# Business Problem Document – Fraud Detection Model

Ben Duong – 20/01/2026

## 1. Business Context

Organizations processing large volumes of financial transactions face continuous risks from fraudulent activities. Although fraud cases represent only a small proportion of total transactions, their financial and reputational impacts are significant. Traditional rule-based detection systems struggle to adapt to evolving fraud patterns and often generate a high volume of false positives, increasing operational costs and investigation workload.

## 2. Problem Statement

The key business challenge is to identify fraudulent transactions within a highly imbalanced dataset, where fraudulent transactions account for approximately 3–5% of total volume. Using accuracy as the primary performance metric is misleading, as models can achieve high accuracy by predicting most transactions as non-fraud. The objective is to improve fraud detection while keeping false alerts at a manageable level.

## 3. Business Objectives

- Increase detection of fraudulent transactions (maximize Recall and F1 score)
- Reduce financial losses caused by missed fraud cases
- Control false positives to avoid overwhelming fraud investigation teams
- Provide a scalable and data-driven solution adaptable to changing fraud patterns

## 4. Analytical Approach

This project applies multiple machine learning approaches (Logistic Regression, SVM and XGBoost) and choose the best model based on metrics. XGBoost classifier is well-suited for tabular data and imbalanced classification problems. Class imbalance is addressed through class weighting techniques. Model performance is evaluated using Recall, Precision, F1-score instead of accuracy alone. The model produces fraud risk scores to support investigation prioritization rather than fully automated decisions.

## 5. Key Insights

- Accuracy alone is not a reliable metric for fraud detection
- There is a clear trade-off between fraud detection rate and false alert volume
- Decision threshold tuning can significantly improve recall depending on business tolerance
- Tree-based models effectively handle missing values and complex feature interactions

## 6. Business Impact

By prioritizing high-risk transactions, the model enables fraud teams to focus on the most suspicious cases first. This approach helps reduce financial losses while improving operational efficiency. The model is designed to support human decision-making and oversight, ensuring appropriate control over critical fraud cases.

## 7. Limitations and Assumptions

- Model performance is constrained by the limited number of fraud examples (Recall and F1 not achieved the benchmark score)
- Hardware restrictions required reducing dataset size and narrowing hyperparameter search ranges.
- Many fields are anonymized or encoded, limiting interpretability and feature engineering opportunities.

## 8. Next steps

- Stakeholder Presentation: Share findings with stakeholders to secure investment in upgraded computational resources and expanded datasets.
- Model Deployment: Integrate the fraud detection model into the bank's application with real-time notification systems for suspicious activity.
- Performance Monitoring: Establish weekly reviews of recall and F1 scores to track improvements, identify weaknesses, and refine the model continuously.
- Implement real-time scoring via API and develop monitoring dashboards for ongoing model performance tracking.