

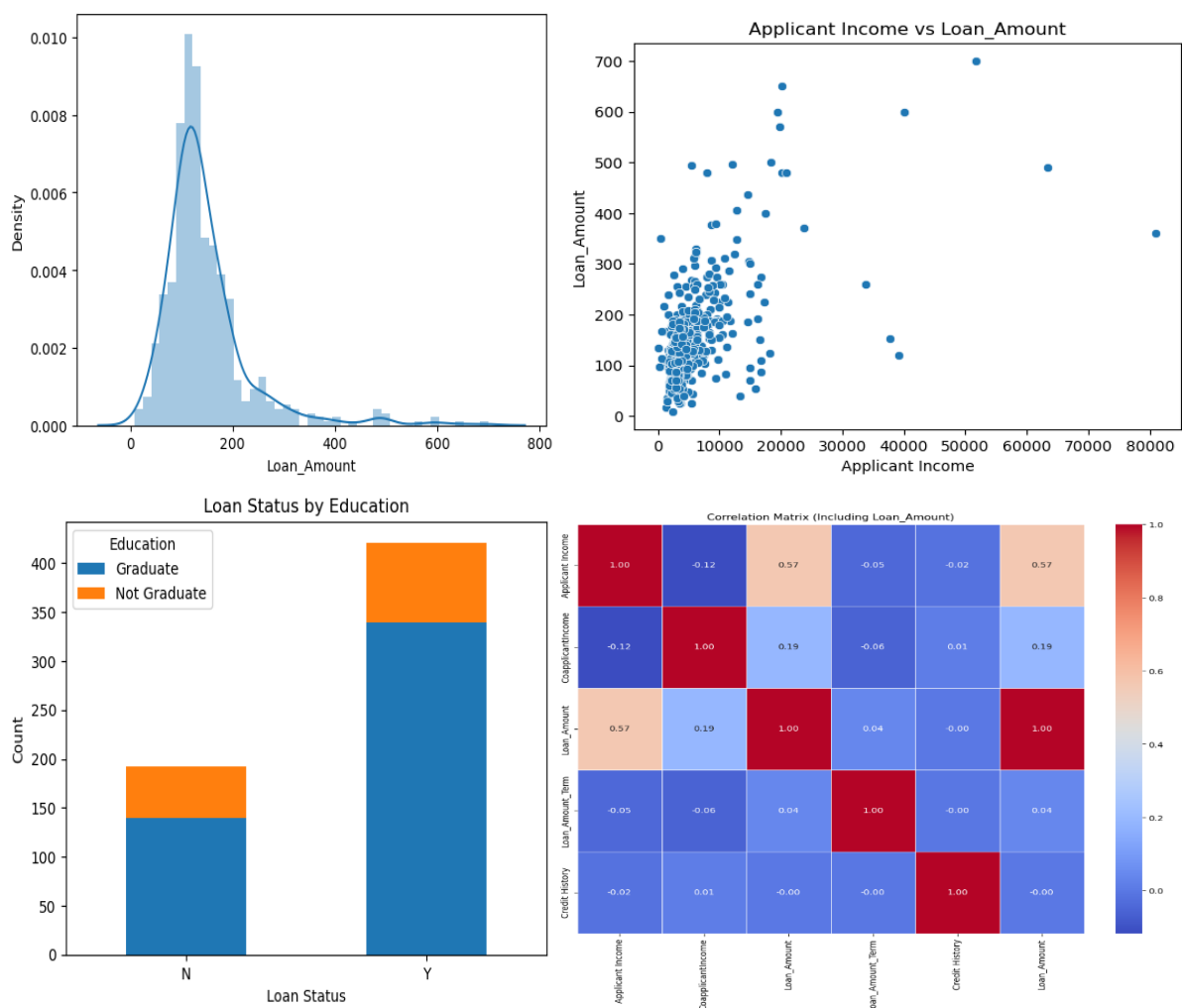
Loan Application Status Prediction

Research Objective:

