

# LOAN APPROVAL PREDICTION

<sup>1</sup> T Manasa, <sup>2</sup> D Prashanth, <sup>3</sup> B Sanjanasri, <sup>4</sup> J Yugendar

<sup>1</sup>AssistantProfessor, <sup>2,3,4</sup>Students

Department of Computer Engineering(Software Engineering)

Siddhartha Institute of Technology & Sciences, Narapally

[thirumanasa@siddhartha.org.in](mailto:thirumanasa@siddhartha.org.in), [23tq1a5643@siddhartha.co.in](mailto:23tq1a5643@siddhartha.co.in), [23tq1a5650@siddhartha.co.in](mailto:23tq1a5650@siddhartha.co.in),  
[23tq1a5604@siddhartha.co.in](mailto:23tq1a5604@siddhartha.co.in),

## Abstract

The loan approval process is a crucial function in financial institutions, where accurate and timely decisions are essential. Traditionally, loan applications are evaluated manually, which can be time-consuming, inconsistent, and influenced by human bias. This project presents a machine learning-based approach to automate the loan eligibility prediction process, improving efficiency and decision-making.

The model predicts whether a loan applicant is eligible for a loan based on various demographic and financial attributes such as gender, marital status, education, applicant income, credit history, and loan amount. The dataset used in this study consists of 614 records with 13 features obtained from a publicly available source.

Data preprocessing techniques were applied, including handling missing values, encoding categorical variables, and feature engineering. A new feature, Total Income, was created by combining applicant and co-applicant incomes to enhance predictive performance. Exploratory Data Analysis (EDA) was conducted using visualization tools to identify patterns and relationships in the data.

## I. Introduction

Financial institutions, including banks and non-banking financial companies (NBFCs), play a vital role in providing loans to individuals and businesses. However, the loan approval process remains a significant challenge due to its complexity and the need for accurate risk assessment. Traditionally, this process involves manual evaluation of an applicant's financial history, credit score, employment status, and other socio-economic factors.

Such manual processes are time-consuming, resource-intensive, and often prone to human errors and inconsistencies. Additionally, subjective judgment and bias may influence decisions, leading to unfair or inaccurate loan approvals or rejections. As the number of loan applications continues to increase, financial institutions require efficient and reliable systems to handle large volumes of data and make consistent decisions.

The core problem lies in determining whether a loan applicant is creditworthy based on multiple factors. This requires analyzing large datasets and identifying patterns that indicate the likelihood of loan repayment. Traditional methods are not sufficient to handle this complexity effectively. With advancements in technology, machine learning has emerged as a powerful solution to automate and improve the loan approval process. Machine learning algorithms can analyze historical data, identify patterns, and make predictions with high accuracy. These models help in assessing loan eligibility in a faster, more consistent, and objective manner.

## II. Literature Survey

A review of existing literature shows a strong trend toward the use of machine learning techniques for credit risk assessment and loan approval prediction. Researchers have explored various algorithms to improve prediction accuracy, reduce risk, and automate the decision-making process in financial institutions.

Author(s)	Year	Methodology Algorithm(s)	Key Findings
Vaidya, A.	2020	Logistic Regression	Highlighted the critical importance of credit history as a major factor influencing loan approval decisions.
Amin et al.	—	C4.5 Decision Tree	Achieved high precision (78%) and recall (96.4%) in predicting loan status using decision tree models.
Arora & Kaur	—	Bolasso (feature selection), Random Forest	Demonstrated that Random Forest performs better when combined with effective feature selection techniques.
Yang et al.	—	Artificial Neural Network (ANN)	Showed that ANN can be used to develop early warning systems for loan risk prediction.

Recent studies consistently emphasize the effectiveness of ensemble learning methods such as Random Forest and XGBoost in handling complex and non-linear relationships in financial data. These models, when combined with proper data preprocessing and feature engineering, significantly improve prediction accuracy.

Additionally, feature selection techniques play a vital role in enhancing model performance by identifying the most relevant attributes, such as credit history, income, and loan amount. Research indicates that optimized machine learning models can achieve accuracy levels ranging from 83% to 99%, depending on the dataset and methodology used.

## III. System Analysis

System analysis focuses on understanding the requirements and workflow of the loan approval prediction system. The system takes input features such as gender, marital status, education, applicant income, co-applicant income, credit history, and loan amount. These inputs are processed through data preprocessing steps including handling missing values, encoding categorical variables, and feature scaling. Feature engineering is applied to create new attributes like Total Income. Exploratory Data Analysis (EDA) is performed to identify patterns and relationships in the dataset. The processed data is then used to train a machine learning model. The system evaluates

different algorithms and selects the best-performing model, such as Random Forest. The trained model predicts whether a loan should be approved or not.

