Credit Card Fraud Detection Report

This report details the process and findings of a project aimed at detecting fraudulent credit card transactions using machine learning techniques. The dataset consisted of 284,807 transactions, of which only 492 were fraudulent, making it a highly imbalanced classification problem.

Key Findings

1. Logistic Regression achieved high accuracy (99.9%) but detected only 56% of fraudulent transactions, highlighting the challenge of imbalanced datasets.

2. Isolation Forest improved recall to 59%, but precision required further optimization.

3. Local Outlier Factor (LOF) underperformed with a recall of 15%, making it less effective.

4. The class imbalance significantly affected model performance, underscoring the importance of appropriate balancing techniques.

Lessons Learned

1. Standard classification models struggle without balancing techniques such as SMOTE or cost-sensitive learning.

2. Precision and recall are more meaningful metrics than accuracy for fraud detection.

3. Anomaly detection methods require fine-tuning to improve precision and recall.

4. Evaluation metrics like AUC-ROC and F1-Score provide a more comprehensive view than overall accuracy.

Next Steps and Recommendations

1. Explore advanced techniques such as ensemble methods (e.g., Random Forest, XGBoost) for better handling of imbalanced data.

2. Engineer new features, such as transaction frequency and time-based aggregates, to improve predictive performance.

3. Experiment with oversampling (SMOTE/ADASYN) or undersampling methods to address class

imbalance.

4. Perform extensive hyperparameter tuning for both supervised and unsupervised models.

5. Develop an ensemble approach that combines supervised and unsupervised methods for robust fraud detection.

6. Build an API (e.g., using Flask or FastAPI) for real-time fraud detection and integrate a feedback loop to refine models.

Closing Remark

This project underscores the complexities inherent in fraud detection and emphasizes the importance of focusing on advanced, tailored techniques to address imbalanced datasets. Future work can build on these findings to develop a more robust and scalable solution for credit card fraud detection.