A MACHINE LEARNING APPLICATION FOR PREDICTION LOAN DEFAULT BASED ON CONSUMER BEHAVIOR

TEAM MEMBERS

VISHVA A 210701314

VINOTH RAJ G 210701509

PROJECT COORDINATOR

Karthick V,

Associate Professor,

Department of Computer Science and Engineering

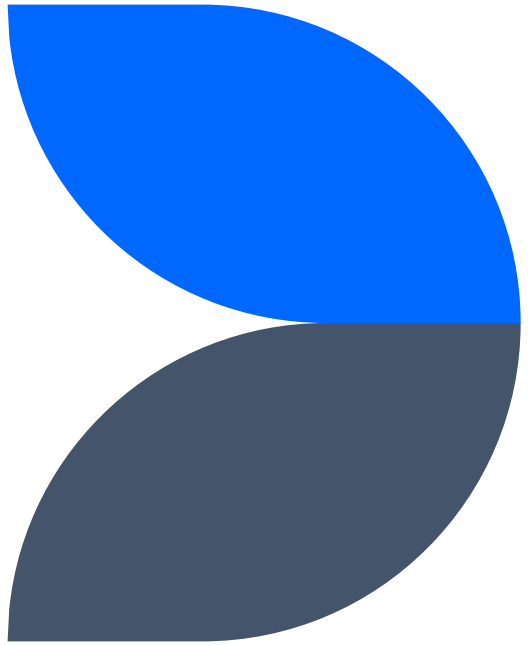
Agenda

Abstract
Introduction
Literature Review
Research Gaps
Proposed Methodology
Results and Discussion
Comparative Analysis
Conclusion and Future Work
References

Abstract

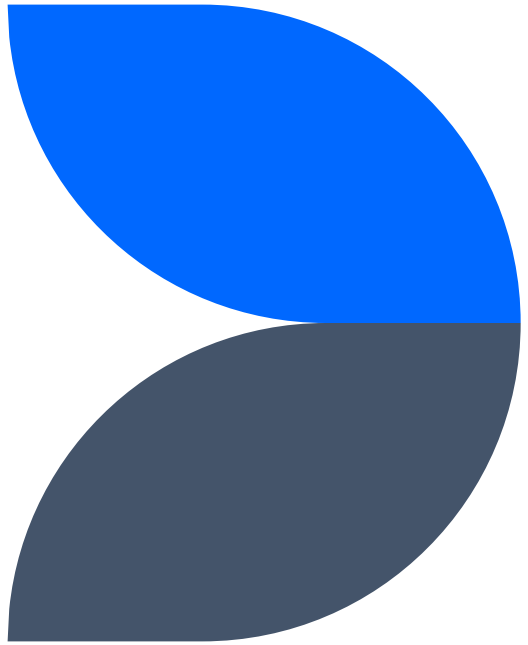
This project presents a Loan Prediction System utilizing machine learning to evaluate and predict loan eligibility for applicants. By leveraging historical loan data, the system employs algorithms like Logistic Regression, Decision Trees, and Random Forests to analyze applicant information such as income, credit history, and employment details. The model is trained and validated to ensure high accuracy and reliability. This automated approach streamlines the loan approval process, reduces human error, and enhances decision-making efficiency for financial institutions, providing a robust tool for assessing credit risk and managing loan portfolios effectively.

Keywords: Machine Learning, Loan Approval Prediction, Algorithms, Random Forest, Logistic Regression, K Nearest Neighbor,



Introduction

A Loan Prediction System using Machine Learning leverages advanced algorithms to predict the likelihood of loan approval for applicants. By analyzing historical data and key applicant features such as credit score, income, and employment history, the system enhances decision-making processes for financial institutions. This approach not only increases accuracy and efficiency but also reduces human bias and processing time. Implementing such a system can lead to improved customer satisfaction and better risk management, ultimately contributing to more robust financial operations and profitability for lenders.



