Loan Default Status Prediction System using Machine Learning

Ansh Joshi, Bhavika Darpe, Bhavik Mundra

Computer Science and Engineering Acropolis Institute of Technology and Research
anshsjoshi07@gmail.com , bhavik.mundra1603@gmail.com bhavika.darpe@gmail.com
DOI: https://doi.org/10.55248/gengpi.5.0524.1134

ABSTRACT

By analyzing extensive historical loan records and employing state-of-the-art machine learning algorithms, this initiative seeks to construct a sophisticated predictive model capable of accurately determining loan approval outcomes based on a diverse array of applicant characteristics. The primary aim is to streamline decision-making processes, reduce processing timelines, bolster risk evaluation strategies, and advocate for responsible lending practices, thereby enhancing overall customer satisfaction and contributing to sustained economic prosperity and stability. Additionally, it also aims to identify patterns in loan applications to optimize resource allocation and minimize potential defaults. By harnessing the power of data-driven insights, the project endeavors to create a more equitable and efficient lending ecosystem.

I. INTRODUCTION

In the contemporary landscape of banking, lending serves as a pivotal revenue stream for financial institutions. Yet, the prospect of borrowers defaulting on loans presents a significant risk. To mitigate this challenge, banks are embracing machine learning (ML) methodologies to refine their credit risk evaluation frameworks. This endeavor seeks to construct a resilient ML model adept at categorizing loan applicants as probable defaulters or non-defaulters, drawing insights from extensive historical loan data. Through the analysis of diverse applicant attributes and historical loan performance data, the model aims to provide accurate predictions, enabling financial institutions to make informed lending decisions. Ultimately, the implementation of such predictive systems not only enhances efficiency but also fosters greater financial stability within the banking sector. By minimizing the incidence of defaults, banks can better allocate resources and extend credit to deserving individuals and businesses, thereby facilitating economic growth and prosperity.

II. PROPOSED WORK

For the purpose of predicting loan status, many approaches may be applied. However, machine learning-based classification is a widely utilized strategy. Here is fundamental process for creating a machine learning model for loan status prediction:

1. Data Collection and Preparation: Gather a comprehensive dataset of information which is required to be provided by the user. We have taken a dataset which consists of columns like credit score, age, loan type, income, co-applicant income, number of dependents etc. The dataset overall consists of 18 columns and approx. 10000 records varying from numerical, floating to categorical values.

2. Preprocessing Data: The preprocessing step includes identifying missing values and outliers if present. The missing values are replaced with mean or mode of the remaining values in the columns. Also unwanted columns are removed if those don't hold any correlation with the target variable.

3. Convert the categorical values into numerical values so that these can be fed into ML model. The conversion is done by the encoding methods like label encoder or one hot encoder. Through the encoding techniques, the categorical values are mapped to numerical values.

4. Split the dataset into test dataset and training dataset respectively.

5. Model Training: The training dataset is used to train the machine learning model by importing scikit-learn library. Through scikit-learn, we can implement various machine learning models for regression, classification, clustering, and statistical tools for analyzing these models.

6. Model Evaluation: Evaluate the trained model's performance using test dataset. To evaluate the model's performance, find the accuracy and precision of the test dataset in comparison to training dataset.

7. Deployment: To be used for loan status prediction, import the model to enable it to integrate with the frontend and deploy it.

III. EXISTING SYSTEMS

1) FICO Score: Widely used credit scoring system assessing credit risk based on factors like payment history, credit utilization, and length of credit history. Scores range from 300 to 850, with higher scores indicating lower risk.

2) Manual Loan Approval: loan applications are manually reviewed by loan officers who assess various factors such as credit history, income, debt-to-income ratio, and collateral to determine the applicant's eligibility for a loan.

IV. RESEARCH METHODOLOGY

Loan status prediction is a system which aims to predict whether an applicant can be granted the loan or not based on the credentials of the applicant provided. The following are key components of this project:

- Data collection: This involves collecting a large dataset which includes columns like credit score, debt to income ratio, income, debts etc. The dataset is used such that it contains labelled classes.

- Data Preprocessing: This involves removing irrelevant columns and converting categorial values to numerical values.

- Model training: The ML model is trained on several classification algorithms like random forest, decision tree, logistic regression and selecting the one with the most accuracy.

- Model evaluation: The trained model will be evaluated on a test set of data to determine its accuracy and effectiveness in predicting the loan status.

- Integration: The model will be integrated with frontend and backend of the web application to give the desired output.

- User interface: The user interface will be provided to users to upload details of the user and accordingly the result will be displayed as approved or not.

The end goal of the Loan Status Prediction project is to build a robust, accurate, and efficient system to enable the banks to know the eligible applicants so that they don't grant loans to the ones that can't repay the loan.