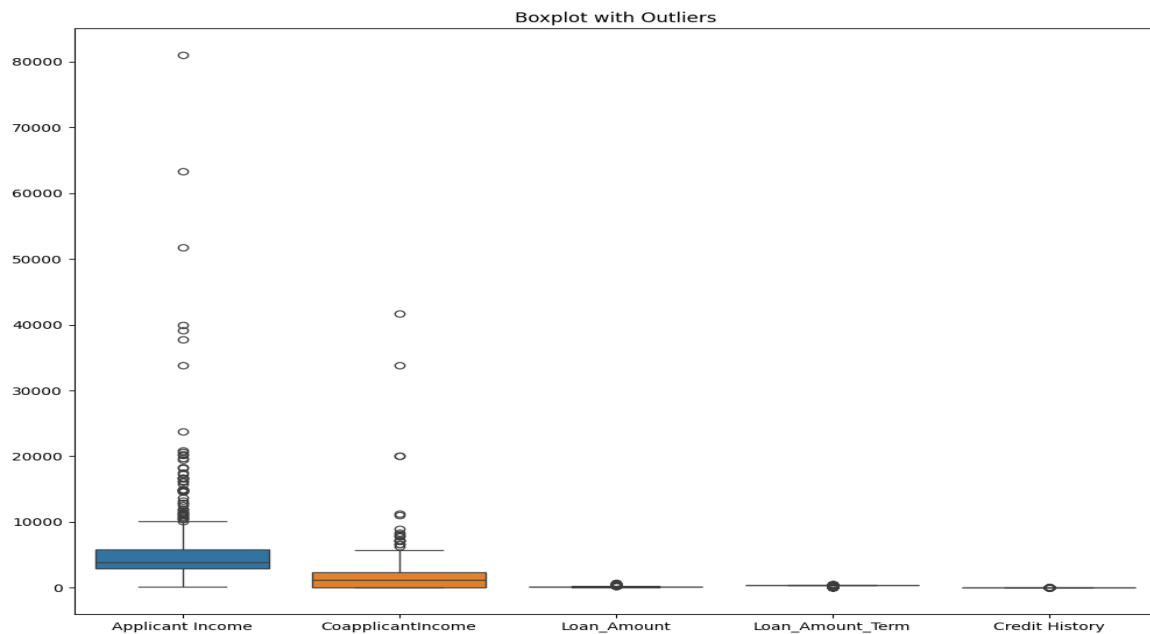
This research aimed to develop a predictive model to accurately forecast loan approvals based on various applicant characteristics. By analyzing key features such as credit history, loan amount, income, and dependents, the goal was to create a reliable model that could assist lenders in making informed decisions.

Data Exploration and Preprocessing:

Exploratory Data Analysis (EDA) was conducted to gain a comprehensive understanding of the dataset. Univariate, bivariate, and multivariate analyses were employed to examine the distribution of variables, identify relationships between features, and assess potential multicollinearity.



Data preprocessing was a critical step to ensure data quality and consistency. Outliers were identified in Applicant Income, Coapplicant Income, Loan Amount, and Loan Amount Term, potentially skewing the analysis. To mitigate their impact, appropriate outlier handling techniques were implemented.



Multicollinearity, a situation where independent variables are highly correlated, can affect model performance. To address this, Variance Inflation Factor (VIF) was calculated for each variable. As all VIF values were below 10, multicollinearity was deemed low, allowing for the inclusion of all variables in the model without concerns.

Model Development and Evaluation:

Logistic Regression was selected as the predictive model due to its suitability for classification problems. This model estimates the probability of an applicant's loan being approved based on the input features.

Hyperparameter tuning was employed to optimize the model's performance. By systematically adjusting key parameters, the model's accuracy and generalization capabilities were enhanced. The final model configuration included a regularization parameter (C) of 10, an L2 penalty, and the liblinear solver.

