customer Churn Prediction

Contents

[Abstract 3](#_Toc533419548)

[What has been Provided by the customer (bank): 3](#_Toc533419549)

[Why this problem is important: 3](#_Toc533419550)

[Common business applications: 4](#_Toc533419551)

[Description of the problem 4](#_Toc533419552)

[Related Work 4](#_Toc533419553)

[Description of dataset. 5](#_Toc533419554)

[Pre-processing steps: 6](#_Toc533419555)

[Exploratory Data Analysis steps: 7](#_Toc533419556)

[Box Plot: 8](#_Toc533419557)

[Bar plots: 8](#_Toc533419558)

[Split Datasets: 9](#_Toc533419559)

[Training Dataset: 9](#_Toc533419560)

[Validation Dataset: 9](#_Toc533419561)

[Test Dataset: 9](#_Toc533419562)

[Feature selection: 9](#_Toc533419563)

[Understanding Outcome [Exited]: 10](#_Toc533419564)

[Metrics 11](#_Toc533419565)

[Building (Training) Model: 11](#_Toc533419566)

[Logistic Regression: 11](#_Toc533419567)

[**Prediction on test data** 11](#_Toc533419568)

[**Accuracy & Confusion matrix** 11](#_Toc533419569)

[**True negative, false negative, true positive & false positive - Confusion matrix** 11](#_Toc533419570)

[**ROC curve** 12](#_Toc533419571)

[Naïve Bayes Classifier: 12](#_Toc533419572)

[**Prediction on test data** 12](#_Toc533419573)

[**Accuracy & Confusion matrix** 13](#_Toc533419574)

[**True negative, false negative, true positive & false positive - Confusion matrix** 13](#_Toc533419575)

[**ROC curve** 13](#_Toc533419576)

[Support Vector Machines (SVM): 13](#_Toc533419577)

[**Prediction on test data** 14](#_Toc533419578)

[**Accuracy & Confusion matrix** 14](#_Toc533419579)

[**True negative, false negative, true positive & false positive - Confusion matrix** 14](#_Toc533419580)

[Decision Trees: 15](#_Toc533419581)

[**Prediction on test data** 15](#_Toc533419582)

[**Accuracy & Confusion matrix** 16](#_Toc533419583)

[**True negative, false negative, true positive & false positive - Confusion matrix** 16](#_Toc533419584)

[**ROC curve** 16](#_Toc533419585)

[K-nearest Neighbours(KNN): 17](#_Toc533419586)

[**Prediction on test data** 17](#_Toc533419587)

[**Accuracy & Confusion matrix** 17](#_Toc533419588)

[**True negative, false negative, true positive & false positive - Confusion matrix** 17](#_Toc533419589)

[**ROC curve** 18](#_Toc533419590)

[Neural Networks: 18](#_Toc533419591)

[**Prediction on test data** 18](#_Toc533419592)

[**Accuracy & Confusion matrix** 19](#_Toc533419593)

[**True negative, false negative, true positive & false positive - Confusion matrix** 19](#_Toc533419594)

[**ROC curve** 19](#_Toc533419595)

[**ROC Comparison of 6 models:** 20](#_Toc533419596)

[Implementation Details: 21](#_Toc533419597)

[Deployment process: 21](#_Toc533419598)

[Hardware requirements: 21](#_Toc533419599)

[Software requierements: 21](#_Toc533419600)

[Detailed steps: 21](#_Toc533419601)

[Hardware/software requirements for client installation: 21](#_Toc533419602)

[Conclusion: 21](#_Toc533419603)

[Future Work 22](#_Toc533419604)

[References 22](#_Toc533419605)

# Abstract

Customers of a big international bank decided to leave the bank. The bank is investigating a very high rate of customer leaving the bank. The dataset contains 10000 records, and we use it to investigate and predict which of the customers are more likely to leave the bank soon. The approach here is supervised classification; the classification model to be built on historical data and then used to predict the classes for the current customers to identify the churn. The dataset contains 13 features, and also the label column (Exited or not). The best accuracy was obtained with the Naïve Bayes model (83.29%). Such churn prediction models could be very useful for applications such as churn prediction in Telecom sector to identify the customers who are switching from current network, and also for Churn prediction in subscription services.

# What has been Provided by the customer (bank):

* The customers dataset with historical data.
* From the column [Exited], we can identify the customers who exited from the bank and who are continuing the services of bank. The value 1 represents the customers who are exited from the bank and 0 represents the customers who are continuing with the bank.

# Why this problem is important:

* Using the solution to this problem, the bank can easily identify the customers who are willing to exit the bank soon.
* From the larger datasets, the bank can easily identify the churn customers using machine learning approach, thus this can reduce the manual intervention and the cost to the bank. Using machine learning solutions, the bank can save processing time and manual intervention to investigate the complete records. The system can take quicker decisions with statistical models with optimal accuracy metrics.
* With the investigation of the customers who are churning soon, the bank has an option to reduce the churn by further investigating the reason of leaving the bank and to convince the customers by providing or improvising the services rendered to them.

# Common business applications:

* The churn model can be integrated with the call center / business software, so that the proper discounting can be provided to the identified customers. Targeted marketing strategies can be used.
* Monitoring of the customer trends and building the alerting mechanism to the business users on a daily / monthly basis.
* This problem can be applicable to any industry (for e.g., Telecom) to identify the churn customers within their organizations.

# Description of the problem

* The leading international bank wants to investigate the reason for the customers leaving the bank. The bank needs an automated way to predict the customers who are more likely to leave the bank soon.
* The manual intervention to identify the customers who are churning costs the bank in terms of money and the effort to identify the large scale.
* So, the bank is looking for a solution with the machine learning algorithms to best fit their historical datasets and to predict the current customers who are likely to churn.
* Given: various features of a customer including surname, credit score, geography, gender, age, tenure, balance, number of products, has credit card or not, is active member, estimated salary.
* Predict: is he/she likely to leave the bank.

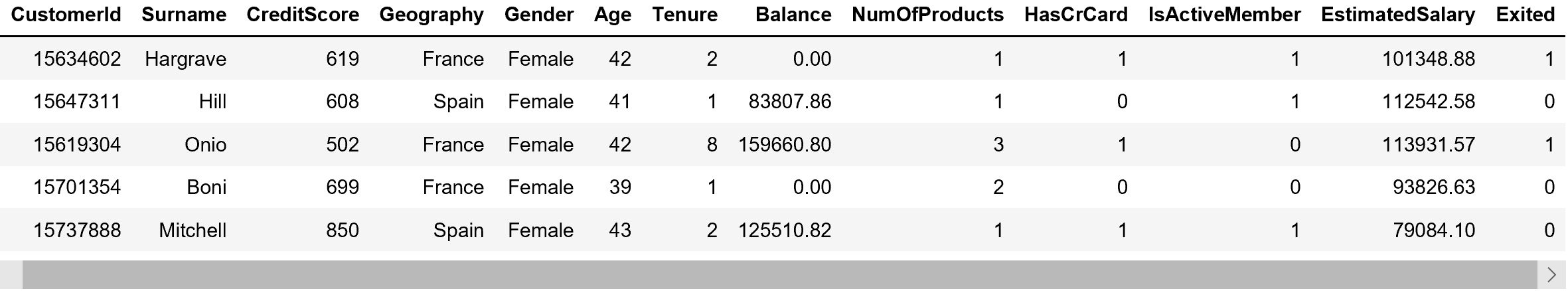
# Related Work

There are numerous predictive modeling techniques for predicting customer churn. These vary in terms of statistical technique (e.g., neural nets versus logistic regression versus survival analysis), and variable selection method (e.g., theory versus stepwise selection). Depending on the domain, many domain specific features have been used for churn prediction in the past. For example, for telecom sector, the following features have been used: 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 evening calls, total evening charge, total night minutes, total night calls, total night charge, total international minutes used, total international calls made, total international charge, number customer service calls made.

Tsai and Lu used two different hybrid models to develop customer churn prediction model. The developed hybrid model is a combination of two artiﬁcial neural networks and the second hybrid model is a combination of self organizing maps and artiﬁcial neural networks. First models are used for data reduction and second models are used for actual classiﬁer [1]. Kechadi and Buckley used attribute derivation process to increase the correct prediction rate [2]. Bayesian Belief Network method is tried in a study which is conducted by Kisioglu and Topcu [3]. Verbeke et al. increased the accuracy by using two different rules extraction method. These methods were AntMiner+ and ALBA [4]. Bock and Poel used two different rotation based ensemble classiﬁers. These are Rotation Forest and Adaboost [5]. Yeshwanth et al. suggested a new hybrid model that combines C4.5 decision tree programming [6]. Zhao et al. used one class support vector machine to increase the performance [7]. Ghorbani etal. created a new hybrid model by combining neural network, tree models and fuzzy modeling [8]. Other recent popular work on churn prediction includes the following: [9], [10], [11], [12].

