TELECOM CUSTOMER CHURN PREDICTION USING MACHINE LEARNING

A UG PROJECT PHASE – 2 REPORT

Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Submitted By

VIJAY KUMAR SRIPADA	**18UK1A0554**
SHRAVYA MADATHA	**18UK1A0551**
POOJA POREDDY	**18UK1A0546**
SAI SHASHANK LINGABATHINI	**18UK1A0528**

Under the guidance of

Mr. V. RANADHEER REDDY

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

(Affiliated to JNTU , HYDERABAD)

BOLLIKUNTA, WARANGAL (T.S) – 506005

2018-2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

BOLLIKUNTA, WARANGAL – 506005

2018 - 2022

CERTIFICATE OF COMPLETION
UG PROJECT PHASE-2

This is to certify that the UG Project Phase – 2 report entitled **"TELECOM CUSTOMER CHURN PREDICTION USING MACHINE LEARNING"** is being submitted by **VIJAY KUMAR SRIPADA (18UK1A0554), SHRAVYA MADATHA(18UK1A0551), POOJA POREDDY (18UK1A0546), SAI SHASHANK LINGABATHINI (18UK1A0526)** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** to **Jawaharlal Nehru Technological University Hyderabad** during the academic year **2018- 2022**, is a record of work carried out by them under the guidance and supervision.

Project Guide	**Head of the Department**
Mr. V. RANADHEER REDDY	**Mr.Dr. R. NAVEEN KUMAR**
(Assistant Professor)	(Professor)

EXTERNAL

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P. Prasad Rao**, Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase – 2 in the institute.

We extend our heartfelt thanks to **Dr. R. Naveen Kumar**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase – 2.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase -2 and for their support in completing UG Project Phase-2.

We express heartfelt thanks to the guide, **Mr. V. Ranadheer Reddy**, Assistant Professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Project Phase – 2.

Finally, we express our sincere thanks and gratitude to our family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

VIJAY KUMAR SRIPADA **(18UK1A0554)**
SHRAVYA MADATHA **(18UK1A0551)**
POOJA POREDDY **(18UK1A0546)**
SAI SHASHANK LINGABATHINI **(18UK1A0528)**

ABSTRACT

Customer churn has become highly important for companies because of increasing competition among companies, increased importance of marketing strategies and conscious behavior of customers in recent years. Customers can easily trend toward alternative services. Companies must develop various strategies to prevent these possible trends, depending on the services they provide. During the estimation of possible churns, data from the previous churns might be used. An efficient churn predictive model benefits companies in many ways. Early identification of customers likely to leave may help to build cost effective ways in marketing strategies.

Telecommunication industry always suffers from a very high churn rates when one industry offers a better plan than the previous there is a high possibility of the customer churning from the present due to a better plan in such a scenario it is very difficult to avoid losses but through prediction we can keep it to a minimal level. A machine learning model is built and this helps to identify the probable churn customers and then makes the necessary business decisions.The model experimented with four algorithms: Decision Tree(classification), Random Forest (classification), Logistic Regression, Support Vector Machine (Classification algorithm). However, the best results were obtained by applying the support vector Machine classification algorithm.This algorithm was used for classification in this churn predictive model.

Keywords: *Customer churn prediction, Churn in telecom, Machine learning, Classification*

TABLE OF CONTENTS

LIST OF CHAPTERS **Page No.**

LIST OF FIGURES PAGE NO:

1. INTRODUCTION

1.1. MOTIVATION

Telecommunications sector has become one of the main industries in developed countries. The technical progress and the increasing number of operators raised the level of competition Companies are working hard to survive in this competitive market depending on multiple strategies. Tree main strategies have been proposed to generate more revenues : acquire new customers, upsell the existing customers, and the retention period of customers. However, comparing these strategies taking the value of return on investment (RoI) of each into account has shown that the third strategy is the most profitable strategy , proving that retaining an existing customer costs much lower than acquiring a new one , in addition to being considered much easier than the upselling strategy . To apply the third strategy, companies have to decrease the potential of customer's churn, known as "the customer movement from one provider to another" . Customers' churn is a considerable concern in service sectors with high competitive services. On the other hand, predicting the customers who are likely to leave the company will represent a potentially large additional revenue source if it is done in the early phase.

Many research confirmed that machine learning technology is highly efficient to predict this situation. This technique is applied through learning from previous data.Te model also was evaluated using a new dataset and the impact of this system to the decision to churn was tested. The model gave good results and was deployed to production.

1.2. DEFINITION

Based on the introduction the key challenge is to predict if an individual customer will churn or not. To accomplish that, machine learning models are trained based on 80% of the sample data. The remaining 20% are used to apply the trained models and assess their predictive power with regards to "churn / not churn". A side question will be, which features actually drive customer churn. That information can be used to identify customer "pain points" and resolve them by providing goodies to make customers stay.

To compare models and select the best for this task, the accuracy is measured. Based on other characteristics of the data, for example the balance between classes (number of “churners” vs. “non-churners” in the data set) further metrics are considered if needed.

1.3. OBJECTIVE OF PROJECT:

The primary and secondary objectives of the study are as follows:

- **Primary objectives**

❖ To explore the customer churn prediction in Telecom using Machine Learning .

- **Secondary objectives**

❖ To investigate the impact of customer churn in the Telecom industry as a whole.

❖ To discuss the significance of customer churn models in the Telecom industry .

❖ To compare the algorithms that are effective in reducing churn rate in Telecom companies.

1.4. PURPOSE:

By telecom customer churn prediction using machine learning we will:

❖ Known churn analysis is the evaluation of a company's customer loss rate in order to reduce it.

❖ One of the ways to calculate a churn rate is to divide the number of customers lost during a given time interval by the number of active customers at the beginning of the period.

2. PROBLEM STATEMENT

The retention and acquisition of users are the major concerns in the telecom industry. The fast growth of the marketplace in every business is giving rise to increased subscriber base. Accordingly, companies have recognized the significance of retaining the customers who are 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 identifying those users who are likely to switch a company over another. Telecom is enduring the problem of an ever-increasing churn rate. Accordingly, the current study employs machine learning algorithms. 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 be such kind of users who might probably churn 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 existing customers are the most beneficial resources for companies than acquiring new one.

3.LITERATURE SURVEY

3.1. EXISTING SYSTEM

Many studies are available for churn problems from different viewpoints with different datasets,algorithms and for different industries where churn analysis is one of the world wide used to analyze the customer behaviors and predict the customers who are about to leave the service agreement from a company.

Studies revealed that gaining new customers is 5 to 10 times costlier than keeping existing customers happy and loyal in today's competitive conditions,and that an average company loses 10 to 30 percent of customers annually.

A large number of researches in the subject of churn prediction have been 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 the telecom industry in the recent period.

- Tanneedi, (2016) pointed out that customer churn has become a dreadful problem for the 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 applications such as Decision trees (DT) have been used for three different datasets. From the analytics of DT, decision trees with accuracy rates of 52%, 70% and 95% have been obtained for three different data sources 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 the telecom industry.
- Huang, (2015) exhibited in terms of telecom industry churn prediction can easily be done with big data and with 3V's such as volume, variety along with velocity. Findings emphasize that the performance of prediction has been enhanced considerably by employing a big amount of training data, a huge number of features from both operations and business support systems, as well as an increased velocity of processing new data.The study has deployed this prediction technique of churn in one of the largest mobile network operators in China. From a large number of active customers, this technique could impart prepaid customers who are about to churn, holding 0.96 accuracy rate for the top 50000 estimated churners in the list. The operations of automated matching retention with the focused essential churners considerably increase their rates of recharge,bringing about a big value for business.
- Nigam, et al (2019) dealt with machine learning techniques for customer churn prediction. A large number of machine learning techniques have been employed in the past for predicting customer churn such as Decision trees, SVM, Logistic Regression, SVM, NN, etc. Machine learning entails constructing algorithms which could learn from dataset available and could be employed to make predictions on information. One of the

most important techniques to predict customer churn in the telecom sector is deep neural networks. By employing it is possible to build a model that corresponds to our data employing different hierarchies of concepts therefore intensifying the performance of the model developed. The current study used the multi-layer artificial neural network (ANN), which is also referred to as deep neural network to predict telecom customer churn. The proposed model in the study has obtained sensitivity of 85% and consequently the findings are satisfactory.

3.2. PROPOSED SOLUTION

In a business scenario predicting customer churn is where a firm is attempting to retain customers which is much probable to leave the services. For reducing the rate of churn this study classifies which customers are likely to churn probably and which will not churn probably. Since obtaining new customers is challenging it is essential to retain present customers. Churn can be decreased by examining the essential customers' past history systematically. Huge amount of data is managed about the customers and on carrying out appropriate examinations on the same it is feasible to find probable customers that might churn. The data that is feasible can be examined in varied ways and thereby offers different ways for operators to imagine the churning of customers and avoid the same. The below figure shows the steps used for proposed system

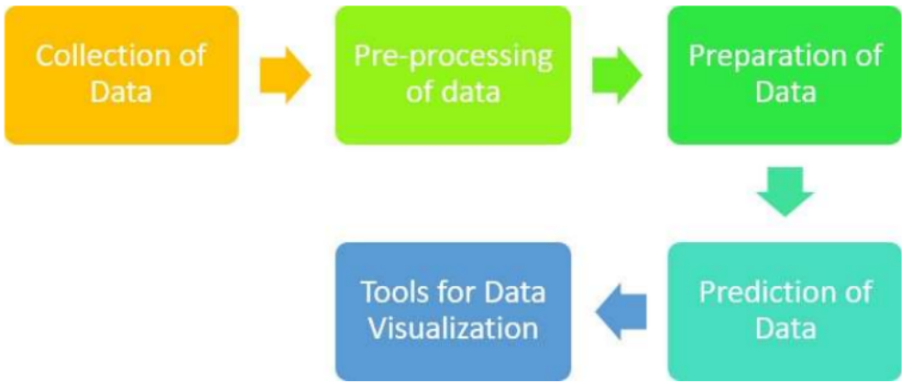


Figure 1: Steps Used For Proposed Solution.

