



TELECOM CUSTOMER CHURN PREDICTION SYSTEM

Mr. K.S. Chandrasekaran^{*1},

D. Abinandhan^{*2}, G. Arun Kumar^{*3}, R. Dhanush Kumar^{*4}, K. Kumaravel^{*5}

¹ Associate Professor, Computer Science and Engineering, Saranathan College of Engineering, Trichy, Tamil Nadu, India

²⁻⁵ Student, Dept of Computer Science and Engineering, Saranathan College of Engineering, Trichy, Tamil Nadu, India

Abstract: Telecom industry has gained a huge growth in the last two decades. Because of the availability of a lot of options, many telecom companies are facing the problem of customer churn in the recent years. Because of the advancement of and indispensable need for internet, customers can easily change from one company to another. It may affect the profits of the Telecom companies if they don't pay enough attention for customer churn. To pay attention to the customer churn, the Telecom companies should be able to predict which customers are likely to leave the company. Manually predicting this is almost impossible. With the help of machine learning algorithms, we can try to predict the customers who could possibly switch over. Once the companies know if a customer is going to churn, they can try to retain those customers through various strategies.

Keywords: Random Forest, Decision Tree, SMOTEENN, ML, Churn

I. INTRODUCTION

The loss of customers of a company is referred to as churn. Recent researches state that gaining a new customer is difficult and expensive than retaining an already existing customer. With the arrival of various new telecom companies, customers are having a lot of choices. Each company is trying to outrun each other by providing aggressive plans. Because of this aggressive competition, companies lose their customers to their competitors.

The aim of the customer churn prediction system is to predict the clients who are having a higher probability of switching over to another company. A recent study shows that in the telecom industry, the monthly customer churn percentage is 2.2% and the yearly customer churn percentage was 27%. Thus, the problem of customer churn has become unavoidable for the telecom companies.

Out of many approaches available, Machine learning algorithms have been used to predict customer churn. Machine Learning involves giving historical data as input, training the model with that data and getting output by providing new data. We have tried five machine learning algorithms and compared the accuracy scores of these algorithms and used the algorithm that has the best accuracy score.

A sampling technique called SMOTEENN, a combination of SMOTE and ENN techniques, is used to balance the dataset. Synthetic Minority Oversampling Technique (SMOTE) is an up sampling technique. Edited Nearest Neighbour (ENN) is a down sampling technique. SMOTEENN combines both these sampling techniques to balance the data in an efficient way.

II. LITERARY WORK

Existing work: In the existing work, python programming has been used for customer churn prediction. The machine learning algorithms used in the existing system are Decision tree and Logistic regression. The model predicted churn with an accuracy 80%. [9]

The authors in this paper pointed out that in order to tackle customer churn, a large number of machine learning (ML) based churn prediction algorithms have been developed in the recently. There are different classification techniques which have been used in studies for model building in churn prediction field, which is ranging from linear classifiers to the state of the art machine learning including boosting and SVM. They have used Support Vector Machine (SVM) algorithm for predicting churn of telecom customers. The system achieved an accuracy percentage of 88%. [2]

The authors of this paper studied the ML applications and suggested the optimal solutions for telecom industry. The authors used Naïve Bayes algorithm for predicting churn. The model achieved an accuracy percentage of 87%. [3]



