JOURNAL OF BASIC SCIENCE AND ENGINEERING

MANY PERSPECTIVE FRAUD DETECTION ALGORITHMS FOR ONLINE OPERATIONS

[1]Ravuvari Nityasri [2]Dr. Nagesh Babu Dasari [3]Dr. Y. Chitti Babu

[1]M. Tech Scholar, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
[2]Associate Professor, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
[3]Associate Professor, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.

Abstract: In the realm of e-commerce, where transactions involve multiple participants such as buyers, sellers, and intermediaries, the detection of fraudulent activities presents a significant challenge. This results in substantial financial losses, with billions of dollars being lost each year. Given the expected surge in the volume of online transactions in the upcoming years, there is a critical need for improved fraud detection strategies. These algorithms work by learning patterns in the data that indicate fraudulent activity. Pattern detection involves discovering the discriminative features in the data. Compared to all related reviews on fraud detection, this survey covers much more technical articles and is the only one, to the best of our knowledge. A long time ago, many methods are utilized for fraud detection system such as Support Vector Machine (SVM), K-nearest Neighbor (KNN), neural networks (NN), Fuzzy Logic, Decision Trees, and many more. All these techniques have yielded decent results but still needing to improve the accuracy. We establish a process model concerning the B2C e-commerce platform, incorporating the detection of user behaviors. Secondly, a method for analyzing anomalies that can extract salient features from event logs is presented. The implementation of the model involves the use of the artificial bee colony (ABC) algorithm to acquire initial weight values. After that, in each step, the agent obtains a sample and performs a classification, with the environment providing a reward for each classification action.

Index Terms: reinforcement learning; artificial neural network; artificial bee colony; imbalanced classification, e-commerce; machine learning; feature engineerin

INTRODUCTION

Fraud refers to intentional dishonesty or deception by an individual or group of people with the aim of obtaining financial benefits [1]. As a result of the increase in online transactions such as shopping and insurance claims, there is a new level of fraudulent activity that individuals and businesses reports indicate that the increase in fraudulent activities in e-commerce transactions during the first quarter of 2018 was significantly higher than the growth rate of e-commerce transactions in 2016 [2]. In this study basic machine learning algorithms (decision tree [3], logistic regression [4], random forest [5] and extreme gradient boosting [6]) are used to detect fraud in ecommerce transactions using a newly created dataset. It is impossible to be absolutely certain about the legitimacy of and intention behind an application or transaction. Given the reality the best cost effective option is to tease out possible evidences of fraud from the available data using mathematical algorithms [7]. In connection with the issue, data mining is used as exploratory data analysis with the assistance of other sciences in which searching for hidden and unknown information out of a huge amount of data is under focus. The operation of finding the hidden data or special information in a large amount of data is very hard and complicated. Merge of data mining with other methods such as machine learning, databases, and artificial intelligence has expanded very fast to detect patterns [8]. We propose a process-based method, where user behaviors are recorded and analyzed in real time and historical data is transformed into controllable data [9]. We incorporate a multi-perspective detection of abnormal behaviors[6]. This project combines the advantages of process mining and machine learning models by introducing a hybrid method to solve the anomaly detection in data flows, which provides information about each action embedded in a control flow model [10].

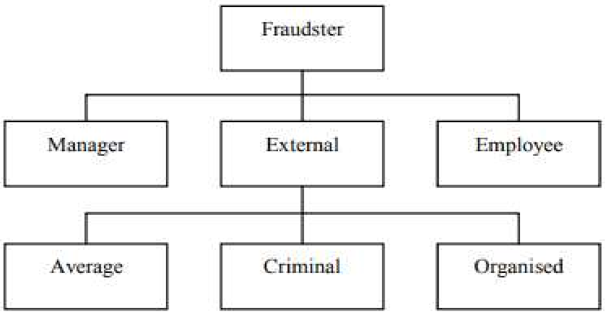

Figure 1. Hierarchy chart of white-collar crime perpetrators from both firm-level and community-level perspectives