4. EXPERIMENTAL ANALYSIS

4.1. PROJECT ARCHITECTURE:

The Project Architecture briefly explains the procedure involved:

- Firstly, Collect the dataset and split them into Training and Testing datasets.
- Preprocess both training and testing datasets.
- Use machine learning algorithms to classify and build the model.
- Now using Python-Flask module and HTML build the app/webpage to classify the churn

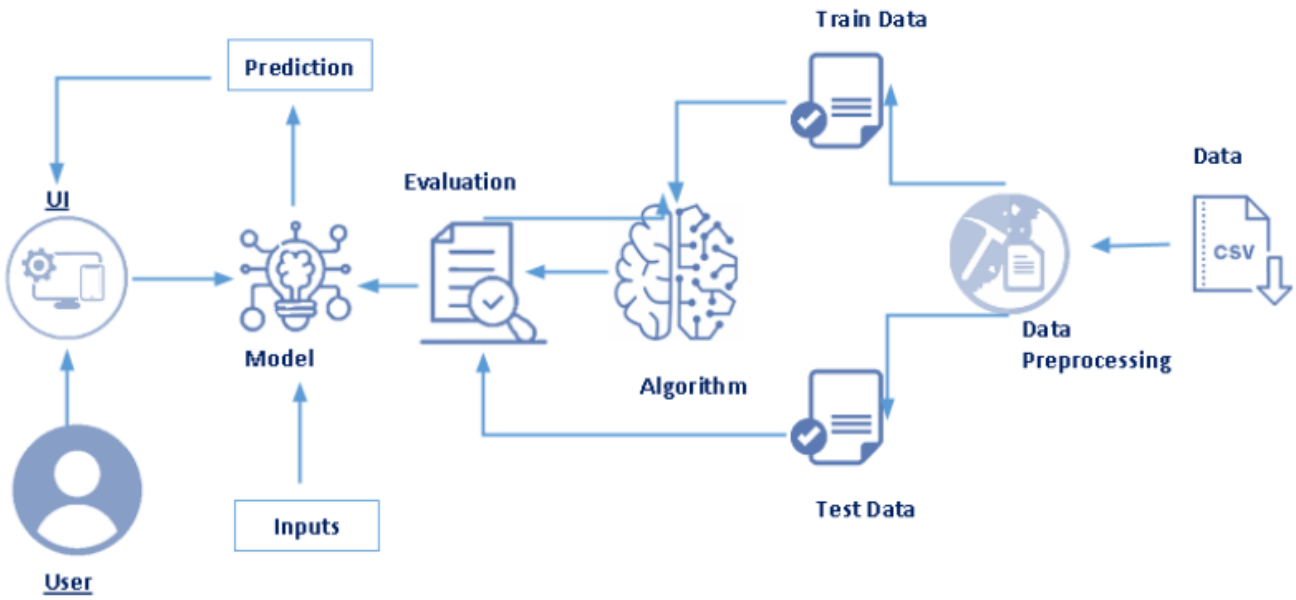


Figure 2: Project Architecture.

4.2. BLOCK DIAGRAM:

Block diagram represents the procedure in systematic and sequential manner with its blocks connected by lines that show the relationship of the blocks.

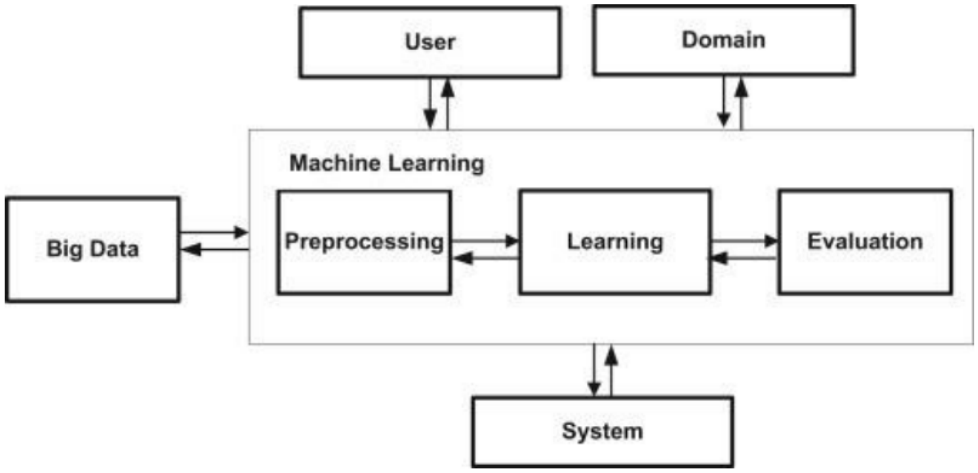


Figure 3: Block diagram representing process of ML.

4.3. SOFTWARE REQUIREMENTS

4.3.1. Python 3.9

4.3.2. Anaconda Environment

4.3.3. Flask

4.3.4. And other python libraries like NumPy, pandas, etc..,

Figure 4: Logos of python and Anaconda

4.4. PROJECT FLOW

You will go through all the steps mentioned below to complete the project.

- User interacts with the UI (User Interface) to enter Data
- The entered data is analyzed by the model which is integrated.
- Once model analyses the input the prediction is showcased on the UI

To accomplish this, we have to complete all the activities and tasks listed below

- **Data Collection.**
  - ➔ Collect the dataset or Create the dataset
- **Data Pre-processing.**
  - ➔ Import the Libraries.
  - ➔ Importing the dataset.
  - ➔ Checking for Null Values.
  - ➔ Data Visualization
  - ➔ Taking care of Missing Data.
  - ➔ Label encoding.
  - ➔ One Hot Encoding.
  - ➔ Feature Scaling
  - ➔ Splitting Data into Train and Test.
- **Model Building**
  - ➔ Training and testing the model
  - ➔ Evaluation of Mode
- **Application Building**
  - ➔ Create an HTML file
  - ➔ Build a Python Code

5. DESIGN

5.1. CLASS DIAGRAM

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application. Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of object oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints. It is also known as a structural diagram.

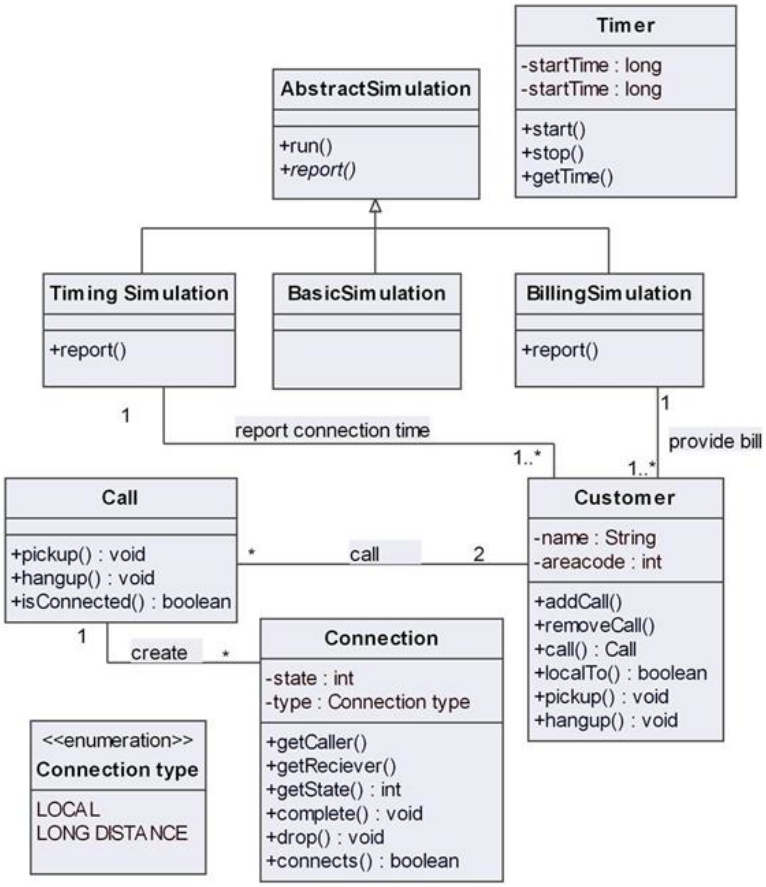


Figure 5: Class diagram

5.2. USE CASE DIAGRAM

In the Unified Modeling Language (UML), a use case diagram can summarize the details of your system's users (also known as actors) and their interactions with the system. UML is the modeling toolkit that you can use to build your diagrams. Use cases are represented with a labeled oval shape. Stick figures represent actors in the process, and the actor's participation in the system is modeled with a line between the actor and use case

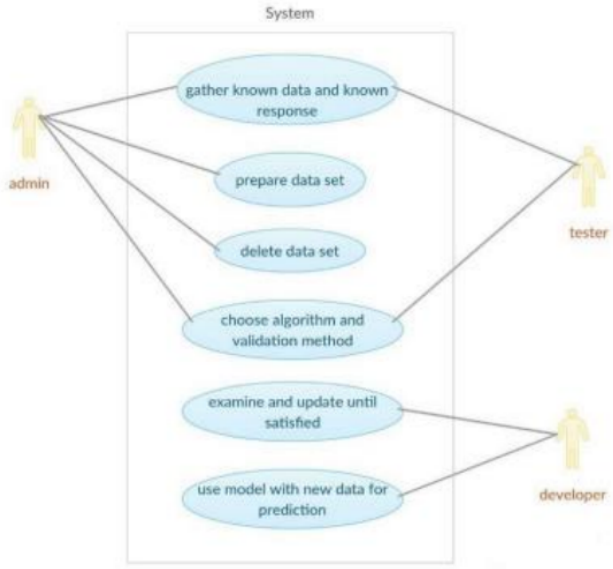


Figure 6: Use Case Diagram

5.3. FLOWCHART

A flowchart is a picture of the separate steps of a process in sequential order.

