Home Loan Konapure Dataset

Navigating Risk

Insights from Machine Learning Models in Loan Approval for Canadian Banks

Presented by
Inge Angelia
Moromisato Renzo
Edward Rodney

Introduction

Owning a home symbolizes financial stability and achievement

The journey often involves a long-term partnership with financial institutitons through motgages

Home Loan Konapure Dataset

History

Historical turbulence in the housing market, exemplified by the 2008 bubble burst.

Lessons learned from the global financial crisis emphasize the dangers of reckless lending

Home Loan Konapure Dataset

Concerns in the Canadian housing market

Historical turbulence in the housing market, exemplified by the 2008 bubble burst.

Lessons learned from the global financial crisis emphasize the dangers of reckless lending

Home Loan Konapure Dataset

Objective

The imperative for financial institutions to manage risk effectively.

Advanced machine learning models complement traditional risk assessment methods.

Home Loan Konapure Dataset

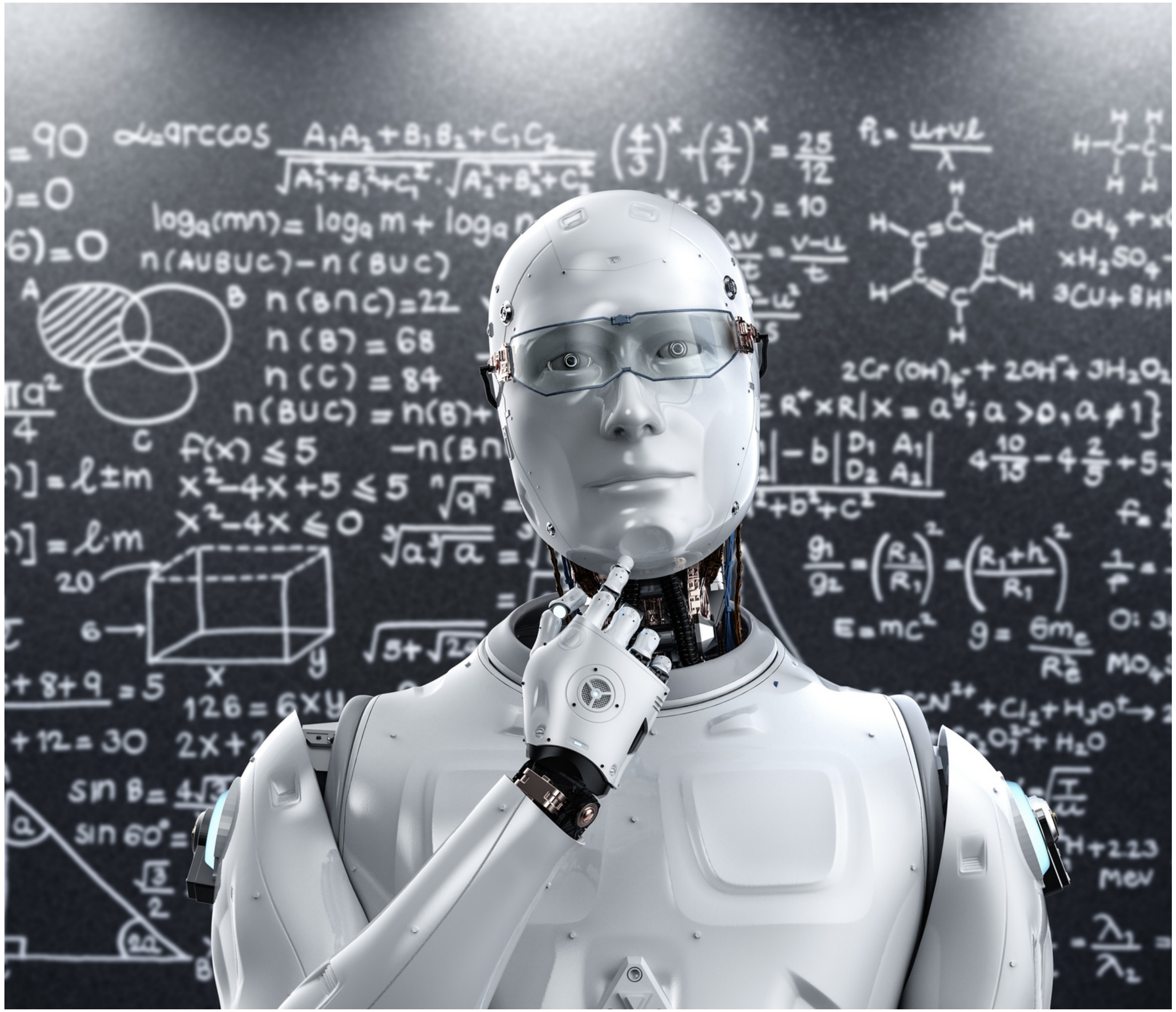
Proposal

Utilizing predictive capabilities of machine learning to identify and mitigate risk.

Augmenting traditional risk assessment methods with advanced algorithms.

Exploring the intersection of risk management and technological innovation in loan approval.

Leveraging machine learning to enhance creditworthiness assessment.



Literature Review

1. Canada's Housing Market Dynamics: A Distinctive Perspective

- Concerns raised by researchers regarding Canada's housing market, contrasting it with the 2008 U.S. housing bubble.
- Primary issues: demand-supply imbalance driven by rapid population growth outpacing housing construction.
- Role of low-interest rates, particularly during the COVID-19 pandemic, in fueling demand.
- Rising interest rates now pose challenges for Canadians, potentially leading to increased mortgage defaults.
- Equifax report highlights alarming mortgage delinquency rates, particularly among younger demographics.

Literature Review

2. Economic Implications of a Bursting Housing Bubble

- Potential for severe economic damage from a bursting housing bubble, with the capacity to induce a national recession.
- Canada's vulnerability emphasized by its high ranking in global real estate indices and housing-risk indicators.
- Numerous studies underline the risk of recession following a housing bubble burst.
- Lessons from the 2008 crisis underscore the extensive repercussions on financial institutions, the broader economy, and individual homeowners.

Literature Review

3. The Role of Machine Learning in Banking

- Machine Learning emerges as a transformative tool in banking, capable of handling complex tasks beyond human capabilities.
- Seamless integration of Machine Learning into banking operations facilitated by real-time data streams.
- Success stories like Danske Bank's Fraud Detection and Postbank's loan administration automation highlight the effectiveness of Machine Learning.
- Significant reductions in false positives, increased accuracy rates, and enhanced operational efficiency demonstrate the potential of Machine Learning in the banking sector.

Eaglewood Realty

Konapure (2023) Dream Housing Finance Company

Dataset

The Konapure Dream Housing Finance Company dataset consists of a train and test set, each with 13 columns. The columns and encoded values are as follow:

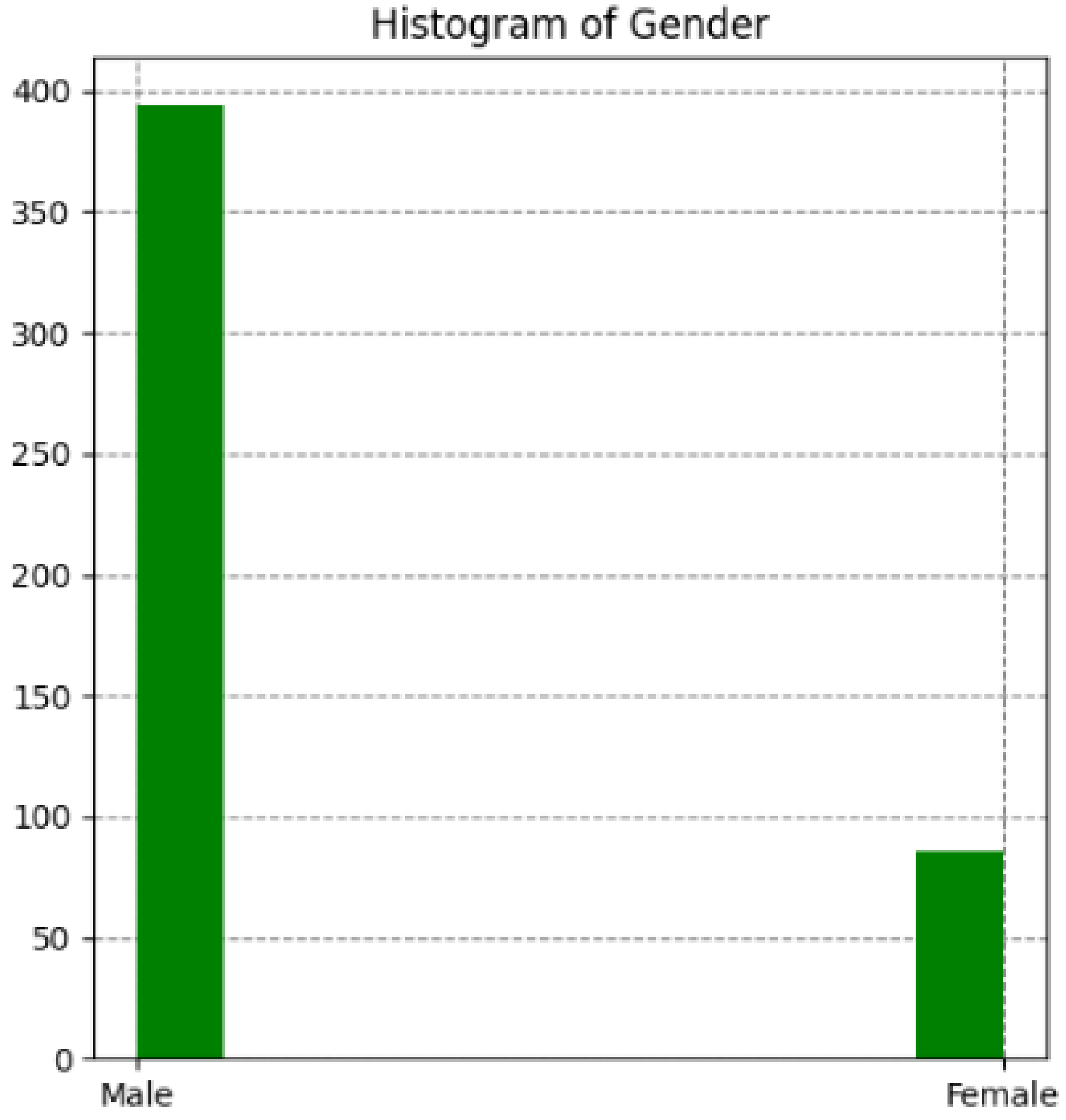
- Loan_ID
- Gender - Female: 0; Male: 1
- Married – No: 0; Yes: 1
- Dependents – 0: 0 ; 1: 1; 2: 2; 3+:3
- Education – Graduate: 0; Not Graduate: 1
- Self_Employed – No: 0; Yes: 1
- ApplicantIncome
- CoapplicantIncome
- LoanAmount
- Loan_Amount_Term
- Credit_History: No: 0; Yes: 1
- Property_Area: Rural: 0; Semiurban: 1; Urban: 2
- Loan_Status (only train set)