### **Existing System**

The existing loan approval system primarily relies on manual evaluation by financial officers. Loan applications are reviewed based on documents such as income statements, credit reports, and employment details. This process is time-consuming and requires significant human effort. Decisions often depend on the experience and judgment of loan officers. The system may involve paper-based or semi-digital processes, which are inefficient. It is difficult to handle a large number of applications quickly. The process is also prone to human errors and inconsistencies. Bias may influence decision-making in some cases. Traditional systems lack automation and advanced data analysis capabilities. As a result, loan approval decisions may not always be accurate or consistent.

### **Disadvantages of Existing System**

- Time-consuming and slow processing of applications
- High dependency on manual work and human judgment
- Prone to errors and inconsistencies
- Possibility of bias in decision-making
- Inefficient handling of large datasets
- Lack of automation and scalability
- Limited use of historical data for prediction

### **Proposed System**

The proposed system is a machine learning-based loan approval prediction system designed to automate the decision-making process. It uses applicant data such as demographic and financial attributes to predict loan eligibility. The system performs data preprocessing including cleaning, encoding, and feature engineering. A new feature called Total Income is created to improve prediction accuracy. The Random Forest algorithm is used to train the model due to its robustness and high performance. The dataset is split into training and testing sets for evaluation. The trained model predicts whether a loan application should be approved. The system provides fast and consistent results. It reduces human intervention and bias in decision-making. Overall, it enhances efficiency, accuracy, and reliability in loan approval processes.

### **Advantages of Proposed System**

- Faster and automated loan approval process
- Improved accuracy using machine learning algorithms
- Reduces human errors and bias
- Efficient handling of large volumes of data
- Consistent and reliable decision-making
- Cost-effective and scalable solution
- Enhances customer satisfaction through quick responses
- Uses historical data for better predictions
- Supports data-driven financial decision-making

## IV. Methodology

The methodology of the Loan Approval Prediction System follows a structured process to build an accurate and efficient model. Initially, the dataset containing applicant details such as gender, marital status, education, applicant income, co-applicant income, loan amount, and credit history is collected from a reliable source.

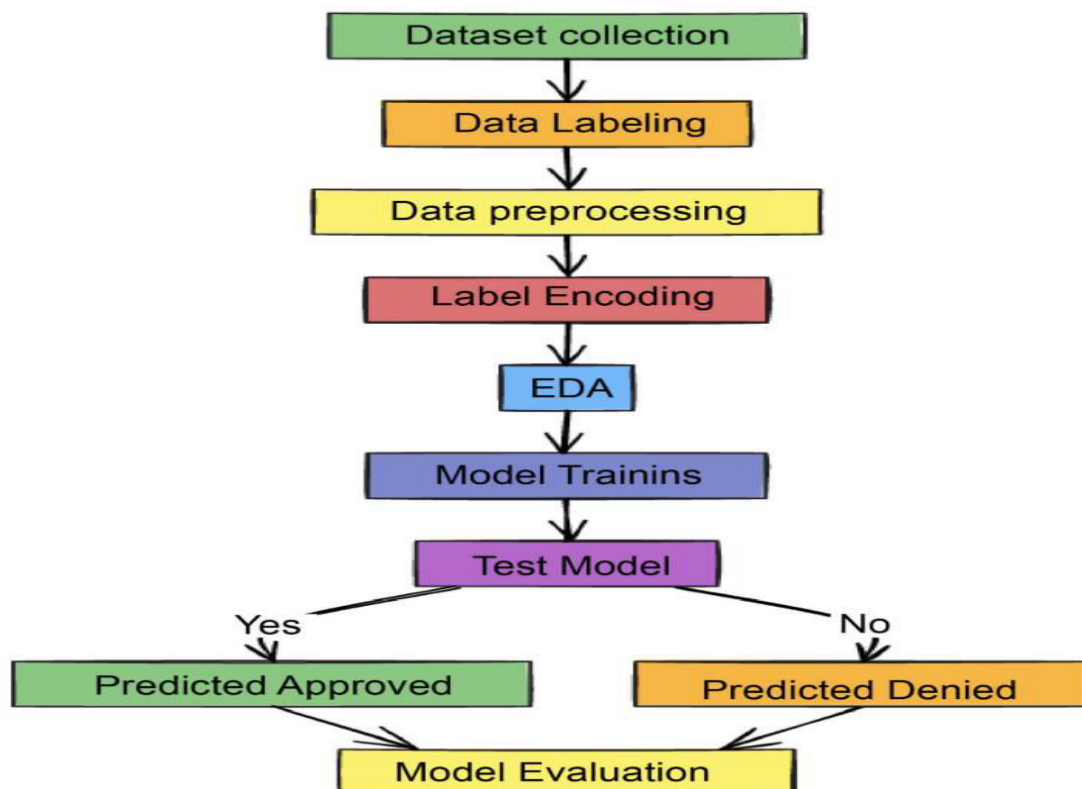
The collected data undergoes preprocessing, which includes handling missing values, removing inconsistencies, and encoding categorical variables into numerical form. Feature engineering is applied to create a new attribute called Total Income by combining applicant and co-applicant incomes. Feature scaling techniques are also used to normalize the data and improve model performance.

