JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Customers Churn Prediction using Machine Learning and Deep Learning

Amol Chole[1], Trupti Mate[2], Rushikesh Chavan[3], Shradha Mandan[4]

Department of E&TC, SKNCOE, SPPU, Pune

Abstract— **Customer churn is a major problem and one of the most important concerns for large companies, especially in the telecom field, companies are seeking to develop means to predict potential customer to churn. The main contribution of our work is to develop a churn prediction model which assists telecom operators to predict customers who are most likely subject to churn. The model developed in this work uses machine learning and Deep Learning techniques on big data platform and builds a model that predicts the Customer Churn. The model was prepared and tested by working on a large dataset created by transforming big raw data provided by GitHub. The dataset contained all customers information, and was used to train, test, and evaluate the system. The model experimented four algorithms: Random Forest, Support Vector Machine, K-Nearest Neighbour and Convolutional Neural Network. However, the best results were obtained by applying Random Forest algorithm. The Big Data used from GitHub contains 56000 user's data from different Telecom Sectors.**

***Keywords*— RF (Random Forest), SVM (Support Vector Machine), KNN (K-nearest Neighbour's), CNN (Convolutional Neural Networks).**

I. INTRODUCTION

Telecommunications is known as one of the fast-growing industries in many countries. However, the average annual churn rate of the telecom industry is between 20–40%, which leads to huge loss of revenue. Customers always have a variety of choices, and they tend to choose the companies that can offer them better quality and less expensive services. To survive in this highly competitive market, telecom companies have to develop strategies in attracting new customers or increasing the customer retention rate. It is noted that acquiring a new user costs 5 to 10 times more than retaining an existing one. Therefore, customer churn prediction has become a popular research area since it has the potential to help telecom companies to identify customers with high potential of terminating their contracts. Then, companies evaluate the situations and design appropriate service packages for these customers to retain them. It is extremely important for the company's long-term development. Many previous works have confirmed that customer churn prediction depended on complex human behaviours, such as customer data, usage, consumption behaviour, and so on. Typical a churn prediction model involves four steps: data gathering and understanding, features selection, design and development of predictive model, validate and evaluate the model. Most previous studies focused on improvements of predictive models using machine learning algorithms such as neural networks, support vector machine, random forest, etc. However, prediction results are not highly accurate due to incorrect classifications of churners and non-churners. In this work, we proposed an integrated framework for churn prediction problem using a Random Forest as the predictive model. Our research is based on big data which is taken from GitHub, it contains 56000 user's data with around 25 variables.

II. METHODOLOGY

A. Flow Chart

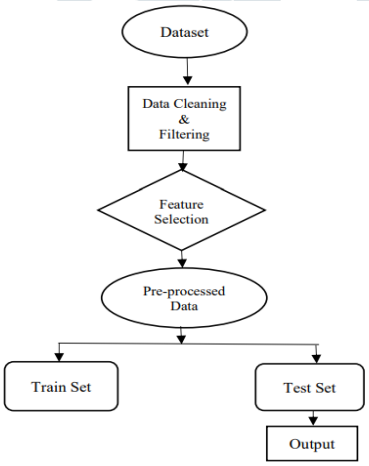

Fig 1. Flow Chart

B. Block Diagram

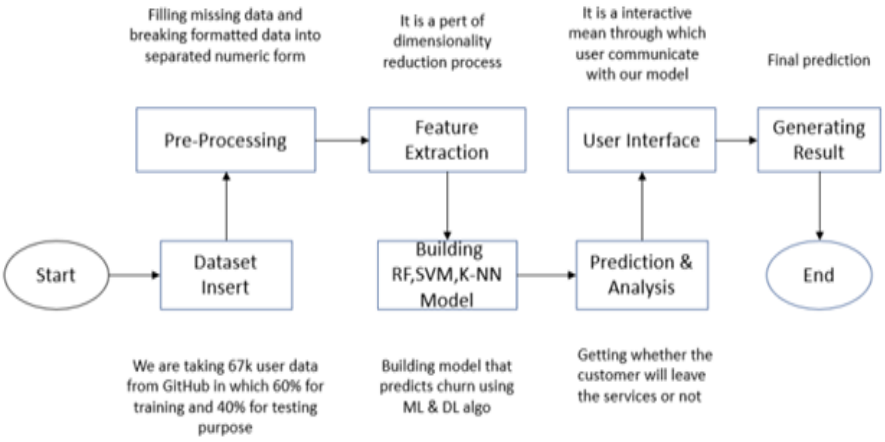

Fig 2. Block Diagram of proposed system

Data Pre-processing: The technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models. Here, we clean and organize the churn data which is present in the dataset we have.

