Credit Card Fraud Detection using Machine Learning Models: Comparative Analysis of various models

Submitted by:
Annesha Naskar (2448306), Harshitha S (2448326), Kiran Guruv (2448333), Parameswaran A (2448343), Saranya M (2448356) of 3MDS B.

Under the guidance of
Dr. Hemanth K.S.

Abstract:

Detection of credit card fraud is an important challenge in financial security, which requires strong models to effectively identify fake transactions. In this study, we analyze the three machine learning models– Logistic Regression, Random Forest, and XGBoost-performance, on a real-world credit card fraud dataset. Our goal is to compare how these models handle fraud detection, given the highly unbalanced nature of the dataset. We preprocess the data by handling the missing values, encoding the categorical features and standardizing numerical features. Each model is trained and evaluated using metrics like accuracy, precision, recall and F1-score. The results suggest that the Logistic Regression performs well but it struggles to determine any complex fraud pattern. Random Forest performs better due to its ability to catch non-linear relationships, while XGBoost achieves the highest recall and F1-score, making it the most effective model to detect fraud. By analyzing model performance across using various metrics, this study provides insights into how different algorithms handle fraud detection, balancing accuracy, recall, and computational efficiency. This will help us to understand which model suits best for real-world fraud detection scenarios.

Keywords: credit card frauds, fraud detection, Logistic Regression, Random Forest, and XGBoost accuracy, precision, recall and F1-score.

1. Introduction

The rise of digital transactions has increased credit card fraud, indulging in significant risks to financial institutions and consumers. The fraud transaction leads to financial loss and affects consumer trust, which leads to the need for fraud detection in banking and financial sectors. Traditional fraud detection depends on the rules-based systems, which are often insufficient due to the development of the pattern of fraud. Machine Learning provides an adaptive approach by learning from historical transactions data and improving the accuracy of detection of fraud over time.

The primary purpose of this study is to compare different machine learning models to determine the best performing model in identifying fraud transactions. In detecting fraud, challenges faced are like serious class imbalances, feature engineering and needs of models that correctly identify fraudulent transactions. The study focuses on behavioral differences of three models – Logistic Regression, Random Forest, and XGBoost - when applied to a dataset.

In particular, we:

- Apply preprocessing techniques, which involves handling missing values, encoding classified features and normalizing numerical features.
- Train Logistic Regression, Random Forest, and XGBoost models on dataset.
- Evaluate and compare their performance using accurate, recall and F1-score.

We analyze how each model treats the dataset and effectively detect fraudulent transactions.

By comparing these models, we provide insight into their strengths and weaknesses, helping the selection of the best approach to detect real -world fraud.

2. Literature Review

Several studies have used machine learning approaches to detect fraud, showing that various models perform diversely based on dataset characteristics. Research indicates that supervised learning models dominate the detection of credit card fraud, appearing in most studies with Support Vector Machines (SVMs) and Logistic Regression, followed by Random Forest and K-Nearest Neighbors (KNN). Large datasets, often exceeding 100,000 transactions, are usually used to detect fraud.

The most frequently used evaluation metrics include precision and recall/sensitivity, which emphasize the need to correctly identify fraudulent transactions while reducing false negatives. Random Forest and boosting-based models perform better having accuracy, precision, and F1-scores ranging between 0.999 and 1.00. In contrast, Logistic Regression displays a broad accuracy range (0.5486–0.9444), which suggests that the model depends on the type of dataset. K-Nearest Neighbors achieves high accuracy (0.9769–0.9993), while one study finds that a fuzzy-based SVM gives an accuracy of 0.9861.

Maja Puh and Ljiljana Brkić,2019 say that there is considerable interest in using machine learning algorithms for credit card fraud detection through data mining techniques. Nevertheless, there are several challenges that come into play, such as the non-availability of public datasets, high class distribution imbalance, and the ever-changing nature of fraudulent activities, making model generalization challenging.

Further, P. Tiwari et al.,2024 also emphasizes the complexity involved in determining the authenticity of transactions. The study brings to light the need for credit card issuing organizations to have proper fraud detection mechanisms in place to help counteract financial losses. Their study also brings to light the need for adaptive and scalable machine learning platforms that can evolve to keep up with newly evolving fraudulent patterns.