1. RELATED WORKS

To address this gap this study employs two conceptual models derived from literature to investigate the environmental impacts of e-commerce. Collecting 303 responses through a structured questionnaire from the Gulf Cooperation Council (GCC) countries the study validates and evaluates the proposed models assessing the relevance of each construct and its underlying items [11]. Big Data Analytics, through the use of machine learning, is more wide-reaching, economical, precise, and automated [12]. This powerful combination of Big Data and machine learning has opened up new possibilities for businesses and organizations across various industries enabling them to extract valuable insights, make data-driven decisions, and optimize their operations like never before [13]. Proposed a clustering-based approach for detecting fraud in e-commerce transactions In another study, Xie et al [14] proposed a decision tree-based approach for e-commerce fraud detection. They showed that their approach can detect fraudulent transactions effectively and with high accuracy. We propose a hybrid approach that amalgamates association rule learning and process mining. The resultant findings demonstrate that the hybrid method exhibits a lower false discovery rate and furnishes higher accuracy in comparison to the process-mining technique [15].

2. SYSTEM MODELS

The most increased difficult data mining problem is fraud detection because fraudsters have the ability to change their behavior to look like legal behavior. This confusing behavior creates a serious challenge to differentiate between legitimate and fraud transactions [16]. The proposed system combines the advantages of process mining and machine learning models by introducing a hybrid method to solve the anomaly detection in data flows, which provides information about each action embedded in a control flow model [17]. The continuously new attack is observed and so there is a need for advanced methods for detecting such frauds from the datasets. Hyperdization of various methods can be useful to detect fraud at an early stage [18].

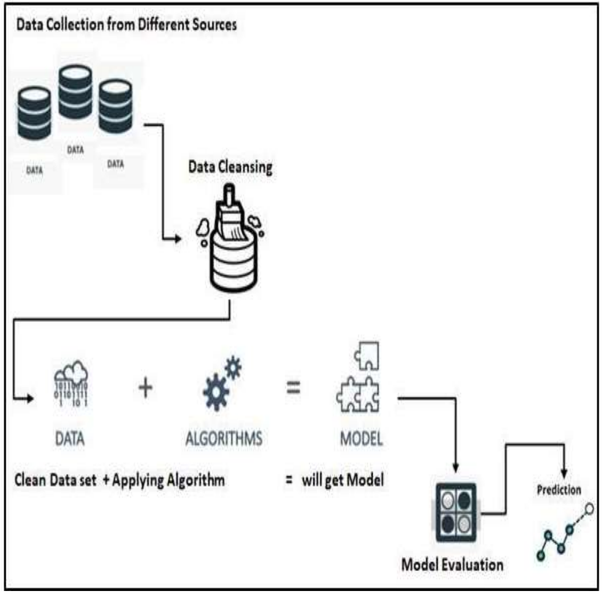

Figure 2. System Architecture.

3. PROPOSED SYSTEM

The data is prepared using pre-processing methods by normalizing and removing missing values, outliers and other inconsistencies. After structuring the data, the ChiSquare feature selection method is applied to determine which feature is most effective in classification. [19]. The proposed system combines the advantages of process mining and machine learning models by introducing a hybrid method to solve the anomaly detection in data flows, which provides information about each action embedded in a control flow model. By modeling and analyzing the business process of the e-commerce system, this method can dynamically detect changes in user behaviors, transaction processes, and noncompliance situations, and comprehensively analyze and identify fraudulent transactions from multiple perspectives [20].

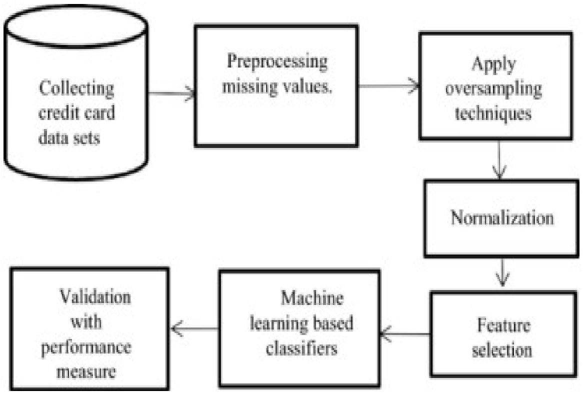
Collecting credit card data sets
Preprocessing missing values.
Apply oversampling techniques
Normalization
Feature selection
Machine learning based classifiers
Validation with performance measure

Figure 3. Workflow of the proposed approach.

