**Loan Amount Prediction Project Report**

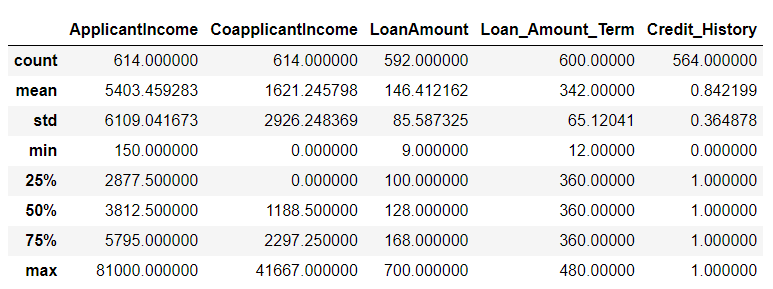
**By Brindhalakshmi**

**Introduction:**

The Loan Amount Prediction project aims to develop a predictive model that estimates the loan amount a borrower is eligible for based on various factors. This project is crucial for financial institutions to streamline their lending processes, assess risks accurately, and provide suitable loan amounts to applicants. The primary objective is to leverage machine learning techniques to build a robust predictive model capable of accurately estimating the loan amount. By analysing historical loan data and identifying patterns, the model will learn to predict loan amounts for future applicants.

**Data Description:**

The data description represents an analysis of loan-related data including Applicant Income, Co applicant Income, Loan Amount, Loan Amount Term, and Credit History.



**Application Income:**

* Slope of Peak: Down sloping
* Description: The distribution of Applicant Income indicates a down sloping trend, with the peak occurring at lower income levels. This suggests that many loan applicants have relatively lower incomes.

**Co applicant Income:**

* Slope of Peak: Down sloping
* Description: Like Applicant Income, the distribution of Co applicant Income also exhibits a down sloping trend, with the peak occurring at lower income levels.

**Loan amount:**

* Slope of Peak: Flat
* Description: The distribution of LoanAmount appears relatively flat, indicating that loan amounts vary across a wide range without a clear peak.

**Loan amount term:**

* Slope of Peak: Flat
* Description: The distribution of Loan\_Amount\_Term shows a relatively flat trend, suggesting that loan terms are diverse without a distinct peak.

**Credit\_history:**

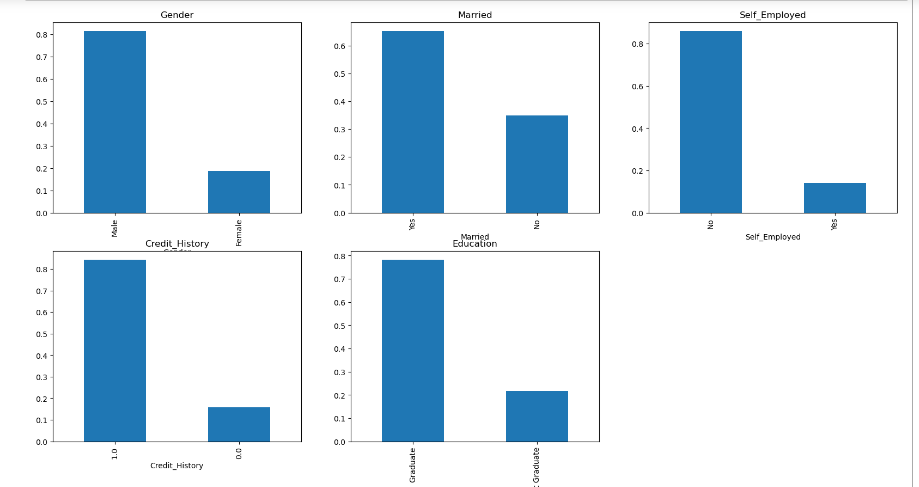
* Slope of Peak: Upsloping
* Description: The distribution of Credit\_History indicates an upsloping trend, with the peak occurring at 1, implying that a significant portion of loan applicants have a credit history meeting the required criteria.

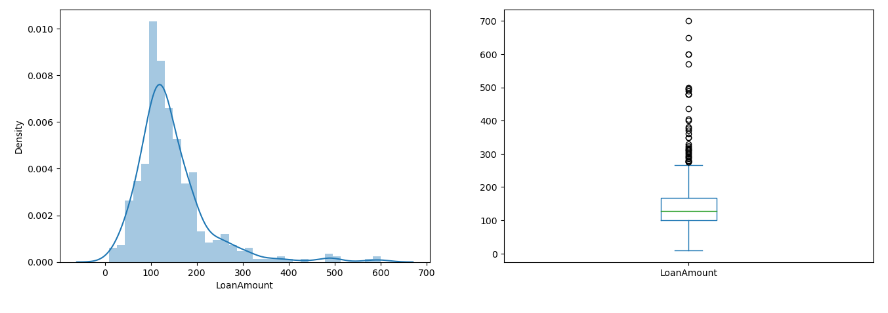
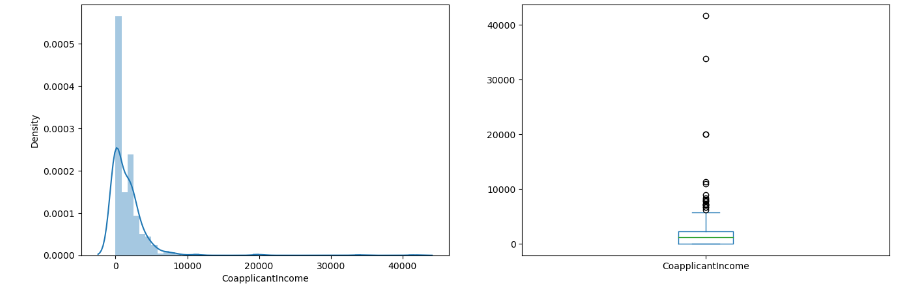
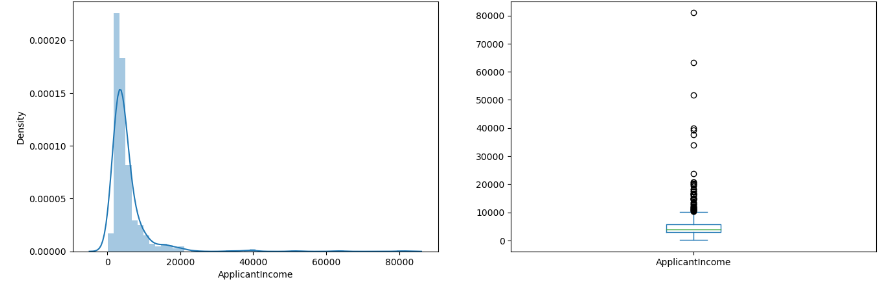
**Approach:**

