



**PREDICTION OF CUSTOMER CHURNING IN TELECOM INDUSTRY**

***Group Members:***

**Sandipa Bhowmick, CALCUTTA INSTITUTE OF ENGINEERING AND MANAGEMENT, 151650110104**

**Sohom Banerjee, CALCUTTA INSTITUTE OF ENGINEERING AND MANAGEMENT, 151650110113**

**Soumita Dutta, CALCUTTA INSTITUTE OF ENGINEERING AND MANAGEMENT, 151650110115**

**Souvik Datta, CALCUTTA INSTITUTE OF ENGINEERING AND MANAGEMENT, 151650110117**

**Souvik Mitra, CALCUTTA INSTITUTE OF ENGINEERING AND MANAGEMENT, 151650110118**

**Swarup Kumar Das, KALYANI GOVERNMENT ENGINEERING COLLEGE, 151020120015**

**Table of Contents**

Acknowledgement....................................................................................................................4

Project Objective.......................................................................................................................5

Project Scope.............................................................................................................................6

Requirement Specification........................................................................................................7

Screenshots...............................................................................................................................8

Future Scope of Improvements...............................................................................................11

Code........................................................................................................................................12

Project Certificate...................................................................................................................13

**Acknowledgement**

I take this opportunity to express my profound gratitude and deep regards to my faculty, **Prof. Arnab Chakraborty** for his exemplary guidance, monitoring and constantencouragement throughout the course of this project. The blessing, help and guidance given by him time to time shall carry me a long way in the journey of life on which I am about to embark.

I am obliged to my project team members for the valuable information provided by them in their respective fields. I am grateful for their cooperation during the period of my assignment.

*(Sandipa Bhowmick,*

*Sohom Banerjee,*

*Soumita Dutta,*

*Souvik Datta,*

*Souvik Mitra,*

*Swarup Kumar Das)*

**Project Objective**

In this project we have a public dataset that consists of a customer usage pattern and the churning status of the customer.

***Target Variable:***

*Churn:* if the customer has churned (1=yes; 0 = no)

***The predictors influencing the churn score are as follows:***

* Account Length
* International Plan
* Voice mail plan
* Number of voice mail messages
* Total day minutes used
* Day calls made
* Total day charge
* Total evening minutes
* Total night calls
* Total night charge
* Total international minutes used
* Total international calls made
* Total international charge
* Number of Customer Service calls made

***Goal:***

* Predict the whether there will be customer churn based on the other features.
* Prepare dataset and perform **K-Fold Cross Validation**.
* Create three different types of models from the data – **Naive Bayes Classifier**, **KNN Classifier** and **Random Forest Classifier** based on the training set.
* Apply on the test set and **compare the differences in the accuracies** of the different models
* **Write the KNN algorithm** in Python from scratch, and apply it again on the data. Compute the accuracy and the confusion matrix.

**Project Scope**

The broad scope of the **Prediction of Customer Churning in Telecom Industry** project includes:

* In this project we analysed a dataset of customer usage pattern in a certain telecom enterprise. The dataset consists of various other factors that influence both the customer usage and customer churn.
* By this project we can only predict the churn; we cannot provide any solution to prevent the churning.

**Requirement Specification**

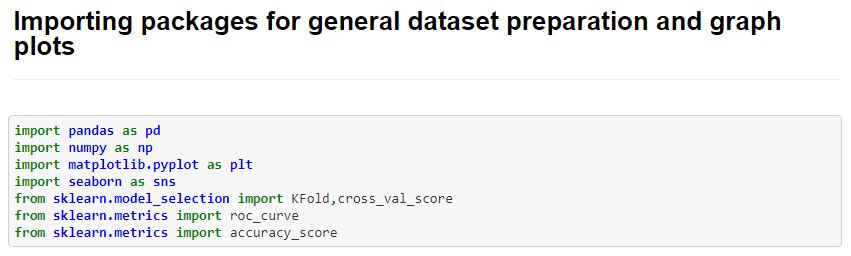
Hardware requirements:

