LOAN INTEREST RATE PREDICTION

[1] T Manasa, [2] Arrolla Vamshi, [3] Erumalla Abilash, [4] A Tushar Singh
[1]AssistantProfessor, [234]Students
Department of Computer Engineering(Software Engineering)
Siddhartha Institute of Technology & Sciences, Narapally
thirumanasa@siddhartha.org.in, 23tq1a5619@siddhartha.co.in, 23tq1a5609@siddhartha.co.in, 23tq1a5652@siddhartha.co.in

Abstract

Loan interest rate prediction plays a pivotal role in the financial sector by enabling banks and lending institutions to optimize their loan offerings and mitigate risks. Accurate forecasting of loan interest rates ensures that financial institutions can make informed decisions while maintaining competitiveness in the market. This study explores the development of a predictive model for loan interest rates, leveraging machine learning techniques to identify key factors influencing interest rate determination. By analyzing historical loan data, including borrower characteristics such as credit scores, loan amounts, and repayment durations, as well as external economic indicators, the model predicts the most appropriate interest rate for new loan applicants. Several machine learning algorithms, including linear regression, decision trees, and random forests, are compared for their efficacy in modeling complex relationships between the input variables and the interest rates. The performance of the models is evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared to determine their accuracy and robustness. The results demonstrate that machine learning-based models outperform traditional statistical methods, providing more accurate and dynamic predictions that can adapt to market fluctuations. This research underscores the importance of data-driven approaches in financial decision-making, offering a robust framework for integrating machine learning into loan interest rate forecasting. Future directions include real-time data integration, deep learning models, and further model refinement to improve prediction accuracy for both lenders and borrowers.

I. Introduction

Loan interest rate prediction is a critical component in the financial sector, as it directly impacts both lenders and borrowers. The interest rate is a key factor that determines the cost of borrowing and the overall financial health of a loan portfolio. For financial institutions, accurately predicting interest rates enables them to balance profitability with risk management. An incorrect interest rate can either drive away potential borrowers or lead to financial losses due to missed opportunities in a competitive market. For borrowers, an accurate and fair interest rate ensures that they are not overburdened by excessive costs, fostering trust between lenders and clients. This predictive modeling approach leverages historical loan data, economic indicators, and borrower-specific characteristics to determine the most appropriate interest rate for each loan application.Traditionally, banks and financial institutions have relied on manual processes, heuristics, and regulatory guidelines to set interest rates. However, these methods are often inefficient and prone to human error. As technology has advanced, especially in the field of data science and machine learning, there has been a shift toward more automated and data-driven approaches. Predicting interest rates through machine learning not only increases efficiency but also enhances accuracy by considering a broader range of factors, including economic conditions, borrower

behavior, and financial market trends. The introduction of advanced algorithms, such as decision trees, support vector machines, and neural networks, has revolutionized this process, allowing for dynamic and real-time predictions that can be tailored to individual borrowers.

II. Literature Survey

1. Uddin, N. et al. [2023]
This paper addresses the limitations of traditional loan approval processes, which often struggle with accuracy and efficiency in predicting loan eligibility. The existing methods, such as Logistic Regression and Decision Trees, fail to provide the desired performance when handling large datasets with complex relationships. The authors propose an ensemble machine learning model that integrates various classifiers such as Extra Trees, AdaBoost, and Gradient Boosting.

2. Salem, A. A. et al. [2024]
This research highlights the challenges faced by traditional statistical models in accurately forecasting interest rates, particularly their inability to capture complex, non-linear relationships in financial data. The authors explore the limitations of conventional models like linear regression and support vector machines. Their proposed approach incorporates deep learning models to improve the accuracy of interest rate forecasts.

3. Monje, L. et al. [2025]
One of the main challenges identified in this study is the lack of interpretability in advanced machine learning models used for loan default prediction. Traditional methods like logistic regression and probit models provide simpler but less accurate predictions. The authors propose an approach that combines eXtreme Gradient Boosting (XGBoost) with linguistic explainability through surrogate models and the 2-tuple fuzzy linguistic model to ensure transparency and interpretability. The dataset used for this research is sourced from Lending Club, one of the largest peer-to-peer (P2P) lending platforms. Machine learning techniques such as XGBoost, along with fuzzy linguistic models for interpretability, are applied. The tools include Python and XAI frameworks. The study uses AUC, accuracy, and F1-score as validation metrics.

4. Kouser, R. et al. [2024]
The primary challenge discussed in this paper is the difficulty in accurately predicting both loan status and interest rates. Traditional credit scoring models are insufficient for accurately predicting the loan status due to their reliance on linear models and limited feature sets. The authors propose a new methodology that applies various machine learning models, including Random Forest, SVM, and Logistic Regression, to predict loan status and compute the interest rate. The dataset used for this research is derived from the Lending Club platform, which includes a variety of borrower information. Techniques such as Random Forest and Support Vector Machine are applied for prediction tasks. The tools used are Python, Scikit-learn, and Pandas. Accuracy, precision, recall, and F1-score are used as evaluation metrics.

