

DEVELOPING PREDICTIVE ANALYTICS MODELS FOR RISK-BASED AUDITING TO IMPROVE FINANCIAL ACCOUNTABILITY AND CORPORATE GOVERNANCE PRACTICES

Dasari Jayanth Kumar ¹, Miss.C.Yamini ²

¹student, Mca 2nd Year Kmmips, Tirupati, Affiliated To S.V. University, Tirupati, A.P, India

² Associate Professor, Dept Of Mca, Kmmips, Tirupati, Affiliated To S.V. University, Tirupati, A.P, India

ABSTRACT - The significant losses that banks and other financial organizations suffered due to new bank account (NBA) fraud are alarming as the number of online banking service users increases. The inherent skewness and rarity of NBA fraud instances have been a major challenge to the machine learning (ML) models and happen when non-fraud instances outweigh the fraud instances, which leads the ML models to overlook and erroneously consider fraud as non-fraud instances. Such errors can erode the confidence and trust of customers. Existing studies consider fraud patterns instead of potential losses of NBA fraud risk features while addressing the skewness of fraud datasets. The detection of NBA fraud is proposed in this research within the context of value-at-risk as a risk measure that considers fraud instances as a worst-case scenario. Value-at-risk uses historical simulation to estimate potential losses of risk features and model them as a skewed tail distribution. The risk-return features obtained from value-at-risk were classified using ML on the bank account fraud (BAF) Dataset. The value-at-risk handles the fraud skewness using an adjustable threshold probability range to attach weight to the skewed NBA fraud instances. A novel detection rate (DT) metric that considers risk fraud features was used to measure the performance of the fraud detection model. An improved fraud detection model is achieved using a K-nearest neighbor with a true positive (TP) rate of 0.95 and a DT rate of 0.9406. Under an acceptable loss tolerance in the banking sector, value-at-risk presents an intelligent approach for establishing data-driven criteria for fraud risk management

Key Words: Detection rate, fraud detection, K-nearest neighbor, skewed instances, value-at-risk.

1. INTRODUCTION

The Association of Certified Fraud Examiners (ACFE) 2022 released a financial fraud report stating that 2,110 fraud cases involving industries in financial sectors in 133 countries resulted in losses of around \$3.6 billion. Financial fraud can be termed as the deliberate employment of unlawful procedures or tactics to obtain financial gain. The consequences of financial fraud can potentially disrupt economies, raise living expenses, and undermine consumer confidence. Forms of financial fraud include insurance

fraud, money laundering, new bank account fraud, credit and debit card fraud, mortgage fraud, and many more. The act of opening an account to commit fraud at banks or other financial organizations is known as “new bank account(NBA) fraud”. Fraud not only results in immediate financial losses and erodes public confidence in institutions, but has broader consequences, affecting customers and financial systems through market instability and contributing to larger macroeconomic downturns. Fraud datasets typically exhibit some properties including skewness, evolving patterns, highly dimensional, and restricted access to relevant information. Specifically, fraud skewness which represents the majority fraud class over the non-fraud class has been a major concern to studies, as it affects the performance of fraud detection model. The Skewed fraud instances can have a bad influence on machine learning algorithms such as distance-based algorithms. Previous efforts in tackling fraud involve developing rule-based expert systems, statistical methods, machine learning, and risk-based methods. Due to the cost of maintenance and the inefficiency of rule-based methods, decision-makers decide to utilize statistical methods such as autoregressive models to handle financial fraud. The complex patterns and high dimensional nature of frauds make the statistical methods less effective, as such machine learning models were deployed. However, some of the studies that utilize machine learning techniques were found to have a high False Positive (FP) rate. Machine learning models can potentially handle high-dimensional data and complex patterns of fraud instances. To evaluate the effectiveness of machine learning model, Jesus et al. presented the first domain-specific and real-world bank account fraud (BAF) dataset. The datasets were generated using generative adversarial networks (GANs) and evaluated using light gradient boosting method (LGBM).The study utilizes 25 sets of hyper parameter configurations to optimize the LGBM model, utility aware reweighing was used to handle the class skewness of BAF dataset. The study utilizes stacking in ensembled learning with majority voting to evaluate the BAF dataset and address the changing fraud patterns. The study [20] uses federated learning in addressing data privacy issues of BAF dataset and deep neural networks to classify fraud instances. These studies achieve good performance in addressing BAF challenges; However, the studies do not

