CUSTOMER CHURN PREDICTION

A PROJECT REPORT

submitted By

AKHIL S S

TVE21MCA-2011

to

the APJ Abdul Kalam Technological University
in partial fullfilment of the requirements for the award of the degree

of

Master of Computer Applications

Department of Computer Applications

College of Engineering

Trivandrum-695016

NOVEMBER 2022

Declaration

I undersigned hereby declare that the project report titled **"CUSTOMER CHURN PREDICTION using Machine Learning"** submitted for partial fulfillment of the requirements for the award of degree of Master of Computer Applications of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Smt Deepa S S, Head of Department , Associate Professor. This submission represents my ideas in my words and where ideas or words of others have been included. I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity as directed in the ethics policy of the college and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and/or University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title.

Place : Trivandrum **AKHIL S S**

Date : 21/11/2022

DEPARTMENT OF COMPUTER APPLICATIONS

COLLEGE OF ENGINEERING
TRIVANDRUM

CERTIFICATE

This is to certify that the report entitled **CUSTOMER CHURN PREDICTION using machine learning** submitted by **AKHIL S S** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of the project work carried out by him under my guidance and supervision. This report in any form has not been submitted to any University or Institute for any purpose.

Internal Supervisor External Supervisor

Head of the Dept

Acknowledgement

First and for most I thank **GOD** almighty and to my parents for the success of this project. I owe a sincere gratitude and heart full thanks to everyone who shared their precious time and knowledge for the successful completion of my project.

I am extremely thankful to **Dr Suresh Babu V**, Principal, College of Engineering Trivandrum for providing me with the best facilities and atmosphere which was necessary for the successful completion of this project.

I am extremely grateful to **Dr. Deepa S S**, HOD, Dept of Computer Applications, for providing me with best facilities and atmosphere for the creative work guidance and encouragement.

I express our sincere thanks to **Smt. Deepa S S**, Asst. Professor, Department of Computer Applications, College of Engineering Trivandrum for her valuable guidance, support and advice that aided in the successful completion of my project.

I profusely thank other Asst. Professors in the department and all other staffs of CET, for their guidance and inspirations throughout my course of study.

I owe my thanks to my friends and all others who have directly or indirectly helped me in the successful completion of this project. No words can express my humble gratitude to my beloved parents and relatives who have been guiding me in all walks of my journey.

AKHIL S S

Abstract

Customers are the base for any business success and that is why firms become aware of the significance of acquiring satisfaction of customers. Customer churn is an essential issue and it is regarded as one of the most essential concerns among firms because of increasing rivalry among firms, increased significance of marketing strategies and customers conscious behaviour in present years. Organizations must develop different strategies to resolve the churn issues relying on the services they offer. Customer churn practice is essential in competitive and rapidly developing in telecom sector. The process of migrating from one service provider to another telecom service provider occurs due to good services or rates or due to various advantages which the rivalry firm provides customers when signing up. Due to the greater cost related with acquiring new customers the prediction of customer churn has developed as an indispensable part of planning process and strategic decision making in telecom sector. The main aim of the study is to explore the customer churn prediction in telecom using machine learning model . Machine learning techniques have been used for estimating the customer probability to churn. This study makes use of Random Forest Classifier,XGboost and Support vector Machine for predicting consumer churn in the telecom sector. This study uses Kaggle website for dataset in predicting and analyzing churn. The results of the study show that the accuracy rate of prediction in consumer churn is found to be 0.86 percent

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Customer churning refers to the action of when a customer chooses to abandon their service provider . The term is relatively new and has gained more relevance with the emergence of online services. Firms across the globe recognize customer churning as a great loss since they have already invested in attracting these customers. This is one of the major reasons that customer retention is beneficial for a firm. Customers can churn for many reasons and it is hard to pinpoint a general reason for churning. The availability of information has given consumers a bargaining power, and nowadays customers can easily find the service provider, which provides the same product with a more satisfying deal . To manage this, firms invest in customer churn prediction, which means that companies try to predict which of their customers will churn, so that they can apply preventative measures. These preventive measures could differ depending on the reason a customer might churn, and could be for example, offering a lower price or including an extra service. As mentioned earlier, analyzing customer behavior serves as the basis for predicting customers who might churn, which is important for many reasons. One reason is that for companies who rely on subscription-based income, it can make a big difference on whether they can keep a steady income level or if they need to make changes to their services to keep customers. Another reason is that, compared to retaining customers, attracting new ones is costlier and firms can save money by retaining their existing customer base.

