Final Project Proposal

Brain Bits

Problem Statement

Financial fraud is a huge problem in today's economy, where bad actors go to great lengths to exploit individuals for personal gain, causing immense damage to financial institutions. Tactics such as identity theft, account takeovers and phishing are all used to take control of an individual's financial account and commit fraudulent transactions. Traditional methods that use simple rule-based systems are insufficient to capture fraudulent transactions today. That is why we propose to use an AI-based fraud detection solution to identify financial transactions that are fraudulent.

Project Objectives

1. Develop predictive AI to detect fraudulent financial transactions.
2. Establish good model performance through metrics like accuracy and F1 score.
3. Reduce false positives.

Stakeholders

- Financial Institutions
- Governmental bodies: Canadian Anti-Fraud Centre (CFAC)
- Customers

Target audience

- Data scientists and AI researchers interested in financial fraud detection.
- Executives and managers in the finance industry evaluating ML solutions for fraud prevention.

AI Methodologies

We aim to test and select the best methodologies (evaluated on F1 score, accuracy, recall and precision) from:

1. DNNs
2. Tree-based methods (Decision Trees, Random Forest, etc.)
3. Multi Logistic Regression

Dataset Justification

Synthetic Financial Datasets for Fraud Detection by Edgar Lopez-Rojas

An open synthetic dataset of financial transactions developed using PaySim, a financial simulator. Each row is a financial transaction, with columns representing key details about the transaction like sender, payee, amount, etc. There are 11 features, with 6 million samples, so the size of the dataset is more than sufficient. Some team members have smaller machines, so the size may be too large, in which case we intend to create smaller datasets. The dataset is rich with class imbalances, different transaction types and large amounts of data which provide a good challenge for our team. Additionally, there are different features we need to strategize about how to encode effectively. One example is the step, or time, feature. Our team will have to discuss how to transform that feature to really leverage it in our models. The data set is also clearly relevant to financial fraud, and it is publicly open to everyone. All in all, this data set fits the criteria of size, complexity, relevance and availability, so we believe that we are justified in using it for this project.

Feasibility

Feasibility Type	**Assessment**
Technical	• We have technical skills to preprocess, develop, and evaluate DNNs, Tree-based and MLR algorithms. • Any challenges can be resolved with guidance from the instructor, alumni, outside sources (videos, books, etc.)
Economical	• Project has no cost, as all work is volunteered by students.
Organizational	• Stakeholders are supportive of this project. • Project insights have a place in stakeholder institutions.

Expected Outcomes

1. Higher detection accuracy than traditional fraud detection methods.
2. Lower false positives rate compared to traditional methods.
3. Reduction in loss due to financial fraud.
4. Increase in financial reputation and customer trust.