

UNIVERSITY OF TARTU
Institute of Computer Science
Data Science Curriculum

Andres Matsin, Kadri Onemar

**Financial Fraud Detection Using Machine Learning
Methods**

Master's Project (15 ECTS)

Supervisor: Kaur Lumiste, PhD

Tartu 2022

Financial Fraud Detection Using Machine Learning Methods

Abstract:

The project was created as a master thesis project and is a joint effort of two data science students from University of Tartu. This practical project is a part of implementing a machine learning-based approach in detecting financial fraud for a RegTech company — Salv. The aim of this project is to detect financial fraud from online transactions using supervised machine learning methods. The main deliverable of the project is a proof of concept for a binary classifier — a predictor that flags transactions to be either as suspicious or authentic.

Keywords:

Financial fraud detection, machine learning

CERCS: P176 - Artificial intelligence

Finantspettuse tuvastamine masinõppe meetoditega

Lühikokkuvõte:

Antud praktiline projekt on kahe Tartu Ülikooli andmeteaduse üliõpilase ühine magistritöö. Projekt on üks osa masinõppe meetodite rakendamisel finantspettuste tuvastamiseks ettevõttes Salv. Projekti eesmärk on juhendatud masinõppe meetodite abil tuvastada maksetehingutel põhinevaid finantspettusi. Peamiseks tulemuseks on binaarse klassifitseerija loomine, mis ennustab maksetehingu põhjal, kas tegu on pettusega või mitte.

Võtmesõnad:

Finantspettuse avastamine, masinõppe

CERCS: P176 - Tehisintellekt

About the project

The project was created as a master thesis project and is a joint effort of two data science students. This practical project is a part of implementing a machine learning-based approach in detecting financial fraud for a RegTech company — Salv¹.

The aim of this project was to detect financial fraud from online transactions using supervised machine learning methods. That aim was achieved by first researching the topic of financial fraud in order to find out which methods and supervised machine learning algorithms have been used in the domain and then begin building a pipeline that can be later used for financial fraud detection.

During the project two documents were created:

1. White paper document;
2. Technical report.

White paper in general is an informational document that addresses issues in different fields and how to solve them. A commercial white paper is issued by a company to educate readers and promote a specific company's solution or products.² The white paper created during the project (Appendix I) aims to educate readers about the financial fraud domain and give an overview of the practical project.

Technical report (Appendix II) was created to give an overview of the practical steps taken during the project and future plans.

The main deliverable of the project is a proof of concept for a binary classifier — a predictor that flags transactions to be either as suspicious or authentic.

¹ <https://salv.com/>

² <https://www-cdn.law.stanford.edu/wp-content/uploads/2015/04/Definitions-of-White-Papers-Briefing-Books-Memos-2.pdf>

Licence

Non-exclusive licence to reproduce the master's project and make the project public

I, **Kadri Onemar**,

1. grant the University of Tartu a free permit (non-exclusive licence) to:
reproduce, for the purpose of preservation, including for adding to the DSpace digital archives
until the expiry of the term of copyright, my master's project

Financial Fraud Detection Using Machine Learning Methods,

supervised by Kaur Lumiste,

2. I grant the University of Tartu the permit to make the project specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work from **09/05/2027** until the expiry of the term of copyright,
3. I am aware that the author retains the rights specified in points 1 and 2.
4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Kadri Onemar
17/05/2022