Entity Relationship Diagram:
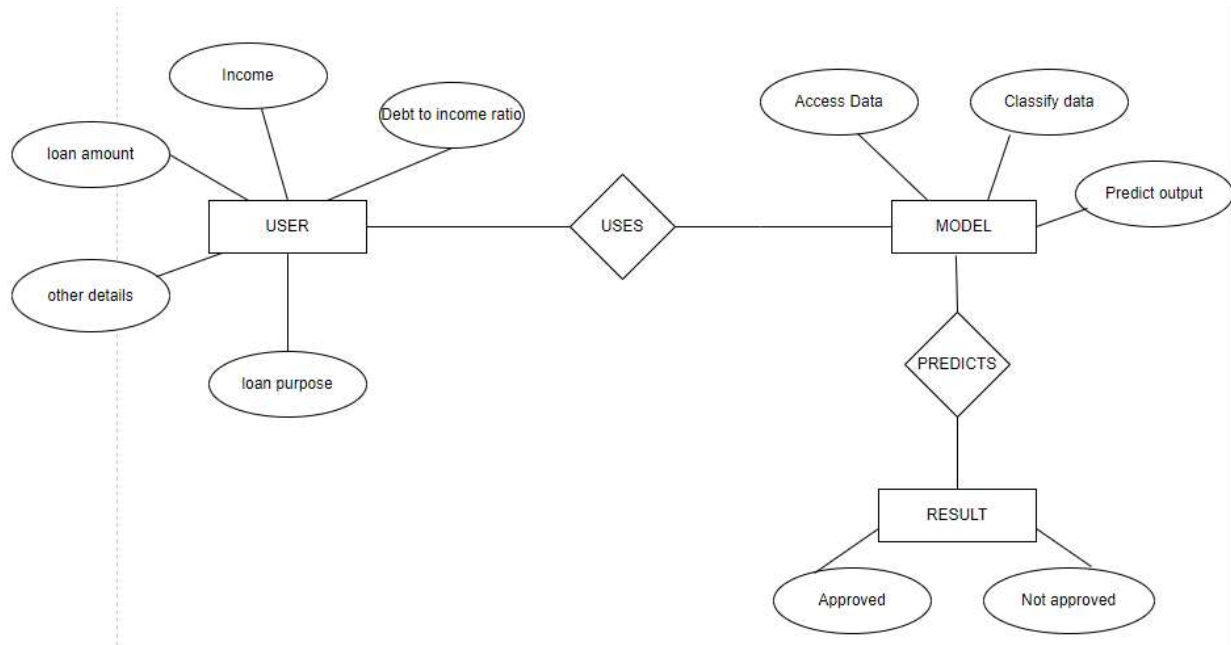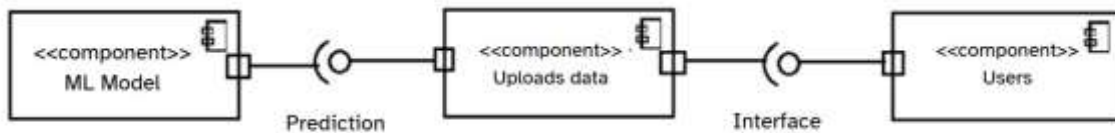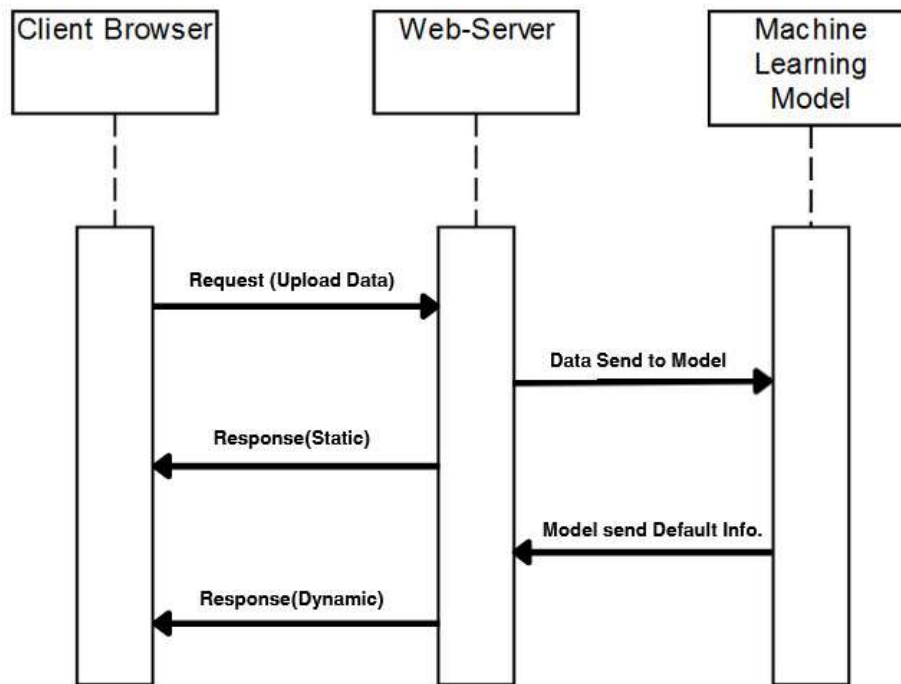
V. CONCLUSION

In summary, the research undertaken in "Loan Status Prediction using Machine Learning" holds significant promise for transforming the loan approval landscape in the financial sector. Through the utilization of historical loan data and cutting-edge machine learning methodologies, the project endeavors to construct a predictive framework adept at reliably determining loan approval outcomes based on diverse applicant characteristics. This endeavor marks a pivotal step towards enhancing the efficiency and accuracy of loan approval processes, thereby offering substantial benefits to both financial institutions and loan applicants alike.

VI. REFERENCE

1. Cyert, R. M., Davidson, H. J., & Thompson, G. L. (1962). search 6, 180–195. Estimation of allowance for doubtful accounts by Markov chains. Management Science 8, 287–303.

2. Safavian, S. R., & Landgrebe, D. (1991). A survey decision tree classifier methodology. IEEE Transactions on Systems, Man and Cybernetics 21, 660–674.

3. Orgler, Y. E. (1971). Evaluation of bank consumer loans with credit scoring models. Journal of Bank Research 1(Spring), 31–37.