Eaglewood Realty

Graphs and Diagrams

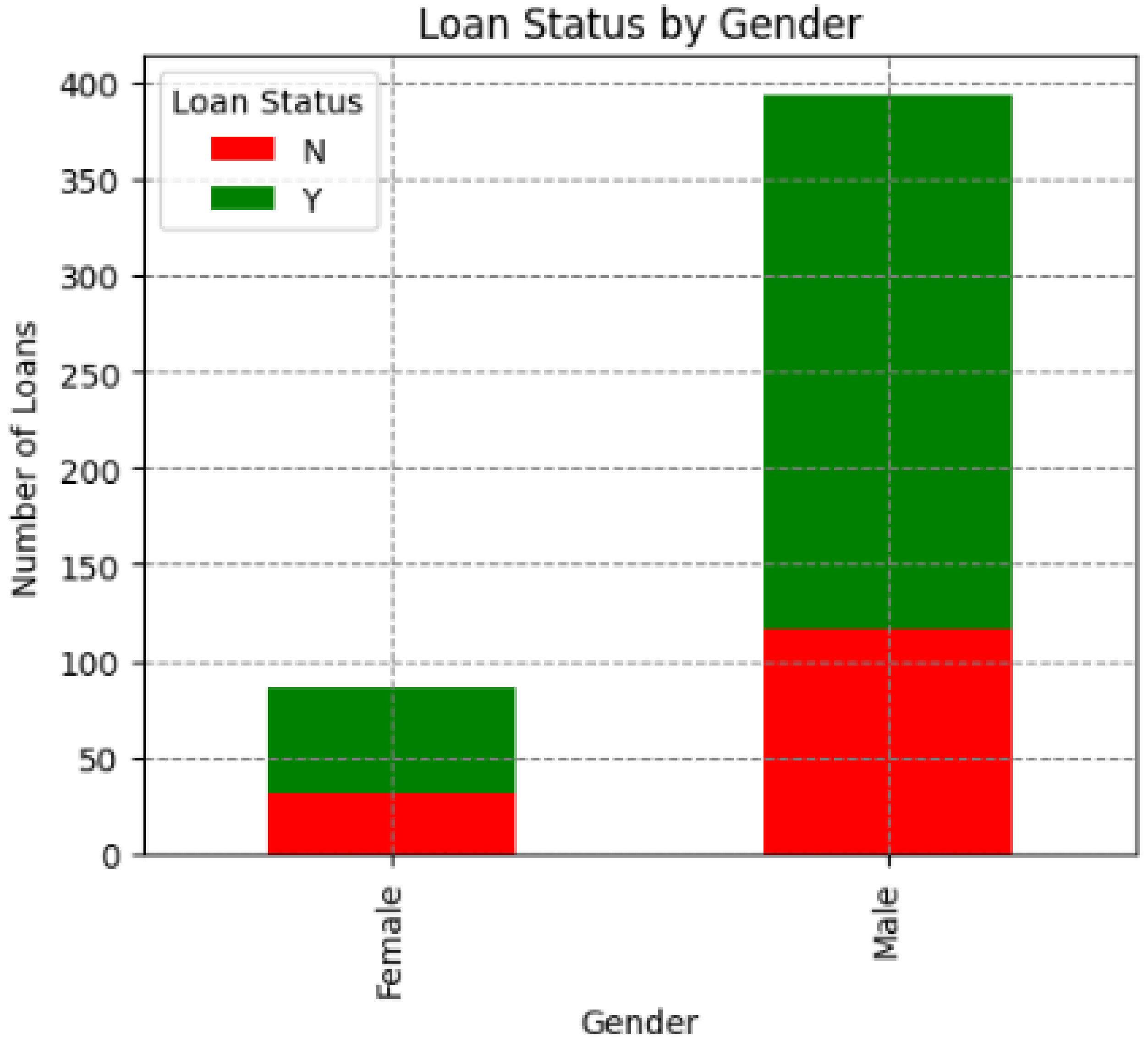
Gender

Training Set



Histogram

This graphical representation highlights a noticeable gender imbalance in home loan applications, with significantly more male applicants than female applicants.

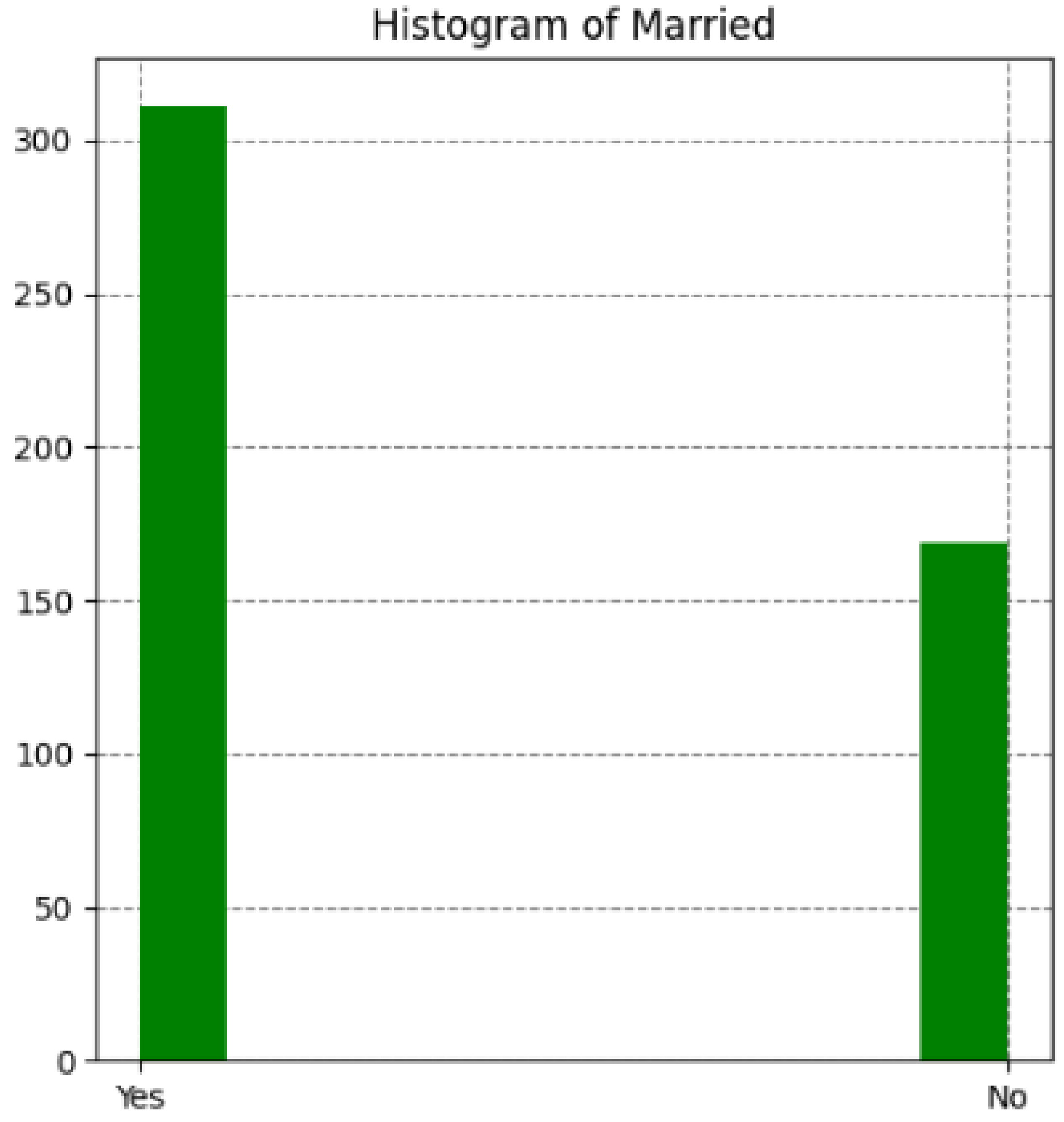


Loan Status Stacked Bar Chart

The approval rate for male applicants is higher than the approval rate for female applicants. 70.56% of male applicants are approved for home loans compared to 62.79% of female applicants.