# Description of dataset.

The dataset [Churn\_Modelling.csv] belongs to banking domain. Which contains the customers with the attributes as listed below.



This dataset is a 10,000 records sample.

Link to Dataset:

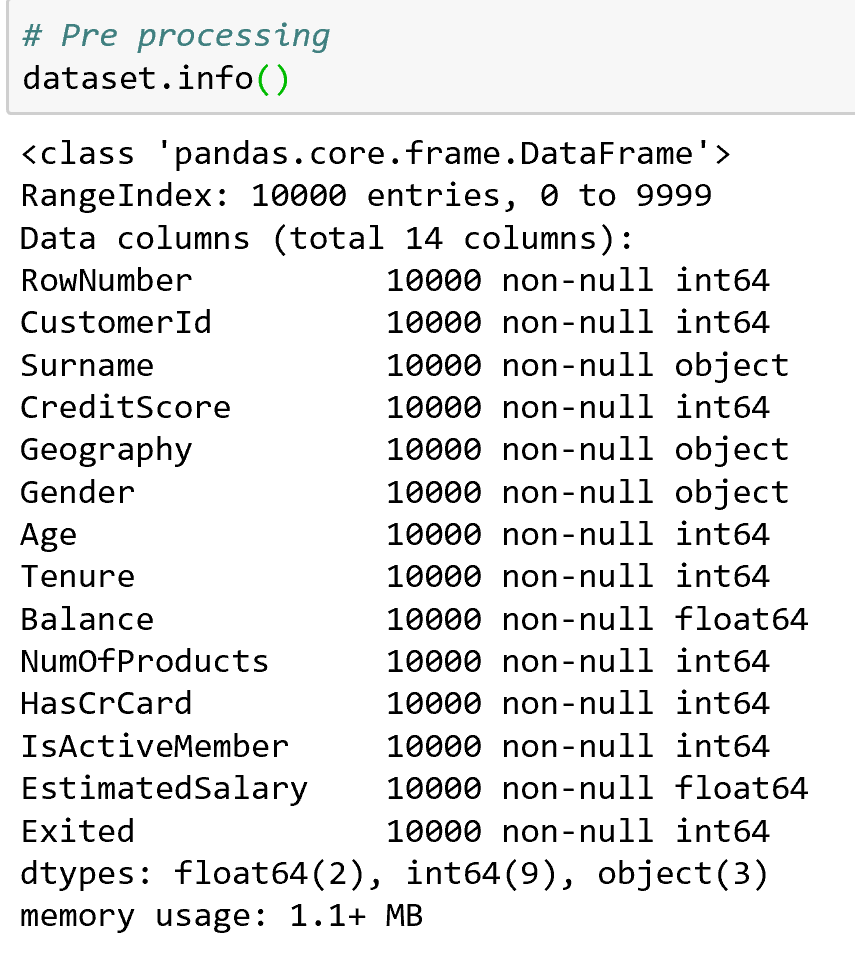


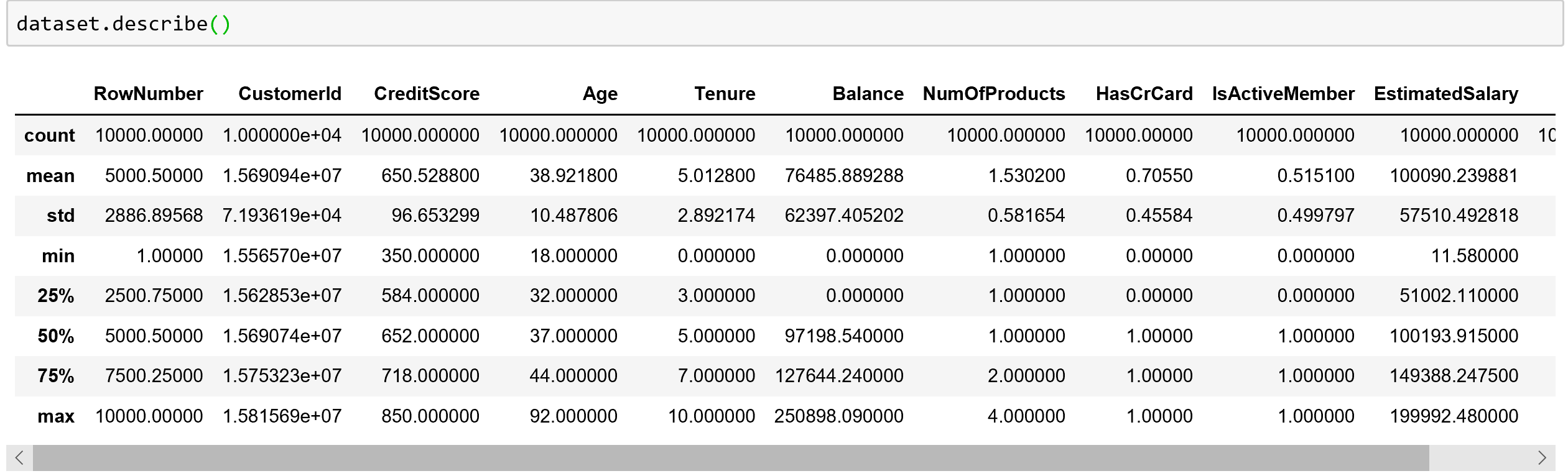
* The dataset contain the columns as listed below
  + customerId – represents the customer unique id provided by bank
  + Surname – sur name of the customer
  + Credit Score – The credit score for the customer.
  + Geography – The region where the customer located.
  + Gender – Male / Female
  + Age – The customer’s age
  + Tenure – The duration in years
  + Balance – The account balance of customer
  + NumOfProducts – The number of products subscribed / used by the customer.
  + HasCrCard – The flag to indicate the credit card
  + IsActiveMember – The activeness of the customer.
  + EstimatedSalary – The estimated salary for the customer
  + Exited – Indicates the customer who churned if the value is 1.

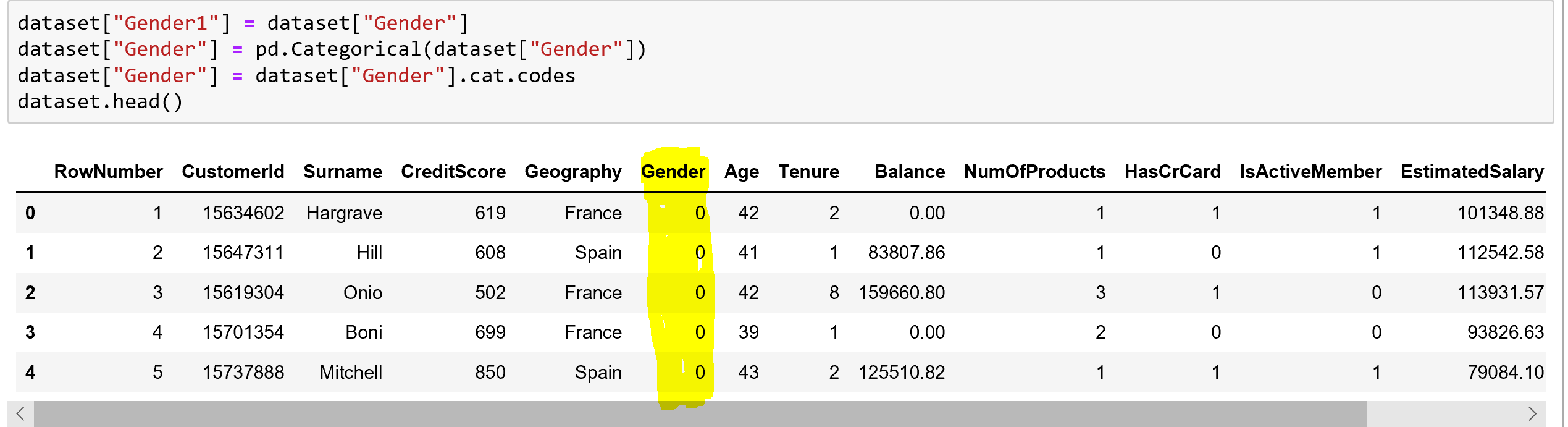
# Pre-processing steps:

The pre-processing techniques used to transform the datasets and to impute the missing data.

* Identify the missing values: From the below command, we can identify the columns with the missing values. From the non-null records count, we can identify the number of missing values present in the respective columns.