consider the potential losses of fraud risk features. To our knowledge, little research exists that employs machine learning techniques in NBA fraud detection. The detection of NBA fraud is proposed in this paper within the context of risk management that uses value-at-risk to consider skewed fraud instances as a worst-case scenario. To adequately estimate the losses of fraud risks, value-at-risk was augmented with expected loss and expected shortfall of frauds which further quantifies the mean and extreme loss effects respectively. These risk measures combination will allow the quantification of risks across mean, worst-case, and extreme scenarios. Value-at risk employs historical simulation to estimate potential losses of risk features. The risk-return features obtained from value at-risk are based on assessing their risk exposure to fraud risk. The risk-return features are sent as input to the NBA fraud detection model. Different machine learning models were trained; However, the K-nearest neighbor outperformed other models. The contributions of this paper are:

- This paper used an extreme value theorem to model the tails (potential losses) instead of the fraud pattern.
- This paper used value-at-risk to model the skewness of fraud instances more efficiently.
- This paper utilized historical simulation to estimate value-at-risk as it makes no assumptions on any distribution.
- This paper used novel detection rate performance metrics to capture the overall performance in detection of NBA fraud instances that incorporate risk fraud factors.

The remainder of the paper is arranged as follows: The study's review of the literature is presented in Section II. The problem definition is presented in Section III. The materials and procedures are presented in Section IV. The experimental setup is presented in Section V. The results are presented in Section VI. The study's conclusions and discussions are presented in Section VII.

2. LITERATURE REVIEW

This section presents related studies in financial fraud detection. Different studies exist that utilize both statistical and artificial intelligence-based methods in the context of a risk and financial fraud perspective.

2.1 STATISTICAL METHODS OF FRAUD DETECTION

Many studies in the literature utilize statistical methods in evaluating financial fraud. Specifically, significant studies were found to utilize ordinary least squares (OLS) regression and autoregressive (AR) models for financial fraud detection. Using the Tehran Stock Exchange dataset, the study uses a regression model to investigate the association between auditor characteristics and fraud

detection in emerging economies. The authors provide useful information for improving the reliability of the findings. Using pooled OLS and panel regressions, the study investigates the effect of political alignment on corporate fraud convictions; offer insights into the connection between politics and fraud. The authors use state-level data from 2003 to 2018 on US corporate fraud convictions and party affiliation. The study utilizes OLS to investigate financial factors of financial fraud, which is attributed to the fraud triangle. The study uses logistic regression to discover that external pressures and financial stability had a favorable impact on financial reporting fraud. On the other hand, collaboration, arrogance, changes in directors, incompetent oversight, and hubris have little bearing on false financial reporting. The study provides evidence for the contribution of gender diversity to fraud commission and detection in Chinese listed businesses between 2007 and 2018 using bivariate probit model. The authors opined that female corporate executives are linked to a stronger ability to detect fraud, which lowers the likelihood of businesses to commit fraud. From the standpoint of external auditors, the study sheds light on the causes of fraud and the function of forensic accounting using regression analysis to analyse Lebanese data. The study discovers that while the overall number of employees engaged in fraud affects the performance of money banks in Nigeria, the number of fraud cases and the total amount lost to fraud had a favorable influence. The use of statistical methods by the author such as OLS regression, Pearson correlation, and descriptive analysis strengthens the findings by the authors. The sales growth index and the depreciation index factors. The M-score are used in the study to analyse the possibility of profit management using the Athens Stock Exchange Market. It is pertinent to know that a large body of literature exists that utilizes the AR model. To handle large-scale non-uniform transactions more quickly, the authors employ the AR model, which makes it appropriate for detecting money laundering operations. The study uses factor analysis to generate the composite indicator, fractional integration (ARFIMA), and fractional counteraction VAR (FCVAR) approaches to evaluate the behavior of the composite suspicion tax fraud indicator about GDP and tax collection. The study employs the AR model, which is appropriate for studying networks with such topologies and applying it to the detection of financial transaction fraud since it considers the block-wise structure of networks. The authors discovered that, in line with reality, there is a risk relationship between fraudulent groups and ordinary loan applicants. The study outlined specific identification indicators that help with the detection of financial fraud using digital distributed laws, and the authors demonstrate that the probability of financial fraud increases significantly as the deviation of financial data distribution from Bedford's law increases. In summary, a large body of literature uses statistical methods to analyse the causes and effects that influence financial

