Predicting Credit Risk in Lending: Data-driven Insight and Modeling

Created by:
Muhammad Iqbal Mudzakky
Email : iqbalmudzakky@gmail.com
LinkedIn II Github

"Iqbal is a junior data scientist with a robust background in exploratory data analysis, preprocessing, modeling, visualization, and providing actionable insights. He possesses extensive experience in handling diverse data types, particularly in the Fintech and E-Commerce industries. His expertise is evident through successful completion of numerous supervised and unsupervised learning projects, which demonstrate his ability to extract valuable information from data, develop accurate models, and effectively communicate findings. With a proven track record in multiple facets of data science, Iqbal serves as a highly valuable asset to any team or organization seeking expertise in data-driven decision making and problem-solving. "

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com

PROJECT BACKGROUND

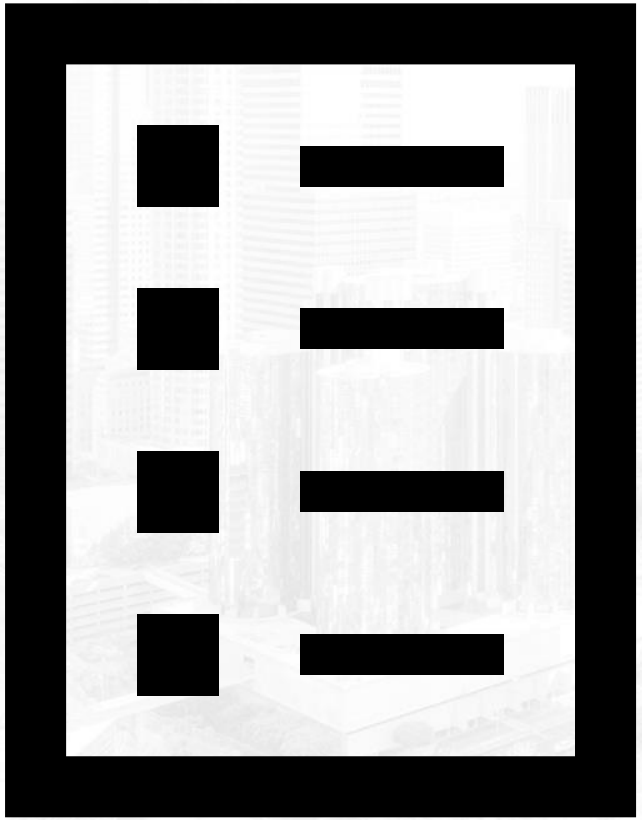
Company Background

"The challenge in the lending business is accurately assessing credit risk. Errors in assessment can lead to financial losses and negative reputation impacts. Technology and data analysis are now crucial in addressing this challenge. Through a credit risk prediction model that utilizes data such as credit history and income, we aim to enhance decision-making. While specific borrower company details are not available, our project's focus is on a solution that improves loan approval efficiency. Thus, this project contributes to business accuracy, sustainability, and customer satisfaction."

The Main Issue/Problem?