Literature Review

Paper Title	Advantages	Disadvantages	Outcomes
Loan Approval Prediction Using Machine Learning Algorithms	High accuracy, reduces human error, quick processing	Requires high-quality data, complex algorithms	Improved loan approval accuracy and reduced processing time
Predictive Analytics for Loan Default Prediction	Early identification of potential defaulters, improved risk management	Possible overfitting, dependency on historical data	Enhanced ability to mitigate risks associated with loan defaults
A Comparative Study of Machine Learning Techniques for Loan Default Prediction	Comparison of multiple algorithms, identification of the most effective models	Requires extensive computational resources, data preprocessing	Identification of the most accurate and efficient algorithm for loan default prediction

Literature Review

Paper Title	Advantages	Disadvantages	Outcomes
Machine Learning Techniques for Credit Risk Evaluation	Improved credit risk assessment, better decision-making for lenders	Data privacy concerns, need for continuous model updates	More reliable credit risk evaluations and better informed lending decisions
Loan Eligibility Prediction System Using Ensemble Learning Techniques	High prediction accuracy, robustness due to ensemble methods	Computationally intensive, may require complex implementation	Increased accuracy and reliability in predicting loan eligibility
Application of Deep Learning in Loan Default Prediction	Ability to capture complex patterns in data, high predictive performance	Requires large datasets, longer training times	Enhanced predictive performance in identifying loan defaults
Predicting Loan Approval using Decision Trees	Easy to interpret, fast and efficient	Prone to overfitting, less effective with noisy data	Quick and interpretable loan approval predictions

