A MULTI-PERSPECTIVE FRAUD DETECTION METHOD FOR MULTI-PARTICIPANT E-COMMERCE TRANSACTIONS

[1]MRS.CH.DEEPTHI, [2] Polisetti Ramya

[1]Assistant Professor, Department of Master of Computer Applications,
QIS College of Engineering & Technology, Ongole, Andhra Pradesh, India
[2]PG Scholar, Department of Master of Computer Applications,
QIS College of Engineering & Technology, Ongole, Andhra Pradesh, India

ABSTRACT

Detection and prevention of fraudulent transactions in e-commerce platforms have always been the focus of transaction security systems. However, due to the concealment of e-commerce, it is not easy to capture attackers solely based on the historic order information. Many researches try to develop technologies to prevent the frauds, which have not considered the dynamic behaviors of users from multiple perspectives. This leads to an inefficient detection of fraudulent behaviors. To this end, this paper proposes a novel fraud detection method that integrates machine-learning and process mining models to monitor real-time user behaviors. First, we establish a process model concerning the B2C e-commerce platform, by incorporating the detection of user behaviors. Second, a method for analyzing abnormalities that can extract important features from event logs is presented. Then, we feed the extracted features to a Support Vector Machine (SVM) based classification model that can detect fraud behaviors. We demonstrate the effectiveness of our method in capturing dynamic fraudulent behaviors in e-commerce systems through the experiments.

I. INTRODUCTION

With the increasing popularity of e-commerce platforms, more and more commercial transactions are now relying on web-based systems than the traditional cash-based approach [1]. Although the entity economy is greatly impacted by the COVID-19 epidemic in recent years, e-commerce remains largely unaffected by the pandemic, whereby aiding a steady market growth [2]. The sales volume of B2C (Business to Customer) e-commerce is expected to reach 6.5 trillion dollars by 2023 [3]. This paper 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. Important contributions of this paper are listed as follows:

1) A conformance checking method based on process mining is applied in the field of e-commerce transactions to capture the abnormalities.

2) A user behavior detection method is proposed to perform comprehensive anomaly detection based on Petri nets.

3) An SVM model is developed by embedding a multi perspective process mining into machine learning methods to automatically classify fraudulent behaviors. The rest of this paper is organized as follows: Section 2 introduces the related work. Section 3 presents a model analysis and a background study. Section 4 forms the theoretical basis and describes our proposed fraud detection method. Section 5 presents and discusses the results of our experiments and Section 6 validates our proposed fraud detection method. Section 7 concludes our paper along with outlining our future research directions.

II.EXISTING SYSTEM

The machine-learning-based methods learn from previously obtained historical data to perform classifications or predictions of future observations to identify potential risky offline or online transactions [6]. Xuetong Niu et al. conducted a comparative study on credit card fraud detection methods that rely on machine-learning algorithms. Most of the machine-learning models perform well on the dataset of credit card transactions. Moreover, supervised models perform slightly better than unsupervised models after additional pre-processing, such as removing outliers [7].

Credit card fraud detection is widely deployed at the application layer, which uses the idea of discovering specific abnormal user behaviors to detect fraud. The supervised learning algorithm is the most commonly used learning method in online fraud monitoring transactions, since it has higher accuracy and coverage. Recent research in [8, 9] has proved that the machine learning method can efficiently capture fraudulent transactions in credit card applications.

Fraudsters often change their behavioral pattern dynamically to overcome existing fraud detection methods. In online credit card fraud detection, SVM can classify user behaviors under complex scenarios and deliver reliable results [10]. Many researchers take the advantage of combining multiple detection

methods for comprehensive fraud detection. For example, focusing on payment fraud applications, Dahee Choi et al. proposed a method by combining supervised and unsupervised learning [11]. Most of the machine learning based methods use historical data to analyze fraudulent transactions. They have not given enough emphasis to the transactional process flow and dynamic user behaviors. The second type of fraud detection methods uses process mining, focusing on extracting knowledge from existing event logs in information systems for the purpose of monitoring and improving the operational process in business IT infrastructure [12]. Process mining specializes in comparing the event log with an established model to further detect, locate, and interpret the deviation between the established model and the actual event log [13].Process mining can detect a large number of abnormal transactions, which are not known to be identifiable by traditional methods. M Jans et al. postulated the emerging process mining approach as an appropriate solution to mitigate against fraud incorporating internal affairs [14]. For example, C Rinner et al. applied conformance checks to monitor the process of melanoma patients [15]. Asare et al. applied alignment and replay to check the conformance of the electronic medical record log and the hospital workflow model [16]. Research has focused on monitoring and evaluating the sequence of processes occurring in the historical medical event log by establishing corresponding training and testing models for conformance checking [17]. Tools such as ProM, Disco and Heustic miner are largely used for conformance checking. Process mining can be an efficient approach for fraud detection. Especially, it is important to be dynamic and multi perspective when detecting fraudulent user behaviors [18]. Process mining helps to compare the actual data against the standard model to identify outliers. Despite existing progress in fraud detection, it is still necessary to develop hybrid learning methods to improve the accuracy of detection [19]. To promote the understanding and development of process mining for anomaly detection, a method of multi-perspective anomaly detection is proposed that goes beyond the perspective of control flow including time and resources [20]. Febriyanti et al. [21] assumed any noticeable changes in business processes as a suspected fraud behavior and proposed a method to detect some suspicious abnormal behaviors using a hybrid method of association rules and process mining. Previous research on using process mining to detect fraudulent transactions showed that process mining is capable of detecting fraudulent transactions, and it can

effectively prevent audit fraud at a much earlier stage due to the continuous monitoring nature of event logs [22].

