**Loan Eligibility Prediction**

Kankana Ghosh

**Introduction:**

In the rapidly evolving landscape of financial services, the accurate assessment of loan eligibility is of paramount importance for both lenders and borrowers alike. This report delves into the intricate realm of loan eligibility prediction, utilizing advanced data analytics and machine learning techniques to forecast whether an individual is qualified for a loan based on various factors and historical trends. In this project, I am going to develop one such model which can predict whether a person will get his/her loan approved or not by using some of the background information of the applicant like the applicant’s gender, marital status, income, etc.

**Methodology:**

The dataset is DataFrame in nature having 559 observations on 5 variates. The columns are Gender, Married, Applicant Income, Loan Amount and Loan Status. Here, the objective of this project is to predict whether a person will get his/her loan approved or not by using the background information of the applicant. The dataset doesnot contain any missing values.

In the initial analysis of the dataset, it is evident from the pie chart representation that a notable imbalance exists among the two classes (‘Yes’ and ‘No’) of Loan Status. One of the main observations we can draw from bar graphs representing the count of Gender and Marital status is that the chances of getting a loan approved for females are quite low compared to males and loan approval for married people are quite low compared to those who are not married.The histogram gives the visual representation of the distribution of 'Applicant Income' and 'Loan Amount' columns. Both the plots reveal concentration of Incomes, common loan amount ranges, potential outliers, and skewness in the distribution. Both are positively skewed. From the boxplots of Applicant Income and Loan Amount shows some extreme outliers present in the data. We need to remove them before further analysis.The loan amount requested by males is higher than what is requested by females. One more interesting observation is that the married people requested loan amount is generally higher than that of the unmarried. This may be one of the reason’s that we observe earlier that the chances of getting loan approval for a married person are lower than that compared to an unmarried person.The heatmap shows that most of the attributes are mildly correlated to each other. The attributes Married-Gender and LoanAmount-ApplicantIncome are moderately positively correlated with values 0.38 and 0.46 respectively. Support Vector Classifier is used for training the model. The dataset is split into train and test set in 80:20 ratio.The evaluation metrics provide insights into the model's performance on the dataset. Precision, indicating the accuracy of positive predictions, is 26% for class 0 (Loan Status: ‘No’) and 72% for class 1(Loan Status:’Yes’). Recall, representing the model's ability to identify true positives, is 29% for class 0 and 71% for class 1. The F1-score, a balance between precision and recall, is 0.28 for class 0 and 0.70 for class 1. The dataset contains 31 instances of class 0 and 81 instances of class 1. Overall accuracy stands at 58%, depicting correct predictions for 58% of instances. Macro average metrics, treating all classes equally, are 0.49, while weighted average metrics, accounting for class distribution, are 0.59. The model performs relatively better for class 1 compared to class 0. The F1-score's weighted average of 0.59 indicates a moderate overall performance. The lower precision and recall for class 0 suggest potential challenges or imbalances affecting the model's predictions for this class.

**Conclusion:**

As this dataset contains fewer features the performance of the model is not up to the mark maybe if we use a better and big dataset we will be able to achieve better accuracy.