This project aimed to figure out if someone gets a loan by using computers to learn from past loan applications. We collected a bunch of info about people asking for loans, like their money situation and credit history, then made sure the info was ready for the computer. We looked at the info to see if there were any clear patterns or signs that could help predict if someone would get a loan. After trying out different computer ways to learn from the info, one method stood out: it guessed right 82% of the time! This method, called XGBoost, was our best bet for predicting loan approvals. By understanding these patterns, we hope to make the loan approval process smoother and fairer for everyone involved.

**Data Visualization:**

For data visualization, numerical data was analyzed using distribution plots (dist plot), providing insights into the spread and shape of the data distribution. Additionally, categorical data underwent visualization through histograms and count plots, allowing for a clear understanding of the frequency distribution and proportions of different categories within the dataset. These visualization techniques collectively facilitated a comprehensive exploration of the dataset, enabling us to identify trends, outliers, and patterns essential for further analysis and decision-making processes.





**Algorithm:**

* **Logistic Regression:**

Utilized StratifiedKFold to evaluate Logistic Regression's performance, leveraging its simplicity and interpretability for binary classification. Logistic Regression serves as a strong baseline model, modeling linear relationships between features and the target variable. StratifiedKFold ensured robust evaluation while maintaining class proportions across folds, facilitating unbiased assessment for loan prediction tasks.

* **Decision Tree:**

Applied StratifiedKFold to assess Decision Tree's performance, exploiting its intuitive decision-making to capture non-linear relationships. Despite interpretability, Decision Trees may overfit; thus, rigorous evaluation was conducted to gauge suitability for loan prediction. This approach ensured fair evaluation, mitigating overfitting risks and yielding reliable results.

* **Random Forest:**

Evaluated Random Forest using StratifiedKFold to harness its ensemble-based approach, combining decision trees to reduce overfitting and improve accuracy. StratifiedKFold facilitated unbiased evaluation, maintaining class balance across folds. This method enabled thorough assessment of Random Forest's potential for accurate loan prediction, considering its ability to handle complex relationships effectively.

* **XGBoost Classifier:**

Assessed XGBoost Classifier's performance with StratifiedKFold, leveraging its gradient boosting technique for capturing intricate data patterns and achieving superior accuracy. XGBoost iteratively enhances model performance by combining weak learners into a strong ensemble. StratifiedKFold ensured unbiased evaluation and class balance, facilitating thorough assessment of XGBoost Classifier's suitability for loan prediction with high accuracy.

**Comparison and Evaluation:**

In evaluating four classification models for loan prediction, Logistic Regression demonstrated a mean validation accuracy of 80.13%, positioning it as a reliable performer. Despite its simplicity, Logistic Regression yielded competitive results, showcasing its effectiveness in capturing linear relationships within the data.

The Decision Tree model achieved a mean validation accuracy of 71.49%, indicating moderate predictive power but potentially suffering from overfitting due to its tendency to capture noise in the data. Random Forest, with a mean validation accuracy of 78.50%, showcased strong predictive performance by leveraging the collective wisdom of multiple decision trees, offering robustness against overfitting and high accuracy. XGBoost Classifier emerged as the top performer with a mean validation accuracy of 80.45%, highlighting its ability to capture complex relationships in the data and achieve superior predictive accuracy.

Despite its computational complexity, XGBoost proved its worth as a powerful tool for loan prediction tasks. Overall, while Logistic Regression provides a solid baseline, Random Forest and XGBoost Classifier stand out as promising models for accurate loan prediction, offering potential for deployment in real-world scenarios to optimize lending decisions and enhance financial outcomes.

**Conclusion:**

Employing advanced machine learning techniques, including Logistic Regression, Decision Tree, Random Forest, and XGBoost Classifier, in conjunction with the StratifiedKFold method facilitated a comprehensive evaluation for loan prediction tasks. While Logistic Regression served as a robust baseline model, XGBoost Classifier emerged as the optimal choice, exhibiting superior performance in capturing complex data patterns and achieving high accuracy. This project underscores the significance of leveraging ensemble-based algorithms like XGBoost in financial applications, showcasing their ability to handle intricate data relationships effectively and enhance predictive performance.

Reference:

<https://www.academia.edu/download/69679375/IRJET_V8I4785.pdf>

<https://www.sciencedirect.com/science/article/pii/S2666307423000293>

<https://ieeexplore.ieee.org/abstract/document/9155614/>