The literature suggests that ensemble models, particularly from Random Forest and boosting algorithms, exhibit high performance in identifying credit card fraud mostly in big and imbalanced datasets. This study uses these results by comparing the performance metrics of Logistic Regression, Random Forest, and XGBoost using a real-world fraud detection dataset and comparing their performance using various performance metrics.

3. Methodology

This section details the systematic approach for detecting credit card fraud, encompassing data preprocessing, feature engineering, model selection, and evaluation. The methodology addresses class imbalance, ensures reproducibility, and optimizes model performance.

A. Dataset

The dataset used for this study is sourced from Kaggle and consists of 1,852,394 credit card transactions. Each transaction is labeled as either fraudulent (is_fraud = 1) or legitimate (is_fraud = 0). The dataset contains 23 features that include transaction details, cardholder information, merchant details, and fraud indicators. The primary objective of this study is to develop machine learning models to accurately identify fraudulent transactions and minimize false positives.

Key features in the dataset include:

- **Transaction Information:** Transaction date and time, unique transaction identifier, and transaction amount.
- **Cardholder Details:** Anonymized personal details such as name, gender, address, and job.
- **Merchant Details:** Merchant location data (latitude and longitude).
- **Fraud Indicator:** Binary classification indicating whether a transaction is fraudulent or legitimate.

B. Mathematical Model

1. Preprocessing Techniques

To ensure data quality and model efficiency, the following preprocessing steps were applied:

- **Handling Missing Values:** Missing data, if present, was analyzed and either imputed using statistical methods or removed based on its impact on model performance.
- **Feature Engineering**: New features were derived to enhance model interpretability, such as:
  (i) Transaction frequency per user.

$$Z_{score} = \frac{x - \mu}{\sigma}$$

where:
x= Transaction amount,
μ = Mean transaction amount,
σ= Standard deviation.

(ii) Aggregated spending behavior per merchant or category.
(iii) Location-based fraud detection patterns.

$$d = R \times \arccos(\sin\phi1 \times \sin\phi2 + \cos\phi1 \times \cos\phi2 \times \cos(\lambda2 - \lambda1))$$

where:
R= Earth's radius (6371 km),
(ϕ1, λ1) = Cardholder's latitude & longitude,
(ϕ2, λ2) = Merchant's latitude & longitude.

- **Encoding Categorical Variables:** Categorical features such as merchant categories and cardholder jobs were encoded using one-hot encoding or label encoding.
- **Scaling Numerical Features:** Features like transaction amount and population size were normalized using Min-Max Scaling:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where:
X = Original feature value
X_{min} = Minimum value in the dataset
X_{max} = Maximum value in the dataset

2. Machine Learning Models

Several machine learning models were implemented and evaluated to identify fraudulent transactions effectively. The models used include:

(i) Logistic Regression

A baseline model to determine the linear separability of fraud cases.

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=0}^{n} \beta_i X_i)}}$$

where:
P (Y = 1|X) = Probability of fraud
β_0 = Intercept
β_i= Coefficients for feature X_i

(ii) Random Forest

Tree-based models to capture complex decision boundaries.
Random Forest model combines multiple decision trees:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(X)$$

where:
T = Number of trees
$h_t(X)$= Prediction from tree t

(iii) XGBoost

Gradient boosting models known for their high predictive accuracy.
XGBoost optimizes a loss function:

$$L = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k} \Omega(f_k)$$

where:

$l(y_i, \hat{y}_i)$ = Loss function (e.g., log loss)

$\Omega\left(f_k\right)$= Regularization term

C. Model Accuracy

Model performance was evaluated using the following metrics:

- ***Accuracy:*** Overall correctness of the model's predictions.

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

- ***Precision:*** Ratio of correctly predicted fraud cases to total predicted fraud cases.

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

- ***Recall (Sensitivity):*** Ability to correctly identify fraudulent transactions.