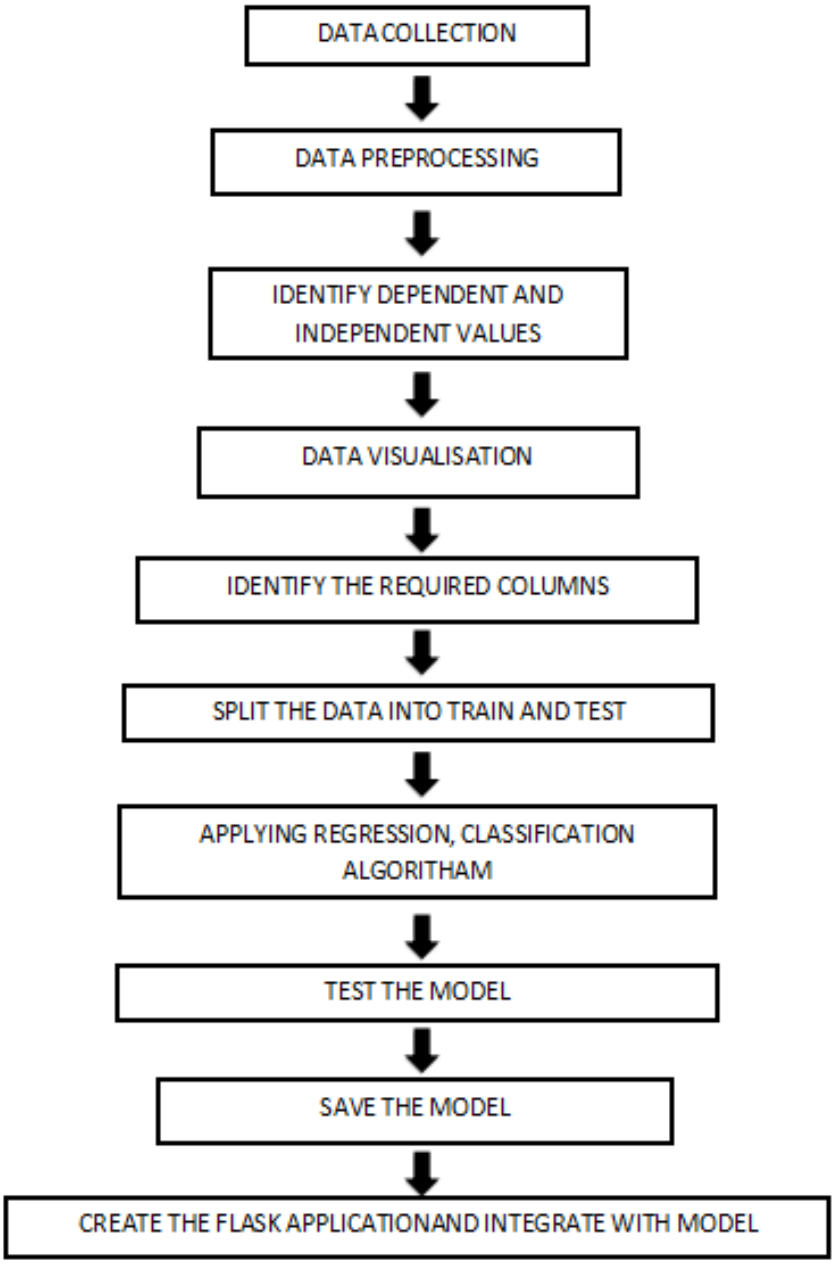


Figure 7: Flowchart

6.CODE SNIPPET

6.1 MODEL CODE

```
In [1]: #import necessary libraries
        import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns

In [2]: #import dataset
        data = pd.read_csv(r"C:\Users\ajaysripada\Desktop\Telecom churn modelling\data\DataSet.csv")
```

Figure 8: importing libraries Dataset.

```
In [3]: data.head()
```

Out[3]:

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

5 rows × 21 columns

```
In [4]: data.drop(["customerID"], axis =1, inplace = True)

In [5]: data.head()
```

Out[5]:

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSuppo
0	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	No	N
1	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	N
2	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	N
3	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes	Y
4	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	N

Figure 9: ipynb code describing important libraries and displaying the few rows from the dataset.

In [6]:
```
data.describe()
```

Out[6]:

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000

In [7]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 20 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   gender            7043 non-null   object
 1   SeniorCitizen     7043 non-null   int64
 2   Partner           7043 non-null   object
 3   Dependents        7043 non-null   object
 4   tenure            7043 non-null   int64
 5   PhoneService      7043 non-null   object
 6   MultipleLines     7043 non-null   object
 7   InternetService   7043 non-null   object
 8   OnlineSecurity    7043 non-null   object
 9   OnlineBackup      7043 non-null   object
 10  DeviceProtection  7043 non-null   object
 11  TechSupport       7043 non-null   object
 12  StreamingTV       7043 non-null   object
 13  StreamingMovies   7043 non-null   object
```

In [8]:
```
#checking for null values
data.TotalCharges = pd.to_numeric(data.TotalCharges, errors='coerce')
data.isnull().any()
```

Out[8]:
```
gender              False
SeniorCitizen       False
Partner             False
Dependents          False
tenure              False
PhoneService        False
MultipleLines       False
InternetService     False
OnlineSecurity      False
OnlineBackup        False
DeviceProtection    False
TechSupport         False
StreamingTV         False
StreamingMovies     False
Contract            False
PaperlessBilling    False
PaymentMethod       False
MonthlyCharges      False
TotalCharges         True
Churn               False
dtype: bool
```

```
In [9]: data.isnull().sum()

Out[9]: gender             0
        SeniorCitizen      0
        Partner            0
        Dependents         0
        tenure             0
        PhoneService       0
        MultipleLines      0
        InternetService    0
        OnlineSecurity     0
        OnlineBackup       0
        DeviceProtection   0
        TechSupport        0
        StreamingTV        0
        StreamingMovies    0
        Contract           0
        PaperlessBilling   0
        PaymentMethod      0
        MonthlyCharges     0
        TotalCharges      11
        Churn              0
        dtype: int64
```

Figure 10: ipynb code describing whether they are any NULL values in the Dataset.

```
In [10]: data["TotalCharges"].fillna(data["TotalCharges"].median() , inplace =True)

In [11]: data.isnull().sum()

Out[11]: gender             0
         SeniorCitizen      0
         Partner            0
         Dependents         0
         tenure             0
         PhoneService       0
         MultipleLines      0
         InternetService    0
         OnlineSecurity     0
         OnlineBackup       0
         DeviceProtection   0
         TechSupport        0
         StreamingTV        0
         StreamingMovies    0
         Contract           0
         PaperlessBilling   0
         PaymentMethod      0
         MonthlyCharges     0
         TotalCharges       0
         Churn              0
         dtype: int64
```

Figure 11: ipynb code describing filling of Null Values.

```
In [12]: data.corr()
```

Out[12]:

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges
SeniorCitizen	1.000000	0.016567	0.220173	0.102652
tenure	0.016567	1.000000	0.247900	0.825464
MonthlyCharges	0.220173	0.247900	1.000000	0.650864
TotalCharges	0.102652	0.825464	0.650864	1.000000

Figure 12: ipynb code describing Area chat and dataset corr.

```
In [13]: sns.heatmap(data.corr(), annot=True)
Out[13]: <AxesSubplot:>
```

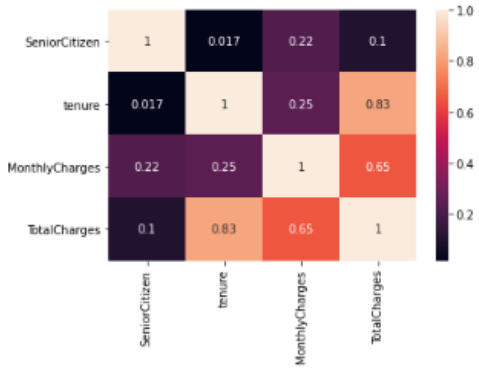


Figure 13: ipynb code describing Heatmap Plot.

```
In [14]: sns.pairplot(data=data, markers=["^","v"], palette="inferno")
Out[14]: <seaborn.axisgrid.PairGrid at 0x29fe0eb5dc0>
```

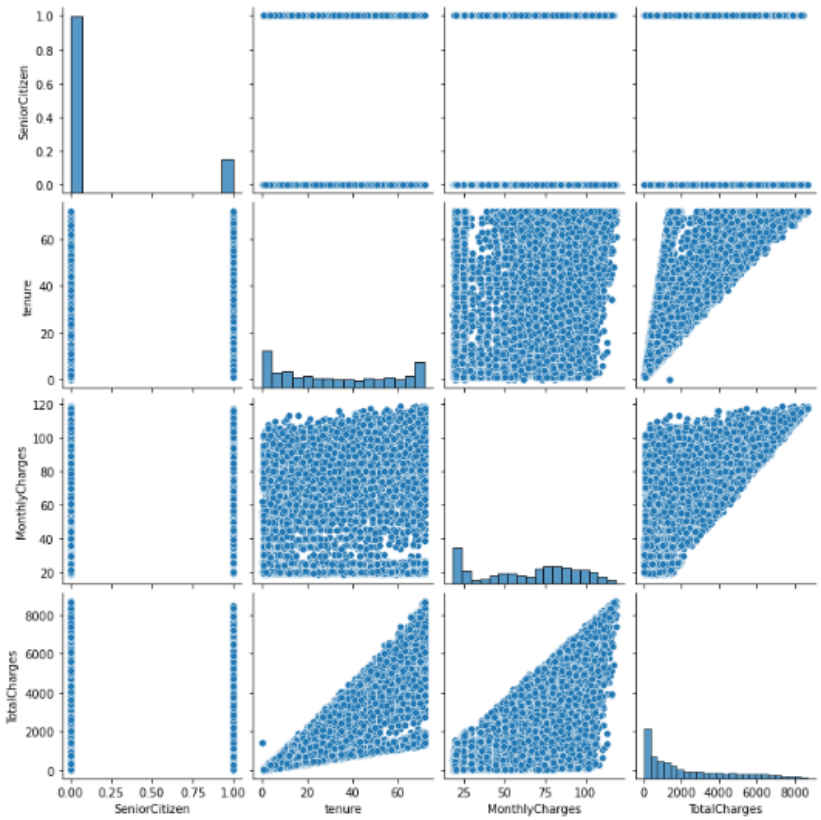


Figure 14: ipynb code describing PairPlot.

```
In [15]: for i in data:
             print(data[i].unique())
```

```
['Female' 'Male']
[0 1]
['Yes' 'No']
['No' 'Yes']
[ 1 34  2 45  8 22 10 28 62 13 16 58 49 25 69 52 71 21 12 30 47 72 17 27
  5 46 11 70 63 43 15 60 18 66  9  3 31 50 64 56  7 42 35 48 29 65 38 68
 32 55 37 36 41  6  4 33 67 23 57 61 14 20 53 40 59 24 44 19 54 51 26  0
 39]
['No' 'Yes']
['No phone service' 'No' 'Yes']
['DSL' 'Fiber optic' 'No']
['No' 'Yes' 'No internet service']
['Yes' 'No' 'No internet service']
['No' 'Yes' 'No internet service']
['No' 'Yes' 'No internet service']
['No' 'Yes' 'No internet service']
['No' 'Yes' 'No internet service']
['Month-to-month' 'One year' 'Two year']
['Yes' 'No']
['Electronic check' 'Mailed check' 'Bank transfer (automatic)'
 'Credit card (automatic)']
[29.85 56.95 53.85 ... 63.1  44.2  78.7 ]
[  29.85 1889.5   108.15 ...  346.45  306.6  6844.5 ]
['No' 'Yes']
```

Figure 15: ipynb code describing Unique Data[i].

```
In [16]: # Label Encoder is used to convert categorical to numeric form
         from sklearn.preprocessing import LabelEncoder
         le = LabelEncoder()
         data["gender"] = le.fit_transform(data["gender"])
         data["Partner"] = le.fit_transform(data["Partner"])
         data["Dependents"] = le.fit_transform(data["Dependents"])
         data["PhoneService"] = le.fit_transform(data["PhoneService"])
         data["MultipleLines"] = le.fit_transform(data["MultipleLines"])
         data["InternetService"] = le.fit_transform(data["InternetService"])
         data["OnlineSecurity"] = le.fit_transform(data["OnlineSecurity"])
         data["OnlineBackup"] = le.fit_transform(data["OnlineBackup"])
         data["DeviceProtection"] = le.fit_transform(data["DeviceProtection"])
         data["TechSupport"] = le.fit_transform(data["TechSupport"])
         data["StreamingTV"] = le.fit_transform(data["StreamingTV"])
         data["StreamingMovies"] = le.fit_transform(data["StreamingMovies"])
         data["Contract"] = le.fit_transform(data["Contract"])
         data["PaperlessBilling"] = le.fit_transform(data["PaperlessBilling"])
         data["PaymentMethod"] = le.fit_transform(data["PaymentMethod"])
         data["Churn"] = le.fit_transform(data["Churn"])
```

Figure 16: ipynb code describing Conversion of categorical to numeric form.

