**Assignment 2**

Course Information

Course Title: Applications of Artificial Intelligence

Course Number EAI6010

Term and Year: Winter Part B:

Start and End Dates: Start and End Dates: March 2 – April 10, 2021

STUDENT NAME: SHIVANI ADSAR

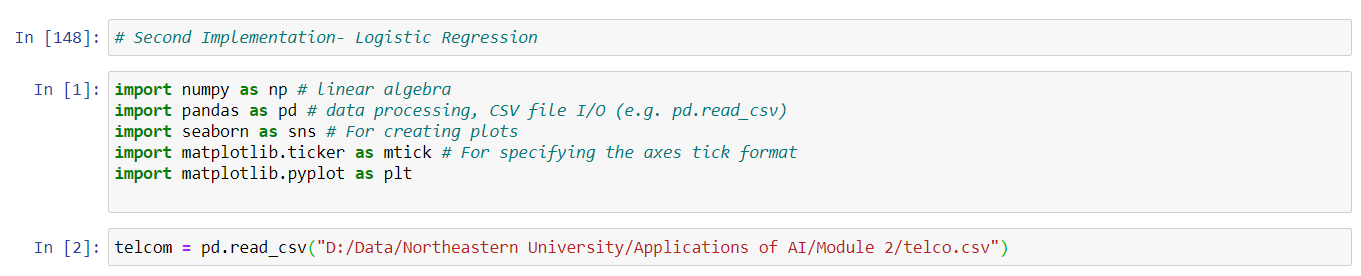
NUID: 001399374

ASSIGNMENT: 2

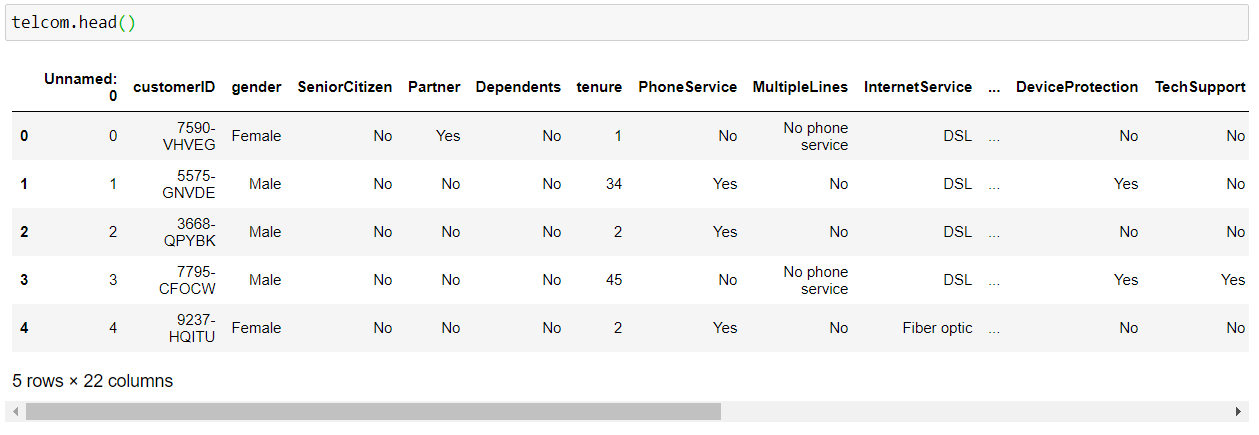
**Analysis**

Reading the CSV file

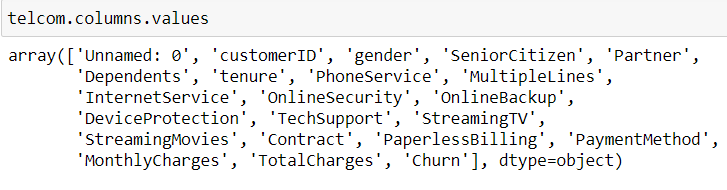
Some of the libraries like, numpy, pandas, seaborn, matplotlib were used for data processing, performing mathematical calculations and creating plots.



