BFSI Credit Score Case Study

Presented by:

- Samrudhi Yeginwar
- Abhishek Pawar
- Pratik Chide

Problem Statement

- SmartLend is a financial institution that recognizes that banking carries credit risk.
- Credit risk just alludes to the likelihood that installments might be made late or not the least bit, which might bring about income issues and may influence a bank's liquidity.
- Credit risk remains the primary cause of bank failures, despite advancements in the financial services sector.
- This aspect of risk management typically takes up more than 70% of a bank's balance sheet.
- As a result, SmartLend isn't the only financial institution that needs to know about customers' credit scores.
- As a result, this project's primary objective is to assist SmartLend in dividing its customers into various credit score brackets based on demographic and transactional data.

Objective

- SmartLend wants to understand the primary factors that contribute to a person's credit score based on their demographics and past transactions data.
- It wants to be able to predict which credit bucket a customer would fall into using data from the past.
- SmartLend intends to reduce their credit risk by utilizing this information to make lending decisions that are more robust and accurate.

Methodology

- Firstly, data understanding and applying steps accordingly to achieve the result.
- Data handling, cleaning and manipulation :

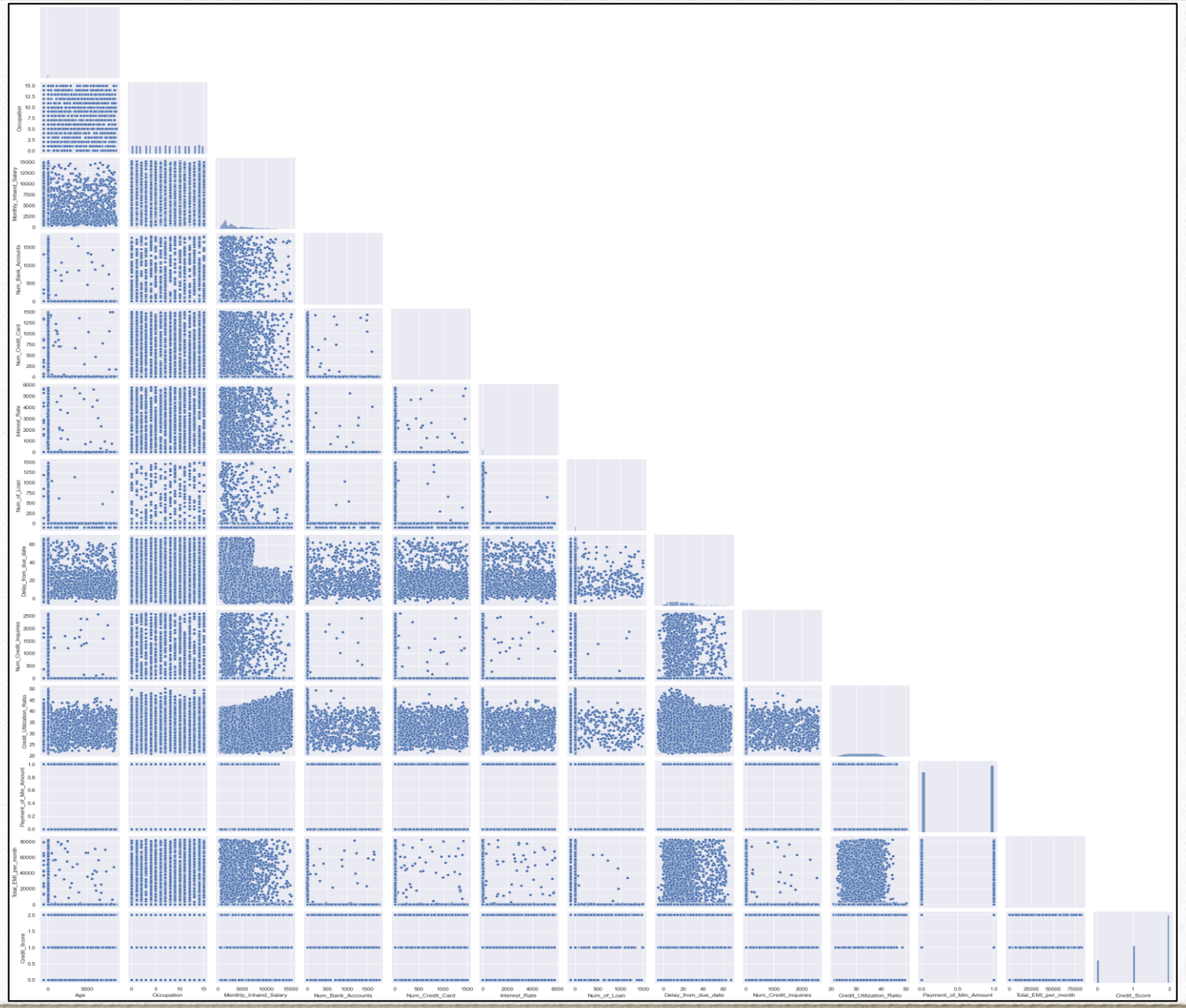
1. Importing all libraries and checking for the missing values and null values.
2. After that, handling them by dropping some "not required" values for analysis.
3. Imputation if necessary and also outliers handled.

