# B F S I-Credit Risk Assignment

By
Kalyani Masina

# B F S I –CREDIT CARD ASSIGNMENT

OBJECTIVE: The objective is to build a statistical model to estimate borrowers' Loss Given Default(LGD)

$$LGD = \frac{Loan\ Amount - (Collateral\ value + Sum\ of\ Repayments)}{Loan\_Amount}$$

# BACKGROUND PROCESS

Credit risk analytics in the context of the banking sector and model a common metric used for estimating the expected credit loss (ECL)

ECL method is used for provisioning the capital buffer to protect banks against possible default of the customers.

Expected credit loss =Exposure at default x Probability of Default x

The loss given default(LGD) isa measure of the amount of loss that a bank is expected to incur in the event of a

# DATA SOURCES

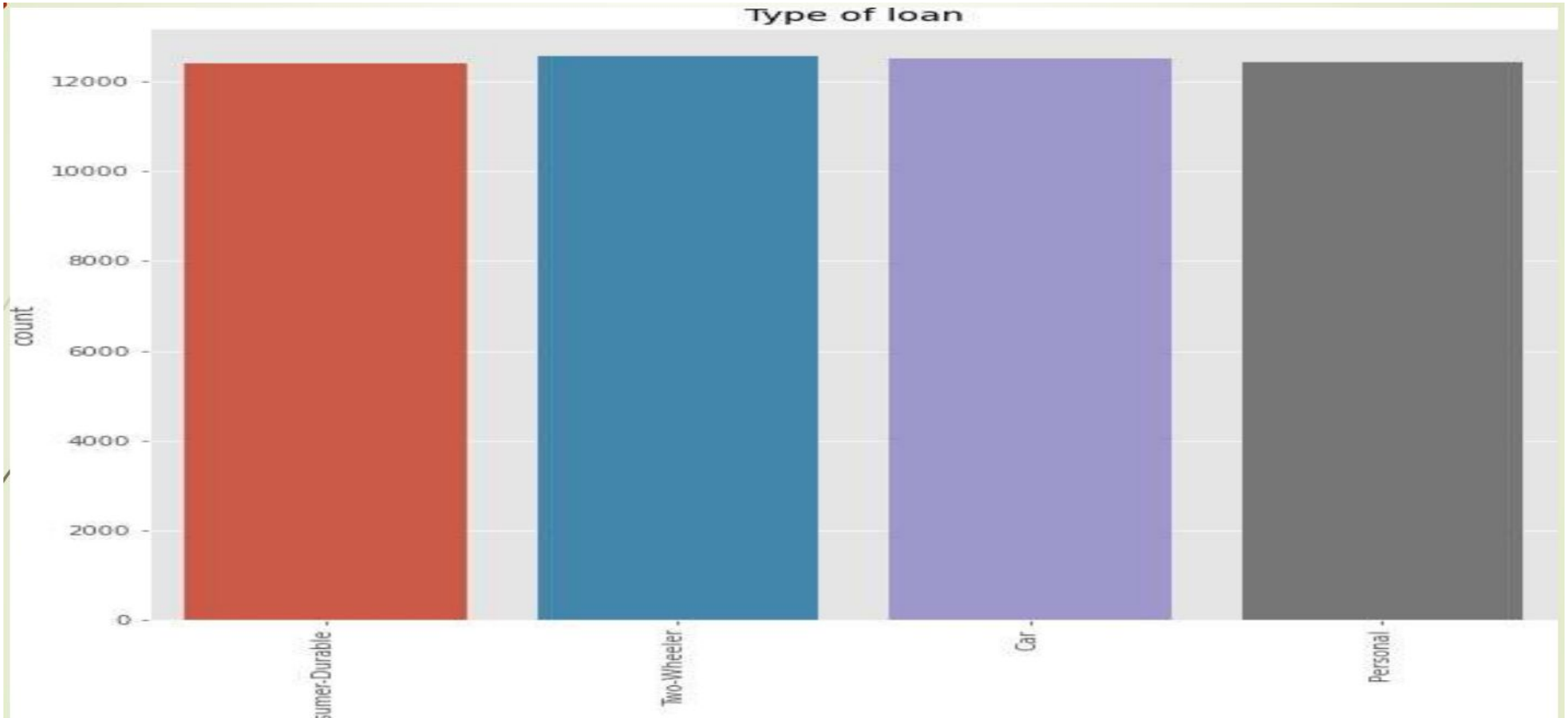
**Used 3 Data sets for model Building:**

