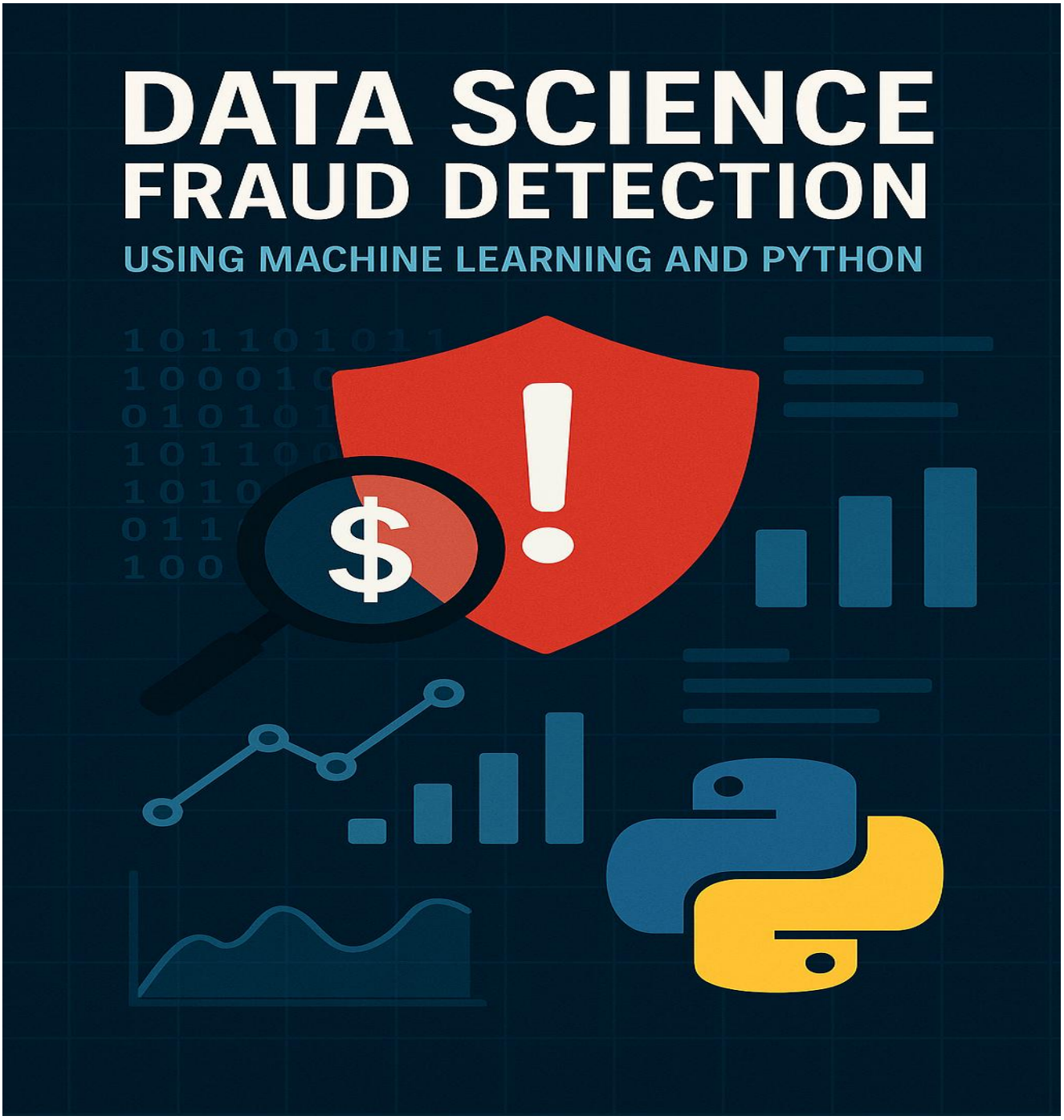


Name: Thato Maelane
Email: thato6216@gmail.com
LinkedIn: https://www.linkedin.com/in/thatomaelane
GitHub: https://github.com/thatomaelane
Kaggle: https://www.kaggle.com/thatomaelane
Date: 09 June 2025

Contents

Aim of the Project

To build a machine learning system that can automatically detect fraudulent credit card transactions and minimize financial losses by flagging suspicious activities.

