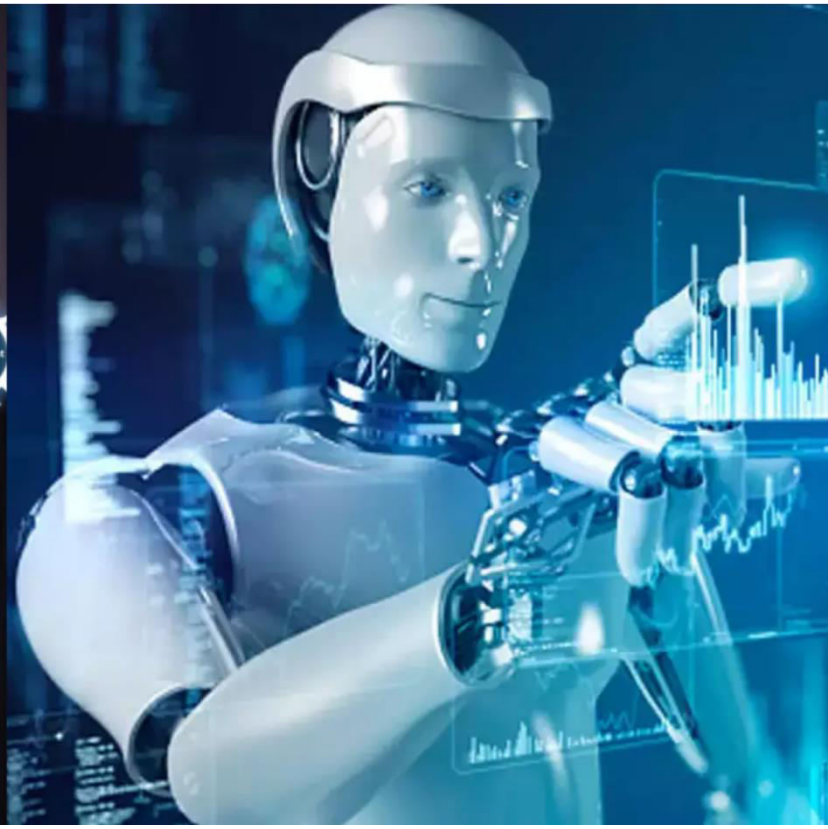
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503229|

Smart Loan Fraud Detection System Using Machine Learning

Katakamsetti Gayathri[1], Dhavaleswarapu Keerthi[2], Garlapati Anand Babu[3], Alla Adi Keshava Reddy[4], Inturi Haranadha Reddy[5], N.R.L.Prasanna[6]

UG Students, Department of CSE (AI & ML), Kallam Haranadhareddy Institute of Technology, Guntur, India[1-5]

Assistant Professor, Department of CSE (AI & ML), Kallam Haranadhareddy Institute of Technology, Guntur, India[6]

ABSTRACT: Loan fraud detection is an important task for banks and financial institutions as fraudulent loan applications can lead to significant financial losses. With the increasing use of digital platforms for loan processing, it has become necessary to develop efficient systems that can identify fraudulent activities at an early stage. Machine learning techniques provide effective solutions by analyzing large amounts of historical data and identifying patterns related to fraud.

In this project, a Loan Fraud Detection System using Machine Learning is developed to analyze applicant data and detect fraudulent loan applications. The dataset is first preprocessed by cleaning the data, handling missing values, and converting categorical data into numerical format so that it can be used for machine learning algorithms. Proper data preprocessing helps improve the performance and accuracy of the model.

Machine learning algorithms such as Decision Tree and Random Forest are used to train the model and classify loan applications as genuine or fraudulent. The trained model helps financial institutions make better loan approval decisions and reduce the risk of financial fraud. This system improves security, increases efficiency in loan processing, and supports reliable financial decision-making.

I. INTRODUCTION

Loan fraud has become a major concern for banks and financial institutions due to the rapid growth of digital financial services and online loan applications. Fraudulent loan activities can cause significant financial losses and affect the trust and stability of financial systems. Traditional methods of fraud detection are often time-consuming and may not be effective in identifying complex fraud patterns. Therefore, there is a need for intelligent systems that can detect fraudulent activities quickly and accurately.

Machine Learning has emerged as an effective approach for detecting fraud by analyzing large amounts of historical data and identifying hidden patterns. By using machine learning techniques, financial institutions can automatically analyze applicant information such as income, credit history, loan amount, and other relevant details to determine whether a loan application is genuine or fraudulent. These techniques help improve the accuracy and efficiency of the fraud detection process.

