Project proposals in Financial Innovation track

Below are five proposed thesis projects. You may use one of these project proposals for your thesis. You are also welcome to choose a different topic for your thesis research, in consultation with your teacher.

1. Enhancing Money Laundering Detection with Semi-Supervised Learning

1.1 Background, assignment, and key objectives

Financial institutions face significant challenges in effectively identifying and preventing money laundering, the illegal process by which illicit money enters legal financial channels [1, 2, 3]. Although machine-learning techniques are used in anti-money laundering (AML) systems to detect fraudulent transactions, traditional methods such as supervised learning and unsupervised learning may have inherent limitations [4]. The methods are mainly limited by unbalanced datasets not representing the diverse spectrum of criminal activity associated with money laundering [5], but also due to the high dimensionality in transaction data [6]. Previous research suggests that combining supervised and unsupervised learning through semi-supervised approaches is promising for improving fraud detection [4].

1.2 Research activities

A tool development approach based on e.g., the use of Machine Learning models (unsupervised, supervised, and semi-supervised learning paradigms), NLP, and big data analysis.

Data source SynthAML:

- See description of dataset in reference [7] (https://www.nature.com/articles/s41597-023-02569-2)
- The tables of the dataset are available for download from Figshare (https://springernature.figshare.com/collections/SynthAML_a_Synthetic_Data_Set_to_Benchmark_Anti-Money_Laundering_Methods/6504421/1).

1.3 Contact lecturer

Kees van Montfort, PhD

1.4 References

[1] Cox, D. (2014). Handbook of anti-money laundering. John Wiley & Sons.

[2] Ofoeda, I., Agbloyor, E.K., Abor, J.Y., and Osei, K.A. (2022). Anti-money laundering regulations and financial sector development. International Journal of Finance & Economics, vol. 27, no. 4, pp. 4085–4104, 2022.

[3] Raweh, B.A., Erbao, C., and Shihadeh, F. (2017). Review the literature and theories on anti-money laundering. Asian Development Policy Review, vol. 5, p. 140–147.

[4] Chen, Z., Soliman, W.M., Nazir, A., and Shorfuzzaman, M. (2021). Variational autoencoders and wasserstein generative adversarial networks for improving the anti-money laundering process. IEEE Access, vol. 9, pp. 83762– 83785.

[5] Altman, E., Egressy, B., Blanuvsa, J., and Atasu, K. (2023). Realistic synthetic financial transactions for anti-money laundering models. ArXiv, vol. abs/2306.16424.

[6] Bakry, A.N., Alsharkawy, A.S., Farag, M.S., and Raslan, K.R. (2024). Combating financial crimes with unsupervised learning techniques: Clustering and dimensionality reduction for anti-money laundering. Al-Azhar Bulletin of Science, vol. 35.

[7] Jensen, R.I.T., Ferwerda, J., Jørgensen, Jensen, E.R., Borg, M., Krogh, M.P., Jensen, J.B., and Iosifidis, A. (2023). A synthetic data set to benchmark anti-money laundering methods, Sci Data, vol. 10, no. 661.

2. Graph Neural Networks Applied to Money Laundering Detection

2.1 Background, assignment, and key objectives

Money laundering is identified as the process through which illegally obtained money is moved through legitimate channels to obscure its origin [1]. This process enables the financing of various criminal activities, making the fight against money laundering an aspect of global security and economic integrity.

The operational, technical and behavioral challenges inherent in current Anti Money Laundering (AML) methodologies necessitate a shift towards more innovative solutions. The potential of advanced technologies, such as AI and machine learning (ML) to enhance AML capabilities already have shown to streamline detection processes, reduce false positives and improve operational efficiency.

Among these AI methodologies, Graph Neural Networks (GNNs) stand out due to their unique capability to model and learn from the graph-structured data typical of financial transactions. GNNs provide a deeper insight into the networks involved in money laundering by analyzing the entire web of financial interactions. This not only potentially enhances the accuracy of identifying illicit activities but also minimizes false positives by considering the contextual nuances of transactions [2]. This research aims to examine the effectiveness and integration of GNNs within AML frameworks.

2.2 Research activities

A tool development approach based on e.g., the use of Machine Learning models (unsupervised, supervised, and semi-supervised learning paradigms), NLP, and big data analysis.

Data source:
E. Altman (2022). “IBM Transactions for Anti Money Laundering (AML):
Use for Foundation Models, GNNs, and More”.
Kaggle dataset, 2022. Available: https://www.kaggle.com/datasets/ealtman2019/ibm-transactions-for-anti-money-laundering-aml/data?select=HI-Large\ Trans.csv

2.3 Contact lecturer

Kees van Montfort, PhD

2.4 References

[1] United Nations, “1988 united nations convention against illicit traffic in narcotic drugs and psychotropic substances,” United Nations Treaty Series, 1988, available online.

[2] Johannessen, F. and Jullum, M. “Finding money launderers using heterogeneous graph neural networks,” Jul 2023, arXiv:2307.13499v1 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2307.13499

3. Natural language processing in detecting fraud on debit card transactions

3.1 Background, assignment, and key objectives

Fraud in financial transactions continues to be a growing challenge affecting both customers and financial institutions. Despite numerous efforts to combat this problem, fraudsters constantly develop new methods and techniques to conceal their fraudulent activities. As a result, banks and other financial institutions have embraced technological advancements to implement more effective methods for detecting and preventing fraudulent activities. This thesis focuses on the development of a Natural Language Processing (NLP)-based approach model for detecting financial fraud in debit card transactions [1][2].