Exploratory Data Analysis (EDA) is performed to understand patterns, distributions, and relationships among different variables. The dataset is then divided into training and testing sets to evaluate the model.

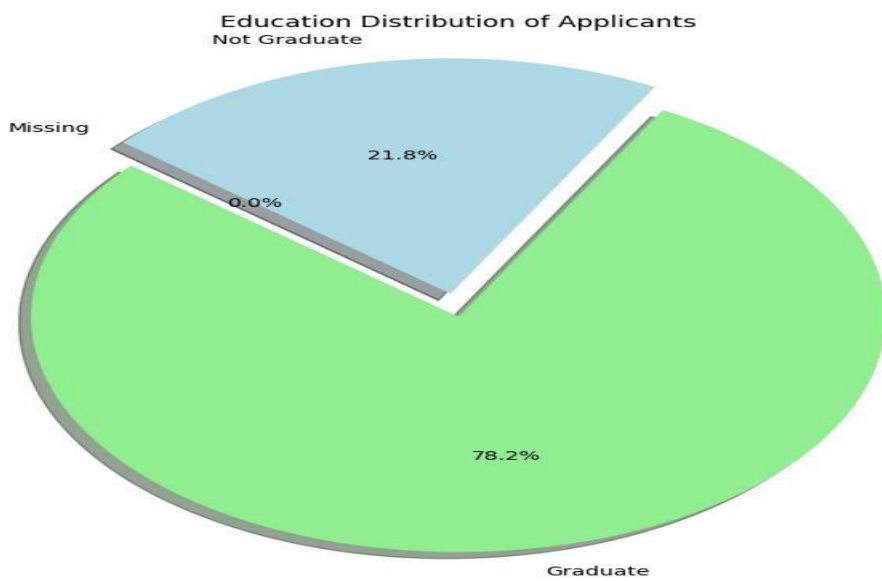
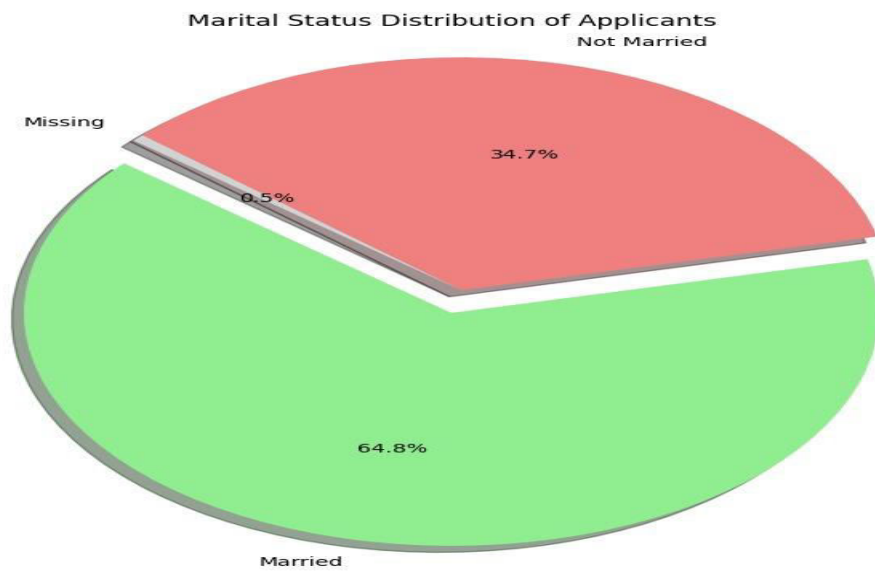
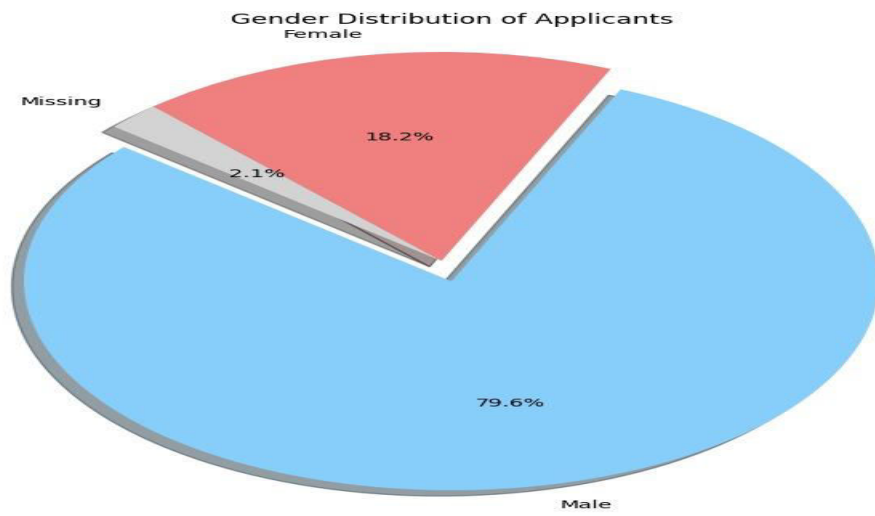
### System Architecture