This project focuses on developing a Loan Fraud Detection System using Machine Learning algorithms such as Decision Tree and Random Forest. The dataset is first preprocessed to handle missing values, clean the data, and convert categorical attributes into numerical form. The machine learning models are then trained using this processed data to predict whether a loan application is legitimate or fraudulent. The proposed system helps financial institutions reduce financial risks, improve decision-making, and enhance the security of the loan approval process.

II. LITERATURE SURVEY

Several researchers have studied the use of machine learning techniques for detecting fraud in financial systems. With the rapid growth of digital banking and online loan services, identifying fraudulent loan applications has become an important task for financial institutions. Machine learning algorithms can analyze large volumes of financial data and

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503229|

identify hidden patterns that indicate fraudulent behavior, helping banks improve the efficiency and accuracy of fraud detection systems.

Bhattacharyya et al. (2011) conducted research on financial fraud detection using machine learning techniques. Their study analyzed various classification algorithms and found that ensemble models such as Random Forest provide better performance in detecting fraud because they can handle complex patterns in financial datasets.

Ngai et al. (2011) reviewed different data mining and machine learning methods used for financial fraud detection. Their research highlighted that classification techniques such as Decision Trees, Neural Networks, and Support Vector Machines are widely used to identify suspicious financial activities and improve fraud detection accuracy.

In recent years, researchers have continued to improve fraud detection systems using advanced machine learning methods. **Ismail and Haq (2024)** proposed a machine learning-based approach for financial fraud detection that uses data analytics and anomaly detection techniques to identify unusual patterns in financial transactions. Their study showed that machine learning models can effectively detect fraud by analyzing large datasets.

Another recent study by **Bhanusri and Ramana Reddy (2025)** focused on fraud detection in banking transactions using machine learning algorithms. Their research demonstrated that machine learning models such as Decision Trees and Random Forest can accurately classify financial transactions and help financial institutions reduce fraud risks.

Overall, these studies show that machine learning techniques play a significant role in financial fraud detection. By analyzing historical financial data and identifying abnormal patterns, machine learning models help banks detect fraudulent loan applications, reduce financial losses, and improve the security of financial systems.

III. METHODOLOGY

The proposed **Loan Fraud Detection System using Machine Learning** follows a structured workflow consisting of dataset preparation, data preprocessing, model training, and prediction. The system analyzes historical loan data to identify patterns that help in detecting fraudulent loan applications.

1. Dataset Collection

The dataset used in this project contains information about loan applicants and their loan details. It includes various attributes such as applicant income, co-applicant income, loan amount, loan term, credit history, property area, education level, and loan status. These features help in understanding the financial background of the applicant and identifying possible fraud patterns.

The dataset contains two main classes:
• Genuine Loan Applications
• Fraudulent or Risky Loan Applications
These classes help the machine learning model learn the difference between normal and suspicious loan applications.

2. Data Preprocessing

Before training the machine learning model, several preprocessing steps are applied to improve data quality and prepare the dataset for analysis.

• Missing values in the dataset are handled by filling or removing incomplete records.
• Categorical features such as gender, education, and property area are converted into numerical values using label encoding.
• Irrelevant or duplicate data is removed to maintain dataset consistency.
• The dataset is organized into a structured format suitable for machine learning algorithms.
These preprocessing steps help improve the performance of the machine learning model and ensure accurate predictions.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503229|

3. Model Architecture

The machine learning models used in this project are **Decision Tree** and **Random Forest classifiers**. These algorithms are widely used for classification problems and are effective in analyzing complex relationships between different features in a dataset.

The Decision Tree model works by creating a tree-like structure of decision rules based on input features. The Random Forest model improves prediction performance by combining multiple decision trees and generating a final prediction based on the majority result.

These models are implemented using the **Python programming language** with libraries such as **Scikit-learn, Pandas, and NumPy**.

4. Model Training and Evaluation

The dataset is divided into training and testing sets to evaluate the performance of the machine learning models. The training dataset is used to train the model and allow it to learn patterns from historical loan data.

The performance of the trained model is evaluated using different metrics, including:

- Accuracy
- Precision
- Recall
- Confusion Matrix
- Classification Report

These evaluation metrics help measure how effectively the model can identify fraudulent loan applications and distinguish them from genuine ones.

5. Prediction System

After training and evaluation, the final trained model is used to predict the status of new loan applications. The system analyzes the input features of a loan applicant and determines whether the loan application is genuine or potentially fraudulent.

