Introduction to the minitrack on fraud detection using machine learning at HICCS 2022

Wouter Verbeke
KU Leuven
wouter.verbeke@kuleuven.be

María Óskarsdóttir
Reykjavík University
mariaoskars@ru.is

Tim Verdonck
KU Leuven
University of Antwerp
tim.verdonck@uantwerpen.be

Bart Baesens
KU Leuven
University of Southampton
bart.baesens@kuleuven.be

Fraud, defined as the wrongful or criminal deception intended for financial or personal gain, is a considerable concern for governments and companies across industries, and notably in - but not limited to - banking, insurance, telecommunication and health care. Fraud is characterized as an uncommon, well-considered, time-evolving, carefully organized and imperceptibly concealed crime which appears in many different types and forms [1], e.g., corruption, money laundering, tax evasion, identity theft, and credit card fraud. Although a variety of advanced systems for preventing and detecting fraud are put in place, fraud remains a substantial challenge, with fraudsters continuously adapting their methods and probing security systems for weaknesses [2].

Over the last two decades, substantial advances in developing data-driven approaches for detecting fraud have been made, for instance, by leveraging machine learning as a powerful instrument to learn patterns of features that are indicative for fraud from historical data and to automatically detect fraudulent instances [3]. Despite these advances and a rapidly growing body of literature presenting cases, methods and experimental results, still a variety of open research questions and pressing issues remain to be addressed. Innovative methods are to be further developed, tested and fine-tuned for increasing the power of fraud detection systems, with the eventual aim to eradicate fraud in the most effective and efficient manner.

For scientists and practitioners to share insights and discuss on the latest developments in the field, this minitrack called for papers on fraud detection using machine learning or predictive analytics techniques across industries and making use of diverse methods and data sources. Authors were encouraged to submit papers that focused on typical challenges related to the development of machine learning systems for fraud detection, such as extreme class imbalance, unlabeled data sets, or the dynamic nature of fraud (i.e., concept drift) [4]. Also, papers were welcomed exploring the use of new and diverse sources of data, such as text, images and transactions, or proposing novel approaches to (pre-) process data, for example using network analysis [5]. Additionally, papers on cost-sensitive learning, evaluation or decision-making for minimizing fraud losses [6] as well as papers that focus on managerial aspects of developing, implementing and maintaining fraud machine learning systems were invited [7].

Four papers were eventually accepted for presentation in this minitrack. The selected papers display a broad diversity in terms of the application domain, the data and the methods that are used.

In the paper entitled 'False Positives in Credit Card Fraud Detection: Measurement and Mitigation', the authors propose a new method for assessing the cost of false positives and evaluate several state-of-the-art fraud detection classifiers using this method. Additionally, they investigate the effectiveness of ensemble learning, as previous work indicated that a combination of diverse, individual classifiers can improve performance. Their results show that cost-based evaluation yields valuable insights for practitioners and that an ensemble learning strategy can cut fraud costs by almost 30%.

In the paper entitled 'A Model for Detecting Accounting Fraud by using Machine Learning', the authors propose a machine learning method that extends upon XGBoost and that enables to predict signs of financial statement fraud by combining accounting domain knowledge and machine learning. The proposed method is empirically evaluated and benchmarked on a large data set of financial statements.

In the paper entitled 'Detecting potential money laundering addresses in the Bitcoin blockchain using unsupervised machine learning', the authors analyze methods that can be used to detect money laundering in Bitcoin using machine learning, with the aim to empower investigators to more accurately and efficiently determine whether a suspicious activity concerns money laundering.

In the paper entitled 'Short Attack: The Roles of Forensic Accountants and Artificial Intelligence in Detecting Fraud', the authors present three fascinating

URI: https://hdl.handle.net/10125/79524
978-0-9981331-5-7

cases of short attacks that alleged target companies committed fraud. They discuss on the role that forensic accountants play for both short seller companies and target companies in these short attacks, and how they can efficiently generate valuable information from large volumes of data to detect fraud and support their claims.

We believe these papers provide a cross-sectional view on the state-of-the-art in fraud detection, with a range of valuable take-aways for practitioners and scientists alike. Moreover, they open new opportunities for further research and present methods results for other researchers to build and improve upon.

Acknowledgements We thank the authors of these papers to share their valuable expertise, insights and experience, and for contributing to the body of knowledge on developing powerful fraud detection systems.

We thank the reviewers of these papers for providing useful comments to help the authors improving their work.

References

[1] V. Van Vlasselaer, T. Eliassi-Rad, L. Akoglu, M. Snoeck, and B. Baesens, "Gotcha! network-based fraud detection for social security fraud," *Management Science*, vol. 63, no. 9, pp. 3090–3110, 2017.

[2] B. Baesens, V. Van Vlasselaer, and W. Verbeke, *Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection*. John Wiley & Sons, 2015.

[3] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Statistical science*, vol. 17, no. 3, pp. 235–255, 2002.

[4] J. Vanhoeyveld, D. Martens, and B. Peeters, "Value-added tax fraud detection with scalable anomaly detection techniques," *Applied Soft Computing*, vol. 86, p. 105895, 2020.

[5] M. Óskarsdóttir, W. Ahmed, K. Antonio, B. Baesens, R. Dendievel, T. Donas, and T. Reynkens, "Social network analytics for supervised fraud detection in insurance," *Risk Analysis*, 2021.

[6] S. Höppner, B. Baesens, W. Verbeke, and T. Verdonck, "Instance-dependent cost-sensitive learning for detecting transfer fraud," *European Journal of Operational Research*, vol. 297, no. 1, pp. 291–300, 2022.

[7] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert systems with applications*, vol. 41, no. 10, pp. 4915–4928, 2014.