fraud, but due to the complex nature and scalability of fraud, statistical methods are not enough to adequately examine financial fraud.

2.2 RISK-BASED METHODS OF FRAUD DETECTION

This section presents the financial fraud assessment from the perspective of risk mitigation. The existing studies utilize different risk measures such as value-at-risk (Var), expected loss, and expected shortfall to assess the level of risk of fraud. The study [29] offers strategies for breaking down the risk of fraud, identifying potential fraudsters, and enabling more targeted anti-fraud measures by tying the motivation of the fraud triangle to human tendencies that lead to spurious actions as well as the meta-model of fraud together. Regression analysis is utilized in the study [30] to look at how enterprises manage risk to determine how control environments, risk assessments, control activities, information and communication, and monitoring contributed to fraud prevention and detection efforts in Indonesian firms. The study defined additional security attributes that might have an impact on the cloud system and carried out an anomaly detection based on risk assessment named parallel processing (PP) that covers cyber threats and exploitation likelihoods. The model checker is then employed to determine the risk exposure rates associated with the respective attacks. The study proposes a framework in which doubly-truncated severity distributions are used to estimate the operational risk and offered a framework that includes database construction and risk modeling. By applying value-at-risk and expected shortfall to identify operational risk sources like external fraud risk and legal risk sections, the authors were able to produce better and consistent results. The study uses the number of compromised records to determine the cost of a data breach; the findings indicate that the total number of affected records has a Fréchet distribution, random forest is used for estimating the number of such records. The study uses the estimate of generalized extreme value paradigms to evaluate competency, digital technology abilities, and personality qualities that may improve the ability of external auditors to identify fraud risk, the efficiency of fraud risk assessment was linked to digital technology abilities through the application of the partial least-squares structural equation model (PLS-SEM). The study identified a positive core relation between fraud risk assessment and management and the efficient use of forensic accounting using chi-square, Fisher test, and correlation, however, there is no relationship between fraud risk assessment and management in terms of techniques causing fraud. The study examines fraud using ensemble learners for anomaly detection and also handles data skewness, a triage model that receives input from the ensemble model, and a risk model that estimates the financial losses. The authors successfully provide an effective fraud risk-based detection, from machine learning techniques to risk assessment, but do not to evaluate fraud

detection by first considering the risk component before subjecting it to machine learning detection. In summary, risk measures are good in the assessment and management of the features associated with fraud for effective fraud prevention and control. However, due to the nonlinearity, high dimension, and complex nature of fraud, these risk measures need to be augmented with other techniques such as machine learning techniques that enable proper and efficient fraud prevention and detection.