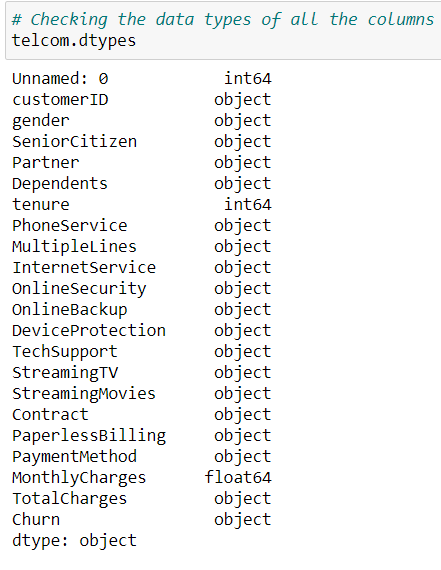
Head Function: This was used to check the data present in the first 5 rows of the dataset



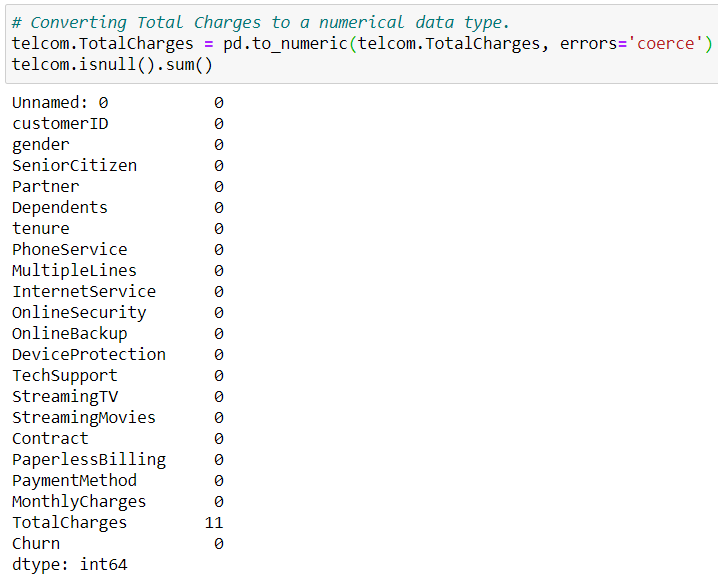
We analyzed the data present in the columns.



Further, we checked the data types of all the columns in the dataset



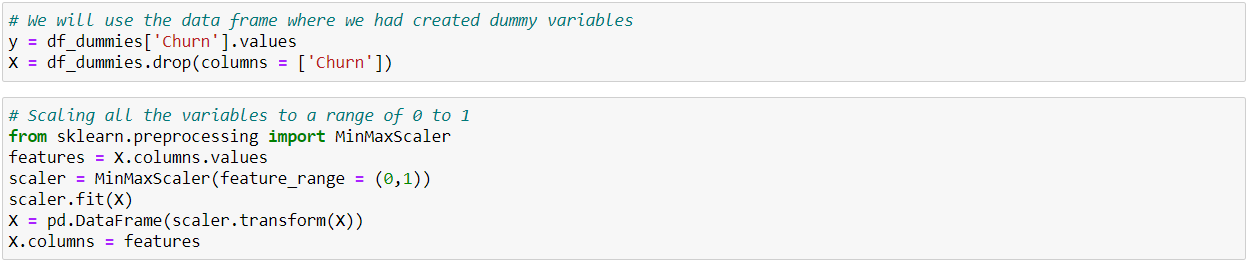
The variable ‘Total Charges’ was converted to the numerical data type



Moreover, we removed the missing values and then, converted the ‘Churn’ target variable into binary values.



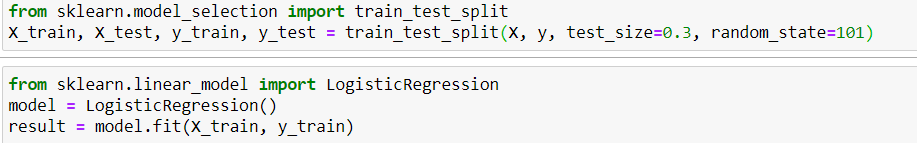
Further, we performed scaling to improve the accuracy, using normalization and standardization.