1. Credit Risk Uncertainty
2. Limitations of Human Decision-Making
3. Predictive Value of Data

Workflow



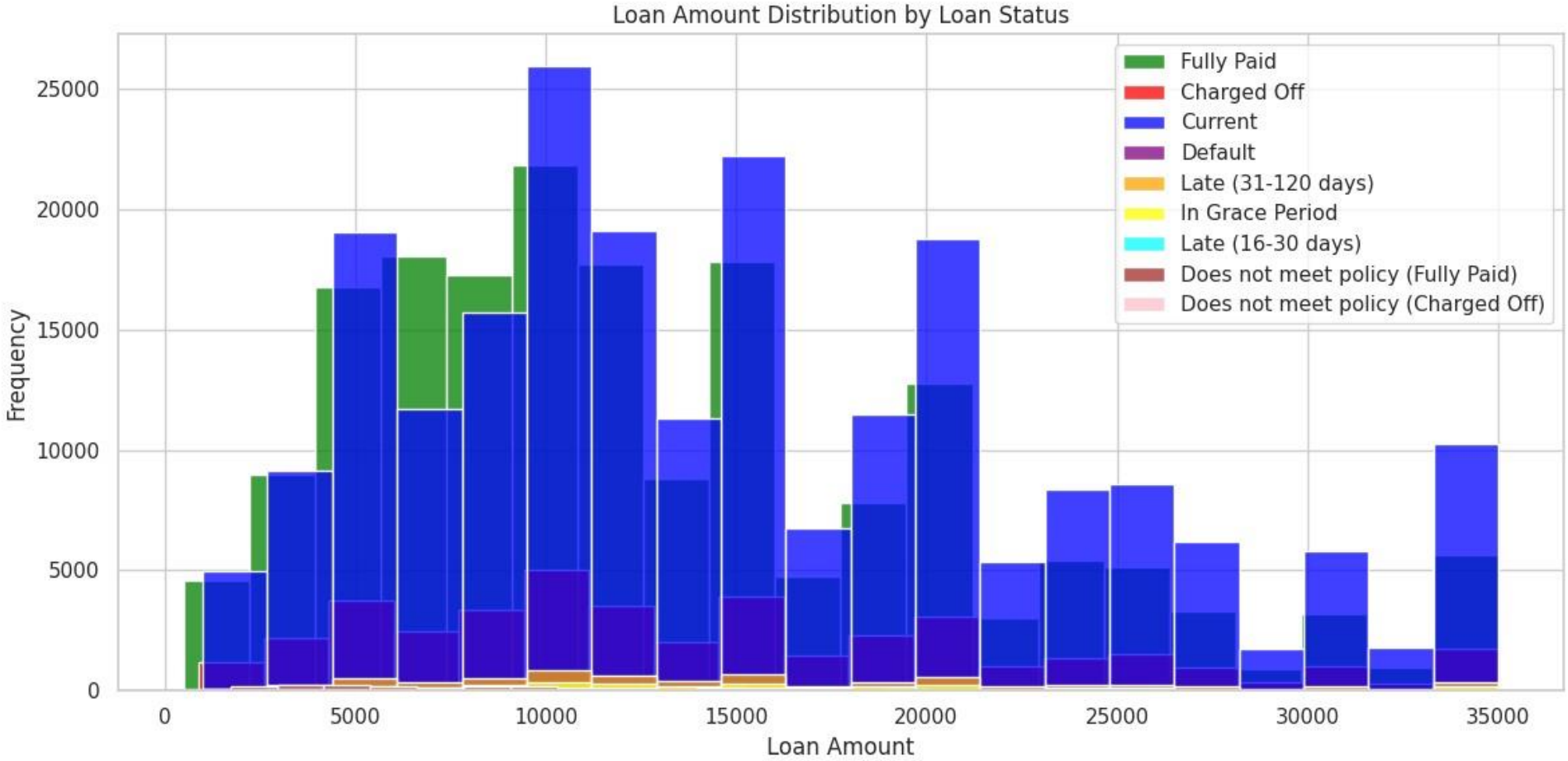
1. Exploratory Data Analysis

2. Modeling

3. Business Recommendation

EXPLORATORY DATA ANALYSIS

Loan Amount Distribution by Loan Status



Loan Amount Distribution by Loan Status

Rakamin Academy

1. Borrower Financial Stability:
The abundance of loans with a "Current" status may indicate that the majority of borrowers possess financial stability, allowing them to make regular installments and maintain a smooth credit status.

2. Risk Level:
Since the majority of loans are categorized as "Current" and "Fully Paid," it could suggest that lending companies tend to select borrowers with lower-risk profiles or have implemented effective selection processes.

