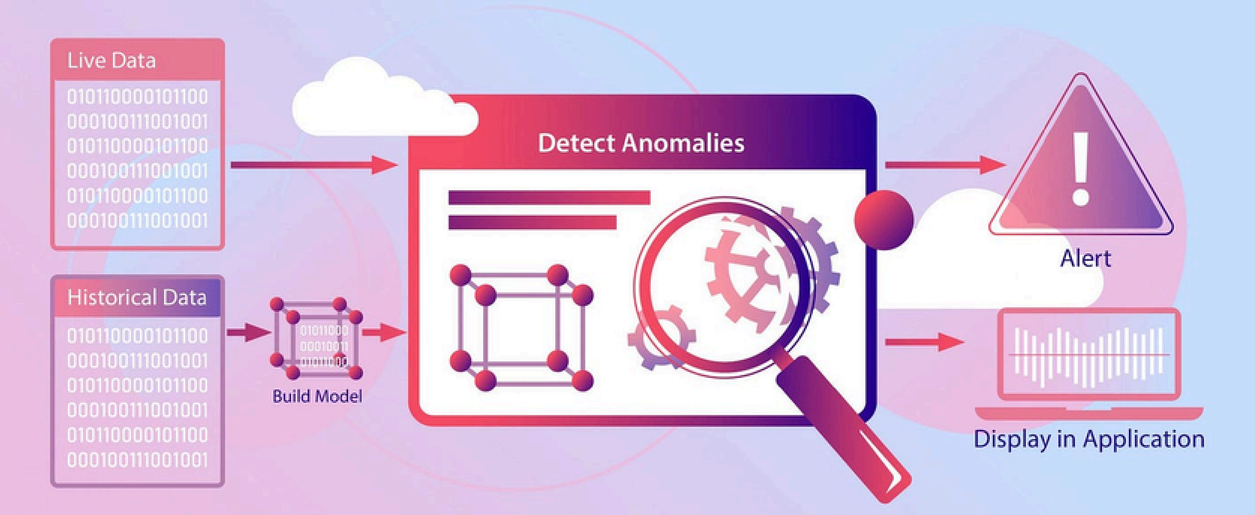
GDS PROJECT FINANCIAL FRAUD DETECTION

NEIL KANTH LAKHANI, SYED NISAR HUSSAIN



Introduction

Financial fraud poses severe risks to individuals, businesses, and system integrity. Detecting fraudulent activities like money laundering and identity theft is critical yet challenging due to transaction complexity. Our project leverages graph data science by modeling transactions as networks in Neo4j to identify fraudsters, it also extracts relevant graph features, and trains a machine learning model on these features for prediction.

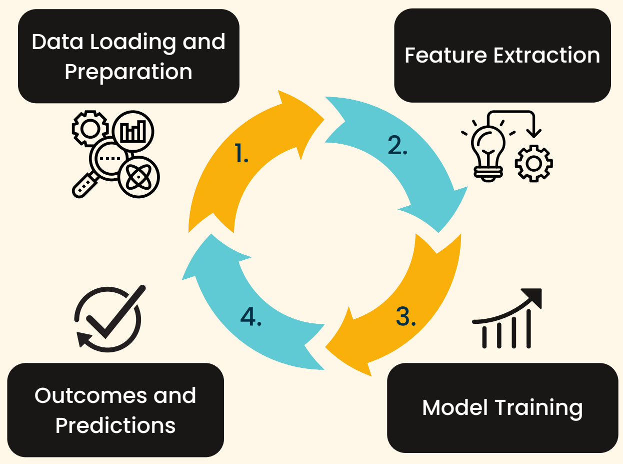
Problem Statement

Traditional fraud detection methods often struggle to keep pace with evolving fraud schemes and fail to capture the intricate relationships within transaction networks. We aim to develop a robust graph-based fraud detection system that can accurately identify fraudsters across interconnected entities using various graph features, and predict whether a transaction performed by a certain person is indicative of fraud or not .

Methodology

1. **Data Loading**: Loaded financial transaction dataset into Neo4j using a dump file.
2. **Graph Feature Extraction**: Ran Cypher queries to extract relevant graph features like Weakly Connected Components (WCC), Degree Centrality, PageRank, and Node Similarity using the Jaccard metric. These features helped us gain insights into Fraud Rings and identify First Party and Second Party Fraudsters.
3. **Model Training**: Leveraged extracted graph features to train a Random Forest Classifier model on to predict the likelihood of a Client being a fraudster.
4. **Outcomes and Prediction**: Used the trained ML model to make predictions on the test set of the dataset

Tools Used: Streamlit for app interface, Neo4j Python Driver and Neo4j GDS client for interacting with the database, scikit-learn for ML model training and predictions, Joblib for storing ML model