Chapter 2

Problem Defnition and Objectives

The retention and acquisition of users are the major concerns in telecom industry. The fast growth of marketplace in every business is giving rise to increased subscriber base. Accordingly, companies have recognized the significance of retaining the customers who is on hand. It has become necessary for service-providers to reduce the churn rate of customers since the inattention might negatively influence profitability of the company. Churn prediction contributes to identify those users who are likely to switch a company over another. Telecom is enduring the problem of ever-increasing churn rate. Accordingly, the current study employs machine learning algorithm on big-data platform. Machine learning algorithm techniques facilitate these telecom firms to be protected with efficient approaches for lessening the rate of churn. Silent churn is one type which is considered complicated to predict since there might have such kind of users who might probably churns in the near future. It must be the aim of the decision-maker and advertisers to lessen the churn ratio since it is a recognized fact that comparatively

The Major Objectives Of this Project are:

- To explore the customer churn prediction in telecom using machine learning
- To investigate the impact of customer churn in telecom industry as a whole.
- To discuss the significance of customer churn models in telecom industry.
- To compare the algorithms that are effective in reducing churn rate in telecom companies.

Chapter 3

Literature Review

A large number of researches in the subject of churn prediction are being investigated employing various statistical and machine learning algorithms since a decade. This chapter deals with the recent and most important publications on churn prediction in telecom industry in the recent period

3.1 Impact Of customer churn in Telecom Industry

Tanneedi, (2016) pointed out that customer churn has become a dreadful problem forthe telecom industry since customers never have a second thought to leave if they don't exactly get what they are expecting. There is no benchmark model that deals with the churning issues of telecom companies precisely. The study emphasized that Big Data analytics with machine learning are considered effective as means for identifying churn. The current study makes an effort to predict customer churn in telecom employing Big Data analytics. Statistical analyses and machine learning application such as Decision trees (DT) have been used for three different datasets. From the analytics of DT, decision trees with accuracy rate of 52sources correspondingly. The findings pointed out that the more the quality and volume increases, the lesser the annoyance and possibility of churn can be expected in telecom industry.

Almana, et al (2014) pointed out that the most complicated problem gone through by telecom sector is customer churn. To effectively deal with the churn prediction challenge, the current study has used machine-learning algorithms along with data mining tools. An effort in retaining existing customers could result in a considerable growth in profits as well as revenues. The necessity of retaining old customers pines for precise prediction of customer churn algorithms

which are both precise and understandable. The study figured out certain factors that impact customer to churn. Comparative to post-paid customers, the prepaid users are not significantly bound by service contracts and they are more likely to churn even for simplest reasons, which make it quite complicated to predict their churn rate. Customer loyalty is another factor which can be defined by quality of service and customer service delivered by the service providers. Issues such as network coverage might impact customers to switch another network. Other factors which intensify likelihood of customers defecting to the rivalry include poor response to complaints and errors relating to billing services.

