Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-10 (2022) 406–411

Fraud Detection in Supply Chain with Machine Learning

Mahdi Seify*, Mehran Sepehri*[1], Amin Hosseinian-Far*, Aryana Darvish**
*School of Management and Law, University of Northampton, UK NN15PH
**Vice-Provost Office, University College London, London, UK WC1E 6BT

Abstract: A variety of fraud in Supply Chains may be detected either in physical parts or in cyber data. We use supervised machine learning to detect various fraud and misinformation in supply chains. The study is based on a car manufacturer concerned with increasing fraud, ranging from fraudulent invoices to inflated prices. Big data is provided for pattern recognition. A macro-level code is presented with actual algorithms developed in Python. The research is continuing, while the current work is presented with promising results.

Keywords: Machine Learning, Supply Chain, Fraud, Detection, Pattern Recognition.

1. INTRODUCTION

Fraud detection means facing the threat of fraud that facilitates revealing and preventing losses (Aziz and Dowling, 2019). In the past, the fraud detection process was mainly based on audit techniques that were with waste of time techniques (Abdallah et al., 2016). However, in recent years, organizations depended on artificial intelligence (AI) technology represented in machine learning (ML) systems (Aziz and Dowling, 2019).

Thus, Behdad et al. (2012) have stated another definition of nowadays fraud detection as it *"tries to discover and identify fraudulent activities as they enter the systems and report them to a system administrator"*. This means that it became an automated process by using ML systems instead of manual techniques. The machine learning systems are implemented by organizations to detect fraud or prevent it by depending on a process of data analytics (Agrawal, 2018).

What is more is that the use of ML systems has increased in organizations because of the significant benefits of its application such as increasing fraud detection success and accuracy. Therefore, examining the efficiency of different types of ML techniques and systems has been widely studied by researchers to support the fraud detection organizations with the highest efficient ML system to detect fraud (Bologa et al., 2013). The use of this modern technology is reflected positively on the performance of organizations which is the main driver of success (Borges et al., 2020). Further, the development of modern technology in the organizations is a source for the achievement of strategic organizational performance objectives (Oh and Pinsonneault, 2007).

2. MACHINE LEARNING

An applied form of AI technologies is machine learning (ML). ML can is defined as using the machine to detect the different types of data that have different patterns in an automated way by computer systems and applications (Murphy, 2012 as cited in Borges et al., 2020). Agrawal (2018) explains the ML as the automated systems which facilitate the prediction process instead of human prediction efforts. ML systems are based on a big data analytics process to perform its prediction task.

According to Baesens et al. (2016), big data are used for prediction analytics that can be found in machine learning systems. This is supported by the previous statement in the previous section by different scholars that big data is the generator of AI and its different technologies. This means that the ML is a system for big data analytics to predict something (Borges et al., 2020). Furthermore, Yaram (2016) states that the learning of ML systems occurs from the input data that are inserted in the system, then analyzed and make the prediction.

There are two types of ML systems that are commonly used. Most of the scholars have agreed that the two types of ML that are implemented at organizations are supervised ML and unsupervised ML (Dwivedi et al., 2019; Abdullah et al., 2016; Eshghi and Kargai, 2019; Hand et al., 2008; Glancy and Yadav, 2011). However, some papers included a third type such as reinforcement ML (Kaplan and Haenlein, 2019) and semi-supervised ML (Abdullah et al., 2016).

The first type of ML is supervised ML. According to Abdullah et al. (2016), supervised ML is the most common type and it differs with its requiry to label the data. Furthermore, this type of ML requires training classifiers. The data inserted in the ML are compared to labeled training data already have been put in to identify an output (Aziz and Dowling, 2019).

The main positive aspect of this type is that the output of the automated system is helpful for human use and classification (ibid.). However, this type might be difficult to implement in where the amount of data is not able to be labeled because of their high volume (ibid.). The algorithms that are used in the supervised ML are not the same as the unsupervised ML. For example, the classification algorithms that are used here are many such as support vector machine (SVM), artificial neural network, K-nearest neighbours and Naive Bayes in addition to the regression algorithms (ibid.).

The development of the ML has different aspects that need to be considered. As stated previously, the data are the generator of the ML system (Agrawal, 2018; Kaplan and Haenlein, 2019; Dwivedi et al., 2019; Shin et al., 2020) in addition to the algorithms as a second generator (Choudhury et al., 2020).

