**Problem statement**

Financial companies are looking for ways to detect fraudulent transactions so that customers are not charged for items they did not purchase. The purpose of this project is to implement machine learning models to detect fraudulent transactions.

**Context**  
Due to the private nature of financial data, there aren't many publicly accessible datasets for analysis. The dataset used in this project was generated using a simulator called PaySim, which is publicly available on Kaggle. A multinational mobile financial services company's private dataset was used to generate the dataset.

**Criteria for success**

Be able to classify fraudulent transaction as accurately as possible (AUC score with 80% or above)

**Scope of solution space**

The solution will be only applied to this dataset--Synthetic Financial Datasets For Fraud Detection

**Constraints**

it’s hard to acquire the publicly available dataset. The dataset was generated a simulator called PaySiml, the model built here is only applied to this dataset, however, we can apply similar methodologies whenever the new source data available

**Stakeholders**

Data science team

Financial department Leadership team

Client (financial companies)

**Data sources**

Kaggle datasethttps://www.kaggle.com/datasets/ealaxi/paysim1

There are 6362620 transactions, The data set has 11 attributes which include is

* Type of transactions
* Amount transacted
* Customer ID and Recipient ID
* Old and New balance of Customer and Recipient
* Time step of the transaction
* Whether the transaction was fraudulent or not