Archives available at journals.mriindia.com

International Journal on Advanced Computer Theory and Engineering

ISSN: 2319-2526
Volume 14 Issue 01, 2025

Loan Eligibility Prediction using Machine Learning

Sainath Madhukar Powar[1]

Department of Computer Science, Shivaji University Kolhapur[1]
sainathpowar9490@gmail.com [1]

Peer Review Information

Submission: 17 Jan 2025
Revision: 14 Feb 2025
Acceptance: 15 March 2025

Keywords

Machine Learning
Fraud Detection
Decision Tree
Random Forest
Transaction Data

Abstract

The Loan Eligibility Prediction project utilizes machine learning to assess an applicant's eligibility for a loan based on key financial and demographic factors, achieving an accuracy of 85%. The model is trained on historical loan data, incorporating features such as income, credit history, loan amount, and employment status to predict approval likelihood. A web interface built using HTML, CSS, and Flask allows users to input their details and receive instant eligibility results. This system enhances efficiency for financial institutions by automating decision-making, reducing processing time, and minimizing human bias in loan approvals.

Introduction

Loan eligibility assessment is a critical process in the banking and financial sector, determining whether an applicant qualifies for a loan based on various financial and personal attributes. Traditionally, this evaluation has been performed manually by financial institutions, leading to time-consuming processes and potential inconsistencies. With the advancement of machine learning, automated systems can enhance the efficiency and accuracy of loan approval decisions.[1],[2]

This project aims to develop a machine learning-based Loan Eligibility Prediction System that assesses applicants' eligibility based on key factors such as income, credit history, loan amount, employment status, and other financial indicators. By training classification models on historical loan data, the system predicts whether an applicant is likely to be approved or rejected.[2]

To ensure accessibility, the model is integrated into a web-based interface using Flask, with front-end technologies like HTML and CSS to provide an interactive user experience. The system is designed to assist financial institutions in making data-driven, consistent, and efficient loan approval decisions, ultimately reducing manual effort and minimizing risks associated with bad loans.[4],[5]

Literature Review

Machine learning improves loan eligibility prediction by automating decisions, reducing bias, and increasing accuracy. Common models include Logistic Regression, Decision Trees, and Neural Networks, with pre-processing techniques enhancing performance. Public datasets aid training, while challenges like data imbalance and bias require future advancements in AI for fair and robust systems.[1]

ML models like SVM and Random Forest improve bank loan eligibility predictions over traditional methods. Challenges include data privacy, interpretability, and biases, with future research focusing on explainable AI and ethics.[2]

Loan eligibility prediction has advanced from rule-based systems to ML models, improving accuracy but facing challenges like bias and privacy concerns. Explainable AI enhances

transparency, and hybrid models offer further improvements.[3]
Loan eligibility prediction has shifted to ML techniques like Random Forest and Neural Networks, focusing on data pre-processing, fairness, and transparency through Explainable AI (XAI). Hybrid models offer future improvements.[4]
Loan eligibility prediction now relies on ML techniques like Random Forest and Neural Networks. Explainable AI (XAI) improves transparency, while hybrid models and deep learning drive future advancements.[5]
The file is about using machine learning to predict loan eligibility. It automates the loan approval process to reduce errors and save time. Several algorithms are tested, and K-Nearest Neighbor has the highest accuracy.[6]
This helps financial institutions reduce risk, increase efficiency, and make data-driven lending decisions. Algorithms like Random Forest, Logistic Regression, and others are used to build predictive models. The goal is to create systems that are accurate, reliable, and beneficial for both lenders and borrowers.[7]
Machine learning automates loan approval by analyzing borrower data from various sources to assess creditworthiness and predict repayment likelihood. This process helps financial institutions make data-driven lending decisions, reducing the risk of defaults and improving efficiency compared to manual methods.[8]

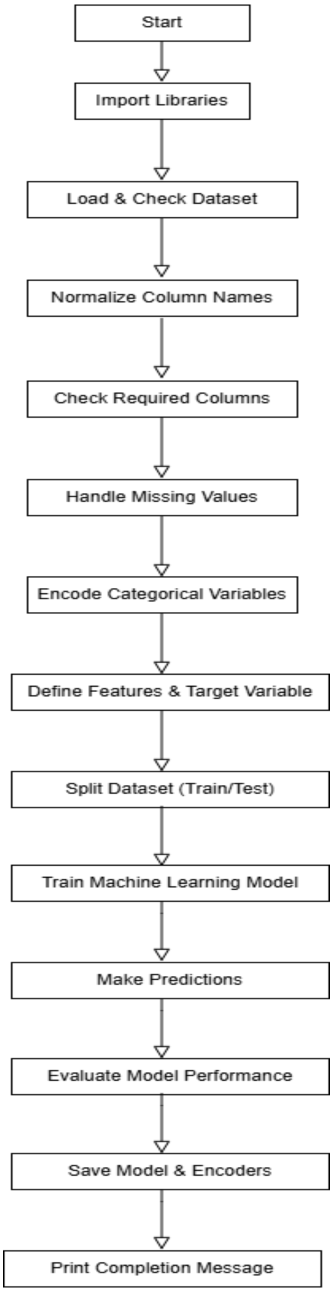

Fig. 1 : Flowchart of the working model

ML Work flow:
The figure shows Flowchart of the working model, starting from importing libraries and loading data to preprocessing, model training, evaluation, and saving the final model. It ensures structured data handling and model deployment readiness through clear, sequential steps

