PNBHFL Credit Risk Analysis

INTERNSHIP PROJECT

JULY'22-AUG'22

Organization and Aim of the Project

PNB Housing Finance Limited (PNB Housing) is a registered housing finance company that It provides housing loans to individuals and corporate bodies for purchase, construction, repair and upgradation. It also provides loans for commercial space, loan against property and loan for purchase of residential plots.

Since the company needs to minimize risk for profit maximization and market expansion it need to carefully evaluate credit scores of the customers to make appropriate lending decisions avoid the incur of losses to the company.

Thus this project will help analyse the same and help the organization in loan lending decisions i.e., whether the loan lent is a good loan or a bad loan for the company.

High level solution architecture to solve the objective

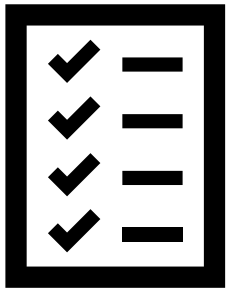
Labelled historic dataset on credit is collected , cleaned, understood and pre-processed.

Building and training the model(Logistic Regression) with train set

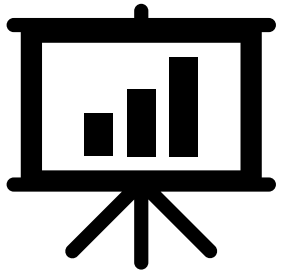
Deriving Analysis and testing the Risk ML Model on the test set.

Lending loan predictions are made for the test set based on the predicted outcome and the actual outcome.

The approach for the architecture

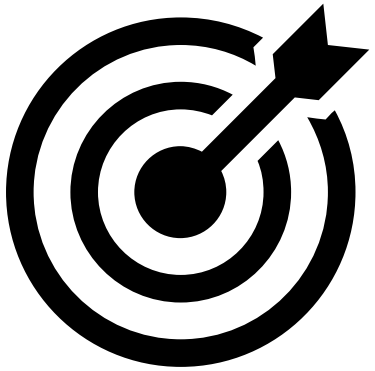


- **Assumptions**
  - Missing values are replaced with mean
  - Customer ID is excluded from analysis

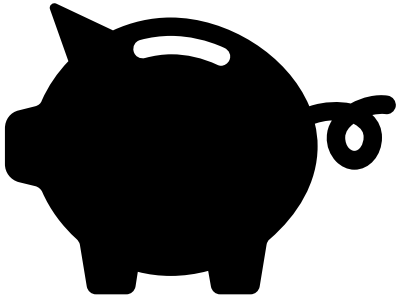


- **Trained a classification model**
  - Predicting likelihood of loans being good/bad
  - Using Logistic Regression Classifier for the prediction

