Fraud Detection in Motor Insurance Claims Using Supervised Learning Techniques: A Review

David Gichohi Maina
Department of Computer Science
Dedan Kimathi University of Technology
Nyeri, Kenya
Email: david.gichohi [AT] dkut.ac.ke

Juliet Chebet Moso
Department of Computer Science
Dedan Kimathi University of Technology
Nyeri, Kenya
Email: juliet.moso [AT] dkut.ac.ke

Patrick Kinyua Gikunda
Department of Computer Science
Dedan Kimathi University of Technology
Nyeri, Kenya
Email: patrick.gikunda [AT] dkut.ac.ke

Abstract—**Fraudulent claims have been a big drawback in motor insurance despite the insurance industry having vast amounts of motor claims data. Analyzing this data can lead to a more efficient way of detecting reported fraudulent claims. The challenge is how to extract insightful information and knowledge from this data and use it to model a fraud detection system. Due to constant evolution and dynamic nature of fraudsters, some approaches utilized by insurance firms, such as impromptu audits, whistle-blowing, staff rotation have become infeasible. Machine learning techniques can aid in fraud detection by training a prediction model using historical data. The performance of the models is affected by class imbalance and the determination of the most relevant features that might lead to fraud detection from data. In this paper we examine various fraud detection techniques and compare their performance efficiency. We then give a summary of techniques' strengths and weaknesses in identifying claims as either fraudulent or non-fraudulent, and finally propose a fraud detection framework of an ensemble model that is trained on dataset balanced using SMOTE and with relevant features only. This proposed approach would improve performance and reduce false positives.**

Keywords— Insurance, Fraud, Class Imbalance, SMOTE, Feature Selection, Ensemble Learning

I. INTRODUCTION

Fraud is the crime of cheating someone in order to get money or goods illegally [1] where the actor benefits with profits and victims get losses [2]. Insurance industry in the world have more than a thousand companies and collect trillions of dollars as premium [3]. In Kenya, insurance plays a critical role in the economy. According to [4], Kenya's insurance industry had a gross written premium of KES 274.98 billion in 2021 noting that there is a growth of the premium every year. In motor insurance, the owner of a motor vehicle (insured) pays premium to an insurance company which will take the financial risk in case of an accidental incidence of vehicular damage or theft [5]. A fraudulent motor insurance claim is a situation where the insured seeks financial gain by either filing forged documents or fabricating an accident or motor theft [6].

In all reported claims, approximately 10% are fraudulent claims [7] but only less than 3% is legally preceded [8]. For this reason, there is need for efficient ways to detect fraudulent claims reported to insurance companies that will reduce their loss adjustment expenses. Insurance companies use some common approaches to fight fraud such as impromptu audits, job rotation, anti-fraud policy [9], external audits, management reviews, code of conduct [10], whistleblowing [11], and biometric systems among others. These approaches used, rely on experts' experiences, intuition, and business or domain knowledge, in selecting statistically significant features to detect fraudulent claims [12].

Various fraud detection techniques have been proposed to identify abnormal activities that occurred in past transactions. However, these techniques become infeasible due to constant evolution of new methods by fraudsters [13]. The huge amount of data available in the insurance industry [14] can be used to collect insightful information using approaches such as machine learning, statistical and mathematical techniques [13]. This information can be applied in fraud detection which will go a long way by significantly reducing losses made by insurance companies, thereby making insurance policies affordable.

This research will primarily focus on review of these techniques, on their success in improving performance for fraud detection. It will aim to provide an insight on the use of supervised learning models.

The rest of the paper is organized as follows: Section II presents the methodology used. In Section III, we present an overview of fraud, fraudsters, and fraudulent claims; in Section IV, we discuss classical approaches used to detect fraud; in Section V and VI a review of machine learning techniques and their use in fraud detection is provided. Section VII discusses a proposed framework and the paper is concluded in Section VIII.

