# Final Report: Loan Approval Prediction Using CRISP-DM Framework

## 1. ****Business Understanding****

The goal of this project is to predict loan approval status based on various applicant features. This will assist financial institutions in making informed lending decisions, minimizing risk, and improving customer satisfaction.

### Key Questions

* What factors most significantly influence loan approval?
* How can we reduce the risk of default on loans?
* What is the acceptable accuracy level for the predictive model?
* How can we ensure the model adapts to changing economic conditions?

## 2. ****Data Understanding****

### 2.1 Data Collection

The dataset used for this project is sourced from a loan approval dataset, containing various features related to loan applications.

### 2.2 Data Exploration

The dataset contains 4269 rows and 13 columns, including:

* **loan\_id**
* **no\_of\_dependents**
* **education**
* **self\_employed**
* **income\_annum**
* **loan\_amount**
* **loan\_term**
* **cibil\_score**
* **residential\_assets\_value**
* **commercial\_assets\_value**
* **luxury\_assets\_value**
* **bank\_asset\_value**
* **loan\_status** (target variable)

### 2.3 Data Quality

Initial checks indicate no missing values in the dataset. Data types include integers and categorical values.

## 3. ****Data Preparation****

### 3.1 Data Cleaning

* Stripped whitespace from column names.
* Coded categorical variables for model compatibility.

### 3.2 Feature Selection

Selected relevant features based on business understanding and exploratory data analysis.

### 3.3 Data Transformation

Applied transformations, including normalization and encoding, to prepare the dataset for modeling.

## 4. ****Modeling****

### 4.1 Model Selection

Various models were tested, including:

* **Logistic Regression**
* **Random Forest**
* **Support Vector Machine**

### 4.2 Model Training

The dataset was split into training and testing sets. Models were trained and validated using appropriate metrics.

### 4.3 Model Evaluation

Performance metrics included:

* Accuracy
* Precision
* Recall
* F1 Score  
  Confusion matrices were generated to assess model performance.

## 5. ****Deployment****

The best-performing model (Logistic Regression) was selected for deployment. A user-friendly interface was developed for loan officers to input applicant data and receive predictions.

## 6. ****Monitoring and Maintenance****

Regular evaluations will be conducted to ensure model performance remains optimal. Updates will be made as new data becomes available or as business requirements change.

## 7. ****Conclusion****

The project successfully developed a predictive model for loan approval, enhancing decision-making processes in lending institutions. Future work will focus on continuous improvement and adaptation to changing market conditions.