Processed Data: Processed data is the data after pre-processing in which raw data is removed, data that is not needed, that that has empty cells, and also the data that in not necessary while building the module for churn prediction and factors identification.

Machine Learning Techniques

1. Support Vector Machine: Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression.

   Step-1: Import the dataset

   Step-2: Explore the data to figure out what they look like

   Step-3: Pre-process the data

   Step-4: Split the data into attributes and labels

   Step-5: Divide the data into training and testing sets

   Step-6: Train the SVM algorithm

   Step-7: Make some predictions

   Step-8: Evaluate the results of the algorithm

2. Random Forest:

   Step-1: Select random K data points from the training set.

   Step-2: Build the decision trees associated with the selected data points (Subsets).

   Step-3: Choose the number N for decision trees that you want to build.

   Step-4: Repeat Step 1 & 2. 3.

3. K-nearest neighbor:

   Step 1 − For implementing any algorithm, we need dataset. So during the first step of KNN, we must load the training as well as test data.

   Step 2 − Next, we need to choose the value of K i.e. the nearest data points.

   Step 3 − For each point in the test data do the following.

   Step 4 − End.

4. Convolutional Neural Network:

   A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and Tasks that involve the processing of pixel data. There are other types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice.

Prediction of Churn & Non-Churn Customers:

At the end of this the modules that are formed with these algorithms will be able to predict the customers that are going to leave the services means Churn Customers and those are staying and not leaving the service means Non-Churn cutomers.

Factor Identification:

Then the factors that are affecting the churning in this process of finding the module that will find the churn customers can be taken as Speed, Network, Smooth working, etc.

System Required

1. Hardware Specification HP intel 7(8gb ram)
2. Software Specification
   Python – to build model
   html – index page, prediction page
   Flask – User Interface
3. Dataset -56000 user dataset from GitHub

III. EXPERIMENTATION

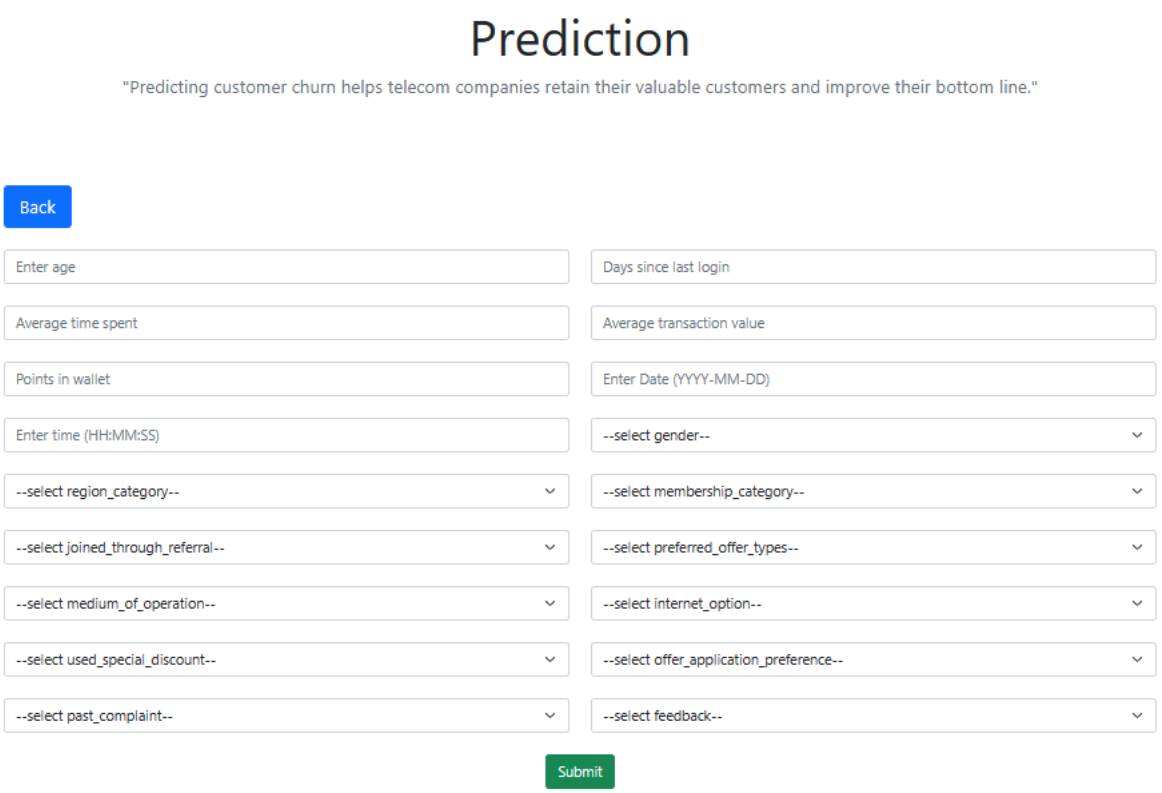

Fig 3. Main Page (Prediction Page)