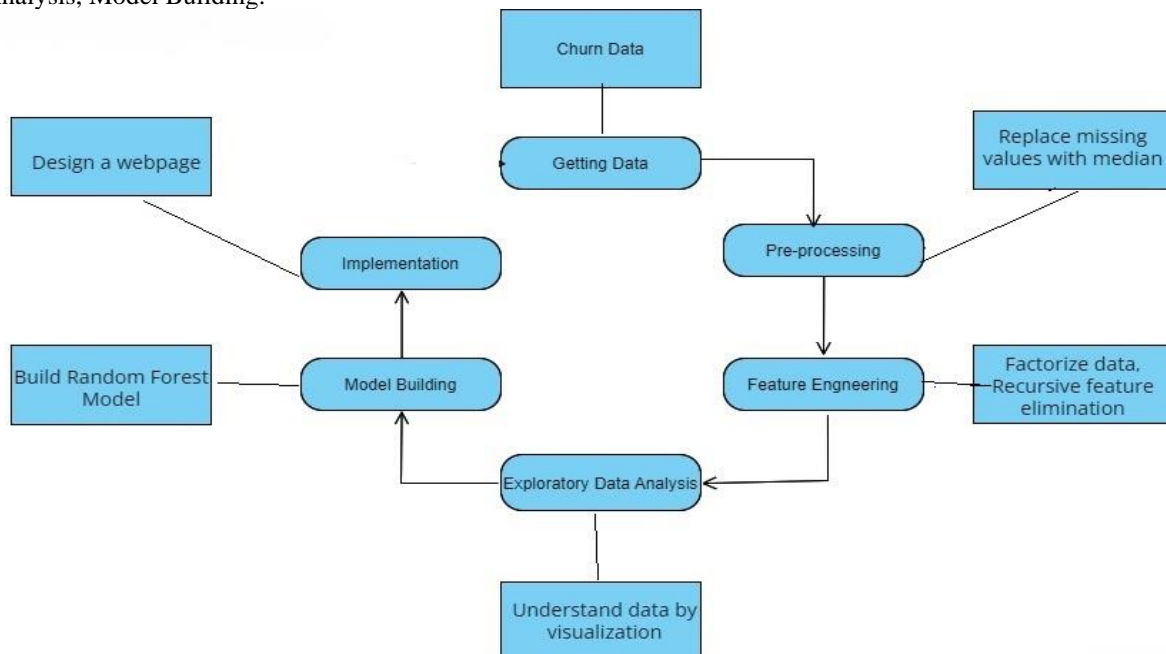
In this paper, the author expresses that churn could be considerably lessened by investigating the earlier history of the potential users analytically. The model has been developed using algorithms such as logistic regression and the neural network. The author also used C5.0 model of the decision tree for predicting churn. The system with the above mentioned model achieved an accuracy of 85 percentage. [1]

The author of this paper expressed that churn prediction models have developed as one of the most critical Business Intelligence (BI) applications which intends at finding customers who are likely to switch to other telecom companies. The author has employed Exhaustive CHAID, a derivation of Decision Tree algorithm. The study managed the issue of data imbalance effectively. The overall accuracy of the model is 75.4 percentage. [6]

Here the authors expressed that customer churn essentially has financial consequences on a firm, since it bears five to ten times much to bring in a new user than retaining an old user within the firm. The Authors tried out K-Nearest Neighbour and Logistic Regression algorithms, compared the accuracy of both the models and used the model with the best accuracy. The model with KNN had an accuracy percentage of 69% and the model with Logistic Regression had an accuracy percentage of 80%. [4]

III. PROPOSED WORK

The proposed work consists of five modules: Collection of Data, Pre-processing, Feature Engineering, Exploratory Data Analysis, Model Building.



Collection of Data

The model uses dataset collected from Kaggle website for predicting churn. The dataset consists of 7043 rows and 21 columns. Of these 70% data will be used for training and 30% data will be used for testing using the train-test split technique.

Pre-processing

In Machine learning, we use Data preprocessing to convert the raw data in the dataset into the clean data. Data preprocessing includes filling the empty data present in the dataset with median values, removing irrelevant features present in the dataset (Example: Customer ID), finding out the important features to train a model with high performance.

Feature Engineering

Feature engineering is a machine learning technique that leverages data to create new variables that were not present initially in the training set. Feature engineering is an important step in machine learning that helps to improve the accuracy of the model, by extracting features that will be most useful predictor variables for the model from raw data. Feature Engineering includes converting categorical data into numerical data; simplifying and speeding up processing and Recursive Feature Elimination. Recursive Feature Elimination is a feature selection algorithm that removes the features that are present repetitively in the dataset.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a method to analyse the data in the dataset and gain insights from their main characteristics. The dataset contains categorical features, numerical features and a predictive feature (Churn). It is not at



all easy to look at a whole spreadsheet and determine important characteristics of the data. So, in those cases EDA plays a major role. When analysing the dataset, the dataset was found to be imbalanced. The ratio of churners to non-churners is 27:73. Some of the insights gained after EDA are, 1. both genders have almost equal number of churners, so there is no significance there, 2. Senior Citizens tend to churn less and younger citizens churn much higher, customers with no partners, and customers with no dependents. 3. Churn decreases with increase in tenure, 4. Churn increases with increase in Monthly Charges, 5. When the customer contract period is shorter churn rate becomes higher, 6. Clients who use electronic checks to pay their bills are more likely to churn.