2.3. MACHINE LEARNING METHODS IN FRAUD DETECTION

This section presents studies that utilize machine learning techniques for the classification of fraud applications. The majority of the presented studies consider the detection while addressing the skewed nature of fraud instances. Sampling methods, hybrid methods, and other novel methods are majorly used to overcome these skewed nature of fraud datasets. The study addresses class skewness in credit card fraud using quantum machine learning (QML) and support vector machines (SVM). The results show that classic machine learning techniques are still useful for non-time series data, whereas Implications can be used for time-series-based and highly skewed data. Quantum neural network (QNN) achieves good performance in fraud detection by the study. The study trained different machine learning models, all of which were using default implementations and parameters, Boost performed more accurately than any other models. The effectiveness of telecom fraud is assessed in the study using a dynamic graph neural network (DGNN), the authors effectively present a suggested method for resolving the issue of telecom fraud detection in extensive phone social networks. To assess credit card fraud while considering the skewness of fraud instances, the study makes use of logistic regression (LR), K-nearest neighbor (KNN), decision tree (DT), random forest (RF), and autoencoder (AE) as they can handle skewed data better than other models, the AE model performs better. KNN, linear discriminant analysis (LDA), and linear regression are used in the study to investigate credit card fraud, by addressing the skewed nature of the credit card fraud data and using cross-validation techniques, KNN showed higher performance. Using ARIMA model for fraud detection based on daily transaction counts, the study carried out anomaly detection, the model is contrasted with four industry-standard anomaly detection algorithms: the box plot, isolation for Est(IF), local outlier factor (LOF), and K-means models. An ensemble classifier (EC) [incorporating bagging and boosting has been used to address the issue of fraud class skewness, the approach is found to perform better when compared to the current methods. The study addresses the issue of skewed datasets by using fuzzy C-means clustering and the selection of related instances. The authors address the issues with conventional under-sampling strategies to enhance the detection performance

and accuracy. To identify fraudulent transactions, the study suggested LSTM ensemble, SMOTE-ENN was used to address the problem of fraud skewness. The method outperformed other algorithms in terms of performance, but, SMOTE method may occasionally produce instances that are not typical instances of the minority class. A dynamic ensemble technique for anomaly identification in the Internet of Things systems is proposed. To address the issue of fraud skewness, the borderline synthetic minority over sampling approach (Borderline SMOTE), One-Sided Selection (OSS), and adaptive synthetic (ADASYN) were applied in the study, OSS were found to be optimal under sampling technique and that adaptive synthetic (ADASYN) performs better when employing the gradient tree boosting (GTB) classifier. Random forest ensemble approach performed exceptionally well on oversampling and under-sampling. Though under-sampling usually led to the loss of important information while on the other hand, oversampling brings information that may not be fully a representative of the training set. It is widely acknowledged that the skewed distribution of fraud instances presents a significant challenge for many machine learning models. The resampling techniques that have been used in effective fraud skewness mitigation may not be free from certain shortcomings. The resampled instances usually suffer from non-representative of the dataset, over fitting, and the loss of important data. Hence, there is a need to augment the machine learning algorithms with novel approach in overcoming this challenge.

2.4. RESEARCH PROBLEM

The problem of NBA fraud keeps increasing daily as the number of online banking service users keeps increasing ML techniques applied in many researches shows promising performance in overcoming NBA fraud. However, most ML struggles when the distribution fraud instances are skewed as in the case of BAF dataset. The studies utilize LGBM to address skewed fraud instances using the True Positive (TP) rate as a performance measure. The study utilizes stacking in ensemble learning with majority voting to evaluate the BAF dataset and address the changing fraud patterns. The study uses federated learning in addressing data privacy issues and deep neural networks to classify fraud with TP rate as a metric. These studies achieve good performance in addressing BAF challenges. However, the studies did not consider the potential losses of fraud risk features. Major problems this paper addresses include: • Most existing studies do not consider potential losses of fraud risk features, but fraud instances happen rarely and cause big losses when they occur Fraud instances are inherently skewed compared to non-fraud instances, producing a highly skewed distribution. • Fraud pattern tends to have more irregular and extreme values, while models like

logistics regression or regression assume normality and predictions may produce an inaccurate result.