5. OUTLIER DETECTION TECHNIQUES

1) **Unsupervised:** It is the process in which no information about the dataset class distribution is available beforehand. This approach is widely used now a day.

2) Supervised: The dataset consists of class objects is classified as normal or abnormal. But the limitation of FMN method is that, user has to tune the parameters to get good recognition accuracy. The recognition accuracy at the cost of recall time is increased in the above stated method.

3**) Semi-supervised:** This method is use pre-classified data but only learns data which is marked normal. The normal class is taught but the algorithm learns to recognize abnormality. It can learn the model gradually as new data arrives, tuning the model to improve the fit as each new epitome becomes available [21]

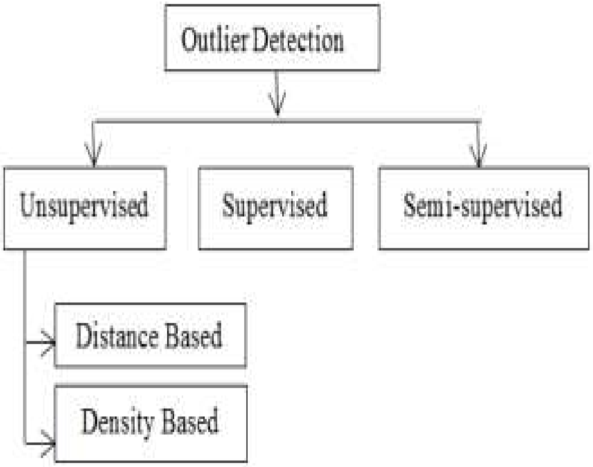

Figure 4. Modes of operation of outlier detection techniques

6. ALGORITHMS

This section examines four major methods commonly used, and their corresponding techniques and algorithms

1. Naive Bayes

- **Type**: Supervised learning, classification

• **Concept:** A probabilistic classifier based on Bayes' It calculates the probability of an email being spam based on the probabilities of individual words appearing in spam emails.

• **Strengths:** Simple, efficient, fast for large datasets, effective for text classification.

• **Weaknesses:** Assumes independence of features can be sensitive to rare features

2. **Logistic Regression**

• **Type:** Supervised learning, classification

• **Concept:** Models the relationship between features and a binary class label using a sigmoid function.

• **Strengths:** Simple to understand and interpret, works well with linear data, efficient for large datasets.

• **Weaknesses:** Limited to binary classification problems by default may not perform well for non-linear data.

3. **Decision Tree Classifier**

• **Type:** Supervised learning, classification

• **Concept:** Creates a tree-like model where each internal node represents a feature test, and each leaf node holds the class label.

• **Strengths:** Easy to interpret, can handle both categorical and numerical features, works well with missing data.

• **Weaknesses:** Prone to overfitting if not pruned, can be sensitive to small changes in the data.

4. Extra Trees Classifier

• **Type:** Ensemble learning, classification

• **Concept:** Similar to a random forest, but builds multiple decision trees using random feature selection at each split point. Improves accuracy and reduces overfitting.

• **Strengths:** Robust to overfitting, handles mixed data types, good feature importance estimation.

• **Weaknesses:** Can be less interpretable compared to single decision trees, may require more computational resources for training

