Indessa Bank Problem:

Please find below the steps in predicting the probabilities of a customer being default or not?

1. Data Cleaning and dealing with null values column by Column
2. Balancing the Data Set
   1. Case 1- The output had 23% Default and 77 % non default I converted to 1:1 by under sampling
   2. Case 2- The Output percent was left as default and analysis was performed.
3. EDA
4. 7 Algorithm Implementations and their accuracy score

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm Implemented | Judging Parameter | Under sampling | Imbalanced Dataset |
| Logistic Regression | AUC Score | 99.79% | 57.90% |
| KNN | AUC Score | 94.10% | 50.00% |
| Naïve Bayes | AUC Score | 79.47% | 53% |
| SVM | AUC Score | 99.47% | 58% |
| Descision Tree | AUC Score | 100% | 74.99% |
| Random Forest | AUC Score | 99.99% | 67.43% |
| Xgboost | AUC Score | 100% | 90% |

1. Testing

After getting the accuracy model was applied on test data using model.predict\_proba which leads to a numpy array of float 32. Which is converted to a data frame. Which is then copied to csv file

1. Submission File

The member\_id of submission file is concatenated with probabilities and submitted for 1st 100 members