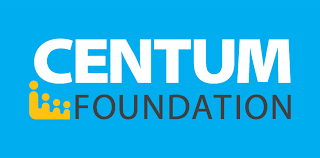
****

**Centum - Data analytics Training**

**Project completion report**

**Project Title: Customer churn using XGBoost**

**Project Lead: HARSHIT**

**Project team members: AAYUSHI AGARWAL**

**CHETNA KUMARI**

**HARSH SHARMA**

**HIMANSHI SINGH**

**HUNNY CHAUDHARY**

**Batch: MCA**

**Date: 31/01/2021**

**Section 1: Project description**

For many businesses, accurately predicting customer churn is critical to long-term success. Accurate prediction of churn probability drives many aspects of a business including proactive customer marketing, sales forecasting, and churn-sensitive pricing models. Therefore even slight improvements in accuracy can lead to dramatic improvements in profit.

Although customer churn models have existed in the business domain for decades, they have recently grown in complexity and accuracy as modern machine learning methods have advanced in recent years. Consequently, modern machine learning libraries such as extreme gradient boosting (XGBoost) can be applied to create very accurate models even on high dimensional data.

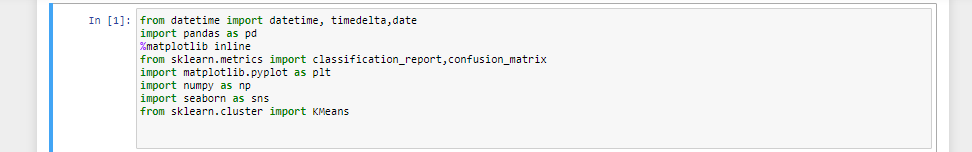
In addition, one of the most challenging aspects of creating churn models is handling issues associated with temporal data and ensuring all features are correctly accounting for time-shifts across the various time-windows that are used when training, cross-validating, and testing a machine learning model. In this model, we implemented a method for handling temporal feature engineering which was validated to be successful.

The datasets consist of several independent variables include:

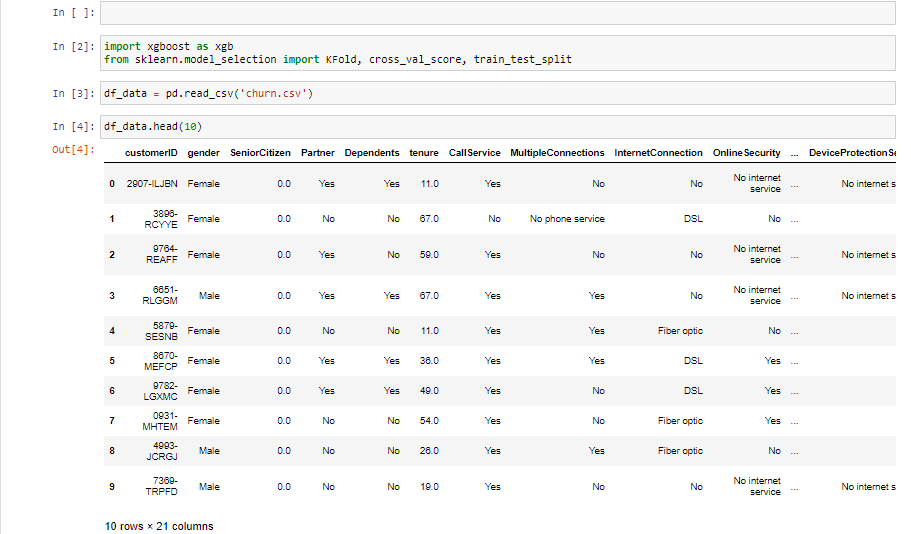
* Customer ID
* gender
* Senior Citizen
* Partner
* Dependents
* Tenure
* Call Service
* Multiple Connections
* Internet Connection
* Online Security
* Online Backup
* Device Protection Service
* Technical Help
* Online TV
* Online Movies
* Agreement
* Billing Method
* Payment Method
* Monthly Service Charges
* Total Amount
* Churn