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

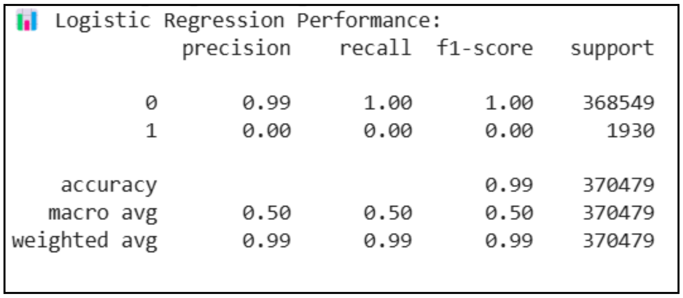
- ***F1-Score:*** Harmonic mean of precision and recall, used for imbalanced classification problems.

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

- Area Under the ROC Curve (AUC-ROC): Measures the ability of the model to distinguish between fraud and legitimate transactions.

5. Result

The ***logistic regression model*** gave an accuracy of 99%, indicating that it correctly classified the majority of transactions.

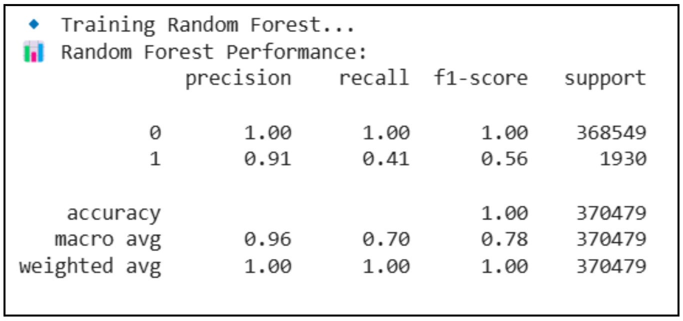


```
📊 Logistic Regression Performance:
              precision    recall  f1-score   support

           0       0.99      1.00      1.00    368549
           1       0.00      0.00      0.00      1930

    accuracy                           0.99    370479
   macro avg       0.50      0.50      0.50    370479
weighted avg       0.99      0.99      0.99    370479
```

However, despite the high accuracy, the macro average precision, Recall and F1-Score are around 50%, which suggests that the model struggles with the imbalanced data.

The precision for the fraudulent class is lower, meaning it misclassifies some fraudulent transactions as non-fraudulent.

The recall for the fraudulent class is also poor, meaning that the model fails to detect some fraudulent transactions. Given that fraud detection is a highly imbalanced classification problem (with a fraudulent transaction than legitimate ones), a model with high accuracy but poor recall is not ideal.

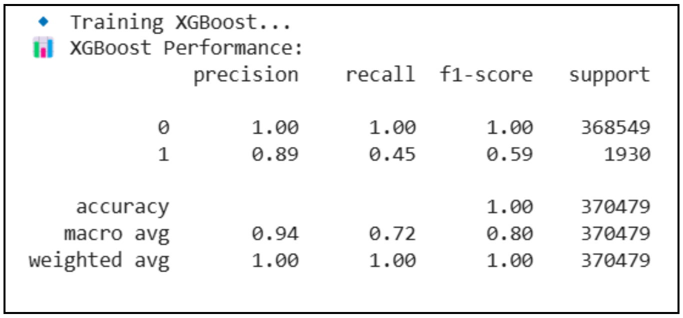


```
🔹 Training Random Forest...
📊 Random Forest Performance:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    368549
           1       0.91      0.41      0.56      1930

    accuracy                           1.00    370479
   macro avg       0.96      0.70      0.78    370479
weighted avg       1.00      1.00      1.00    370479
```

The ***Random Forest model*** achieved an accuracy of 100%, meaning that it correctly classified all transactions in the test set. It had a macro-average Precision of 95% which implies that it is highly confident in its fraud detection. The macro-average Recall was 70%, which suggests that while it detects better than Logistic Regression, there can be an improvement in this.

The macro-average F1 score was 78%, indicating a good balance between precision and recall. The Random Forest Model reduced false negatives (missed fraud cases) compared to Logistic Regression but still did not capture all fraud cases effectively.

