# Loan Approval Prediction - Project Documentation

## 1. Introduction

Loan approval prediction is a critical task in the financial sector, helping banks and lending institutions assess the likelihood of loan repayment by applicants. This project focuses on building a predictive model to determine loan approval status based on applicant data using machine learning techniques.

## 2. Problem Statements

1. Can we predict loan approval status with high accuracy using historical data?

2. What are the key factors influencing loan approval decisions?

3. How can predictive analytics improve the efficiency of the loan approval process?

## 3. Data Sources

The dataset used in this project contains information about loan applicants, such as their income, credit history, loan amount, and more. This dataset was sourced from an open Kaggle competition and includes the following features:  
- Applicant Income  
- Co-applicant Income  
- Loan Amount  
- Credit History  
- Loan Term  
- Loan Status (Target Variable)

## 4. Methodology

The project methodology involves the following steps:

1. **Data Preprocessing**: Handling missing values, encoding categorical variables, and scaling numerical features.

2. **Exploratory Data Analysis (EDA)**: Identifying patterns and relationships between features using visualization techniques.

3. **Feature Engineering**: Creating new features and selecting the most important ones for the predictive model.

4. **Model Building**: Training and evaluating machine learning models such as Logistic Regression, Random Forest, and XGBoost.

5. **Evaluation Metrics**: Measuring model performance using accuracy, precision, recall, and F1-score.

## 5. Results

The final model achieved the following performance metrics:  
- Accuracy: 85%  
- Precision: 80%  
- Recall: 78%  
- F1-Score: 79%  
The model demonstrates reliable performance, making it suitable for deployment in real-world loan approval scenarios.

## 6. Insights

1. Applicants with a clear credit history and higher income levels are more likely to have their loans approved.

2. Loan Amount and Loan Term have a significant influence on approval decisions.

3. Proper feature engineering and handling of missing values substantially improved model accuracy.

## 7. Conclusion

This project successfully built a predictive model for loan approval status using machine learning techniques. By leveraging applicant data, the model helps financial institutions streamline their loan approval process, improve decision-making, and reduce manual effort. Further improvements can be made by incorporating additional data and exploring advanced models.