

Title: Fraud Detection in Financial Transactions

1. Problem Statement



What problem are you trying to solve?
What larger issues do the problem address?
Financial institutions struggle to identify and prevent fraudulent transactions due to the high volume of transactions and advanced fraud tactics. The goal is to develop a machine learning model that can detect fraudulent transactions in real-time.

2. Outcomes/Predictions



What prediction(s) are you trying to make?
Identify applicable predictor (X) and/or target (y) variables.
The project aims to create a predictive model that assigns a risk score to each transaction, indicating the likelihood of it being fraudulent. This score will be used to flag suspicious transactions for further investigation, reducing the incidence of fraud and minimizing false positives.

3. Value Propositions



What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

Reduced losses due to fraudulent activities, leading to higher profitability and a better reputation, and increased trust and confidence in customers.

4. Data Acquisition



Where are you sourcing your data from?
Is there enough data? Can you work with it?
The project will utilize historical transaction data provided by financial institutions, which include features such as transaction amount, time, location, payment method, and account history. Challenge would be ensuring the data is anonymized to protect customer privacy and dealing with class imbalance

6. Model Evaluation



How can you evaluate your model performance?
The model’s performance will be evaluated using metrics like precision, recall, F1-score, and the area under the ROC curve (AUC-ROC). Special attention will be given to minimizing false positives

5. Modeling



What models are appropriate to use given your outcomes?
A variety of machine learning models will be explored, including decision trees, random forests, and gradient boosting machines. Given the potential complexity of fraud detection, ensemble methods or deep learning models like neural networks might also be considered.

7. Data Preparation



What do you need to do to your data in order to run your model and achieve your outcomes?
Data cleaning, normalization, balancing the dataset