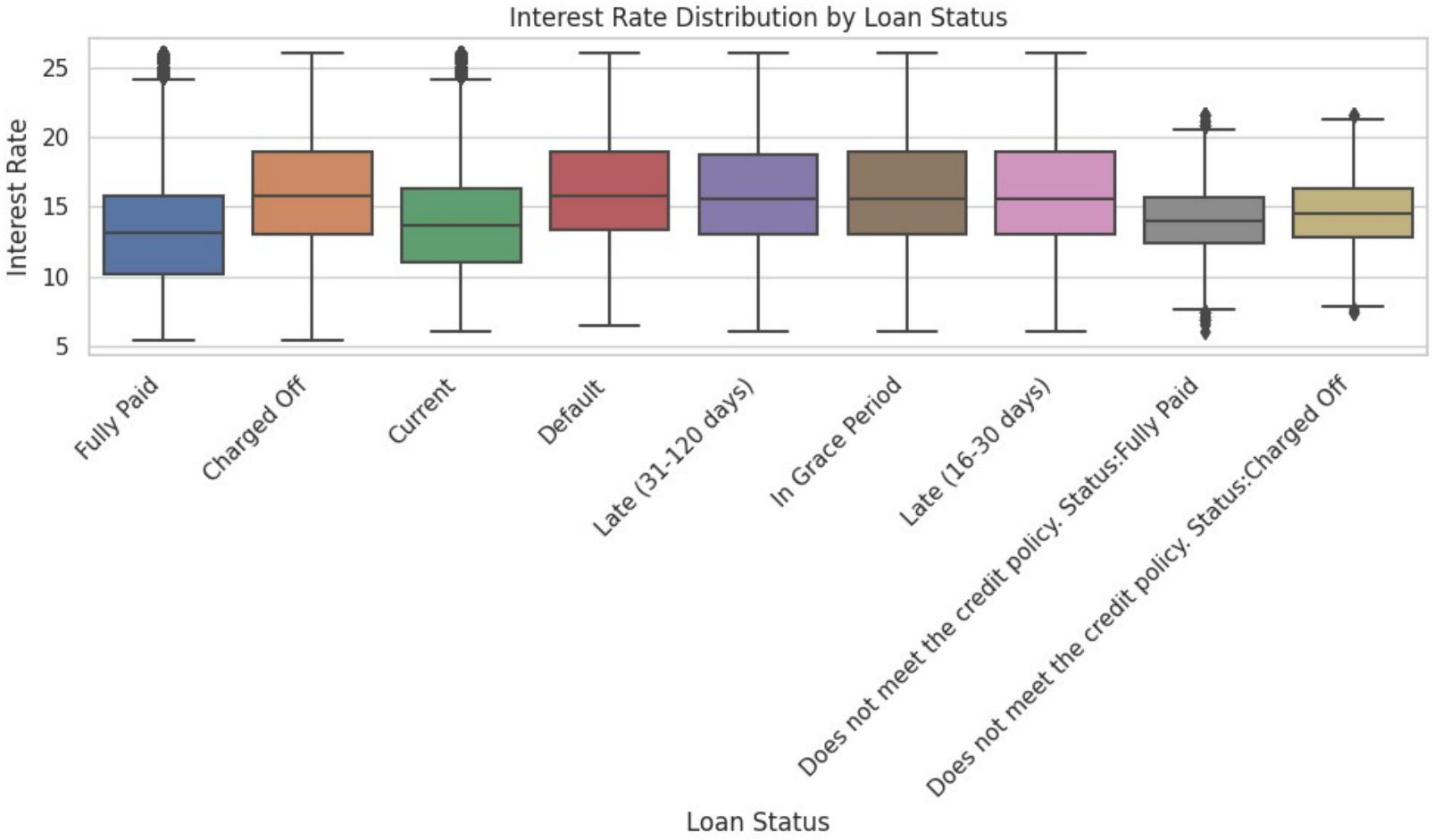
3. Payment Patterns:
The increased number of loans with a "Fully Paid" status might indicate that most borrowers are capable of managing their payment obligations and completing their loans according to plan.

4. Credit Management Effectiveness:
With more loans marked as "Fully Paid" compared to those that have encountered issues like "Charged Off," "Default," or "Late," it provides insights into how well the company manages credit risks.

5. Potential Revenue Growth:
Focusing on loans that are "Current" and "Fully Paid" could signal opportunities to develop products or services that are more appealing to borrowers with low-risk profiles.

6. Payment Speed:
From here, further analysis can delve into how long, on average, loans remain in the "Current" status before transitioning to other statuses. This can aid in identifying payment patterns and borrower behavior.

Interest Rate Distribution by Loan Status



Interest Rate Distribution by Loan Status

Rakamin Academy

1. Credit Risk Patterns:
A lower range of interest rates on "Fully Paid" and "Current" statuses may indicate that loans with lower risk or more regular repayments tend to have lower interest rates.