```
In [17]: data.head()
```

Out[17]:

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSuppc
0	0	0	1	0	1	0	1	0	0	2	0	
1	1	0	0	0	34	1	0	0	2	0	2	
2	1	0	0	0	2	1	0	0	2	2	0	
3	1	0	0	0	45	0	1	0	2	0	2	
4	0	0	0	0	2	1	0	1	0	0	0	

```
In [18]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 20 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   gender            7043 non-null   int32
 1   SeniorCitizen     7043 non-null   int64
 2   Partner           7043 non-null   int32
 3   Dependents        7043 non-null   int32
 4   tenure            7043 non-null   int64
 5   PhoneService      7043 non-null   int32
 6   MultipleLines     7043 non-null   int32
 7   InternetService   7043 non-null   int32
 8   OnlineSecurity    7043 non-null   int32
 9   OnlineBackup      7043 non-null   int32
 10  DeviceProtection  7043 non-null   int32
 11  TechSupport       7043 non-null   int32
 12  StreamingTV       7043 non-null   int32
 13  StreamingMovies   7043 non-null   int32
 14  Contract          7043 non-null   int32
 15  PaperlessBilling  7043 non-null   int32
 16  PaymentMethod     7043 non-null   int32
 17  MonthlyCharges    7043 non-null   float64
 18  TotalCharges      7043 non-null   float64
 19  Churn             7043 non-null   int32
dtypes: float64(2), int32(16), int64(2)
memory usage: 660.4 KB
```

Figure 17: ipynb code describing Converted Data.

```
In [19]: data.corr()
```

Out[19]:

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	Device
gender	1.000000	-0.001874	-0.001808	0.010517	0.005106	-0.006488	-0.006739	-0.000863	-0.015017	-0.012057	
SeniorCitizen	-0.001874	1.000000	0.016479	-0.211185	0.016567	0.008576	0.146185	-0.032310	-0.128221	-0.013632	
Partner	-0.001808	0.016479	1.000000	0.452676	0.379697	0.017706	0.142410	0.000891	0.150828	0.153130	
Dependents	0.010517	-0.211185	0.452676	1.000000	0.159712	-0.001762	-0.024991	0.044590	0.152166	0.091015	
tenure	0.005106	0.016567	0.379697	0.159712	1.000000	0.008448	0.343032	-0.030359	0.325468	0.370876	
PhoneService	-0.006488	0.008576	0.017706	-0.001762	0.008448	1.000000	-0.020538	0.387436	-0.015198	0.024105	
MultipleLines	-0.006739	0.146185	0.142410	-0.024991	0.343032	-0.020538	1.000000	-0.109216	0.007141	0.117327	
InternetService	-0.000863	-0.032310	0.000891	0.044590	-0.030359	0.387436	-0.109216	1.000000	-0.028416	0.036138	
OnlineSecurity	-0.015017	-0.128221	0.150828	0.152166	0.325468	-0.015198	0.007141	-0.028416	1.000000	0.185126	
OnlineBackup	-0.012057	-0.013632	0.153130	0.091015	0.370876	0.024105	0.117327	0.036138	0.185126	1.000000	
DeviceProtection	0.000549	-0.021398	0.166330	0.080537	0.371105	0.003727	0.122318	0.044944	0.175985	0.187757	
TechSupport	-0.006825	-0.151268	0.126733	0.133524	0.322942	-0.019158	0.011466	-0.026047	0.285028	0.195748	
StreamingTV	-0.006421	0.030776	0.137341	0.046885	0.289373	0.055353	0.175059	0.107417	0.044669	0.147186	
StreamingMovies	-0.008743	0.047266	0.129574	0.021321	0.296866	0.043870	0.180957	0.098350	0.055954	0.136722	
Contract	0.000126	-0.142554	0.294806	0.243187	0.671607	0.002247	0.110842	0.099721	0.374416	0.280980	
PaperlessBilling	-0.011754	0.156530	-0.014877	-0.111377	0.006152	0.016505	0.165146	-0.138625	-0.157641	-0.013370	
PaymentMethod	0.017352	-0.038551	-0.154798	-0.040292	-0.370436	-0.004184	-0.176793	0.086140	-0.096726	-0.124847	
MonthlyCharges	-0.014569	0.220173	0.096848	-0.113890	0.247900	0.247398	0.433576	-0.323260	-0.053878	0.119777	
TotalCharges	-0.000002	0.102652	0.318364	0.063593	0.825464	0.113013	0.452849	-0.175588	0.253935	0.375063	
Churn	-0.008612	0.150889	-0.150448	-0.164221	-0.352229	0.011942	0.038037	-0.047291	-0.289309	-0.195525	

Figure 18: ipynb code describing Area chat and dataset corr for converted Data.

```
In [20]: sns.heatmap(data.corr(), annot=False)
Out[20]: <AxesSubplot:>
```

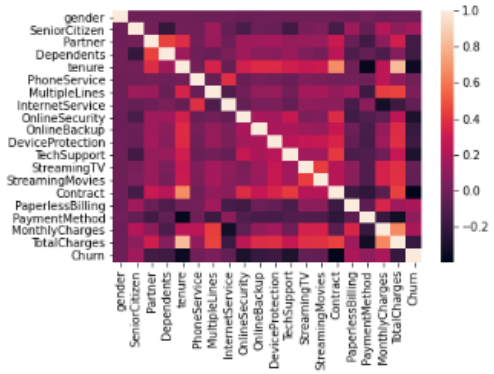


Figure 19: ipynb code describing Heatmap Plot for Converted Data.

```
In [22]: x= data.iloc[:,0:19].values
         y= data.iloc[:,19:20].values

In [23]: x
Out[23]: array([[0.0000e+00, 0.0000e+00, 1.0000e+00, ..., 2.0000e+00, 2.9850e+01,
                 2.9850e+01],
                [1.0000e+00, 0.0000e+00, 0.0000e+00, ..., 3.0000e+00, 5.6950e+01,
                 1.8895e+03],
                [1.0000e+00, 0.0000e+00, 0.0000e+00, ..., 3.0000e+00, 5.3850e+01,
                 1.0815e+02],
                ...,
                [0.0000e+00, 0.0000e+00, 1.0000e+00, ..., 2.0000e+00, 2.9600e+01,
                 3.4645e+02],
                [1.0000e+00, 1.0000e+00, 1.0000e+00, ..., 3.0000e+00, 7.4400e+01,
                 3.0660e+02],
                [1.0000e+00, 0.0000e+00, 0.0000e+00, ..., 0.0000e+00, 1.0565e+02,
                 6.8445e+03]])

In [24]: y
Out[24]: array([[0],
                [0],
                [1],
                ...,
                [0],
                [1],
                [0]])

In [25]: x.shape
Out[25]: (7043, 19)

In [26]: y.shape
Out[26]: (7043, 1)
```

Figure 20: ipynb code describing Loc, type(x) and type(y).

```
In [32]: from sklearn.preprocessing import StandardScaler
         sc = StandardScaler()
         x_train = sc.fit_transform(x_train)
         x_test = sc.fit_transform(x_test)

In [33]: from sklearn.linear_model import LogisticRegression
         lr = LogisticRegression(random_state=0)
         lr.fit(x_train,y_train)

Out[33]: LogisticRegression(random_state=0)

In [34]: lr_pred = lr.predict(x_test)

In [35]: lr_pred

Out[35]: array([0, 0, 0, ..., 0, 0, 0])

In [36]: y_test

Out[36]: array([[0],
                [0],
                [0],
                ...,
                [1],
                [0],
                [0]])
```

Figure 21: ipynb code describing label encoding and splitting the dataset into independent and the dependent variables.

```
In [37]: from sklearn.metrics import accuracy_score
         lr_acc = accuracy_score(lr_pred,y_test)