II. METHODOLOGY

The main aim of this work was to systematically examine various fraud detection techniques that can be used in identifying a motor insurance claim as either fraudulent of non-fraudulent. We evaluate various approaches used by different

authors and explain their strengths and weaknesses. Our data source for published articles were searched from
Google Scholar, IEEE, ScienceDirect, SpringerLink, and Elsevier. The search criteria was structured in a manner that we obtain papers discussing fraud detection in financial and motor insurance industries using machine learning.

When selecting review articles, we read through the abstracts of all papers and pick those that mention fraud or outlier detection and use of machine learning approaches. Specifically, we ensured that we picked papers that address topics such as: fraudsters and fraudulent activities in financial and insurance industry, approaches for fraud detection, use of machine learning for fraud detection.
To ensure study quality, we excluded book chapters, magazines and review paper older than seven years i.e papers of 2016 and earlier on exclusion of classical papers like review of machine learning. Subsequently, we settled on 30 papers to use for this review.

III. Fraudulent Claims and Fraudsters

People can perpetrate fraud in various ways. The act of fraud can be committed by either the vendor or the client [15]. Fraud from the vendor side arises from non-existent companies used by brokers who offer service to clients, failure by the insurance agent to submit full information to the insurance company, failure to remit premiums by brokers, and so on. From the client side, fraud is encountered by overestimation by garages and spare parts suppliers, exploited accidents [16], fabricated claims, and provoked accidents [17], among others.
Frauds can be categorized into different categories like financial frauds and auto indemnity frauds [16]. They are divided further into either soft fraud or hard fraud [16]. In hard fraud, the vehicle will have total damage for the fraudster to deliberately get rid of it or get more money than its market value. This may include theft, fire, or loss under excluded risks. Soft fraud includes instances like double claim in a single loss, inflating costs of repair, damages caused earlier, and replacement rather than repair [18], among others. Most motor vehicle insurance fraudulent claims are soft frauds where fraudsters utilize different approaches to breach the claim process [19].
There are also various types of fraudsters. They can be classified into opportunist, amateur, and professional [20]. Opportunist fraudsters will sometimes embellish or inflate valid claims to increase the value of the payout [21], for example claiming items not broken alongside items broken in an accident. In rare cases, they fabricate an entire claim. An amateur fraudster can go beyond opportunists such as filling a claim for an accident that never occurred. Professional fraudsters are difficult to identify and take frauds either as individuals or in an organized network.

IV. Classical Approaches to Detect Fraud

These are the various approaches that have been used by insurance companies to predict, detect and identify fraudulent claims. Insurance companies adopt these approaches depending on human experts, intuition, and domain knowledge to combat the menace of fraudulent claims. Fraud hot lines, forensic accountants and whistle-blowing [22] have high effectiveness rating but are barely utilized [23]. Other preventive controls observed to deter fraud include internal audit [23], staff rotation [24], proper due diligence on customers, and code of conduct [25]. Red flags coupled with internal audit are also effective in detecting fraud. However, this audit requires analysis of data and transactions to identify fraud indicators including patterns, statistics, and other relevant abnormalities [26].
While these approaches have high effectiveness rating, digital analysis, discovery sampling and data mining are less prevalent since they are costly in terms of resources [25] and time [27]. These approaches are also developed manually and their success ratio is low due to their complexity and undetectable nature.
Implementing strong authentication systems and security frameworks make it difficult for fraudsters to carry out fraud [28] but some fraudsters use legitimate clients' information and manage to slip past the security net. New technologies would help insurance companies in increasing the possibility of detecting fraudulent claims [29].

V. Machine Learning Techniques

Machine learning is a process by which a computer learns, without being explicitly programmed [30], from training data and creates a prediction model based on learned data [31]. The learning can be classified as supervised, unsupervised, semi-supervised, or deep learning [32]. Machine learning technology is an approach for fraud detection that would identify a fraudulent claim earlier without human intervention.