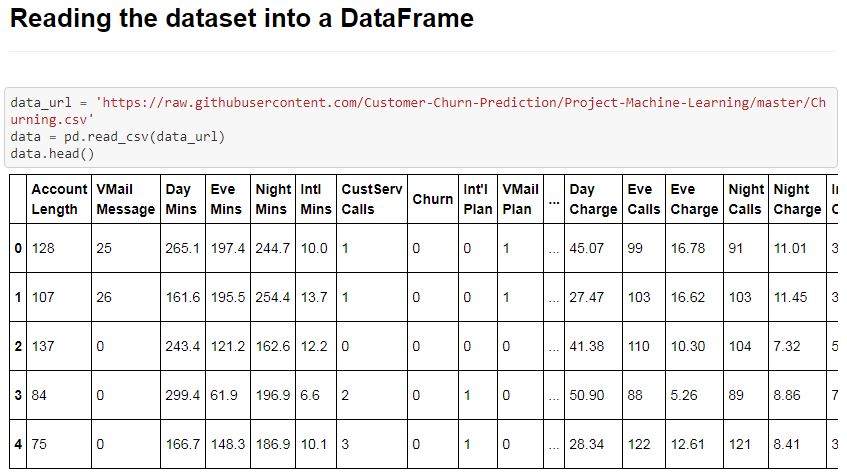
* CPU: 2 x 64-bit 2.8 GHz 8.00 GT/s CPUs
* RAM: 2 GB RAM (recommended 4 GB RAM)
* Storage: 2 GB for installation of Anaconda Navigator.
* Internet access to download the files from Anaconda Cloud or a USB drive containing all of the files you need with alternate instructions for air gapped installations.

Software requirements:

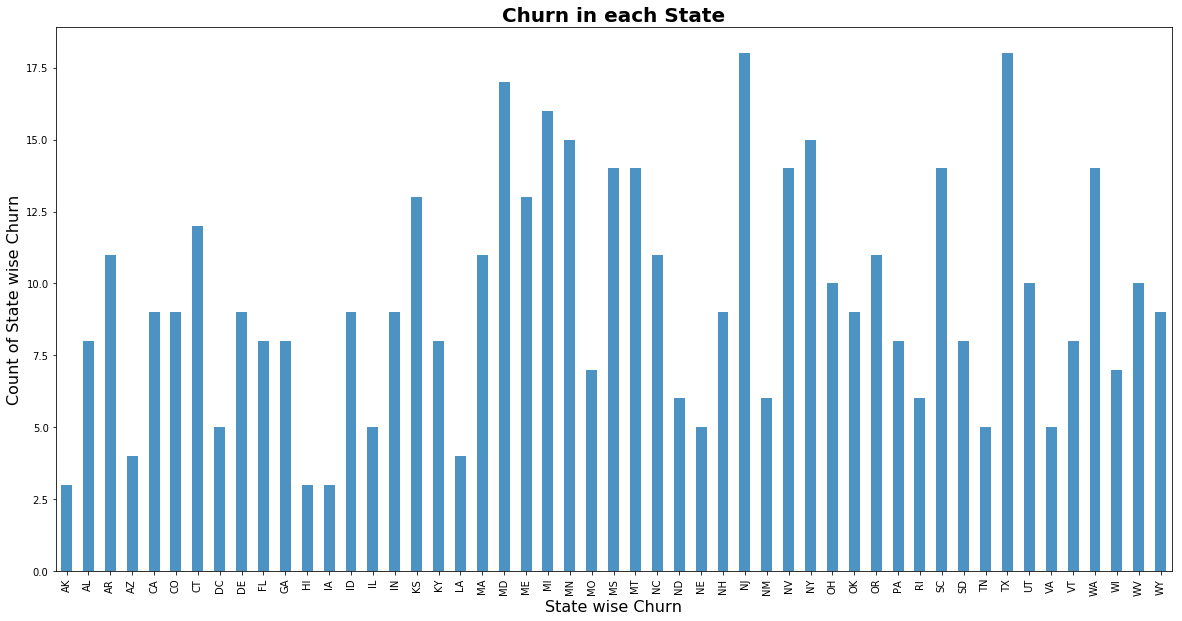
* Anaconda Navigator v3.6.4
* SciKit learn package for Python
* Any web browser like Google Chrome