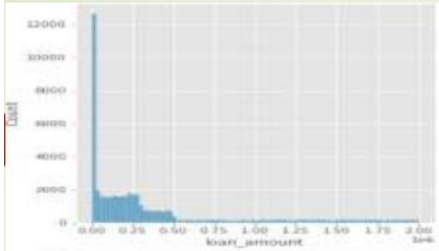
- The main_loan_base data set contains information about loan accounts and other relevant information for the corresponding borrowers.
- The repayment base data set contains information about the repayments received by the banks in the form of EMIs or through other collection efforts
- The monthly_balance_base contains the information pertaining to the monthly balance statements in the borrower's accounts

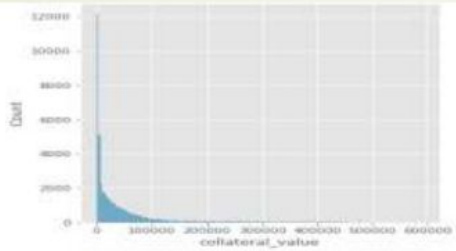
# PROCESSING OF DATA

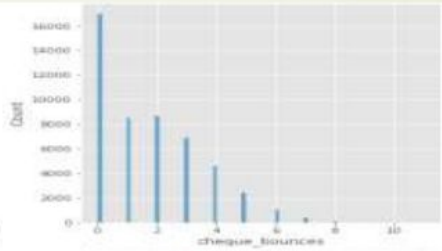
- For each data set converted Data types if necessary
- Merging the data sets and created target variable(LGD)
- Exploratory Data Analysis has been performed
- Variable Transformation
- Dummy Encoding Scaling
- using Standard Scale

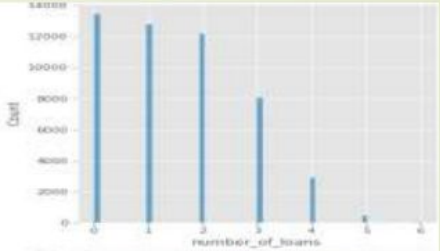
# EDA:Monthly EMI also car loan is much higher compared to other loans