1) Supervised Learning: In supervised learning, the dataset is properly labelled and each claim instance tagged as either fraudulent or non-fraudulent. A supervised model is trained by extensive amount of tagged training data [33] and is used for classification and regression problems.

2) Unsupervised Learning: Unsupervised learning model uses training dataset that is not labelled. The model groups data into sets according to their similarities (clustering) and also mines hidden relationships between data (association) [34].

3) Semi-supervised Learning: Semi-supervised learning is where some of the training data is not labelled.

4) Deep Learning: In deep learning, various representations of the data are learned in different layers of an artificial neural network to automate feature extraction [35].

Since fraud detection in motor insurance claims is to flag a claim as either fraudulent or valid, the models to detect fraud are trained with labelled dataset. Therefore supervised learning techniques are used.

VI. Supervised Learning Techniques For Fraud Detection

While building a fraud detection model, the sampling methodology employed, parameter selection, and identification techniques used all have a significant impact on the effectiveness of the model [36].

In this section, several researches relating to fraud discovery, detection and prediction using supervised learning techniques have been reviewed.

Use of decision trees (DT) is one common method for detecting motor insurance fraud. Dataset is split into smaller subsets using classification rules [37] and represented in form of nodes and leaves, where each node represent a feature and each leaf node ends with the outcome [38].A research for detecting automobile insurance fraud using supervised classifiers by [39] used DT C4.5 which had an accuracy of 93.6% with a sensitivity of 100%. However the specificity was 93.5% meaning that some legitimate cases were classified as fraudulent. Gradient boosted decision tree improved performance in fraud detection for medicare [40].

Extreme Gradient Boost (XGBoost) algorithm applied by [41] achieved an accuracy of 99.25% though it took more training and evaluation time than decision tree. A case study for fraud detection in automobile's body insurance observed that DT had a better efficiency, accuracy of 92.5% than Naive Bayes (NB), 90.28%, and Support Vector Machine (SVM) 30.28% [2]. There is a big difference between accuracy achieved by DT and SVM. The dataset used had only 360 damage instances with 91 being fraudulent and 269 were non-fraudulent.

In an approach to detect credit card fraud by [42], feature engineering is performed to create new attributes from existing features. This increased classification accuracy. XGBoost algorithm was used and performed well compared to other algorithms and approaches such as RF, LR and DT. The research attributed the good performance to XGBoost being an ensemble model that uses boosting method. Despite the outstanding performance, the dataset set was highly imbalanced.

Results by [3] in detecting insurance claims also indicate that DT and RF (Random Forest) outperformed NB. A research by [10] for identification of fraud cases introduced into insurance organizations using data mining methods of Decision Tree, Nearest Neighbor, and Neural Networks. Using the same variables for each method, the study indicates that decision tree method had the best performance with 66.90% accuracy. Kho and Vea [43] evaluated several classifiers, BayesNet, NB, J48, RandomTree, libSVM, and MODLEM, for fraud detection in credit card using transaction behavior and used 66% split validation. J48 and Random Tree (RT) were observed to have the highest accuracy rate of 93.5% and 94.32% respectively. DT has a higher classification accuracy compared to SVM and Random Forest (RF) as observed by [44]. However, imbalanced dataset was used and accuracy could be attributed to oversampling.

While analysing effect of imbalanced data in detecting automobile insurance fraud, [45] used adaptive oversampling technique (ADASYN) to delete imbalance classes. The research also used 10-fold cross validation in SVM, Multi Layer Perceptron (MLP) and DT. After balancing the dataset, the sensitivity of SVM increased from 70.76% to 94.74% while that of DT increased from 86.94% to 94.52%. SVM technique maximizes margin hyper plane which categorizes input samples into two classes [33]. SVMs determine the best separating hyperplane by mapping the input space into a higher dimensional feature space. Due to their capacity for extracting important and relevant features, SVMs are suitable for detection problems with a highly unbalanced data set [46]. In a systematic review for anomaly detection using machine learning observed SVM is commonly used by most researchers [47]. Sundarkumar et al [48] used a one-class support vector machine (OCSVM) based on undersampling method in a study for improving identification of fraud in insurance companies. The original dataset with 12,335 cases was divided into two in the ratio 80:20. 80% was subjected to under-sampling for model training while 20% was set aside to validate the effectiveness of the model. Classification techniques used included Logistic Regression, Group Method of Data Handling, SVM and Decision Tree. The research showed that Decision Tree decreased model's complexity and outperformed other techniques.

