Vivekanand Education Society's
Institute of Technology

(Autonomous Institute Affiliated to University of Mumbai, Approved by AICTE & Recognised by Govt. of Maharashtra)

NAAC accredited with 'A' grade

V.E.S.
Since 1962

Semester: VI Review: 6.1

Title of the Project: Loan Approval Prediction

Group Members:
Member 1: Abhinav Swaminathan (1)
Member 2: Aditya Ahuja (2)
Member 3: Ganesh Gupta (13)

Mentor Name: Mrs.Ravita Mishra

Content

- Overview
- Introduction
- Problem Statement
- Objectives
- Requirements
- Literature Survey
- Proposed System
- Proposed Design (along with UML Diagrams)
- Implementation
- Results and Analysis
- Conclusion
- References

Introduction to Project

Finance raising and lending for real estate, consumer, mortgage and companies' loans is the central part of almost every bank's business model. Lending money to inappropriate customers forms the major source of credit risk. The major share of the bank's assets comes directly from the profit derived from the bank's loans.

The primary goal of the banking community is to safely invest their capital. In the current scenario, many NBFCs and banks approve loans after a clear verification and authentication process, however, it remains uncertain whether the candidate selected is the worthy correct of all the applicants. In the case of housing finance companies, they can use this to find their target demographic of customers who can readily acquire and pay loans for houses.

Problem Statement

The system will analyze factors like income, loan amount, credit history, and demographics to identify eligible customer segments in real-time, enabling targeted marketing and efficient processing.Dream Housing Finance aims to automate loan eligibility verification based on customer details from online applications.

Objectives of the project

1. The first objective is to build a machine learning model that will predict the Loan Granted Status of a user with the highest accuracy.This will be done by building multiple ML models and comparing their performance.

2. The second objective is to help the housing finance companies with customer segmentation to find out which customers will successfully acquire and pay back loans, so that they can focus on marketing to these customers.

Requirements of the system (Hardware, software)

◆ **Software Requirements**

- **OS:** Windows 10/11, Linux (Ubuntu), macOS
- **Languages & Frameworks:** Python
- **ML Technologies:**
  - Regression Model
  - EDA
- **ML Libraries:**, Pandas,NumPy.
- **Dataset:** loan_prediction.csv
- **Development:** Jupyter Notebook, Google Colab.

Requirements of the system (Hardware, software)

◆ **Hardware Requirements**

● **Basic Training:** i5/i7 (10th Gen+), 8GB RAM, 100GB SSD, No GPU needed

● **Advanced Training:** i7/i9, 16GB+ RAM, NVIDIA RTX 3060+, 500GB SSD

● **Cloud GPU:** Google Colab, AWS EC2 GPU

● **Backend Server:** 2-4 Core CPU, 4GB RAM, 50GB SSD (AWS EC2, Render)

● **Frontend Hosting:** Vercel, Netlify

Literature Survey

sr.No	Title	Authors	Year	Objective	Key points	Conclusion	Drawbacks
[1]	Loan Approval Prediction based on Machine Learning Approach	Kumar Arun, Garg Ishan, Kaur Sanmeet	2023	Predict loan safety for a borrower.	- Data collection from relevant sources. - Comparison of various machine learning models on the data. - Training the system on the most promising model. - Testing the trained system for accuracy and reliability.	The most promising model effectively predicts loan eligibility.	Lacks real-time implementation details and scalability analysis.

Literature Survey

sr.No	Title	Authors	Year	Objective	Key points	Conclusion	Drawbacks
[2]	Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process	E. Chandra Blessie, R. Rekha	2019	Address challenges in lending for banks & NBFCs	- The increasing number of loan applications places significant challenges on banks and non-banking financial companies (NBFCs) with limited capital.	The paper emphasizes the urgent need for a safe, standardized loan approval process using machine learning algorithms to streamline lending practices and ensure financial stability.	The paper lacks a concrete machine learning-based solution to the problem and does not provide detailed examples of how machine learning can be applied in the loan sanctioning process.

Literature Survey

sr.No	Title	Authors	Year	Objective	Key points	Conclusion	Drawbacks
[3]	An Approach for Prediction of Loan Approval using Machine Learning Algorithm	M. A. Sheikh, A. K. Goel, T. Kumar	2020	Reduce risk in selecting reliable borrowers to minimize NPAs.	- Uses historical loan records to predict loan repayment behavior. - Trains a machine learning model to predict whether a borrower will be able to repay the loan on time.	Machine learning models significantly help in minimizing non-performing assets (NPAs) by accurately predicting which borrowers are likely to repay their loans on time, ensuring financial stability.	The paper does not provide sufficient discussion on the possible bias in datasets or address regulatory challenges in applying machine learning models for loan approval in real-world scenarios.