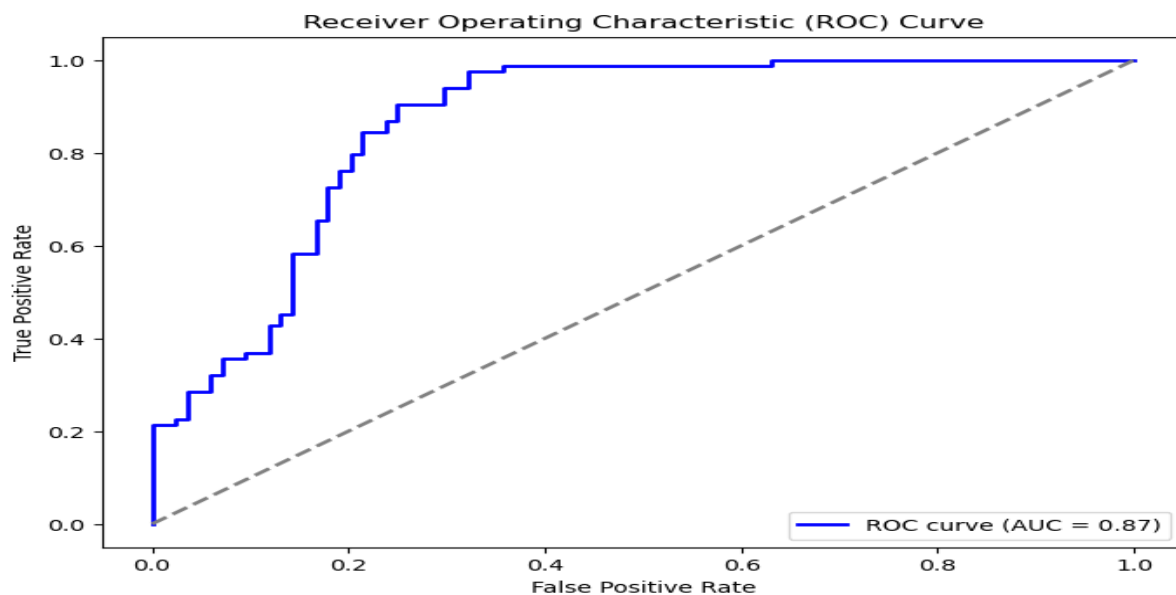
```
Best Classifier: LogisticRegression
Best Parameters: {'C': 10, 'penalty': 'l2', 'solver': 'liblinear'}
Accuracy: 0.8274
Classification Report:
      precision    recall  f1-score   support

     N       0.97     0.68     0.80        84
     Y       0.75     0.98     0.85        84

 accuracy              0.83       168
 macro avg       0.86     0.83     0.82       168
 weighted avg     0.86     0.83     0.82       168
```

The model's performance was evaluated using various metrics, including accuracy, precision, recall, and F1-score. These metrics provide insights into the model's ability to correctly classify loan

applications, minimize false positives and negatives, and achieve a balance between precision and recall.



Results and Insights:

The Logistic Regression model achieved an impressive accuracy of 82.74%, indicating its effectiveness in predicting loan approvals. The precision, recall, and F1-score metrics further demonstrated the model's ability to accurately classify both approved and rejected loans.

Key findings from the analysis include:

Applicants with lower incomes were more likely to apply for loans.

The distribution of loan amounts was skewed to the right, indicating a concentration of smaller loans with a few larger outliers.

Outlier handling and addressing multicollinearity were essential for improving model performance.

Logistic Regression, with appropriate hyperparameter tuning, emerged as the most effective classifier for this task.

Conclusion:

This research successfully developed a predictive model for loan approval based on applicant data. The model, built using Logistic Regression, demonstrated strong performance in accurately classifying loan applications. The findings provide valuable insights for lenders in assessing loan applications and making informed decisions.

Reference:

Myers, J.L., Well, A.D., and Lorch, R.F. Jr, 2010. Research Design and Statistical Analysis. 3rd ed. New York: Routledge. eBook [Online]. Available at: <https://doi.org/10.4324/9780203726631>

Sarker, I.H., 2021. *Machine Learning: Algorithms, Real-World Applications and Research Directions*. *SN Computer Science*, 2, p.160. [Online] Available at: <https://doi.org/10.1007/s42979-021-00592-x>

Petersen, S.E. and Ostendorf, M., 2009. A machine learning approach to reading level assessment. *Computer Speech & Language*, 23(1), pp.89-106. [Online] Available at: <https://doi.org/10.1016/j.csl.2008.04.003>