The Problem We're Solving

- Credit card fraud is **rare** but **very costly**.
- Fraudulent transactions make up **only ~0.17%** of all transactions in the dataset — making it hard for traditional systems to detect them.
- The challenge is to accurately **identify fraud** without generating too many false alarms.

Dataset Information

- Source: Kaggle – Credit Card Fraud Detection
- Contains: 284,807 transactions
- Only 492 are frauds (class 1), rest are normal (class 0)
- Features are anonymized as V1, V2, ..., V28 plus Time, Amount, and Class

Language Used

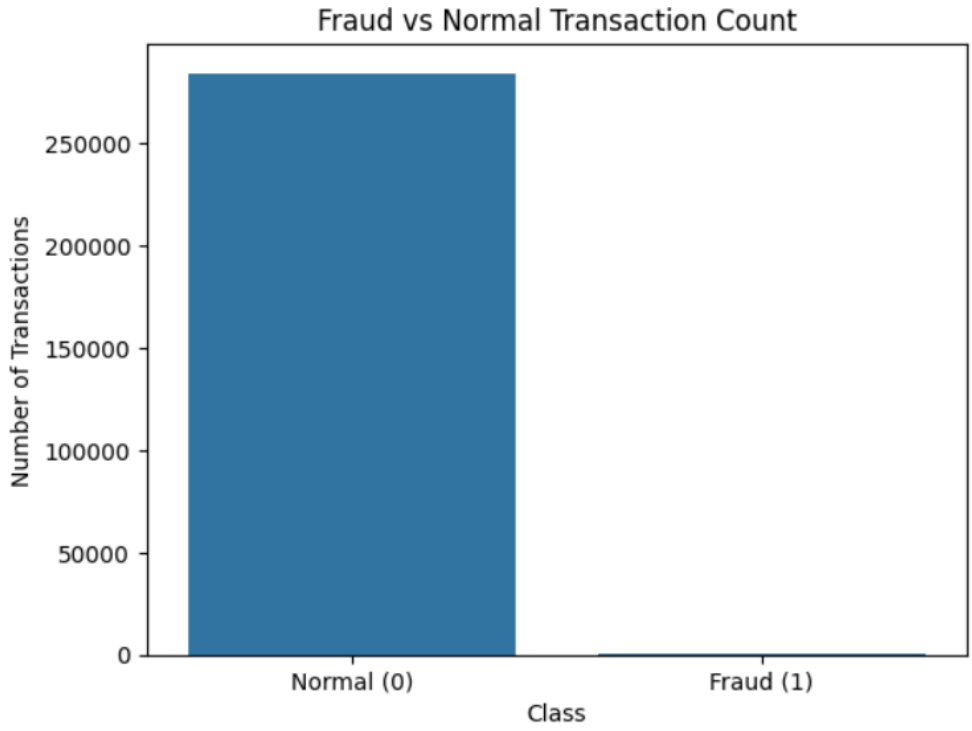
- Python