[1] Corresponding author: Mehran.Sepehri@Northampton.ac.uk

10.1016/j.ifacol.2022.09.427

3. FRAUD DETECTION

One of the issues using AI is fraud. Fraud nowadays became an issue that has a significant negative impact. According to Levi and Burrows (2008, as cited in Aral et al., 2012), fraud is a technique that leads to an illegal profit for a fraudster or an illegal loss, in all sectors and fields such as insurance, customs, corporate, social security, and supply chain fraud (Pourhabibi et al., 2020; Ngai et al., 2011; Galeotti et al., 2020; Glancy and Yadav, 2011; Goode and Lacey, 2011; Abdallah et al., 2016).

Furthermore, the main reason for classifying the fraud as a crime and a crucial issue that needs to be seriously faced is the high financial losses as a result of this crime (Galeotti et al., 2020; Quah and Sriganesh, 2008). Moreover, it has an impact on the society values, law and economy (Abdallah et al., 2016). Therefore, many organizations, authorities, corporates and businesses are trying to detect the fraud and fraudulent by enhancing their detection efforts (Craja et al., 2020; Goode and Lacey, 2011; Cecchini et al., 2010).

The classifiers, algorithms and data mining techniques are all different names for the same categories which are the techniques that are used in the supervised ML systems (Bhattacharyya et al., 2011; Triepels et al., 2018; Craja et l., 2020). The data mining techniques are the algorithms that generate the process of data analytics in the fraud detection ML (Bhattacharyya et al., 2011). Firstly, support vector machines (SVM) have been widely used to detect fraud as a classifier technique (Cecchini et al., 2010; Bhattacharyya et al., 2011). Furthermore, it uses the statistical learning theory that was founded by Vapnik in 1995 which makes it a statistics-based technique that generates a solution by using a linear model (ibid.). Therefore, SVMs are widely used in detecting fraud (Vanhoeyveld et al., 2020). Secondly, Random forests is a second classifier used in fraud detection. According to Bhattacharyya et al. (2011), random forests is a classifier model that can be defined as aggregated decision trees models which are characterized with flexibility and ability to analyse complex data.

Furthermore, each decision tree involves ensembles and nodes which individually generate a prediction output for the fraud detection system which gives it a strength advantage in addition to its ease of use (ibid.). Thirdly, logistic regression (LR) as a third classifier has been used in fraud detection as well (Carneiro et al., 2017; Bhattacharyya et al., 2011; Zou and Kapoor, 2011). According to Ngai et al. (2011) and Shen et al. (2007) the LR technique is a mathematical technique that is used for the processing of the task of ML prediction before knowing the outcome. Specifically, it helps in predicting the fraud case before occurring by giving a probability percentage of having a fraud case (Shen et al., 2007). Furthermore, Bhattacharyya et al. (2011) argued that the task of prediction that this technique delivers is crucial in fraud detection and prevention.

Four, artificial neural networks (NN) are effective classifiers for fraud detections well because they are useful for future prediction (Krauss et al., 2017). According to Shen et al. (2007), the NNs are non-linear mapping classifiers which are better than the statistical classifiers such as LR and SVMs in detecting financial fraud. One NN types is deep learning (DL) (Brahma et al., 2016). Moreover, DL as a process involves many layers which are learned from data and produce object recognition or detection (Shin et al., 2020). The recognition of DL has many examples such as voice recognition, image recognition, text recognition (King et al., 2020; Yaram, 2016).

What is more is that the behavior mining is a crucial concept and factor in nowadays fraud detection systems because of its ability to profile, store and analyse the users' behaviour which is a main component in nowadays feud detection systems (Sung and Liu, 2007). Additionally, the behavior consideration in this process means grouping and measuring the total of the repeated raw features which creates aggregated features (Dal Pozzolo et al., 2014). The paper of Bahnsen et al. (2016) has found that by combining the aggregated features with the raw features a has significant positive impact on the performance of the detection system by increasing it 200%. However, Eshghi and Kargari (2019) claim that the use of aggregated features in the detection system might have challenges such as decreasing accuracy and increasing time in addition to the probability of having a lack of historical data which can disrupt the aggregated features. According to Bhattacharyya et al. (2011), many studies concluded the efficiency and effectiveness of performance for the use of aggregated features with the random forests compared to other classifiers. Additionally, the use of aggregated features with SVMs and logistic regression has resulted in a good performance as well but not better than random forests (ibid.).