3. PROBLEM DEFINITION

This paper considers $X_i = x_1, x_2, \dots$ as a vector of observation in a respective raw feature. The X_i transformed log return $X_p = x_1, x_2, \dots$, which is a vector of log returns. The log return X_p is computed using $\log(1 + x_i)$. The log returns are assessed using value-at-risk to determine the risk of fraud for each respective feature. The fraud instances are considered as the worst-case scenario and beyond. The value at-risk model is the tail of a distribution i.e., extreme quantiles where fraud occurs. The historical simulation was conducted to estimate potential losses distribution $\tilde{\ell}_p = \ell(X_p) = -(f(t) + 1, Z_t + x_p - f(t, Z_t))$. The extreme value theorem is applied to estimate the tail distribution based on fraud instances skewness. The value-at-risk V_A is a measure that assesses the risk of the features is the sum of expected loss ℓ and expected shortfall C , as seen in (3). The risk-return features were obtained as log return passes through the formulation comprising ℓ, V , and C as given in (9-12) and the equations are derived based on the event of fraud instances. The value-at-risk quantified risks across mean, worst-case, and extreme scenarios. This study aims to detect NBA fraud based on risk-return features using the KNN model.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (17) \quad \text{Prate} = \frac{TP}{TP+FN} \quad (18)$$

$$\text{Prate} = \frac{FP}{FP+TN} \quad (19) \quad \text{F-score} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (20)$$

3.1 PROPOSED METHOD

The proposed design of this research is illustrated in Fig. 1 which describes the steps and process involved in NBA fraud detection. Value-at-risk being an important part of this research is designed to model the severe and extreme fraud risk features; it also focuses on rare fraud instances that are detrimental and very costly when occurred. However, the rare cases that are mostly skewed can distort machine learning algorithms especially distance based like KNN. The value-at-risk can handle the fraud skewness through the utilization of adjustable threshold probability ranges (confidence level) unlike the conventional methods that employ constant fraud probability weight that's attached to the skewed fraud instances. The preprocessed, extracted and engineered features were sent as input to value-at-risk for simulation. Meanwhile, a distance based KNN is designed for adjustability to detect fraudulent features through identifying rare clusters with nearest neighbor distance k . The confidence level chosen considers the rare fraud cases as higher risk features that would result in fewer training sets, particularly for the KNN model with hyper parameter k . The fraud detection model requires the optimization of k to a lower setting to

sufficiently model the fraudulent features in the rare cluster. The distance weight of KNN is imperative in inhibiting fraud skewness by assigning a higher weight to near instances which in turn facilitates efficient detection of skewed instances. Additionally, this paper put forward a novel approach to NBA fraud detection through the utilization of value-at-risk that appropriately models the fraud skewness. The selection of a 99.5% confidence level highlighted the need to capture 0.5% of extreme fraud risk instances which fit to fall under the subset of 1% fraud rate (detection effectiveness) as shown in Fig. 2. The value-at-risk which is finance and risk management tools model the tail of fraud events that are extreme. Consequently, a novel detection rate performance metrics that incorporate the risk of skewed fraud instances into the overall performance measure of detecting rare instances were put forward which will later be seen in (21). The metrics provides the model with capacity to identify and attach more weight to rare and extreme fraud instances by including the fraud rate and confidence level in detection process. Therefore, this research put forward a single metrics that capture overall rate of fraud detection based on risk exposure. Under an acceptable loss tolerance in the banking sector, value-at-risk presents an intelligent approach for establishing data-driven criteria for fraud risk management.