Models Used

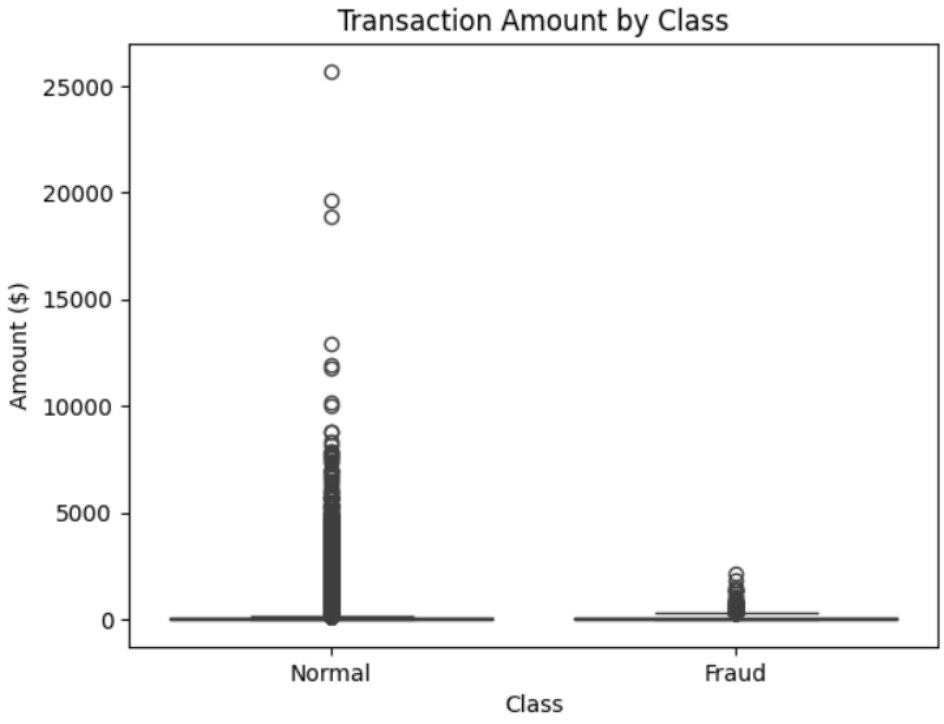
1. Logistic Regression
2. Random Forest Classifier

Results & Findings

Class Distribution Plot



Amount of Money in Fraud VS Normal Plot



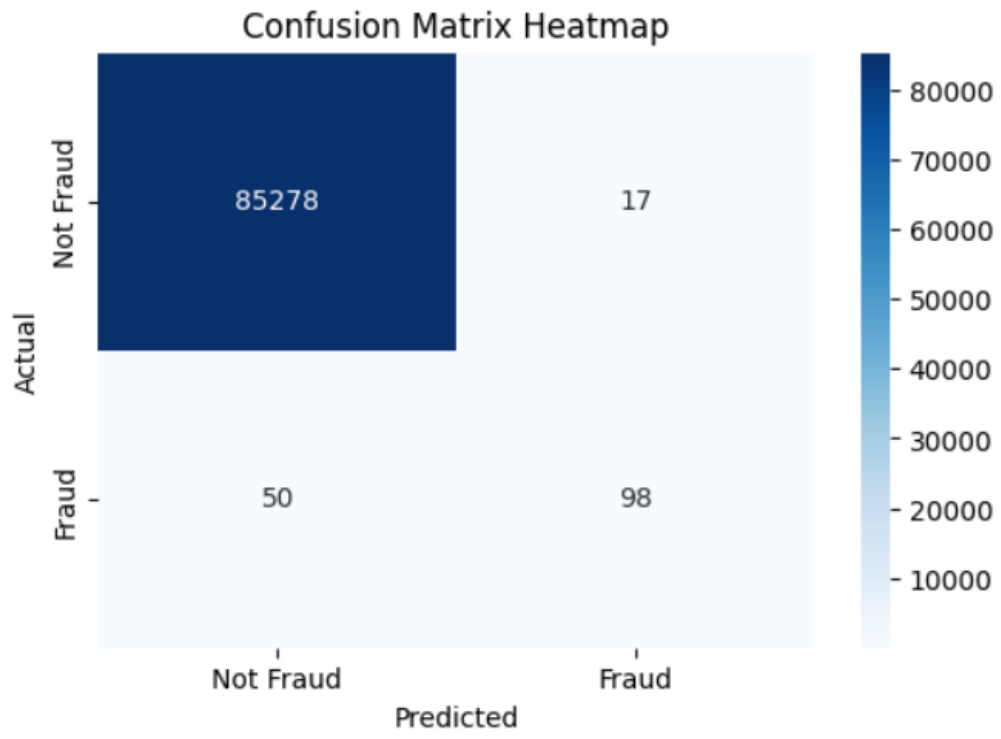
Logistic Regression Confusion Matrix:

```
Confusion Matrix:
[[85278    17]
 [   50    98]]

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     85295
           1       0.85      0.66      0.75       148

    accuracy                           1.00     85443
   macro avg       0.93      0.83      0.87     85443
weighted avg       1.00      1.00      1.00     85443
```