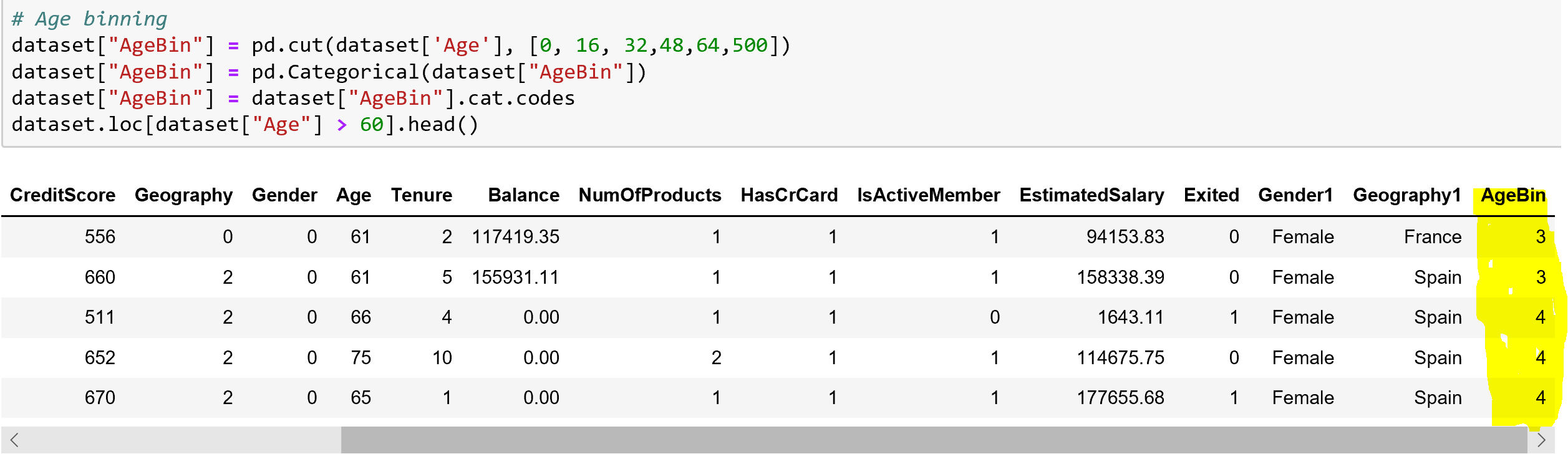


* Get the summary of statistics: From this, we can able to understand how all the datapoints were distributed in comparison with mean. 
* Generally, if we identify the missing values in a dataset, then we will use the imputation techniques to impute the null or blank values with the above calculated metrics (any of mean, median, mode and quartiles) based on the business impact.
* If the column is sensitive enough to impute the data points then better to ignore those records.
* Transform the categorical variables into factors. For the attributes, Gender, Geography the categorical variables need to be converted to factor values for other statistical models other than Decision tree model.



From the above, the categorical values of Gender were converted into factors of levels 0, 1. 1 represents male and 0 represents female.

* Transform the Age column into categorical bins.



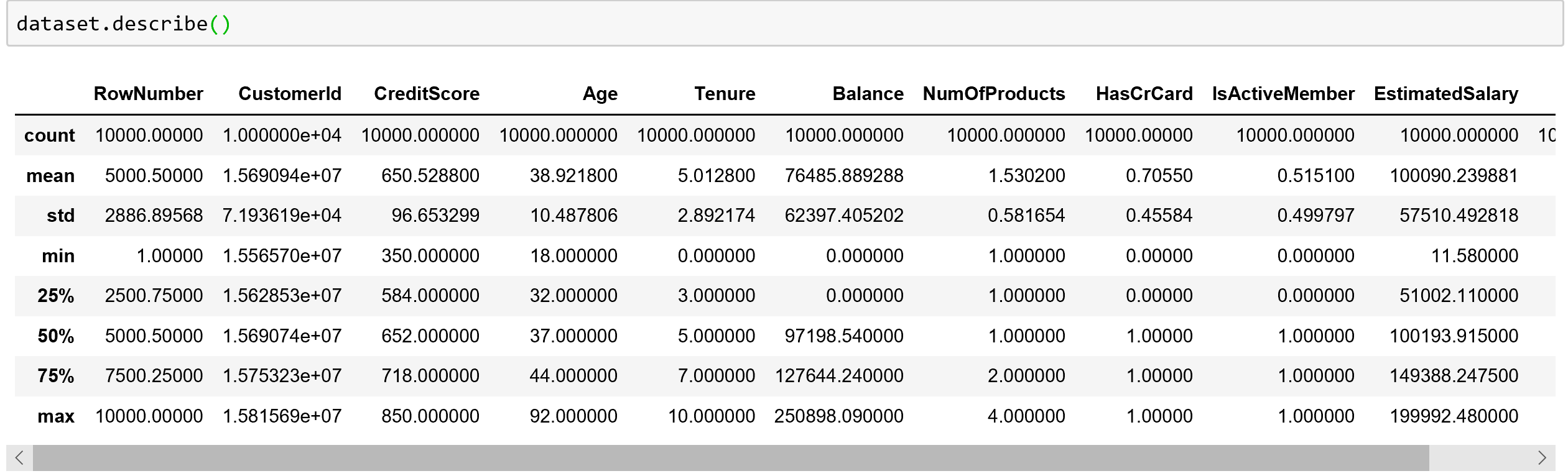
From the above snapshot, the continuous variable Age is binned into multiple factors based on the ranges as mentioned in the cut function and based on the ranges the age can be binned into 5 categories from 0 to 4. Such as 0-16,16-32,32-48,48-64, >64

* Similarly, the binning can be applied to the variables CreditScore, EstimatedSalary, Balance columns.

# Exploratory Data Analysis steps:

Exploratory Data analysis is an approach to analyze the datasets with the visualizations to identify the trends in the datasets.

First, we simply used describe to understand the type and range of values per column.



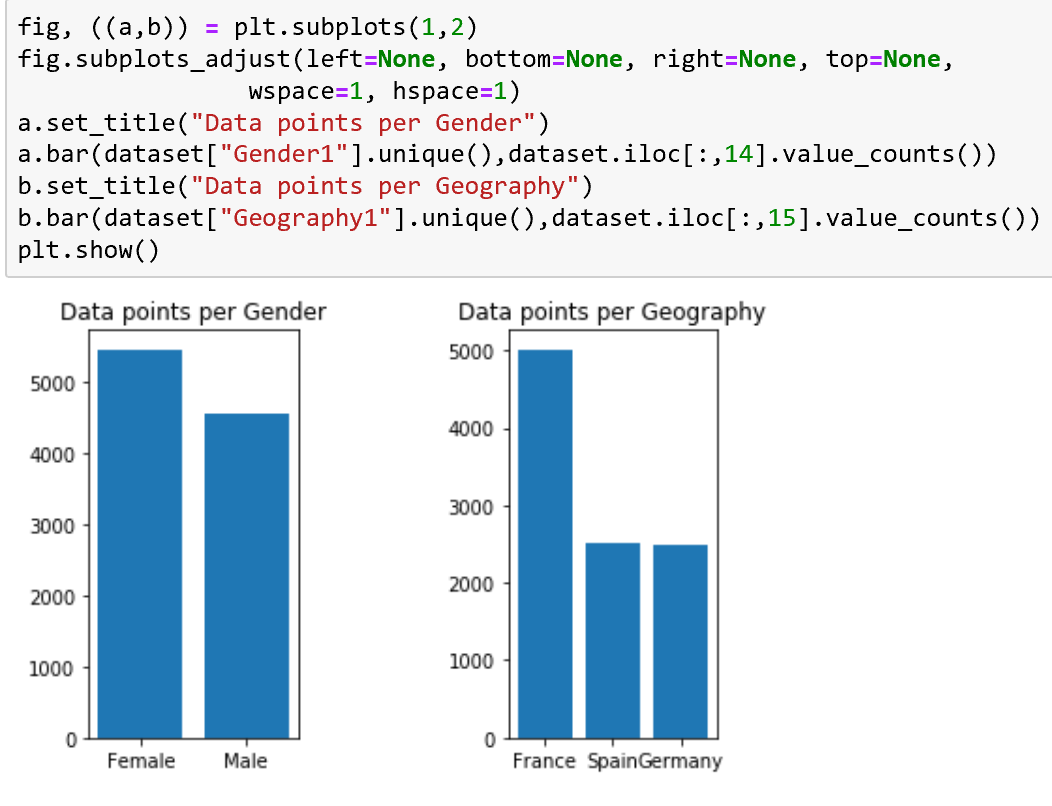
## Box Plot:

Using box plots, we can identify the distribution of all the data points over mean. These charts are used to identify the median (in red lined bar) in the charts and the respective quartiles in the visualized way.



## Bar plots:

These plots can be used to compare the attributes among one another.



# Split Datasets:

In any machine learning model, we need to split the dataset as listed below:

## Training Dataset:

This dataset can be used to train the statistical model.