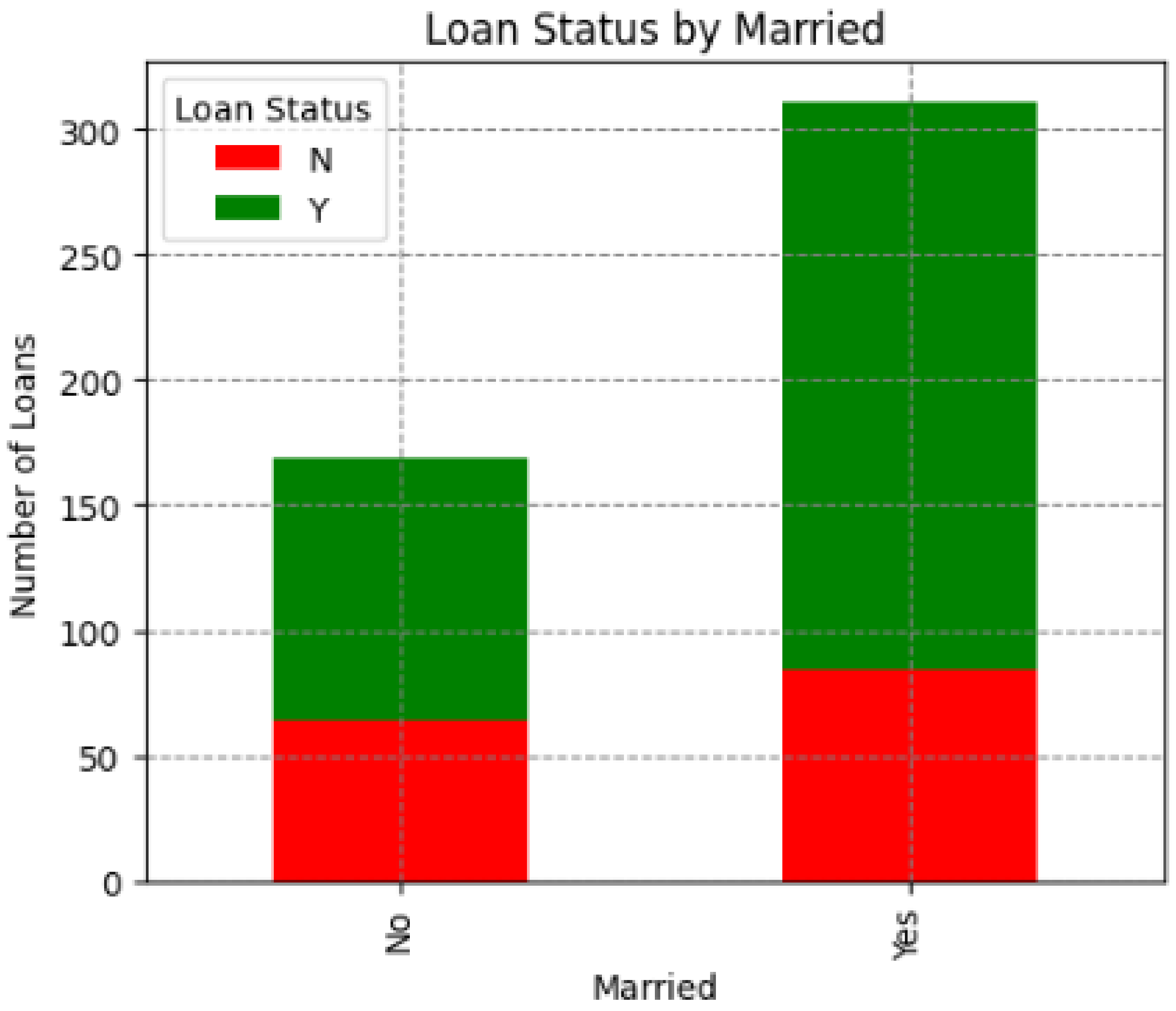
Marital Status

Training Set



Histogram

The histogram clearly shows a significantly higher number of married applicants compared to unmarried ones, with approximately 320 "Yes" responses and 170 "No" responses.

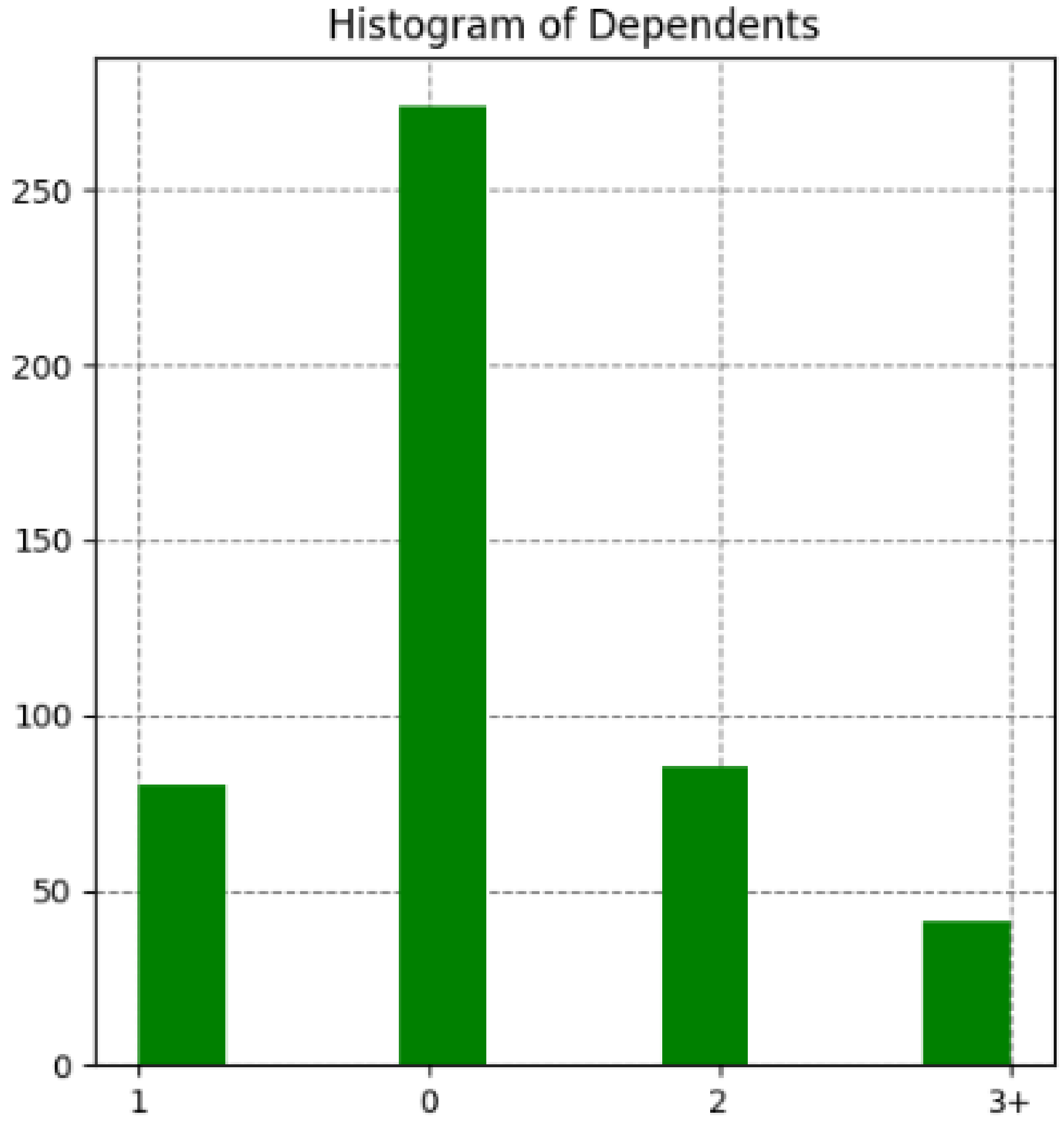


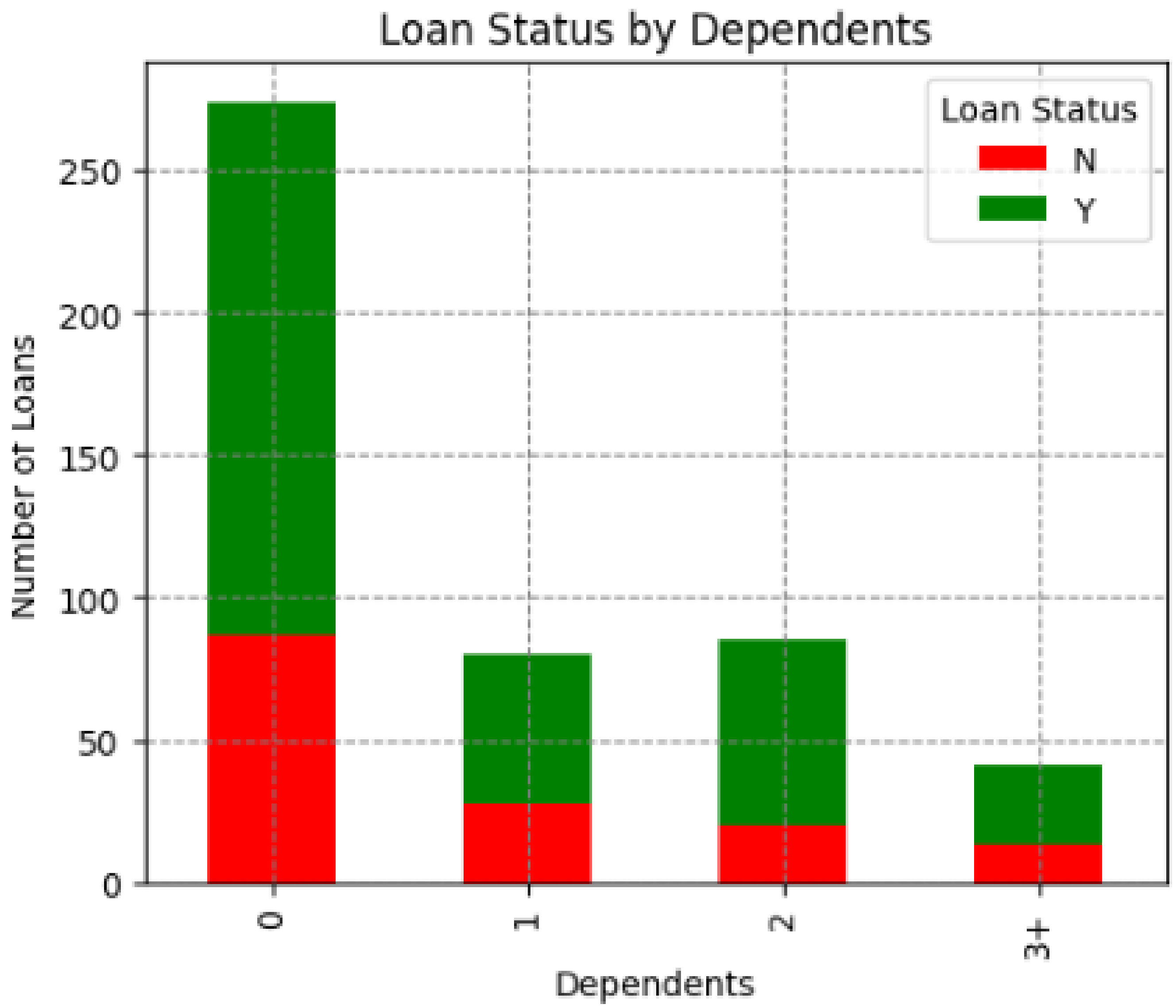
Loan Status Stacked Bar Chart

Married applicants have a higher approval rate than applicants who are not married. 72.99% of married applicants are approved for home loans compared to 62.13% of unmarried applicants.

