AMRITA
VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY UNDER SECTION 3 OF THE UGC ACT, 1956
AMRITA SCHOOL OF ENGINEERING, BANGALORE
AMRITA VISHWA VIDYAPEETHAM
BANGALORE

Name: CHARAN KALYAN

Reg.No: BL.EN.U4ECE19072

Branch: Electronics and Communication Engineering

Sem: VI

Section: B

Course Code: 19ECE363

Course Name: Machine Learning

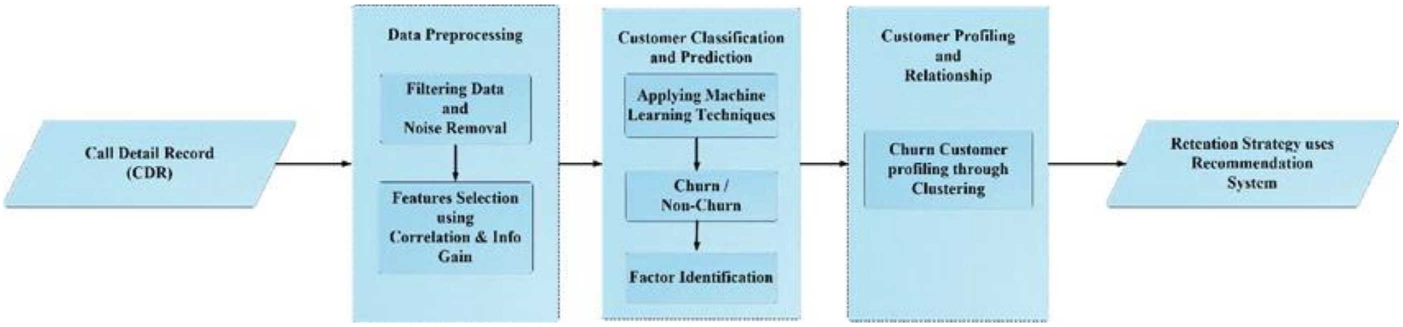
CHURN PREDICTION

Objectives:

CUSTOMER CHURN PREDICTION ANALYSIS IS USED TO UNDERSTAND WHY CUSTOMERS DONT RETURN FOR REPEAT BUSINESS OR CANCEL THE SUBSCRIPTION.

Introduction:

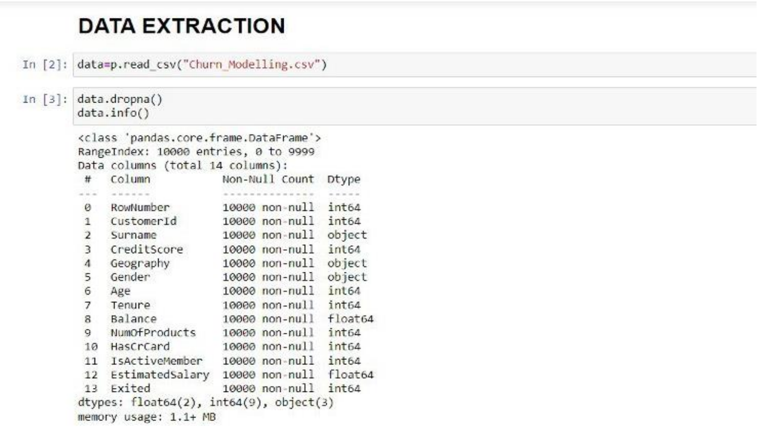
Customer churn is one of the main problems in the telecommunications industry. Several studies have shown that attracting new customers is much more expensive than retaining existing ones. Therefore, companies are focusing on developing accurate and reliable predictive models to identify potential customers that will churn in the near future.


Call Detail Record (CDR)
Data Preprocessing
Filtering Data and Noise Removal
Features Selection using Correlation & Info Gain
Customer Classification and Prediction
Applying Machine Learning Techniques
Churn / Non-Churn
Factor Identification
Customer Profiling and Relationship
Churn Customer profiling through Clustering
Retention Strategy uses Recommendation System

Methodology:

To find a possible solution to the problem of churn prediction. i.e., successfully apply a machine learning to the available data, one needs a deep understanding of the business rules of the telecommunications company and their specificity.

Once a final dataset is derived, the classification algorithms can be successfully trained and their performances correspondingly evaluated.