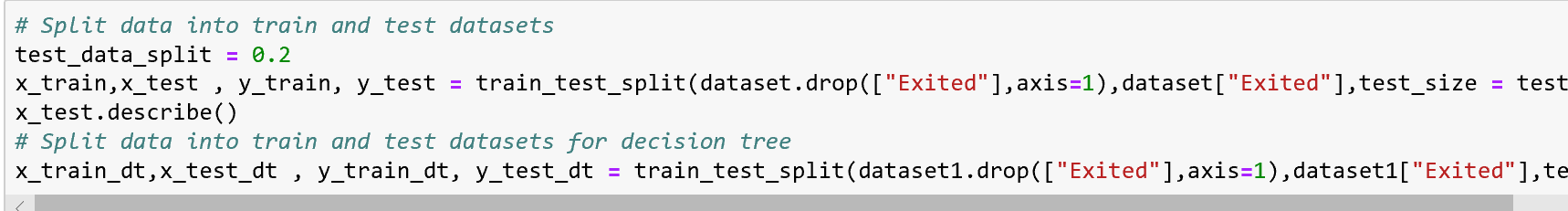
## Validation Dataset:

This dataset can be used to predict the test results and for calculation of accuracy metrics and ROC curves to determine the model performance. The validation dataset contains the known classes, and these original classes can be compared with the predicted classes to calculate the accuracy.

## Test Dataset:

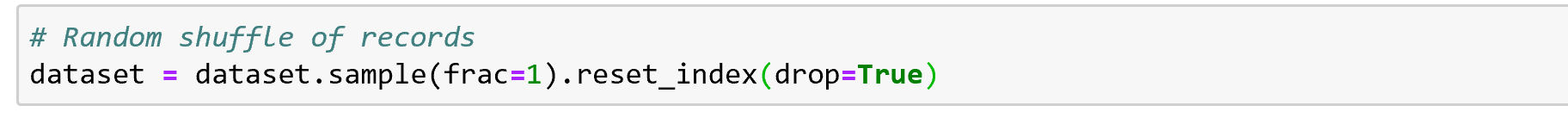
This is the resultant dataset, This dataset can be used to predict the classes.

In our case, for model evaluation, we have split the dataset into 2 parts as training dataset and test dataset.



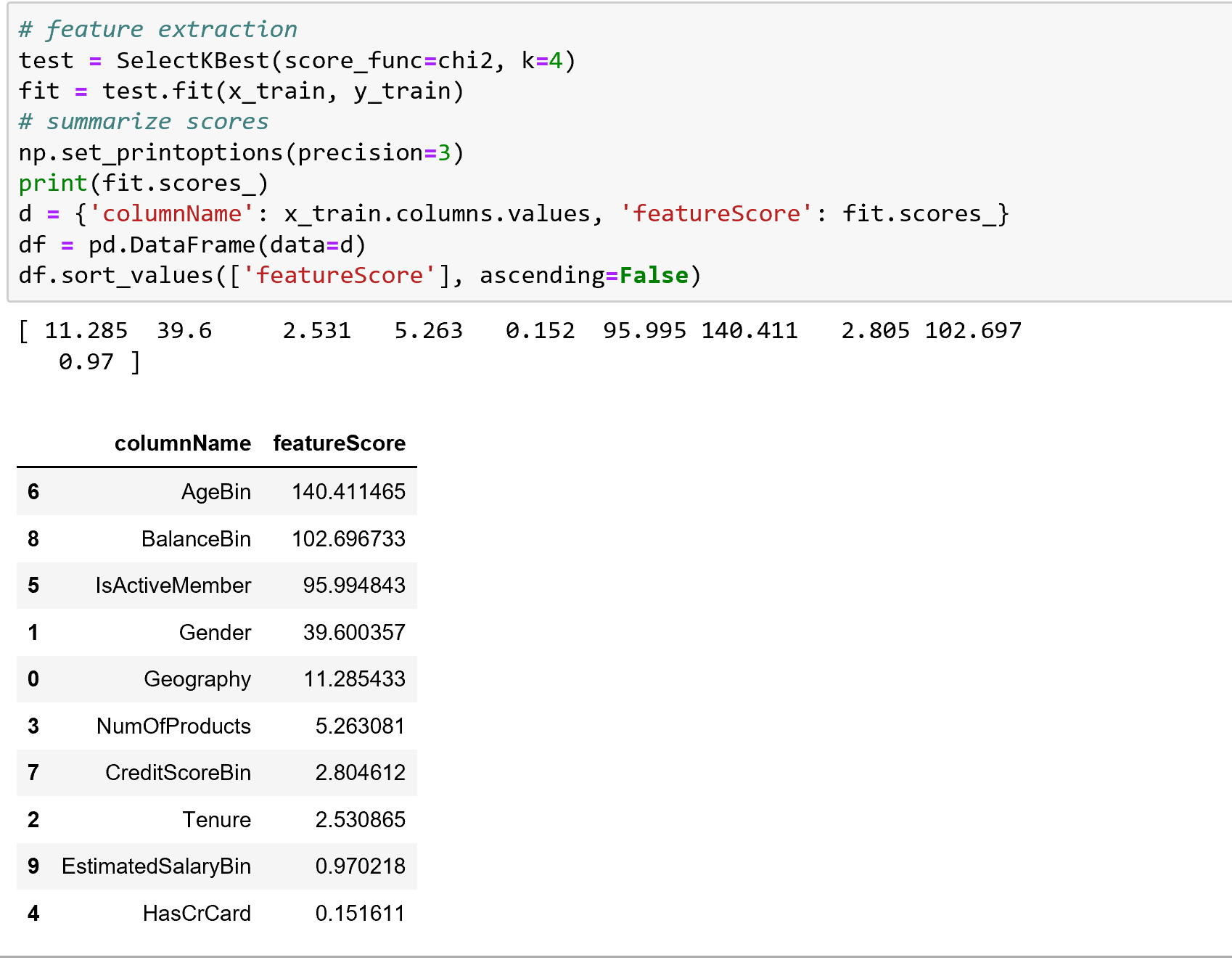
Here, we split the dataset into two samples with 80% of records as training dataset and 20% of records as test dataset.

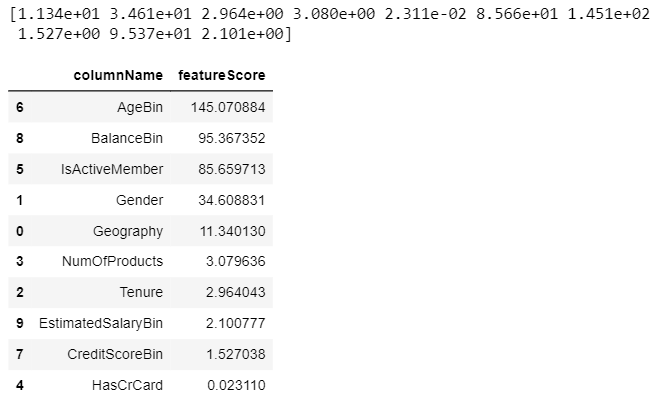
In order to get the best combination of records split over training and test datasets, we need to random mix the samples from the below command.



# Feature selection:

To determine the importance of variables over the predicted outcome. Here, in our case we need to identify the impact of a variable over Exited column. Ideally, the variables with less importance can be ignored in model building.





The higher the feature score, the higher the importance.

# Understanding Outcome [Exited]:

Number of customers with their Exited status spread in training, test and the complete dataset.



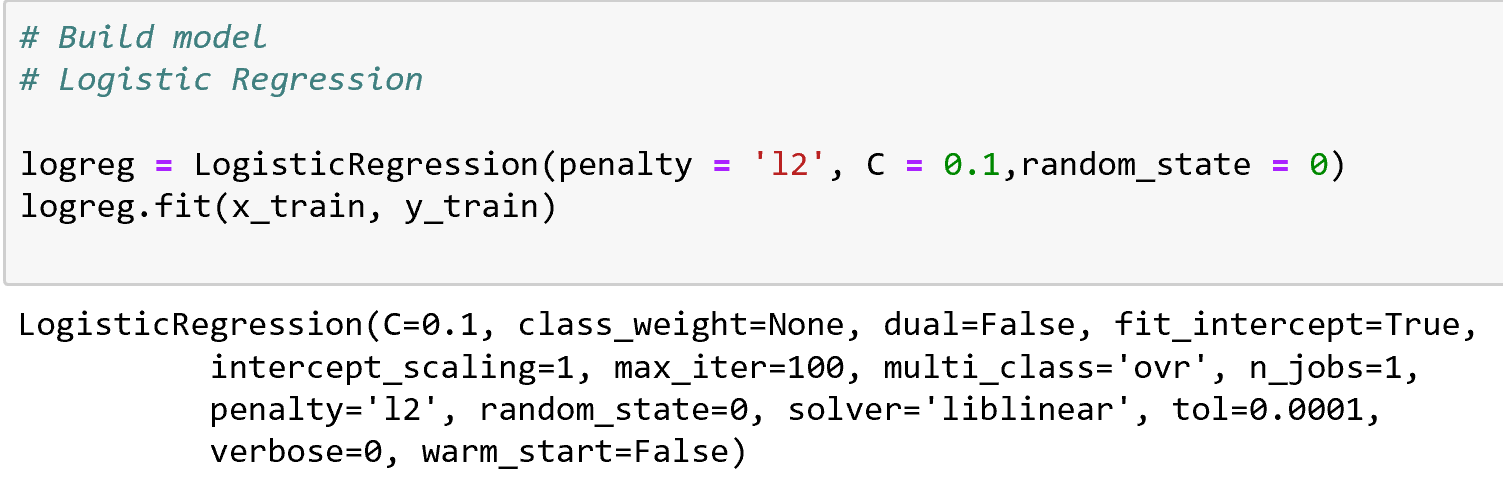
# Metrics

* **Accuracy:**  This metric represents the accuracy of the model by calculating the successful calculation of positives as true positives and negatives as true negatives.
* **ROC curve:** ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate against the false positive rate at various threshold settings