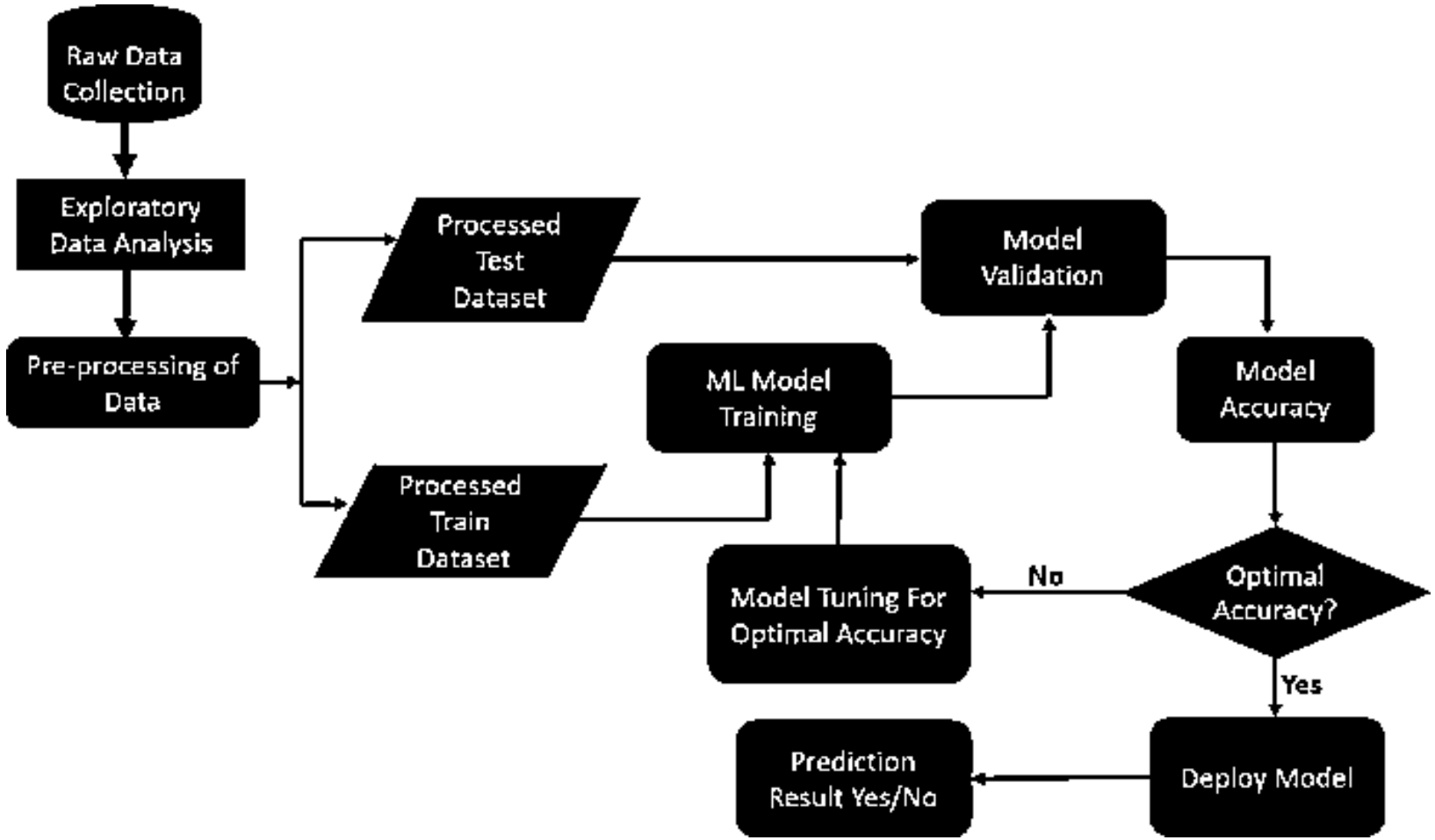
Literature Review

Paper Title	Advantages	Disadvantages	Outcomes
Using Random Forest Algorithm for Loan Approval Prediction	High accuracy, handles large datasets well, robustness to overfitting	Interpretability issues, need for parameter tuning	Improved loan approval prediction accuracy with robustness against overfitting
Credit Scoring and Loan Default Prediction Using SVM	High accuracy in classification, effective with small to medium-sized datasets	Sensitive to parameter settings, can be computationally expensive	Effective credit scoring and loan default prediction with high classification accuracy
Neural Network Approaches for Predicting Loan Defaults	Ability to model complex relationships, high accuracy	Requires large datasets, black-box nature leading to lack of interpretability	High accuracy in predicting loan defaults, capturing complex non-linear relationships in data

Research Gaps

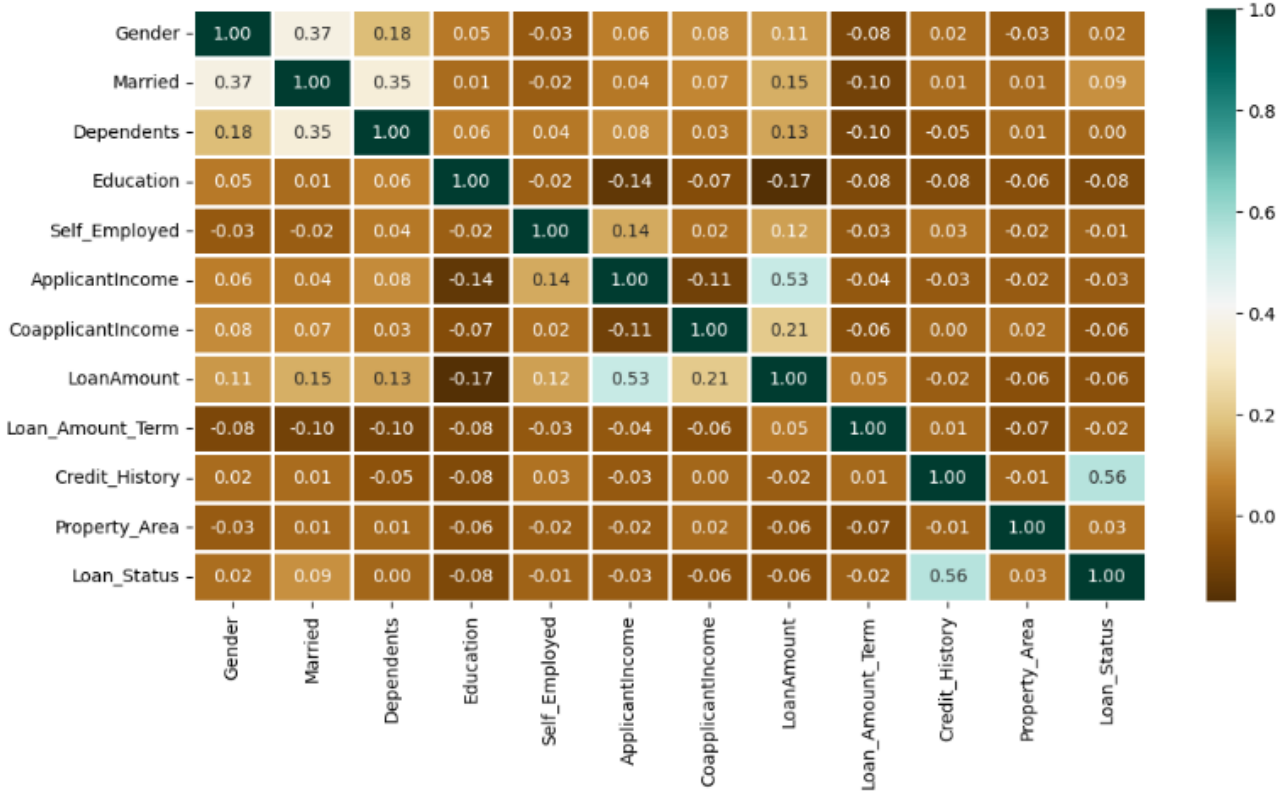
- **Data Quality and Availability:** Many studies rely on high-quality, extensive datasets, which are often not available in real-world scenarios, leading to a gap in the applicability of these models.
- **Model Interpretability:** Advanced machine learning models, such as neural networks and ensemble methods, often lack interpretability, making it difficult for financial institutions to understand and trust the predictions.
- **Handling Imbalanced Data:** Loan datasets are often imbalanced, with fewer defaults than approvals. Many existing models struggle to effectively handle this imbalance, leading to biased predictions.
- **Generalization Across Regions:** Models trained on data from specific regions or financial institutions may not generalize well to other regions with different economic conditions and borrower behaviors.
- **Dynamic Economic Factors:** Existing models often do not account for dynamic economic factors and market conditions, which can significantly impact loan repayment behaviors.
- **Ethical and Bias Concerns:** There is a need for more research into ensuring that machine learning models do not perpetuate existing biases in lending practices, which can lead to unfair treatment of certain demographic groups.