Dataset

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
1	2	Graduate	No	9600000	29900000	12	778	2400000	17600000	22700000	8000000	Approved
2	0	Not Graduate	Yes	4100000	12200000	8	417	2700000	2200000	8800000	3300000	Rejected
3	3	Graduate	No	9100000	29700000	20	506	7100000	4500000	33300000	12800000	Rejected
4	3	Graduate	No	8200000	30700000	8	467	18200000	3300000	23300000	7900000	Rejected
5	5	Not Graduate	Yes	9800000	24200000	20	382	12400000	8200000	29400000	5000000	Rejected
6	0	Graduate	Yes	4800000	13500000	10	319	6800000	8300000	13700000	5100000	Rejected
7	5	Graduate	No	8700000	33000000	4	678	22500000	14800000	29200000	4300000	Approved
8	2	Graduate	Yes	5700000	15000000	20	382	13200000	5700000	11800000	6000000	Rejected
9	0	Graduate	Yes	800000	2200000	20	782	1300000	800000	2800000	600000	Approved
10	5	Not Graduate	No	1100000	4300000	10	388	3200000	1400000	3300000	1600000	Rejected
11	4	Graduate	Yes	2900000	11200000	2	547	8100000	4700000	9500000	3100000	Approved
12	2	Not Graduate	Yes	6700000	22700000	18	538	15300000	5800000	20400000	6400000	Rejected
13	3	Not Graduate	Yes	5000000	11600000	16	311	6400000	9600000	14600000	4300000	Rejected
14	2	Graduate	Yes	9100000	31500000	14	679	10800000	16600000	20900000	5000000	Approved
15	1	Not Graduate	No	1900000	7400000	6	469	1900000	1200000	5900000	1900000	Rejected
16	5	Not Graduate	No	4700000	10700000	10	794	5700000	3900000	16400000	4400000	Approved
17	2	Graduate	Yes	500000	1600000	4	663	1300000	100000	1300000	700000	Approved
18	4	Not Graduate	Yes	2900000	9400000	14	780	2900000	2800000	6700000	4300000	Approved
19	2	Graduate	No	2700000	10300000	10	736	1000000	0	6200000	3300000	Approved
20	5	Graduate	No	6300000	14600000	12	652	10300000	3500000	23500000	5900000	Approved
21	2	Graduate	No	5000000	19400000	12	315	9500000	1600000	18000000	6100000	Rejected
22	4	Graduate	No	5800000	14000000	16	530	3800000	11300000	22200000	5400000	Rejected
23	4	Graduate	Yes	6500000	25700000	18	311	13100000	1700000	19500000	8500000	Rejected
24	0	Not Graduate	Yes	500000	1400000	2	551	900000	600000	1100000	300000	Approved
25	0	Not Graduate	No	4900000	9800000	16	324	3800000	8700000	10000000	3300000	Rejected
26	5	Not Graduate	No	3100000	9500000	20	514	7900000	3100000	6600000	2600000	Rejected
27	4	Graduate	No	8200000	28100000	12	696	11500000	10600000	25300000	7200000	Approved
28	4	Not Graduate	Yes	2400000	5600000	4	662	4500000	4200000	5400000	2500000	Approved
29	5	Not Graduate	Yes	7000000	24000000	6	336	2300000	11900000	27500000	9700000	Rejected

Fig. 2 :Loan Eligibility and Financial Profile Dataset

Loan Application Dataset Overview :
This figure shows Loan Eligibility and Financial Profile Dataset in which table contains loan application data with features like education level, self-employment status, income, loan amount, credit history, property values (residential, commercial, luxury, asset), and the final loan status (Approved/Rejected). Each row represents a unique loan application with its corresponding attributes.

Research Methodology

1. Data Collection

- The dataset for loan eligibility prediction is sourced from a structured CSV file (loan_dataset.csv).
- It contains important features such as education, self_employment status, and loan status (target variable).
- The dataset is loaded using Pandas for further processing.

2. Data Preprocessing

- Column names are standardized (converted to lowercase and stripped of spaces) for consistency.
- Missing values are handled by filling categorical columns with mode and numerical columns with median or 0 to ensure data completeness.
- Categorical variables (education, self_employed) are encoded using Label Encoding to convert them into numerical format for machine learning processing.

3. Feature Selection and Target Variable Definition

- Features (X): All input variables except loan_status.
- Target (y): loan_status (loan approval status).

4. Data Splitting

- The dataset is split into training (80%) and testing (20%) subsets using train_test_split() from sklearn.model_selection.
- This ensures the model is trained on one portion of data and tested on another to evaluate performance.