In [38]: lr_acc

Out[38]: 0.7970191625266146
```

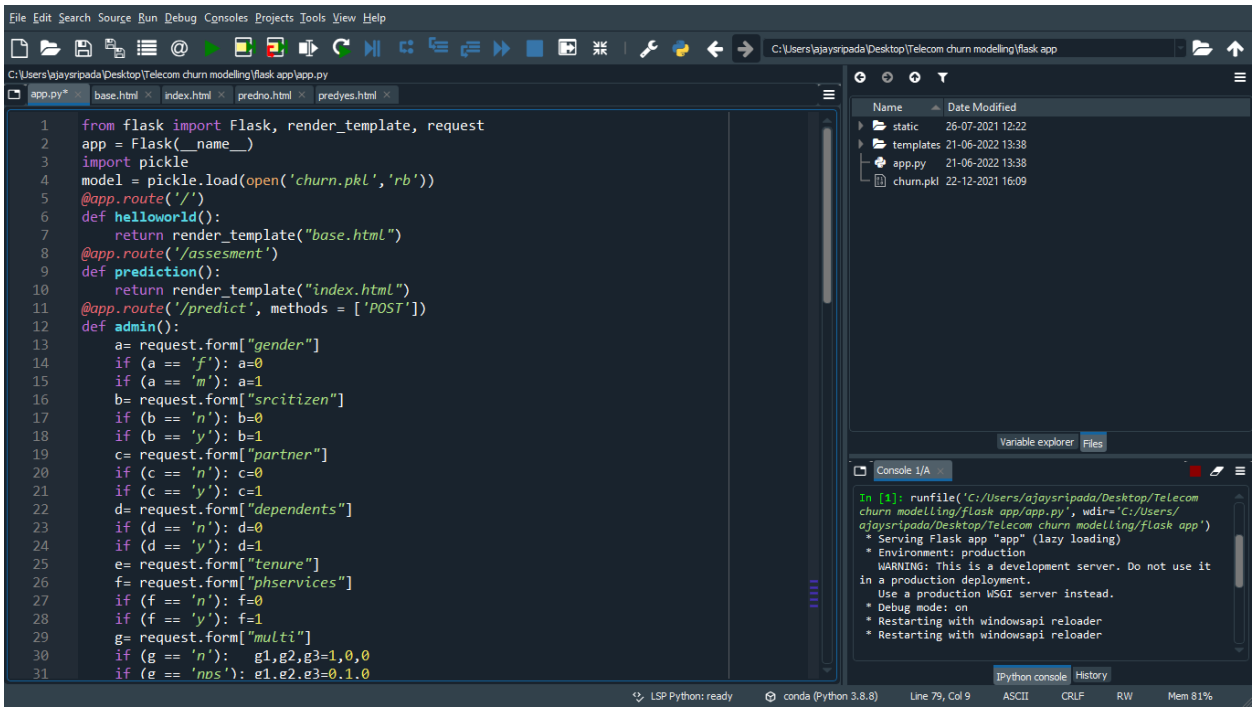
Figure 22: ipynb code describing finding Accuracy.

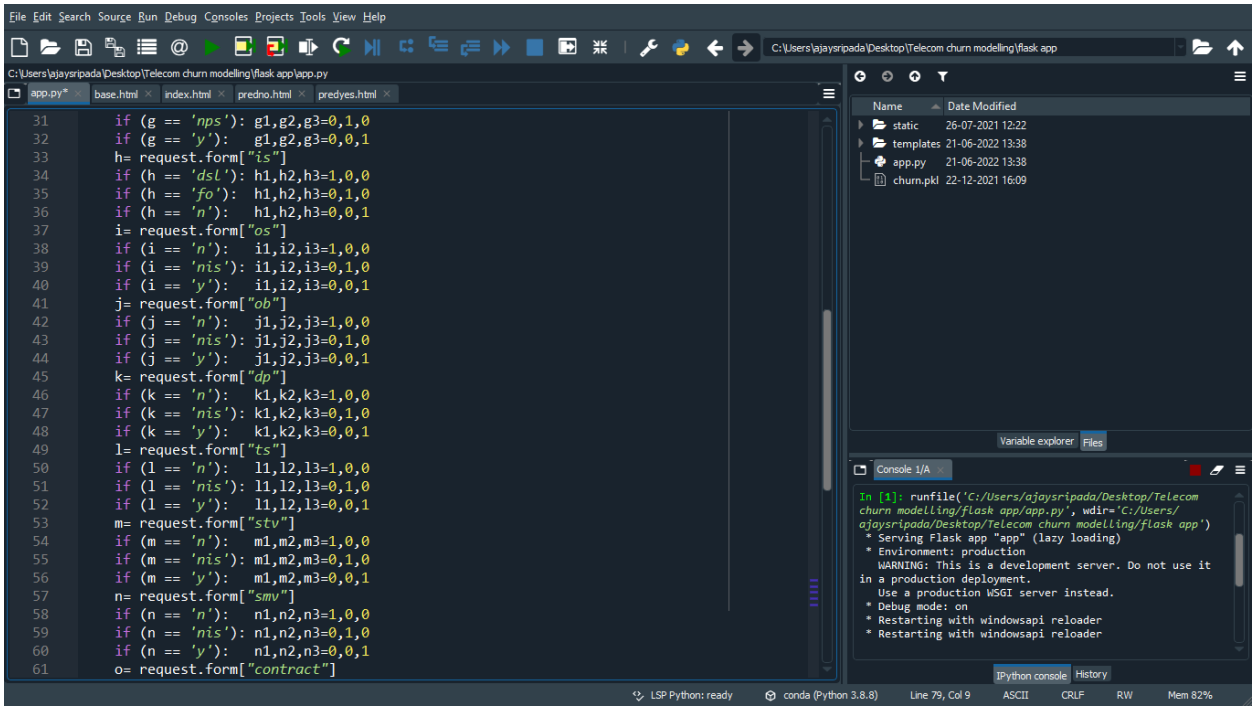
```
In [56]: import pickle
         pickle.dump(rfc,open("churn.pkl" , "wb"))
```

Figure 23: ipynb code describing pickle file.

6.2 HTML CODE AND PYTHON CODE

1. App.py code:





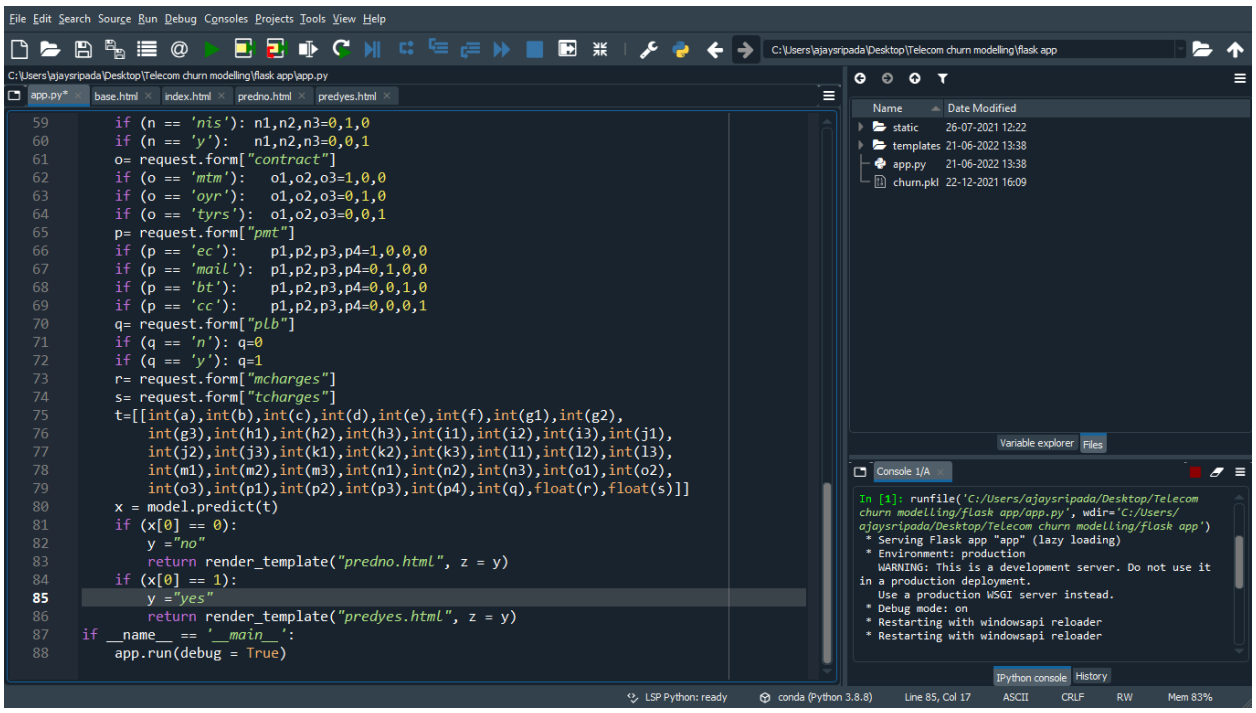


Figure 24: Python code used for rendering all the HTML pages.

2.BASE.HTML

app.py | base.html* | index.html* | predno.html | predyes.html

```
<!DOCTYPE html>
<html lang="en">
<head>
    <!-- Required meta tags-->
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
    <meta name="description" content="Colorlib Templates">
    <meta name="author" content="Colorlib">
    <meta name="keywords" content="Colorlib Templates">
    <!-- Title Page-->
    <title>Telecom Churn Modelling</title>
    <!-- Icons font CSS-->
    <link href="{{ url_for('static', filename='vendor/font-awesome-4.7/css/font-awesome.min.css') }}" rel="stylesheet">
    <link href="{{ url_for('static', filename='vendor/mdi-font/css/material-design-iconic-font.min.css') }}" rel="stylesheet">
    <!-- Font special for pages-->
    <link href="https://fonts.googleapis.com/css?family=Roboto:100,100i,300,300i,400,400i,500,500i,700,700i,900,900i" rel="style
    <!-- Vendor CSS-->
    <link href="{{ url_for('static', filename='vendor/select2/select2.min.css') }}" rel="stylesheet">
    <!-- Main CSS-->
    <link href="{{ url_for('static', filename='css/main.css') }}" rel="stylesheet">
</head>
<body>
    <div class="page-wrapper bg-red p-t-100 p-b-100 font-robo">
        <div class="wrapper wrapper--w960">
            <div class="card card-2">
                <div class="card-heading"></div>
                <div class="card-body">
                    <h2 class="title">Telecom Customer Churn Prediction</h2>
                    <div class="row row-space">
     <p>Customer churn has become highly important for companies because of increasing
        competition among companies, increased importance of marketing strategies and conscious
        behaviour of customers in the recent years. Customers can easily trend toward alternative
        services. Companies must develop various strategies to prevent these possible trends,
```

app.py | base.html* | index.html* | predno.html | predyes.html