The ***XGBoost model*** gave an accuracy of 100%, similar to Random Forest. The macro-average Precision was 94%, indicating that when the model predicted fraud, it was correct 94% of the time. The macro-average Recall was 72%, which was the highest among the models, meaning it identified more fraudulent cases correctly.



```
 ◆ Training XGBoost...
 📊 XGBoost Performance:
               precision    recall  f1-score   support

           0       1.00      1.00      1.00    368549
           1       0.89      0.45      0.59      1930

    accuracy                           1.00    370479
   macro avg       0.94      0.72      0.80    370479
weighted avg       1.00      1.00      1.00    370479
```

The macro-average F1-Score was 80%, showing a strong balance between precision and recall. The model had the best performance in detecting fraud cases with the lowest number of false negatives, making it the most effective in identifying fraud. However,
The XGBoost model had the best Recall (72%)
And F1-Score (80%), making it the most effective model for fraud detection. While Random Forest was a close competitor, its Recalls was slightly lower (70%), making it slightly less effective in detecting fraud. Logistic Regression was the weakest model due to its recall and F1-Score, making it unsuitable for fraud detection despite its accuracy.

6.Conclusion

The analysis was conducted using three primary models - Logistic Regression, Random Forest and XGBoost – each evaluated on key metrics such as accuracy, precision, recall and F1-Score. Among these models XGBoost outperformed the others with an accuracy of 100% and a macro-average recall of 72%, demonstrating its capability in fraud detection. Random Forest also performed well with similar accuracy but had a slightly lower recall (70%), making it slightly less efficient in identifying fraudulent transactions. Logistic Regression, while achieving high accuracy (99%), struggled with recall and failed to detect many fraudulent transactions, making it less suitable for real-world fraud detection applications.

This study highlights the challenges of working with imbalanced datasets, as fraud cases are much rarer than legitimate transactions. To address this oversampling, under sampling, and synthetic data generation such as SMOTE could be explored in future research to further enhance model performance. Additionally, ensemble learning methods combining multiple models, deep learning-based fraud detection and anomaly detection techniques can be investigated for even greater accuracy and robustness.

This underscores the importance of Machine Learning in financial security and fraud detection, providing a solid foundation for further advancements in AI- driven fraud detection.

7. Future Scope of Study

1. Deep Learning-Based Approaches
Investigating deep learning architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTMs), and Autoencoders to enhance fraud detection by capturing complex transaction patterns. Exploring transformer-based models like BERT to analyze transaction sequences and identify fraudulent patterns more effectively.

2. Anomaly Detection Techniques
Implementing unsupervised learning techniques such as Isolation Forests and One-Class SVMs to detect novel fraud patterns without relying on labeled data.

3. Federated Learning for Fraud Detection
Exploring federated learning to enable fraud detection across multiple financial institutions while preserving data privacy and security.

4. Real-Time Fraud Detection Systems
Investigating the integration of fraud detection models into real-time banking systems using streaming analytics tools such as Apache Kafka and Spark Streaming for instant fraud detection.

5. Adversarial Attacks & Security in Fraud Detection
Analyzing potential adversarial attacks on machine learning-based fraud detection systems and developing countermeasures to improve model robustness and security.

8. References

[1] P. Tiwari, Saksham Mittal, Deepak Upadhyay (2024). A Comparative Analysis of Credit Card Fraud Detection Machine Learning Algorithms. 2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT)

[2] Maja Puh, Ljiljana Brkić (2019). Detecting Credit Card Fraud Using Selected Machine Learning Algorithms. International Convention on Information and Communication Technology, Electronics and Microelectronics

[3] Khalid, A.R.; Owoh, N.; Uthmani, O.; Ashawa, M.; Osamor, J.; Adejoh, J. Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach. Big Data Cogn. Comput. 2024, 8, 6. https://doi.org/10.3390/bdcc8010006

[4] Credit Card Fraud Detection Dataset by Tushar Bhadouria https://www.kaggle.com/datasets/tusharbhadouria/credit-card-fraud-detection

Contribution:

ML Model:

- EDA: Parameswaran A (2448343), Annesha Naskar (2448306)
- Preprocessing and Encoding: Parameswaran A (2448343)
- Model Building: Parameswaran A (2448343), Harshitha S (2448326)

Presentation: Annesha Naskar (2448306), Harshitha S (2448326), Kiran Guruv (2448333), Parameswaran A (2448343), Saranya M (2448356)

Report: Annesha Naskar (2448306), Harshitha S (2448326), Kiran Guruv (2448333), Parameswaran A (2448343), Saranya M (2448356)