# Development of Predictive Model to detect fraudulent self-checkout transactions

## Free A hooded figure engaged in hacking using a laptop and smartphone in low light. Stock PhotoAnalytics Project:

**Authors**:

Ryan Kreis, Ranuthan Sittampalam, Gabriel Schnelli & Taoufik Brinis

**Lecturer**:

Prof. Andreas Reber

**City, Date:**

Basel, 05 March 2025

# Table of Contents

[Table of Content 2](#_Toc191998508)

[1 Heading 3](#_Toc191998509)

[1.1 Subheading 3](#_Toc191998510)

[Bibliography 4](#_Toc191998511)

[List of Figures 5](#_Toc191998512)

[List of Aids 6](#_Toc191998513)

# Introduction

## Business Understanding

## Data Understanding

In this paragraph, we look at the data as well as the features that are present in the dataset. After downloading the dataset and importing it we saw the following information:

The dataset contains 498121 entries. The key feature of the dataset for us will be the feature: **fraud** as it will be our target variable. We can see that of the 498121 entries, 474394 of the entries are **not** fraudulent, with 23727 entries being fraudulent.

In the dataset there are 10 unique attributes. These features can be seen with a brief description of their meaning below.

|  |  |  |
| --- | --- | --- |
| Column name | Description | Value range |
| trustLevel | A customer’s individual trust level. 6: Highest trustworthiness | {1,2,3,4,5,6} |
| totalScanTimeInSeconds | Total time in seconds between the first and last product scanned | Positive whole number |
| grandTotal | Grand total of products scanned | Positive decimal number  with maximum two  decimal places |
| lineItemVoids | Number of voided scans | Positive whole number |
| scansWithoutRegistration | Number of attempts to activate the scanner without actually scanning anything | Positive whole number or 0 |
| quantityModification | Number of modified  quantities for one of the  scanned products | Positive whole number or 0 |
| scannedLineItemsPerSecond | Average number of  scanned products per  second | Positive decimal number |
| valuePerSecond | Average total value of  scanned products per  second | Positive decimal number |
| lineItemVoidsPerPosition | Average number of item  voids per total number of  all scanned and not  cancelled products | Positive decimal number |
| fraud | Classification as fraud (1)  or not fraud (0) | {0,1} |

Table , Feature Descriptions

## Modeling

## Evaluation

## Deployment

|  |  |
| --- | --- |
| Hello | Da |
| D |  |
|  |  |

Table , testtable

# Bibliography

**There are no sources in the current document.**

# List of Figures

No table of figures entries found.

# List of Tables

[Table 1, testtable 9](#_Toc191999165)

# List of Aids

|  |  |  |
| --- | --- | --- |
| **Aid** | **Usage** | **Affected areas** |
| Private proofreading | Spell check | Entire work |
| Paid proofreading | Spell check | Entire work |
| DeepL Write | Optimization of individual text passages according to feedback from proofreading | Chapter 2, paragraphs 1-4  Entire chapter 5 |
| DeepL Translate | Translation of quoted text passages from Turkish to English | p. 16, paragraph 2 |
| Zotero | Preparation of in-text citations and reference list | In-text citations  Reference list |
| Litmaps | Literature search using "Seed Map" function | Entire work |
| Dream (by Wombo) | Creation of graphics | Figure 3, p. 10  Figure 8, p. 23 |
| ChatGPT | ChatGPT was asked to identify possible adjacent topic areas to the research question. The results were used to search for further literature and to delineate the research question as precisely as possible. | Entire chapter 1.2 (screenshots of prompts and answers in appendix B) |
| … | … | … |

# Appendix

The following documents are submitted with this documentation:

Data input/output files:

• marketing.csv

• euribor3m\_ratesMAY2008toNOV2010.csv

• OUTPUT Data Preparation.csv

Quarto scripts to be run in R Studio with its associated wordfile outputs:

• DataUnderstaningwithR.qmd

• DataUnderstaningwithR.docx • DataPreparationwithR.qmd

• DataPreparationwithR.docx

Orange workflows:

• Data Preparation FINAL.ows • Modelling FINAL.ows