Logistic Regression

- Precision (Fraud): 85%
- Recall (Fraud): 66%
- Missed frauds: 50
- Conclusion: Good start, but misses too many frauds

Confusion Matrix Heatmap (Logistic Regression):



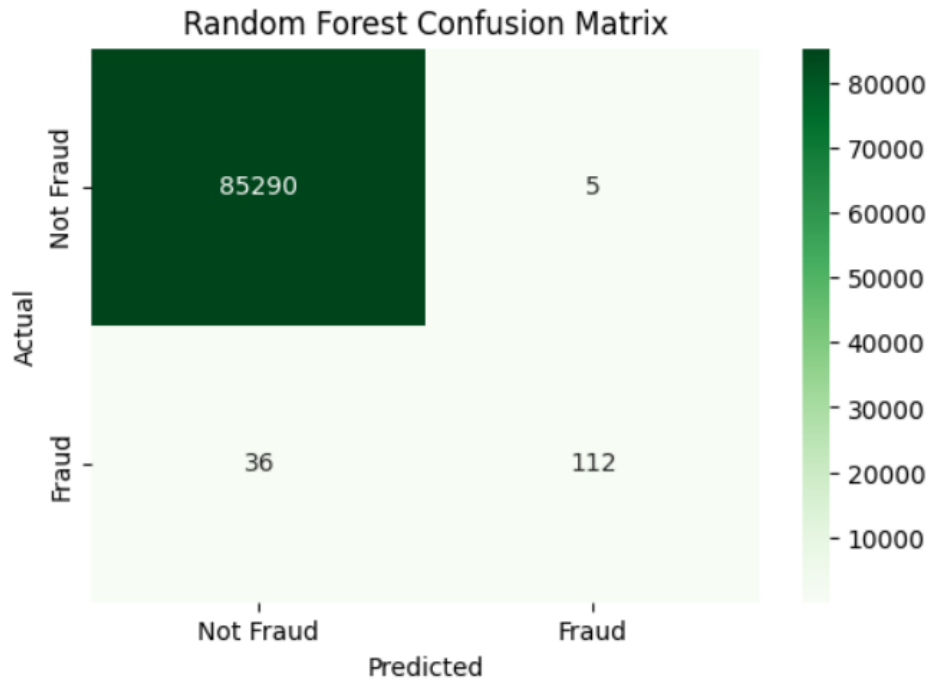
Random Forest Classifier Confusion Matrix:

```
Random Forest Confusion Matrix:
[[85290     5]
 [   36   112]]

Random Forest Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     85295
           1       0.96      0.76      0.85       148

    accuracy                           1.00     85443
   macro avg       0.98      0.88      0.92     85443
weighted avg       1.00      1.00      1.00     85443
```

Random Forest Classifier (Better Model)

- Precision (Fraud): 96%
- Recall (Fraud): 76%
- Missed frauds: 36
- Conclusion: Random Forest Classifier Performs better than Logistic Regression

Confusion Matrix Heatmap (Random Forest Classifier):



Summary & Conclusion

This project successfully demonstrated how machine learning can be applied to detect fraudulent transactions from a large and imbalanced credit card dataset. By training and evaluating multiple models, we found that the Random Forest Classifier performs best, achieving 96% precision and 76% recall, making it both accurate and reliable for real-world use.

This model can now be used as a foundation for a real fraud detection system, and improved further with techniques like:

- Feature engineering
- Scaling/normalization
- Ensemble models or deep learning

Reference

Kaggle, 2016. *Credit Card Fraud Detection*. [online] Available at: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud [Accessed 6 Jun. 2025].