Number of Dependents

Training Set





Histogram

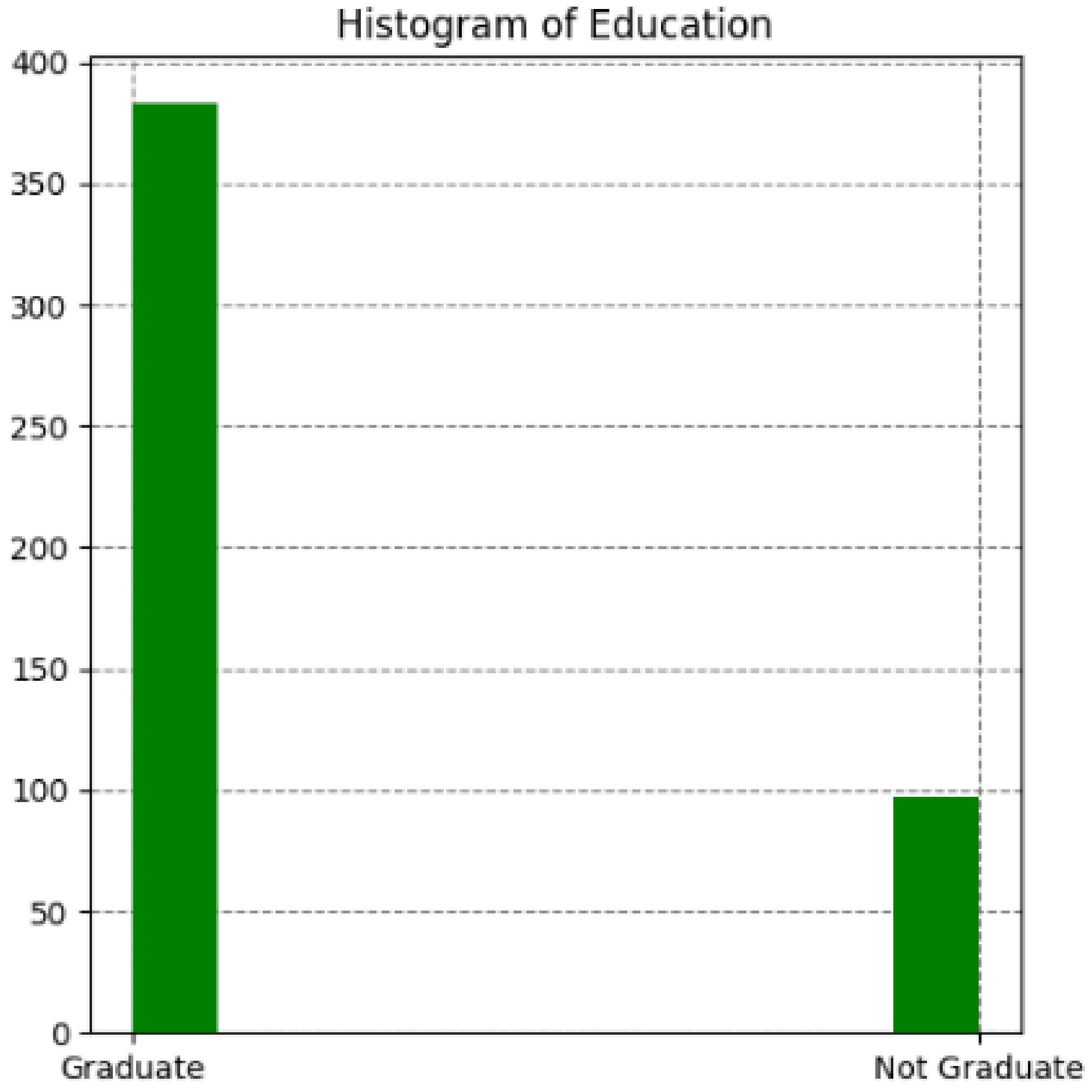
The tallest green bar corresponds to individuals or entities with no dependents. This indicates a higher frequency of such cases within the dataset.

Loan Status Stacked Bar Chart

Applicants with 2 dependents had the highest approval rate of 76.47% followed by an approval rate of 68.29% for applicants with 3 or more dependents, 68.25% for applicants with no dependents, and 65% for applicants with 1 dependent.

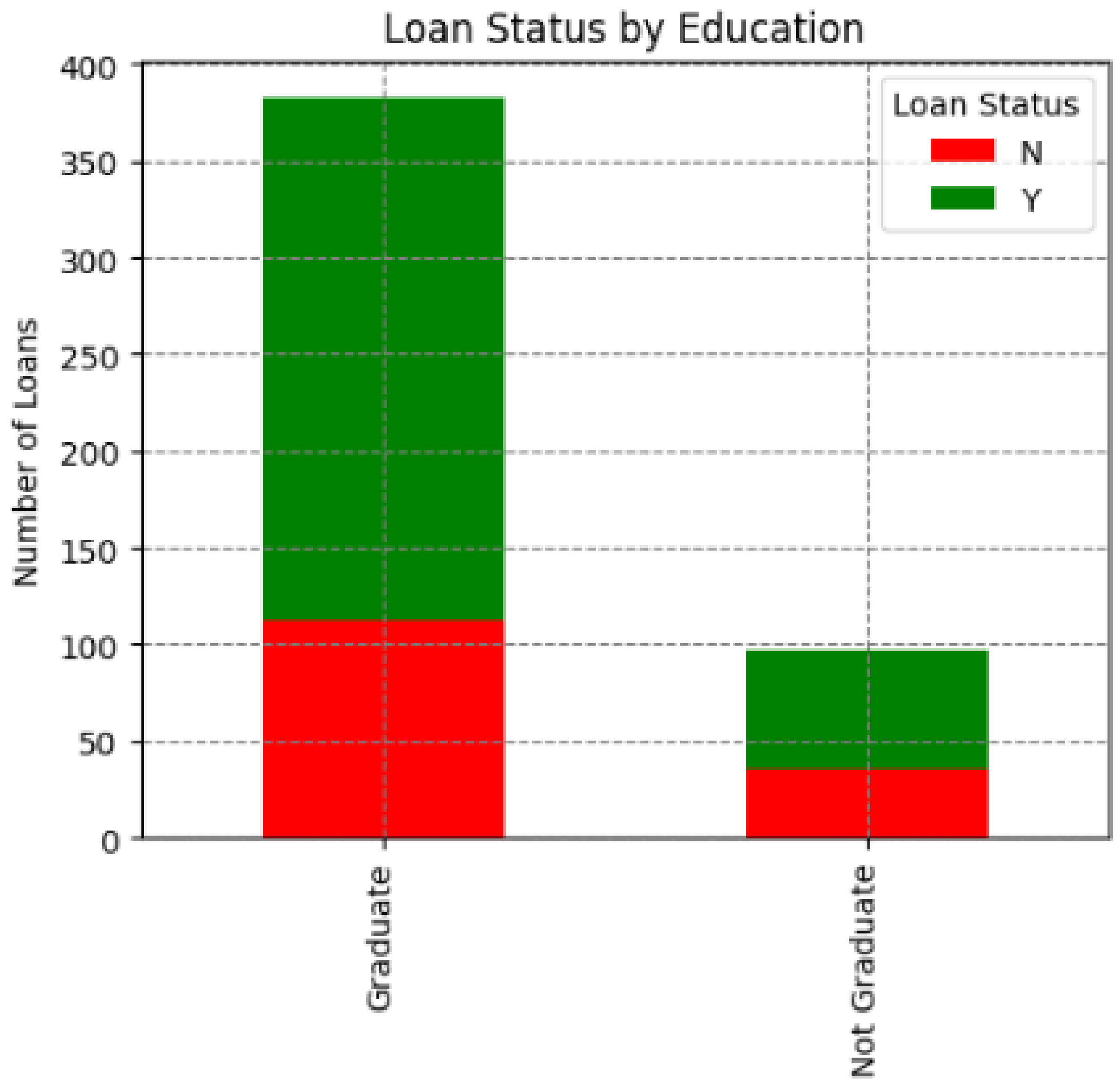
Education

Training Set



Histogram

Graduate-level applicants reach a count of approximately 380 while applicants that are not graduates reach a count of approximately 90 .