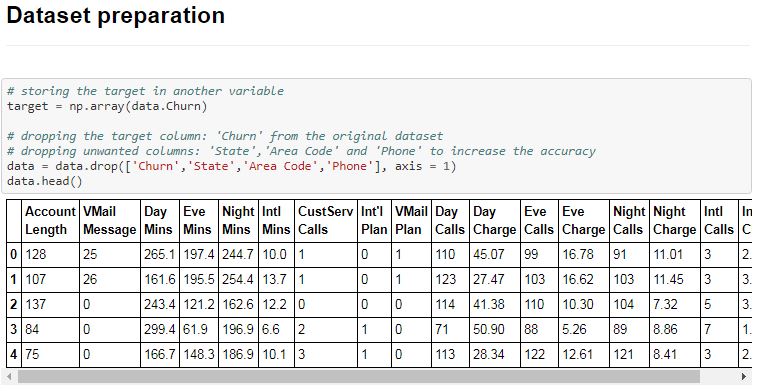
**Screenshots**

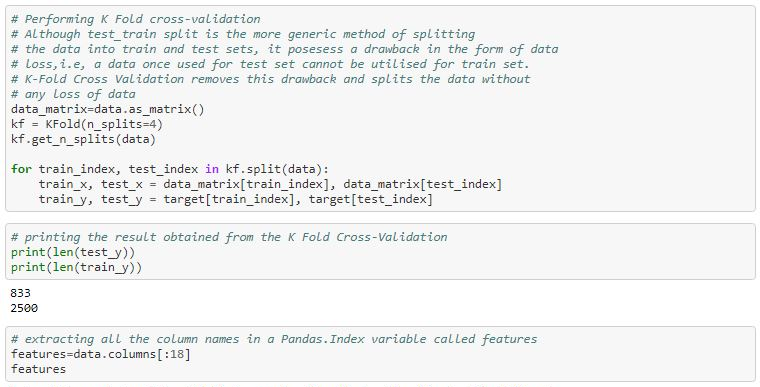


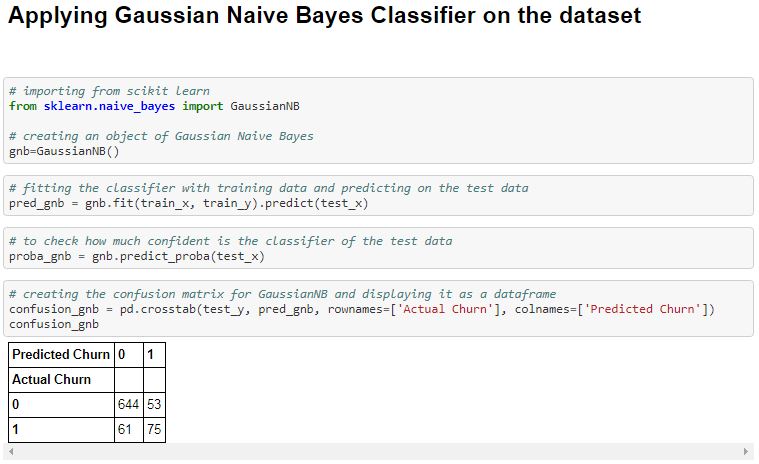


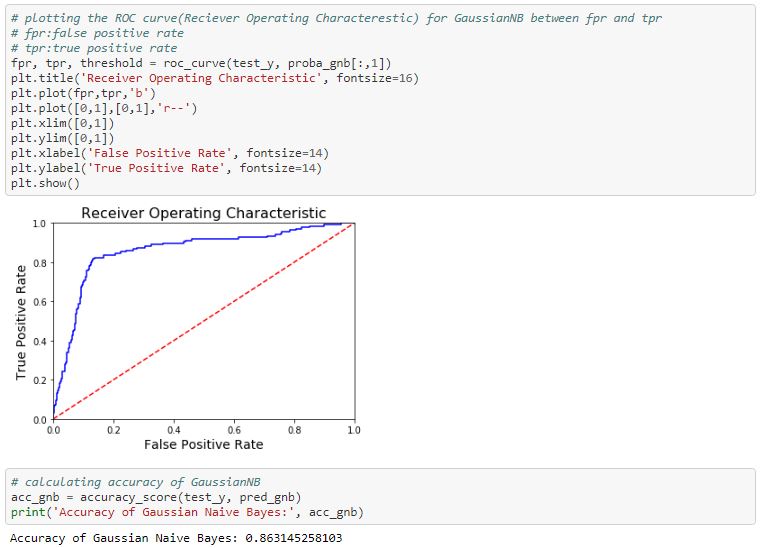


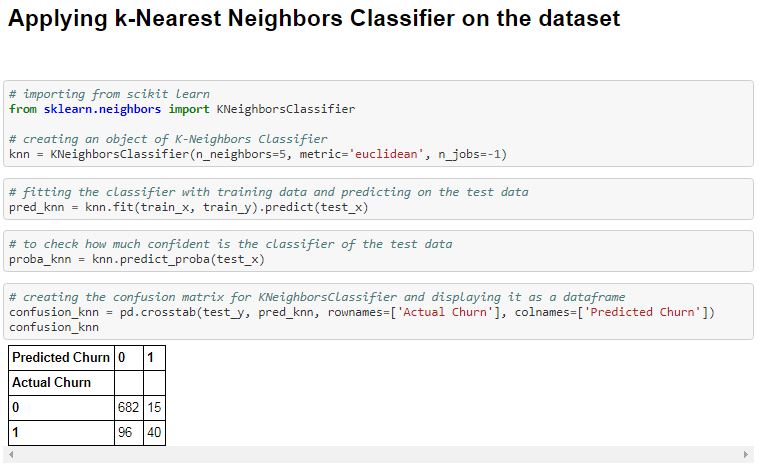


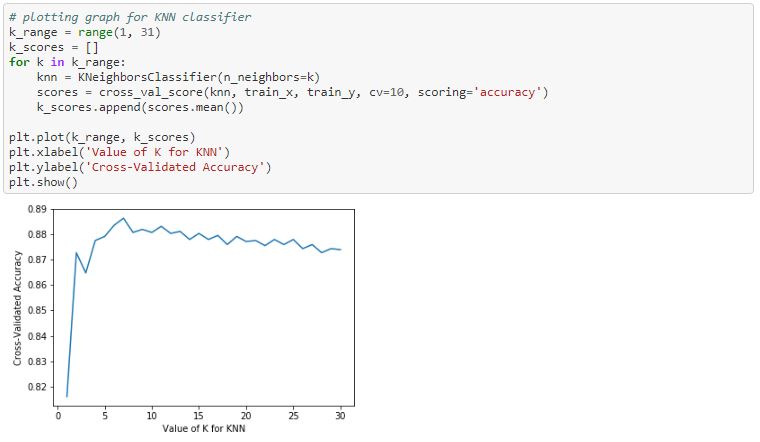


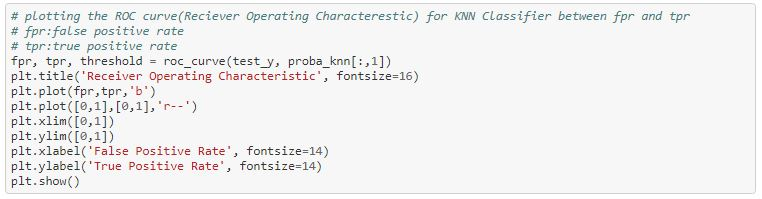


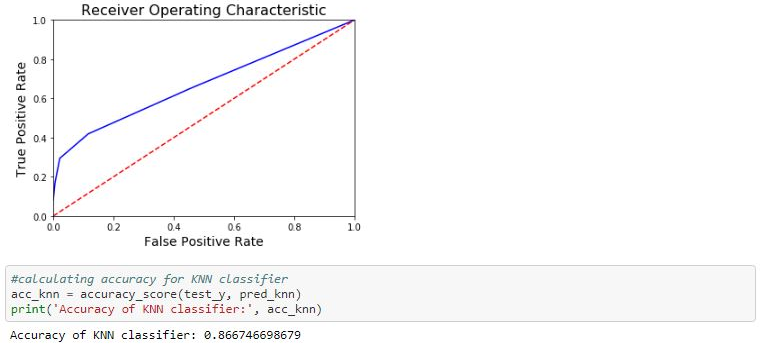


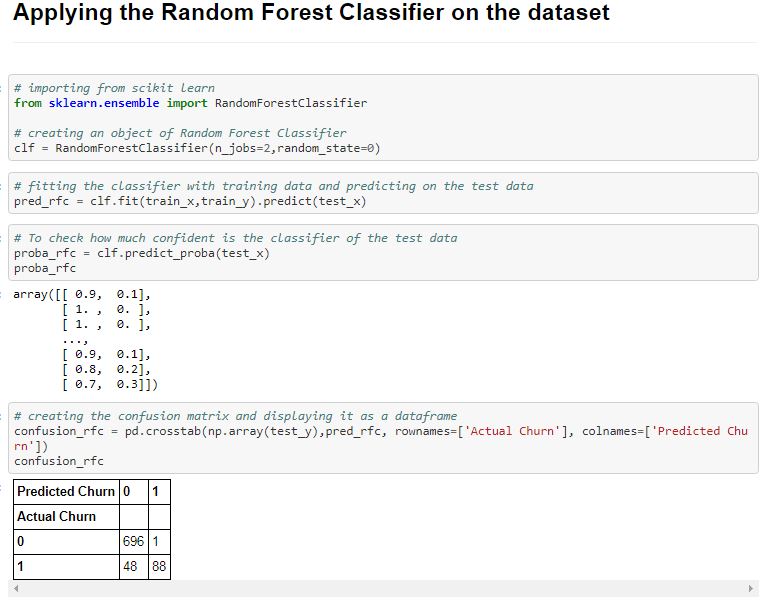




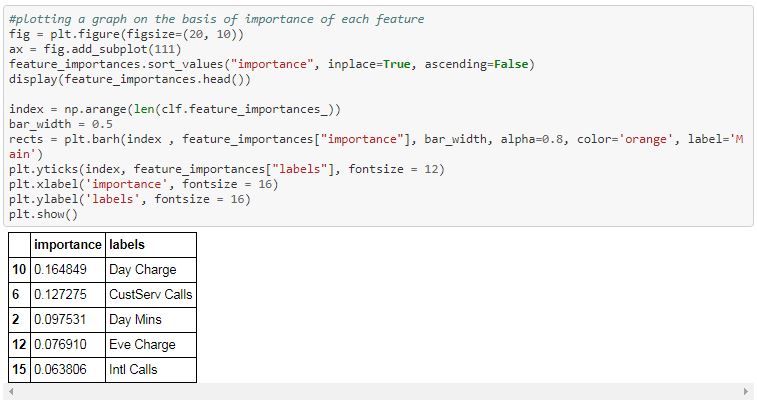


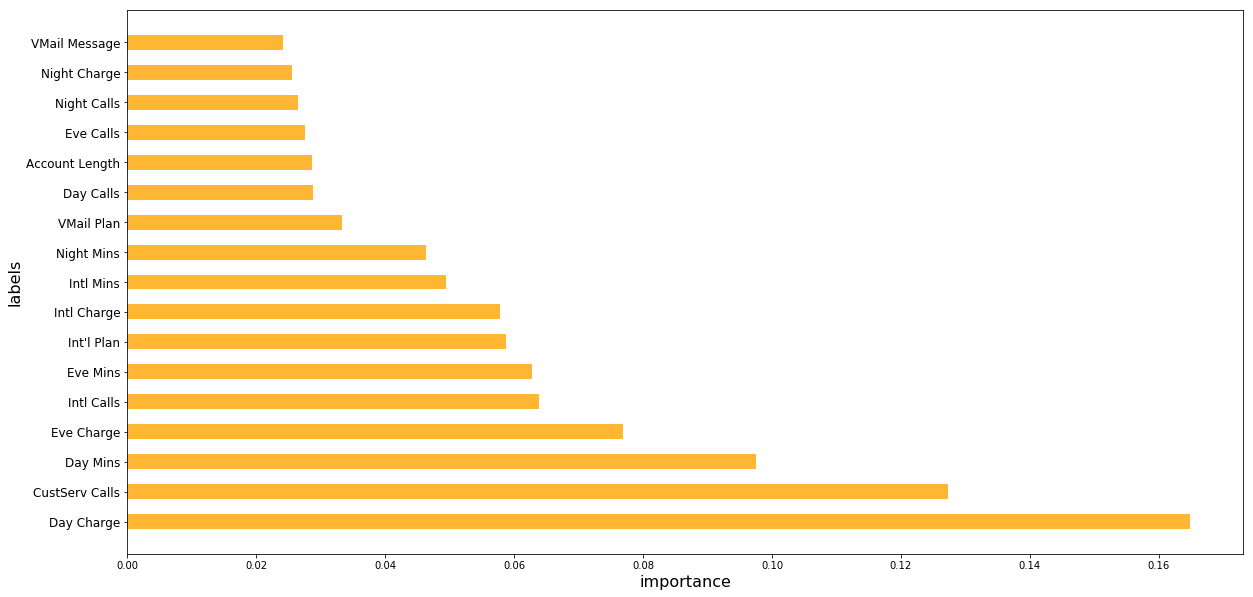


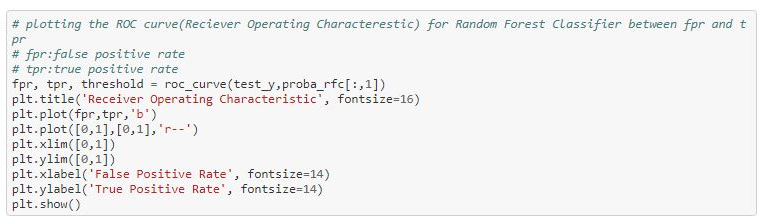


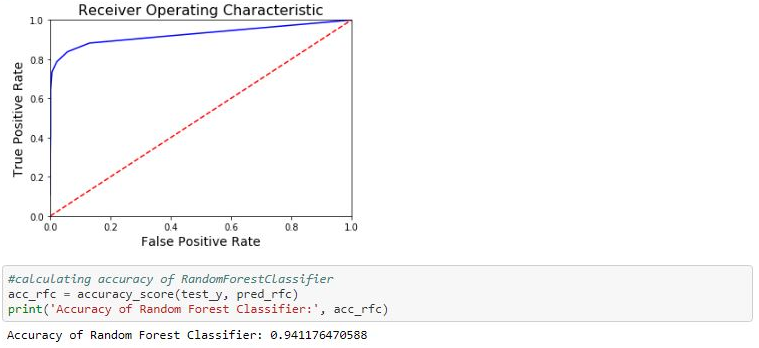


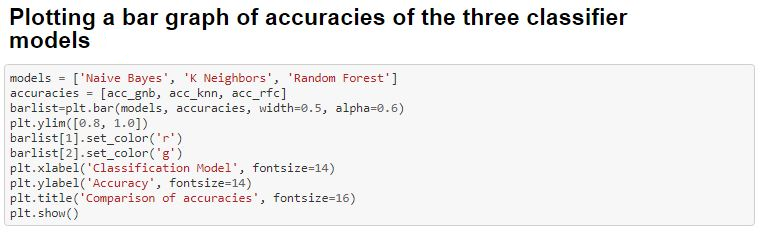


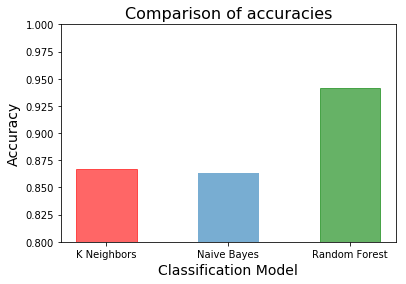