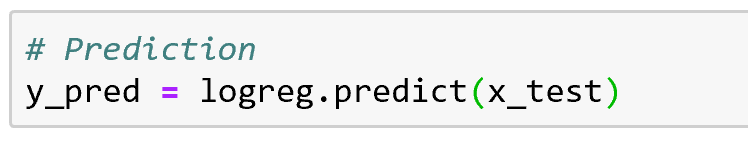
# Building (Training) Model:

## Logistic Regression:

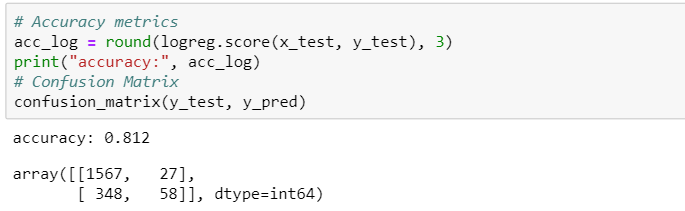
* + This model can be used to model a binary dependent variable where the two outcomes are 0 and 1.



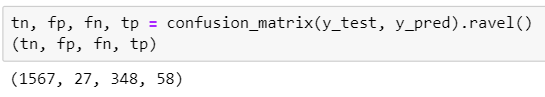
### **Prediction on test data**



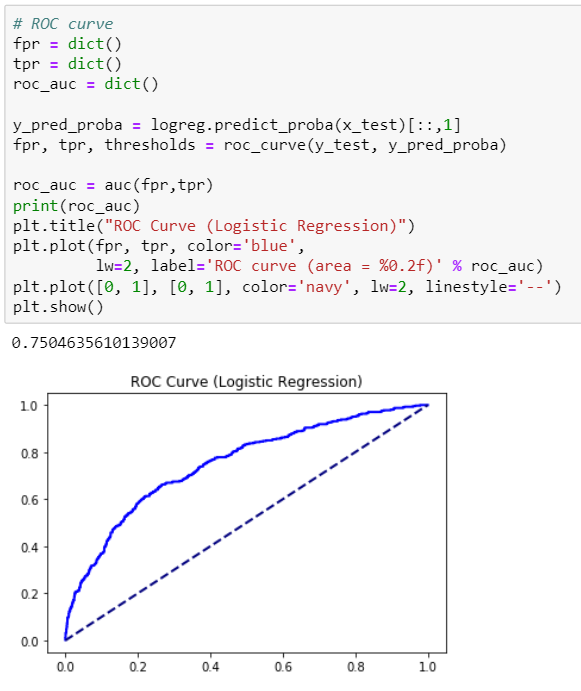
### **Accuracy & Confusion matrix**



### **True negative, false negative, true positive & false positive - Confusion matrix**

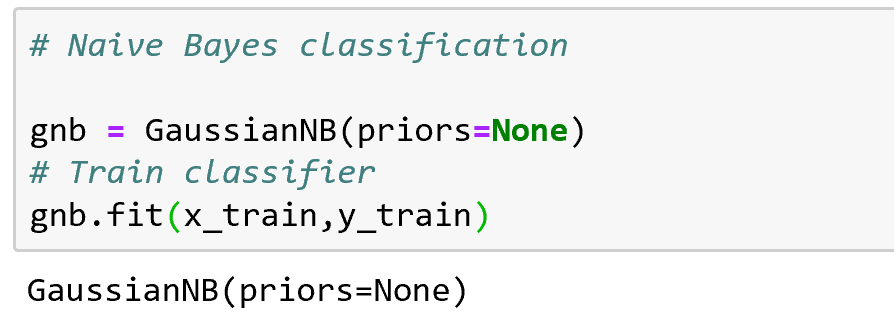


### **ROC curve**



## Naïve Bayes Classifier:

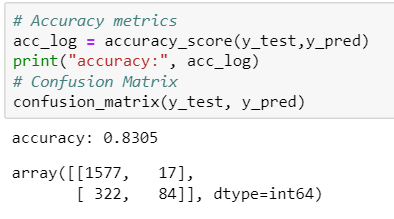
This model is basically a simple probabilistic classifier considering all the features are independent with each other.



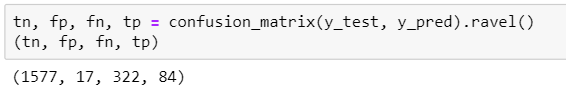
### **Prediction on test data**



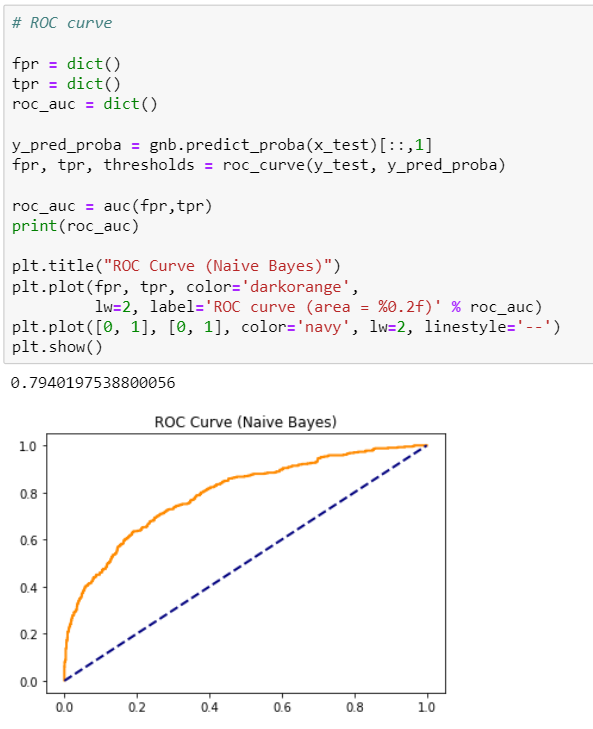
### **Accuracy & Confusion matrix**



### **True negative, false negative, true positive & false positive - Confusion matrix**

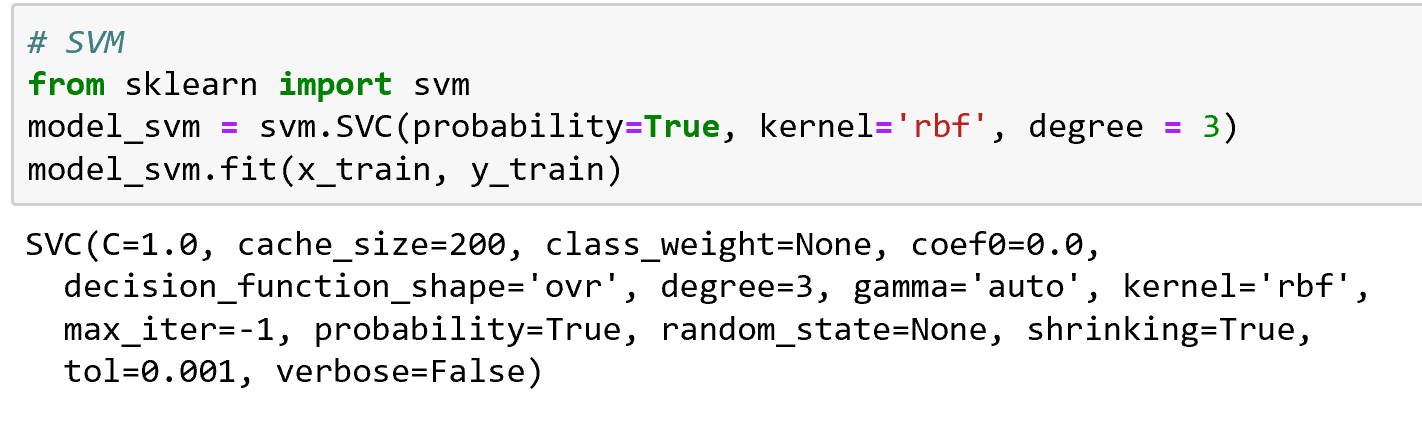


### **ROC curve**

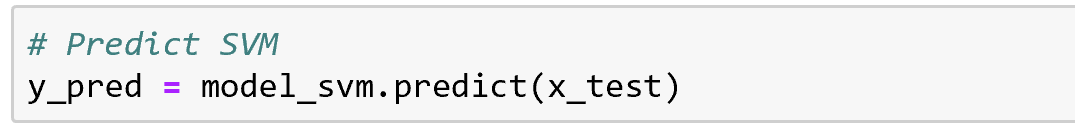


## Support Vector Machines (SVM):

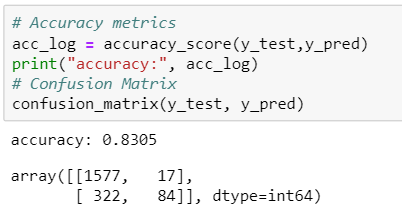
* + This model can be used for supervised learning for a classification and regression analysis. SVM model is a representation of points in space and considers as a hyperplane equation to solve the problem.



