

**Loan Status Prediction**

**Name:**

**Shoaib Alam Khan**

**Roll No.**

**041**

**Department:**

**Software engineering**

**Section:**

**BSAI (3A)**

**Professor:**

**Sir Rasikh Ali**

# Project Report: Loan Status Prediction

## 1. Introduction

Loan status prediction is a critical application in the banking sector. By predicting whether a loan applicant is likely to default or repay, financial institutions can make more informed decisions. This report outlines the development of a predictive model for loan status prediction based on customer attributes.

## 2. Problem Statement

The goal of this project is to predict the loan status (“Approved” or “Not Approved”) based on the provided customer information:  
- Features:  
 - Gender  
 - Married  
 - Dependents  
 - Education  
 - Self\_Employed  
 - ApplicantIncome  
 - CoapplicantIncome  
 - LoanAmount  
 - Loan\_Amount\_Term  
 - Credit\_History  
 - Property\_Area  
- Target:  
 - Loan\_Status

## 3. Dataset

### 3.1 Data Description

The dataset contains customer demographic details, financial information, and loan status. A brief description of the features is as follows:

|  |  |
| --- | --- |
| Column | Description |
| Gender | Male or Female |
| Married | Applicant marital status |
| Dependents | Number of dependents |
| Education | Applicant education level |
| Self\_Employed | Whether the applicant is self-employed |
| ApplicantIncome | Applicant’s monthly income |
| CoapplicantIncome | Co-applicant’s monthly income |
| LoanAmount | Loan amount requested |
| Loan\_Amount\_Term | Loan repayment term in months |
| Credit\_History | Credit history of the applicant (1 or 0) |
| Property\_Area | Area type (Urban, Semiurban, Rural) |
| Loan\_Status | Loan approved or not (‘Y’ or ‘N’) |

### 3.2 Preprocessing Steps

1. Handling Missing Values:  
 - Missing values were removed to ensure data integrity.  
2. Label Encoding:  
 - Encoded categorical variables such as `Married`, `Gender`, and `Property\_Area` with numerical equivalents.  
 - Replaced `Loan\_Status` (‘Y’ and ‘N’) with 1 and 0, respectively.  
3. Feature Selection:  
 - Dropped irrelevant columns such as `Loan\_ID`.

## 4. Data Visualization

### 4.1 Key Visualizations

1. Education vs Loan Status:  
 - A count plot showed that graduates had a higher probability of loan approval.  
2. Marital Status vs Loan Status:  
 - Married applicants were more likely to get loan approval.  
3. Self-Employment vs Loan Status:  
 - Self-employed applicants showed a slightly lower approval rate.

## 5. Methodology

### 5.1 Train-Test Split

The dataset was split into training and testing sets using a 90:10 ratio to train and evaluate the model effectively.

### 5.2 Model Selection

The Support Vector Machine (SVM) algorithm with a linear kernel was chosen for its simplicity and effectiveness in classification tasks.

## 6. Results

### 6.1 Evaluation Metrics

1. Training Accuracy: 86.48%  
2. Testing Accuracy: 83.33%  
These metrics demonstrate that the SVM model performed well without significant overfitting.

## 7. Conclusion

The project successfully predicted loan approval status using demographic and financial data. The SVM model achieved high accuracy and revealed critical insights, such as the strong influence of credit history and applicant income on loan status.

### 8. Predictive System

A predictive system was developed to take user input and predict loan approval status:

Example Input:  
  
input\_data = (1, 0, 0, 1, 0, 4583, 1508.0, 128.0, 360.0, 1, 2)  
  
Prediction Output:  
- Loan Approval Prediction: Yes