3.2 Importance Of Customer Churn In Telecom Industry

Sjarif, et al (2019) emphasized that it is important for Telecom Company to have a churn prediction model in order to prevent their user from moving to another operatorservices. Consequently, the underlying principle of this study is to develop the customer churn prediction model. Machine learning can possibly be the sort of tools which could help telecom companies in churn prediction model. Machine learning is a kind of artificial intelligence tools which give the capability to let computer learns the algorithm instinctively without human contribution. Comparatively, churn prediction in telecom has been considered as unique application domain to churn prediction than othersubscriptionbased industry as a result of the variety, volume and biases of the information. On the basis of the findings, the study noticed that the KNN algorithm surpasses the others with the accurateness for training and testing is the ratio of 80.45

3.3 Summary

There are three underlying objectives of the study. One is to explore the "customer churn prediction in telecom using in big machine learning data platform." Second, was to analyze the importance of churn prediction model in telecom industry in order to help telecom industry to inhibit customer churn rate in its initial stage itself. Third, was to compare the algorithms that are effective in reducing churn rate in telecom industry. In this case, various algorithms relating to customer churn prediction have been compared and discussed its effectiveness. The study positively contributes to telecom industry to suggest the optimal solutions in reducing churn rate in telecom industry

Chapter 4

Requirement Analysis

4.1 Purpose

In order to achieve a working project in the python environment some major requirements were done. Software Development require some important steps to be taken. In developing our Machine Learning model was also done with similar steps

4.2 Problem Definition

In short, the work aims at Predicting the likeliness of a customer leaving an Organization and thereby allowing the Organization to go with the proper steps to make the customers not to churn and make the customer churn. This project uses a machine learning algorithm can help Organization in finding the customer will leave or not.

4.2.1 Objective

- The following are the main objectives of the System:
- • The user can input the value according to the Datas
- • The trained model using the Machine learning algorithm will process the inputs.
- • Using the Inputs it then shows the resulting predicted value to the user.

4.2.2 Hardware Requirements

- Processor : Intel Core i3
- Storage : 512 GB Hard Disk space
- Memory : 4 GB RAM

4.2.3 Software Requirements

- Operating System : Linux/Windows
- Platform : Python
- Librarie used : flask, pandas, matplotlib, numpy, sklearn,

4.3 Functional Requirements

The functional requirements includes all the activities or processes that should be achieved by the proposed system. It includes

- **flask:** is a micro web framework written in Python. It is classified as a micro framework because it does not require particular tools or libraries.It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools
- **sklearn:** sk learn (formerly sci-kit learn and sometimes called sk learn) is a machine learning library can be used in python programming language. By using this library, we can implement various regression, classification and clustering algorithms such as random forest, support vector machine, k-means and DBSCAN. And the sk learn library is built in a way that it can work with various scientific and numeric libraries of python such as scipy and numpy.

- **matplotlib:** It's used for the visualisation of data in python programming language. It's implemented to work with the wider scipy stack and it's built on numpy arrays. It's a multi platform data visualization technique. It was developed in 2002 by John Hunter. Visualization is the most efficient way to understand the data. Using this library, we can represent our data in various plots such as line, bar, histogram, scatter etc.

4.4 Non Functional Requirements

4.4.1 Performance Requirements

- Accuracy : Accuracy in functioning and the nature of user-friendly should be maintained by the system.
- Speed : The system must be capable of offering speed.
- Low cost: This system is very cheap to implement and is also user-friendly.
- Less Time consuming: It uses very less time comparing to the existing sysytem .
- User Friendly: This proposed system is highly user friendly they enables to create a good environment.

4.4.2 Quality Requirements

- Scalability : The software will meet all of the functional requirements.
- Maintainability : The system should be maintainable. It should keep backups to atone for system failures, and should log its activities periodically.
- Reliability : The acceptable threshold for down-time should be large as possible. i.e. mean time between failures should be large as possible. And if the system is broken, time required to get the system backup again should be minimum.
- Availability: This system is easily available as the core equiments in building the sofware is easily obtained.
- High- Functionality: This system is highly functional in all environment since, They are highly adaptable.