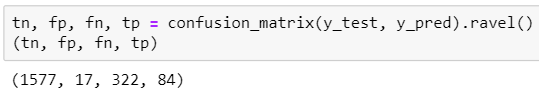
### **Prediction on test data**



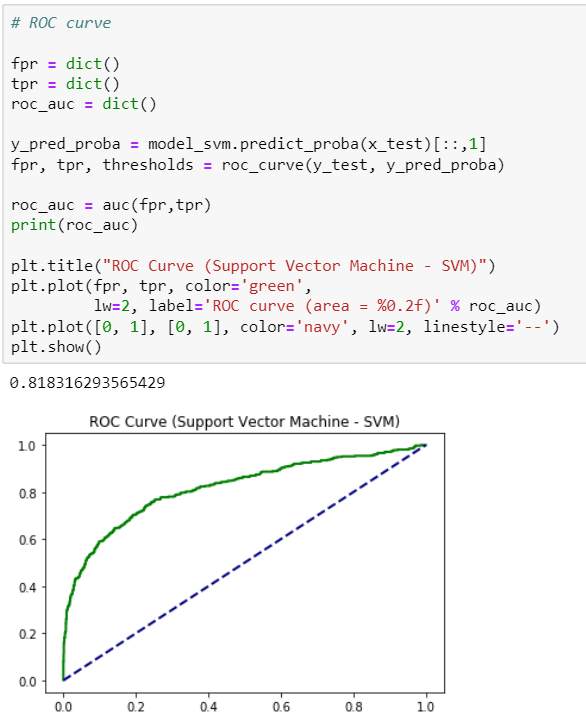
### **Accuracy & Confusion matrix**



### **True negative, false negative, true positive & false positive - Confusion matrix**

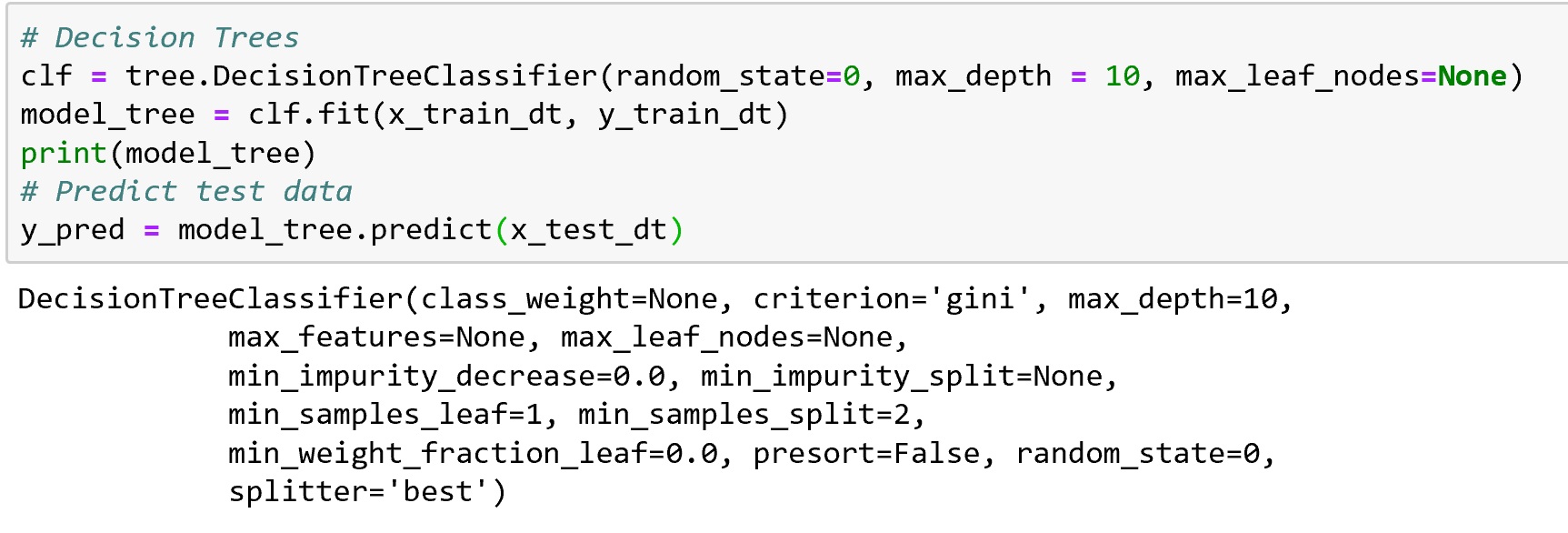


**ROC curve**

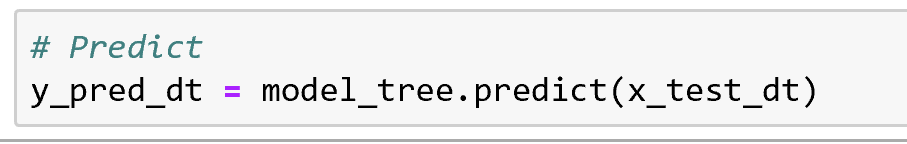


## Decision Trees:

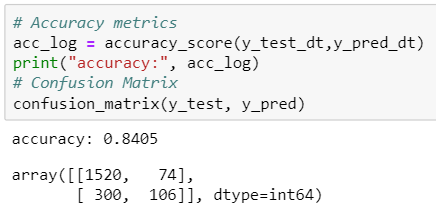
* + This model is used to build a decision making using the tree like structure. The corresponding outcome can be derived by traversing over the tree structure.