An improved technique for detecting credit card fraud that uses an SVM and RF as a feature selection algorithm to identify transactions that are anomalous was proposed by [49]. Relevant features were selected from the dataset using RF Classifier Method. This dataset is highly imbalanced with only 492 fraudulent transactions out of 284,807 transactions (0.17%). SVM is then used to classify the transactions as legitimate or fraudulent. The research observed that SVM based on RFC has a good accuracy of 95% , precision of 91% and sensitivity of 87%. However, the model did not address class imbalance of the dataset

A health insurance claim fraud detection approach introduced by [50] integrates three mining approaches: Association Rule Mining- on the basis of data correlation analysis to find frequent patterns, K-Means Clustering that increases performance and reduces time complexity, and Outlier Detection to expose insurance claim frauds. The study classified fraudulent behavior into two categories: period based claim anomalies and disease based anomalies. A study by [51] for fraud detection in auto insurance used Nearest Neighbor based Method (Distance and Density Based) and Statistics Methods (Interquartile range). When using original dataset with 33 features, the methods had an accuracy of: SVM- 82%, Distance based- 94.4%, Density based- 35.2% and Interquartile range- 92.1%. The research further applied feature selection process to simplify computational time and increased performance. The attributes reduced to 7 (seven). With the selected features, the accuracy changes to 82%, 99.9%, 82%,

and 98% respectively. It is observed that using feature selection would increase performance of models in detecting the occurrence of fraud. Feature selection also reduces complexity of the model.

While analyzing insurance claims using machine learning for fraud detection, [52] observed that feature selection reduces number of independent states from the features that are having very high unique values. Also converting categorical values to numerical ones improves the results of the algorithm. Using F1 Score as evaluation metric, XGBoost had a score of 81% while DT and KNN had 71.86% and 68% respectively. The research did not address class imbalance in the dataset.

In performing exploratory analysis for credit card fraud detection, [53] used several techniques such as DT,KNN,Neural Network, LR and their mixtures. The experiment used k-fold cross validation process to ensure equal representation of all data. LR was observed to have higher accuracy, majority of techniques under fit. Feature selection and class balancing was not done to the dataset.

Research by [54] for credit card fraud first extracted some features to assist in determining behavioural patterns using Sliding-Window Method. Experiments showed that SMOTE dataset provides better results compared to imbalanced dataset. An alternative for handling class imbalance by use of one-class classifiers, using OCSVM, also improved the results. Evaluation was done using MCC (Matthews Correlation Coefficient) metric. RF was observed to perform better than DT, LR and Local Outlier Factor.

Baga et al [55] applied investigated performance of 9 different classifiers in credit card fraud detection. The classifiers include LR,NB, KNN, MLP, RF, Pipelining and Ensemble Learning. Dataset was balanced using ADASYN method.In examination of performance, Ensemble Learning and Pipelining performed significantly better than all other models with accuracy of 99.99% and 99.999% respectively. Random Forest had an accuracy of 99.7%. Pipelining classification started with series of transformation followed by RF as the classifier, thus improving the accuracy. In Ensemble Learning, bagging classification was applied with RF as the base classifier.

VII. PROPOSED FRAMEWORK

Despite remarkable performance by models used to detect fraud, some valid claims are classified as fraudulent. The problems of the complexity of data, behavioral analysis for feature selection, and class imbalance also affect the performance of the techniques.