The workflow of the prediction system includes:

1. User enters loan applicant details.
2. The input data is preprocessed and converted into numerical format.
3. The trained machine learning model analyzes the input data.
4. The system predicts whether the loan application is genuine or fraudulent.

This system helps financial institutions make better loan approval decisions, reduce financial risks, and improve the overall security of the loan approval process.

IV. RESULTS

The proposed Loan Fraud Detection System using Machine Learning was implemented using Decision Tree and Random Forest algorithms. After preprocessing the dataset, the models were trained using historical loan data to learn patterns related to genuine and fraudulent applications. The trained models were then tested using a separate testing dataset.

The performance of the system was evaluated using metrics such as accuracy, precision, recall, and confusion matrix. These evaluation measures help determine how well the model can correctly classify loan applications. Among the models used, Random Forest showed better prediction accuracy compared to Decision Tree.

The results show that the system can effectively analyze applicant details and identify possible fraudulent loan applications. By using machine learning techniques, the system improves the efficiency of fraud detection. This helps financial institutions reduce financial risk and make better loan approval decisions.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503229|

User Evaluation Results:
The user interface starts with an home page.

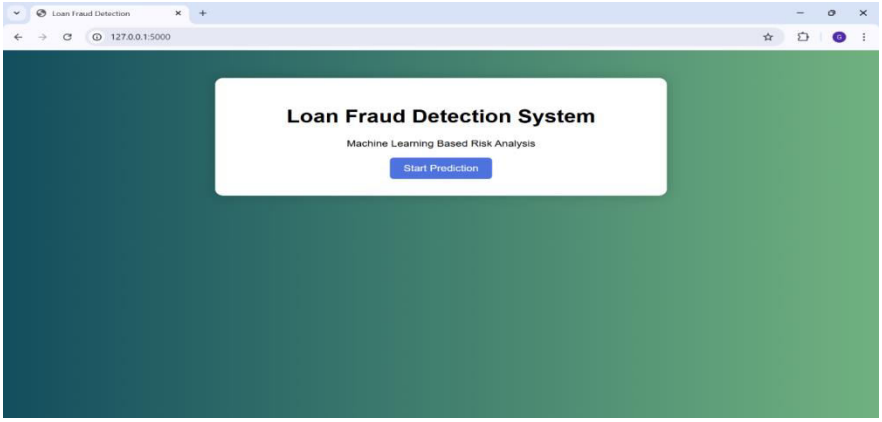


Figure 1: Home page

After click the start prediction the user is redirected to the figure2 where the user can enter the details of the applicant

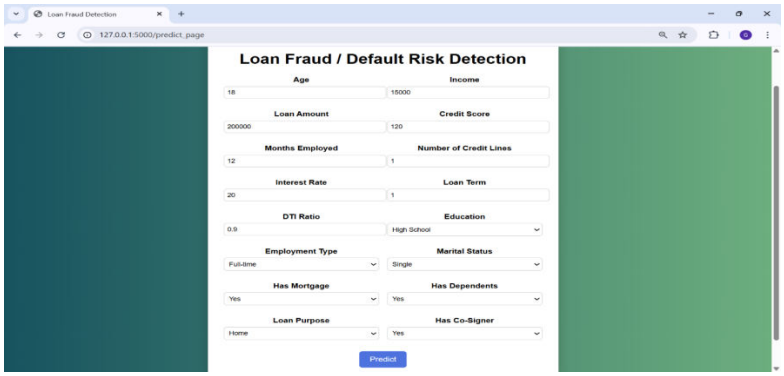


Figure 2: Prediction page

Here the figure3 shows the predicted result whether the loan application is fraud or not

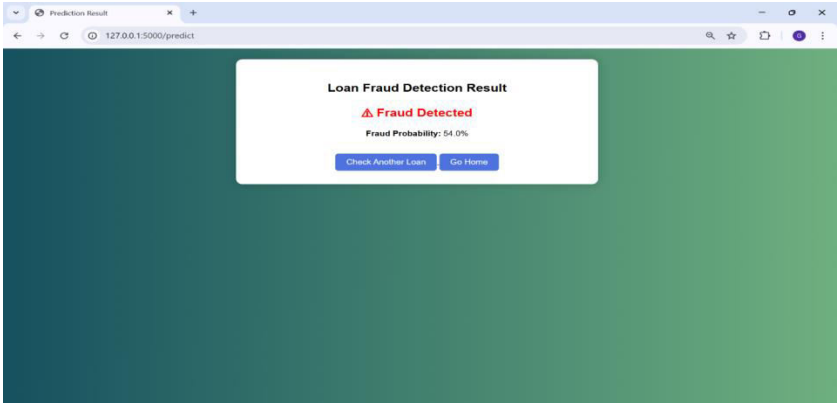


Figure 3: Result

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503229|

V. LIMITATIONS

The performance of the Loan Fraud Detection System depends on the quality and size of the dataset used for training. If the dataset contains missing, incomplete, or inaccurate information, the prediction accuracy of the model may decrease.