3.2 Research activities

A tool development approach based on e.g., the use of Machine Learning models (unsupervised, supervised, and semi-supervised learning paradigms), NLP, and big data analysis.
Data source:
Data are available on https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.

3.3 Contact lecturer

Kees van Montfort, PhD

3.4 References

[1] Fernández Rodríguez, J., Papale, M., Carminati, M., and Zanero, S. (2022). A Natural Language Processing Approach for Financial Fraud Detection. Available: https://re.public.polimi.it/bitstream/11311/1224432/1/paper10.pdf

[2] Di Milano, P., Zanero, R., Carminati, M., and Fernández, J. (2020). A Natural Language Processing Approach to Fraud Detection. Available: https://www.politesi.polimi.it/bitstream/10589/170662/3/thesis.pdf

4. Assessment of SME Credit Risk

4.1 Background

Small and medium enterprises (SMEs), the cornerstone for economic growth in many Western European countries, encounter difficulties in financing their activities as the assessment of the corresponding credit risk often becomes a serious constraint. The lack of assets to secure the loan (collateralization), a proper credit history, and no financial statements, potentially lead to higher default risk and limiting access to credit (Dastile et al., 2020; Guégan & Hassani, 2018; Nwachukwu, 2024).

In addition, information asymmetry leads to an imbalance of power in transactions because the parties engaged in the transaction have access to different levels of information, e.g., actuaries and insured, borrowers and sellers, which in the end leads to inefficiencies in business transactions.

Peer-to-peer (P2P) lending to either consumers (B2C) or businesses (B2B) bypasses conventional intermediaries, processes, and overall requirements to contact borrowers and lenders. Loans are granted by lenders, not by these lending platforms, which 'only' transfer the credit risk to the lenders. There is no need, therefore, to meet capital requirements as banks do. But, what about the risk for these lending platforms and the lenders when onboarding the customers?

There is a need to systematically evaluate different ML models on their ability to predict default risk accurately and transparently. This evaluation should include both performance metrics and interpretability assessments to provide a more comprehensive understanding of each model's trade-offs and strengths (Chang et al., 2024; Mestiri, 2024; Robisco & Carbó, 2022).

4.2 Assignment/Key Objectives

Improve SME credit risk assessment:

- Systematically evaluate different ML models on their ability to predict default risk accurately and transparently.
- Lower the risk exposure and reduce potential losses by financial institutions granting the loans.
- Increase credit supply and alleviate credit constraints by the SMEs.

4.3 Research activities

A phenomenon investigation based on literature study, and data analytics on the use of publicly available databases. The analyses use Machine Learning models (unsupervised, supervised, and semi-supervised learning paradigms), NLP, and big data analysis.
Data source: Loan Prediction Dataset of Kaggle; contains data on loan applications, including approved and rejected loans.

4.4 Contact lecturer

Kees van Montfort, PhD

4.5 References

Chang, V., Sivakulasingam, S., Wang, H., Wong, S. T., Ganatra, M. A., & Luo, J. (2024). Credit risk prediction using machine learning and deep learning: A study on credit card customers. Risks, 12(11), 174. https://doi.org/10.3390/risks12110174

Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing, 91, 106263. https://doi.org/10.1016/j.asoc.2020.106263

Guégan, D., & Hassani, B. (2018). Regulatory learning: How to supervise machine learning models? An application to credit scoring. The Journal of Finance and Data Science, 4(3), 157–171. https://doi.org/10.1016/j.jfds.2018.04.001

Mestiri, S. (2024). Credit scoring using machine learning and deep learning-based models. Data Science in Finance and Economics, 4(2), 236–248. https://doi.org/10.3934/DSFE.2024009

Nwachukwu, G. (2024). Enhancing credit risk management through revalidation and accuracy in financial data: The impact of credit history assessment on procedural financing. International Journal of Research Publication and Reviews, 5(11), 631–644. ISSN 2582-7421. https://www.ijrpr.com

Robisco, A., & Carbó, J. M. (2022). Measuring the model risk-adjusted performance of machine learning algorithms in credit default prediction. Financial Innovation, 8(1), Article 366. https://doi.org/10.1186/s40854-022-00366-1

5. Cybersecurity in Fintech

5.1 Background

FinTech platforms are frequently targeted by cyberattacks due to their large amounts of financial data and digital transactions ([1],[2]). Common attacks include malware (e.g., Trojans, ransomware), phishing (credential abuse, social engineering), API attacks, and account takeovers.

Using datasets that describe malware and phishing behavior allows us to model patterns in attack techniques relevant to FinTech contexts [3].

5.2 Assignment/Key Objectives

Recognizing patterns in cyberattacks on FinTech platforms and developing countermeasures.

5.3 Research activities

Literature overview of cyber threats in FinTech and relevant machine learning techniques in cybersecurity.
Use machine learning models to identify patterns in cyberattacks.
Analyze the impact of different security measures using statistical techniques.
Kaggle datasets:

- Malware Classification Dataset
  Contains data on malware attacks, including characteristics of malicious software.
- Phishing Websites Dataset
  Contains data on phishing websites, including characteristics that can be used for detection.

5.4 Contact lecturer

Kees van Montfort, PhD

5.5 References

[1] Aleroud, A. & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. Computers & Security, 68(9). DOI:10.1016/j.cose.2017.04.006.

[2] Javaheri, D., Fahmideh, M., Chizari, H., Lalbakhsh, P., & Hur, J. (2023). Cybersecurity threats in FinTech: A systematic review. Expert Systems with Applications, 241(8), 122697. DOI:10.1016/j.eswa.2023.122697.

[3] Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. Computers & Security, vol. 81, March, 123-147.