Deliverable of in-house risk model

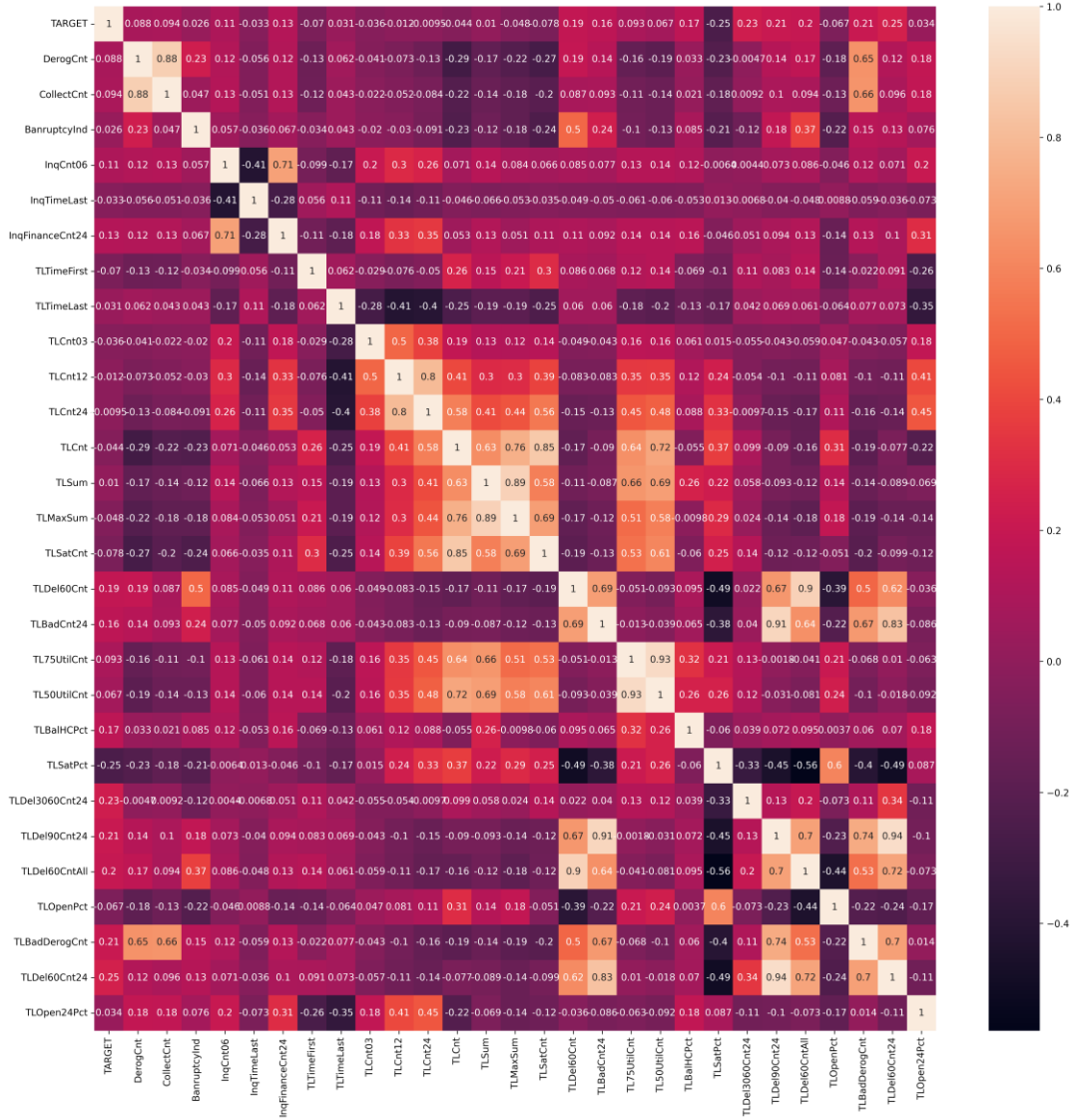


82.66%
Model accuracy achieved on the test set

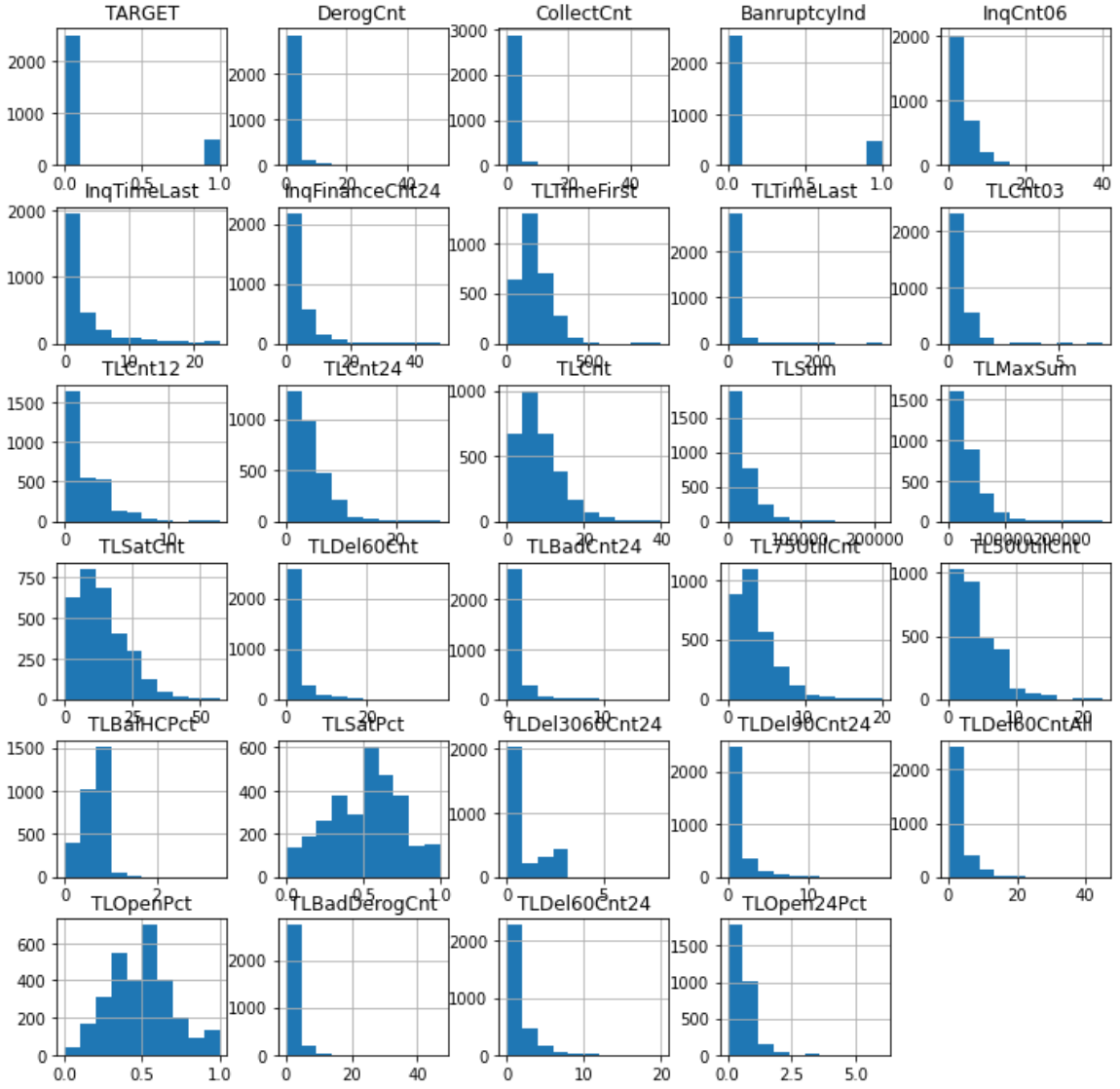


Zero
Operational cost to the business

Exploratory Data Analysis

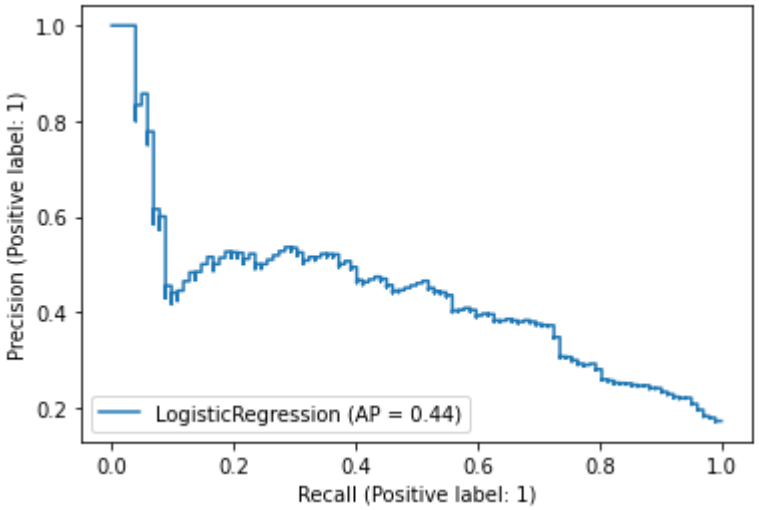


Heatmap to show the distribution of data across the dataset

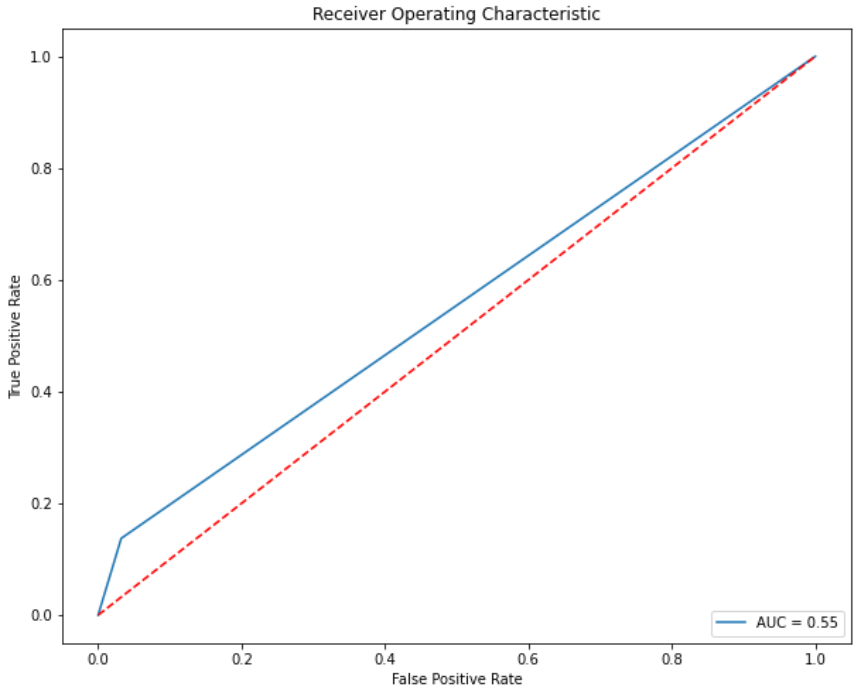


Distribution of all the attributes for understanding of the dataset