Proposed System

First step would be to pre-process the data, separate the data set into Training and Testing data.
After which we clean data by removing null and nan values, handling outliers.
Then using data visualization techniques to look for the necessary factors that we need for prediction.
This is a Classification problem and can be solved by using any Classification Algorithm. Algorithms that we will use for predictions are
● Logistic Regression
● Random Forest
● K-Nearest Neighbors
● Gaussian Naive Bayes Classifier
SVM

Then use the algorithm with highest accuracy

Proposed Design

The system design follows the standard CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, ensuring a systematic approach to problem-solving and model development:
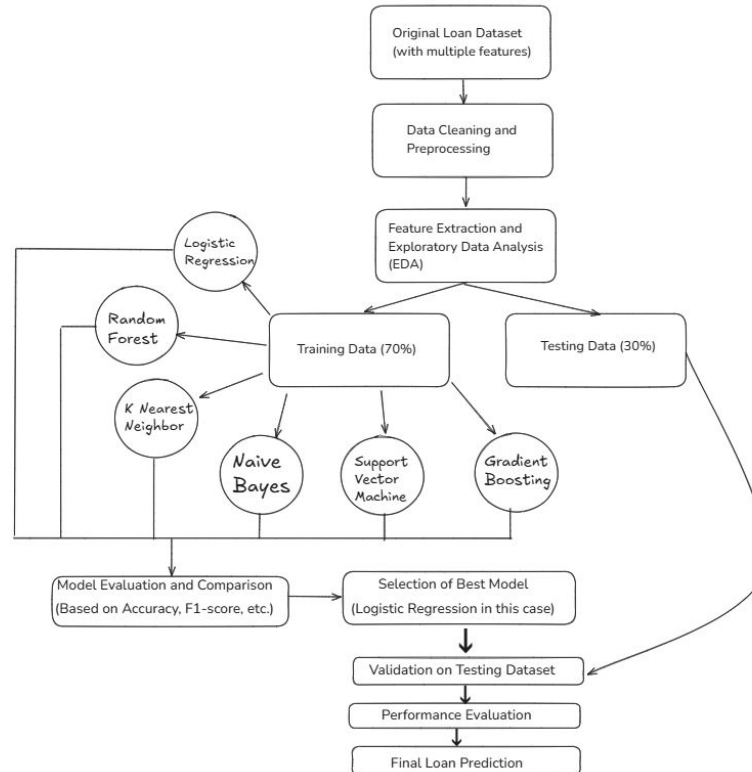
1. Data Collection

2. Data Preprocessing

3. Feature Engineering(New features)

4. Model Building (Logistic Regression , Random Forest Classifier, K-Nearest Neighbors (KNN)Gaussian Naive Bayes ,Support Vector Machine (SVM),Gradient Boosting Classifier)

5. Model Evaluation

6. Visualization

Proposed Design

Implementation

Loan Approval Prediction System

Enter your details to check if your loan would be approved.

Personal Information

Gender
Female

Marital Status
Yes

Number of Dependents
3+

Education
Graduate

Self Employed
No

Credit History ⊘
1

Financial Information

Annual Applicant Income (₹) ⊘
12,00,000

Annual Co-applicant Income (₹) ⊘
0

Loan Amount (₹) ⊘
3,00,000

Loan Term (months)
60

Property Area
Urban

Add Entry

✅ Congratulations! Your loan is likely to be APPROVED

Interest rate impact:

Base model probability:
71.75%

Adjusted with interest rate:
92.13%

Factors Affecting Decision

Key Metrics:

Credit History: Good

Income Level: High

Loan-to-Income Ratio: High

Property Area: Urban

Dependents: 3+ (High)

Interest Rate: 15.0%

Analysis:

• You have a good credit history

• Your combined annual income is high

• The loan amount is high compared to your income

• You have multiple dependents which increases financial responsibility

• Higher interest rate (15.0%) increases approval probability by approximately 50.0% in odds.

Odds = P / (1 - P)

Summary: Your approval is mainly due to your good credit history, though the loan amount is high relative to income.

Bank Settings

Interest Rate (%) ⊘

15.00

8.00   24.00

Result Analysis

● **Confusion Matrix Analysis:** The confusion matrix showed that the model had high precision (0.94) for predicting non-eligible applicants, but with low recall (0.31), indicating it missed many true negatives. In contrast, it achieved excellent recall (0.99) for eligible customers, ensuring minimal rejection of genuinely eligible applicants.

● **Model Accuracy Comparison:** Logistic Regression outperformed other classifiers with an accuracy of 97%, followed by Naive Bayes (95.92%) and Gradient Boosting (93.84%). SVM and K-NN underperformed with lower accuracy scores (71.89% and 61.08% respectively), suggesting their limitations in handling this dataset.

● **F1-Score and Class Imbalance Impact**: The weighted F1-score of 0.76 indicated balanced overall performance, but the macro-average F1-score (0.67) reflected class imbalance effects. Models tended to favor class '1' (loan approved), leading to higher false positives for class '0'.

Conclusion

The results demonstrate that machine learning techniques can significantly enhance the loan approval process by reducing manual effort, improving accuracy, and minimizing the risk associated with loan defaults. Our approach enables banks and financial institutions to make informed, data-driven decisions, ensuring better financial stability and customer segmentation.

Future enhancements could include integrating real-time data streams, optimizing models using advanced deep learning techniques, and expanding the dataset to incorporate additional features for better predictions. This system serves as a foundational step towards automated, intelligent loan approval systems that align with modern financial needs

References

[1]   Kumar Arun, Garg Ishan, Kaur Sanmeet, ― Loan Approval Prediction based on Machine Learning Approach‖, IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 18, Issue 3, pp. 79-81, Ver. I (May-Jun. 2016)

[2]   E. Chandra Blessie, R. Rekha, ― Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process (2019)

[3]   M. A. Sheikh, A. K. Goel and T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 490-494, doi: 10.1109/ICESC48915.2020.9155614.

Publications (if any)