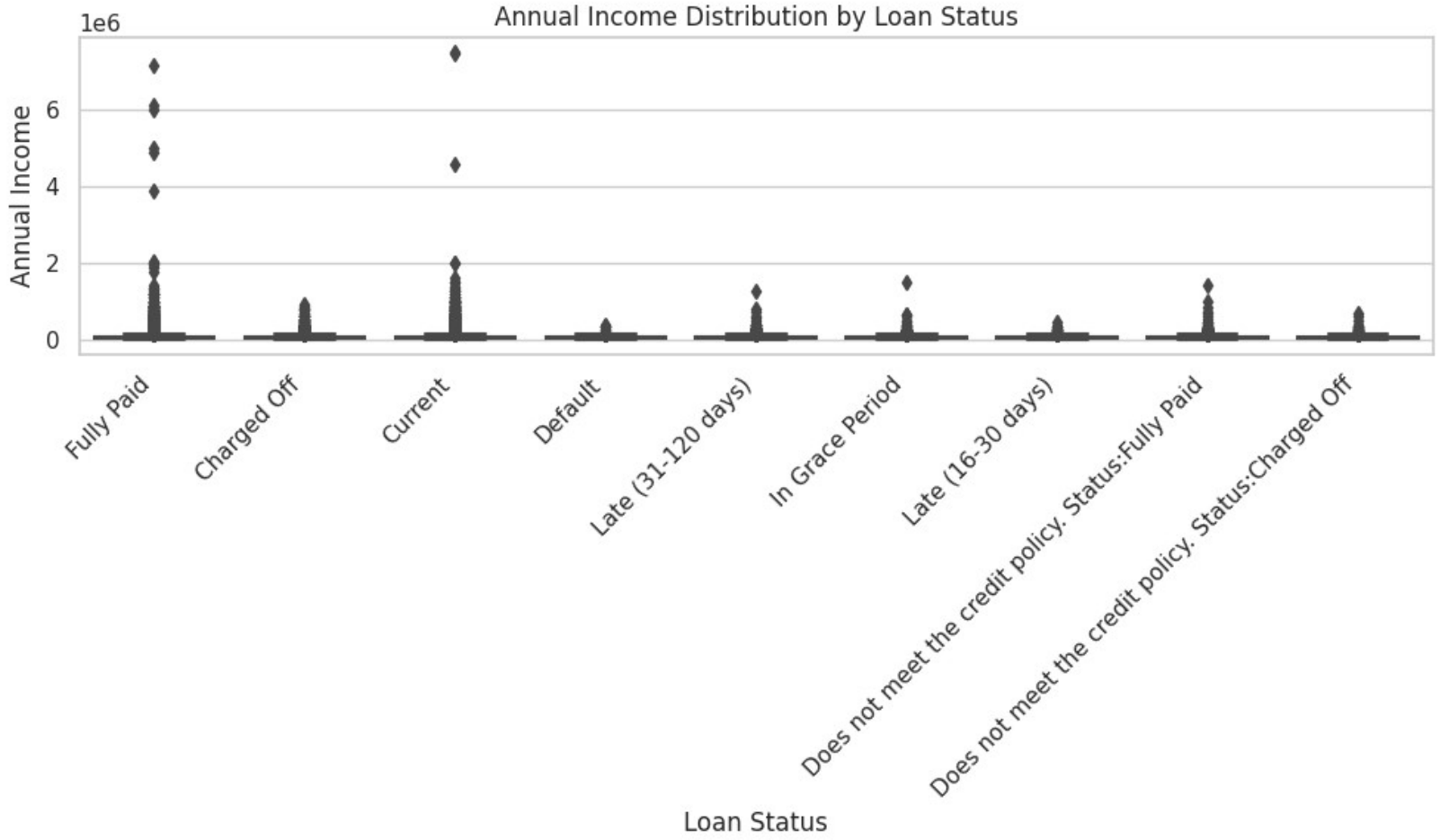
2. Borrower Profile:
Borrowers with better risk profiles or payment histories might receive lower interest rates. Conversely, borrowers with higher risk or poor payment histories might be charged higher interest rates.

3. Borrower Trends:
Borrowers with higher-risk statuses might understand that they are likely to pay higher interest rates, so they might prefer to avoid payment delays or defaults.

4. Business Strategy:
A higher range of interest rates on statuses like "Charged Off," "Default," "Late (31-120 days)," "In Grace Period," and "Late (16-30 days)" may reflect the company's policy to mitigate risk and prevent further losses by offering higher payment incentives.

5. Relationship Between Interest Rates and Risk:
Significant differences in interest rate ranges between statuses like "Fully Paid"/"Current" and other risk statuses (such as "Charged Off" and "Default") could suggest that interest rates are used as a tool to manage credit risk. In this case, higher interest rates might be applied to compensate for higher risk in those loans.

Annual Income Distribution by Loan Status



Annual Income Distribution by Loan Status

1. Financial Resilience:
Borrowers with higher annual incomes likely possess better financial resilience to handle challenging financial situations, thereby minimizing the risk of delays.