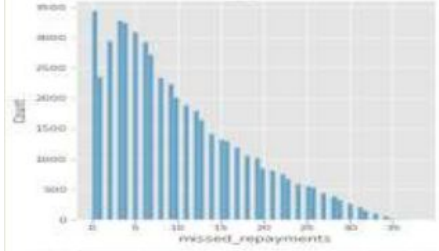


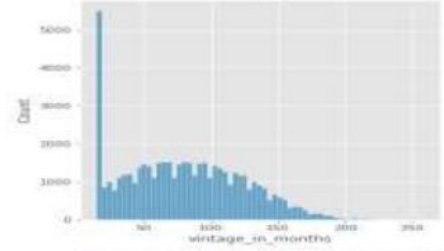

Count
loan_amount

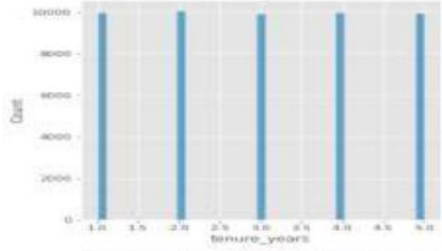

Count
collateral_value

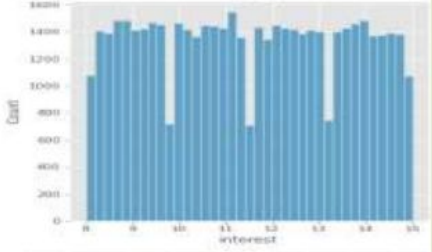

Count
cheque_bounces

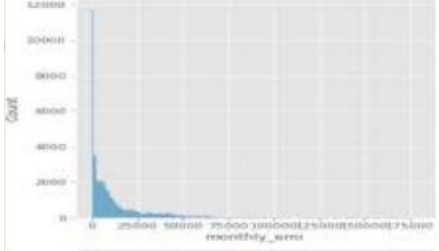

Count
number_of_loans

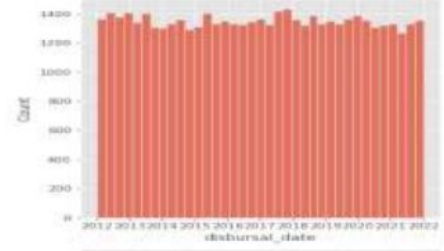

Count
missed_repayments

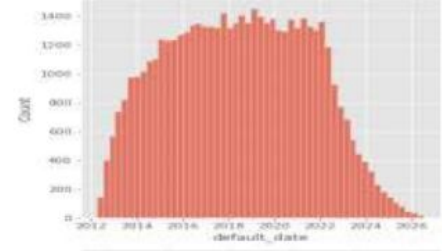

Count
vintage_in_months

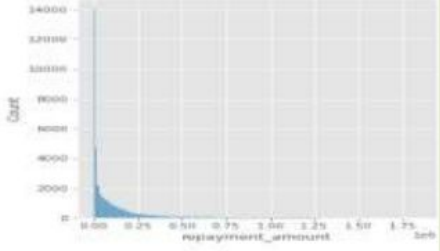

Count
tenure_years

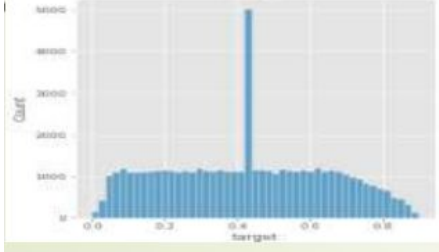

