6. TESTING

In the proposed system, the dataset is prepared upon collection of data from various resources. The dataset contains data of hundreds of customer details which include their customer-id, credit score, location, gender, age, tenure with the bank, bank balance, their credit card status, whether he is an active member or not. These data is separated as the train and test data, where the train data is 80% of the complete data and the test data is of 20% of complete data.

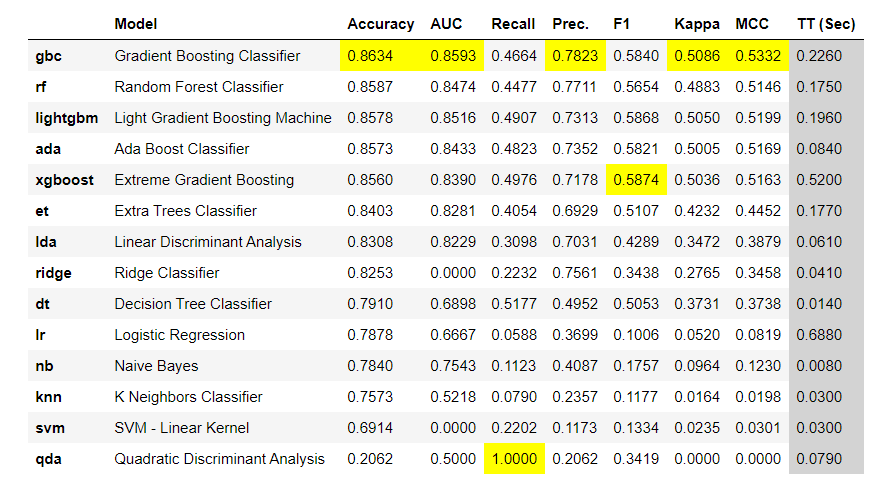
With the models used, the accuracies we obtained are:

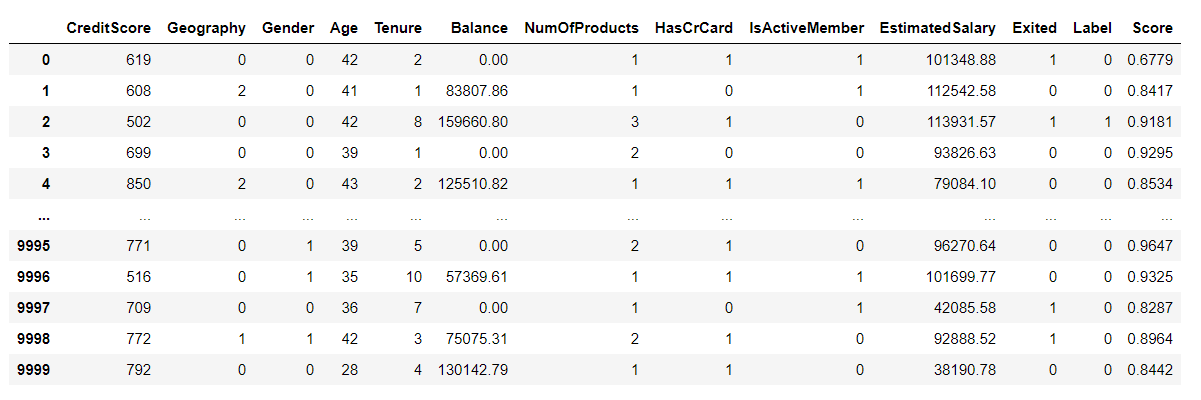
|  |  |
| --- | --- |
| Models Used | Accuracy |
| Logistic Regression | 77.30% |
| Decision Tree Classifier | 78.93% |
| Random Forest Classifier | 87.22% |
| K-NN Classifier | 83.52% |
| AdaBoost Classifier | 83.55% |
| Gradient Boosting Classifier | 84.99% |
| eXtreme Gradient Boosting Classifier | 86.72% |

Out of these all the models used, Random Forest classifier gave us the highest accuracy of 87.2%.

But the validation technique used here was Train/Test split validation. We also checked with another validation technique.

LOOCV(Leave one out Cross Validation)

Using LOOCV, we obtained the following accuracy scores:

We also displayed the predictions in a tabular format:

Also, as part of our proposed system, we also designed a real-time web application in which the bank can feed in the customer details which in turn will help the bank to understand which customer is at a risk of being churned.