3.2 DATA PREPROCESSING

This paper carries out preprocessing tasks to improve the quality of features and ensure model accuracy. The redundant feature device fraud count contains zero instances all of which were manually removed from the data making the model less complex. The categorical features device_os, employment status, payment_type, and housing status were labeled to make them easier to learn because machine learning cannot process features that are non-numeric. The features zip count, keep alive session and bank months count were eliminated to avoid noise and collinearity issues. The features foreign request, has other card, and email is free were eliminated as they give too much undefined log returns

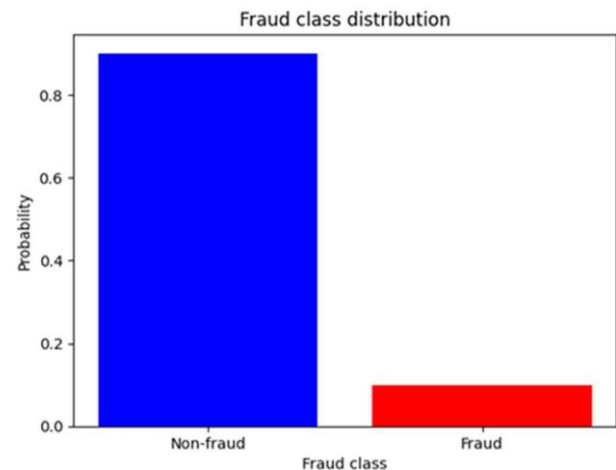


Chart 1. Fraud class distribution

3.3 FEATURE EXTRACTION AND ENGINEERING

The development of a NBA fraud detection model was based on the selection of relevant features from demographic, behavioral, risk management and transactional per spec Demographic features such as income, customer age, and employment status were selected. Behavioral features such as bank_branch_count_8w and housing status were selected. Risk-based features that include credit_risk_score and proposed_credit_limit were selected. Transactional features such as days_since_request, total velocity, and payment type were also selected. Additionally, two or more existing features are combined to form a new feature as given in Table 2. The features engineered are based on location, velocity of transactions, default risk, and ability to repay loans to determine the likelihood of fraudulent behaviors. The selected features were used along with the other raw features for accurate model training.

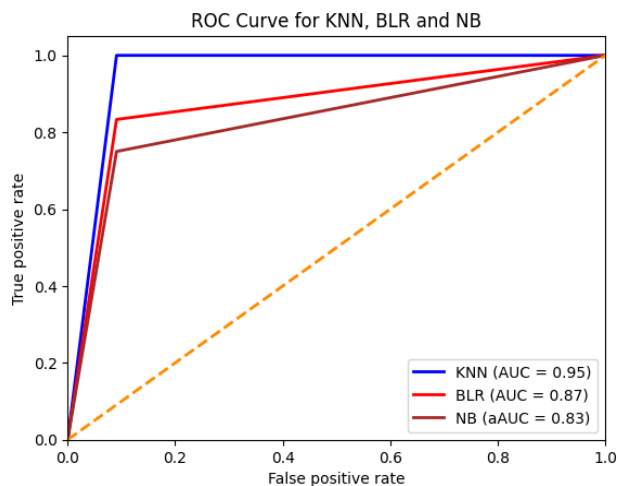
4. RESULTS ANALYSIS AND DISCUSSION

This section presents the general results obtained from expert amental research with skewed fraud instances and risk-return features with their validation. The 10-fold cross validation was used to evaluate NBA fraud detection models.

4.1. RESULT OF NBA FRAUD DETECTION MODEL WITH SKEWED FRAUD INSTANCES

This section presents the result of the NBA fraud detection model with skewed instances using BLR, KNN, and NB. The results are presented in Table 3, the best metric results among the models were written in bold number. The accuracy result of BLR, KNN, and NB are 0.9869, 0.9884, and 0.9743 respectively. The TP rate results of BLR, KNN, and NB are 0.0016, 0.0061, and 0.1355 respectively. observed that the results of accuracy and FP rate were

good. However, the results of the TP rate and f-score were not very good. The TP rate is a very important metric especially in fraud detection, robust and accurate fraud detection must attain a good TP rate. The poor performance of the fraud detection model, particularly in TP rate and f score, using fraud skewed instances highlighted the need for model improvement. We employ to improve the fraud detection model using value.



