers is considerably more expensive than keeping old ones. One of the most serious challenges for broadcast telecom providers is client retention. Churn is typically characterised as the attrition of subscribers as a consequence of them switching service providers, triggered by better interest rates or benefits or by the advantages offered at information sharing by a competitor, depreciating the shareholders' benefits and minimizing revenues to a substantial degree. So, for retention of existing customer companys need to implement various marketing strategy.

II. LITERATURE REVIEW

This chapter contains some of our previous research paper in the prediction domain.

- Anujkumar Tiwari, Reuben Sam, and Shakila Shaikh proposed a paper on "Analysis and prediction of churn customers for telecommunication industry" in which with the help of Apache PIG, their project was able to successfully evaluate and anticipate churn customers. Pig is a Hadoop engine that allows to run data flows in parallel. Pig Latin has been added as a data flow language. It offers operators for many common data operations, as well as the flexibility for users to create their own data reading, handling, and writing functions.
- Sanket Agrawal, Aditya Das, Amit Gaikwad, and Sudhir Dhage published a paper in which they make use of deep learning ideas in this. They used a multi-layered Artificial Neural Network that was

created expressly for this project to feed the data into it.

- Shrisha Bharadwaj, B.S. Anil, Abhiraj Pahargarh, Adhi- raj Pahargarh, P.S. Gowra, and Sharath Kumar proposed paper on” Customer churn prediction in mobile networks using logistic regression and multilayer perceptron(mlp).” They created a Logistic Regression model and an Artificial Neural Network model for predicting churn in the mobile telecoms industry as part of this study. These models are independent of other clients’ data and anti- pate customer attrition based on the client’s behaviour.
- Ngurah Putu Oka H and Ajib Setyo Arifin proposed a paper on in which Machine learning has been utilised to predict outcomes in this study. By looking at three different machine learning algorithms. Random Forest, XGBoost, and Deep Neural Network are the algorithms employed.
- Abhishek Gaur and Ratnesh Dubey proposed a paper on” Predicting customer churn prediction in telecom sector using various machine learning techniques.” In this paper, to train the machine learning model, they used Logistic Regression, SVM, Random Forest, and Gradient Boosted Tree.

III. PROBLEM STATEMENT

Client churn forecasting is a challenging but critical business challenge, especially in businesses where customer engagement costs are high, such as technology, telecommunications, and finance. The ability to recognize which customers are most likely to stop reporting but if there is still time to intervene then company will able to stop them from churning.

Customer churn is a worry in the telecom sector these days since it’s important for telecommunication companies to monitor the behaviours of various consumers in order to predict which customers are going to cancel their subscriptions. Churn forecasting is an important indicator of a company’s long-term accomplishment.

IV. METHODOLOGY

Machine learning will be used in this project to anticipate outcomes. By examining two machine

learning algorithms. Random Forest and decision tree classifier are the algorithms implemented.

The technique used in this study will be based on data provided by the company about consumer user records. In order to deal with this raw data, it must first be cleansed. To detect the variance in the data and apply strategies to normalise the data, a high-level analysis is necessary. Following that, characteristics will be chosen to become element of the feature set that will be leveraged to train a model. This information is then passed into the model that was created specifically for this scenario. After the model has been trained on the training data, it is confirmed using validation data and then tested on the testing set. This would result in a percent accuracy score that could be used to evaluate the model under consideration. A comparison graph will be created in order to determine the primary reasons that caused the churn. Numeric parameters will be subjected to qualitative analysis to see how useful they are in action recognition and overall churn trends.

A. Software Requirements

- Programming Languages: Python, HTML
- Windows: 10(64 bit)
- IDE or Editor: Spyder, Jupyter notebook

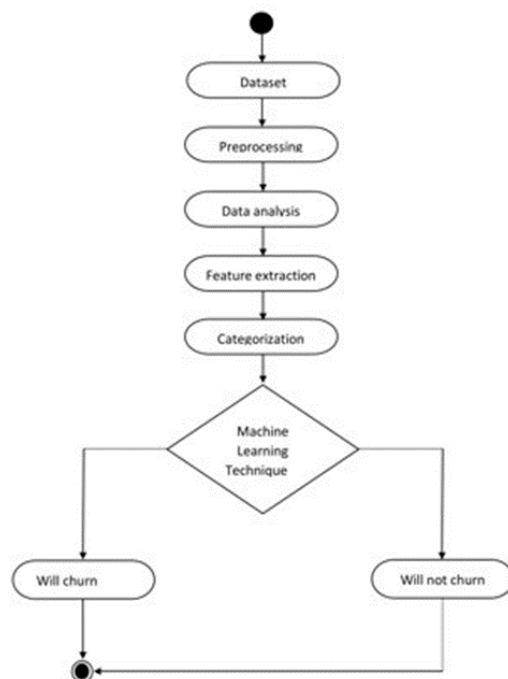


Fig. 1. Flow diagram for proposed methodology

B. Data pre-processing

After collection of data, numerous procedures are undertaken to determine the data. The objective of this step is to learn well about data structure, execute preliminary preprocessing, data cleaning, find insights and incomplete data such as tracking error, anomalies, and null values and develop and test conclusions.

In this process we came to know that data is highly imbalanced. There are 5174 non-churners and 1869 are churners. So, approximately 73.46% are non-churners and 26.53% are churners, this creates imbalance in dataset, this affects accuracy of our model. So, we have balanced this difference in model building part.

C. Exploratory data analysis

The main aim of EDA is to obtain confidence in a data to an extent where we are ready to engage a machine learning model.

Univariate analysis is done to deal with analyzing one feature at a time.