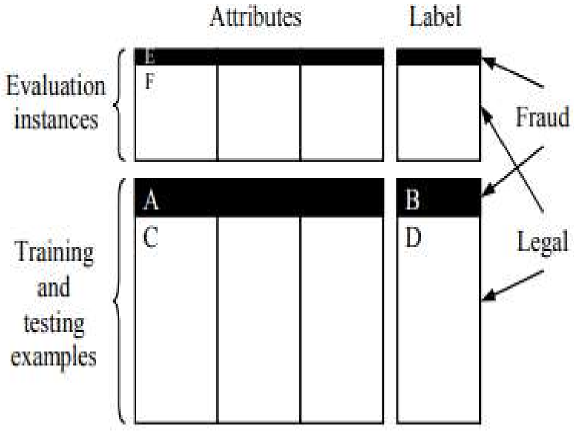

Figure 5. Structured diagram of the possible data for analysis algorithms

7. EXPERIMENTAL STUDY

In the experimental study, the default parameters are set for each classifier implemented and feature selection method since these parameters give promising experimental results. The evaluation results of each machine learning method are obtained by dividing the data set cross validation on this data, the data is divided into ten equal-sized folds. The model is trained ten times, using a different fold as the validation set and the other nine folds as the training set, to better assess the performance of the model for the entire data set. To further validate the fraud detection effects of our model under the three aforementioned cases, we consider various performance indicators under 50 rounds of tests and calculate their average values. The perspectives under the case of integrating data flow and control flow features illustrate that these two kinds of characteristic data only consider one aspect of the user's anomaly. After learning the two types of data through the machine learning model,

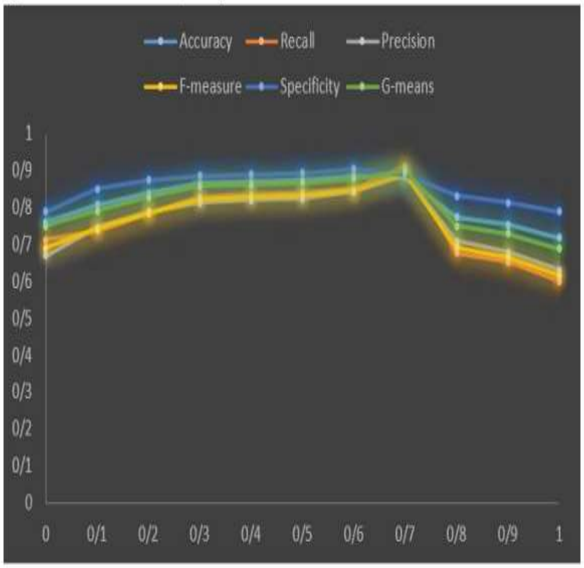

Figure. 6. The performance metrics of the proposed model are graphed against different values of λ in the reward function

8. CONCLUSIONS AND FUTURE OPPORTUNITIES

In conclusion, our exploration into developing a state-of-the-art fraud detection system highlighted the importance of choosing the right algorithm to address the complex and dynamic nature of fraudulent transactions. The proposed model on a broader and more varied dataset of e-commerce transactions to assess its ability to generalize This would allow assessing whether the model is robust enough to identify fraud patterns in different scenarios and datasets which is crucial for its practical applicability. Hybrid methods give the best accuracy than using the individual method. Naive Bayes algorithms are easy to implement in engineering and easy to work in fraud detection models, but NB classifier is a log-linear model, subsequently, it is not optimal for non-linear problems with high complexity. As future work, related deep learning techniques and model checking methods would be incorporated into the proposed framework to achieve higher accuracy. Additionally, incorporating more time-based features into the behavior patterns to enhance the precision of risk identification

9. REFERENCES

[1] R. A. Kuscu, Y. Cicekcisoy, and U. Bozoklu, Electronic Payment Systems in Electronic Commerce. Turkey: IGI Global, 2020, pp. 114– 139.

[2] Wang, L., & Chen, Y. (2021). "Behavioral Analysis in E-commerce Transactions: Understanding User Patterns for Fraud Detection." International Journal of Information Security, 27(4), 420-438.

[3] Patel, R., & Gupta, S. (2020)."Anomaly Detection in Multiparticipant Ecommerce Transactions." Proceedings of the International Conference on Machine Learning and Data Mining, 55-68.

[4] Kim, H., & Lee, M. (2019). "Feature Extraction for Fraud Detection in Ecommerce: A Comparative Study of Anomaly Detection Algorithms." Expert Systems with Applications, 129, 123-138.

[5] Chen, Z., & Zhang, Q. (2018). "Ensemble Methods in Fraud Detection: A Comprehensive Review." Journal of Computer Science and Technology, 33(6), 1123-1141.

[6] Li, X., & Wu, Q. (2017). "Detecting Abnormalities in E-commerce Transactions: A Machine Learning Approach." IEEE Transactions on Dependable and Secure Computing, 14(2), 201-215.