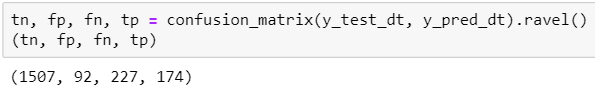
### **Prediction on test data**



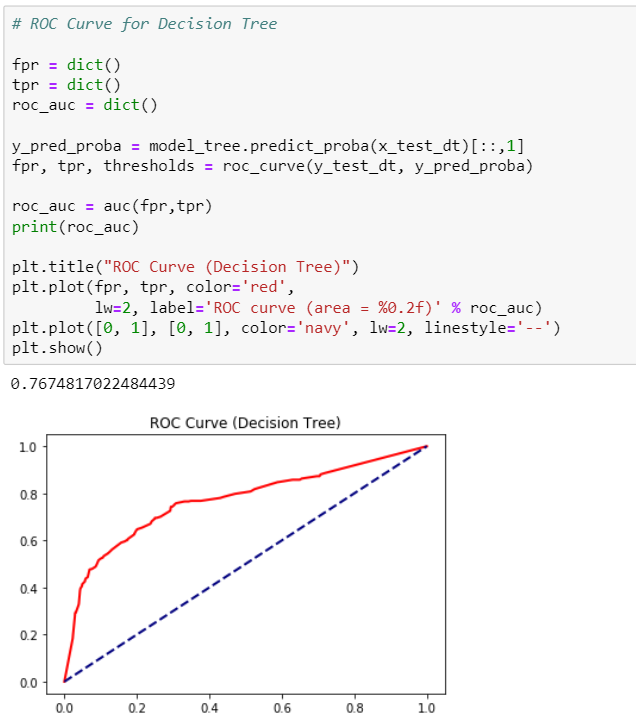
### **Accuracy & Confusion matrix**



### **True negative, false negative, true positive & false positive - Confusion matrix**

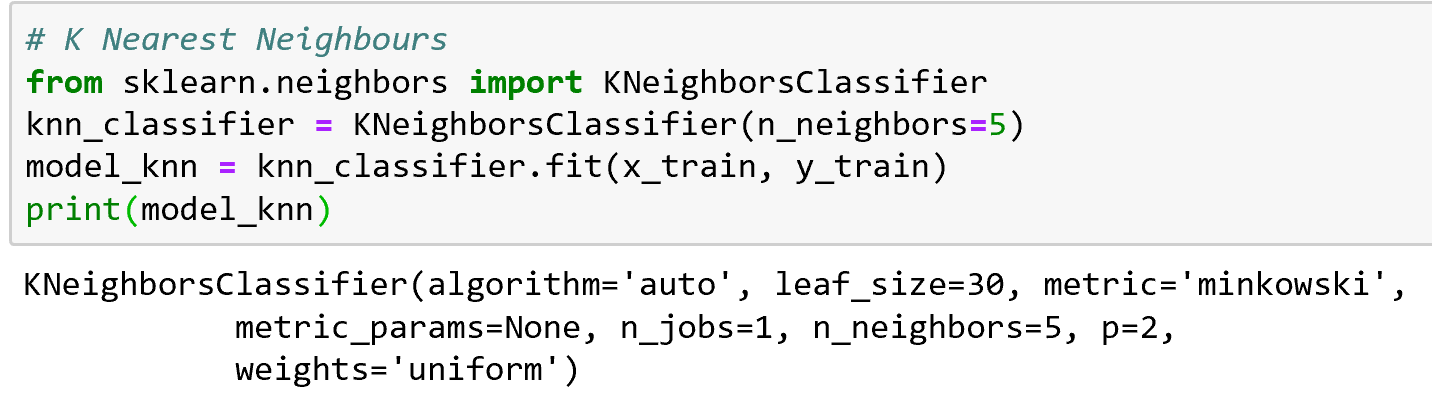


### **ROC curve**

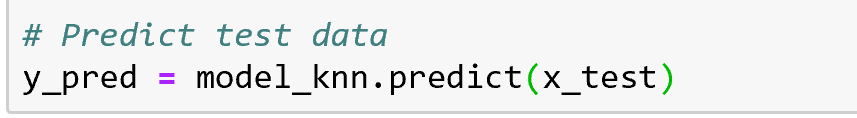


## K-nearest Neighbours(KNN):

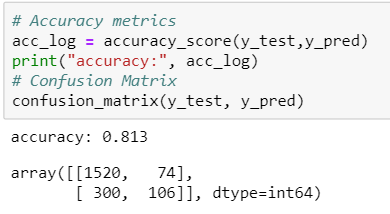
* + K-nearest neighbors algorithm is a non-parametric method used for classification and regression problems. The outcome result can be classified by majority of nearest neighbor points.



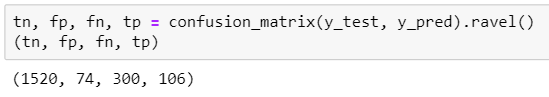
### **Prediction on test data**



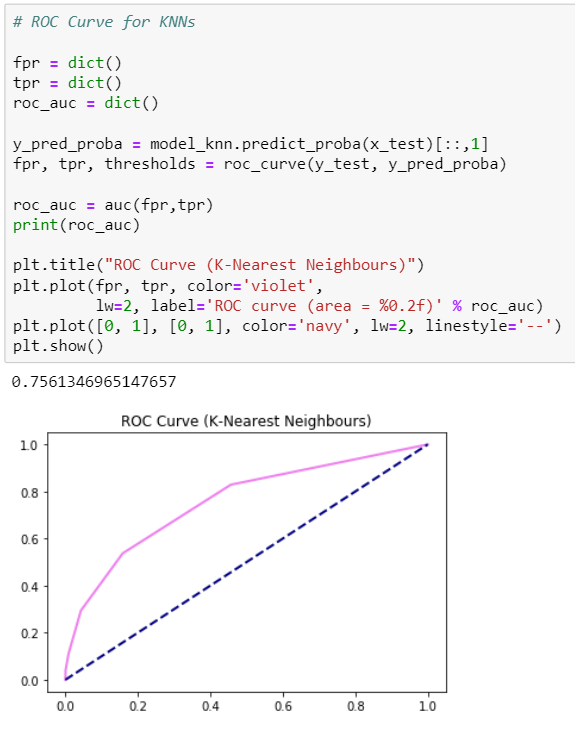
### **Accuracy & Confusion matrix**



### **True negative, false negative, true positive & false positive - Confusion matrix**

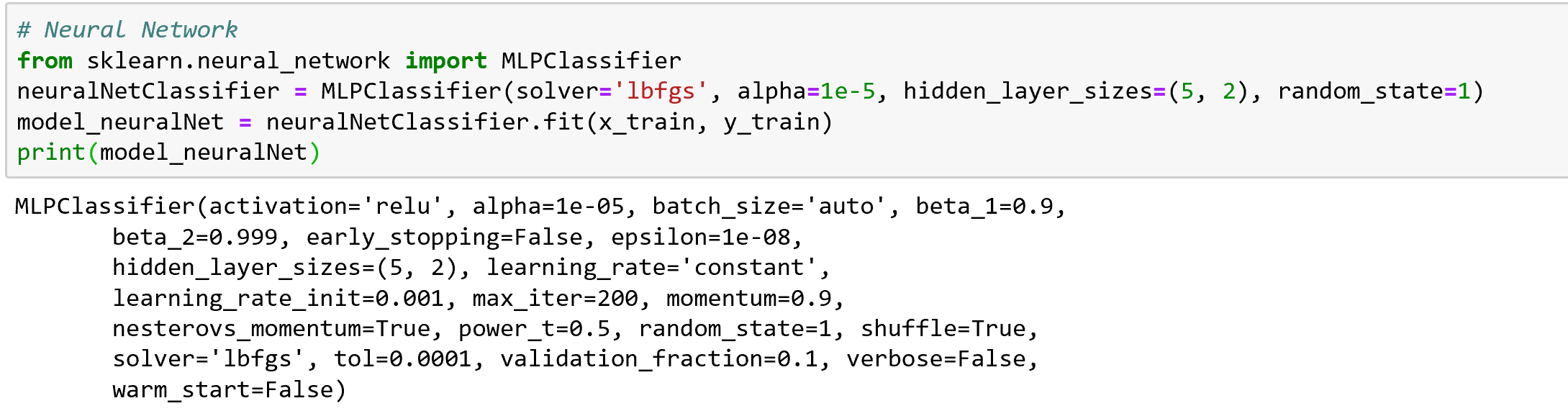


### **ROC curve**

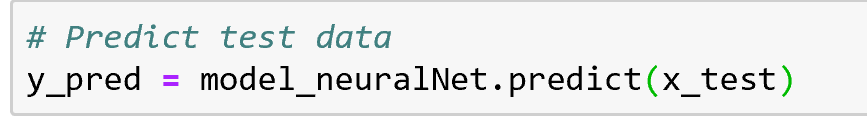


## Neural Networks:

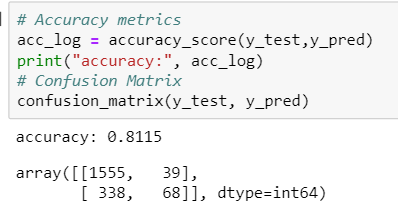
* + This algorithm used to identify the relationships in a set of data through a process as similar to human brain operates. Neural network generates the best possible result without redesign the output criteria.



