International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING MODELS

Tharun K[1],Saravanan R[2], Seeralan G[3], Jagan D[4], C.CHINNIMA [5]

Department of Computer Science, Kingston Engineering College, Vellore
Email: **tharun.kesavan0704@gmail.com**,**saravanansai714@gmail.com**
seeralan24122001@gmail.com,
jagan2003js12345@gmail.com,
[5] Under the guidance of:
AP/CSE
Kingston Engineering College, Vellore
Email: **chinni16.engineering@kingston.ac.in**

ABSTRACT :

This research introduces a state-of-the-art web-based tool for online credit card fraud detection in real-time based on ensemble machine learning algorithms, i.e., Random Forest, XGBoost, and Gradient Boosting. The tool is implemented using Streamlit with a user-friendly interface for manual entry, CSV upload, and synthetic data generation. The main contributions are robust data preprocessing, real-time visualization of predictions, interpretability of models using feature importance, and responsive design for both light and dark themes. The system consists of trained models and a scaler such that fraud prediction can be performed on new data without retraining. Experiments confirm the effectiveness and accuracy of ensemble models in identifying fraudulent transactions. The platform is applicable both for research exploration and real-world deployment within financial fraud prevention systems.

Keywords :Credit Card Fraud Detection, Machine Learning, Random Forest(RF), eXtreme Gradient Boosting (XGBoost), Gradient Boosting (GB),Streamlit, Real-Time Detection

1. Introduction

The worldwide increase in web-based financial transactions has resulted in a rise in fraudulent activities, and this calls for efficient fraud detection systems. Conventional rule-based static systems tend not to change according to changing fraud patterns. In this paper, an interactive fraud detection system based on ensemble machine learning algorithms is proposed for real-time prediction and data-driven analysis.

2. Related Work

Prior research in fraud detection emphasizes supervised learning, anomaly detection, and neural networks. Ensemble methods like Random Forest and XGBoost have shown promise in imbalanced data scenarios such as fraud detection. This work extends existing solutions by integrating these models into a user-friendly web interface that supports real-time decision-making.

3. Methodology

The platform integrates:

- *Data Preprocessing:* Validates input, handles missing values, and scales features.
- *Models Used:*
    - Random Forest (RF)
    - eXtreme Gradient Boosting (XGBoost)
    - Gradient Boosting (GB)
- *Model Training:* Models are trained with balanced datasets (SMOTE or undersampling externally) and saved via Pickle.
- *Interface Design:* Built using Streamlit with advanced CSS for responsiveness and dark/light theme support.
- *Prediction Modes:* Manual input, batch CSV upload, and synthetic data generation.

4. System Architecture

The system architecture includes:

- *Frontend:* Developed with Streamlit, includes form-based inputs & dynamic visualization.
- *Backend:* Handles model loading, data preprocessing, and prediction logic.
- *Visualization:* Displays confusion matrix, fraud probability, and feature importance via Plotly charts.

5. Results and Evaluation

In order to assess the performance of the implemented credit card fraud detection platform, a rigorous set of experiments was performed using the open-source Credit Card Fraud Detection Dataset from Kaggle, having 284,807 transactions with 492 fraudulent ones. Because of this extreme class imbalance, the usual classifiers are prone to perform very poorly. Therefore, ensemble methods like Random Forest, XGBoost, and Gradient Boosting were used due to their stability and effectiveness in dealing with such imbalanced distributions.

Evaluation Metrics:

To have a balanced evaluation, the following metrics were employed:

- Accuracy: General accuracy of the model
- Precision: Correct frauds predicted out of all frauds predicted (positive predictive value)
- Recall (Sensitivity): Positive frauds predicted out of all actual frauds
- F1-Score: Harmonic mean of precision and recall
- AUC-ROC: Area under the Receiver Operating Characteristic curve, indicating the model's capability to differentiate between classes

Performance on User Inputs:

To mimic actual use, a few user interaction modes were evaluated:

- Manual input: Made near-real-time predictions (<100 ms).
- CSV batch upload: Processed ~1,000 transactions with predictions in less than 2 seconds.
- Synthetic data: Synthetic data had realistic distributions and validated the model's generalizability.

Interface Usability:

The Streamlit-based interface was examined for:

- Responsiveness: The app responded well to various screen sizes and devices.
- Dark/Light Mode: Visual aspects (charts, metrics) were readable and clean in both themes.
- Loading Speed: Cached model and scaler loading reduced waiting time very much.

6. Conclusion and Future Work

This paper introduces an extensive and easy-to-use Streamlit-based system for credit card fraud detection based on state-of-the-art machine learning methods. The proposed system combines three of the best-performing models—Random Forest, XGBoost, and Gradient Boosting—trained on the publicly released imbalanced credit card fraud dataset. Using SMOTE for oversampling and relying on strong ensemble approaches, the system efficiently overcomes the class imbalance problem that is a distinguishing feature of real-life fraud detection problems.The platform showed high precision, recall, and AUC-ROC values for all models, with the best performance overall coming from XGBoost .Overall, the project successfully balances model interpretability, speed, and usability, offering a scalable and extensible framework for real-time fraud detection.

Future work includes:

- Integration with real-time payment APIs
- Enhanced explainability (SHAP, LIME)
- Expansion to multi-label fraud scenarios and neural models

7. Limitations and Recommendations

Limitations: **The proposed system, though effective, has several limitations.**

- *Data Imbalance*: The model was trained on an imbalanced dataset, which may affect generalization to real-world fraud cases.
- *Static Dataset*: Lacks real-time behavioral or contextual data
- *Anonymized Features*: Makes interpretation and domain insight difficult.
- *Model Complexity*: Ensemble methods used are less interpretable.
- *No Real-Time Integration*: The system does not support live transaction monitoring.
- *Prototype Stage*: It is not integrated into any live financial environment.

Recommendations: **To improve system performance and deployment.**

- Use real-time, behavior-rich, and diverse datasets.
- Implement online learning to adapt to new fraud trends.
- Add explainability tools like SHAP or LIME for better transparency.
- Build real-time APIs and integrate with live payment systems.
- Ensure security, compliance, and cloud-based scalability.
- Establish a feedback loop for continuous learning and performance tuning.