This research proposes a framework (see FIgure 1) of an ensemble model, trained with a balanced dataset and with the most relevant features. It proposes use of Synthetic Minority Oversampling Technique (SMOTE) to balance the dataset. This approach will create synthetic instances of minority class and ensures important details of the majority class are not lost. The ensemble model will be built by combining classifiers towards improving the performance of these individual classifiers. By selecting a suitable base classifier and performing a series of transformations on the training process; the ensemble model improves the performance and efficiency of supervised learning techniques and reduces the false positives [42].

VIII. CONCLUSION

The primary objective of this paper is to review some supervised learning techniques and their performance in motor insurance claims fraud detection. It has been observed that motor insurance claim fraud has become a huge problem for the insurance industry. Many researchers have been working actively in building and developing fraud detection systems that will help mitigate this problem. It was important to measure the performance of various supervised learning techniques in classifying claims as either fraudulent or not. Most of the research was based on a comparative study of the performance of various techniques.

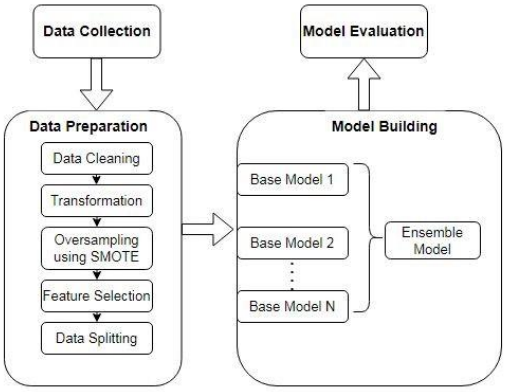

Fig. 1. Proposed Framework

Decision Trees (DT) are observed to decrease model complexity. It splits data by using nodes to represent features and branches will represent values that the feature can hold. It is able to predict highly non-linear data. However, it has a weakness of overfitting and takes more time to train. To reduce overfitting and enhance accuracy, boosting ensemble learning techniques can be used such as Gradient boosting DT. An alternative ensemble learning technique is bagging such as Random Forests.

XGBoost (Extreme Gradient Boosting) is an ensemble technique that uses boosting approach with more precise approximations and improved model prediction. Although the model's performance and accuracy increase, more training and evaluation time is needed.

Random Forests (RF) uses DT as base learners and creates a tree from bootstrap sample of the primary dataset. Multiple trees are combine by averaging their results thus decreasing

overfitting. More computational power and resources are need to create the multiple trees. It is able to handle outliers and works well even when the dataset have noise. Naive Bayes (NB) is a probabilistic approach that requires prior probabilities and performs poorly when features are related. It is ideal for a limited training data and less sensitive to missing data.

Logistic Regression (LR) is based on sigmoid function. It is faster and has low computing speed. Training this model does not make assumptions about class distribution in the feature space. However, if the feature space is large, its performance decreases. It may also be affected by underfitting.

K-Nearest Neigbours (KNN) works by locating distances within side, either averages of maximum common label. The advantage is that it only requires two parameters: k and distance, thus easy to understand and implement. Though, in large unbalanced datasets, it is computationally expensive and performs poorly.

Support Vector Machine (SVM) uses set of rules to create a hyperplane that separates input vector in a highly dimensional feature space into classes. The technique is memory efficient and tolerates redundant and irrelevant attributes. Its weakness include slow learning when the number of attributes and instances increase.

The strengths and weaknesses of techniques reviewed are summarized in Table I.

TABLE I SUMMARY: SUPERVISED LEARNING TECHNIQUES FOR FRAUD DETECTION

Technique	Strength	Weakness
Decision Tree (DT)	Reduced complexity. Able to predict highly non-linear data.	Overfitting. More training time.
XGBoost	More precise approximations and improved prediction.	Higher training and evaluation time.
Random Forest (RF)	Decreased overfitting for base decision trees. Ability to handle outliers. Works well with noisy data.	More computational power and resources.
Naive Bayes (NB)	Less sensitive to missing data in dataset. Ideal for limited training data.	Underfitting. Performs poorly when features are related.
Logistic Regression (LR)	Fast training and low computing power.	Underfitting. Performance reduces and feature space increases.
K-Nearest Neighbour (KNN)	Requires only two parameters; k and distance.	Computationally expensive for large unbalanced datasets.
Support Vector Machine (SVM)	Tolerates redundant and irrelevant attributes. Memory efficient.	Increase in attributes and instances reduces its learning speed.

