



✕
Challenges
Knowledge Base
Job Board
Marc Bickel
Profile
Account Settings
Log Out

Transaction Monitoring Challenge



By [Credit-Suisse-LauzHack](#) [Private Challenge]

Ending
2 hours left (14:00 UTC)

0
Submissions
288
Views
4
Participants



1

FOLLOW

[Overview](#) [Leaderboard](#) [Discussion](#) [Dataset](#) [Submissions](#)

CREATE SUBMISSION

Introduction

Identifying suspicious clients undertaking abnormal behaviors, fraudulent activities or terrorist financing is a major challenge in Compliance at Credit Suisse. Nowadays, rule based scenarios are not relevant anymore and data science is seen as a game changer for the bank. In a world where there are more and more regulations and where the reputational risk is more and more at stake for Credit Suisse, we need you to find out the methods for the future and help us identify our suspicious clients using state of the art machine learning techniques and your smart analysis.

You will be given a set of clients with a list of features regarding their profiles and activity with the bank such as the total transaction amount during the year, or the age of the client.

You will then have to derive a list of suspicious clients to investigate for the financial crime compliance team - taking also into account not only the features but also regulators rules.

The submission will be evaluated both on the algorithmic approach, the scoring results and the novelty in visualizing the data and the model results.

The Challenge

LZHCK Bank is required to monitor some suspicious client activity by the regulators. Unfortunately, the scenarios are getting deprecated and returns more and more false positives - LZHCK is getting some pressure from the Regulators to have better results and the Investigation Team is asking for help from the data science team.

You are in charge of the data science team and you are required to provide a list of clients which are worth being investigated. To find this list of suspicious clients, you are given client data from the previous year :

Year 2016:

Client ID	SUSPICIOUS?	Turnover	Age	...
C00100	Yes	\$1600	32	...
C00200	No	\$12	55	...
C00300	No	\$1	null	...

You then have to guess the suspicious clients for the next year 2017.

Your Mission

- Along the 3 rounds, the investigation team requires you not only to provide your results, but also to help them understand your decision by building a tool or a visualisation interface or whatever you can imagine to help them make a decision on who has a fraudulent behaviour vs who doesn't.
- Round 1: Get the first batch - 2 hours
The bank committed to the regulators that a new program of transaction monitoring will be put in place in the horizon end-of-the-year. However, deadlines are tight and regulators expect to have some customers to investigate already.

Therefore, you have to provide a list of 200 customers that are worth being investigated. The Data Sourcing team and Transformation Team already provided you with some features, now it's up to you to decide!

- Round 2: More Data, More Batch - 4 hours

The first batch ran successfully. Now you are asked to provide a set of 1000 customers.

It's now your turn to decide which customers should the investigation team have a careful look at!

- Round 3: Business As Usual - Rest of the challenge

The regulators are very happy with the performance of the bank. You successfully managed to build a model to identify risky clients - so much so that the project receives more fund to maintain it for the next year. You are now given a new dataset and you have to provide a full list of customers that needs to be investigated. You can choose the size of the list but you have to take the following into consideration:

1 The cost of an investigation is inflexible and is I.

2 The code of a Fine by the regulators is F. Each undiscovered fraudulent behavior has a probability P of being discovered during the year by the regulators.

Therefore, you have to provide a full list of N suspicious clients with M suspicious customers that you missed, your final score will be:

$$S = I * N + F * P * M$$

$$I = 1'000 \text{ chf}$$

$$F = 6'000'000 \text{ chf}$$

$$P = 1\%$$

Good Luck!

Evaluation

The evaluation will be based on multiple aspects of your work:

- The scores from the 3 rounds of the hackathon
- The quality and originality of your innovative data visualization tool to help the investigation team to understand your results and have a deeper understanding of why you raised some specific alerts.
- The presentation you will make to the jury.

Rules

- As your results are being investigated by an internal team, you can only produce a limited number of datasets per round: round 1: 5 round 2: 2 round 3: 4
- If the dataset is wrongly formatted and raise an evaluation error, you still have the possibility to re-submit as many times as you want.

Prizes

At the end of the Hackathon.

Resources

You will find a training dataset of 1 million rows located here:

[train.csv.zip](#)

You will have to evaluate the following datasets:

[test.csv.zip](#)



© 2018 crowdAI



[FAQ](#) [Contact](#) [Privacy](#) [Terms of Service](#) [Cookie consent](#)