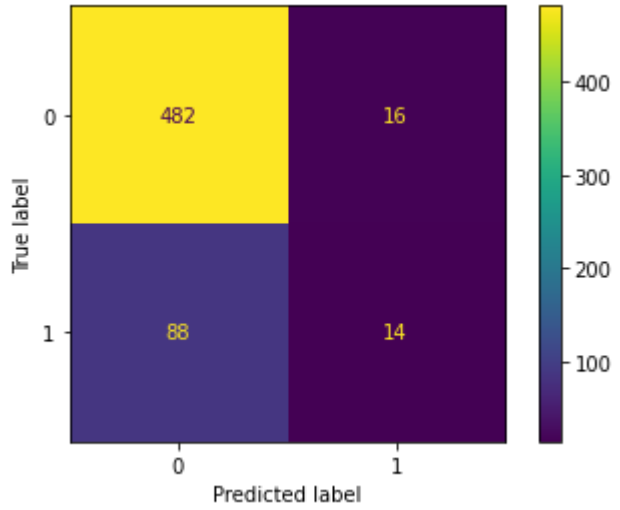
Exploratory Data Analysis



Precision recall graph showing the performance of the classifier



Plot of ROC Curve for the model that shows the tradeoff between the true positive rate and the false positive rate



The confusion matrix shows that the classifier has a accuracy rate of 82.66 and an error rate of 17.34

Overall Outcome

- Predictions for good/bad loans scores are made for each customer in the test set
- Predicting the overall score whether the loan is good or bad loan.
- Writing all the above along with the actual outcome in an excel file.
- Many other upgradations can be made to the project according to the client's requirements

	probGL_0	probBL_1	predicted_TARGET	Actual Outcome
0	0.040435	0.959565	1	1
1	0.937798	0.062202	0	0
2	0.711622	0.288378	0	0
3	0.906753	0.093247	0	0
4	0.877058	0.122942	0	0

Sample of how the output file(excel) looks like with first few results

Thank You!