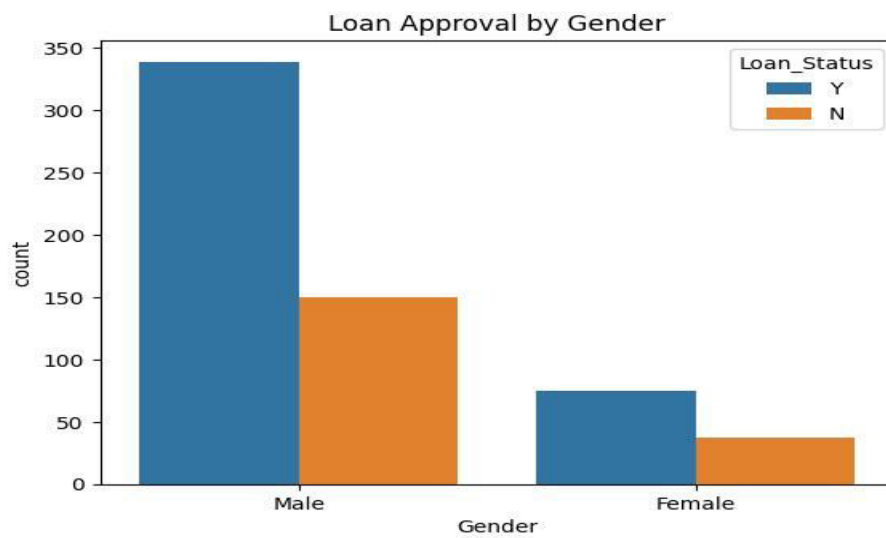
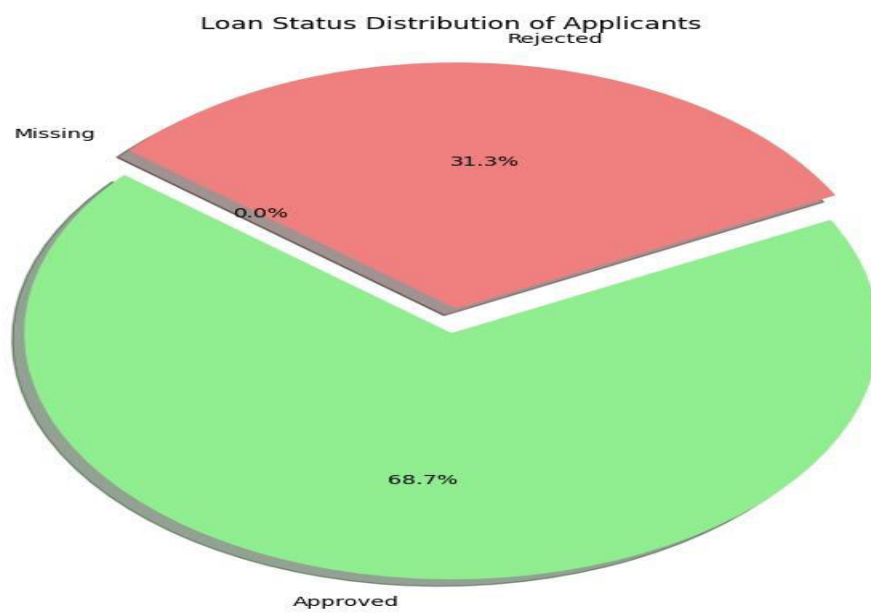
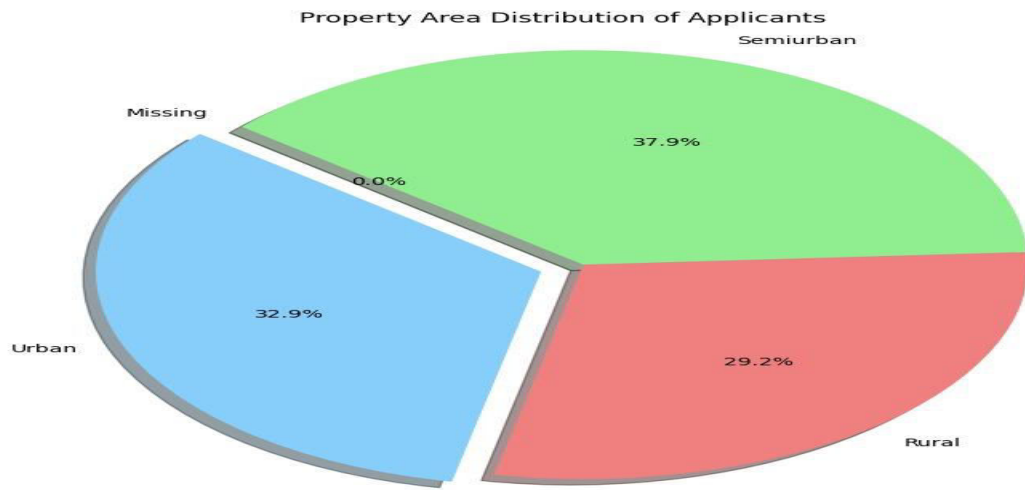
The system architecture defines the workflow and interaction between different components of the Loan Approval Prediction System.