IV. RESULTS

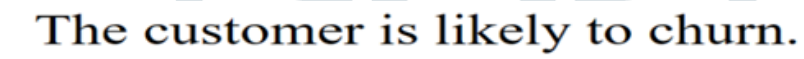

Fig 4. Churner Prediction

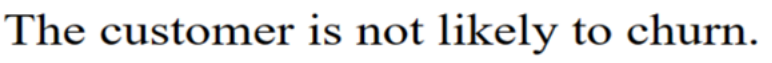

Fig 5. Non-churner's Prediction

```
# make predictions
y_pred = model.predict(x_test)


# checking training accuracy
print("training accuracy is : ", rf.score(x_train, y_train)*100)

training accuracy is :  100.0

# checking accuracy of test dataset
print("testing accuracy is : ", rf.score(x_test, y_test)*100)

testing accuracy is :  83.11407460803748
```

Fig 6. Prediction Accuracy

V.CONCLUSIONS

This study aims to minimize the number of misclassifications in churn prediction by developing the Random Forest models. Several algorithms like Support Vector Machine, K-Nearest Neighbours, Convolutional Neural Network are involved in these techniques including data cleaning, filtering, feature selection and pre-processing, based on results obtained Random Forest has given 83.11% of accuracy. Our future research will be to expand the interval to search for better values of the hyper parameters. Furthermore, different machine learning algorithms can be applied in feature selections and resampling data.

ACKNOWLEDGMENT

It is a great pleasure to present the report on "Comparative Analysis of Churn Prediction and Factor Identification of Subscribers Using ML & DL Technique". We would like to convey our gratitude to our respected principal sir Dr. A.V. Deshpande, who have provided all the facilities to us. We would like to thanks our respected Head of Department Dr. S.K. Jagtap, Department of Electronics and Telecommunication for giving us support and suggestions during our project. With our deep sense of gratitude, we would like to thank our respected project guide Mrs. T.A. Mate and our respected project coordinator Mr. P.S. Kokare and Ms. M. M. Sonkhaskar (Department of Electronics and Telecommunication) for their guidance, support, valuable time and encouragement during the project. We also wish to thank all the teaching and non-teaching staff members of the Department of Electronic and Telecommunication Engineering for their valuable suggestions and support and co-operation during project.

REFERENCES

[1] Zhang, Kun Liu, William Wang, Tai Zhang and Jie Lu, "*A personalised recommender system for Telecom products and services, Zui*", Volume 5, Issue 4th March 2021

[2] Dr. M. Balasubramaniam, M. Selvarani, "*Churn Prediction in Mobile Telecom System Using Data Mining Techniques*", International Journal of Scientific and Research Publications, Volume 4, Issue 4, April 2020.

[3] Saad Ahmed Qureshi, Ammar Saleem Rehman, Ali Mustafa Qamar, Aatif Kamal, Ahsan Rehman, "*Telecommunication Subscribers' Churn Prediction Model Using Machine Learning*", IEEE International Conference on Digital Information Management (ICDIM), 2013 8th International Conference on, Volume 4, Issue 5, April 2016.

[4] Georges D. Olle and Shuqin Cai, "*A Hybrid Churn Predictio Model in Mobile Telecommunication Industry*", International Journal of e-Education, e-Business, eManagement and e-Learning, Volume 4, Issue 1, February 2015

[5] Kiran Dahiya and Kanika Talwar*," Customer Churn Prediction in Telecommunication Industries using Data Mining Techniques- A Review*", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, March 2015.

[6] Nisha Saini and Monika," *Churn Prediction in Telecommunication Using Classification Techniques Based on Data Mining: A Survey*,", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 3, March 2015.

[7] Hu X, Yang Y, Chen L, Zhu S. Research on "*A Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network*", 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA). 2020

[8] anjasuchat M, Limpiyakorn Y "*Applying Reinforcement Learning for Customer Churn Prediction*", Journal of Physics: Conference Series. 2020; 1619:012016. https://doi.org/10.1088/1742-6596/1619/1/ 012016