Over all datasets, there is no specific technique that would perform better than all other techniques in fraud detection. The research also observed that the measure used to evaluate and compare the performance is confusion matrix where metrics such as accuracy, recall, precision, F1-score, and MCC (Matthews Correlation Coefficient) are obtained.

REFERENCES

[1] B. Bart, H. Sebastiaan, and C. Oppner, "Verdonck.(2021)," *Data engineering for fraud detection. Decision Support Systems Jurnal. https://doi. org/10.1016/j. dss*, 2021.

[2] L. Goleiji and M. J. Tarokh, "Fraud detection in the insurance using decision tree, naive bayesian and support vector machine data mining algorithms (case study-automobile's body insurance)," 2016.

[3] R. Roy and K. T. George, "Detecting insurance claims fraud using machine learning techniques," in *2017 international conference on circuit, power and computing technologies (ICCPCT)*. IEEE, 2017, pp. 1–6.

[4] A. Association of Kenya Insurers, "Insurance Industry Report 2021," Tech. Rep., 2021.

[5] N. Remli, F. Salleh, and J. Arifin, "Motor insurance fraudulent claims: An overview reconnaissance," *International Journal of Business, Economics and Law*, vol. 25, 2021.

[6] S. Subudhi and S. Panigrahi, "Detection of automobile insurance fraud using feature selection and data mining techniques," *International Journal of Rough Sets and Data Analysis (IJRSDA)*, vol. 5, no. 3, pp. 1–20, 2018.

[7] B. Itri, Y. Mohamed, Q. Mohammed, and B. Omar, "Performance comparative study of machine learning algorithms for automobile insurance fraud detection," in *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)*. IEEE, 2019, pp. 1–4.

[8] G. Kowshalya and M. Nandhini, "Predicting fraudulent claims in automobile insurance," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, 2018, pp. 1338–1343.

[9] M. A. Rashid, A. Al-Mamun, H. Roudaki, and Q. R. Yasser, "An overview of corporate fraud and its prevention approach," *Australasian Accounting Business & Finance Journal*, vol. 16, no. 1, pp. 101–118, 2022.

[10] M. H. AYBOGA and F. Ganji, "Detecting fraud in insurance companies and solutions to fight it using coverage data in the covid 19 pandemic," *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 18, no. 15, pp. 392–407, 2021.

[11] U. Rani, O. L. Pramudyastuti, and A. P. Nugraheni, "Disclosing the practice of whistleblowing system in indonesiaˆG™ s public listed companies," *INOVASI*, vol. 18, pp. 79–87, 2022.

[12] N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud, "A secure ai-driven architecture for automated insurance systems: Fraud detection and risk measurement," *IEEE Access*, vol. 8, pp. 58546–58558, 2020.

[13] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019," *Computer Science Review*, vol. 40, p. 100402, 2021.

[14] M. Guillen, J. P. Nielsen, and A. M. Perez-Mar´ ´ın, "Near-miss telematics in motor insurance," *Journal of Risk and Insurance*, vol. 88, no. 3, pp. 569–589, 2021.

[15] L. Rukhsar, W. H. Bangyal, K. Nisar, and S. Nisar, "Prediction of insurance fraud detection using machine learning algorithms," *Mehran University Research Journal Of Engineering Technology*, vol. 41, no. 1, p. 33–40, 2022. [Online]. Available: https://search.informit.org/doi/10.3316/informit.263147785515876

[16] A. Kini, R. Chelluru, K. Naik, D. Naik, S. Aswale, and P. Shetgaonkar, "Automobile insurance fraud detection: An overview," in *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, 2022, pp. 7–12.