```
            <div class="card card-2">
                <div class="card-heading"></div>
                <div class="card-body">
                    <h2 class="title">Telecom Customer Churn Prediction</h2>
                    <div class="row row-space">
     <p>Customer churn has become highly important for companies because of increasing
        competition among companies, increased importance of marketing strategies and conscious
        behaviour of customers in the recent years. Customers can easily trend toward alternative
        services. Companies must develop various strategies to prevent these possible trends,
        depending on the services they provide. During the estimation of possible churns, data from
        the previous churns might be used. An efficient churn predictive model benefits companies in
        many ways. Early identification of customers likely to leave may help to build cost effective
        ways in marketing strategies. Customer retention campaigns might be limited to selected
        customers but it should cover most of the customer. Incorrect predictions could result in a
        company losing profits because of the discounts offered to continuous subscribers.</p>
                        <div class="p-t-30">
                         <img src="../static/images/yes.jpg" width="450" height="200"  >
                        </div>
                       <div class="p-t-30">
                           <form action="/assesment">
       <button class="btn btn--radius btn--green" type="submit">Click me to continue with prediction</button>
                           </form>
                      </div></div>
                </div></div>
         </div></div>
    <!-- Jquery JS-->
    <script src="{{ url_for('static', filename='vendor/jquery/jquery.min.js') }}" type="text/javascript"></script>
    <!-- Vendor JS-->
    <script src="{{ url_for('static', filename='vendor/select2/select2.min.js') }}" type="text/javascript"></script>
    <!-- Main JS-->
    <script src="{{ url_for('static', filename='js/global.js') }}" type="text/javascript"></script>
</body>
</html>
```

Figure 25: Base.html is the page that displays the introduction content.

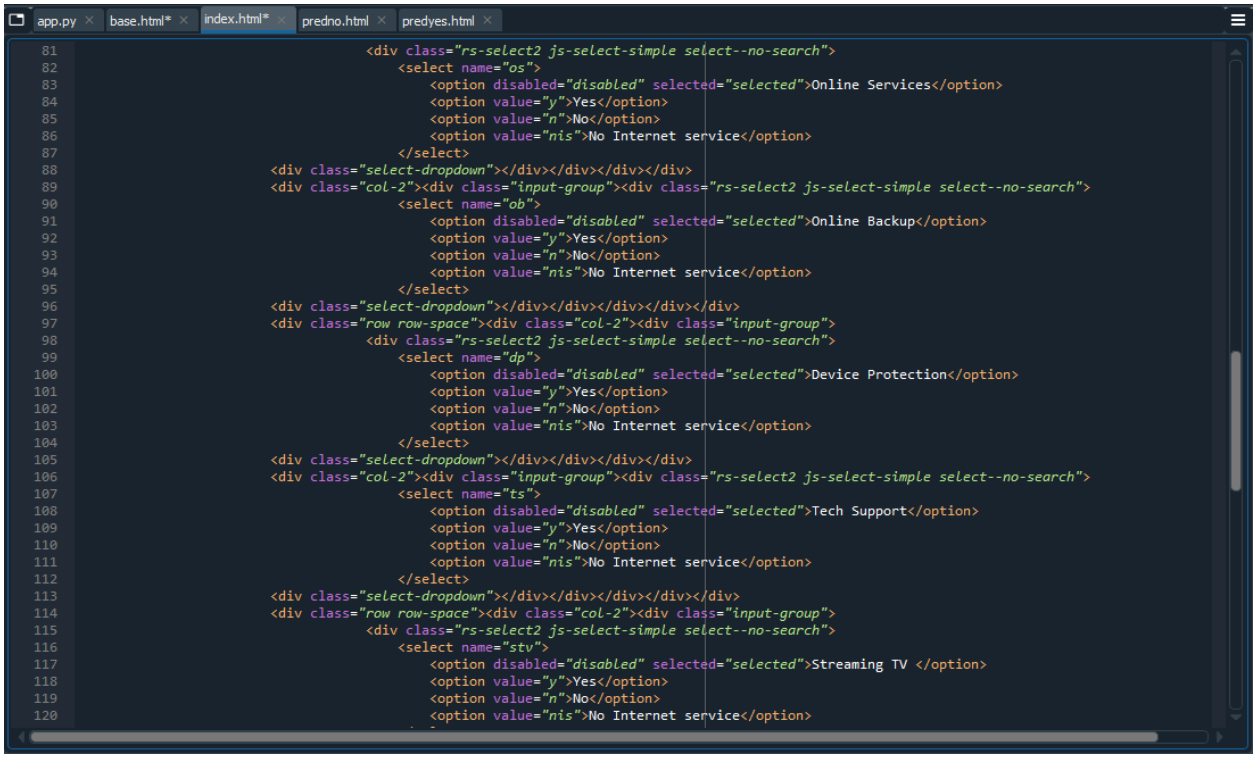
3.INDEX.HTML

app.py | base.html* | index.html* | predno.html | predyes.html

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
    <meta name="description" content="Colorlib Templates">
    <meta name="author" content="Colorlib">
    <meta name="keywords" content="Colorlib Templates">
    <title>Telecom Customer Churn Prediction</title>
    <link href="{{ url_for('static', filename='vendor/font-awesome-4.7/css/font-awesome.min.css') }}" rel="stylesheet">
    <link href="{{ url_for('static', filename='vendor/mdi-font/css/material-design-iconic-font.min.css') }}" rel="stylesheet
    <link href="https://fonts.googleapis.com/css?family=Roboto:100,100i,300,300i,400,400i,500,500i,700,700i,900,900i" rel="stylesheet">
    <link href="{{ url_for('static', filename='vendor/select2/select2.min.css') }}" rel="stylesheet">
    <link href="{{ url_for('static', filename='css/main.css') }}" rel="stylesheet">
</head>
<style>
</style>
<body>
    <div class="page-wrapper bg-red p-t-100 p-b-100 font-robo"><div class="wrapper wrapper--w960">
            <div class="card card-2"><div class="card-heading"></div><div class="card-body">
                    <h2 class="title">Prediction form</h2>
                    <form action="/predict" method="post">
            <div class="row row-space"><div class="col-2"><div class="input-group"> <div class="rs-select2 js-select-simple select--no-search">
                                    <select name="gender">
                                        <option disabled="disabled" selected="selected">Gender</option>
                                        <option value="f">female</option>
                                        <option value="m">Male</option>
                                    </select>
                        <div class="select-dropdown"></div></div> </div></div>
                             <div class="col-2"><div class="input-group"><div class="rs-select2 js-select-simple select--no-search">
                                    <select name="srcitizen">
                                        <option disabled="disabled" selected="selected">Senior Citizen</option>
                                        <option value="y">Yes</option>
                                        <option value="n">No</option>
                                    </select>
                        <div class="select-dropdown"></div></div></div></div></div>
                        <div class="row row-space"><div class="col-2"><div class="input-group">
                        <div class="rs-select2 js-select-simple select--no-search">
                                    <select name="partner">
                                        <option disabled="disabled" selected="selected">Partner</option>
```

app.py | base.html* | index.html* | predno.html | predyes.html

```
                                        <option value="y">Yes</option>
                                        <option value="n">No</option>
                                    </select>
                        <div class="select-dropdown"></div></div></div></div></div>
                            <div class="col-2"><div class="input-group"><div class="rs-select2 js-select-simple select--no-search">
                                    <select name="dependents">
                                        <option disabled="disabled" selected="selected">Dependents</option>
                                        <option value="y">Yes</option>
                                        <option value="n">No</option>
                                    </select>
                        <div class="select-dropdown"></div></div></div></div></div>
                        <div class="row row-space"><div class="col-2">
                                <div class="input-group">
                                    <input class="input--style-2" type="number" placeholder="Tenure" name="tenure">
                        </div></div>
                        <div class="col-2"><div class="input-group"><div class="rs-select2 js-select-simple select--no-search">
                                    <select name="phservices">
                                        <option disabled="disabled" selected="selected">Phone Services</option>
                                        <option value="y">Yes</option>
                                        <option value="n">No</option>
                                    </select>
                        <div class="select-dropdown"></div></div></div></div></div>
                        <div class="row row-space"><div class="col-2"><div class="input-group">
                                    <div class="rs-select2 js-select-simple select--no-search">
                                    <select name="multi">
                                        <option disabled="disabled" selected="selected">Multiple Lines</option>
                                        <option value="y">Yes</option>
                                        <option value="n">No</option>
                                        <option value="nps">No Phone service</option>
                                    </select>
                        <div class="select-dropdown"></div></div></div></div>
                        <div class="col-2"><div class="input-group"><div class="rs-select2 js-select-simple select--no-search">
                                    <select name="is">
                                        <option disabled="disabled" selected="selected">Internet services</option>
                                        <option value="dsl">DSL</option>
                                        <option value="fo">Fibre Optics</option>
                                        <option value="n">No</option>
                                    </select>
                        <div class="select-dropdown"></div></div></div></div></div>
                        <div class="row row-space"><div class="col-2"><div class="input-group">
```



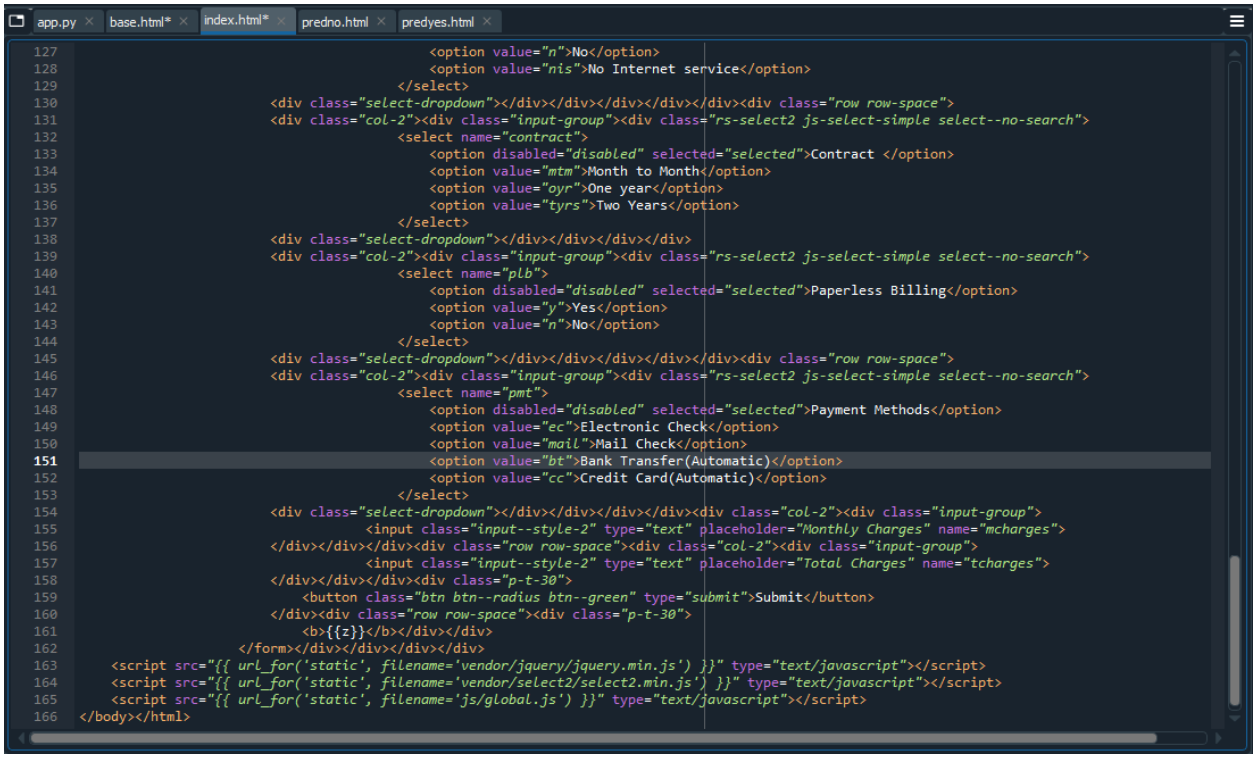


Figure 26: Index.html is the page that displays all the inputs that are needed to be given by the user.