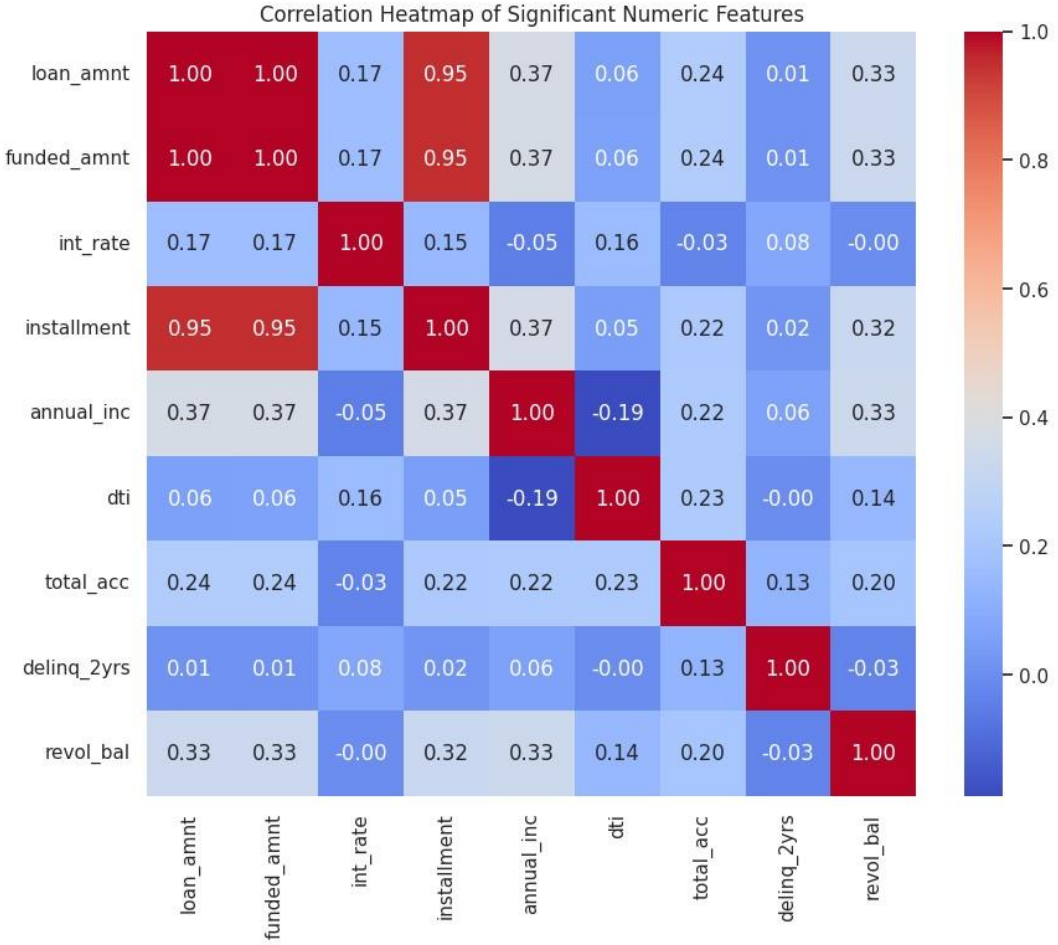
2. Low Credit Risk:
The presence of outliers and a higher distribution of annual incomes in "Fully Paid" and "Current" statuses might suggest that borrowers with lower-risk profiles (evidenced by good payment history and high income) are more likely to successfully repay loans.

3. High-Quality Borrower Profile:
Borrowers with "Fully Paid" and "Current" statuses and high incomes can be considered high-quality borrowers, potentially having undergone a stricter selection process before obtaining loans.

4. Business Strategy: The company might have offered lower interest rates to borrowers with high incomes and low-risk profiles as an incentive for them to repay loans successfully. This approach could help maintain a low credit risk level.

5. Payment Capacity:
Outliers in feature annual_inc for "Fully Paid" and "Current" statuses could indicate that borrowers with higher incomes tend to have better payment capacity, making them more capable of repaying loans on time or ahead of schedule.