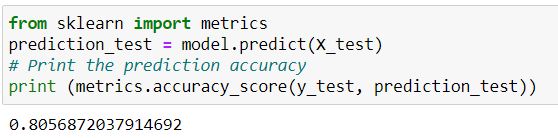
**Logistic Regression**

The data was split into ‘Test’ and ‘Train’ datasets for model building, with 30% in the Test data and 70 % in the Training data.

We used logistic regression function to perform logistic regression on our model.

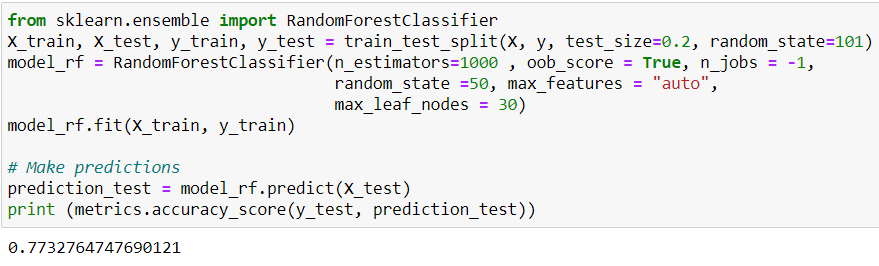


The accuracy of the logistic regression model was predicted to be 80.56 %.



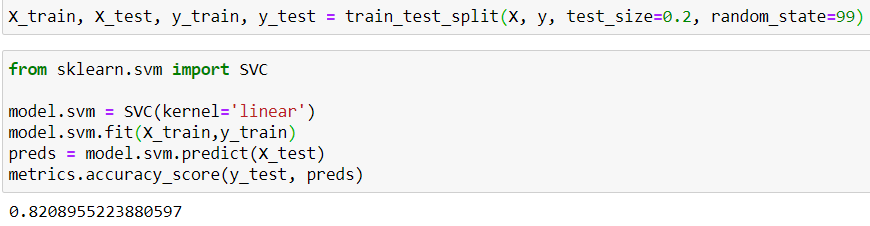
**Random Forest**

We used the RandomForestClassifier () function to implement and build the model on data.   
Further, we predicted the accuracy of the model on test data to be 77.32 %.



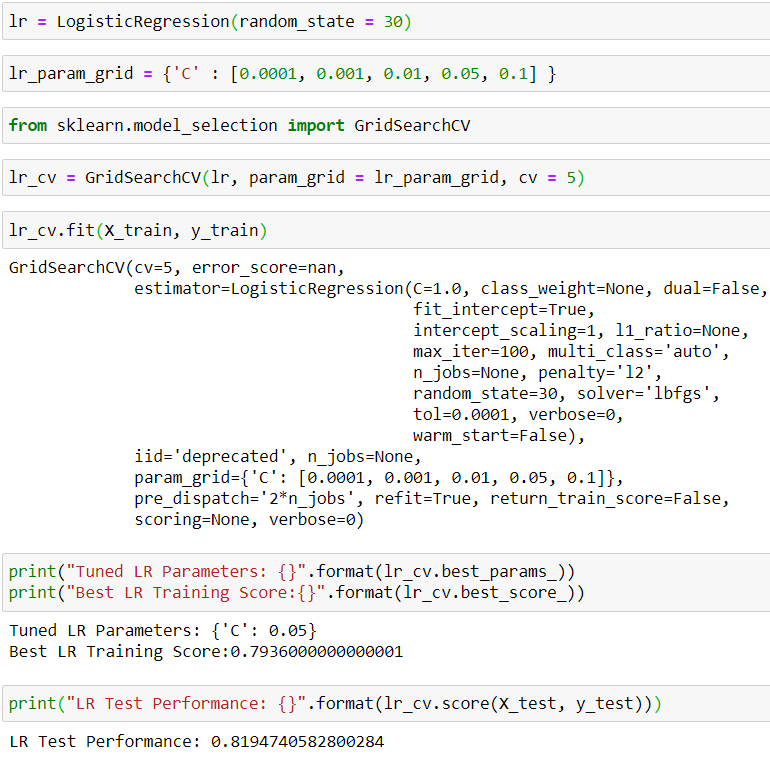
**Support Vector Machine**

We have used the SVC () function for building the Support Vector Machine classifier, and the accuracy was predicted to be 82.08 %.