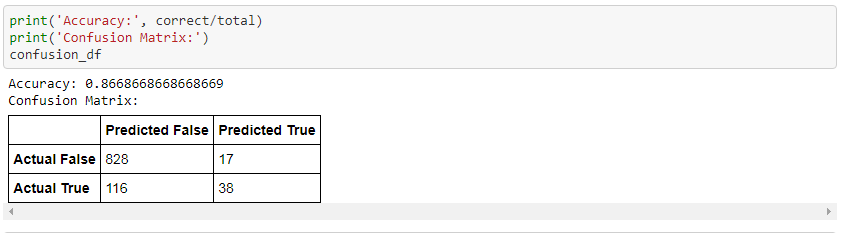








**Computed accuracy and confusion matrix from the KNN algorithm written from scratch**

****

**Code**

*# all imports required*

**import** **numpy** **as** **np**

**from** **collections** **import** Counter

**import** **pandas** **as** **pd**

**import** **random**

*# defining the k nearest neighbors classifier algorithm*

*# we pass the training data and the test data as list*

*# we pass k number of neighbors with 3 neighbors as default*

**def** k\_neighbors\_classifier(train\_set, test\_set, k=3):

*# declaring a list for calculating and storing euclidean distances*

distances = []

*# calculating the euclidean distances and storing it in the list 'distances'*

**for** group **in** train\_set:

**for** features **in** train\_set[group]:

euclidean\_dist = np.linalg.norm(np.array(features)-np.array(test\_set))

*# calculated distances are stored in the list along with their respective groups*

distances.append([euclidean\_dist, group])

*# sorting the 'distances' and storing the names of the k nearest neighboring groups in a list*

neighbors = [i[1] **for** i **in** sorted(distances)[:k]]

*# prediction of the most common neighbor and finding the probability of the prediction*

*# the most\_common() method stores the data in a list of a tuple,*

*# in this case we could write the tuple of list as: [(group\_name, frequency)]*

*# therefore we choose the first element as the predicted group*

prediction = Counter(neighbors).most\_common(1)[0][0]

*# we know k is the total no. of neighbors*

*# so we find the probability of our predicted group by division: (frequency / k)*

probability = Counter(neighbors).most\_common(1)[0][1] / k

**return** prediction, probability

*# url of the source file*

url = 'https://raw.githubusercontent.com/Customer-Churn-Prediction/Project-Machine-Learning/master/Churning.csv'

*# reading the source file and storing it in a pandas DataFrame*

df = pd.read\_csv(url)

*# the 'Churn' column is not the last column in the original dataset*

*# appending the 'Churn' data i.e the target containing the two groups 0 and 1 as the last column*

*# storing the 'Churn' in another DataFrame variable*

target = df['Churn']

*# deleting unnecessary columns and the target column from the DataFrame*

df.drop(['Churn', 'Phone', 'State','Area Code'], axis = 1, inplace=**True**)

*# now we append the target data as the last column*

df['Churn'] = target

*# storing the datas present in the DataFrame as list*

data\_list = df.astype(float).values.tolist()

*# shuffling the data so that the training and testing data can be chosen at random*

random.shuffle(data\_list)