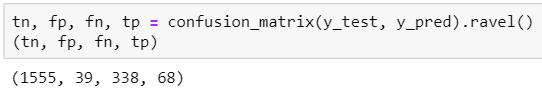
### **Prediction on test data**



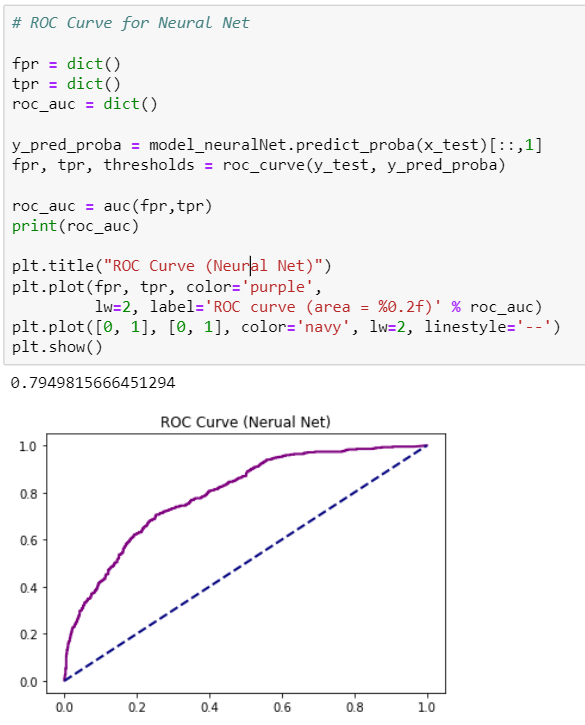
### **Accuracy & Confusion matrix**



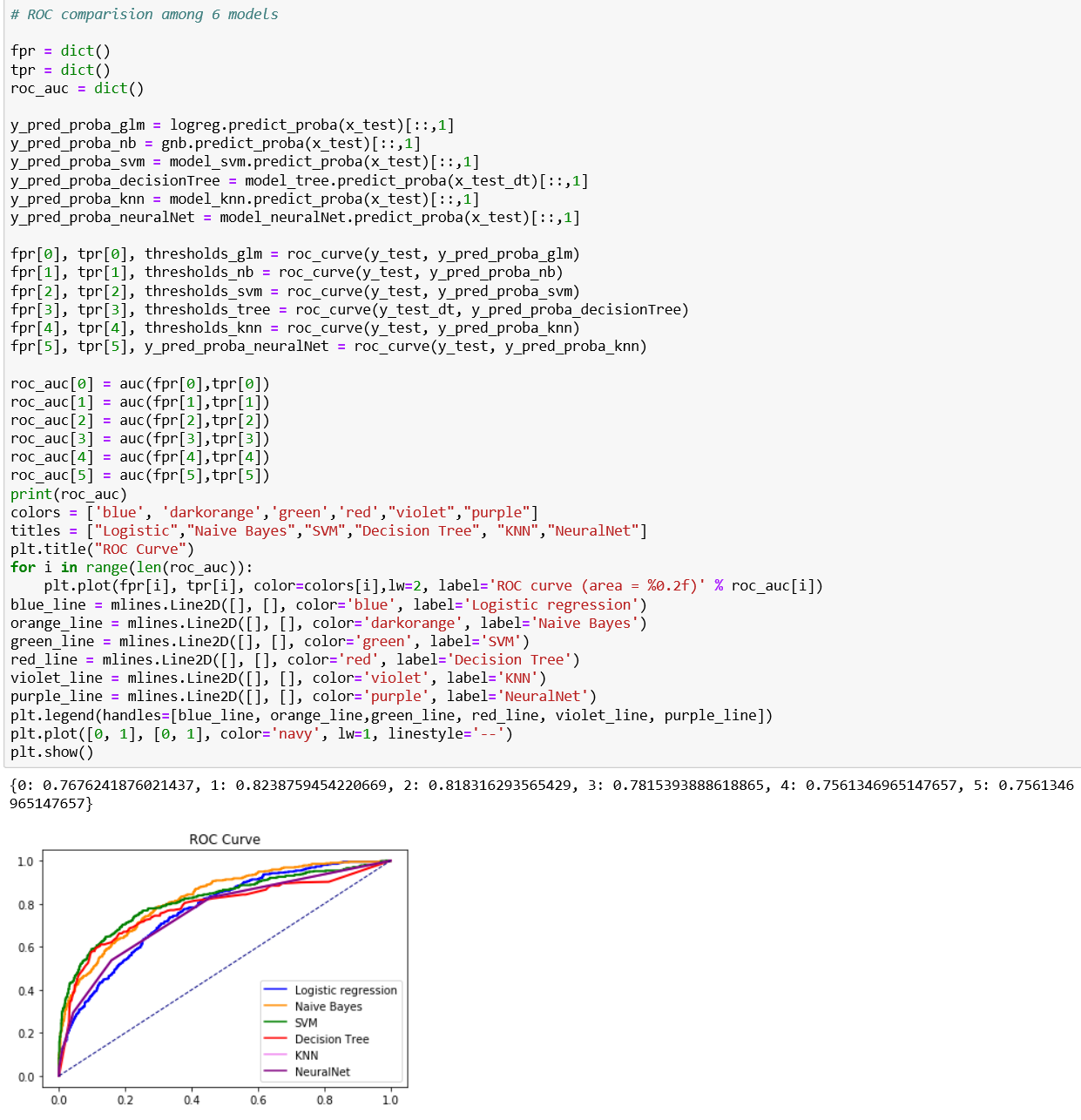
### **True negative, false negative, true positive & false positive - Confusion matrix**



### **ROC curve**



## **ROC Comparison of 6 models:**



* As this dataset columns are independent with each other, the Naïve Bayes classification gave good results.

# Implementation Details:

* Python Libraries used: Pandas, matplotlib, numpy, sklearn
* Tools/software used: Python Jupyter notebook.
* Code files:
  + ChurnPrediction.ipynb – This code was built using Jupyter notebooks, and by uploading this script to jupyter notebooks to run this module.
  + ChurnPrediction.py – A python file to run it as a python application.

# Deployment process:

## Hardware requirements:

* Windows (or) Linux
* RAM (min 4GB)

## Software requierements:

* Python
* Browser

## Detailed steps:

* Using the pickle library to serialize a model object.
* Build a web service for prediction using the flask library.
* Invocation of web service by passing the test dataset to the web service using rest api calls.

## Hardware/software requirements for client installation:

* Depends on the amount of datasets and the performance time. In general the Python applications will run with minimum configuration of 2GB RAM. It can be configured to even more to improvise the processing time.

# Conclusion:

* Analyzed the problem statement for building the churn analytics model for the leading bank to identify the customer who are willing to churn.
* Detailed understanding of the dataset and loading the dataset into python environment for analysis.
* Pre processing of the dataset is done.
* Data exploratory analysis is done and their visualization of data points was done.
* Built the statistical models and calculation of their accuracy metrics.
* Perform the prediction on the test dataset.
* Confusion metrics and their associated true positives, true negatives, false positives and false negatives was identified.
* Built the ROC curve and compare all the models in terms of accuracy metrics.

# Future Work

1. The models can be improvised by continuous feedback implementation to identify the future trends and with the additional attributes.
2. We can request the additional data from the customer to use the extra features to improvise the solution.
3. The customer datasets might have missing values, those can be handled with imputation techniques as a pre-processing steps and the proper imputation methods and other pre-processing steps can be considered to best improvise the model.
4. Hyper-parameter tuning can be done for various classifiers.

# References

|  |  |
| --- | --- |
| [1] | C. F. T. a. Y. H. Lu, "Customer churn prediction by hybrid neural networks," *Expert Systems Application,* vol. 36, no. 10, p. 12547—12553, 2009. |
| [2] | M. T. K. a. B. B. B. Huang, "Customer churn predictionin telecommunications," *Expert Systems with Applications,* vol. 39, no. 1, pp. 1414-1425, 2012. |
| [3] | P. K. a. I. Y. Topcu, "Applying Bayesian belief network approach to customer churn analysis: a case study on the telecom industry of Turkey.," *Expert Systems with Applications,* vol. 38, no. 1, pp. 7151-7157, 2010. |
| [4] | D. M. C. M. a. B. B. W. Verbeke, "Building comprehensible customer churn prediction models with advanced rule inductiontechniques," *Expert Systems with Applications,* vol. 38, no. 1, p. 2354—2364., 2011. |
| [5] | K. W. B. a. D. V. Poel, "An empirical evaluation of rotation-based ensemble classiﬁers for customer churn prediction.," *Expert Systems with Applications,* vol. 38, no. 10, pp. 12293-–12301, 2011 . |
| [6] | V. V. R. a. M. S. V. Yeshwanth, "Evolutionary churn prediction in mobile networks using hybrid learning," in *Proc. of XXIVFlorida Artiﬁcial Intelligence Research Society Conference*, Florida, 2011. |
| [7] | B. L. a. X. L. Y. Zhao, "Customer churn prediction using improvedone-class support vector machine.," *Lecture Notes in Artiﬁcial Intelligence,* vol. 3584, no. 1, pp. 300-306, 2005. |
| [8] | F. T. a. C. L. A. Ghorbani, "The application of the locally linear model tree on customer churn prediction," in *Proceedings of theInternational Conference of Soft Computing and Pattern Recognition(SOCPAR’09)*, Malaysia, 2009. |
| [9] | C.-P. a. C. I.-T. Wei, "Turning telecommunications call details to churn prediction: a data mining approach," *Expert systems with applications,* vol. 23, no. 2, pp. 103-112, 2002. |
| [10] | W.-H. a. C. K. C. a. Y. X. Au, "A novel evolutionary data mining algorithm with applications to churn prediction," *IEEE transactions on evolutionary computation,* vol. 7, no. 6, pp. 532-545, 2003. |
| [11] | J. a. V. d. P. D. Burez, "Handling class imbalance in customer churn prediction," *Expert Systems with Applications,* vol. 36, no. 3, pp. 4626--4636, 2009. |
| [12] | K. a. V. d. P. D. Coussement, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," *xpert systems with applications,* vol. 34, no. 1, pp. 313--327, 2008. |