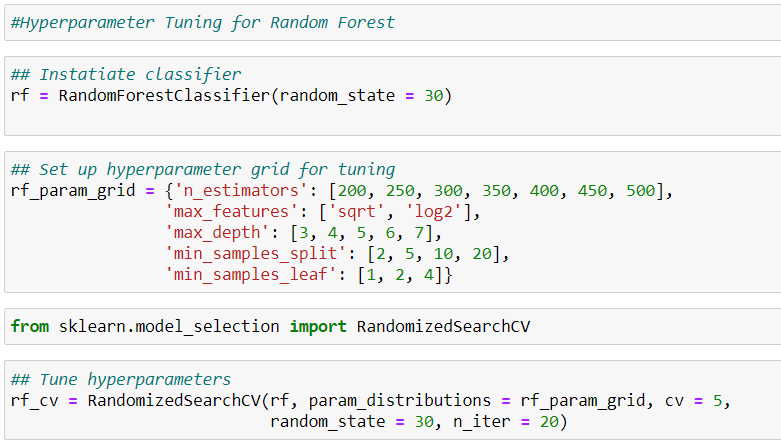
**Hyperparameter Tuning for Logistic Regression**

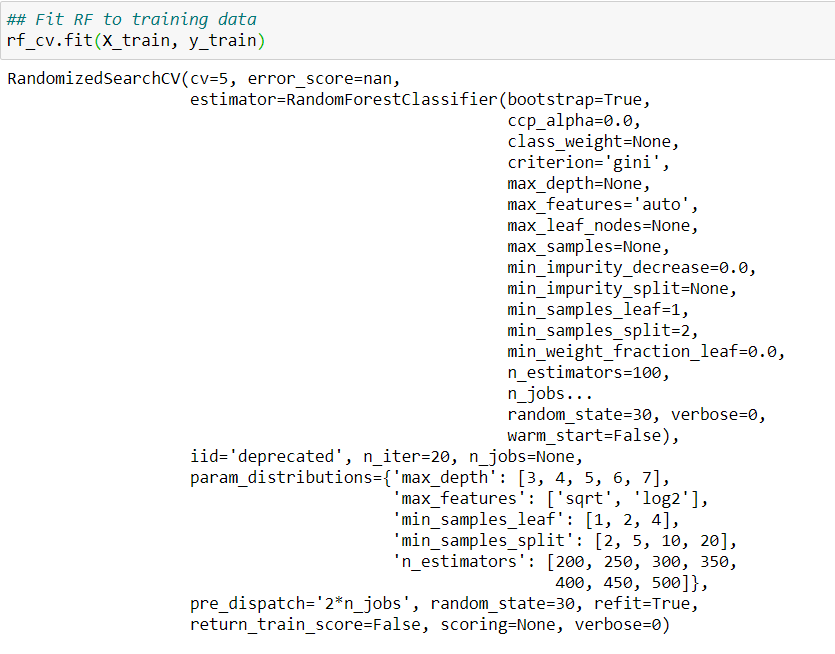
The Hyperparameter Tuning was performed on the logistic regression model and the accuracy was predicted to be 81.94 %.

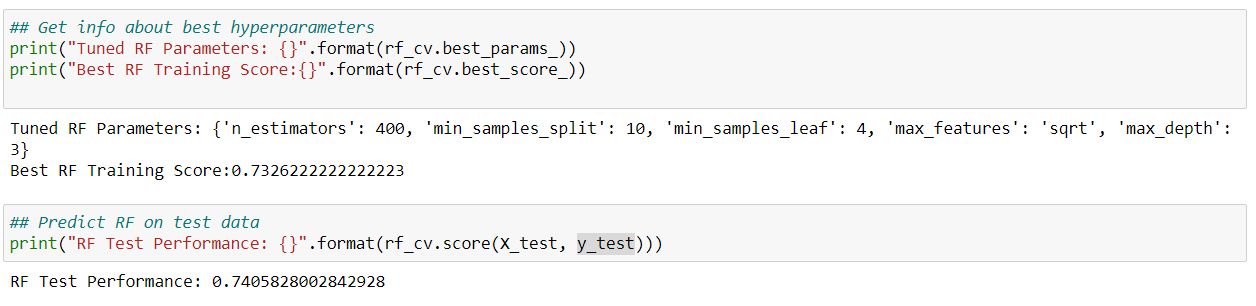


**Hyperparameter Tuning for Random Forest**

The Hyperparameter Tuning was performed on the Random Forest model and the accuracy was predicted to be 74.05 %.

