Bank Transaction Fraud Detection Project Proposal

By: Lachezar Mitov

Introduction

What?

This project aims to develop a machine learning system that can accurately identify potentially fraudulent bank transactions in real-time. The system will analyse transaction patterns, and behavioural indicators to flag suspicious activities before they are completed, reducing financial losses for both banking institutions and their customers. The end product will be a predictive model with an accompanying dashboard for visualizing and explaining fraud predictions.

Why?

Financial fraud costs the global banking industry billions of dollars annually, with increasing sophistication of scam techniques making traditional rule-based detection systems inadequate. Machine learning offers the potential to identify subtle patterns indicative of fraud that human analysts or static rules might miss. As a student entering the AI field, this project addresses a significant real-world problem with measurable impact while demonstrating practical machine learning skills in classification, anomaly detection, and interpretable AI.

Who?

The primary stakeholders for this project include:

- **Banking representatives** (simulated stakeholder): The primary users of the system who need to review flagged transactions
- **Banking customers** (indirect stakeholders): Beneficiaries of improved fraud detection

For the purposes of this educational project, I will research to find as much information provided from banking professionals or financial security experts to understand the current challenges in fraud detection.

When?

Iteration 1: Foundation

- Research current fraud detection approaches
- Collect and prepare datasets
- Conduct stakeholder interviews
- Develop baseline models

Iteration 2: Model Development

- Feature engineering
- Implement and test multiple model architectures
- Address class imbalance issues
- Establish evaluation metrics

Iteration 3: Refinement

- Optimize model performance
- Implement explainability features
- Create visualization dashboard
- Gather stakeholder feedback

Iteration 4: Finalization

- Fine-tune based on feedback
- Develop final demonstration

- Complete documentation
- Present project results

How?

The end product will be a machine learning pipeline that:

1. Ingests transaction data
2. Processes and transforms the data through feature engineering
3. Applies ensemble models to detect anomalies and classify transactions
4. Generates risk scores with explanatory factors
5. Presents results through an interactive dashboard

The demonstration will showcase the system analysing batches of transactions in near real-time, highlighting how it identifies suspicious patterns and providing explanations for its decisions. The inferencing will happen through a Python application where new transaction data can be input either individually or in batches, with the model providing immediate classification results and confidence scores.

Domain Understanding

Central Research Question

How can machine learning techniques be effectively applied to identify fraudulent bank transactions with high accuracy while minimizing false positives that could negatively impact legitimate customer transactions?

Methods of Research

1. **Literature Review**: Examine current research papers and industry reports on financial fraud detection to identify common patterns and effective approaches.
2. **Data Analysis**: Explore available transaction datasets to understand the characteristics and patterns of both legitimate and fraudulent transactions.
3. **Competitive Analysis**: Evaluate existing commercial fraud detection solutions to identify industry best practices and potential areas for innovation.

Stakeholder Interview Plan

Key questions will include:

- What are the most challenging types of fraud to detect currently?
- How has fraud evolved in recent years?
- What information is most valuable when determining whether a transaction is fraudulent?
- What are the limitations of current detection systems?
- What would an ideal fraud detection system provide that current systems don't?

Analytic Approach

Target Variable

The primary target variable is binary: whether a transaction is fraudulent (1) or legitimate (0). However, the model will also output a fraud probability score (0-100%) to allow for different thresholds based on risk tolerance.

Type of Problem

This is primarily a supervised binary classification problem with imbalanced classes. The model must learn to distinguish between legitimate and fraudulent transactions based on labelled historical data.

However, the project also incorporates elements of:

- **Anomaly detection**: Identifying unusual patterns that deviate from normal transaction behaviour
- **Time series analysis**: Considering the sequence and timing of transactions
- **Cost-sensitive learning**: Accounting for the different costs associated with false positives versus false negatives

Proposed Modelling Approach

I will implement an ensemble approach combining multiple models:

1. **Transaction-level classifiers**: Gradient Boosting (XGBoost, LightGBM) and Random Forest models to evaluate individual transactions based on their features
2. **Account-level anomaly detectors**: Isolation Forests and Autoencoders to identify unusual patterns compared to the account's history
3. **Sequence models**: LSTM or GRU networks to capture temporal patterns in transaction sequences

These models will be combined using a meta-learner that weighs their predictions appropriately. The ensemble approach mitigates the weaknesses of individual models and improves overall robustness.

Data Requirements

Required Data Types

1. **Transaction Data**:
    - Transaction amount
    - Transaction type (purchase, withdrawal, transfer)
    - Merchant category code (MCC)
    - Transaction location and time
    - Device and channel used (online, in-person, mobile)
    - Transaction velocity (frequency of transactions)
2. **Account Information**:
    - Account age and type
    - Average balance and transaction history
    - Previous fraud flags or reports
3. **Customer Behaviour**:
    - Typical transaction patterns (locations, times, amounts)
    - Changes in behaviour patterns
    - Relationship with transaction recipients

Data Sources

Primary data will come from publicly available datasets:

- IEEE-CIS Fraud Detection dataset (Kaggle)
- Credit Card Fraud Detection dataset
- Synthetic Financial Dataset for Fraud Detection

These datasets will be supplemented with synthetic data generated to represent specific fraud patterns not well-represented in the public datasets.

Data Diversity Requirements

To ensure the model is robust and generalizable, the data should include:

- Diverse transaction types and amounts
- Various customer demographics and account types
- Seasonal patterns and time-of-day variations
- Both obvious and subtle fraud indicators

Privacy and Ethical Considerations

1. **Data Anonymization**: All personal identifiers must be removed or encrypted, even when working with public datasets.
2. **Bias Mitigation**: The model must be tested across different demographic groups to ensure it doesn't disproportionately flag transactions from certain populations.
3. **Transparency**: The system should provide explanations for fraud predictions to allow for human review and justification.
4. **False Positive Impact**: The model development must consider the negative customer experience caused by false fraud alerts.
5. **Regulatory Compliance**: The approach should align with relevant financial regulations and data protection standards like GDPR or CCPA.

Data Preparation Challenges

1. **Class Imbalance**: Fraudulent transactions typically make up less than 1% of all transactions. Techniques such as SMOTE, class weighting, and custom loss functions will be used to address this imbalance.
2. **Feature Engineering**: Creating meaningful features from raw transaction data will require domain knowledge and iterative refinement.
3. **Data Quality**: Missing values, inconsistent formats, and outliers will need to be addressed through robust preprocessing pipelines.

4. **Temporal Validation**: Standard cross-validation is insufficient; time-based validation approaches will be needed to reflect how the model would perform in production.