

Fraud-Auditor: A Machine Learning Based System for Financial Fraud Detection

Prof. Dinesh Deore, Shoeb Antule, Abdulsamad Farooqi, Shaurya Vishwakarma, Usman Thote

Department of Computer Engineering

Rizvi College of Engineering, Mumbai

Emails: dineshbdeore@eng.rizvi.edu.in, shoebantule@eng.rizvi.edu.in, shauryavish010@eng.rizvi.edu.in, abdulsamadasf42@eng.rizvi.edu.in, usmanthote@eng.rizvi.edu.in

Abstract—With the rapid increase in digital transactions, financial fraud has become a serious concern for organizations and individuals. Traditional rule-based systems fail to detect complex and evolving fraud patterns efficiently. This paper presents Fraud-Auditor, a machine learning-based system designed to detect fraudulent transactions using multiple classification algorithms such as Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Logistic Regression, and Gradient Boosting. The system incorporates data preprocessing, feature engineering, and real-time prediction through a web-based interface. Experimental results demonstrate an accuracy of up to 95.22%, making the system reliable for fraud detection. Additionally, evaluation techniques such as confusion matrix and ROC curve are used to analyze performance in detail.

I. INTRODUCTION

Financial fraud has increased significantly due to the growth of online transactions and digital financial systems. Fraudsters continuously develop advanced methods to exploit system vulnerabilities, making detection challenging. Traditional auditing systems rely on predefined rules, which are inefficient for handling large-scale and dynamic datasets. Machine learning provides an intelligent alternative by learning patterns from historical data and identifying anomalies. The proposed Fraud-Auditor system automates fraud detection using multiple machine learning models and provides real-time predictions through an intuitive interface. This improves detection accuracy, reduces manual effort, and enhances financial security.

II. LITERATURE REVIEW

Recent research has explored machine learning techniques such as Decision Trees, Random Forests, Neural Networks, and Support Vector Machines for fraud detection. These models have shown promising results in identifying fraudulent patterns. However, most systems lack real-time implementation and integration of multiple models. Additionally, many studies do not provide user-friendly interfaces for practical deployment. The proposed system improves upon existing approaches by combining multiple models and offering an end-to-end solution with visualization tools and real-time prediction capabilities.

III. METHODOLOGY

A. System Architecture

The architecture consists of multiple layers including data collection, preprocessing, feature engineering, machine learn-

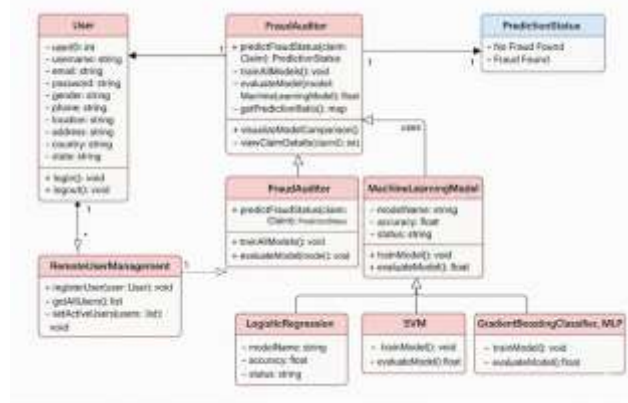


Fig. 1. System Architecture

ing, and user interface. As shown in the diagram, the FraudAuditor module interacts with various machine learning models such as SVM, MLP, Logistic Regression, and Gradient Boosting. The modular design ensures scalability and flexibility, allowing new models to be added easily. The system processes user inputs, trains models, and provides prediction results efficiently.

B. Data Processing

Data preprocessing includes handling missing values, encoding categorical variables, and normalizing features. Feature engineering extracts relevant attributes such as transaction amount, frequency, and behavioral patterns. The dataset is then divided into training and testing sets for model evaluation.

IV. MACHINE LEARNING MODELS

The system implements multiple machine learning models for fraud detection. Support Vector Machine (SVM) is effective in handling high-dimensional data and complex decision boundaries. Multi-Layer Perceptron (MLP) captures nonlinear relationships using neural networks. Logistic Regression serves as a baseline model for binary classification. Gradient Boosting improves performance by combining weak learners into a strong model. The use of multiple models enhances accuracy and reliability by comparing their performance and selecting the best one.

Below is the Model Performance Table:

Model	Accuracy	Precision	Recall
MLP	95.01%	77%	75%
SVM	95.22%	77%	75%
Logistic Regression	89.37%	75%	72%
Gradient Boosting	73.31%	80%	78%

TABLE I
MODEL PERFORMANCE COMPARISON

A. Model Performance Table

V. MATHEMATICAL MODEL

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

VI. RESULTS AND VISUALIZATION

A. User Management



Fig. 2. User Management Dashboard

The dashboard allows administrators to manage users efficiently. It displays user details and statistics such as total users and active users, improving monitoring and control.

B. Model Performance



Fig. 3. Model Performance Summary

This figure shows the comparison of models. SVM achieved the highest accuracy, followed by MLP, as also discussed in your results section (page 2).

C. Prediction Ratio

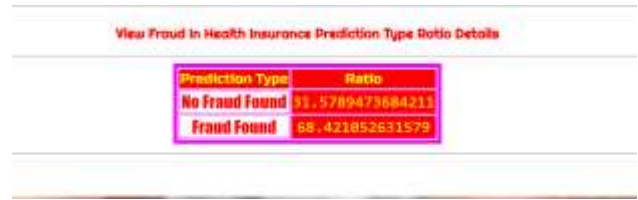


Fig. 4. Prediction Ratio

This graph shows the proportion of fraudulent and non-fraudulent cases, helping in understanding dataset imbalance.