*# setting the percentage of test data as 30%*

test\_size = 0.3

*# train\_set and test\_set are two dictionaries and their keys represent the two groups or classes*

*# here it is 0(churn = False) or 1(churn = True)*

train\_set, test\_set = {0:[], 1:[]}, {0:[], 1:[]}

*# splitting the data into train and test sets*

*# selecting from the beginning upto the last 30% of data i.e. the first 70%*

train\_data = data\_list[:-int(test\_size\*len(data\_list))]

*# selecting the last 30% of the data*

test\_data = data\_list[-int(test\_size\*len(data\_list)):]

*# storing the grouping the data according to their group or class 0 or 1*

**for** i **in** train\_data:

train\_set[i[-1]].append(i[:-1])

**for** i **in** test\_data:

test\_set[i[-1]].append(i[:-1])

correct = 0

total = 0

confusion\_matrix = [[0, 0], [0, 0]]

*# calculating the accuracy and computing the confusion matrix*

**for** group **in** test\_set:

**for** data **in** test\_set[group]:

vote,confidence = k\_neighbors\_classifier(train\_set, data, k=5)

**if** group == vote:

correct += 1

confusion\_matrix[group][group] += 1

**else**:

confusion\_matrix[group][vote] += 1

total += 1

confusion\_df = pd.DataFrame(data=confusion\_matrix,

columns=['Predicted False', 'Predicted True'],

index = ['Actual False', 'Actual True'])

print('Accuracy:', correct/total)

print('Confusion Matrix:')

**Output:**

Accuracy: 0.8668668668668669

Confusion Matrix:

Predicted False Predicted True

Actual False 828 17

Actual True 116 38

**Certificate**

This is to certify that Ms. Sandipa Bhowmick of Calcutta Institute of Engineering and Management, registration number: 151650110104, hassuccessfully completed a project on Prediction of Customer Churning in Telecom Industry using Machine Learning with Python under the guidance of Prof. Arnab Chakraborty.

--------------------------------------------------

Prof. Arnab Chakraborty

**Globsyn Finishing School**

**Certificate**

This is to certify that Mr. Sohom Banerjee of Calcutta Institute of Engineering and Management, registration number: 151650110113, hassuccessfully completed a project on Prediction of Customer Churning in Telecom Industry using Machine Learning with Python under the guidance of Prof. Arnab Chakraborty.

--------------------------------------------------

Prof. Arnab Chakraborty

**Globsyn Finishing School**

**Certificate**

This is to certify that Ms. Soumita Dutta of Calcutta Institute of Engineering and Management, registration number: 151650110115, hassuccessfully completed a project on Prediction of Customer Churning in Telecom Industry using Machine Learning with Python under the guidance of Prof. Arnab Chakraborty.

--------------------------------------------------

Prof. Arnab Chakraborty

**Globsyn Finishing School**

**Certificate**

This is to certify that Mr. Souvik Datta of Calcutta Institute of Engineering and Management, registration number: 151650110117, hassuccessfully completed a project on Prediction of Customer Churning in Telecom Industry using Machine Learning with Python under the guidance of Prof. Arnab Chakraborty.

--------------------------------------------------

Prof. Arnab Chakraborty

**Globsyn Finishing School**

**Certificate**

This is to certify that Mr. Souvik Mitra of Calcutta Institute of Engineering and Management, registration number: 151650110118, hassuccessfully completed a project on Prediction of Customer Churning in Telecom Industry using Machine Learning with Python under the guidance of Prof. Arnab Chakraborty.

--------------------------------------------------

Prof. Arnab Chakraborty

**Globsyn Finishing School**

**Certificate**

This is to certify that Mr. Swarup Kumar Das of Kalyani Government Engineering College, registration number: 151020120015, hassuccessfully completed a project on Prediction of Customer Churning in Telecom Industry using Machine Learning with Python under the guidance of Prof. Arnab Chakraborty.

--------------------------------------------------

Prof. Arnab Chakraborty

**Globsyn Finishing School**