Loan Sanction Predictor

- **Sathvik N Shendige**

Introduction:

The loan sanction predictor project aims to develop a robust and reliable machine learning model capable of predicting the likelihood of loan approval based on various applicant attributes. This predictive tool is crucial for financial institutions seeking to streamline and enhance their loan approval processes, ultimately improving efficiency and customer satisfaction.

Project Overview:

1. *Automated Decision-Making*: The project addresses the need for an automated decision-making system that assesses loan applications efficiently, reducing manual effort and processing times.

2. *Risk Mitigation:* By leveraging machine learning algorithms, the project aims to identify key factors influencing loan approval, helping financial institutions mitigate risk associated with lending.

3. *Data-Driven Insights:* The analysis of applicant data provides valuable insights into patterns and trends affecting loan outcomes, empowering decision-makers with data-driven information.

4. *Optimized Approval Process:* The ultimate goal is to optimize the loan approval process, ensuring that qualified applicants receive timely approvals while maintaining risk management standards.

Dataset Overview:

The project utilizes a comprehensive dataset containing various attributes such as applicant income, co-applicant income, credit history, property area, and more. Exploratory data analysis and feature importance analysis have been conducted to understand the dataset's characteristics and identify significant predictors.

Key Components and Concepts:

1. _Machine Learning Models_: The project employs machine learning models, including Random Forest, to predict loan approval outcomes. These models have been selected for their ability to handle a mix of categorical and numerical features and provide interpretable insights.

2. _Feature Importance Analysis_: Understanding the importance of each feature in the prediction process is crucial. Feature importance analysis guides model refinement and provides valuable insights for decision-makers.

3. _Visualization and Reporting_: Conclusions drawn from the dataset are presented through visualizations, including the distribution of applicant and co-applicant income. The project concludes with a detailed report summarizing key findings and recommendations.

Detailed Explanation:

1. _Overview of Loan Sanction_: Total Loan Approves – 422 and Total Loan Denials – 192

2. _Distribution of Applicant Income_:

    0k - 169 applicants: This group represents applicants with little to no income.

    3k - 299 applicants: A significant number of applicants fall into this income range.

    6k - 77 applicants: Moderate-income applicants.

    9k - 30 applicants: Higher moderate-income applicants.

    12k - 10 applicants: Higher-income applicants.

    15k - 13 applicants: Applicants with substantial income.

    18k - 7 applicants: High-income applicants.

    21k - 1 applicant: Very high-income applicant.

    24k - 1 applicant: Very high-income applicant.

    33k - 1 applicant: Exceptionally high-income applicant.

    36k - 1 applicant: Exceptionally high-income applicant.

    39k - 2 applicants: Exceptionally high-income applicants.

    50k - 1 applicant: Exceptionally high-income applicant.

    62k - 1 applicant: Exceptionally high-income applicant.

    80k - 1 applicant: Exceptionally high-income applicant.

3. *Distribution of Co-applicant Income*:

   0k - 389 applicants: Most applicants do not have a co-applicant or have a co-applicant with no income.

   2k - 158 applicants: A significant number of applicants have a co-applicant with a moderate income.

   4k - 42 applicants: Some applicants have a co-applicant with a higher moderate income.

   5k - 13 applicants: Few applicants have a co-applicant with substantial income.

   7k - 6 applicants: A small number of applicants have a co-applicant with a high income.

   11k - 2 applicants: Very few applicants have a co-applicant with a very high income.

   20k - 2 applicants: Very few applicants have a co-applicant with a very high income.

   33k - 1 applicant: Exceptionally high-income co-applicant.

   42k - 1 applicant: Exceptionally high-income co-applicant.

4. *Feature Importance Analysis*:

   Credit_History (26.5%): The most crucial feature influencing loan approval.

   ApplicantIncome (14.8%): Significant impact on the model's predictions.

   Loan_ID (14.4%): Uniqueness of Loan IDs, might not have a meaningful impact.

   LoanAmount (13.9%): The requested loan amount affects predictions.

   CoapplicantIncome (8.5%): Contributes to predictions but less important.

   Loan_Amount_Term (4.8%): The term of the loan has a smaller impact.

   Property_Area_Semiurban (2.5%) and Property_Area_Urban (2.3%): Property location affects predictions.

   Married_Yes (2.2%): Married applicants may have a slightly higher chance of approval.

   Education_Not Graduate (2.2%): Non-graduates may face a slightly lower chance of approval.

   Dependents_1 (1.9%) and Dependents_2 (1.3%): The number of dependents has a moderate impact.

   Gender_Male (1.8%): Gender has a slight impact.

   Self_Employed_Yes (1.7%): Self-employed applicants may have a slightly lower chance of approval.

   Dependents_3+ (0.9%): Applicants with more dependents have a lower impact.

5. *Model Performance*: Training accuracy – 100% and Test accuracy – 79%

Conclusion:

The model shows a high training accuracy, indicating a potential risk of overfitting to the training data.The test accuracy of 79% suggests reasonable generalization to new data. Credit history, applicant income, and loan amount are the most influential factors in predicting loan approval. Geographic location (semiurban and urban) plays a notable role in loan approval. Marital status, education level, and the number of dependents also impact predictions to a lesser extent. The model demonstrates a strong ability to discriminate between approved and denied loans.