According to Sung and Liu (2007), the fraud detection systems in general were based on classification algorithms in the past. For example, Naive Bayes classification has been introduced by Elkan et al in 1997 and it has been an effective and accurate classifier in analysing big data and learning (ibid.). Moreover, C4.5 is another classifier introduced by Quinlan in 1998 and was widely used in detecting fraud because of its ability to explain patterns in addition to the prediction task (ibid.). Further, Back-propagation is another classifier that was used and performed well in detecting fraud (ibid.). However, the use of these techniques is not effective in nowadays fraud detection because of its training set which are created and entered manually and leading to more complexity and effort (ibid.). Furthermore, the use of the past techniques is not appropriate for nowadays detection because of its inability to analyse behaviours data (ibid.).

The modern classification tools have been used within different types of fraud which required developing fraud detection systems. The system generally is required to distinguish between the legitimacy data and the data which shows that there is fraud (Ngai et al., 2011). Therefore, to deliver the previous task, ML has been used to detect different types of fraud (Aparicio et al., 2020; Ngai et al., 2011). This can be shown as an automated computer system under the cover of ML (ibid.). According to Bologa et al. (2013), the data that generates the fraud detection systems are described as high volume and variety data which means that they are complex data and needs consideration. Furthermore, Aparicio et al. (2020) state that the data in the fraud detection system are called features which generate the system.

4. CASE STUDY

The research used case study approach to create a conceptual design approach for fraud detection. The selected case was from a supply chain for a car manufacturer in an Asian country suffering from fraud and misbehaviour. The investigation may be based on qualitative or quantitative types of data, using a case. In the pilot study for this research, qualitative data was used for many reasons. If research is restricted to narrow the investigation of the research into a particular context as a case study, it will decrease time and resources in the research search (Rashid et al., 2019).

The supply chain for the car manufacturer is concerned about fraudulent information from the vendors and distributors. The misinformation ranges from inflated invoices to an inaccurate accounting practice. In additions, parts may not be of adequate quality or with wrong specifications. This may be caused by intentional fraud by the vendors or by careless practice. Such misinformation can cause damages later in the supply chain or in the product quality. Detecting fraud in infrequent fraudulent practices may be difficult and time consuming with traditional tools. Historical big data, where some frauds were discovered, may be fed to the algorithms with machine supervised learning discovering patterns and leads to misinformation detection.

This research project, limited in time, requires choosing the most efficient research method in terms of time. The primary qualitative data was collected by conducting interviews at the case study organization. According to Chesebro and Borisoff (2007), conducting interviews to support a research paper is a qualitative data collection. Furthermore, this qualitative tool is used to link the case study organization with the theoretical framework of the research, analyse and conduct a discussion later. According to Baskarada (2014), the case study research must be supported by a comprehensive theoretical framework which gives a background about the research questions' fields that the case study will answer. The theoretical framework of the project is based on the high ranked journals identified.

A pilot case was conducted initially to examine the signatures on invoices and purchase orders to be authentic and original. Supplier and client signature and stamp, sources of fraud have been mandated by law for each invoice or purchase order.

5. MACHINE ALGORITHM

The macro-level code below describes the logic and procedure used in our supervised ML. This is an artificial and informal language that helps researchers develop algorithms. It is a description of computer programming algorithms that uses the structured convention of programming languages but omits detailed subroutines or language specific syntax.

The actual code is later developed in Python, which is a high-level interpreted general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Reason for using the Python language in Machine Learning are its simple syntax. the development of applications with Python is fast when compared to many programming languages. Furthermore, it allows the developer to test algorithms without implementing them. Readable code is also vital for collaborative coding.

Domain 1: Business problem (question) framing

- Task 1: Obtain or receive problem statement & usability statement
- Task 2: Identify stakeholders
- Task 3: Determine if problem is amendable to analytics solution
- Task 4: Refine problem statement and delineate constrains
- Task 5: Define an initial set of business benefits
- Task 6: Obtain stakeholder agreement on problem statement.