**Section 2: Program Execution & results summary**

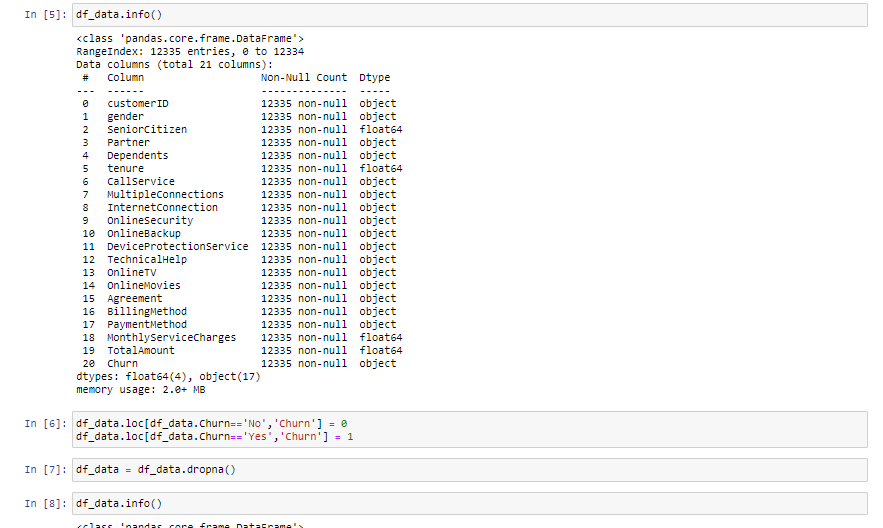
1. Loading Data and Explanation of Features



The above lines of codes the different libraries imported into the project.

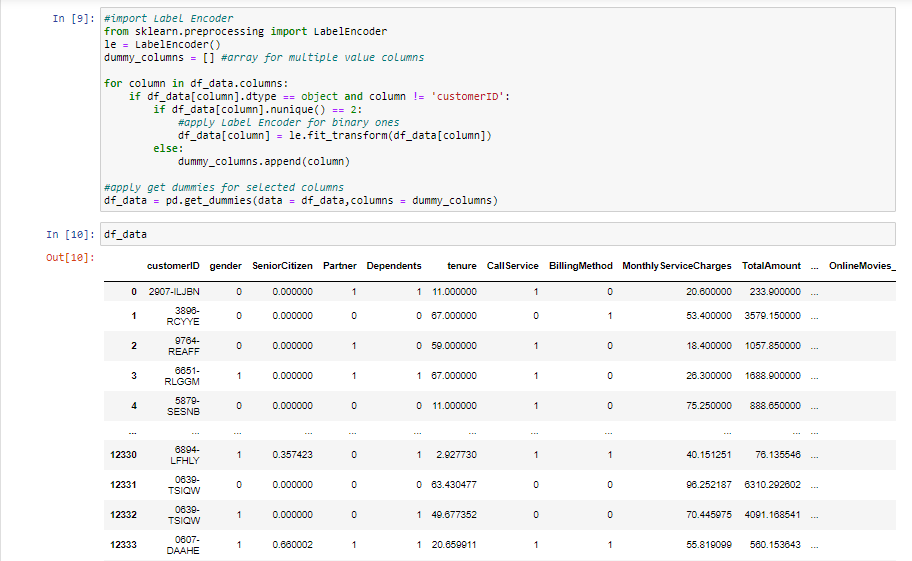
****

The above image describes about the top ten rows of customers and their description.

****

The above image shows the description of different features of datasets involved in the process.