Chapter 5

Design And Implementation

5.1 introduction

This chapter design customer churn prediction in telecom using machine learning models. This study makes use of Random Forest for predicting consumer churn in the telecom sector

5.2 System Design

Design of a system can be defined as the process of applying various techniques and principles for defining a device, a process or a system in sufficient detail to permit its physical realization. Thus, system design is a solution, a "how to" approach to the creation of a new system. This important phase provides the understanding and procedural details necessary for implementing the system recommended in the feasibility study. The design step produces a data design, an architectural design and a procedural design. The data design transforms the information domain model created during analysis into the data structures that will be required to implement the software. The architectural design defines the relationships among major structural components into a procedural description of the software. Source code is generated, and testing is conducted to integrate and validate the software. From the project management point of view software design is conducted in two steps, preliminary design is concerned with the transformation of requirements into data and software architecture. Detailed data structure and algorithmic representation for software.

5.2.1 Random Forest Classifier

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

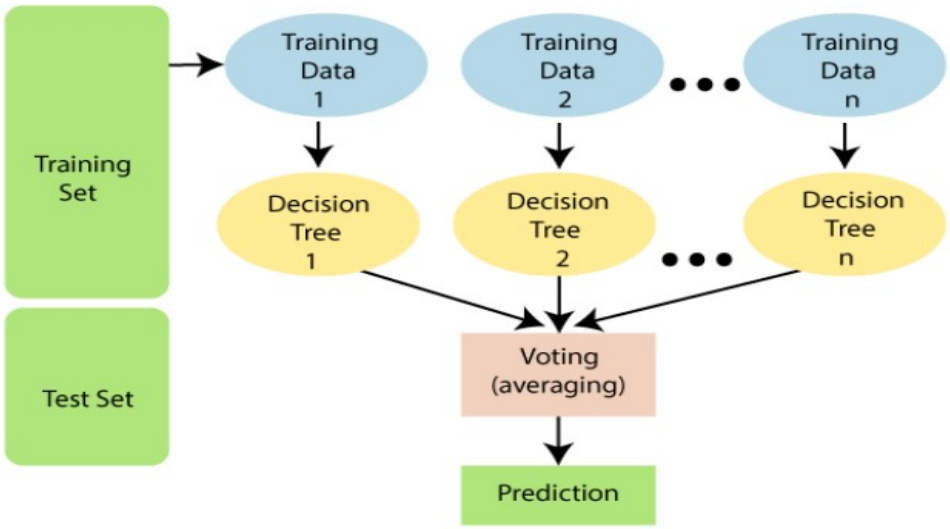

Figure 5.1: Architecture of model Creation

5.2.2 Assumptions For Random Forest

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random Forest classifier:

- The predictions from each tree must have very low correlations.

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.

5.2.3 Reason For Using Random Forest

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing. Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase. The Working process can be explained in the below steps and diagram:
- Select random K data points from the training set.
- Build the decision trees associated with the selected data points
- Choose the number N for decision trees that you want to build.
- Repeat Step 1 2.
- For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes. The working of the algorithm can be better understood by the below example

5.3 Dataset

The Dataset is one of the most Important part of any machine learning algorithm.The Dataset I used here contains 7043rows and 21 columns that is made availiable through the kaggle.for convenience the dataset has been divided into three groups.
1)Demographic Information

- gender: Whether the client is a female or a male (Female, Male).
- SeniorCitizen: Whether the client is a senior citizen or not (0, 1).
- Partner: Whether the client has a partner or not (Yes, No).

- Dependents: Whether the client has dependents or not (Yes, No).

(2) Customer Account Information

- tenure: Number of months the customer has stayed with the company (Multiple different numeric values).
- Contract: Indicates the customer's current contract type (Month-to-Month, One year, Two year).
- PaperlessBilling: Whether the client has paperless billing or not (Yes, No).
- PaymentMethod: The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit Card (automatic)).
- MontlyCharges: The amount charged to the customer monthly (Multiple different numeric values).
- TotalCharges: The total amount charged to the customer (Multiple different numeric values).