Correlation Heatmap of Significant Numeric Features



Correlation Heatmap of Significant Numeric Features

Rakamin Academy

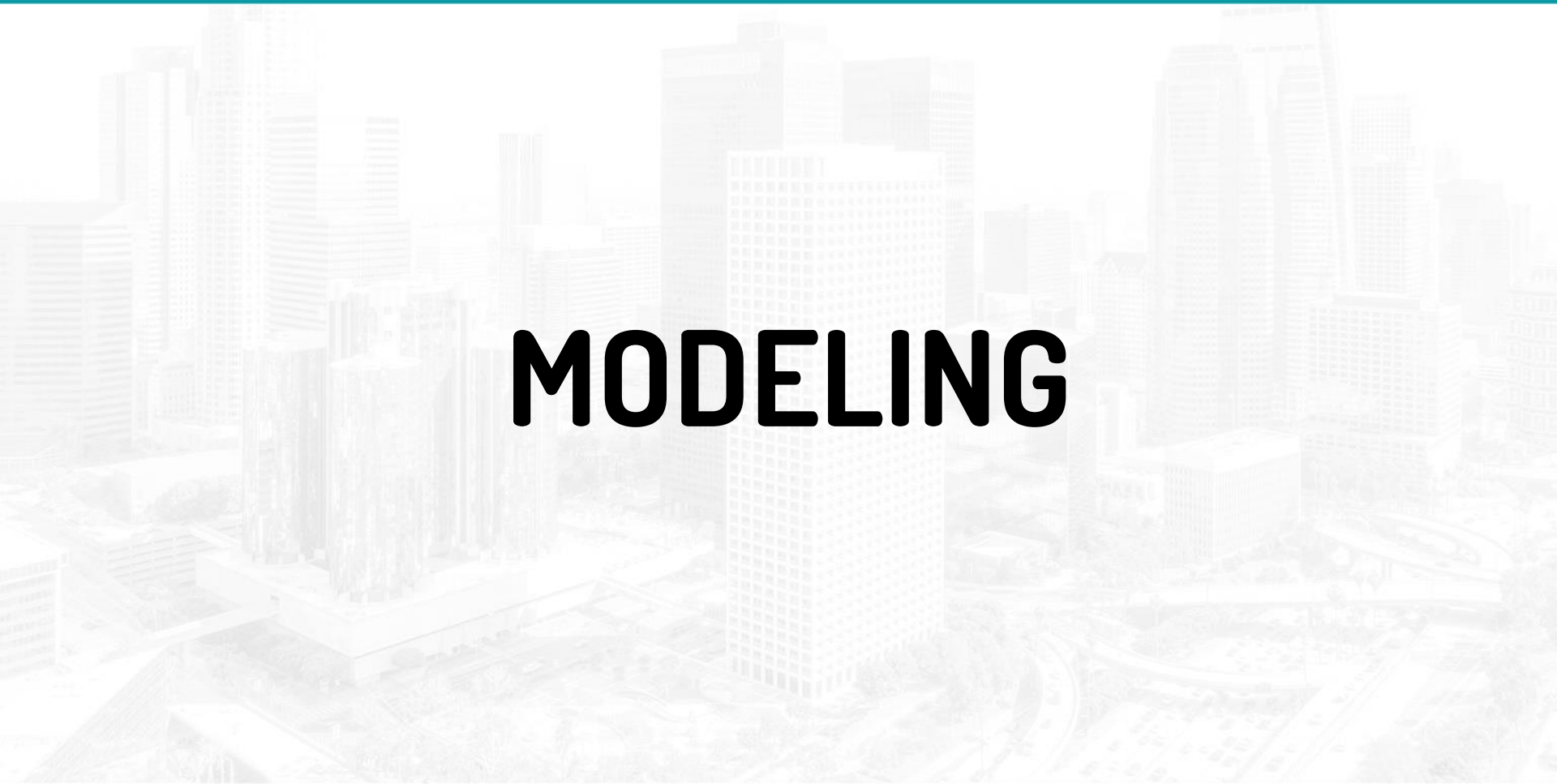
1. Consistency in Loans and Repayments:
The strong correlation between "loan_amnt" (loan amount), "funded_amnt" (funded amount), and "installment" (monthly installment) indicates that larger loans tend to have higher funded amounts provided to borrowers, along with larger monthly installments to repay the loan.

2. Linear Relationship:
The high correlation among these features signifies a robust linear relationship between the loan amount, funded amount, and monthly installment. As the loan amount increases, the funded amount also tends to increase, along with the corresponding monthly installment that needs to be paid.

3. Loan Approval Decision:
This correlation suggests that the company is likely to provide an appropriate amount of funds based on the requested loan amount by borrowers. Additionally, the monthly installment will also be adjusted based on the loan amount.

4. Importance of Payment Feasibility Evaluation:
In the context of credit risk, this correlation emphasizes the significance of evaluating the payment feasibility of borrowers seeking larger loans. Larger loans with higher monthly installments can pose a risk if borrowers are unable to consistently make the payments.

5. Risk Modeling:
This correlation must also be considered when building a credit risk prediction model. The interdependence between "loan_amnt," "funded_amnt," and "installment" can influence how the model predicts credit risk, and it should be carefully accounted for in feature selection and modeling.



