

**Fraud Detection in Mobile Payment Data**  
**Capstone 1 Project Proposal**  
**Springboard Data Science Career Track**  
Eric Cruz

Problem Statement: Online fraud is increasing and spreading rapidly across geographies and industries, and Mobile Payments represent a significant portion of the growth in both overall transaction and fraud rates. While attempting to detect and prevent fraud, the accuracy of the prediction models can have a significant impact on the ability to strike the right balance between true detection and false positives. The customer experience can be severely compromised by security measures enacted on the basis of a false positive. The fraud rate has a measurable impact on revenue, and new types and methods of fraud are evolving in response to successful detection and prevention efforts.

Dataset: Due to privacy concerns, there is little if any publicly available data for real transactions. PaySim.csv is a simulation of mobile money transactions with the objective to generate a synthetic transactional data set that can be used for research into fraud detection.

The dataset for this project is from the following link on Kaggle: <https://www.kaggle.com/ntnu-testimon/paysim1>

Clients: The intended clients are financial institutions and merchants who use mobile payments, and will incorporate the findings into a Fraud Detection and Prevention program which covers various types of fraud attacks.

Approach: As the dataset is synthetic and designed for fraud detection research, there are several approaches that have been published as Capstone projects using the same data. Part of the exploratory analysis will involve replicating results made from contributors on github. For example, due to the imbalanced nature (e.g. fraud proportion is small relative to the population) of the data, under-sampling and over-sampling methods will be explored. As the data does not contain features other than amounts, balances, account types, and fraud flags, the explanatory features must be engineered by combining characteristics of the given fields. For example, a target condition is to check for accounts that wash balances down to zero. After reviewing the approaches taken by others, the project approach will be developed and updated.