Bivariate analysis in which Data which has two variables, when we want to measure the relationship that exists between the two variables.

From this we have got some insights like

- The churn rate among older citizens is considerably higher than that of non-seniors.
- Customers with children and spouses may churn less frequently in order to maintain their family's services operating.
- The electronic check mode of payment has a substantially greater churn rate than other payment methods.
- Contract type-monthly churners are more likely to switch because of no contract terms, as they are free to go customers.
- If the contract is for a longer amount of time, the subscriber becomes less likely to switch.

D. Model Building

1) Decision Tree Classifier:

- The use of a decision tree as a classification algorithm to proceed from observations about an item, that is characteristic represented in the

branches, to inferences about the object's goal value, that is churn or will not churn.

- To create a prediction, this method divides a data sample into two or more undifferentiated sets based on the most relevant differentiation in model parameters. A component of a tree is formed with each split. As a result, a tree gets created, including decision nodes and leaf nodes. A tree begins with a root node, which is the most important determinant.
- Decision tree forecasting findings are simple to interpret and depict. Even someone without a background in analytics or data science can grasp how a particular result came to be. Decision trees require less data preparation than other algorithms, which is a benefit. However, if any slight data modifications are done, they may become unpredictable. To put it another way, data differences can result in trees that are significantly different. divergent trees being created.
- Decision Tree gives very low accuracy about 78% because of imbalanced dataset and after balancing the dataset it gives about 91% accuracy.

2) Random Forest Classifier:

- Random Forest is a classifier technique that uses the averaging of multiple decision trees on distinct subsets of a dataset to enhance prediction results. It employs modelling in the kind of a decision tree and afterwards allocates value to each tree. It can be used to deal imbalanced data.
- Initially, random forest gives about 80% accuracy because data in dataset is not in well proportion and after balancing the dataset by using smote analysis it gives 93% accuracy.

E. Model Balancing

- As we have seen earlier data is highly imbalanced with 73:26 ratio for non-churners and churners respectively.
- Due to imbalance, both the classifier has given low accuracy initially. So, to resolve this issue we have resampled the dataset using SMOTE analysis.
- We have used smote analysis by importing imblearn library.
- Smote analysis internally does up-sampling or down-sampling and balances the dataset.

F. Model Deployment

- Model deployment is done using flask. By using this, our model gets deploy to localhost.
- Flask is a Python-based web application framework. It features a number of modules that make writing web apps easier for web developers.
- Flask provides us with a number of options for constructing web apps, as well as the tools and libraries we'll need to get started.

V. RESULTS

Accuracy Table		
Classification Model	Before Balancing Dataset	After Balancing Dataset
Decision Tree classifier	78%	91%
Random Forest Classifier	80%	93%

- We obtained an accuracy of 93% with Random Forest Classifier and 91% with Decision Tree Classifier. So, it is noted that Random Forest classifier has given highest accuracy.
- So, we have continued with Random Forest Classifier as our model for churn predication.



Fig. 2. Output

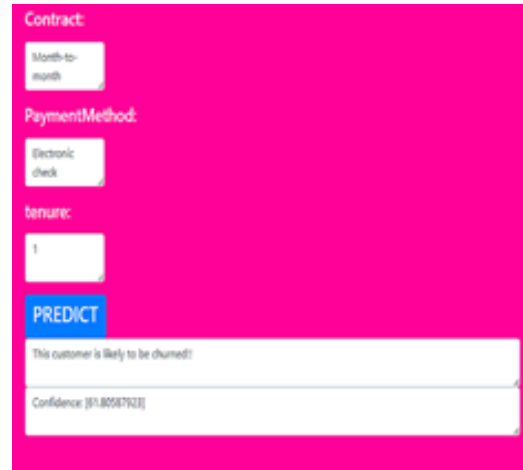


Fig. 3. Output

CONCLUSION

- Churn prediction is among the most significant retention methods employed in the telecom industry. It leads to enhanced cost allocation in customer service management activities, allowing income and profits to be retained in the future.
- Telecom companies must consider the motive for churn in order to maintain existing customers, which can be accomplished using knowledge gathered from Telecom data.
- This study becomes a valuable tool for businesses to choose which characteristics to focus on in order to keep clients and prevent losing them to rivals.
- Telecommunications companies should understand the reasons behind churn in effort to keep existing customers, which can be accomplished using knowledge gathered from Telecom data.
- The Customer's behaviour is directly linked to the fact that customers with long-term contracts or who have been associated with the company for a longer period of time are more likely to be kept than those with short-term contracts or limited tenures.

REFERENCES

[1] Anujkumar Tiwari, Reuben Sam, and Shakila Shaikh. Analysis and prediction of churn customers for telecommunication industry. In 2017 International Conference on I-SMAC (IoT

- in Social, Mobile, Analytics and Cloud) (I-SMAC), pages 218–222, 2017.
- [2] Sanket Agrawal, Aditya Das, Amit Gaikwad, and Sudhir Dhage. Customer churn prediction modelling based on behavioural patterns analysis using deep learning. In 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), pages 1–6, 2018.
- [3] Sanket Agrawal, Aditya Das, Amit Gaikwad, and Sudhir Dhage. Customer churn prediction modelling based on behavioural patterns analysis using deep learning. In 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), pages 1–6, 2018.
- [4] Ngurah Putu Oka H and Ajib Setyo Arifin. Telecommunication service subscriber churn likelihood prediction analysis using diverse machine learning model. In 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT), pages 24–29, 2020.
- [5] Abhishek Gaur and Ratnesh Dubey. Predicting customer churn prediction in telecom sector using various machine learning techniques. In 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), pages 1–5, 2018