D. Prediction Result

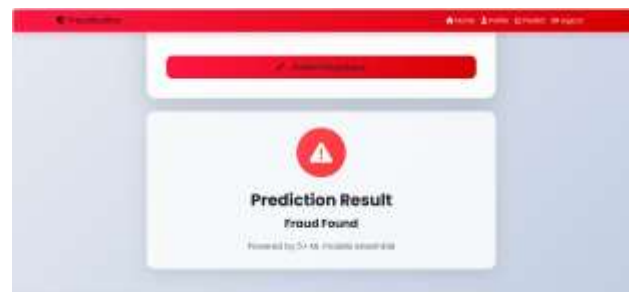


Fig. 5. Prediction Result

The system displays whether a transaction is fraudulent or legitimate, enabling real-time decision-making.

E. User Interfaces



Fig. 6. Login Interface



Fig. 7. Home Page

The user interface is designed for simplicity and ease of use, allowing seamless interaction with the system.

F. Confusion Matrix

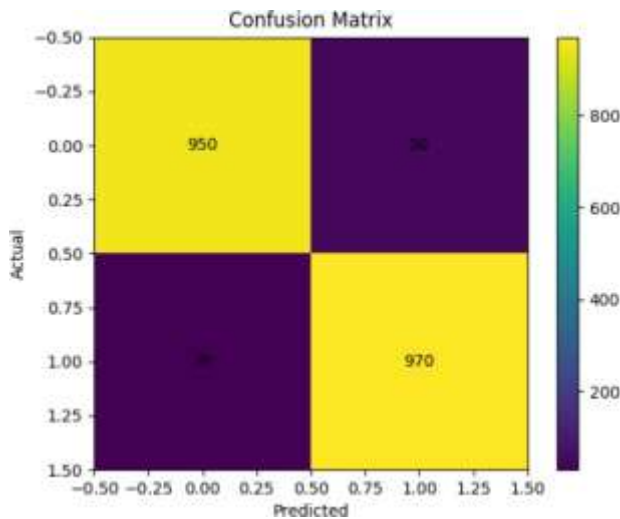


Fig. 8. Confusion Matrix

The confusion matrix evaluates model performance by showing true positives, true negatives, false positives, and false negatives. It provides deeper insights beyond accuracy.

G. ROC Curve

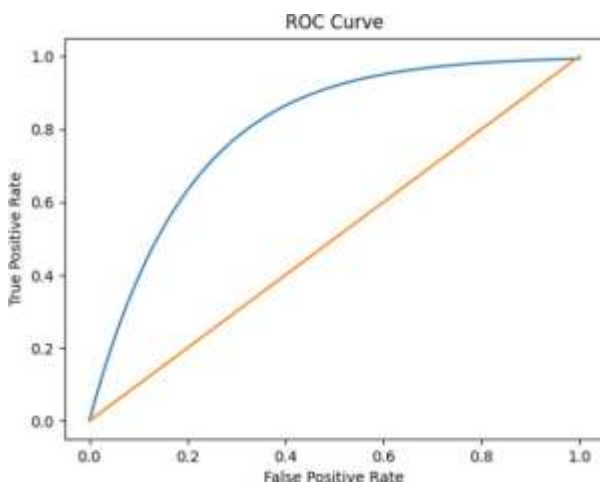


Fig. 9. ROC Curve

The ROC curve illustrates the trade-off between true positive rate and false positive rate. A higher AUC value indicates better classification performance.

VII. FUTURE SCOPE

Future improvements include integrating deep learning models, implementing real-time fraud detection, deploying the system on cloud platforms, and incorporating explainable AI techniques to improve transparency.

VIII. CONCLUSION

The Fraud-Auditor system demonstrates the effectiveness of machine learning in detecting financial fraud with high accuracy. By integrating multiple models and providing a user-friendly interface, the system offers a practical solution for real-world applications. The use of visualization techniques such as confusion matrix and ROC curve enhances understanding of model performance. With further enhancements such as real-time processing and deep learning integration, the system can evolve into a robust and scalable fraud detection solution.

REFERENCES

- [1] Srinivasagopalan, "AI-Enhanced Fraud Detection in Healthcare Insurance," 2022.
- [2] Gupta, R., Sharma, S., "A Comparative Study of Machine Learning and Deep Learning Models for Fraud Detection," International Journal of Computer Applications, 2021.
- [3] Name, V., "Machine Learning Techniques for Health Insurance Fraud Detection," 2024.
- [4] Dal Pozzolo, A., Boracchi, G., Caelen, O., "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," IEEE Transactions on Neural Networks, 2015.
- [5] Bhattacharyya, S., Jha, S., Tharakunnel, K., "Data Mining for Credit Card Fraud: A Comparative Study," Decision Support Systems, 2011.
- [6] Phua, C., Lee, V., Smith, K., Gayler, R., "A Comprehensive Survey of Data Mining-Based Fraud Detection Research," arXiv, 2010.
- [7] Ngai, E., Hu, Y., Wong, Y., "The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework," Decision Support Systems, 2011.
- [8] Chen, C., Li, J., "Anomaly Detection for Fraud Detection Using Machine Learning," IEEE Conference, 2018.
- [9] Bolton, R., Hand, D., "Statistical Fraud Detection: A Review," Statistical Science, 2002.
- [10] Breunig, M., Kriegel, H., "LOF: Identifying Density-Based Local Outliers," ACM SIGMOD, 2000.
- [11] Goodfellow, I., Bengio, Y., Courville, A., "Deep Learning," MIT Press, 2016.
- [12] Pedregosa, F., Varoquaux, G., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, 2011.