5. Zhang, X. et al. [2025]
This study focuses on the issue of imbalanced datasets, which often lead to poor model performance when predicting loan defaults. Traditional statistical methods fail to address the challenges posed by imbalanced datasets, where the majority class (non-default) dominates. The authors propose a machine learning approach

using Random Forest, XGBoost, and LightGBM, combined with techniques like SMOTE to handle class imbalance.

III. System Analysis

The Loan Interest Rate Prediction system is designed to analyze and predict appropriate interest rates for loan applicants based on various financial and personal attributes. The system takes into account multiple input parameters such as applicant income, credit history, employment status, loan amount, loan term, and existing liabilities. These factors are carefully examined to understand their influence on the interest rate assigned to a borrower. The system uses historical loan data to identify patterns and relationships between borrower profiles and previously assigned interest rates.

In the analysis phase, data is collected from reliable financial datasets and undergoes preprocessing steps such as handling missing values, encoding categorical variables, and normalizing numerical features. Exploratory Data Analysis (EDA) is performed to identify correlations and trends among variables. Based on these insights, relevant features are selected to build an efficient prediction model. Machine learning algorithms such as Linear Regression, Decision Trees, or Random Forest are applied to model the relationship between input features and interest rates.

Existing System

The existing system for determining loan interest rates is largely based on traditional banking practices and manual evaluation processes. Financial institutions typically rely on predefined rules, basic statistical methods, and the judgment of loan officers to assign interest rates to borrowers. These decisions are primarily influenced by limited factors such as credit score, income level, employment status, and repayment history. The process often involves reviewing documents submitted by applicants and verifying their financial background through standard procedures.

In many cases, the system lacks advanced data analysis capabilities and does not fully utilize historical data to identify deeper patterns or relationships between borrower characteristics and interest rates. As a result, the decision-making process can be time-consuming and inconsistent, depending on the experience and discretion of the loan officer. Moreover, traditional systems may not adapt quickly to changing economic conditions or customer behavior, leading to less accurate interest rate assignments.

Disadvantages of Existing System

· **Manual and Time-Consuming Process:**
The traditional system relies heavily on manual verification and decision-making, which increases processing time and delays loan approvals.
· **Limited Data Utilization:**
Only a few parameters like income, credit score, and employment status are considered, ignoring many other important factors that could improve prediction accuracy.

· **Human Bias and Errors:**
Decisions depend on loan officers, which may introduce personal bias, inconsistency, and human errors in assigning interest rates.
· **Lack of Accuracy:**
Without advanced analytical models, the system may assign inaccurate interest rates that do not truly reflect the borrower's risk level.
· **Inflexibility to Changing Conditions:**
The system uses fixed rules and does not adapt quickly to economic changes, market trends, or evolving customer behaviors.

Proposed System

The proposed system introduces an automated and intelligent approach for predicting loan interest rates using machine learning techniques. Unlike traditional methods, this system leverages historical loan data and advanced algorithms to accurately determine interest rates based on multiple borrower attributes. Key input features include applicant income, credit score, employment status, loan amount, loan tenure, existing debts, and repayment history. These inputs are processed through a data-driven pipeline to ensure precise and unbiased predictions.

The system begins with data collection and preprocessing, where missing values are handled, categorical variables are encoded, and numerical features are normalized. Exploratory Data Analysis (EDA) is performed to identify important patterns and correlations among variables. Based on this analysis, relevant features are selected to train the machine learning model.

Advantages of Proposed System

· **High Accuracy:**
The use of machine learning algorithms enables precise prediction of interest rates by analyzing complex relationships between multiple factors.
· **Automation:**
The system automates the entire process, reducing manual effort and speeding up loan approval and interest rate assignment.
· **Reduced Human Bias:**
Decisions are data-driven, minimizing the chances of personal bias and ensuring fair interest rate allocation.
· **Efficient Decision-Making:**
Instant predictions improve the speed and efficiency of the decision-making process in financial institutions.
· **Better Risk Assessment:**
The system evaluates multiple variables simultaneously, providing a more accurate assessment of borrower risk.