MODELING

Step For Modeling

Result for Selection Model

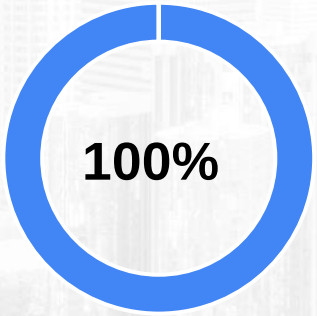
Rakamin Academy

Data Train : Data Test
80 : 20

Overfitting, too perfect

Gradient Boosting

Recall Score :

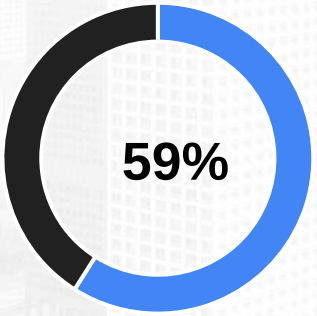


Evaluation Model:

- Accuracy : 100%
- Precision : 100%
- F1 Score : 100%
- ROC AUC : 100%

Logistic Regression

Recall Score :



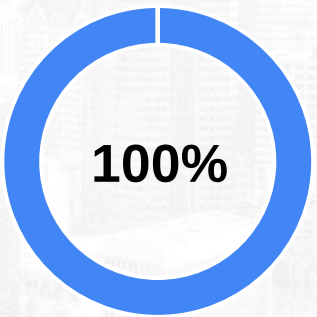
Evaluation Model:

- Accuracy : 95%
- Precision : 96%
- F1 Score : 73%
- ROC AUC : 92%

Overfitting, too perfect

Random Forest

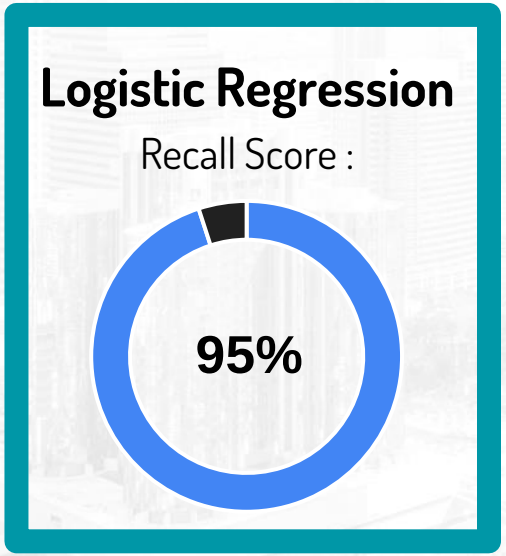
Recall Score :



Evaluation Model:

