Maayan Hagbi  
Steigerweg 7

3006 Bern

[Maayan-hagbi@outlook.com](mailto:Maayan-hagbi@outlook.com)

Data Science Project

Friendly Fraud Detection Model for Enhancing Security in Payment Service Provider Transactions

Conceptual Design Report

11/23/2023

# Abstract

The threat of fraud, particularly in the form of friendly fraud, also known as chargebacks, is a pressing concern for online businesses. Despite its seemingly harmless name, friendly fraud occurs when the person who make the purchase, often the original credit card cardholder, falsely claims a transaction as fraudulent in order to avoid payment. Addressing this challenge effectively is significant challenge for many businesses where often the solution is to apply strict fraud prevention rules on an entire banks or even regions. According to a 2022 sift report, in 2022 online businesses experienced a 35% increase in dispute rates with the average disputed dollar amount increasing close to 16%.

Leveraging data from Fibonatix Ltd, an online payment service provider, our aim is to develop a machine-learning model adept at predicting the likelihood of transactions falling victim to friendly fraud. This initiative is designed to elevate the accuracy and effectiveness in identifying instances of friendly fraud, where seemingly normal customer behavior conceals potentially financially harmful intent.

Contents

[Abstract 2](#_Toc151643591)

[1. Project Objectives 4](#_Toc151643592)

[2. Methods 4](#_Toc151643593)

[3. Data 5](#_Toc151643594)

[4. Metadata 5](#_Toc151643595)

[5. Data Quality 5](#_Toc151643596)

[6. Data Flow 6](#_Toc151643597)

[1. Unsupervised framework: 6](#_Toc151643598)

[1. Supervised framework: 6](#_Toc151643599)

[7. Data Model 7](#_Toc151643600)

[8. Risks 7](#_Toc151643601)

[9. Conclusions 8](#_Toc151643602)

# Project Objectives

Fibonatix Ltd functions as a payment service provider (PSP), offering credit card processing services to a wide array of online businesses spanning across many industries. Notably, within their business operations, friendly fraud emerges as a significant and intricate challenge. The costs associated with friendly fraud extend beyond the direct loss of transactions fees, encompassing operational, processing, reputational costs.

Furthermore, to sustain its operational license, any Payment Service Provider (PSP) must comply with designated chargeback and fraud thresholds set by credit card companies. This underscores the potential harm that chargebacks can inflict on such businesses, emphasizing the imperative need for a robust fraud detection solution.

Our aim is to harness the transactional data supplied by Fibonatix to create a machine learning model that can predict the likelihood of a transaction being reported as fraudulent as close to the time of purchase as possible. This initial step in our initiative is geared towards creating business value by providing the company with sufficient time to investigate potential instances of fraud before they are formally reported. This will essentially give the business time to take the necessary measures to reduce overall risk.

# Methods

Both the analysis and modeling phases of this project will be conducted using the Anaconda Distribution for Python, specifically utilizing Jupyter Notebook and Google Collab for specific tasks as well as a Git repository for version control. The following libraries will be essential for various aspects of the project:

* Pandas: Used for data analysis and management.
* NumPy: Applied for analysis, data cleansing, and preparation for machine learning models.
* Matplotlib: Utilized for data visualization.
* SciPy: Employed for statistical analysis.
* Scikit-learn: Utilized for machine learning modeling and performance evaluations.
* TensorFlow: Applied for machine learning functionalities.
* Seaborn: Used for data visualization enhancement.
* Time: Employed for performance evaluation metrics.

# Data

Access will be granted to the Google BigQuery database, where data will be queried and manipulated to construct the modeling dataset. This is still a work in progress and will be addressed in the future.

# Metadata

Data needs to be evaluated still.

# Data Quality

Will be filled once access to data is granted.

# Data Flow

## Unsupervised framework:

## Supervised framework:

# Data Model

At its core, the model is designed to assess the likelihood of a transaction falling into one of two categories: potentially fraudulent or not fraudulent. We intend to conduct two distinct operations within this model.

Supervised learning framework:

While the original dataset is inherently supervised, our initial strategy revolves around treating the data as unsupervised. The aim is to unveil distinctive clusters that differentiate between fraudulent and genuine transactions.

To achieve this, we intend to employ Principal Component Analysis (PCA) for pattern detection and K-means for cluster identification. This unconventional approach to an initially supervised dataset seeks to reveal unique patterns and structures that might not be immediately evident within traditional supervised frameworks.

Supervised learning framework:

In the subsequent supervised learning framework, our focus shifts to employing advanced modeling techniques such as Logistic Linear Regression, Random Forest, or Neural Networks.

The objective is to enhance our ability to predict fraudulent transactions within the dataset. This approach leverages the labeled data to train models capable of capturing intricate relationships and patterns, contributing to a more accurate and targeted identification of fraudulent activities.

# Risks

The data provided by Fibonatix is of a sensitive nature, encompassing both customer identifying information (CIDs) and confidential business-related data. In order to mitigate the risk of potential data leaks, a precautionary step is taken during the data extraction stage where all CIDs are anonymized. Strict user rights protocols are implemented, allowing only viewing privileges. Any extraction of data for modeling purposes requires prior approval from Fibonatix, the rightful data owner. Additionally, to fortify Fibonatix's interests, a Non-Disclosure Agreement (NDA) and property rights agreement have been duly executed. These measures collectively ensure the stringent protection and confidentiality of Fibonatix's sensitive business data.

# Conclusions

Implementing machine learning models in an e-commerce setting to mitigate various forms of fraudulent activities, including friendly fraud, is a widespread practice in many businesses. Several Software as a Service (SaaS) providers specialize in precisely addressing this challenge. However, industry experts argue that the concept of a one-size-fits-all solution is fundamentally flawed. Each business possesses distinct requirements, making off-the-shelf fraud prevention software often suboptimal as they fail to account for the unique characteristics and needs of individual businesses.

This underscores the potential for us to develop a tailored fraud detection solution specific to the needs of Fibonatix, generating significant business value through effective fraud mitigation.

# Sources:

1. <https://pages.sift.com/rs/526-PCC-974/images/Report_2022_Q4_Digital-Trust-Safety-Index.pdf>