(3) Services Information

- PhoneService: Whether the client has a phone service or not (Yes, No).
- MultipleLines: Whether the client has multiple lines or not (No phone service, No, Yes).
- InternetServices: Whether the client is subscribed to Internet service with the company (DSL, Fiber optic, No)
- OnlineSecurity: Whether the client has online security or not (No internet service, No, Yes).
- OnlineBackup: Whether the client has online backup or not (No internet service, No, Yes).
- DeviceProtection: Whether the client has device protection or not (No internet service, No, Yes).
- TechSupport: Whether the client has tech support or not (No internet service, No, Yes).
- StreamingTV: Whether the client has streaming TV or not (No internet service, No, Yes).

- StreamingMovies: Whether the client has streaming movies or not (No internet service, No, Yes).

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport	StreamingTV
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	No	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	No	No
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	Yes	No
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	No	No

Figure 5.2: DATASET

5.4 Flowchart

A flowchart is a type of diagram that represents a workflow or process. A flowchart can also be defined as a diagrammatic representation of an algorithm, a step-by-step approach to solving a task. The flowchart shows the steps as boxes of various kinds, and their order by connecting the boxes with arrows.

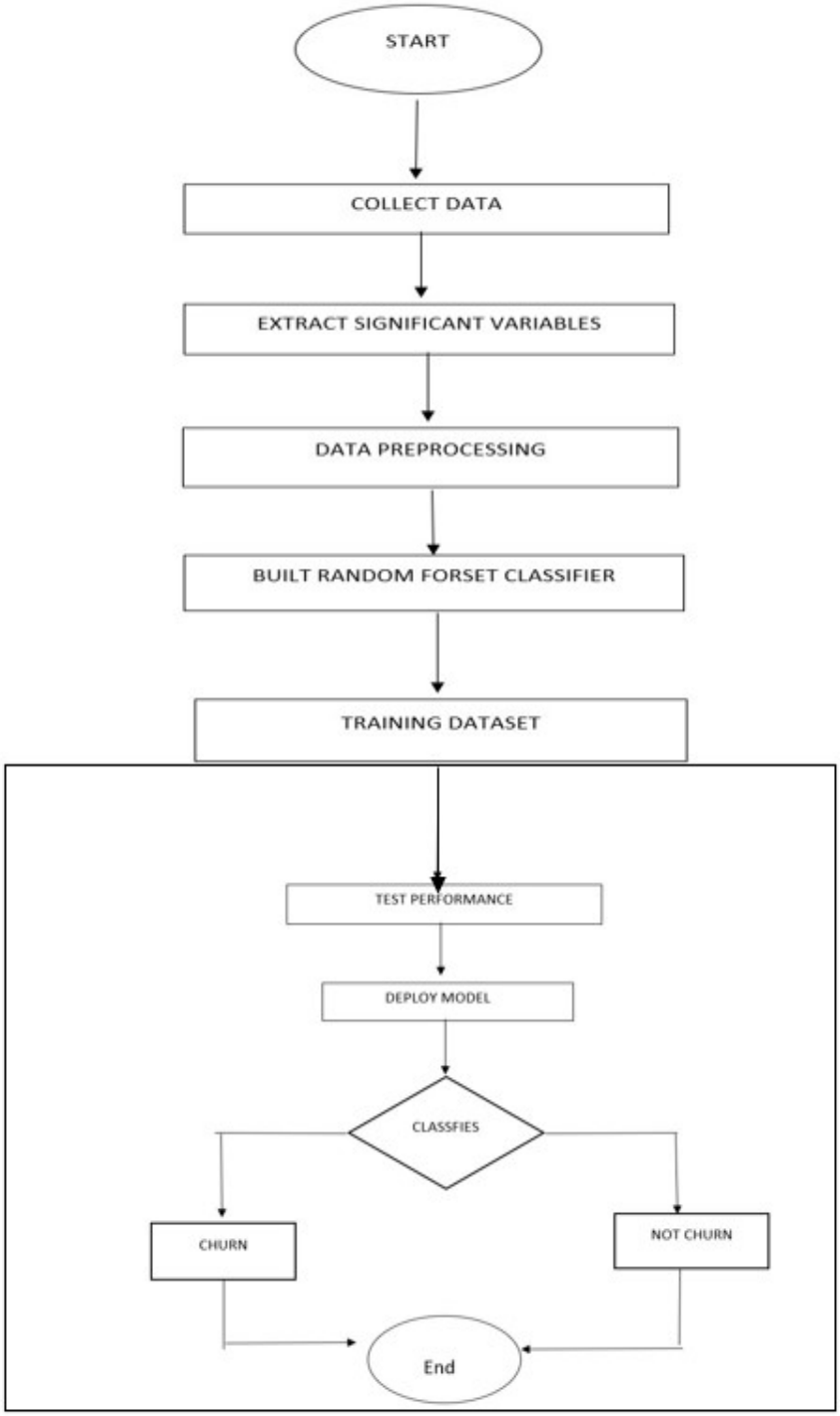