Count
interest

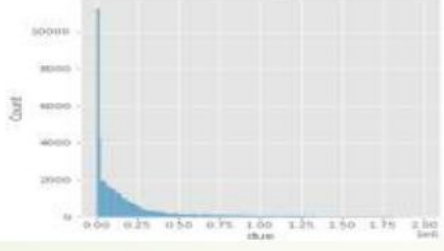

Count
monthly_emi

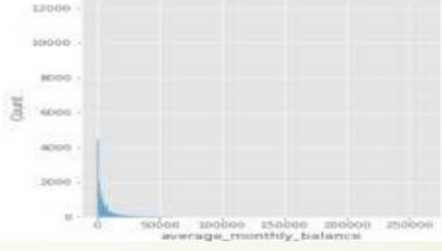

Count
disbursal_date

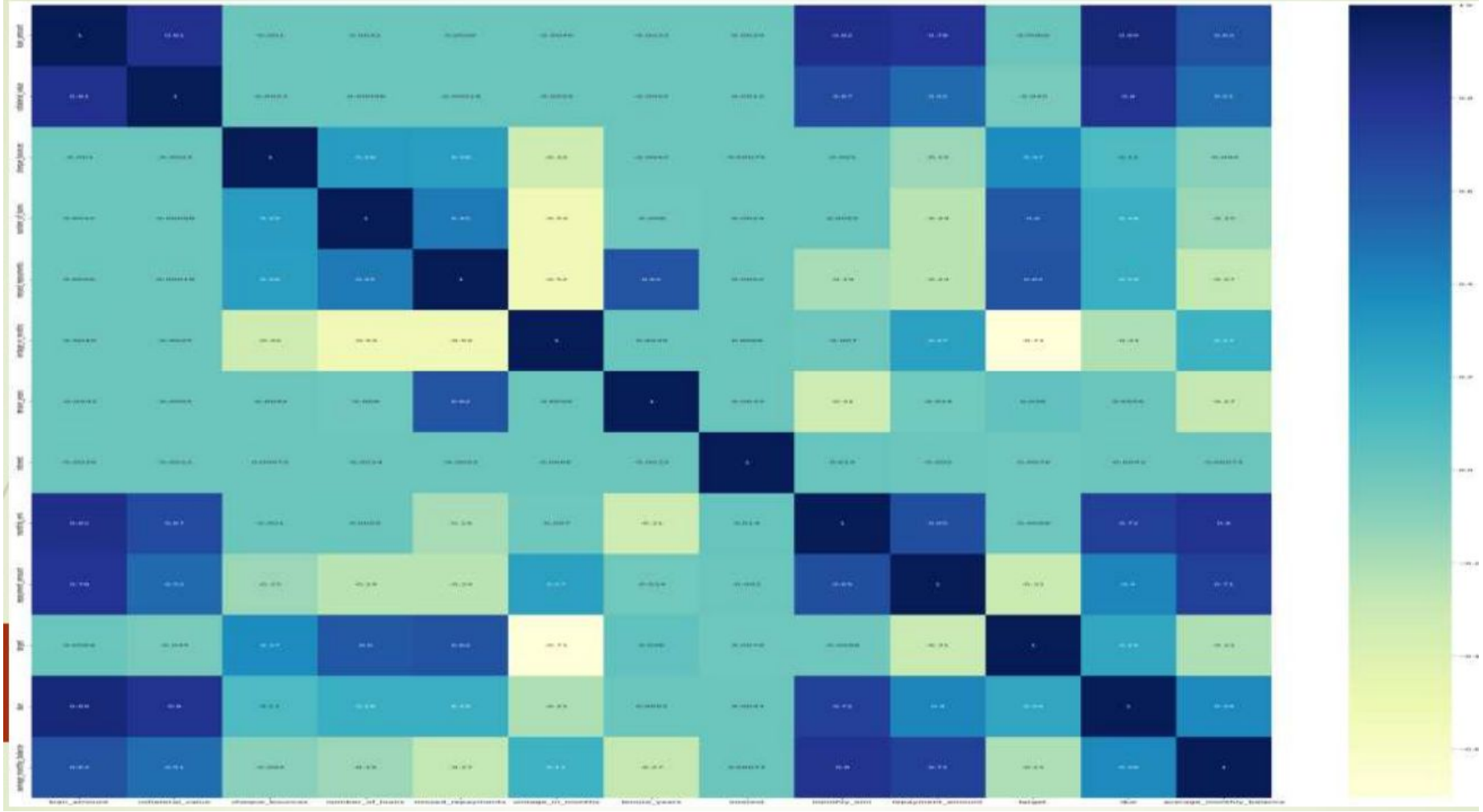

Count
default_date


Count
repayment_amount


Count
target

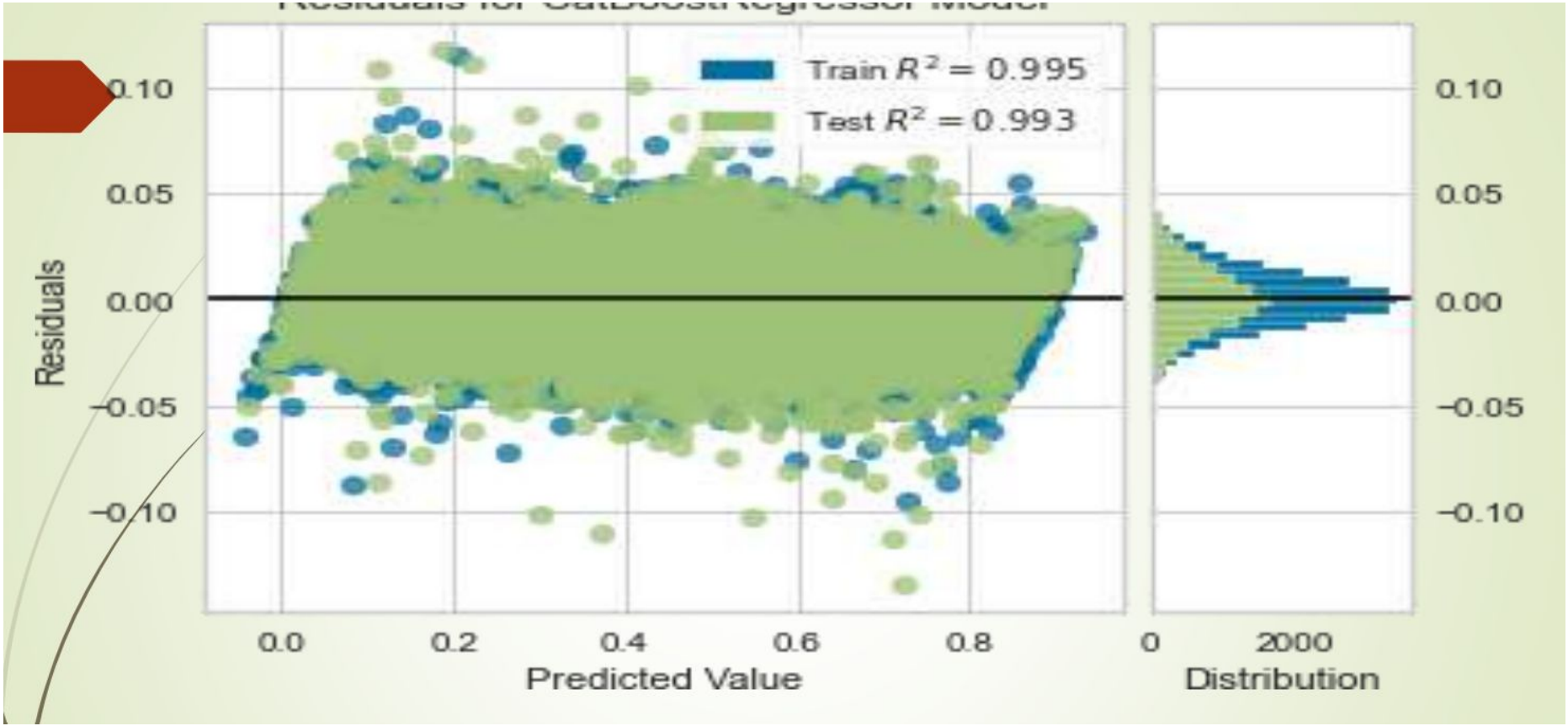

Count
due


Count
average_monthly_balance



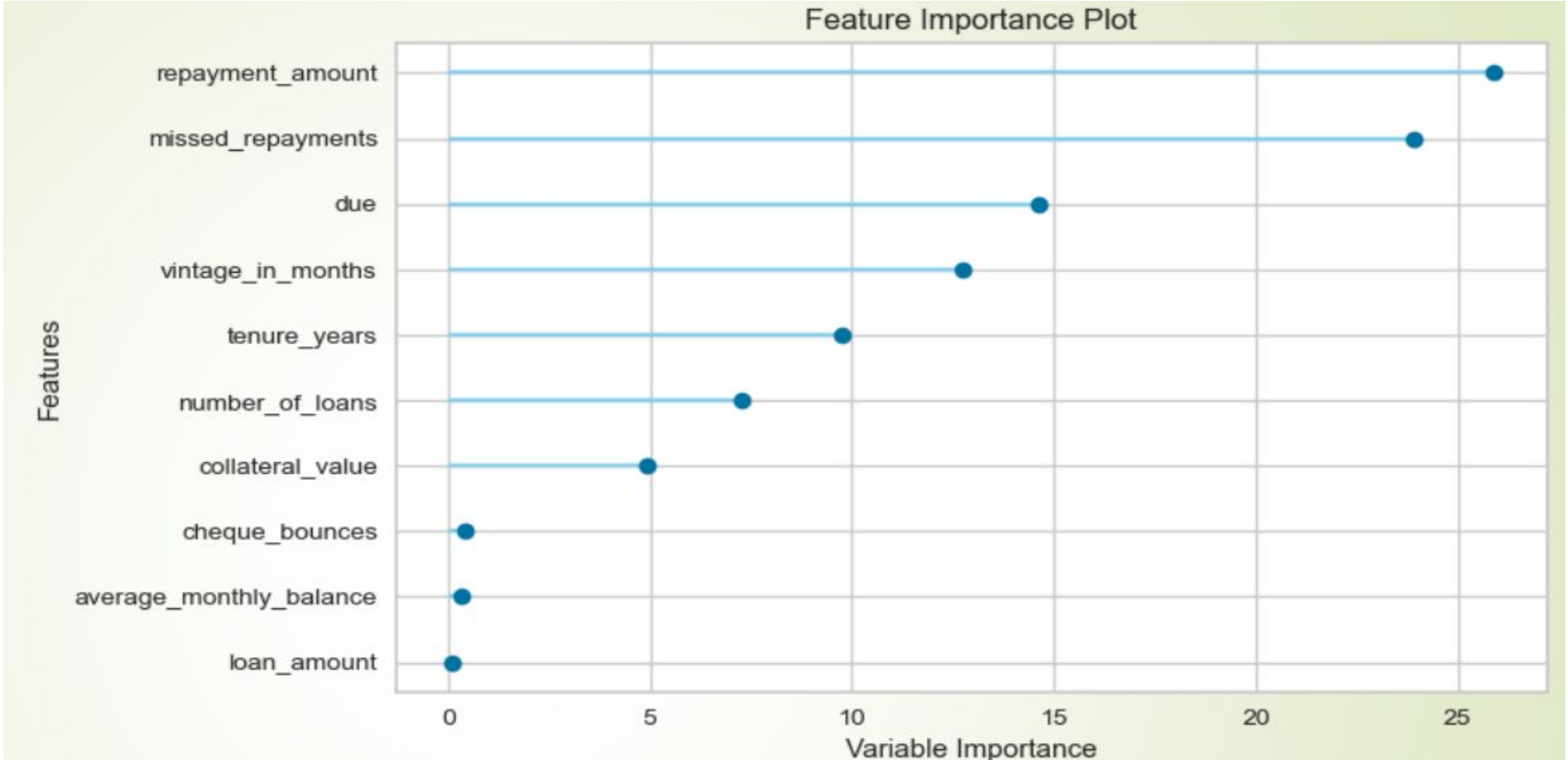
# MODEL BUILDING

- Used various models like Multiple Linear Regression, Randomforestregressor,Gradient Boosting Regressor, XGBoost Regressor, Adaboostregressor, Elastic Net: Hybrid Regularized Model,LightGBM for model building.
- Used R Squared as a performance metrics.
- XGBoost has given us99.5%Rsquaredon testdata across the models.




Residuals for catboost regressor

# REGRESSION INTERPRETATION



# RECOMMENDATIONS

- We should focus more on car and two-wheeler
- Missed Repayment customers with high repayment amount should be highlighted
- Customer’s due factors and tenure are another subset of influencers to predict the LGD of the customers