- EDA(Exploratory data analysis).

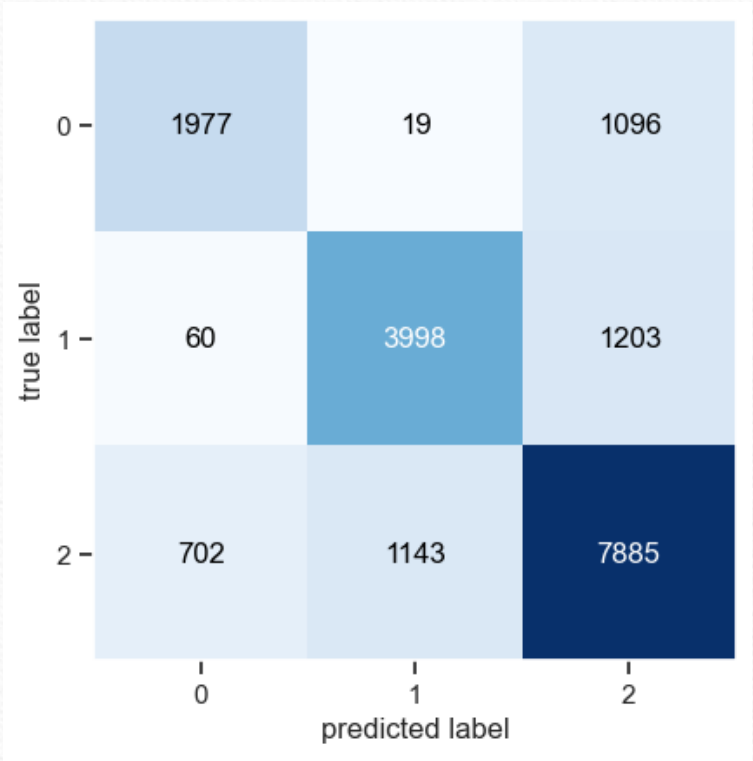
1. Sampling is done.
2. Bivariate data analysis also done with correlation coefficients and pattern between variables.

- Sampling is done – Under Sampling , Over Sampling.
- Classifiers are used.
- Validation of the models is done .
- Model prediction and presenting them with accurate graphs using visualization
- Conclusion

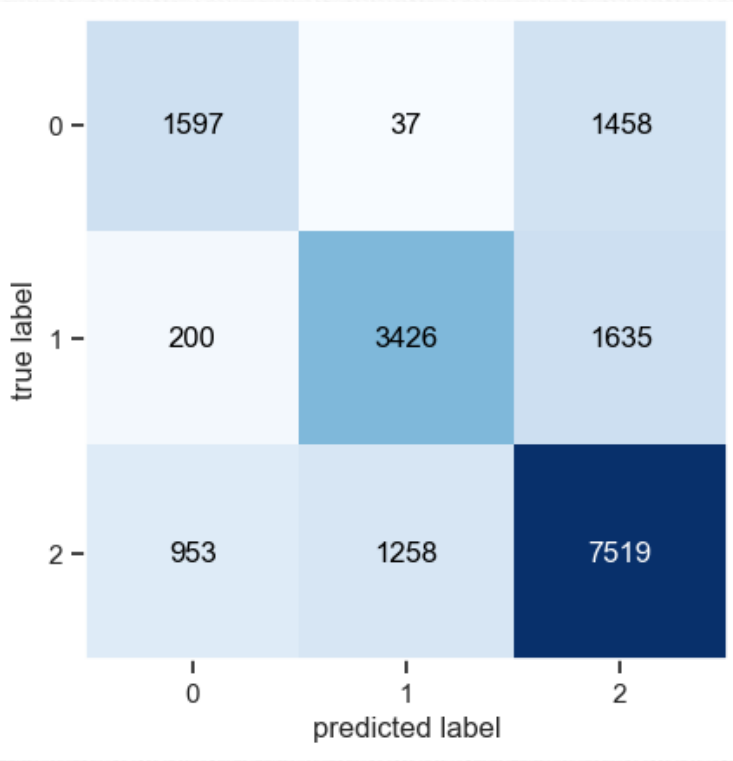
Pair-plot shows a visual representation of relation between all the features in the data