Figure 5.3: FLOWCHART

5.5 Screenshots of user interface

Figure 5.4: input

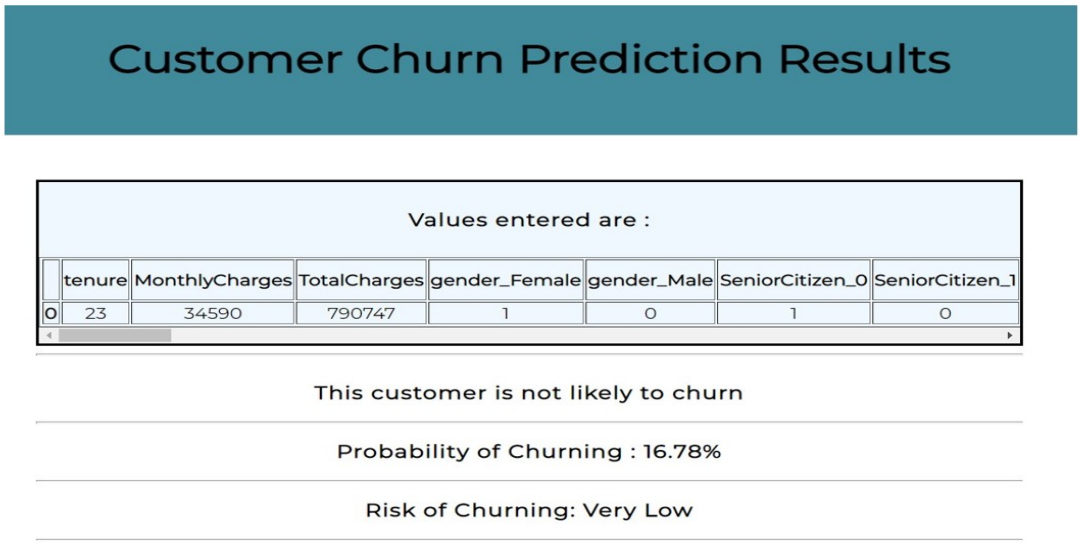

Figure 5.5: output

Chapter 6

Coding

6.1 Algorithm

Algorithm 1 Algorithm for Creating the model:

1: Creating a dataset for the model. The overall result and quality of the project is heavily dependent on the quality of the dataset that is used to train the system.

2: Reading, formatting, cleaning the dataset,The way by which a system interprets the data set can create a huge impact

3: Determining and identifying the class to predict

4: Training the model based on the dataset which is formatted, The model is trained with dataset in the ratio 80:20 in order to check its accuracy without this we can't say that if the system is predicting correctly

5: Obtaining the model, The model after training and finding the accuracy is stored for future implementation

6: Providing data to the pre trained model.To make the desired outcome the data need to be provided to the model Implementing the model using machine learning. stop

6.2 Source Code

6.3 Screenshots of user interface

```
#EDA Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
from matplotlib import rcParams

#Model Building Libraries
from sklearn import metrics
from sklearn.model_selection import train_test_split
from sklearn.metrics import recall_score, confusion_matrix, classification_report, accuracy_score
from sklearn.tree import DecisionTreeClassifier
from imblearn.combine import SMOTEENN
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import AdaBoostClassifier
```