- Accuracy : 100%
- Precision : 100%
- F1 Score : 100%
- ROC AUC : 100%

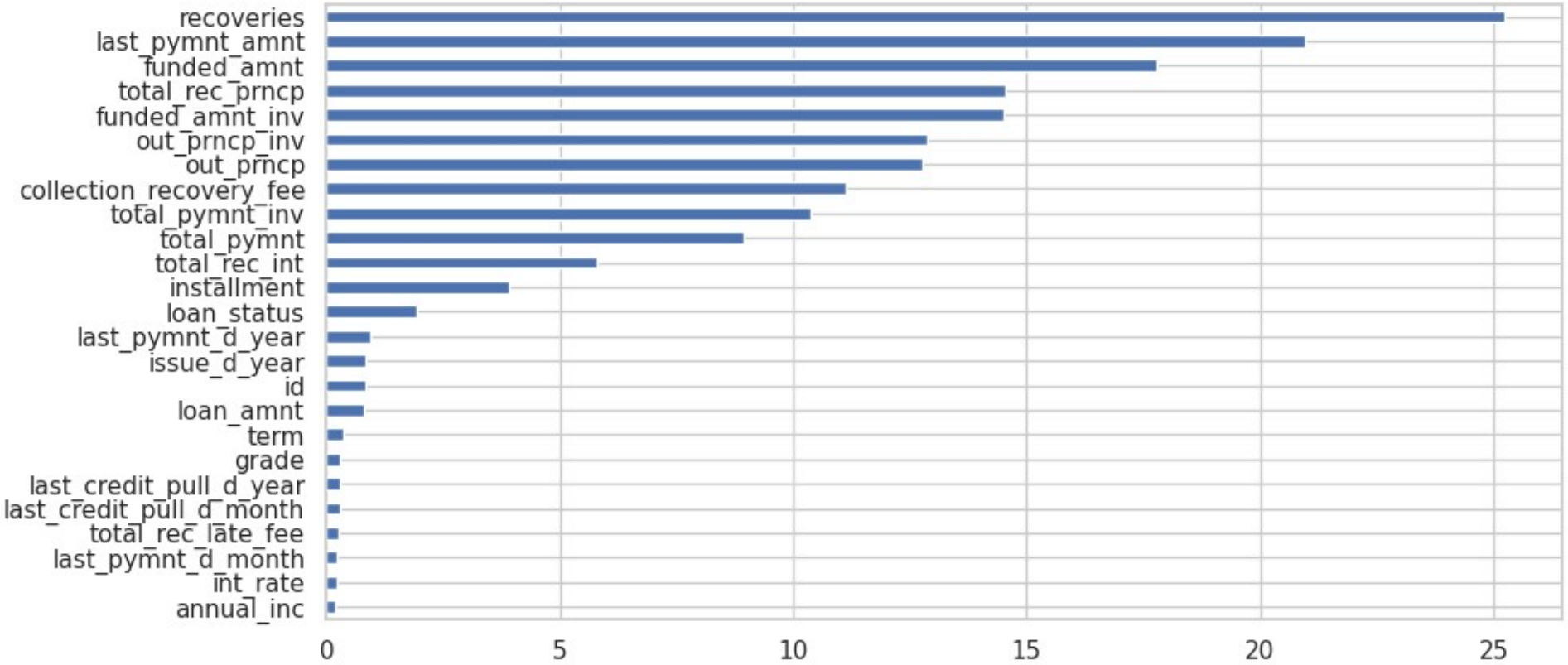
Hyperparameter Tuning



Best Parameter :

- C : 10
- Penalty : l2

Feature Importance



Feature Importance

Rakamin Academy

1. Recoveries:

This feature may have a strong influence on credit risk prediction. High values of this feature may indicate that customers with higher recovery rates are more likely to have a higher risk of default or failure to pay.

2. Last Payment Amount (last_pymnt_amnt):

This feature also holds significant influence on credit risk prediction. High values of this feature may suggest that customers who make higher last payment amounts have a lower likelihood of default.

3. Funded Amount (funded_amnt):

The amount of loan provided to customers also appears to have a significant impact on credit risk prediction. This could imply that the higher the loan amount provided, the higher the risk that customers may not be able to repay.

BUSINESS RECOMMENDATION

Business Recommendation

1. Enhanced Risk Management:
The company can further improve risk management by focusing more on borrowers with higher credit risks. This may involve increasing oversight and collection management to address potential payment issues that may arise.

2. Development of Specialized Products:
- Based on borrower behavior patterns, the company could consider developing specialized products or services targeted at borrowers with low-risk profiles.
- These products might include lower interest rates, longer loan terms, or more flexible payment options.

3. Refinement of Borrower Selection Process:

Considering the majority of loans are "Current" and "Fully Paid," the company can continue enhancing its borrower selection process by paying closer attention to factors impacting credit risk, such as payment history and income.

Business Recommendation

4. Financial Education Programs:
Based on the analysis of income and credit risk importance, the company could launch financial education programs to help borrowers improve their understanding of financial management, timely loan repayment, and risk minimization.

5. Further Model Development:

- With a predictive modeling evaluation yielding a recall of 95%, the company can further develop more accurate risk prediction models to identify borrowers with high-risk potential.
- These models can assist the company in taking preventive actions against potential payment delays or defaults.

6. Partnerships with Other Financial Institutions:

Based on the analysis of high-quality borrower profiles, the company can partner with other financial institutions to offer additional financial products, such as investments or savings.

Business Recommendation

Rakamin Academy

7. Dynamic Interest Rate Adjustments:

- Observing the relationship between interest rates and credit risk, the company could consider dynamic strategies for setting interest rates based on borrower risk profiles.
- This could help optimize revenue and reduce credit risk more efficiently.

8. Expanded Customer Monitoring:

Considering the EDA results regarding payment speed, the company can build a stronger customer monitoring system to identify payment patterns and potential issues early on.

9. Personalized Services:

Based on the evaluation of significant features, the company can consider personalized services and offers based on borrower characteristics to enhance customer satisfaction and business outcomes.

10. Development of Financial Education Apps:

Considering the potential low-credit risk customers, the company can develop financial education apps to help borrowers track and manage their payments, as well as provide relevant financial advice.

THANK YOU

Click icon for the code :