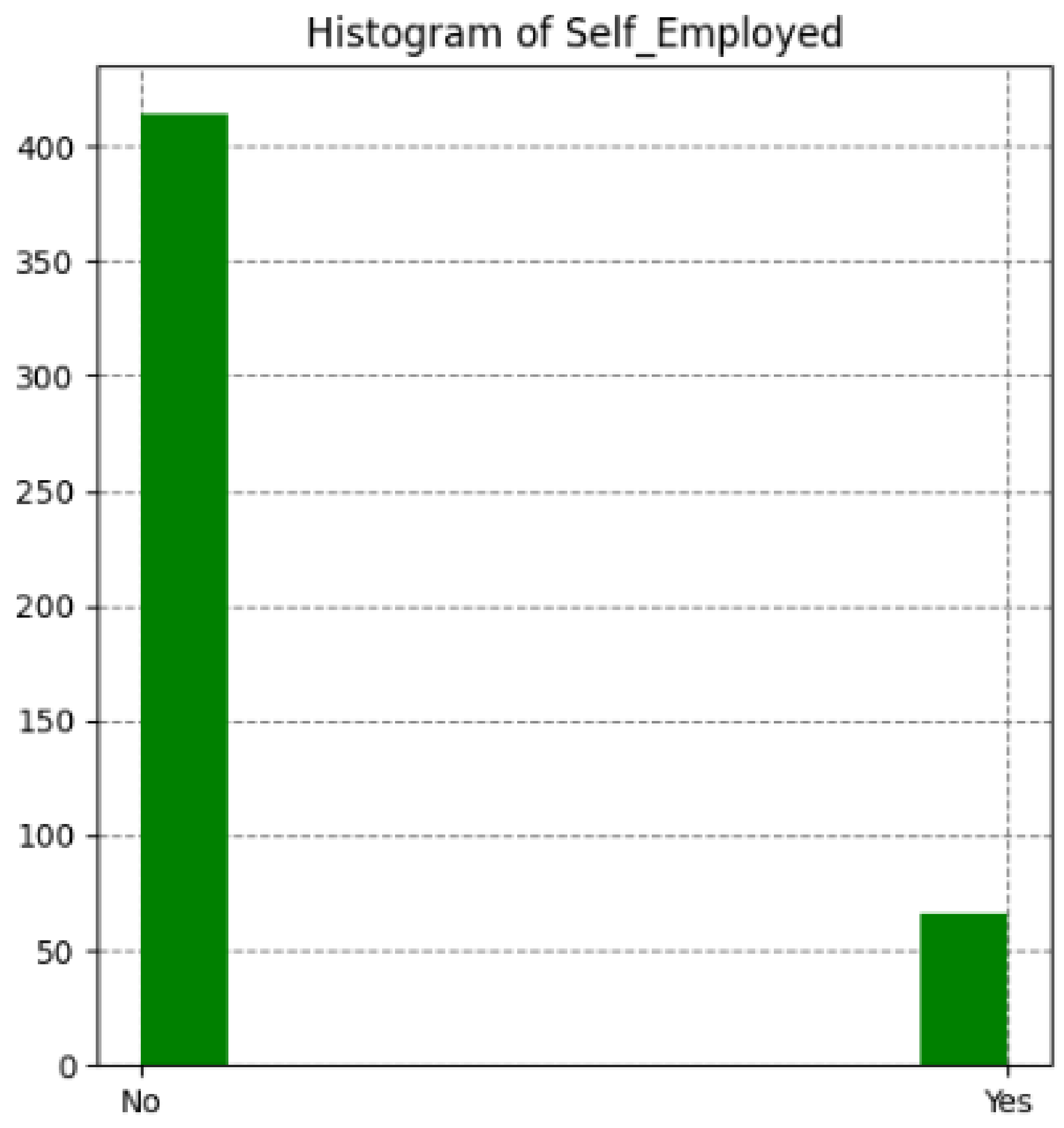


Loan Status Stacked Bar Chart

The approval rate for graduates is higher, standing at 70.76%, in contrast to the 62.89% approval rate for non-graduates

Self-Employment Status

Training Set



Histogram

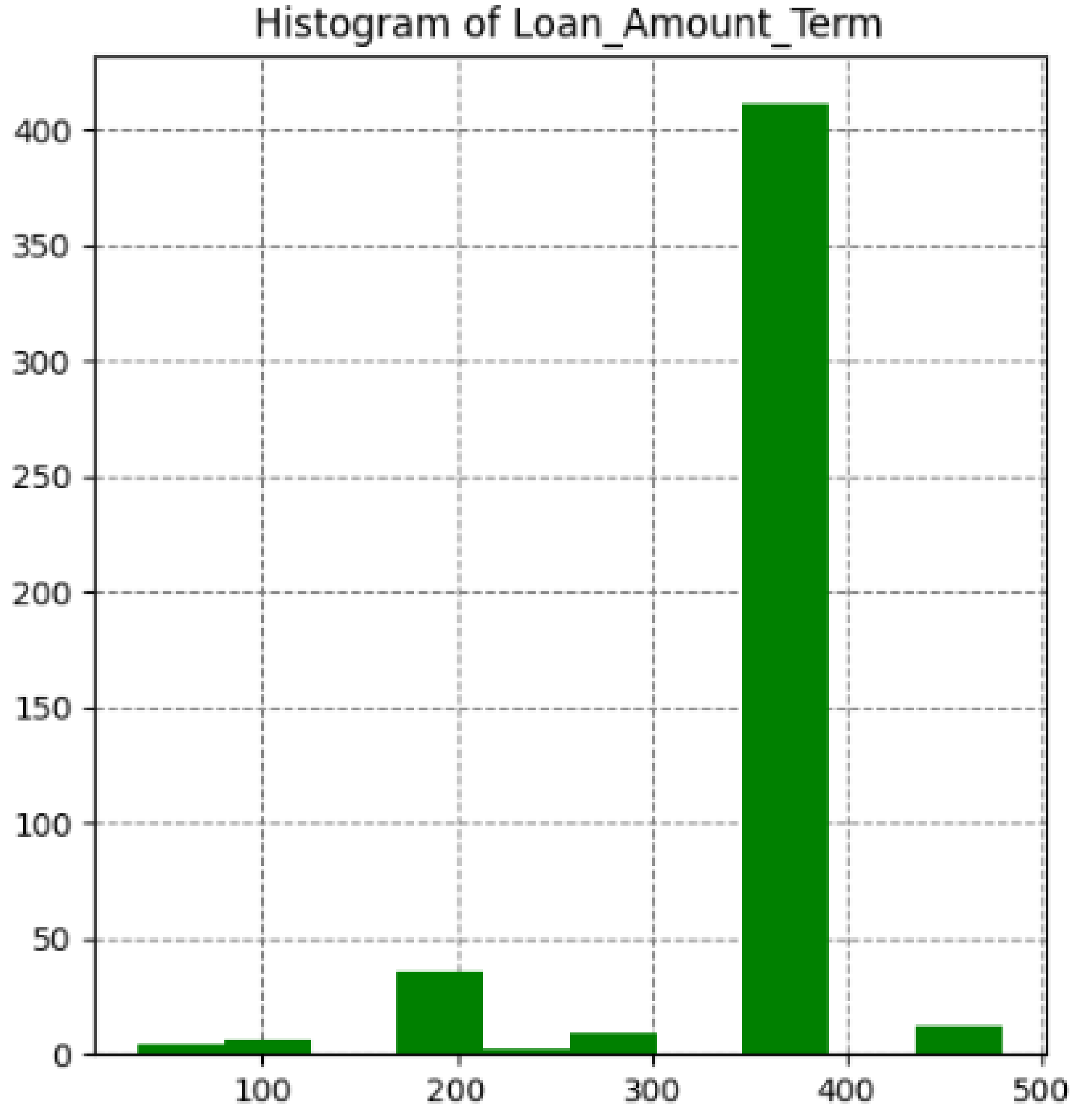
The number of applicants who are not self-employed is notably higher than those that are

Loan Status Stacked Bar Chart

the approval rate for home loans is higher among non-self-employed applicants, standing at 69.81%, compared to 65.35% for self-employed applicants.

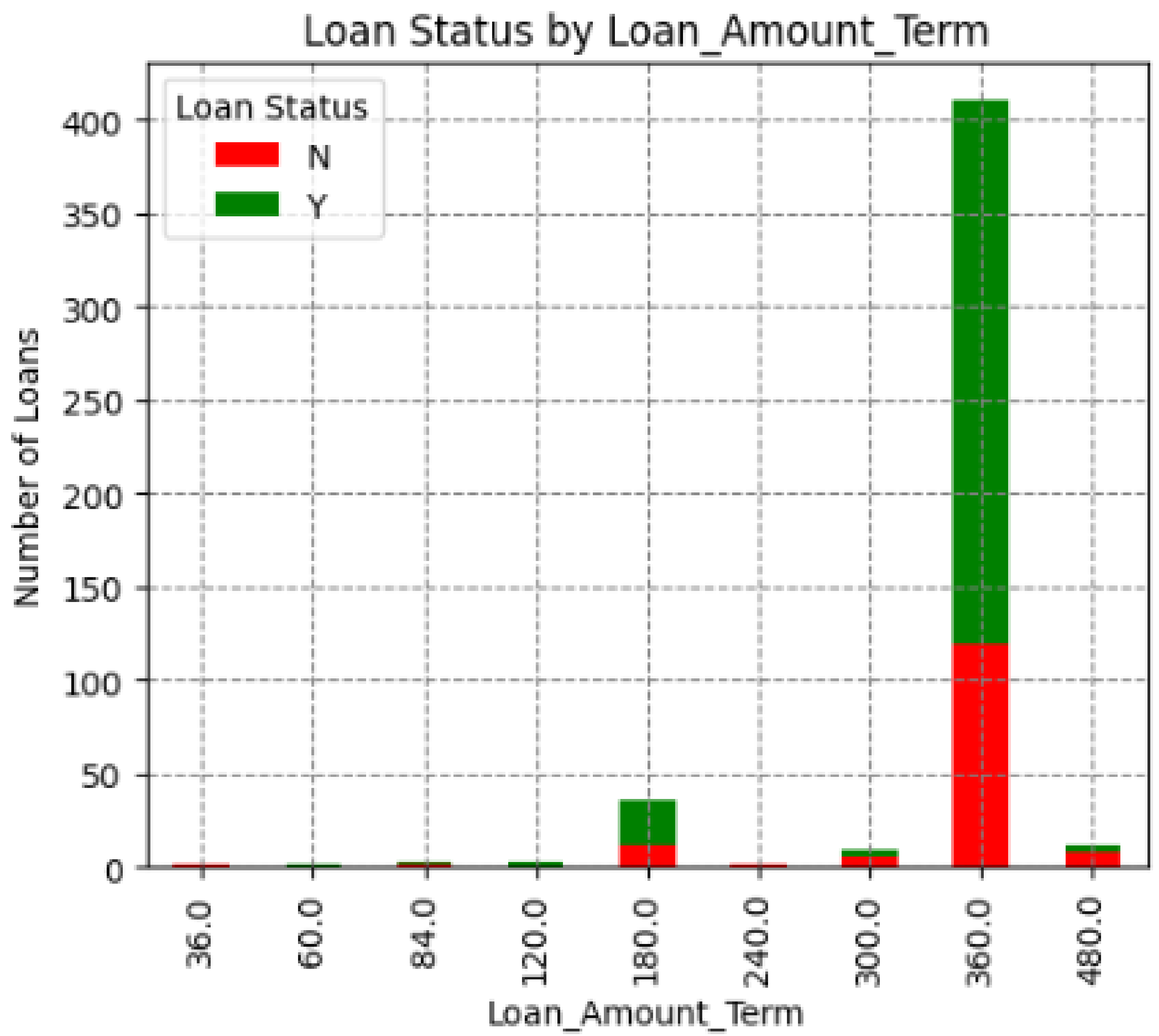
Loan Amount Term

Training Set



Histogram

Loan terms of 400 carry the largest number of applicants compared to loan terms of 100, 200, 300, and 500.