Graf 1. Receiver operating curve for the fraud models

5. CONCLUSION

The value-at-risk-based fraud detection model presented in this paper enables the quantification and mitigation of fraud risk features and at the same time overcome the influence of skewed fraud instances which is very crucial in solving financial fraud challenges. The value-at-risk attach confidence probability weight to the rare fraud cases with KNN is imperative in inhibiting class skewness by assigning a higher weight to near instances which in turn facilitates efficient detection of skewed instances. The deployment of expected short fall and expected loss by value at risk allows quantification of risk across mean, worst-case, and extreme scenarios enabling aggregation of their strengths. Therefore, an accurate fraud detection system assists organizations in making effective choices and reducing the overall expense of fraud detection and prevention. This paper does not consider the time windows in the experiment. However, the major challenge is the lack of data availability in NBA fraud detection.

6. REFERENCES

[1] ACFE. Association of Certified Fraud Examiners (ACFE) 2022 Report to the Nations. Accessed: 2023. [Online]. Available: <https://legacy.acfe.com/report-to-the-nations/2022/>

[2] T. Ashfaq, R. Khalid, A. S. Yahaya, S. Aslam, A. T. Azar, S. AL safari, and I. A. Hameed, "A machine learning and

blockchain based efficient fraud detection mechanism," *Sensors*, vol. 22, no. 19, p. 7162, Sep. 2022.

[3] N. S. Alfaiz and S. M. Fati, "Enhanced credit card fraud detection model using machine learning," *Electronics*, vol. 11, no. 4, p. 662, 2022.

[4] A. Alfaadhel, I. Almomani, and M. Ahmed, "Risk-based cybersecurity compliance assessment system (RC2AS)," *Appl. Sci.*, vol. 13, no. 10, p. 6145, May 2023.

[5] D. Sarma, W. Alam, I. Saha, M. N. Alam, M. J. Alam, and S. Hossain, "Bank fraud detection using community detection algorithm," in *Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2020, pp. 642–646.

[6] A. Pagano, "Digital account opening fraud on demand deposit accounts: An assessment of available technology," Ph.D. thesis, Utica College, Utica, NY, USA, 2020.

[7] Shuftipro. New Account Fraud—A New Breed of Scams. Accessed: 2023. [Online]. Available: <https://shuftipro.com/reports-whitepapers/newaccount-fraud.pdf>

[8] R. Sasirekha, B. Kanisha, and S. Kaliraj, "Study on class imbalance problem with modified KNN for classification," in *Intelligent Data Communication Technologies and Internet of Things*, vol. 101. Singapore: Springer, 2022, pp. 207–217, doi: https://doi.org/10.1007/978-981-16-7610-9_15.

[9] P. Vanini, S. Rossi, E. Zizic, and T. Domenic, "Online payment fraud: From anomaly detection to risk management," *Financial Innov.*, vol. 9, no. 1, p. 66, Mar. 2023, doi: [10.1186/s40854-023-00470-w](https://doi.org/10.1186/s40854-023-00470-w).

[10] X. Zhu, X. Ao, Z. Qin, Y. Chang, Y. Liu, Q. He, and J. Li, "Intelligent financial fraud detection practices in post-pandemic era," *Innovation*, vol. 2, no. 4, Nov. 2021, Art. no. 100176, doi: [10.1016/j.xinn.2021.100176](https://doi.org/10.1016/j.xinn.2021.100176).

