

Capstone Project: Detecting Anomalies in Financial Transactions (Fraud Detection)

Problem Statement

Credit card fraud costs banks and customers billions each year. The project will build a machine learning model to detect fraudulent transactions in real-time, catching fraud early while keeping false alarms low.

Context

- Fraud is rare, making detection hard.
- Fraud patterns change over time.
- Current rule-based systems miss new fraud; ML can improve detection and explain results.
- Dataset: **Kaggle Credit Card Fraud Detection** (highly imbalanced, anonymized).

Stakeholders

- **Fraud Analysts** – need alerts they can trust and understand.
- **Risk/Finance Teams** – want to reduce fraud costs.
- **Engineering Teams** – care about fast, reliable systems.
- **Customers** – affected if fraud slips through or if good transactions are blocked.

Success Criteria

- **High Recall** → catch most fraud.
- **Reasonable Precision** → avoid too many false alarms.
- **AUC-PR & AUC-ROC** → good evaluation on imbalanced data.
- **Business Value** → show money saved compared to costs of false alerts.

Scope

In-scope: ML models (Logistic Regression, Random Forest, XGBoost, Isolation Forest), class imbalance handling, visualizations, model explainability.

Out-of-scope: full banking system integration, real customer data.

Constraints

- Data is anonymized (limits feature meaning).
- Highly imbalanced dataset.
- Limited compute for training.
- Privacy and compliance rules must be followed.

Data

- **Source:** Kaggle Credit Card Fraud Detection dataset.
- **Extra (future use):** bank transaction logs, device info, merchant risk scores.

Approach

1. Explore and visualize the data.
2. Handle imbalance with class weights/SMOTE.
3. Build and test models (LogReg, RF, XGBoost, Isolation Forest).
4. Evaluate using precision, recall, ROC/PR curves.
5. Explain results with SHAP/LIME.

Deliverables

- **GitHub repo** with: notebooks, code, results, and documentation.
- **Report:** problem, methods, results, insights.
- **Slide deck:** short presentation for business/technical audience.

Potential Insights

- Which features signal fraud the most.
- How early fraud can be detected.
- Trade-offs between catching more fraud vs. false alarms.
- Example explanations for flagged transactions.