4.PREDNO.HTML:

app.py | base.html* | index.html* | predno.html* | predyes.html

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
    <meta name="description" content="Colorlib Templates">
    <meta name="author" content="Colorlib">
    <meta name="keywords" content="Colorlib Templates">
    <title>Financial Risk Management</title>
    <link href="{{ url_for('static', filename='vendor/font-awesome-4.7/css/font-awesome.min.css') }}" rel="stylesheet">
    <link href="{{ url_for('static', filename='vendor/mdi-font/css/material-design-iconic-font.min.css') }}" rel="stylesheet">
    <link href="https://fonts.googleapis.com/css?family=Roboto:100,100i,300,300i,400,400i,500,500i,700,700i,900,900i" rel="stylesheet">
    <link href="{{ url_for('static', filename='vendor/select2/select2.min.css') }}" rel="stylesheet">
    <link href="{{ url_for('static', filename='css/main.css') }}" rel="stylesheet">
</head>
<style>
.center {
  margin: auto;
  width: 70%;
  padding: 10px;
}
</style>
<body>
    <div class="page-wrapper bg-red p-t-100 p-b-100 font-robo">
        <div class="wrapper wrapper--w960">
            <div class="card card-2">
                <div class="card-heading"></div>
                <div class="card-body">
                    <h2 class="title">Telecom Customer Churn Prediction</h2>
                    <div class="center">
                        <img src="../static/images/stays.jpg" width="400" height="300">
                    </div>
                    <h3 class="title"><p>The Churn prediction says  <b><u>{{z}}</u></b></p> </h3>
    </div></div></div>
    </div>
    <script src="{{ url_for('static', filename='vendor/jquery/jquery.min.js') }}" type="text/javascript"></script>
    <script src="{{ url_for('static', filename='vendor/select2/select2.min.js') }}" type="text/javascript"></script>
    <script src="{{ url_for('static', filename='js/global.js') }}" type="text/javascript"></script>
</body>
</html>
```

Figure 27: Predno.html is the page that displays the NO as output.

5.PREDYES.HTML:

app.py | base.html* | index.html* | predno.html* | predyes.html*

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
    <meta name="description" content="Colorlib Templates">
    <meta name="author" content="Colorlib">
    <meta name="keywords" content="Colorlib Templates">
    <title>Telecom Customer Churn Prediction</title>
    <link href="{{ url_for('static', filename='vendor/font-awesome-4.7/css/font-awesome.min.css') }}" rel="stylesheet">
    <link href="{{ url_for('static', filename='vendor/mdi-font/css/material-design-iconic-font.min.css') }}" rel="stylesheet">
    <link href="https://fonts.googleapis.com/css?family=Roboto:100,100i,300,300i,400,400i,500,500i,700,700i,900,900i" rel="stylesheet">
    <link href="{{ url_for('static', filename='vendor/select2/select2.min.css') }}" rel="stylesheet">
    <link href="{{ url_for('static', filename='css/main.css') }}" rel="stylesheet">
</head>
<style>
.center {
  margin: auto;
  width: 70%;
  padding: 10px;
}
</style>
<body>
    <div class="page-wrapper bg-red p-t-100 p-b-100 font-robo">
        <div class="wrapper wrapper--w960">
            <div class="card card-2">
                <div class="card-heading"></div>
                <div class="card-body">
                    <h2 class="title">Telecom Customer Churn Prediction</h2>
                    <div class="center">
                        <img src="../static/images/churn.jpg" width="400" height="300">
                    </div>
                    <h3 class="title"><p>The Churn Prediction says  <b><u>{{z}}</u></b></p></h3>
    </div></div></div></div>
    </div>
    <script src="{{ url_for('static', filename='vendor/jquery/jquery.min.js') }}" type="text/javascript"></script>
    <script src="{{ url_for('static', filename='vendor/select2/select2.min.js') }}" type="text/javascript"></script>
    <script src="{{ url_for('static', filename='js/global.js') }}" type="text/javascript"></script>
</body><!-- This templates was made by Colorlib (https://colorlib.com) -->
</html>
```

Figure 28: PredYes.html is the page that displays the YES as output.

7. CONCLUSION

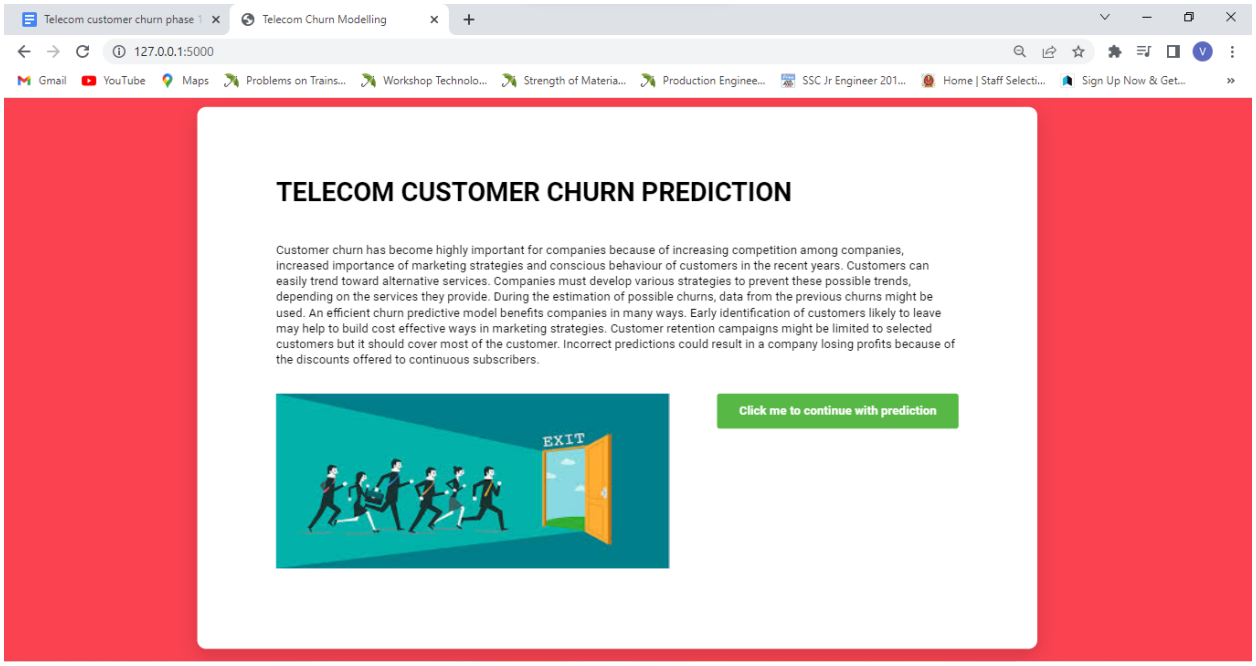


Figure 29:Introduction page.

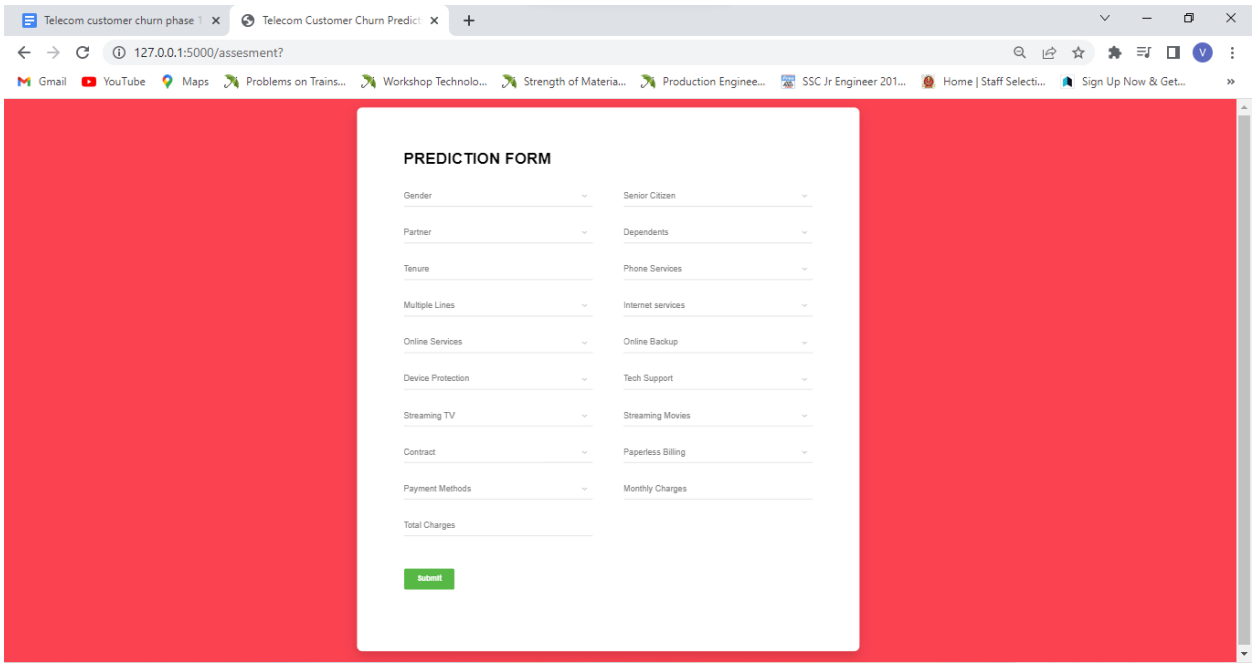


Figure 30: Input pages (Which take the inputs from the User).

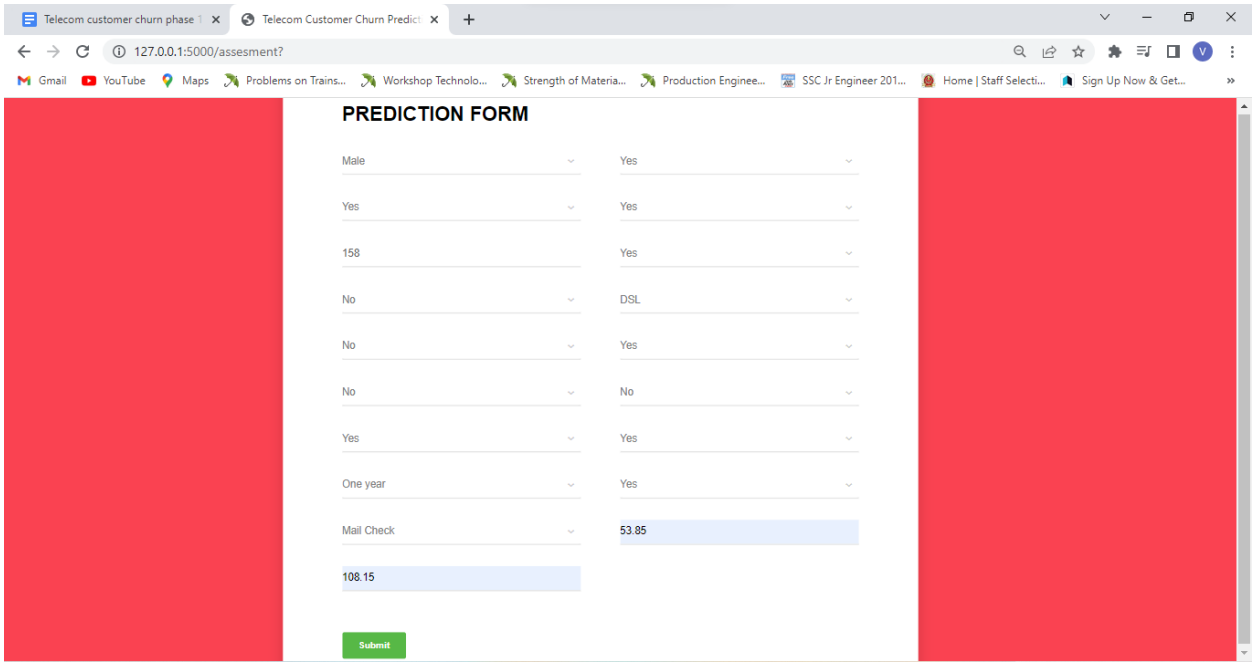


Figure 31: Input pages (Which are given by the User).