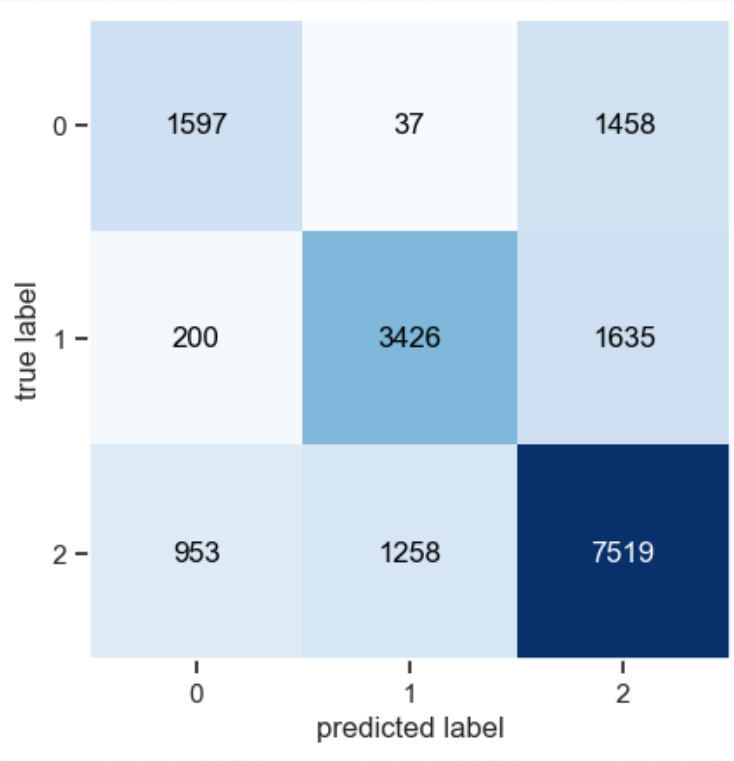
Selecting classifier for Sampling Methods