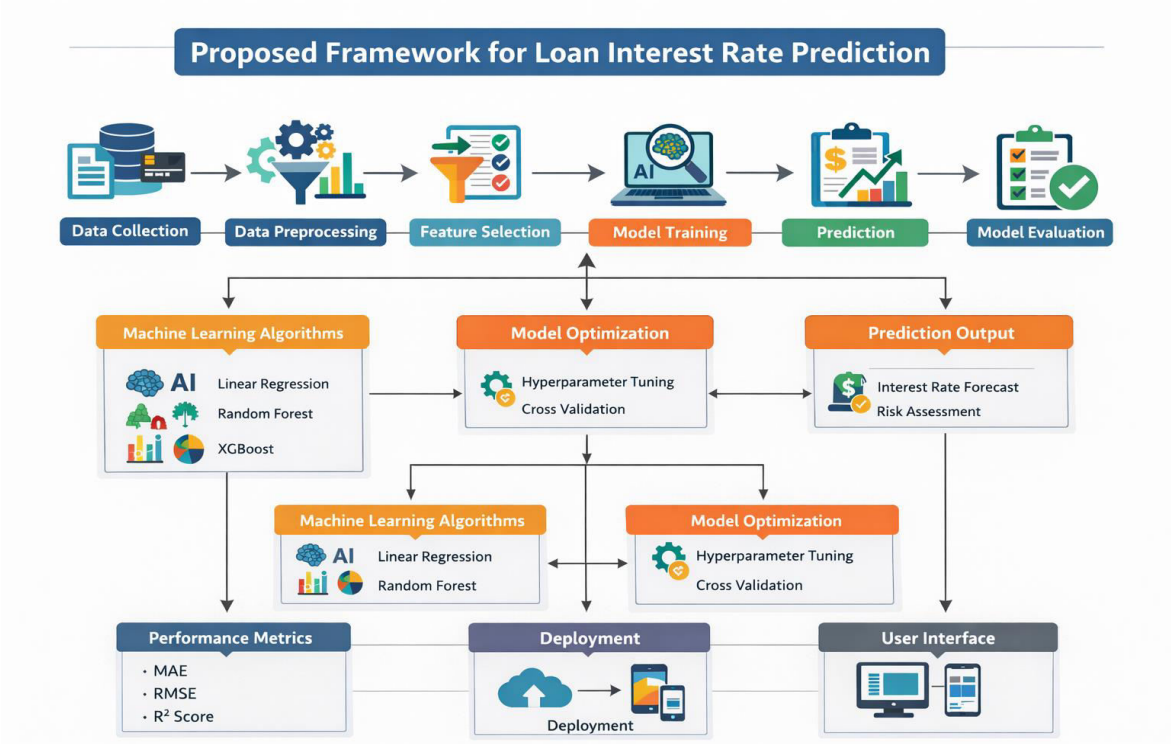
IV. Methodology

The methodology for the Loan Interest Rate Prediction system follows a structured machine learning pipeline to ensure accurate and reliable predictions. Initially, relevant historical loan data is collected from financial datasets, which include attributes such as income, credit score, employment status, loan amount, and

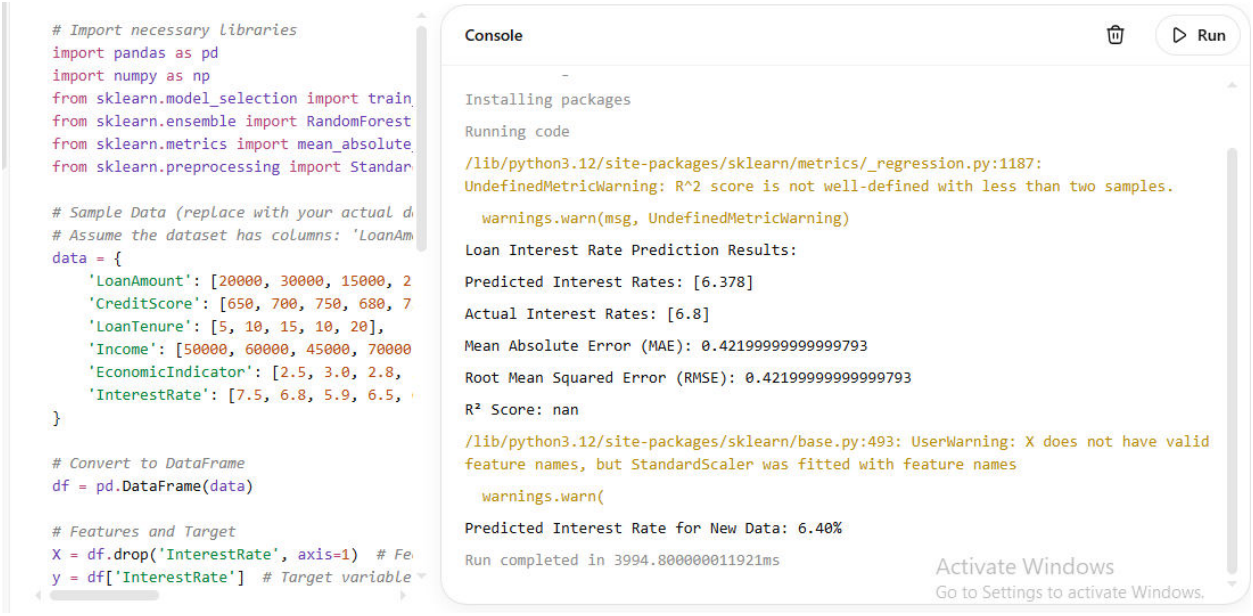
repayment history. The collected data undergoes preprocessing, where missing values are handled, outliers are treated, and categorical variables are encoded into numerical formats. Feature scaling techniques such as normalization or standardization are applied to improve model performance.
Next, Exploratory Data Analysis (EDA) is conducted to identify patterns, correlations, and trends among variables that influence interest rates. Based on these insights, important features are selected to reduce dimensionality and enhance model efficiency. The processed dataset is then divided into training and testing sets.