Domain 2: Analytics problem framing

- Task 1: Reformulate the problem statement as an analytics problem
- Task 2: Develop a proposed set of drivers and relationships to outputs
- Task 3: State the set of assumptions related to the problem
- Task 4: Define key metrics of success
- Task 5: Obtain stakeholder agreement on the approach

Domain 3: Data

- Task 1: Identify and prioritise data needs and resources
- Task 2: Acquire data
- Task 3: Harmonise, rescale, clean and share data
- Task 4: Identify relationships in data
- Task 5: Document and report findings
- Task 6: Refine business and analytics problem statements

Domain 4: Methodology selection

- Task-1: Identify available problem solving approaches (methods)
- Task-2: Select software tools
- Task-3: Test approaches (methods)
- Task-4: Select approached (methods)

Domain 5: Model building

- Task-1: Identify model structures
- Task-2: Run and evaluate models
- Task-3: Calibrate models and data
- Task-4: Integrate the models
- Task-5: Document and communicate findings

Domain 6: Deployment

- Task 1: Perform business validation of the model
- Task 2: Deliver report with finding
- Task 3: Create model, usability & system requirements for production
- Task 4: Deliver production model/system
- Task 5: Support deployment

Domain 7: Lifecycle Maintenance

- Task-1: Document initial structure
- Task-2: Track model quality
- Task-3: Recalibrate and maintain the model
- Task-4: Support training activities:
- Task-5: Evaluate business benefit of model over time

6. PILOT RESULTS

An audit company has been chosen as a pilot case study of this research that audits more than 12 million invoices per month. Mandated by law to verify signatures and stamp for invoices, an invoice without signature is not valid and the organization has the right to avoid payment. The challenge of the manually checking of 12 million invoice per month is not feasible. To cope with this challenge, we leveraged AI with one innovative idea to detect fraud in the fastest way.

The available sample size to teach artificial brain is about 8000 invoice that is not enough but to boost learning process we had to generate some fake data by adopting different augmentation methods such as rotation, random erasing or fading and increase the sample size to 56000 invoices. Finally, the Yolv5 as a pre-trained CNN on a platform that equipped with GPU3090 utilized and we got result. The final model supports us to scan monthly 12 million invoices just in 5 days that manually took about 60 months. Now targeted contractor benefits AI to detect improper invoice and fundamentally return invoices that submit improper or fraudulent claims.

7. CONCLUSION

Supply chains in many instances prone to fraud which may go undetected and cause financial and operational damages to the organizations. As big data becomes available through ERP and other supply chain information systems, computer algorithms are instrumental to discover any physical fraud such as wrong or low-quality parts or information fraud such as invoices with false signatures or numbers.

With the recent advances in AI and ML heuristics and languages, fraud detection is imbedded in the systems to discover and examine any possibility of fraud in supply chains. This paper illustrated the application of ML in SC fraud detection after reviewed the basic concepts and the literature. Results of a pilot study proves the concept here.

REFERENCES

Abdallah, A., Maarof, M.A. and Zainal, A., (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, *68*, pp.90-113.

Agarwal, P. K. (2018). Public Administration Challenges in the World of AI and Bots. *Public Administration Review*, *78*(6), 917–921.

Akkermans, H. A., & Van Oorschot, K. E. (2005). Relevance Assumed: A Case Study of Balanced Scorecard Development Using System Dynamics. *Journal of the Operational Research Society*, *56*(8), 931–941.

Al Sawalqa, F., Holloway, D.A. and Alam, M., (2011). Scope and aims of performance measurement practices: Evidence from Jordan.

Ansar, S.A., (2021). A Critical Analysis of Fraud Cases on the Internet. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(12), pp.2164-2186.

Aparício, D., Barata, R., Bravo, J., Ascensão, J.T. and Bizarro, P., (2020). ARMS: Automated rules management system for fraud detection. *arXiv preprint arXiv:2002.06075*.

Aral, K.D., Güvenir, H.A., Sabuncuoğlu, İ. and Akar, A.R., (2012). A prescription fraud detection model. *Computer methods and programs in biomedicine*, *106*(1), pp.37-46.

Aziz, S. and Dowling, M., (2019). Machine learning and AI for risk management. In *Disrupting Finance*(pp. 33-50). Palgrave Pivot, Cham.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational Issues of Big Data and Analytics in Network Business. *MIS Quarterly: Management Information Systems*, *40*(4), 807–818.

Bahnsen, A.C., Aouada, D., Stojanovic, A. and Ottersten, B., (2016). Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications*, *51*, pp.134-142.

Baskarada, S., (2014). Qualitative case study guidelines. *Baškarada, S.(2014). Qualitative case studies guidelines. The Qualitative Report*, *19*(40), pp.1-25.