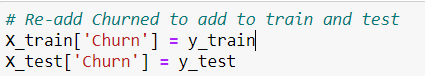




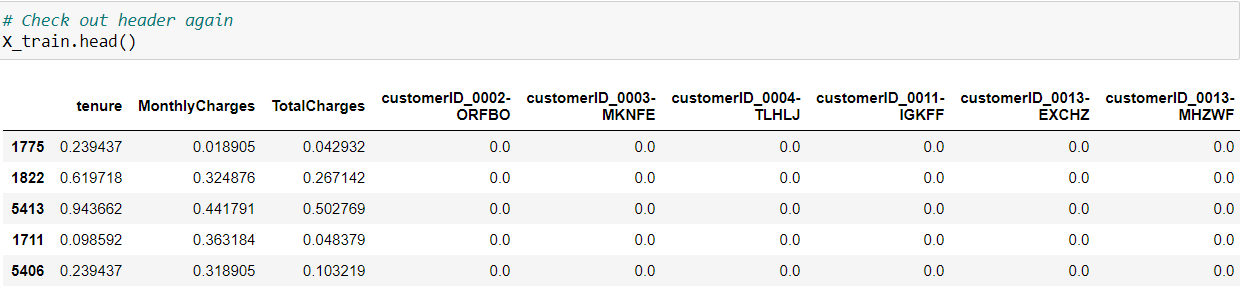


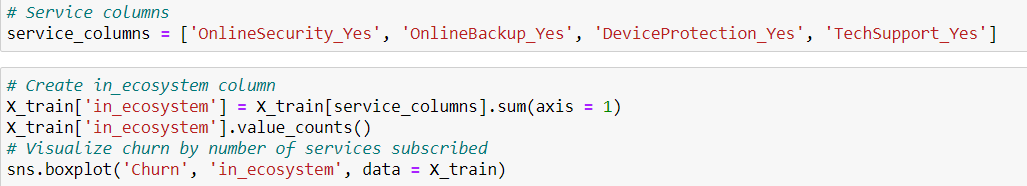
**Feature Engineering**

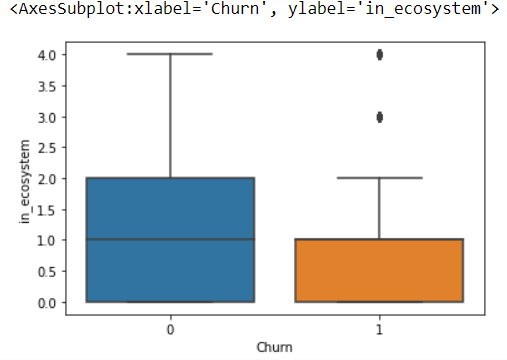
The train and test was separated on the Churn variable.

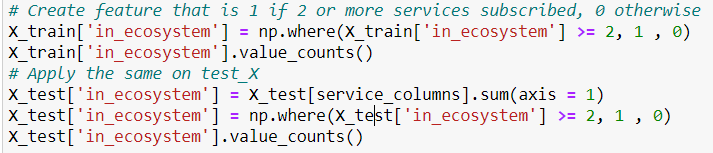


We checked the data present in train dataset by using the head() function.

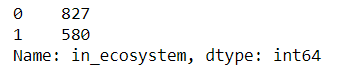








Output:



**Analytical Plan for increasing the threshold accuracy by 80%**

In order to increase the threshold accuracy, the following steps were performed:

* Reading the CSV file
* Checked for the data types of all the fields in the dataset
* Converted the non-numeric data to numeric data types
* Removed missing values
* Conversion of categorical variables to dummy variables
* Scaling of variables: We used Normalization which scales the input data into range between 0 – 1. Moreover, Standardization helps with scaling the variables, which shifts the standard deviation in a way to have zero mean and 1 as the standard deviation.
* Implementation of the Machine Learning model