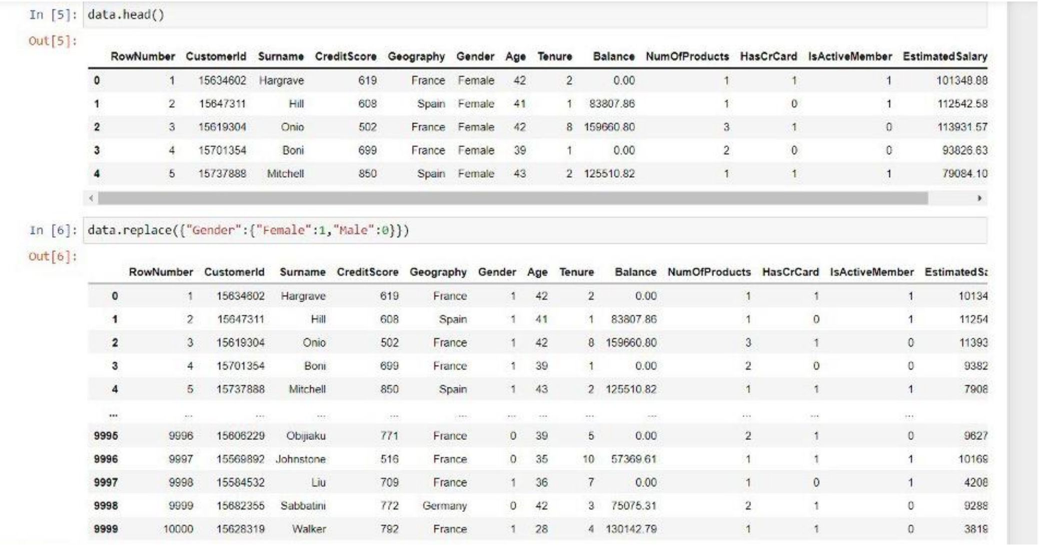


DATA EXTRACTION

```
In [2]: data=p.read_csv("Churn_Modelling.csv")
```

```
In [3]: data.dropna()
        data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   RowNumber        10000 non-null  int64  
 1   CustomerId       10000 non-null  int64  
 2   Surname          10000 non-null  object 
 3   CreditScore      10000 non-null  int64  
 4   Geography        10000 non-null  object 
 5   Gender           10000 non-null  object 
 6   Age              10000 non-null  int64  
 7   Tenure           10000 non-null  int64  
 8   Balance          10000 non-null  float64
 9   NumOfProducts    10000 non-null  int64  
 10  HasCrCard        10000 non-null  int64  
 11  IsActiveMember   10000 non-null  int64  
 12  EstimatedSalary  10000 non-null  float64
 13  Exited           10000 non-null  int64  
dtypes: float64(2), int64(9), object(3)
memory usage: 1.1+ MB
```



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

Out[5]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10

```
In [6]: data.replace({"Gender":{"Female":1,"Male":0}})
```

Out[6]:

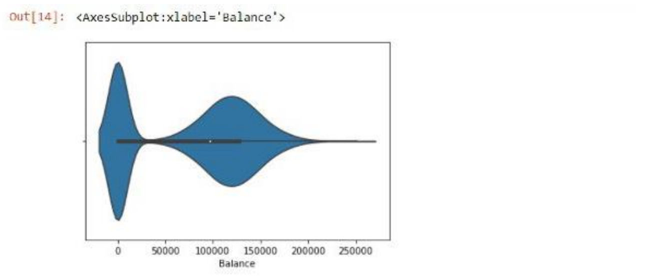
	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSa
0	1	15634602	Hargrave	619	France	1	42	2	0.00	1	1	1	10134
1	2	15647311	Hill	608	Spain	1	41	1	83807.86	1	0	1	11254
2	3	15619304	Onio	502	France	1	42	8	159660.80	3	1	0	11393
3	4	15701354	Boni	699	France	1	39	1	0.00	2	0	0	9382
4	5	15737888	Mitchell	850	Spain	1	43	2	125510.82	1	1	1	7908
...	...	...	...	...	...	...	...	...	...	...	...	...	
9995	9996	15606229	Obijiaku	771	France	0	39	5	0.00	2	1	0	9627
9996	9997	15569892	Johnstone	516	France	0	35	10	57369.61	1	1	1	10169
9997	9998	15584532	Liu	709	France	1	36	7	0.00	1	0	1	4208
9998	9999	15682355	Sabbatini	772	Germany	0	42	3	75075.31	2	1	0	9288
9999	10000	15628319	Walker	792	France	1	28	4	130142.79	1	1	0	3819

The algorithm that has been used is the supervised learning's algorithm decisiontree classifier

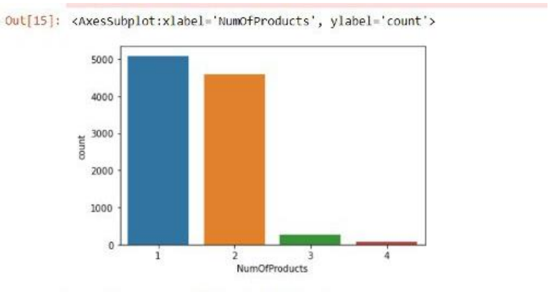
Tools being used are Jupyter notebook

Python Programming Language is used

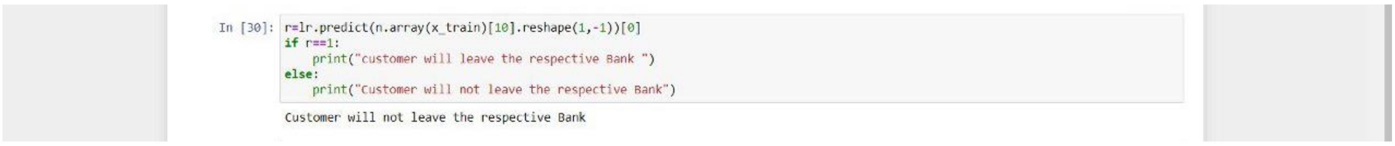
Result:



```
Out[14]: <AxesSubplot:xlabel='Balance'>
```



```
Out[15]: <AxesSubplot:xlabel='NumOfProducts', ylabel='count'>
```



```
In [30]: r=lr.predict(n.array(x_train)[10].reshape(1,-1))[0]
         if r==1:
             print("customer will leave the respective Bank ")
         else:
             print("Customer will not leave the respective Bank")

         Customer will not leave the respective Bank
```

Conclusion:

To attain the output of the objective the required dependencies are imported as pandas, NumPy, matplotlib etc. Loaded data from csv files namely Churn_Modelling.csv. Combined both the csv files based on user_id. Then the dataset has been checked for missing values. Later the plots of the features in the data had been plotted and a heatmap for correlation has been constructed. Features and targets have been selected from the data set. The whole data set will be divided into four arrays x_train, y_train, x_test, y_test. x_train and x_test will be the containing the features selected. Y_train and y_test will be containing the target. The model has been trained using x_train and y_train. Further the model is evaluated using x_test and the mean absolute error between the y_test array and predicted will be calculated.

Reference :

https://www.kaggle.com/code/kmalit/bank-customer-churn-prediction