Data Model and Findings

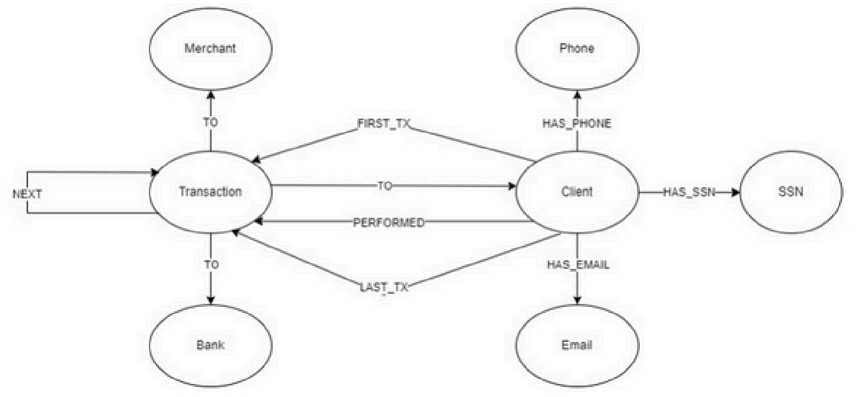
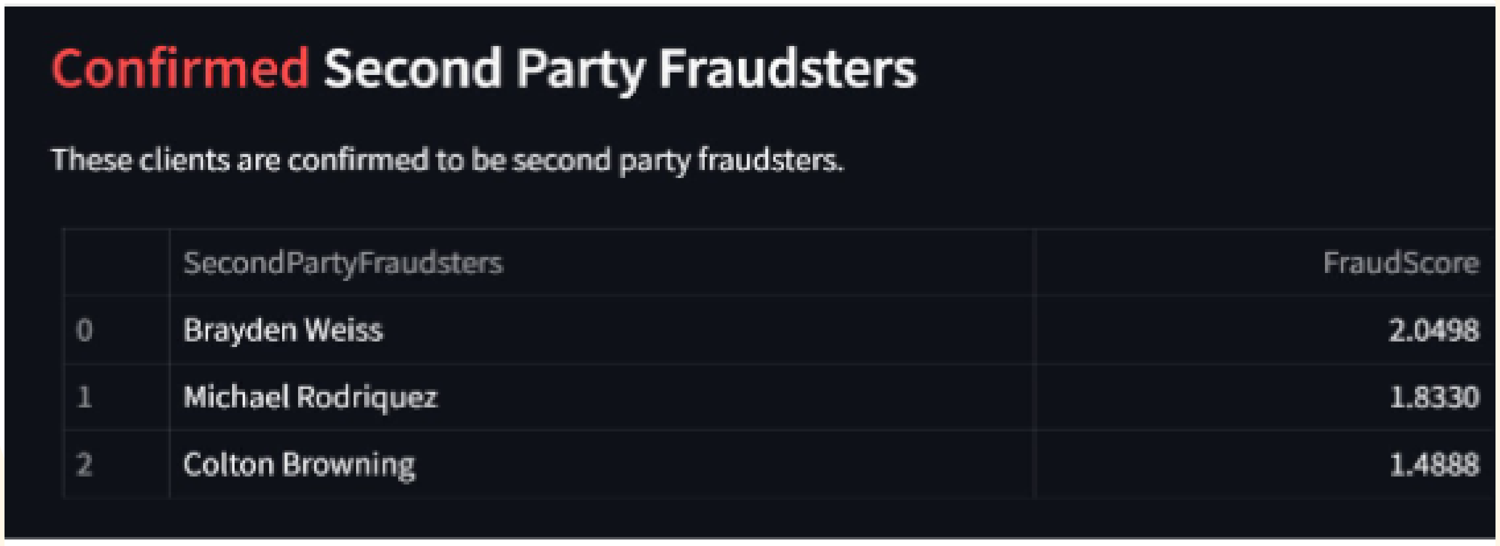


Figure 1: Labelled Graph Property Model

- The dataset contains information about transactions conducted by a Client with other Clients, Banks, or Merchants
- It contains information about Client's PII (Personally Identifiable Information)
- Learn more about it here: https://www.sisu.io/posts/paysim/



Confirmed Second Party Fraudsters

These clients are confirmed to be second party fraudsters.

	SecondPartyFraudsters	FraudScore
0	Brayden Weiss	2.0498
1	Michael Rodriquez	1.8330
2	Colton Browning	1.4888

Figure 2: Output of Fraud Detection App

Conclusion

- The graph-based fraud detection system provides a unique way to approach fraud detection in general.
- Having financial data represented as nodes and edges allows us to view insights into the data that are either not possible or very difficult to get through traditional methods.
- The performance of traditional ML models can further be enhanced by utilizing features extracted from the graph

Results

Graph Analysis

- From a total of 2433 clients in the database, 174 were suspected of being involved in First Party Fraud while using our graph queries, we classified 16 of them as confirmed fraudsters.
- Second Party Fraud is much more difficult to catch, we were able to successfully classify 3 Clients as Second Party Fraudsters

Machine Learning

- Our Random Forest Classifier achieved 93% accuracy after making predictions on the test set