

<LOAN ELIGIBILITY PREDICTION USING MACHINE LEARNING>

Submitted for

Intelligent Model Design Using AI CSET225

Submitted by:

(<E23CSEU1142>) ABHIJEET SINGH

Submitted to

MR. AMIT CHAUHAN SIR (Add your Lab Faculty Name)

Aug-Dec 2025

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING



INDEX

Sr. No	Content	Page No
1.	ABSTRACT, INTRODUCTION , RELATED WORK	1
2.	PROBLEM STATEMENT , CONTRIBUTION , DATASET DESCRIPTION	2
3.	MODEL DESCRIPTION , RESULTS , CONCLUSION,	3
4.	FUTURE WORK, REFERENCES, GITHUB LINK	4

ABSTRACT

This project focuses on predicting loan eligibility using machine learning algorithms. The objective is to automate the loan approval process by analyzing various applicant attributes such as income, loan amount, credit history, and education. The system utilizes supervised learning techniques to classify applicants as eligible or ineligible for loans. By using models like Logistic Regression, Decision Tree, and Random Forest, the project demonstrates how data-driven predictions can improve efficiency, reduce human bias, and streamline decision-making in financial institutions.

.

INTRODUCTION

Loan eligibility prediction plays a significant role in the financial sector, particularly in banking, where determining a customer's creditworthiness is essential. Traditionally, loan assessment relies on manual evaluation which is prone to human error and inconsistency. This project aims to develop an automated system that predicts loan eligibility based on applicant details using machine learning models implemented in Python. The approach enhances accuracy and reduces processing time.

RELATED WORK

Previous studies in financial data prediction have used various machine learning models for credit risk and loan eligibility. Techniques such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines (SVM) have been commonly applied. Ensemble methods like Random Forest have shown improved accuracy due to their ability to handle complex relationships among features. Recent works also emphasize the role of data preprocessing and feature selection in enhancing prediction performance.

PROBLEM STATEMENT

The primary goal of this project is to predict whether a loan applicant is eligible for a loan or not, based on key financial and personal attributes. The problem involves a binary classification task where the target variable indicates loan approval status. The challenge lies in ensuring accurate predictions despite possible imbalances in the dataset and variations in applicant profiles.

CONTRIBUTION

- Collected and analyzed the loan dataset from open-source repositories (Kaggle).
- Performed data preprocessing to handle missing values and categorical encoding.
- Conducted Exploratory Data Analysis (EDA) using visualizations to understand feature relationships.
- Implemented multiple classification algorithms including Logistic Regression, Decision Tree, and Random Forest.
- Evaluated models based on accuracy, precision, recall, and F1-score.
- Selected the best-performing model for loan eligibility prediction.

DATASET DESCRIPTION

The dataset includes various features such as:

- Gender, Marital Status, Education
- Applicant Income and Co applicant Income
- Loan Amount and Loan Term
- Credit History and Property Area

The dataset was preprocessed by filling missing values, encoding categorical variables, and normalizing numerical features to ensure consistent model training.

MODEL DESCRIPTION

The project utilizes machine learning models such as Logistic Regression, Decision Tree, and Random Forest.

- Logistic Regression: A statistical model used for binary classification.
- Decision Tree: A tree-based model that splits data into subsets based on feature conditions.
- Random Forest: An ensemble of multiple decision trees that improves prediction accuracy and reduces overfitting.

The models were trained and evaluated using Scikit-learn in Python.

RESULTS

The Random Forest model achieved the best performance among the models tested.

Key performance metrics:

- Accuracy: ~82–85%
- Precision: ~80%
- Recall: ~83%
- F1-score: ~81%

The model successfully predicts loan eligibility with balanced accuracy and robustness, indicating its suitability for real-world deployment.

CONCLUSION

The Loan Eligibility Prediction System effectively automates the process of loan approval decisions. By using machine learning algorithms, it reduces manual intervention and enhances accuracy. The project demonstrates how AI and ML techniques can be applied to financial systems to optimize operations and minimize human error in decision-making.

FUTURE WORK

- Integrate deep learning models for improved prediction performance.
- Build a web application interface for real-time loan eligibility prediction.
- Include more financial and behavioral attributes for better decision-making.
- Deploy the system on cloud platforms for scalability and accessibility.

REFERENCES

1. Kaggle Loan Prediction Dataset
2. Scikit-learn: Machine Learning in Python – Pedregosa et al. (2011)
3. Ensemble Methods in Machine Learning – Dietterich (2000)
4. Matplotlib and Seaborn documentation for visualization
5. Research papers on ML-based financial prediction systems

GITHUB LINK

<https://github.com/Abhi20037500/LOAN-ELIGIBILITY-PREDICTION-1>