[17] C. Eckert and K. Osterrieder, "How digitalization affects insurance companies: overview and use cases of digital technologies," *Zeitschrift fur die gesamte Versicherungswissenschaft¨* , vol. 109, 10 2020.

[18] H. L. Sithic and T. Balasubramanian, "Survey of insurance fraud detection using data mining techniques," *arXiv preprint arXiv:1309.0806*, 2013.

[19] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *Journal of Network and Computer Applications*, vol. 68, pp. 90–113, 2016.

[20] N. J. Morley, L. J. Ball, and T. C. Ormerod, "How the detection of insurance fraud succeeds and fails," *Psychology, Crime & Law*, vol. 12, no. 2, pp. 163–180, 2006.

[21] O. Tajudeen and R. Abdur, "Control of insurance fraud in nigeria: an exploratory study (case study)," *J. Financ. Crime*, vol. 16, no. 4, pp. 418–435, 2009.

[22] E. J. Efiong, I. O. Inyang, and U. Joshua, "Effectiveness of the mechanisms of fraud prevention and detection in nigeria," *Advances in Social Sciences Research Journal*, vol. 3, no. 3, 2016.

[23] R. Othman, N. A. Aris, A. Mardziyah, N. Zainan, and N. M. Amin, "Fraud detection and prevention methods in the malaysian public sector: Accountants' and internal auditors' perceptions," *Procedia Economics and Finance*, vol. 28, pp. 59–67, 2015.

[24] S. W. Mwangi and J. Ndegwa, "The influence of fraud risk management on fraud occurrence in kenyan listed companies," International Journal of Finance amp; Banking Studies (21474486), vol. 9, no. 4, p. 147–160, 2020. [Online]. Available: https://www.ssbfnet.com/ojs/index.php/ijfbs/article/view/943

[25] J. O. Otieno, "Insurance stakeholders' perceptions on effectiveness and usage of fraud detection and prevention techniques in motor insurance sector," Ph.D. dissertation, Strathmore University, 2018.

[26] W. S. Albrecht, C. Albrecht, C. Albrecht, and M. Zimbelman, "Fraud examination 2e," *Baski. Thomson South-Western*, 2006.

[27] S. Viaene, M. Ayuso, M. Guillen, D. Van Gheel, and G. Dedene, "Strategies for detecting fraudulent claims in the automobile insurance industry," *European Journal of Operational Research*, vol. 176, no. 1, pp. 565–583, 2007.

[28] Z. A. Soomro, J. Ahmed, M. H. Shah, and K. Khoumbati, "Investigating identity fraud management practices in e-tail sector: a systematic review," *Journal of Enterprise Information Management*, 2019.

[29] E. Fernando, "Machine learning approaches on motor insurance fraud detection," Ph.D. dissertation, 2022.

[30] R. Garcia-Dias, A. Mechelli, W. L. Pinaya, and S. Vieira, "Autoencoders," *Machine Learning: Methods and Applications to Brain Disorders, Academic Press, Cambridge*, 2019.

[31] P. Singh, S. P. Singh, and D. S. Singh, "An introduction and review on machine learning applications in medicine and healthcare," *2019 IEEE Conference on Information and Communication Technology, CICT 2019*, 2019.

[32] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.

[33] M. N. Ashtiani and B. Raahemi, "Intelligent Fraud Detection in Financial Statements using Machine Learning and Data Mining: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 72504–72525, 2021.

[34] J. Liu, "From statistics to data mining: A brief review," *Proceedings 2020 International Conference on Computing and Data Science, CDS 2020*, vol. 7, pp. 343–346, 2020.

[35] D. Sarkar, R. Bali, and T. Sharma, *Practical Machine Learning with Python A Problem-Solver's Guide to Building Real-World Intelligent Systems*. Library of Congress Control Number: 2017963290, 2018.

