Fraudulent Transactions Data

SAPIENZA
UNIVERSITÀ DI ROMA

Claudio Gheorghiu
Giuseppe Di Poce
Angelo Mandara
Enrico Grimaldi
Tito Tamburini

By the order of Peppi Blinders

Task and Motivation
What? why?

Task:

1. Clean the data.
2. Logistic regression for fraud detection.
3. Evaluate the model performance.
4. Make conclusions on key factors.

Motivation:

Digital transactions are exponentially increasing, so the security around these context is fundamental. Making digital payments more secure will increase the use of digital money and thus encourage the shift from cash to digital, and thus limit the black market.

Related work:
https://www.kaggle.com/code/dibyendupatra/fraudulent-transactions-data
https://www.kaggle.com/code/kartik2khandelwal/predicting-fraudulent-transactions
https://www.kaggle.com/code/sasakitetsuya/fraudulent-transactions-analysis-by-xgboost

Models,Tools
How?

Tentative material:
We decided to work on this kaggle dataset, related to auditing.
https://www.kaggle.com/datasets/chitwanmanchanda/fraudulent-transactions-data

Model:
We create a logistic regression model to classify fraudulent transactions.
Then we study the key features we used to classify the data, to find the most informative variables.

Methods:
- Pre-processing data
- Logistic regression
- Linear regression
- KNN
- Extreme Gradient Boosting

Analysis

We have to handle categorical and numerical values.
The dataset we are examining is very unbalanced and the output will be affected by this.

Benchmark

- Improve accuracy and recall in fraud detection.
- Evaluate the VIP (Variable Importance Plot) for features selection.
- Given the results, we try to develop a model for predicting fraudulent transactions for a financial company and use insights from the model to build an actionable plan.