Figure 6.1: importlibraries

```
df_base=pd.read_csv('Telco-Customer-Churn.csv')
```

Figure 6.2: Load Dataset

```
df= df_base.copy()
df['TotalCharges']=pd.to_numeric(df_base['TotalCharges'],errors='coerce') #coerce puts NaN values if there are any parsing errors
df.isnull().sum()
```

Figure 6.3: Coversion to Numeric and Checking Null values

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=11)
model_rf=RandomForestClassifier(n_estimators=100, criterion='gini', random_state = 100,max_depth=6, min_samples_leaf=8)
model_rf.fit(X_train,y_train)
y_pred = model_rf.predict(X_test)
print_report(model_rf,X_test,y_test,y_pred)
rf = round(accuracy_score(y_test, y_pred)*100, 2)
```

Figure 6.4: Splitting The Data

```
sm = SMOTEENN()
X_resampled2, y_resampled2 = sm.fit_resample(X_train,y_train)

Xr_train,Xr_test,yr_train,yr_test=train_test_split(X_resampled, y_resampled,test_size=0.2)
model_rf_smote=RandomForestClassifier(n_estimators=100, criterion='gini', random_state = 100,max_depth=6, min_samples_leaf=8)

model_rf_smote.fit(Xr_train,yr_train)
yr_pred = model_rf_smote.predict(Xr_test)

print_report(model_rf_smote,Xr_test,yr_test,yr_pred)
rf_smote = round(accuracy_score(yr_test, yr_pred)*100, 2)
```

Figure 6.5: Applying Smote

```
Confusion Matrix :
 [[392  35]
 [ 15 508]]

Accuracy : 94.74
Classification report :

              precision    recall  f1-score   support

           0       0.96      0.92      0.94       427
           1       0.94      0.97      0.95       523