The model is trained using historical loan data, so it may not always detect new or unknown fraud patterns that were not present in the training dataset. Fraud techniques can change over time, which may affect the model's effectiveness.

Another limitation is that the dataset may be imbalanced, where genuine loan applications are much higher than fraudulent cases. This imbalance can make it difficult for the model to correctly identify fraud cases.

The system relies only on the available features in the dataset. If important applicant or financial details are not included, the model may not capture all patterns related to fraudulent behavior.

The current system focuses mainly on prediction based on input data and does not include real-time monitoring of loan applications. This may limit its ability to detect fraud instantly in real-world financial systems.

VI. FUTURE WORK

In the future, the system can be improved by using larger and more diverse datasets to increase the accuracy of fraud detection. A larger dataset will allow the model to learn more patterns related to genuine and fraudulent loan applications.

Advanced machine learning and deep learning algorithms can be implemented to enhance the performance of the system. Techniques such as neural networks, gradient boosting, and ensemble learning can help detect complex fraud patterns more effectively.

The system can also be improved by incorporating real-time data analysis so that loan applications can be monitored instantly. This would help financial institutions identify suspicious activities quickly and take preventive actions.

Another possible improvement is integrating additional features such as customer transaction history, credit score, and behavioral data. Including more relevant attributes can help the model make more accurate predictions.

Future work may also focus on improving the user interface of the web application. A more interactive and user-friendly interface can make the system easier to use and provide better visualization of prediction results.

Additionally, the model can be regularly updated and retrained with new data to adapt to changing fraud patterns. Continuous model improvement will ensure that the system remains effective in detecting new types of fraudulent activities.

VII. CONCLUSION

The Loan Fraud Detection System using Machine Learning was developed to identify fraudulent loan applications and support secure loan approval processes. Machine learning algorithms such as Decision Tree and Random Forest were used to analyze historical loan data and detect patterns related to fraud.

The dataset was preprocessed by handling missing values and converting categorical data into numerical form to improve model performance. The trained models were evaluated using different metrics, which helped measure their effectiveness in classifying genuine and fraudulent loan applications.

The results show that machine learning techniques can effectively assist financial institutions in detecting suspicious loan applications. This system helps reduce financial risks, improve decision-making, and enhance the overall security of loan approval systems.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503229|

REFERENCES

1. **Vuppala, S. K. (2023).**
Modeling Fraud Detection in Community Development Banking Through Machine Learning.
International Journal of Intelligent Systems and Applications in Engineering.
Link: https://www.ijisae.org/index.php/IJISAE/article/view/7547
This study applies machine learning algorithms to detect fraud in banking systems and improve loan-related fraud identification.
2. **Rao, R. K., & Mandhala, V. N. (2024).**
Unveiling Financial Fraud: A Comprehensive Review of Machine Learning and Data Mining Techniques.
Intelligent Systems and Applications.
Link: https://www.iieta.org/journals/isi/paper/10.18280/isi.290620
This paper reviews machine learning and data mining methods used to detect fraud in financial and banking systems.
3. **Cardona, L. F., Guzmán-Luna, J. A., & Restrepo-Carmona, J. A. (2024).**
Machine Learning Applications in Fraud Detection on Crowdfunding Platforms.
Journal of Risk and Financial Management.
Link: https://www.mdpi.com/1911-8074/17/8/352
The research explains how machine learning models analyze financial data and detect fraud in funding and lending platforms.
4. **Talukder, M. A., Hossen, R., Uddin, M. A., Uddin, M. N., & Acharjee, U. K. (2024).**
Hybrid Ensemble Machine Learning Model for Fraud Detection.
Link: https://arxiv.org/abs/2402.14389
This study proposes an ensemble machine learning model combining Decision Tree, Random Forest, and KNN to improve fraud detection accuracy.
5. **Das, S., Meghanath, A., Behera, B. K., Mumtaz, S., Al-Kuwari, S., & Farouk, A. (2025).**
Quantum Feature Deep Neural Networks for Fraud Detection and *Loan Prediction.*
Link: https://arxiv.org/abs/2504.19632
The research proposes an advanced deep learning model for fraud detection and loan prediction using financial datasets

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details