Becker, K., Antuar, N. and Everett, C., (2011). Implementing an employee performance management system in a nonprofit organization. *Nonprofit management and leadership*, *21*(3), pp.255-271.

Bhattacharyya, S., Jha, S., Tharakunnel, K. and Westland, J.C., (2011). Data mining for credit card fraud: comparative study. *Decision support systems*, *50*(3), pp.602-613.

Bhowmik, R., (2008). Data mining techniques in fraud detection. *Journal of Digital Forensics, Security and Law*, *3*(2), p.3.

Bologa, A.R., Bologa, R. and Florea, A., 2013. Big data and specific analysis methods for insurance fraud detection. *Database Systems Journal*, *4*(4), pp.30-39.

Borges, A.F., Laurindo, F.J., Spínola, M.M., Gonçalves, R.F. and Mattos, C.A., (2020). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, p.102225.

Bose, I., Piramuthu, S. and Shaw, M.J., (2011). On Quantitative Methods for Detection of Financial Fraud. *Decision support systems*, *50*(3).

Brahma, P. P., Wu, D., & She, Y. (2016). Why Deep Learning Works: A Manifold Disentanglement Perspective. *IEEE Transactions on Neural Networks and Learning Systems*, *27*(10), 1997–2008.

Carneiro, N., Figueira, G. and Costa, M., (2017). A data mining based system for credit-card fraud detection in e-tail. *Decision Support Systems*, *95*, pp.91-101.

Carroll, A. B. (1979). A Three-Dimensional Conceptual Model of Corporate Performance. *Academy of Management Review*, 4(4), 497–505.

Cecchini, M., Aytug, H., Koehler, G.J. and Pathak, P., (2010). Detecting management fraud in public companies. *Management Science*, *56*(7), pp.1146-1160.

Chesebro, J.W. and Borisoff, D.J., (2007). What makes qualitative research qualitative?. *Qualitative research reports in communication*, *8*(1), pp.3-14.

Choudhury, P., Starr, E. and Agarwal, R., (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, *41*(8), pp.1381-1411.

Craja, P., Kim, A. and Lessmann, S., (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems*, *139*, p.113421.

Dal Pozzolo, A., Caelen, O., Le Borgne, Y.A., Waterschoot, S. and Bontempi, G., (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert systems with applications*, *41*(10), pp.4915-4928.

Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A. and Galanos, V., (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, p.101994.

Eshghi, A. and Kargari, M., (2019). Introducing a new method for the fusion of fraud evidence in banking transactions with regards to uncertainty. *Expert Systems with Applications*, *121*, pp.382-392.

Galeotti, M., Rabitti, G. and Vannucci, E., (2020). An evolutionary approach to fraud management. *European Journal of Operational Research*,*284*(3), pp.1167-1177.

Glancy, F.H. and Yadav, S.B., (2011). A computational model for financial reporting fraud detection. *Decision Support Systems*, *50*(3), pp.595-601.

Goode, S. and Lacey, D., (2011). Detecting complex account fraud in the enterprise: Role of technical and non-technical controls. *Decision Support Systems*, *50*(4), pp.702-714.

Hand, D.J., Whitrow, C., Adams, N.M., Juszczak, P. and Weston, D., (2008). Performance criteria for plastic card fraud detection tools. *Journal of the Operational Research Society*, *59*(7), pp.956-962.

Hu, N., Liu, L. and Sambamurthy, V., (2011). Fraud detection in online consumer reviews. *Decision Support Systems*, *50*(3), pp.614-626.

Jan, C.L., (2018). An effective financial statements fraud detection model for the sustainable development of financial markets: Evidence from Taiwan. *Sustainability*, *10*(2), p.513.

Jiang, J., Chen,J., Gu, T., Choo, K.., Liu, C., Yu, M., Huang, W. and Mohapatra, P., (2019). Anomaly detection with graph convolutional networks for insider threat and fraud detection. In *MILCOM 2019-2019 IEEE Military Communications Conference (MILCOM)*(pp. 109-114). IEEE.

Journeault, M. (2016). The Integrated Scorecard in Support of Corporate Sustainability Strategies. J*ournal of Environmental Management*, 182, 214–229.

Kane, G. C., Young, A. G., Majchrzak, A., & Ransbotham, S. (2021). Avoiding an Oppressive Future of Machine Learning: a Design Theory for Emancipatory Assistants. *MIS Quarterly: Management Information Systems*, *45*(1), 371–396.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15–25.
Kaplan, R. S., & Norton, D. P. (1996). Linking the Balanced Scorecard to Strategy. *California Management Review*, 39(1), 53–79.