[11] M. Monge, C. Poza, and S. Borgia, "A proposal of a suspicion of tax fraud indicator based on Google Trends to foresee Spanish tax revenues," *Int. Econ.*, vol. 169, pp. 1–12, May 2022, doi: [10.1016/j.inteco.2021.11.002](https://doi.org/10.1016/j.inteco.2021.11.002). [12] S. Kannan and K. Somasundaram, "Autoregressive-based outlier algorithm to detect money laundering activities," *J. Money Laundering Control*, vol. 20, no. 2, pp. 190–202, May 2017, doi: [10.1108/jmlc-07-2016-0031](https://doi.org/10.1108/jmlc-07-2016-0031).

[13] B. Xiao, B. Lei, W. Lan, and B. Guo, "A blockwise network autoregressive model with application for fraud detection," *Ann. Inst. Stat. Math.*, vol. 74, no. 6, pp. 1043–1065, Dec. 2022, doi: [10.1007/s10463-022-00822-w](https://doi.org/10.1007/s10463-022-00822-w).

[14] G. Moschini, R. Houssou, J. Bovay, and S. Robert-Nicoud, "Anomaly and fraud detection in credit card transactions using the ARIMA model," in *Proc. 7th Int. Conf.*

Time Forecasting, Jul. 2021, p. 56, doi: 10.3390/engproc2021005056.

[15] A. A. Alhashmi, A. M. Lashae, A. A. Dare, A. F. Alanazi, and R. Effie, "An ensemble-based fraud detection model for financial transaction cyber threat classification and countermeasures," *Eng., Technol. Appl. Sci. Res.*, vol. 13, no. 6, pp. 12433–12439, Dec. 2023, doi: 10.48084/etasr.6401.

[16] R. M. Aziz, R. Mahto, K. Goel, A. Das, P. Kumar, and A. Saxena, "Modified genetic algorithm with deep learning for fraud transactions of Ethereum smart contract," *Appl. Sci.*, vol. 13, no. 2, p. 697, Jan. 2023, doi: 10.3390/app13020697.

[17] M. Hegazy, A. Madian, and M. Rajaei, "Enhanced fraud miner: Credit card fraud detection using clustering data mining techniques," *Egyptian Compute. Sci. J.*, vol. 40, no. 3, pp. 1–10, 2016.

[18] S. Jesus, J. Pombal, D. Alves, A. Cruz, P. Saleiro, R. Ribeiro, J. Gama, and P. Bizarro, "Turning the tables: Biased, imbalanced, dynamic tabular datasets for ML evaluation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 33563–33575.

[19] J. Pombal, P. Saleiro, M. A. T. Figueiredo, and P. Bizarro, "Fairness-aware data valuation for supervised learning," 2023, arXiv:2303.16963.

[20] T. Awosika, R. Mani Shukla, and B. Pranggono, "Transparency and privacy: The role of explainable AI and federated learning in financial fraud detection," 2023, arXiv:2312.13334.

[21] J. Khaksar, M. Salehi, and M. Lari DashtBayaz, "The relationship between auditor characteristics and fraud detection," *J. Facilities Manage.*, vol. 20, no. 1, pp. 79–101, Jan. 2022, doi: 10.1108/jfm-02-2021-0024.

[22] A. Cordis, "Political alignment and corporate fraud: Evidence from the United States of America," *J. Appl. Accounting Res.*, Oct. 2023, doi: 10.1108/jaar-06-2022

[23] M. J. Rahman and X. Jie, "Fraud detection using fraud triangle theory: Evidence from China," *J. Financial Crime*, vol. 31, no. 1, pp. 101–118, Jan. 2024, doi: 10.1108/jfc-09-2022-0219.

[24] T. Achmad, I. Ghazali, and I. D. Pamunkeys, "Hexagon fraud: Detection of fraudulent financial reporting in state-owned enterprises Indonesia," *Economies*, vol. 10, no. 1, p. 13, Jan. 2022.

[25] Y. Wang, M. Yu, and S. Gao, "Gender diversity and financial statement fraud," *J. Accounting Public Policy*, vol. 41, no. 2, Mar. 2022, Art. no. 106903.