# Credit Card Fraud Detection Project Report

**Author:** Jean Michel Warren Nelson

**Program:** GenAI & Machine Learning Bootcamp 2025 - Part Time 2025

**Submitted on:** October 1, 2025

## I. Executive Summary

This project successfully created a highly effective system for detecting fraudulent credit card transactions. We fixed the main challenge—having very little fraud data—by using the **SMOTE** technique. The final model (Random Forest) is very reliable, keeping **false alarms low (Precision 86%)** while finding **85% of actual fraud (Recall)**.

## II. Methodology

### 1. Data Preparation

* **Feature Engineering:** We cleaned up the raw **Time** and **Amount** values. We used a Log Transformation on **Amount** to handle large values, and created a cyclical **Time-of-Day** feature to capture repeating patterns.
* **Data Splitting:** The data was separated into Training and Testing sets. The Test set was kept completely untouched for a fair evaluation.

### 2. Handling Imbalance

* **Solution:** We used **SMOTE (Synthetic Minority Over-sampling Technique)**.
* **Application:** SMOTE was applied **only to the Training Data** to create a balanced dataset for the model to learn from. This crucial step prevented the model from being biased and ensured the test results are trustworthy.

## III. Results and Evaluation

The model was tested on new, unseen data. The key focus was minimizing false alarms (Precision) to keep operational costs low.

| Metric | Value | What It Means (Class 1 - Fraud) |

| Precision | 0.86 (86%) | When the model said 'Fraud', it was correct 86% of the time. (Low False Alarms) |

| Recall | 0.85 (85%) | The model successfully caught 85% of all actual fraud cases. |

| AUC-ROC Score | 0.973 | The model is extremely good at telling the difference between fraud and normal transactions. |

### Confusion Matrix Snapshot

The final matrix showed only **14 False Positives** (unnecessary alerts) and only **15 False Negatives** (missed fraud) on the test set—excellent results for this challenge.

## IV. Conclusion

The model demonstrates **strong and reliable performance** by successfully solving the extreme class imbalance problem.

* **Low False Alarms (High Precision):** When the model flagged a transaction as fraud, it was correct **of the time**. This means only 14% of alerts were incorrect, saving the operational team time.
* **High Catch Rate (High Recall):** The model simultaneously maintained a strong detection rate, identifying 85% **of all actual fraud cases**.

**Summary:** The system provides an **efficient and balanced** solution for real-time fraud detection. It reliably catches fraud while making very few mistakes.