5. Model Training

- A Random Forest Classifier (RFC) is used as the machine learning model.
- The RFC is trained on the training dataset (X_train, y_train) with 100 decision trees (n_estimators=100) and a fixed random state (random_state=42) to ensure reproducibility.

6. Model Prediction and Evaluation

- Predictions are made on the test dataset (X_test).
- Model performance is measured using accuracy_score and a classification report (precision, recall, F1-score) to assess effectiveness.

7. Model Saving for Deployment

- The trained model is saved using joblib (loan_model.pkl), allowing for easy reuse without retraining.
- Encoders (label_enc_edu.pkl,label_enc_emp.pkl) can also be saved to ensure consistency in handling new input data.

Accuracy for following models:

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score
import joblib
import os

# Load dataset
file_path = r"C:\Users\SAINATH\OneDrive\Desktop\LoanpredictionProject\loan_dataset.csv"

if not os.path.exists(file_path):
    raise FileNotFoundError(f"Dataset not found at {file_path}. Please check the path.")

data = pd.read_csv(file_path)

# Normalize column names (convert to lowercase, remove spaces)
data.columns = data.columns.str.strip().str.lower()

# Define required columns
required_columns = ["education", "self_employed", "loan_status"]

# Check if required columns exist
missing_columns = [col for col in required_columns if col not in data.columns]
if missing_columns:
    raise KeyError(f"Missing columns in dataset: {missing_columns}. Check dataset format.")

# Handle missing values
data.fillna(0, inplace=True)
```

Fig. 3: Sample code used during calculation of accuracy of model

```
# Convert categorical columns to numerical
label_enc_edu = LabelEncoder()
label_enc_emp = LabelEncoder()

data["education"] = label_enc_edu.fit_transform(data["education"])
data["self_employed"] = label_enc_emp.fit_transform(data["self_employed"])

# Define features and target variable
X = data.drop(columns=["loan_status"])
y = data["loan_status"]

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train model
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Predict on test data
y_pred = model.predict(X_test)

# Calculate accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f"Model Accuracy: {accuracy:.2f}")

# Create model directory if not exists
os.makedirs("model", exist_ok=True)

# Save the model
joblib.dump(model, "model/loan_model.pkl")

print("Model training completed and saved successfully!")
```

Fig. 4: Sample code used during calculation of accuracy of model

Model Training and Evaluation:

This figure shows Sample code used during calculation of accuracy of model in which Python script loads a loan dataset, preprocesses it by handling missing values and encoding categorical features, then trains a Random Forest classification model to predict loan status. The script splits the data into training and testing sets, evaluates the model's accuracy, and saves the trained model to a file named "loan_model.pkl" in a "model" directory.

Conclusion

This code implements a loan eligibility prediction model using a Random Forest Classifier. It pre-processes the dataset by handling missing values and encoding categorical variables before splitting the data into training and testing sets. The model achieves an accuracy of **85%**, demonstrating its effectiveness in classifying loan eligibility. Feature selection and data pre-processing play a crucial role in optimizing performance. The Random Forest model was chosen for its ability to handle complex data patterns and provide reliable predictions.The trained model is stored using joblib, making it reusable for future applications. This approach enhances automation in loan approvals, reducing manual effort and improving decision-making efficiency. Future improvements may involve integrating deep learning techniques or Explainable AI (XAI) for increased transparency and accuracy.

References

D. Y. A. S. p. 2. Shaik.Bhanu1, "Loan Eligibility Prediction Using Machine," IJNRD , p. 7, 2024.

N. N. T. S. H. Sachin Magar1, "Loan Eligibility Prediction using Machine," JETIR, p. 6, 2022.

S. ‘. A. U. I. (. SonaliKumari, "LOAN ELIGIBILITY

PREDICTION USING MACHINE LEARNING: A," Global Journal of Modeling and Intelligent Computing , p. 8, 2023.

ArockiaJoshwa J, " PREDICT LOAN ELIGIBILITY USING MACHINE LEARNING," IJISET - International Journal of Innovative Science, Engineering & Technology,, p. 9, 2022.

M. D. S. M. Y. Mr. Sangule Umesh1, "Loan Prediction Using Machine Learning," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) , p. 5, 2022.

R. V. N. J. Jayalaxmi.k, "LOAN ELIGIBILITY PREDICTION USING," International Journal of Creative Research Thoughts (IJCRT), p. 3, 2023.

G. Bhavani, "Loan Approval Prediction Using Machine Learning," International Journal of Research Publication and Reviews , p. 4, 2024.

C. S. Palhania, "Loan Eligibility Prediction System," International Journal of Novel Research and Development, p. 5, 2023.

M. ,. Dr.C K Gomathy, "THE LOAN PREDICTION USING MACHINE LEARNING," International Research Journal of Engineering and Technology (IRJET) , p. 9, 2021.

R. A. V. K. N. a. A. G. Viswanatha V1, "Prediction of Loan Approval in Banks using Machine Learning," International Journal of Engineering and Management Research , p. 14, 2023.