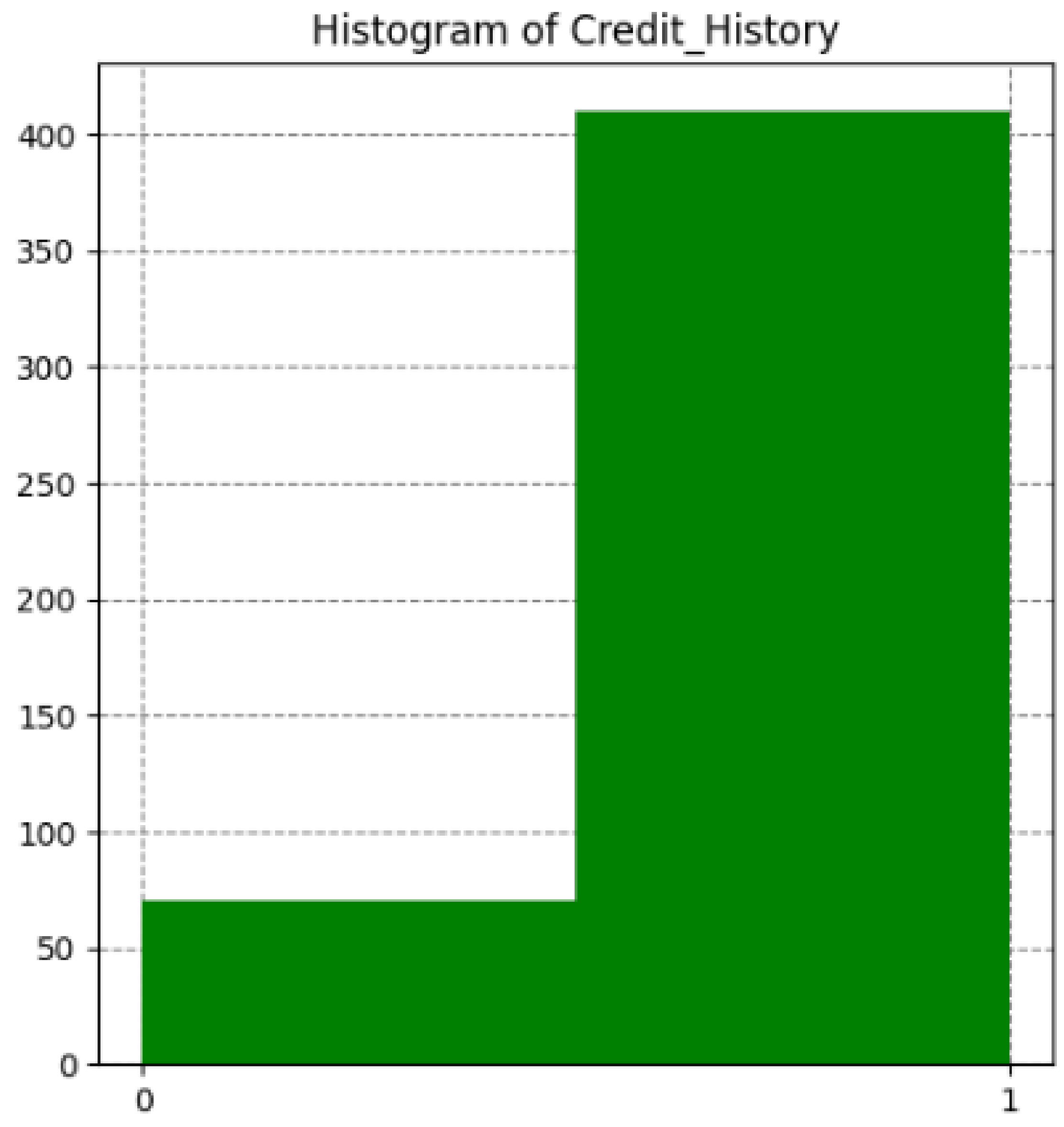


Loan Status Stacked Bar Chart

The loan approval rate is the greatest for terms of 360 at 71.05%, while the loan approval rate for terms of 180 is at 66.67%.

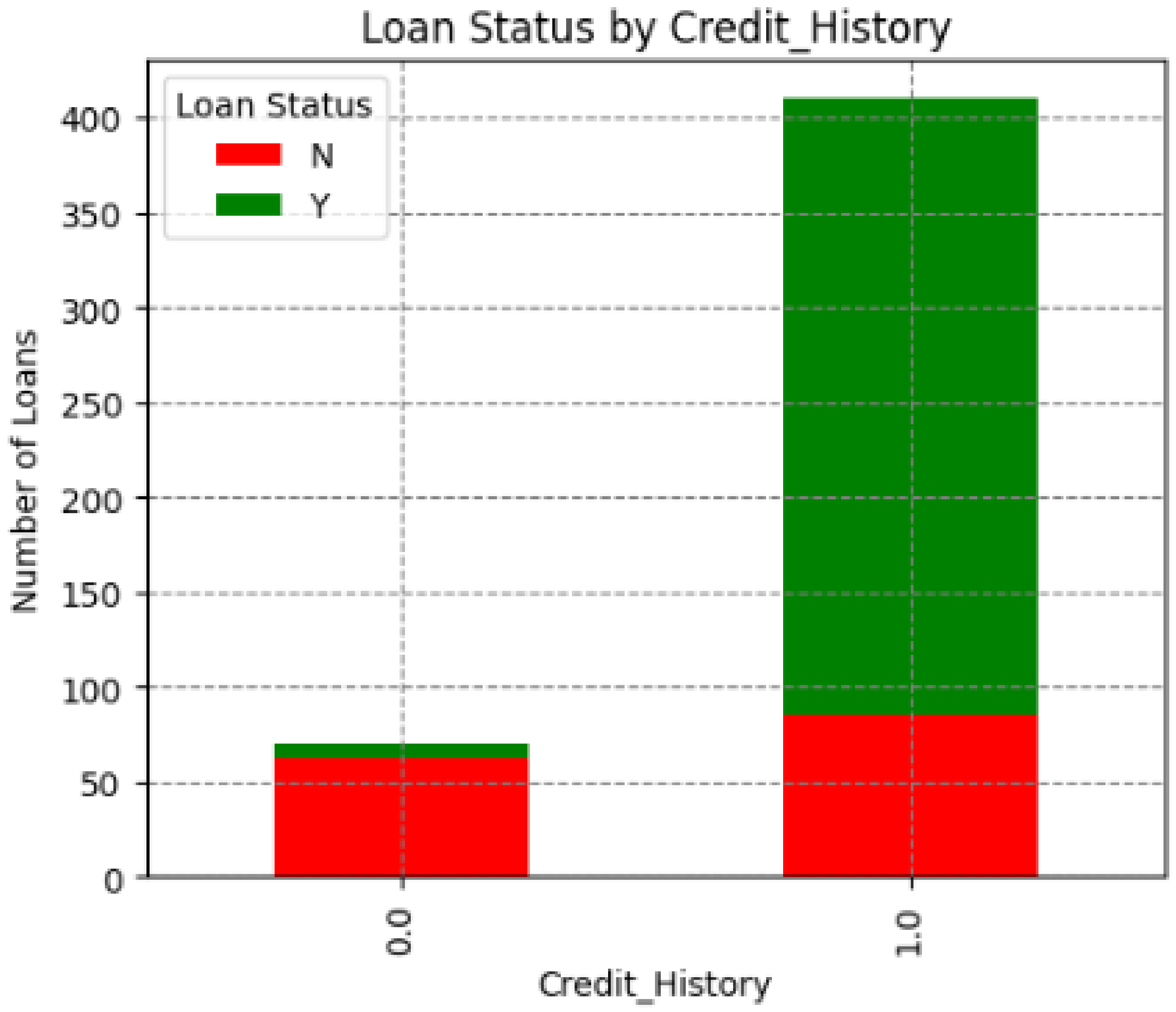
Credit History

Training Set



Histogram

applicants with a credit history are substantially more numerous than those with out a credit history.