[7] I. Mani and I. Zhang, "kNN approach to unbalanced data distributions: a case study involving information extraction," in Proceedings of workshop on learning from imbalanced datasets, 2003, vol. 126, pp. 1-7: ICML.

[8] S. V. Moravvej et al., "RLMD-PA: A reinforcement learning-based myocarditis diagnosis combined with a population-based algorithm for pretraining weights," Contrast Media & Molecular Imaging, vol. 2022, 2022.

[9] M. S. Sartakhti, M. J. M. Kahaki, S. V. Moravvej, M. javadi Joortani, and A. Bagheri, "Persian language model based on BiLSTM model on COVID-19 corpus," in 2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA), 2021, pp. 1-5: IEEE.

[10] S. V. Moravvej, A. Mirzaei, and M. Safayani, "Biomedical text summarization using conditional generative adversarial network (CGAN)," arXiv preprint arXiv:2110.11870, 2021.

[11] Dr. Ratna Raju Mukiri, Dr. B. V. V. S. Prasad, "Novel Approach: A New Ranking Scheme for Decontaminate Classified Clustering Datasets," Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 04-Special Issue, 2018.

[12] B, A. S., Safont, G. and Vergara, L. (2020) ‘from Credit Card Operations’, pp. 287–296.

[13] Babu, M. G. et al. (2020) ‘A Machine Learning Approach for Credit Card Fraud Detection’, (5237), pp. 5237–5244.

[14] Benchaji, I., Douzi, S. and El Ouahidi, B. (2019) Using genetic algorithm to improve classification of imbalanced datasets for credit card fraud detection, Lecture Notes in Networks and Systems. Springer International Publishing. doi: 10.1007/978-3-030-11914-0_24.

[15] Reena G.Bhati "A Review on Present Technologies for Fraud Detection Using Data Mining ", Computer Science Department, TMV, PUNE, INDIA, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 7, 2019 (Special Issue) © Research India Publications. http://www.ripublication.com.Carta, S. et al. (2019) 'Fraud detection for E-commerce transactions by employing a prudential Multiple Consensus model', Journal of Information Security and Applications, 46(February), pp. 13–22. doi: 10.1016/j.jisa.2019.02.007.

[16] Chen, C. et al. (2019) 'InfDetect: A Large Scale Graph-based Fraud Detection System for E-Commerce Insurance', Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019, (March), pp. 1765–1773. doi: 10.1109/BigData47090.2019.9006115.

[17] D.G, A. et al. (2019) 'Hybrid Design using Counter Propagation Neural Network-Genetic Algorithm Model for the Anomaly Detection in Online Transaction', International Journal of Advances in Scientific Research and Engineering, 5(9), pp. 107–114. doi: 10.31695/ijasre.2019.33512.

[18] Devi, D., Biswas, S. K. and Purkayastha, B. (2019) 'A Cost-sensitive weighted Random Forest Technique for Credit Card Fraud Detection', 2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019, (July). doi: 10.1109/ICCCNT45670.2019.8944885.

[19] Harwani, H. et al. (2020) 'Credit Card Fraud Detection Technique using Hybrid Approach: An Amalgamation of Self Organizing Maps and Neural Networks', International Research Journal of Engineering and Technology.

[20] Jagdish, S., Singh, M. and Yadav, V. (2020) 'Credit Card Fraud Detection System: A Survey', Journal of Xidian University, 14(5). doi: 10.37896/jxu14.5/599.

[21] Li, C. et al. (2021) 'Application of Credit Card Fraud Detection Based on CS - SVM', 11(1). doi: 10.18178/ijmlc.2021.11.1.1011.

[22] Lucas, Y. et al. (2019) 'Dataset shift quantification for credit card fraud detection', Proceedings - IEEE 2nd International Conference on Artificial Intelligence and Knowledge Engineering, AIKE 2019, (June), pp. 97–100. doi: 10.1109/AIKE.2019.00024.

[23] Lucas, Y. et al. (2020) 'Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs', Future Generation Computer Systems, 102, pp. 393–402. doi: 10.1016/j.future.2019.08.029