Proposed Methodology



Results And Discussion

The above heatmap is showing the correlation between Loan Amount and ApplicantIncome. It also shows that Credit_History has a high impact on Loan_Status.



Conclusion and Future Work

The table below provides the accuracy score for training and testing prediction.

Algorithm	Training Accuracy	Testing Accuracy
RandomForestClassifier	98	82.5
KNeighborsClassifier	78	63.7
SVC	68	69.16
LogisticsRegression	79	80.0

References

[1] Vaidya and Ashlesha, Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval, 2021 13th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2021.

[2] Amin, Rafik Khairul and Yuliant Sibaroni, Implementation of decision tree using C4. 5 algorithm in decision making of loan application by debtor (Case study: Bank pasargad of Yogyakarta Special Region), 2022 8rd International Conference on Information and Communication Technology (ICoICT). IEEE, 2022.

[3] Arora, Nisha and Pankaj Deep Kaur, A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment, Applied Soft Computing 86 (2020), 105936.

[4] Yang, Baoan, et al, An early warning system for loan risk assessment using artificial neural networks, Knowledge-Based Systems 14.5-6 (2024), 303-306.

[5] Metawa, Noura, M. Kabir Hassan and Mohamed Elhoseny, Genetic algorithm based model for optimizing bank lending decisions, Expert Systems with Applications 80 (2022), 75-82.

References

[6] Hassan, Amira Kamil Ibrahim and Ajith Abraham. "Modeling consumer loan default prediction using ensemble neural networks, 2020 International Conference On Computing, Electrical And Electronic Engineering (ICCEEE). IEEE, 2020.

[7] X.Frencis Jensy, V.P.Sumathi, Janani Shiva Shri, “An exploratory Data Analysis for Loan Prediction based on nature of clients”, International Journal of Recent Technology and Engineering (IJRTE),Volume-7 Issue-4S, November 2018.

[8] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma,Namburi Vimala Kumari, k Vikash,“Loan Prediction by using Machine Learning Models”, International Journal of Engineering and Techniques.Volume 5 Issue 2, Mar-Apr 2023

[9] 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-June. 2021).

References

[10] Aboobyda Jafar Hamid and Tarig Mohammed Ahmed, —Developing Prediction Model of Loan Risk in Banks using Data Mining‖, Machine Learning and Applications: An International Journal (MLAIJ), Vol.3, No.1, pp. 1-9, March 2023.

[11] S. Vimala, K.C. Sharmili, —Prediction of Loan Risk using NB and Support Vector Machine‖, International Conference on Advancements in Computing Technologies (ICACT 2022), vol. 4, no. 2, pp. 110-113, 2022.

[12] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, kVikash,“Loan Prediction by using Machine Learning Models”, International Journal of Engineering and Techniques.Volume 5 Issue 2, Mar-Apr 2023

[13] Nikhil Madane, Siddharth Nanda,”Loan Prediction using Decision tree”, Journal of the Gujarat Research History, Volume 21 Issue 14s, December 2022.

Thank you