Model Building

Random Forest Algorithm is used to build the model to predict churn. Random forest algorithm compares the predictions of multiple decision trees and returns the answer that was predicted by majority decision trees. Since the dataset is imbalanced, a sampling technique called SMOTEENN is used to balance the dataset. SMOTE-ENN combines the SMOTE ability to generate synthetic examples for minority class and ENN ability to delete some observations from the majority class. In simpler words, SMOTEENN is a combination of both up sampling and down sampling. After using SMOTEENN, the ratio of churners to non-churners is 49:51.

IV.IMPLEMENTATION

The system is implemented as a webpage using an open-source python library called Streamlit. Streamlit helps in providing an interface for machine learning models without using regular frontend technologies like HTML and CSS. The webpage will have two operational methods: Online prediction and Batch prediction.

Online prediction

Here, the user needs to type the input values for all the 16 fields manually. This mode can be used when the user needs to predict the churn probability of a single customer.

Batch prediction

This use is for generating predictions for a set of observations instantaneously. Here, the user needs to provide a csv (Comma Separated Values) file as input. The webpage will return output in a batch i.e. If data of 20 customers are present in the file, the application will return the churn probability of all the 20 customers.

V.CONCLUSION

In nowadays highly competition between firms and in digital world the factors of customer churn are an essential undertaking for every service provider to make profitable and long-term relation with particular customers. Some customers are much cost sensitive and move to other telecom service provider if they get better cost and also, they chose the service provider chosen by their friends and family. The cost of obtaining new customers can be greater than that of customer retention. Thus, by implementing a system to predict customer churn, the work helps telecom companies to identify customers who are more likely to churn. The companies will be able to reduce churn rate through various retention strategies targeted specifically at the probable churners.

REFERENCES

1. Alwin, P.K.D.N.M., Kumara, B.T.G.S., Hapuarachchi, H.A.C.S. (2018). Customer Churn Analysis and Prediction in Telecommunication for Decision Making. 2018 International Conference on Business Innovation (ICOBI), pp. 40- 45.
2. Azeem, Muhammad, Muhammad Usman and Alvis Cheuk M. Fong. "A churn prediction model for prepaid customers in telecom using fuzzy classifiers." *Telecommunication Systems* 66 (2017): 603-614.
3. Ebrah, Khulood and Selma Elnasir. "Churn Prediction Using Machine Learning and Recommendations Plans for Telecoms." *Journal of Computer and Communications* (2019).
4. Joolfoo, Muhammad & Jugurnauth, Rameshwar & Joolfoo, Khalid. (2020). Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform. *Journal of Critical Reviews*.
5. Qureshi, S.A., Rehman, A.S., Qamar, A.M., Kamal, A. (2013). Telecommunication Subscribers' Churn Prediction Model Using Machine Learning. *IEEE International Conference on Digital Information Management (ICDIM)*, Pakistan
6. Saini, N, Monika, S, Garg, K. (2017). Churn Prediction in Telecommunication Industry using Decision Tree. *International Journal of Engineering Research & Technology (IJERT)*, 6(4), pp. 439-443.
7. Talwar, K, Dahiya, K. (2015). Customer Churn Prediction in Telecommunication Industries using Data Mining Techniques- A Review. *International Journal of Advanced Research in Computer Science and Software Engineering*.
8. Tanneedi, N.N.P.P. (2016). Customer Churn Prediction Using Big Data Analytics. Master Thesis, Blekinge Institute of Technology.
9. Venkatesh Hariharan, Irshad Khan, Pravin Katkade, Prof.Sejal D' mello. (2020). Customer Churn analysis in Telecom Industry. *International Research Journal of Engineering and Technology (IRJET)*.