System Architecture

The architecture of the Loan Interest Rate Prediction system is designed to efficiently handle the data flow from input collection to prediction output. At the core of the system is the Prediction Engine, which is responsible for processing input data, applying machine learning models, and generating the predicted loan interest rate. The system starts with the User Interface (UI), where financial institutions input data related to the loan applicant, such as credit score, loan amount, repayment period, and relevant economic factors. Once the user inputs the data, it is passed to the Data Preprocessing Module, which handles tasks such as data cleaning, feature extraction, normalization, and handling missing values. This ensures that only high-quality, relevant data is fed into the prediction engine.The Prediction Engine utilizes machine learning algorithms trained on historical loan data to generate the interest rate prediction. Various models such as linear regression, random forests, or neural networks are used to calculate the predicted loan interest rate. The system also includes a Model Evaluation Module to assess the performance of the chosen model using validation metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared. The predicted rate is then displayed on the Output Layer, where financial institutions can review the results. The architecture is designed to be modular and scalable, enabling future upgrades like integrating real-time economic data, adding new features, or improving prediction models.



V. Result and Output



```
# Import necessary libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train
from sklearn.ensemble import RandomForest
from sklearn.metrics import mean_absolute
from sklearn.preprocessing import Standar

# Sample Data (replace with your actual d
# Assume the dataset has columns: 'LoanAm
data = {
    'LoanAmount': [20000, 30000, 15000, 2
    'CreditScore': [650, 700, 750, 680, 7
    'LoanTenure': [5, 10, 15, 10, 20],
    'Income': [50000, 60000, 45000, 70000
    'EconomicIndicator': [2.5, 3.0, 2.8,
    'InterestRate': [7.5, 6.8, 5.9, 6.5,
}

# Convert to DataFrame
df = pd.DataFrame(data)

# Features and Target
X = df.drop('InterestRate', axis=1)  # Fe
y = df['InterestRate']  # Target variable
```

```
Console                                                    Run

Installing packages
Running code
/lib/python3.12/site-packages/sklearn/metrics/_regression.py:1187:
UndefinedMetricWarning: R^2 score is not well-defined with less than two samples.
  warnings.warn(msg, UndefinedMetricWarning)
Loan Interest Rate Prediction Results:
Predicted Interest Rates: [6.378]
Actual Interest Rates: [6.8]
Mean Absolute Error (MAE): 0.42199999999999793
Root Mean Squared Error (RMSE): 0.42199999999999793
R² Score: nan
/lib/python3.12/site-packages/sklearn/base.py:493: UserWarning: X does not have valid
feature names, but StandardScaler was fitted with feature names
  warnings.warn(
Predicted Interest Rate for New Data: 6.40%
Run completed in 3994.800000011921ms
```


VI. Conclusion

The Loan Interest Rate Prediction model showcases the transformative potential of machine learning in the financial sector by automating the complex task of predicting loan interest rates. Traditional methods of determining interest rates often rely on static rules and subjective judgment, which can lead to inconsistencies and inefficiencies. By utilizing historical data, machine learning algorithms such as Random Forest, XGBoost, and Linear Regression, this system can analyze multiple variables and predict loan interest rates more accurately. It not only improves prediction reliability but also helps financial institutions make data-driven decisions, reducing the risk of errors that may occur with manual rate-setting processes.This project underscores the growing significance of predictive analytics in the finance industry, offering a more dynamic and accurate alternative to traditional methods. The use of machine learning opens up opportunities for continuous improvement and adaptation to changing market conditions, allowing financial institutions to stay competitive. The model's ability to process large datasets and generate insights in real time paves the way for future enhancements, such as incorporating real-time economic data and using more advanced algorithms like deep learning. As the system evolves, it will become an essential tool for optimizing lending strategies and ensuring fairer, more efficient loan pricing.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The

International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.