Report on Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	**Petra Kohoutová**
Advisor:	**Ladislav Krištoufek**
Title of the thesis:	**Comparative Analysis of Outlier Detection Models for Transaction Monitoring**

OVERALL ASSESSMENT *(provided in English, Czech, or Slovak):*

Short summary

This thesis investigates various machine learning models applied to transaction monitoring, a key area in anti-money laundering (AML) efforts. Specifically, the thesis compares models such as Isolation Forest, K-Means, and COPOD for their outlier detection capabilities in financial transactions. The research makes a significant contribution to understanding how different models perform on both artificial and real-world data, focusing on the ability to detect fraudulent activities. The thesis employs advanced machine learning methods, contributing both to practical applications and academic knowledge in the field of financial crime detection.

Contribution

The thesis contributes significantly to both theory and practice by conducting a comparative analysis of cutting-edge outlier detection techniques. The topic of applying machine learning models to AML in transaction monitoring is novel and timely. The thesis not only introduces models like Isolation Forest and COPOD but also evaluates their performance on real-world data from a medium-sized financial institution, adding relevance to the findings. The thesis addresses a critical issue in financial monitoring, and its applicability in combating fraud makes it highly impactful.

Methods

The methods employed in the thesis are robust, making excellent use of machine learning techniques. The comparative analysis involves rigorous testing of the models, with a clear explanation of each algorithm's strengths and weaknesses. The use of both artificial and real-world data allows for a comprehensive evaluation of the models under different conditions. On top of that, the implementation of certain more advanced algorithms, such as COPOD, is particularly well-done, given the complexity of transaction monitoring datasets. The selection of models is appropriate, and the analysis is thorough, with a variety of performance metrics used to ensure reliable results.

Literature

The literature review is comprehensive, covering both foundational and cutting-edge studies in machine learning and outlier detection. The thesis effectively places the research within the context of previous work, noting gaps in the literature regarding the application of these models to transaction monitoring. However, the review could have been expanded to include more recent advances in outlier detection techniques in related fields such as cybersecurity or healthcare. Nevertheless, the depth of the review is sufficient for understanding the thesis's contributions.

Manuscript form

The manuscript is well-organized and clearly written, with logical progression from the problem statement to the conclusion. The figures and tables are well-designed and support the text effectively, making complex concepts easier to understand. The formatting is professional, and the use of LaTeX ensures that the thesis is visually appealing. There are no major typographical

Report on **Master Thesis**

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	**Petra Kohoutová**
Advisor:	**Ladislav Krištoufek**
Title of the thesis:	**Comparative Analysis of Outlier Detection Models for Transaction Monitoring**

or grammatical errors, though a few sentences could have been phrased more clearly. Overall, the manuscript meets high academic standards.

Overall evaluation and suggested questions for the discussion during the defense

This is an excellent thesis, both in terms of content and form. The topic is novel and relevant to the financial sector, and the use of machine learning models in transaction monitoring is a valuable contribution. The methods are rigorous, and the analysis is thorough, ensuring that the findings are robust and applicable in practice. The thesis is highly likely to be well-received both academically and in practical financial applications.

Suggested Questions:

- How do you think the performance of the outlier detection models would change if applied to larger datasets, given the scalability challenges associated with real-time transaction monitoring?
- Given the complexity of COPOD, how would you justify its computational cost compared to simpler models like Isolation Forest or K-Means in a real-world application?
- What steps would you take to ensure that the models you've tested continue to perform well as transaction patterns evolve over time?
- Could you elaborate on how the synthetic data you generated for model testing was designed to reflect real-world transaction behavior, and what limitations this might introduce?

The TurnItIn analysis has not uncovered any suspicious practices (score of 16% with no large portions of overlapping text).

SUMMARY OF POINTS AWARDED *(for details, see below):*

CATEGORY		**POINTS**
Contribution	*(max. 30 points)*	29
Methods	*(max. 30 points)*	30
Literature	*(max. 20 points)*	18
Manuscript Form	*(max. 20 points)*	19
TOTAL POINTS	*(max. 100 points)*	**96**
GRADE	**(A – B – C – D – E – F)**	**A**

NAME OF THE REFEREE: Ladislav Kristoufek

DATE OF EVALUATION: 10 September 2024

prof. PhDr. Ladislav Krištoufek, Ph.D.

Digitally signed by prof. PhDr. Ladislav Krištoufek, Ph.D.
Date: 2024.09.10 14:00:13 +02'00'

Referee Signature