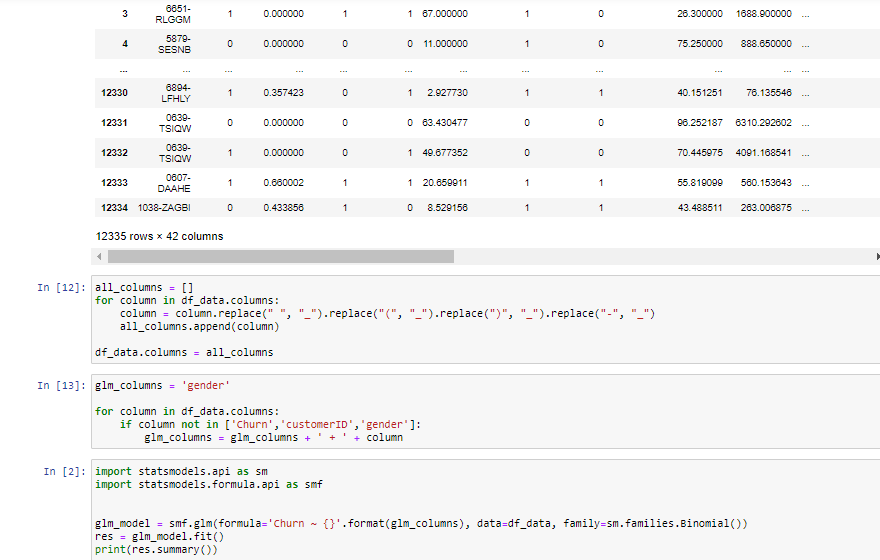
**Looks like our data is properly distributed. There is no NaN values and also the features’ types are proper.**



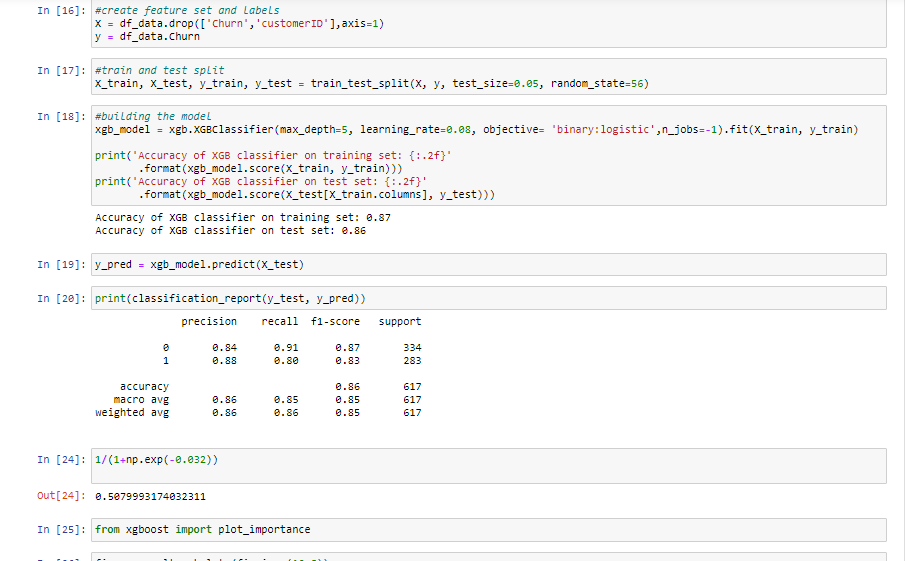
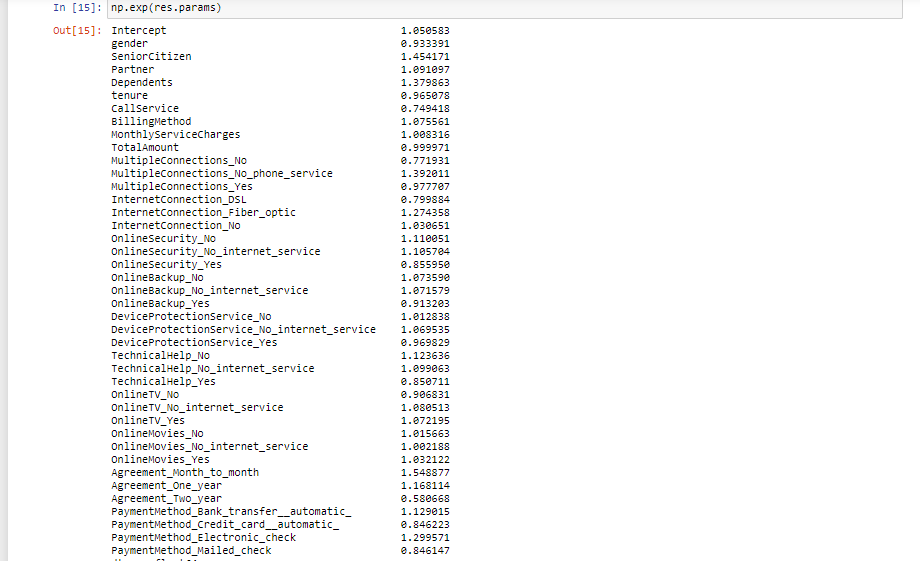
In the above image , the dummy columns have been added where the appropriate values are not inserted in the imported dataset.