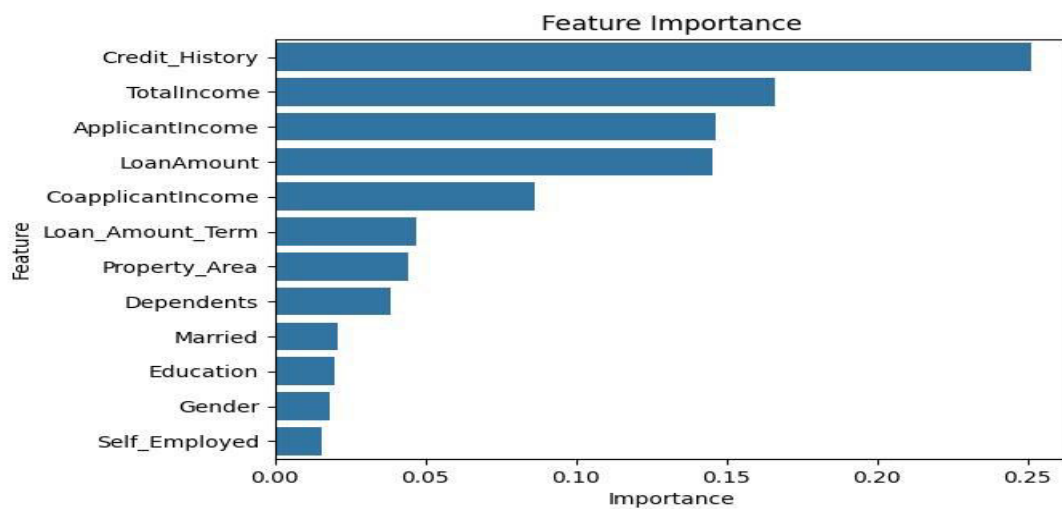
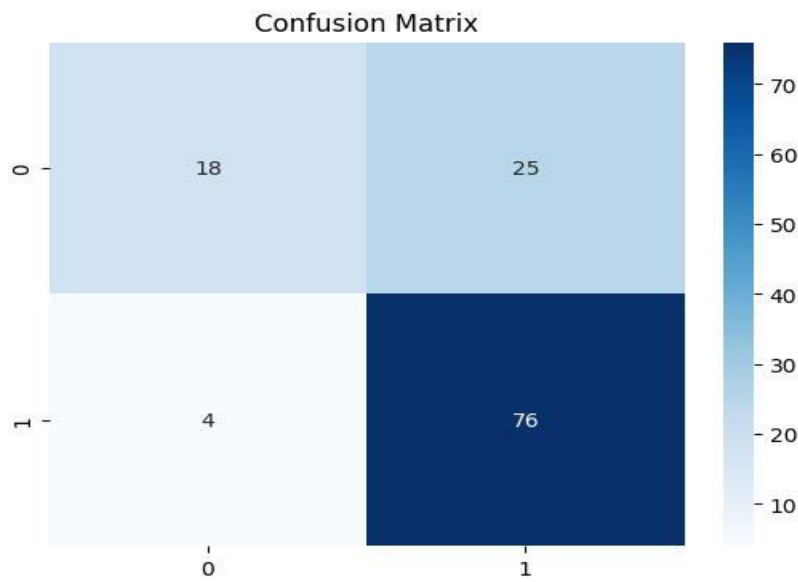
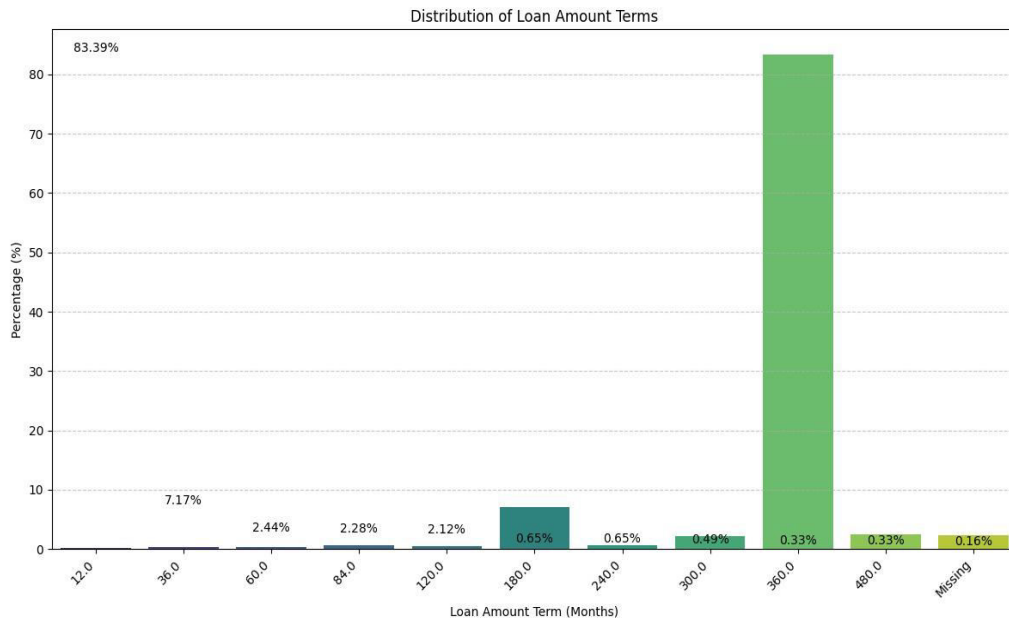
The process begins with the User Interface, where applicants or bank officials input the required details such as income, education, and credit history. The input data is passed to the Data Preprocessing Module, where cleaning, encoding, and feature engineering are performed.



### V. Result and Output







## VI. Conclusion

This project successfully demonstrates the development of an automated loan eligibility prediction system using machine learning techniques. By following a structured pipeline that includes data preprocessing, exploratory data analysis, feature engineering, and model training, a Random Forest classifier was effectively built and evaluated.

The model achieved an accuracy of 76.4%, indicating a reliable level of performance in predicting loan approval outcomes. Feature importance analysis revealed that Credit History is the most significant factor influencing loan approval decisions, followed by total income and loan amount. These insights highlight the importance of financial credibility and repayment capacity in loan evaluation.

The proposed system effectively overcomes the limitations of traditional manual loan processing by providing a scalable, objective, and efficient solution. It reduces human intervention, minimizes bias, and ensures faster decision-making, making it highly beneficial for financial institutions handling large volumes of applications.

## References

- [1] Kumar, R. D., Prudhviraaj, 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.