King, T.C., Aggarwal, N., Taddeo, M. and Floridi, L., (2020). Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. *Science and engineering ethics*, *26*(1), pp.89-120.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep Neural Networks, Gradient-Boosted Trees, Random Forests: Statistical Arbitrage on the S&P 500. *European Journal of Operational Research*, *259*(2), 689–702.

Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems, 50*(3), 559–569.

Oh, W., & Pinsonneault, A. (2007). On the Assessment of the Strategic Value of Information Technologies: Conceptual and Analytical Approaches. MIS Quarterly: Management Information Systems, 31(2), 239–265.

Ozturk, E. and Coskun, A., (2014). A strategic approach to performance management in banks: The balanced scorecard. *Accounting and Finance Research*, *3*(3), pp.151-158.

Pourhabibi, T., Ong, K.L., Kam, B.H. and Boo, Y.L., (2020). Fraud detection: A systematic literature review of

graph-based anomaly detection approaches. *Decision Support Systems*, *133*, p.113303.

Psychoula, I., Gutmann, A., Mainali, P., Lee, S.H., Dunphy, P. and Petitcolas, F.A., (2021). Explainable Machine Learning for Fraud Detection. *arXiv preprint arXiv:2105.06314*.

Quah, J.T. and Sriganesh, M., (2008). Real-time credit card fraud detection using computational intelligence. *Expert systems with applications*, *35*(4), pp.1721-1732.

Rashid, Y., Rashid, A., Warraich, M.A., Sabir, S.S. and Waseem, A., (2019). Case study method: A step-by-step guide for business researchers. *International Journal of Qualitative Methods*, *18*, p.1609406919862424.

Ravisankar, P., Ravi, V., Rao, G.R. and Bose, I., (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision support systems*, *50*(2), pp.491-500.

Ryman-Tubb, N.F., Krause, P. and Garn, W., (2018). How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Engineering Applications of Artificial Intelligence*, *76*, pp.130-157.

Schleicher, D.J., Baumann, H.M., Sullivan, D.W., Levy, P.E., Hargrove, D.C. and Barros-Rivera, B.A., (2018). Putting the system into performance management systems: A review and agenda for performance management research. *Journal of Management*, *44*(6), pp.2209-2245.

Shen, A., Tong, R. and Deng, Y., (2007), June. Application of classification models on credit card fraud detection. In *2007 International f conference on service systems and service management*(pp. 1-4). IEEE.

Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., & Lee, K.-C. (2020). Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach. *MIS Quarterly: Management Information Systems*, *44*(4), 1459–1492.

Sumalatha, M.R. and Prabha, M., (2019). Mediclaim Fraud Detection and Management Using Predictive Analytics. In *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*(pp. 517-522). IEEE.

Sung, A.H. and Liu, Q., (2007). Behaviour mining for fraud detection. *Journal of research and practice in Information Technology*, *39*(1), pp.3-18.

Triepels, R., Daniels, H. and Feelders, A., 2018. Data-driven fraud detection in international shipping. *Expert Systems with Applications*, *99*, pp.193-202.

UN (n.d.). *Transforming our world: the 2030 Agenda for Sustainable Development.* Retrieved 15/08/2021. From: https://sdgs.un.org/2030agenda

Vanhoeyveld, J., Martens, D. and Peeters, B., (2020). Customs fraud detection. *Pattern Analysis and Applications*, *23*(3), pp.1457-1477.

Wang, Y. and Xu, W., (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems*,*105*, pp.87-95.

West, J. and Bhattacharya, M., (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & security*, *57*, pp.47-66.

Yaram, S., (2016). Machine learning algorithms for document clustering and fraud detection. In *2016 International Conference on Data Science and Engineering (ICDSE)*(pp. 1-6). IEEE.

Yesilkanat, A., Bayram, B., Köroğlu, B. and Arslan, S., (2020). An adaptive approach on credit card fraud detection using transaction aggregation and word embeddings. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*(pp. 3-14). Springer, Cham.

Zhang, X., Han, Y., Xu, W. and Wang, Q., (2019). HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Information Sciences*.

Zhou, W. and Kapoor, G., (2011). Detecting evolutionary financial statement fraud. *Decision support systems*, *50*(3), pp.570-575.