[36] N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, "An efficient credit card fraud detection model based on machine learning methods," *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 3414–3424, 2020.

[37] N. Fatima, L. Liu, S. Hong, and H. Ahmed, "Prediction of breast cancer, comparative review of machine learning techniques, and their analysis," *IEEE Access*, vol. 8, pp. 150360–150376, 2020.

[38] R. Hegde, G. Anusha, S. Madival, H. Sowjanya, and U. Sushma, "A review on data mining and machine learning methods for student scholarship prediction," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2021, pp. 923– 927.

[39] I. M. N. Prasasti, A. Dhini, and E. Laoh, "Automobile insurance fraud detection using supervised classifiers," in *2020 International Workshop on Big Data and Information Security (IWBIS)*. IEEE, 2020, pp. 47–52.

[40] J. T. Hancock and T. M. Khoshgoftaar, "Gradient boosted decision tree algorithms for medicare fraud detection," *SN Computer Science*, vol. 2, no. 4, p. 268, 2021.

[41] N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud, "Extreme gradient boosting machine learning algorithm for safe auto insurance operations," in *2019 IEEE international conference on vehicular electronics and safety (ICVES)*. IEEE, 2019, pp. 1–5.

[42] D. X. Cho, D. N. Phong, and N. Duy Phuong, "A new approach for detecting credit card fraud transaction," *International Journal of Nonlinear Analysis and Applications*, 2023.

[43] J. R. D. Kho and L. A. Vea, "Credit card fraud detection based on transaction behavior," in *TENCON 2017-2017 IEEE Region 10 Conference*. IEEE, 2017, pp. 1880–884.

[44] J. V. Devi and K. Kavitha, "Fraud detection in credit card transactions by using classification algorithms," in *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*. IEEE, 2017, pp. 125–131.

[45] S. Subudhi and S. Panigrahi, "Effect of class imbalanceness in detecting automobile insurance fraud," in *2018 2nd International Conference on Data Science and Business Analytics (ICDSBA)*. IEEE, 2018, pp. 528–531.

[46] W. Hilal, S. A. Gadsden, and J. Yawney, "Financial fraud: A review of anomaly detection techniques and recent advances," *Expert Systems with Applications*, vol. 193, p. 116429, 2022.

[47] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: A systematic review," *Ieee Access*, vol. 9, pp. 78658–78700, 2021.

[48] G. G. Sundarkumar, V. Ravi, and V. Siddeshwar, "One-class support vector machine based undersampling: Application to churn prediction and insurance fraud detection," in *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. IEEE, 2015, pp. 1–7.

[49] N. Rtayli and N. Enneya, "Selection features and support vector machine for credit card risk identification," *Procedia Manufacturing*, vol. 46, pp. 941–948, 2020.

[50] A. Verma, A. Taneja, and A. Arora, "Fraud detection and frequent pattern matching in insurance claims using data mining techniques," in *2017 tenth international conference on contemporary computing (IC3)*. IEEE, 2017, pp. 1–7.

[51] T. Badriyah, L. Rahmaniah, and I. Syarif, "Nearest neighbour and statistics method based for detecting fraud in auto insurance," in *2018 International Conference on Applied Engineering (ICAE)*. IEEE, 2018, pp. 1–5.

[52] A. Urunkar, A. Khot, R. Bhat, and N. Mudegol, "Fraud detection and analysis for insurance claim using machine learning," in *2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*, vol. 1. IEEE, 2022, pp. 406–411.

[53] M. Madhurya, H. Gururaj, B. Soundarya, K. Vidyashree, and A. Rajendra, "Exploratory analysis of credit card fraud detection using machine learning techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 31–37, 2022.

[54] V. N. Dornadula and S. Geetha, "Credit card fraud detection using machine learning algorithms," *Procedia computer science*, vol. 165, pp. 631–641, 2019.

[55] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, "Credit card fraud detection using pipeling and ensemble learning," *Procedia Computer Science*, vol. 173, pp. 104–112, 2020.