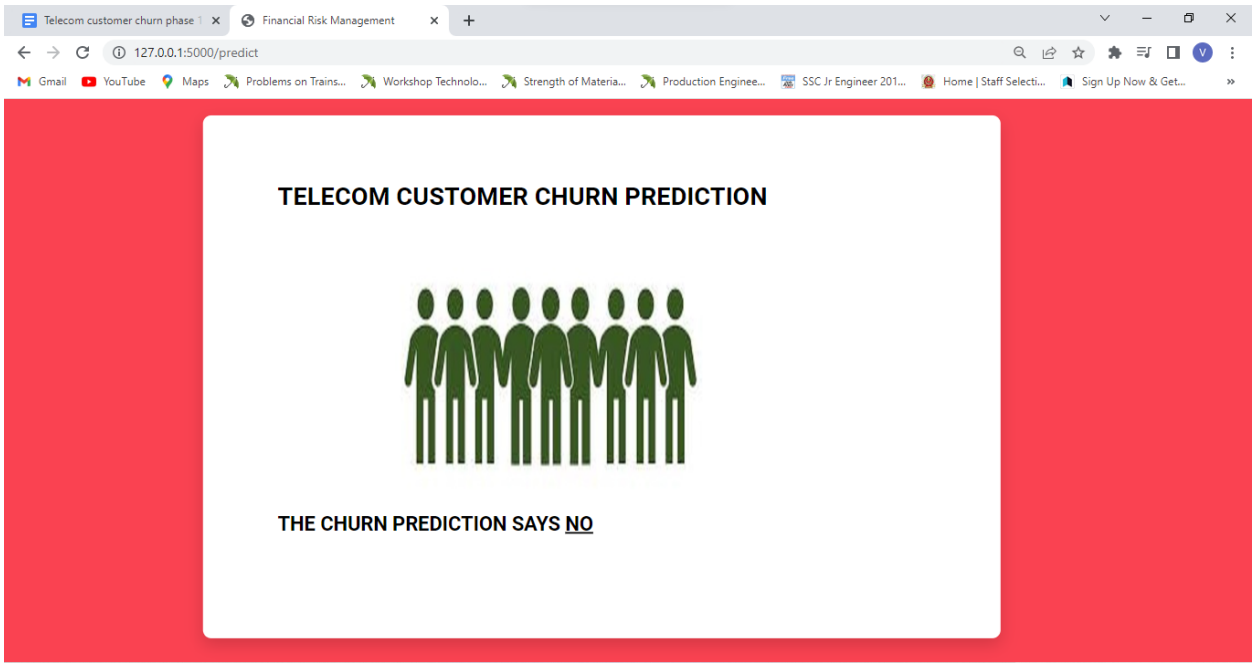


Figure 32: Output page (Displays the Churn Prediction says NO).

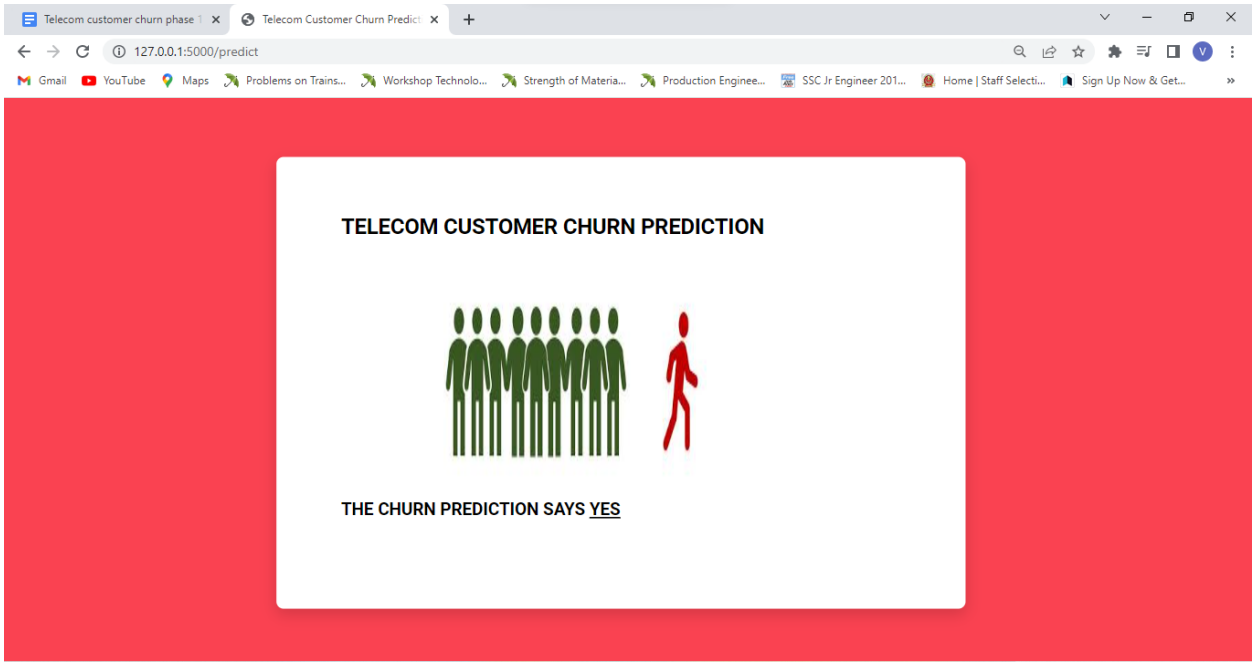


Figure 33: Output page (Displays the Churn Prediction says YES).

8. APPLICATION

The areas where this solution can be applied:

- Banking sectors.
- Saas business.

9. ADVANTAGES

- Having the ability to accurately predict future churn rates is predict future churn rates is necessary because it helps your business gain a better understanding of future expected revenue.
- Predicting churn rates can also help your business identify and improve upon areas where customer service is lacking.

10.DISADVANTAGES

- The insights gained from churn prediction helps them to focus more on the customers that are at a high risk of leaving.
- The output in the case of churn prediction is a simple yes or a no.
- That makes it a classification problem where you have to predict 1 if the customer is likely to churn and 0 otherwise.

11. FUTURE SCOPE

From this entire findings we know fundamental concepts and can work on machine learning.

- Customer churn is one of the major problem which the telecom sector is facing nowadays. It is essential to recognize possible customer churn so that the losses can be avoided.
- In order to maintain a loyal base of customer the service providers in telecom sector aims to retain customers with themselves. Since the costs related with obtaining a new customer is much greater than retaining older customer the prediction of churn becomes even more essential.
- The big data analysis with machine learning makes the churn prediction much easier in telecom sector. Thus, it can be concluded that the big data analytics with machine learning techniques have proven to be accurate and effective to predicts customer churn in nearby future.

Enhancements that can be made in the future:

- This model can be further developed to predict the job position among all the possible positions based on skills a person have.
- We can scope the better job in future with easy experience.

12. BIBILOGRAPHY

[1] Gerpott TJ, Rams W, Schindler A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. Telecommun Policy. 2001.

[2] Wei CP, Chiu IT. Turning telecommunications call details to churn prediction: a data mining approach. Expert Syst Appl. 2002.

[3] Qureshii SA, Rehman AS, Qamar AM, Kamal A, Rehman A. Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management. 2013.

[4] Ascarza E, Iyengar R, Schleicher M. The perils of proactive churn prevention using plan recommendations: evidence from a feld experiment. J Market Res. 2016.

[5] Bott. Predicting customer churn in telecom industry using multilayer preceptron neural networks: modeling and analysis. Igarss. 2014.

[6] Umayaparvathi V, Iyakutti K. A survey on customer churn prediction in telecom industry: datasets, methods and metric. Int Res J Eng Technol. 2016.

[7] Yu W, Jutla DN, Sivakumar SC. A churn-strategy alignment model for managers in mobile telecom. In: Communication networks and services research conference, vol. 3. 2005.

[8] Burez D, den Poel V. Handling class imbalance in customer churn prediction. Expert Syst Appl. 2009..

[9] Zhan J, Guidibande V, Parsa SPK. Identifcation of top-k infuential communities in big networks. J Big Data. 2016.

[10] Barthelemy M. Betweenness centrality in large complex networks. Eur Phys J B. 2004.

[11] Elisabetta E, Meyerhenke H, Staudt CL. Approximating betweenness centrality in large evolving networks. CoRR. 2014..

[12] Brandusoiu I, Toderean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016.

[13] He Y, He Z, Zhang D. A study on prediction of customer churn in fxed communication network based on data mining. In: Sixth international conference on fuzzy systems and knowledge discovery, vol. 1. 2009.

[14] Idris A, Khan A, Lee YS. Genetic programming and adaboosting based churn prediction for telecom. In: IEEE international conference on systems, man, and cybernetics (SMC). 2012.

13. HELP FILE

PROJECT EXECUTION:

STEP-1: Go to Start, search and launch **ANACONDA NAVIGATOR**.

STEP-2: After launching of **ANACONDA NAVIGATOR**, launch **JUPYTER NOTEBOOK**.

STEP-3: Open "Major project code" **IPYNB** file.

STEP-4: Then run all the cells.

STEP-5: All the **data preprocessing, training and testing, model building, accuracy** of the model can be showcased.

STEP-6: And a pickle file will be generated.

STEP-7: Create a Folder named **FLASK** on the **DESKTOP**. Extract the pickle file into this Flask Folder.

STEP-8: Extract all the html files (base.html, index.html, predno.html, predyes.html) and python file(app.py) into the **FLASK** Folder.

STEP-9: Then go back to **ANACONDA NAVIGATOR** and the launch the **SPYDER**.
STEP-10: After launching Spyder, give the path of **FLASK FOLDER** which you have created on the **DESKTOP**.

STEP-11: Open all the app.py and html files present in the Flask Folder.

STEP-12: After running of the app.py, open **ANACONDA PROMPT** and follow the below steps:

cd File Path-> click enter.

python app.py->click enter (we could see running of file).

STEP-13: Then open **BROWSER**, at the URL area type "**localhost:5000**".

STEP-14: Home page of the project will be displayed.

STEP-15: Click on "**CLICK ME TO CONTINUE WITH PREDICTION**". Directly it will be navigated to index page.

STEP-16:A index page will be displayed where the user needs to give the inputs and then click on "**SUBMIT**". Output will be generated "**TELECOM CUSTOMER CHURN PREDICTION"**..