Random Forest

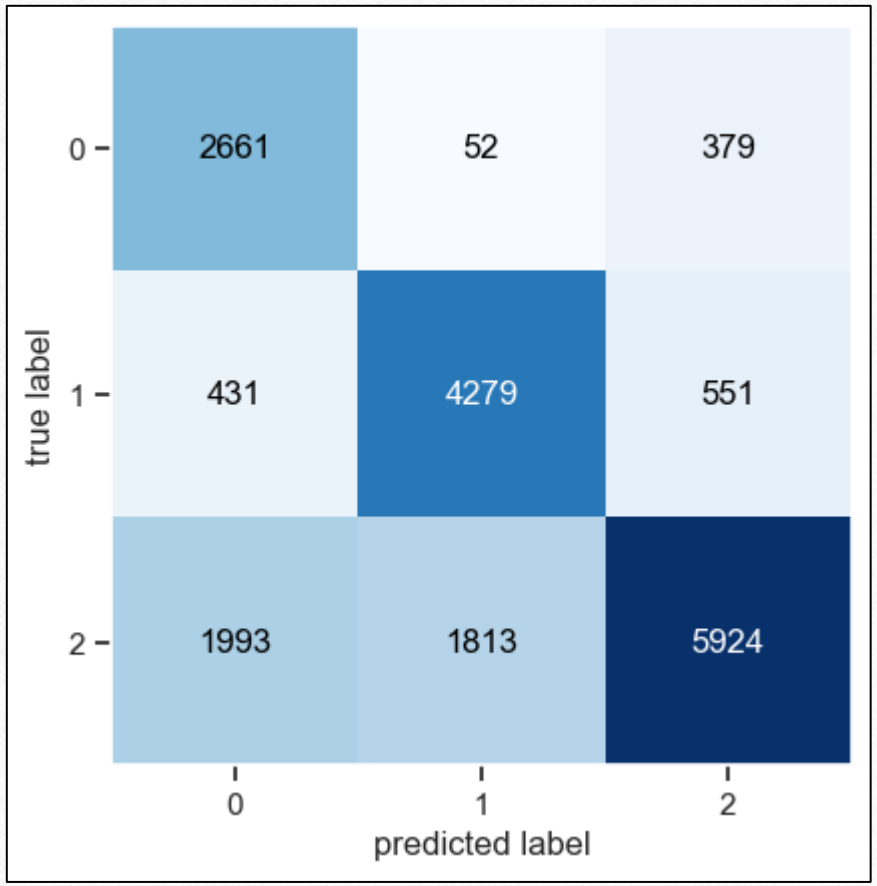


Gradient Boost

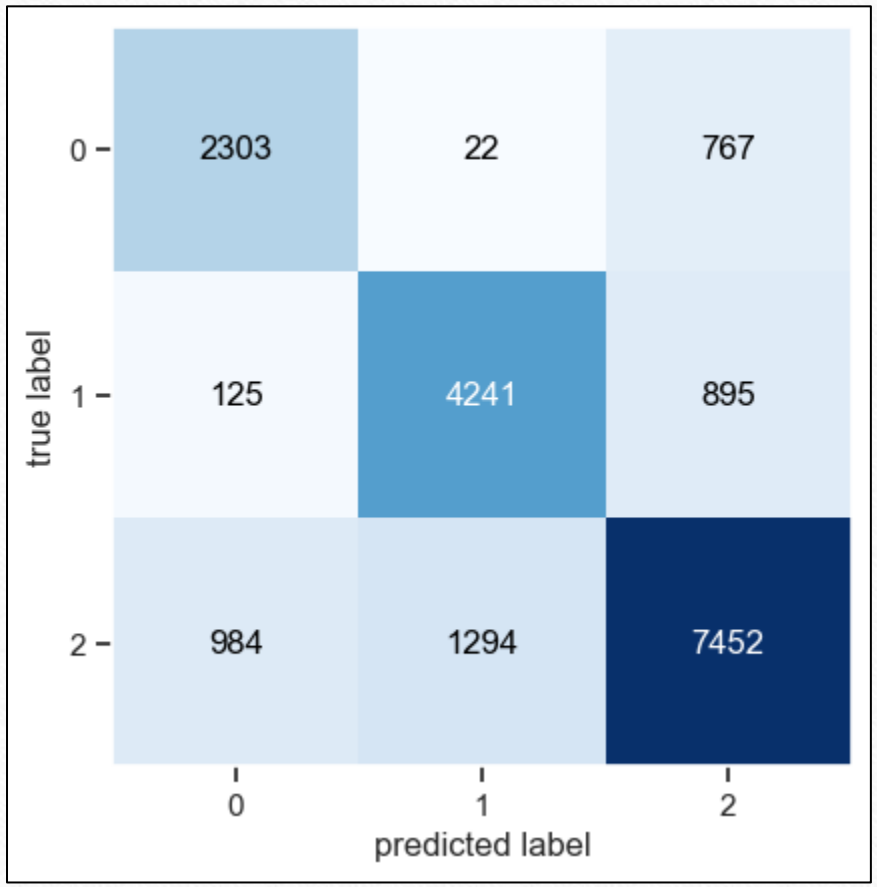


Adaboost

Sampling Methods



Under Sampling



Over Sampling

Recommendations

- Below idea is to find the Customer_ID which are predicted to have only poor Credit_Score in all the instances in data.
- For example-

1. Scenario 1: if Customer_ID 123 appeared 3 times in the data set and all the 3 times the predicted credit score is poor then it is definately a risky lending.
2. Scenario 2: if Customer_ID 123 appeared 3 times in the data set and all the 1 or 2 times the predicted credit score is poor and 2 or 1 times standard then it may not be a risky lending. Then further investigation can be done from the lender.
3. Scenario 3: if Customer_ID 123 appeared 3 times in the data set and all the 1 or 2 times the predicted credit score is poor and 2 or 1 times Good then it may not be a risky lending. Then further investigation can be done from the lender. then it is definitely a risky lending.

- In case of 1 Poor and 2 Good credit rating, there is a very high chance that it is not a risky lending.
- When both Good and Standard column gives null values, then it is definitely a risky lending.

Conclusion

- The factors affecting a person's Poor credit score are identified.
- The model differentiates between Poor, Standard, Good Credit scores.
- The model correctly predicts customers having Poor credit score
- Based on all this observations the model is able to predict the faulty customers, ensuring that the institution is only lending money to customers with a relatively good credit score.

THANK YOU!!!