    accuracy                           0.95       950
   macro avg       0.95      0.94      0.95       950
weighted avg       0.95      0.95      0.95       950
```

Figure 6.6: Accuracy and Confusion matrix

Chapter 7

Testing and Implementation

7.1 Testing and various types of testing used.

Once a software is developed, the major activity is to test whether the actual results match with the experimental results. This process is called testing. It's used to make sure that the developed system is defect free. The main aim of testing is to find the errors and missing operations by executing the program. It also ensure that all of the objectivs of the project are met by the developer. The objective of testing is not only to evaluate the bugs in the created software but also finding the ways to improve the efficiency, usability and accuracy of it. It aims to measure the functionality, specification and performance of a software program. Tests are performed on the created software and their results are compared with the expected documentation. When there are too much errors occurred, debugging is performed. And the result after debugging is tested again to make sure that the software is error free. The major testing processes applied to this project are unit testing, integration testing and system testing. In unit testing, our aim is to test all individual units of the software. It makes sure that all of the units of the software works as it intended. In integration testing, the combined individual units are tested to check whether it met the intended function or not. It helps us to find out the faults that may arise when the units are combined. In system testing the entire software is tested to make sure that it satisfies all of the requirements. The tables shown below describes the testing process occurred during the development of this project "Automated essay scoring". This defines the various steps took to create the project error free.

7.1.1 Unit Testing

Text Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	create the user interface	To load the web page with required fields	Same as expected	Pass
2	pre-processing	clean the dataset for feature extraction	same as expected	Pass
3	extract features from dataset	extract various features from dataset and store it in a csv file	csv file generated	Pass
4	training and testing of model	create the model and store it in a pickle file	pickle file generated	Pass
5	prediction	predict the result accurately	same as expected.	Pass
6	python server program	set up a python flask server to run the program	Same as expected	Pass

Table 7.1: Unit test cases and results

7.1.2 Integration Testing

Text Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	load the user interface from python	the user interface is loaded when we run the flask program	Same as expected	Pass
2	pass input essay from User to server	To pass the input entered by the user to the python program to and receive it there.	Same as expected	Pass
3	evaluation	load the previously generated pickle file to the server and predict the result with it and extracted features.	Same as expected	Pass
4	display results	pass the result and display it	Same as expected	Pass

Table 7.2: Integration cases and result

7.1.3 System Testing

Text Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	to run python server	Server program executed successfully, hence the entire program worked without any crash	Same as expected	Pass
2	evaluation	allow user to input and output generated according to the input essay.	Same as expected	Pass

Table 7.3: System test cases and results

Chapter 8

Results and Discussion

The main aim of the project was to predict the customers will churn or not churn by using a machine learning model. And it is observed that the system performs all the functionalities as expected. By using this machine learning model the we can help their Employees from being churned.

8.1 Advantages and Limitations

The proposed system is a machine learning model to evaluate the input essays and predict it's score. The proposed system posses more advantages over the existing system. The proposed system save a huge amount of time. Like every other system, this system also have it's own disadvantages. But they are negligible while comparing with the advantages and they can be overcame in future.

8.1.1 Advantages

- Companies can Identify at-risk customers.
- Identify the pain points,What the customers needed
- Can reduce the Human resources.
- Identify the Methods to reduce the Customer churn.
- Since the process of evaluation is Using Machine Learning, the Companiess can Reduce their Investment Cost.

8.1.2 Limitations

- The main limitation of the system is that in some rare cases the system will miss classify the target value. Our future goal is to make our system 100percent accurate.
- The feature extraction process can be found as slow. but it can be improved using high performing systems.

Chapter 9

Conclusion and Future Scope

The importance of churn prediction will help many companies, mainly in telecom industries, to have a profitable income and achieve good revenue. Customer churn prediction is the major issue in the Telecom Industry, and due to this, companies are trying to keep the existing ones from leaving rather than acquiring a new customer. Three tree-based algorithms were chosen because of their applicability and diversity in this type of application. By using Random Forest, XGBoost, and Support vector Machine, we will get more accuracy comparing other algorithms. Here we are using the dataset of some customers about their service plan and checking the values of them and have a precise prediction, which will help to identify the customers who are going to migrate to other company services. By this, the Telecom Company can have a clear view and can provide them some exiting offers to stay in that service. The obtained results show that our proposed churn model produced better results and performed better by using machine learning techniques.Random Forest produced better accuracy among the various methods

The results obtained by the created model seems encouraging and can be improved in future. The rate of errors in the machine learning model is very minimum . The majority of the project was built in python. It uses a flask server to connect to the user interface built using HTML, JavaScript and CSS. The project was built with the help of various python libraries such ass sklearn, pandas, numpy etc.

In the coming days, we will further research on lazy learning approaches to have better customer churn prediction.To know the changing behavior of the customers, the study can be extended by using Artificial Intelligence techniques for trend analysis and customer prediction.

Bibliography

[1] V. Lazarov and M. Capota, "Churn prediction," Bus. Anal. Course, TUM Comput. Sci, Technische Univ. München, Tech. Rep., 2007. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.462.7201rep=rep1type=pdf

[2] S. A. Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, and A. Rehman, "Telecommunication subscribers' churn prediction model using machine learning," in Proc. 8th Int. Conf. Digit. Inf. Manage., Sep. 2013, pp. 131–136

[3] S. A Survey on Customer Churn Prediction using Machine Learning Techniques.https://www.researchgate.net/publication/310757545A Su rveyon Customer Churn Prediction using MachineLearning Tech niques

[4] Effectual Predicting Telecom Customer Churn using Deep Neural Network.https://www.ijeat.org/wp-content/uploads/papers/v8i5/D6745048419.pdf

[5] Customer Churn Prediction in Telecommunication with Rotation Forest Methodhttps://www.researchgate.net/publication/282981765 Customer churn prediction in telecommunication

[6] Customer churn prediction in telecom using machinelearninghttps://journalofbigdata.springeropen.com/articles/10.1 186/s40537-019-0191-6