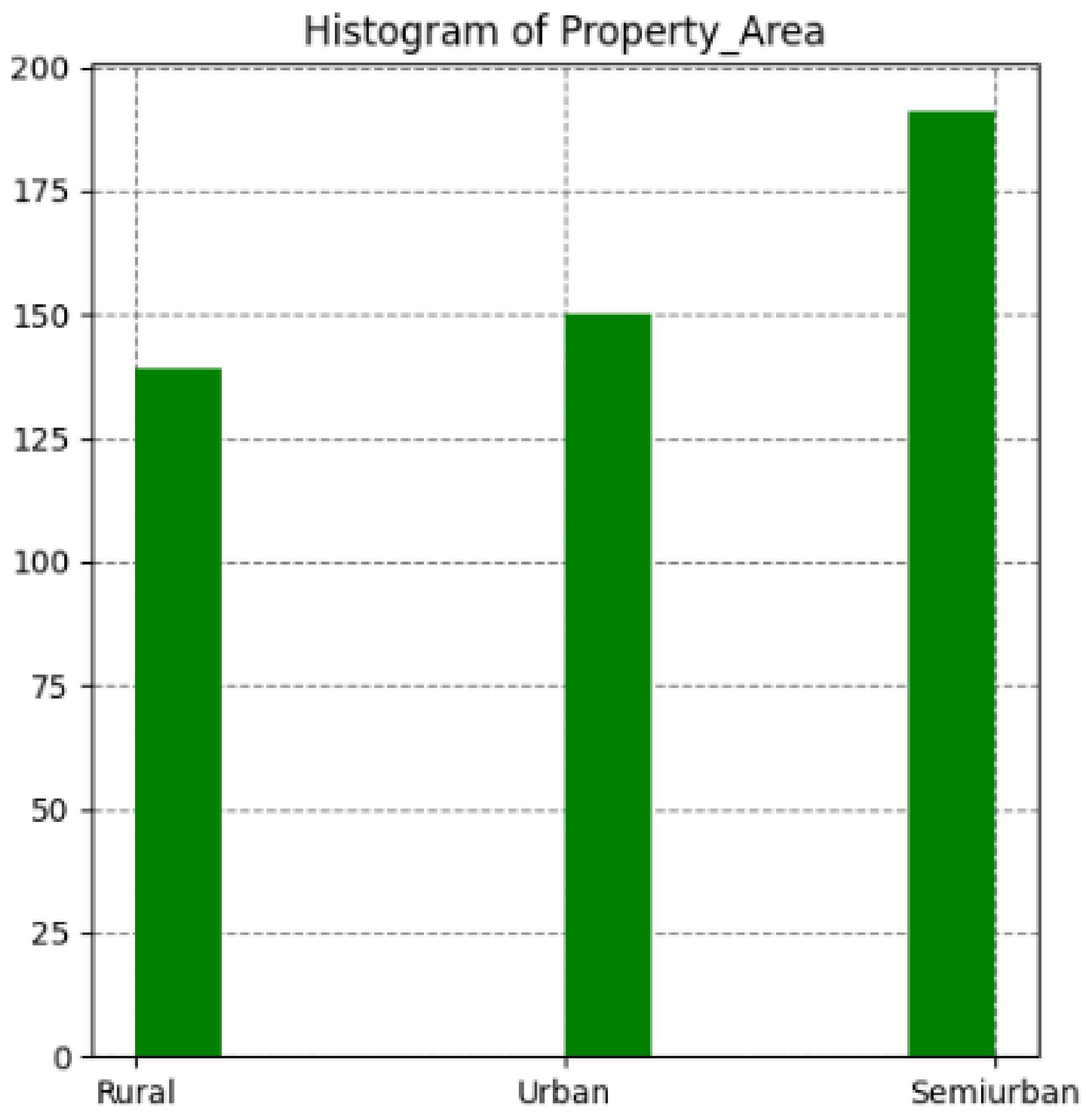


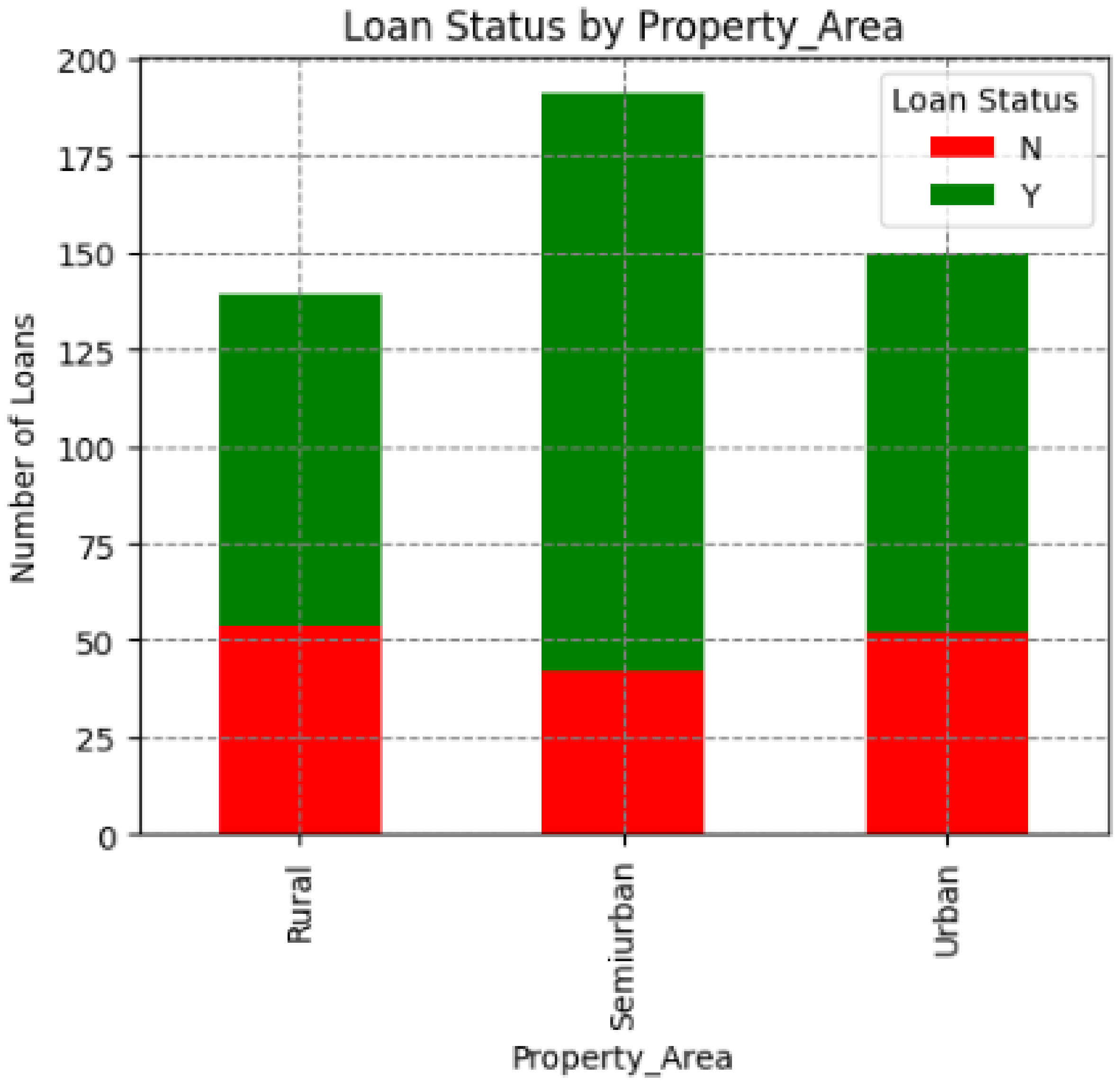
Loan Status Stacked Bar Chart

The loan approval rate is at 79.27% for applications with history of credit while the loan approval rate is only 10% for applications with no history of credit.

Property Area

Training Set





Histogram

Most of the applications submitted are for semiurban properties with a count of approximately 190. This is followed by applications for properties located in urban areas and rural areas.

Loan Status Stacked Bar Chart

The approval rate follows the same pattern as the number of applications submitted with semiurban properties having a home loan approval rate of 78.01%, urban properties having an approval rate of 65.33%, and rural properties having an approval rate of 61.15%.

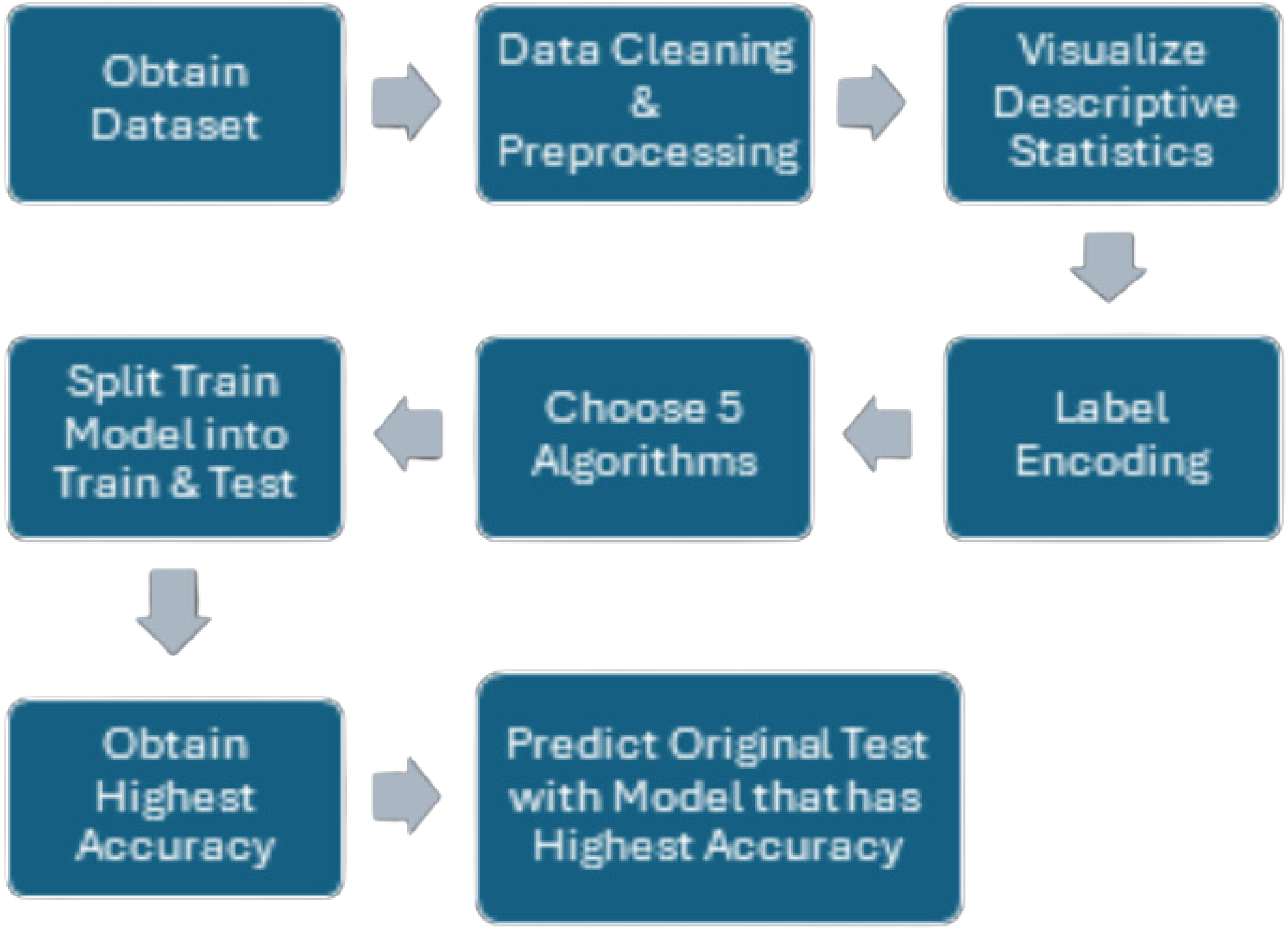
Eaglewood Realty

Methodology

Data Cleaning

The train set contained 614 rows, reduced to 480 rows after cleaning. The test set had 367 initial rows, reduced to 289 rows after cleaning.

Data Processing for Modelling



Algorithms Used

1

Classification and Regression Trees (CART)

CART operates akin to a decision tree, utilizing GINI impurity to determine optimal split points, thereby enhancing predictive accuracy by segmenting data into homogeneous groups.

2

Random Forest Classification

Random Forest, akin to CART, employs decision trees making splits at optimal points, yet it differentiates by aggregating multiple trees trained on varied dataset subsets, using random feature subsets to combat overfitting and enhance diversity.

3

K-Nearest Neighbour (KNN)

KNN, a data-driven model, computes Euclidean distances between predictor variables of test and training data, identifying nearest neighbors for classification based on the majority class among the nearest K neighbors.

4

Naïve Bayes

Naïve Bayes, a probabilistic classifier, assesses each predictor independently, assigning probabilities to classes based on individual feature presence.