In the above lines of code, the replace function is used to edit the names of the attributes.



In the above code, statistics formulae have been applied to fit the datasets into the model.



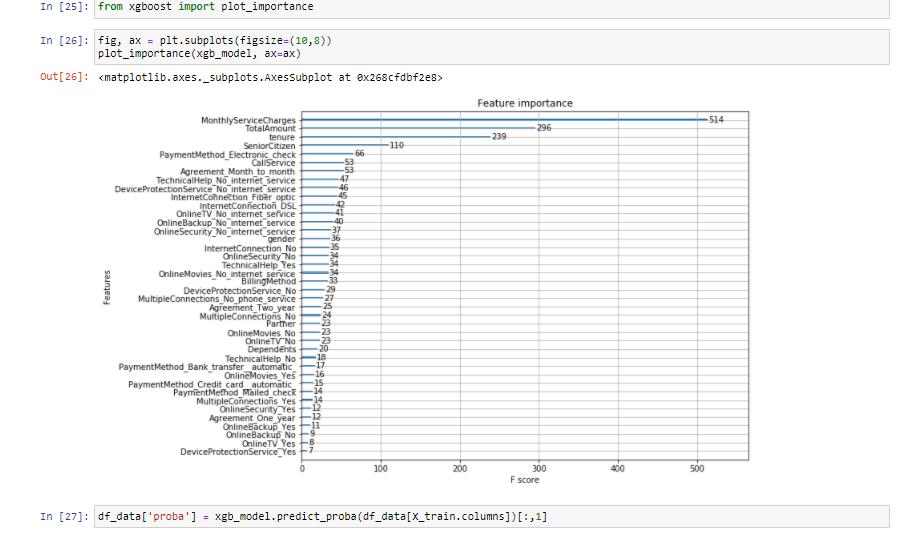
In the above lines of code, the feature set and labels are created and further thus splitted and trained with the use of algorithms.

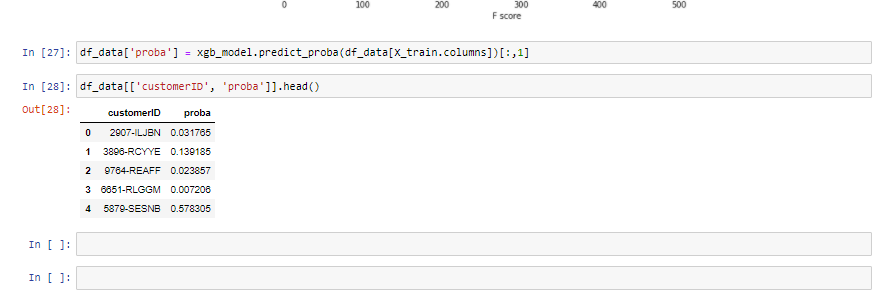
The predicted model is further printed with the set features and labels.

**Section 3: Conclusion**

We used a supervised machine learning implemented in the modern XGBoost library, to build a highly accurate classification model for the purpose of predicting customer churn.

Let’s see the results together in data frame.





XGBOOST machine learning techniques have been used to predict customer churn. By performing different models, it was aimed to get different perspectives and eventually compared their performance. With the help of the data visualizations and exploratory data analysis, the dataset was uncovered and features were explored deeply. The relation between features were examined.