5

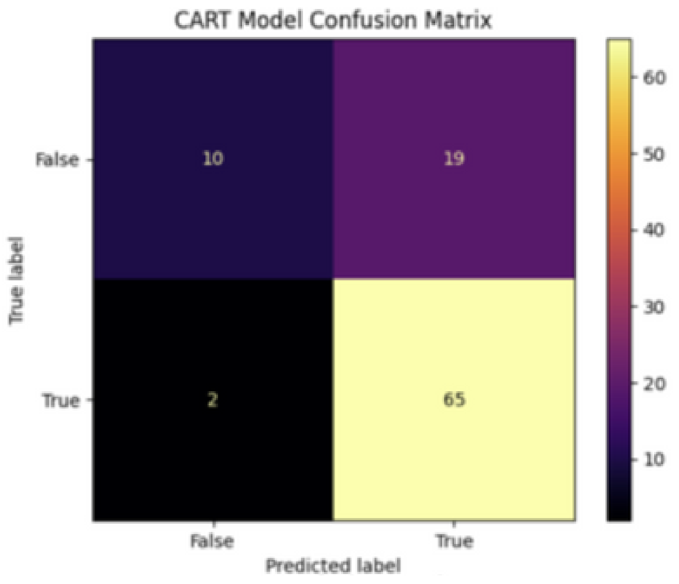
Logistic Regression

Logistic Regression, akin to linear regression, predicts binary outcomes (0 or 1), applicable for both explanatory and predictive tasks.

Algorithm Confusion Matrix

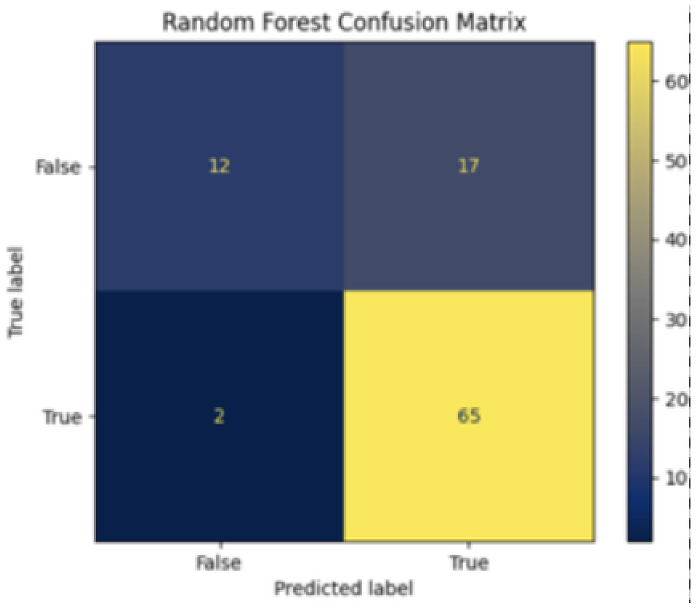
1

Classification and Regression Trees (CART)



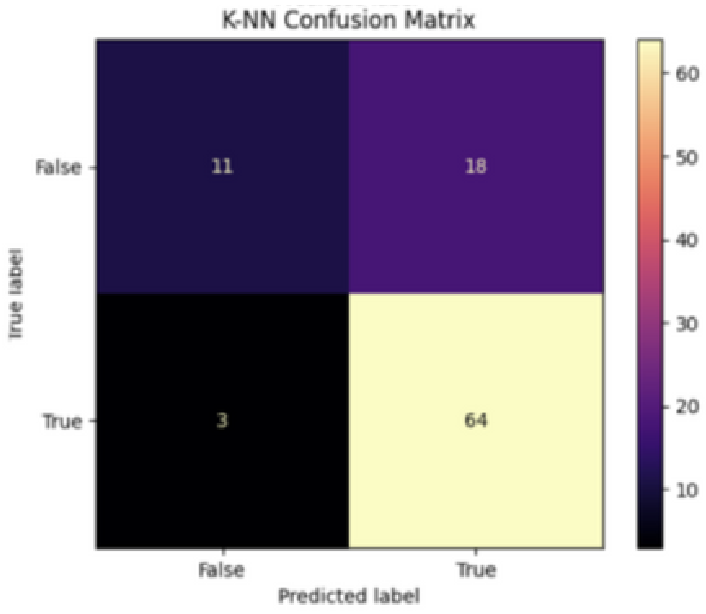
2

Random Forest Classification



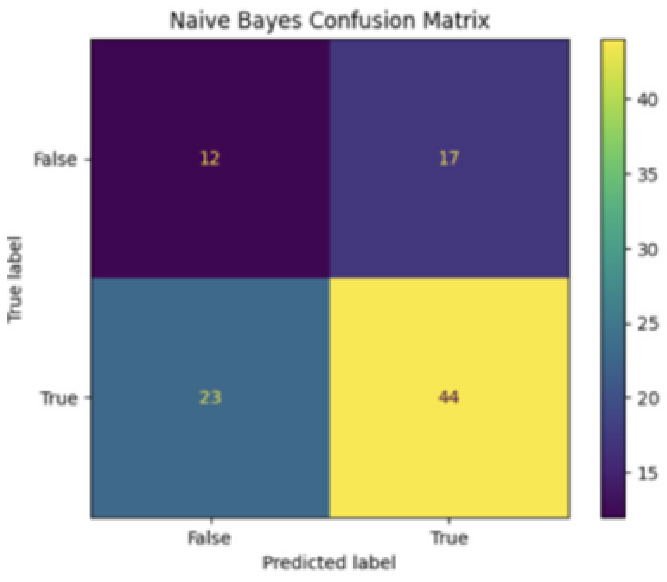
3

K-Nearest Neighbour (KNN)



4

Naïve Bayes



5

Logistic Regression

Top Performing Model

```
Logistic Regression Accuracy: 0.8021
Multinomial Naive Bayes Accuracy: 0.5833
CART Accuracy: 0.7812
Random Forest Accuracy: 0.8021
KNN Accuracy: 0.7917
```

Predicted Values on Test Set

```
['Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y'
 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y'
 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'N' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'N'
 'N' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'N' 'Y' 'N' 'Y' 'Y' 'Y'
 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y'
 'Y' 'Y' 'Y' 'N' 'N' 'N' 'Y' 'Y' 'N' 'N' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y'
 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y'
 'Y' 'N' 'Y' 'N' 'Y' 'Y' 'Y' 'N' 'N' 'Y' 'N' 'Y' 'N' 'Y' 'N' 'N' 'Y' 'Y'
 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'N' 'Y' 'Y' 'N' 'Y' 'N' 'Y' 'Y' 'Y'
 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y'
 'N' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'N' 'N' 'Y' 'Y' 'N' 'Y' 'N' 'Y'
 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'N'
 'Y' 'Y' 'Y' 'Y' 'N' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y'
 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'N' 'N' 'Y' 'Y' 'Y'
 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N' 'Y' 'Y' 'N' 'N'
 'Y' 'Y' 'N' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'Y' 'N'
 'Y']
```

Test Set with Predicted Values

ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Predicted_Loan_Status
5720	0	110.0	360.0	1.0	2	Y
3076	1500	126.0	360.0	1.0	2	Y
5000	1800	208.0	360.0	1.0	2	Y
3276	0	78.0	360.0	1.0	2	Y
2165	3422	152.0	360.0	1.0	2	Y
...	...	...	...	...	...	...
2269	2167	99.0	360.0	1.0	1	Y
4009	1777	113.0	360.0	1.0	2	Y
4158	709	115.0	360.0	1.0	2	Y
5000	2393	158.0	360.0	1.0	0	N
9200	0	98.0	180.0	1.0	0	Y

Eaglewood Realty

Conclusion

Conclusion

Machine learning shows promise in automating loan approval processes, enhancing accuracy, and mitigating risks.

Achieving an 80% accuracy rate in our trained model demonstrates potential benefits.

Limitations include a short timeframe and a relatively small dataset, impacting generalizability.

Variations in predictor variables across banks may further affect applicability.

Despite limitations, insights from this study offer valuable assistance to Canadian banks, enabling more precise and efficient loan approval decisions.

Eaglewood Realty

References

References

[1] S. Rosa, "Canadian mortgage-holders increasingly missed payment in Q4, Equifax says", CTV News, 2024. [Online]. Available: https://www.ctvnews.ca/business/canadian-mortgage-holders-increasingly-missed-payments-in-q4-equifax-says-1.6794802

[2] Mortgage Sandbox, "Is the canadian real estate market a bubble? Here are the risks to consider", 2024. [Online]. Available: https://www.mortgagesandbox.com/risk-in-the-canadian-real-estate-market

[3] B. LaCerda, A. Singh, S. Mintah, and G. Pinel, "Canada housing market outlook: more struggles ahead", Moody's Analytics., 2023. [Online]. Available: https://www.moodysanalytics.com/whitepapers/pa/2023/rps-ma-canada-housing-market-outlook-more-struggles-ahead

[4] I. Poshnjari, "A housing bubble burst would be worse in Canada than U.S.: Rosenberg", BNN Bloomberg, 2022. [Online]. Available: https://www.bnnbloomberg.ca/a-housing-bubble-burst-would-be-worse-in-canada-than-u-s-rosenberg-1.1841896

[5] G. Suhanic, "Posthaste: the coming recession will be a tale of housing versus commodities", Financial Post, 2024. [Online]. Available: https://financialpost.com/news/canada-recession-about-housing-versus-commodities

[6] J. Weinberg, "The great recession and its aftermath", Federal Reserve History, 2013. [Online]. Available: https://www.federalreservehistory.org/essays/great-recession-and-its-aftermath#:~:text=Effects%20on%20the%20Broader%20Economy,-The%20housing%20sector&text=The%20decline%20in%20overall%20economic,recession%20since%20World%20War%20II

[7] R. Merle, "A guide to the financial crisis – 10 years later", The Washington Post, 2018. [Online]. Available: https://www.washingtonpost.com/business/economy/a-guide-to-the-financial-crisis--10-years-later/2018/09/10/114b76ba-af10-11e8-a20b-5f4f84429666_story.html

[8] Projectpro, "15 projects on machine learning applications in finance", 2024. [Online]. Available: https://www.projectpro.io/article/projects-on-machine-learning-applications-in-finance/510

[9] R. Konapure, "Home loan approval", Kaggle, n.d. [Online].Available: https://www.kaggle.com/datasets/rishikeshkonapure/ho me-loan-approval