Disadvantages

1) Fraud mode one - an order is tempered by a malicious actor: The malicious actor may deceive the victim merchant by sending a fake formal payment order F

A to the cashier server. The malicious actor obtained the order items that do not match the payment value by tampering with the order information, such as the total amount.

2) Fraud mode two - subcontract the order: The victim pays the malicious actor's order instead of his order. To achieve their goals, the malicious actors impersonate the duties of sellers and buyers. The order information changes before and after the payment.

III.PROPOSED SYSTEM

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. Important contributions of this paper are listed as follows:

1) A conformance checking method based on process mining is applied in the field of e-commerce transactions to capture the abnormalities.

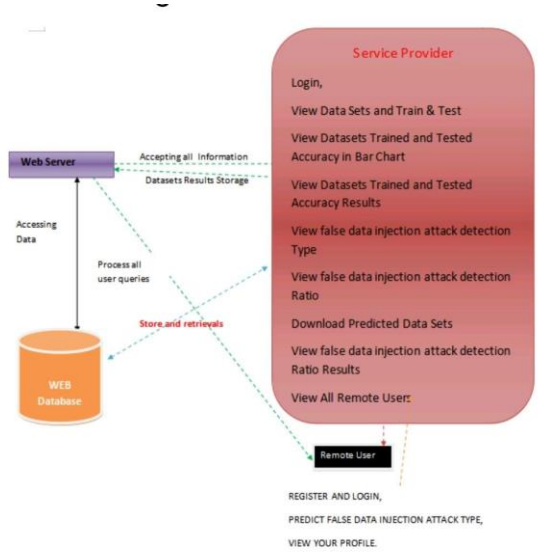
2) A user behavior detection method is proposed to perform comprehensive anomaly detection based on Petri nets.

3) An SVM model is developed by embedding a multi perspective process mining into machine learning methods to automatically classify fraudulent behaviors.

Advantages

- To arrive at a clearer result, the plug-in Multi-Perspective Process Explorer and Conformance Checking are used to match and analyze the event log and the DPN. The result is shown in this system, where each action is represented with different colors. For instance, green represents the move both on model and log, purple means move on the model only, and grey represents invisible actions, that is, skipped actions.
- By clicking on a given action, we can obtain the matching information between the model and the event log in the data flow of each action. The data marked in red indicates a

mismatch. We extract these suspicious anomalies and use them as the basis for subsequent training using machine learning models.



IV.ALGORITHMS

1. Logistic Regression:

Logistic Regression is a fundamental statistical method used for binary classification tasks. Despite its name, it is used for classification rather than regression. Logistic Regression models the probability of a binary outcome based on one or more predictor variables. It estimates the probability that a given input data point belongs to a particular class using a logistic (sigmoid) function, which transforms the output into a range of [0, 1]. The decision boundary is typically set at 0.5, classifying inputs with probabilities above 0.5 into one class and below into the other. Logistic Regression is simple yet effective for linearly separable data and provides interpretable results by estimating coefficients for each feature.

Algorithm 2: PSEUDO code for logistic regression algorithm

```
Step1: Function grad (predictor_attributes, target_attribute, weights)
            {
              Calculate gradient_descent;
              Return weights + learning_rate * gradient_descent;
            }
Step2: Normalize the dataset;
Step3: Repeat
    {
      Weights = grad (params);
      Update weights;
    } until convergence
Step4: z = dot product of predictor variables and updated weights;
Step5: prediction_limit = sigmoid function (z);
Step6: Predict the target class
```

2. Naive Bayes:

Naive Bayes is a probabilistic classifier based on Bayes' theorem with the "naive" assumption of independence between features. Despite its simplifying assumptions, Naive Bayes can be surprisingly effective in many real-world applications, especially in text classification and spam filtering. It calculates the probability of each class given a set of input features and selects the class with the highest probability as the prediction. The algorithm computes these probabilities using Bayes' theorem:

$$P(y|\mathbf{x}_1,\ldots,\mathbf{x}_n) = \frac{P(\mathbf{x}_1,\ldots,\mathbf{x}_n|y)P(y)}{P(\mathbf{x}_1,\ldots,\mathbf{x}_n)}$$

3. Support Vector Machine (SVM):

Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification and regression tasks. SVM finds the optimal hyperplane that best separates classes in the feature space, maximizing the

margin between classes. For linearly separable data, SVM aims to find a hyperplane that maximizes the distance between the closest data points of different classes. For non-linearly separable data, SVM uses kernel functions to map the input space into a higher-dimensional feature space where classes are separable. The decision function of SVM for classification can be represented as:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{N_{SV}} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b\right)$$

Algorithm 1: SVM

1. Set $Input = (x_i, y_i)$, where $i = 1, 2, \cdots, N, x_i = R^n$ and $y_i = \{+1, -1\}$.
2. Assign $f(X) = \omega^T x_i + b = \sum_{i=1}^{N} \omega^T x_i + b = 0$
3. Minimize the QP problem as, $min\ \varphi(\omega, \xi) = \frac{1}{2}\|\omega\|^2 + C.(\sum_{i=1}^{N} \xi_i)$.
4. Calculate the dual Lagrangian multipliers as $min\ L_P = \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{N} x_i y_i(\omega x_i + b) + \sum_{i=1}^{N} x_i$.
5. Calculate the dual quadratic optimization (QP) problem as $max\ L_D = \sum_{i=1}^{N} x_i - \frac{1}{2}\sum_{i,j=1}^{N} x_i x_j\ y_i\ y_j\ (x_i, x_j)$.
6. Solve dual optimization problem as $\sum_{i=1}^{N} y_i x_i = 0$.
7. Output the classifier as $f(X) = sgn(\sum_{i=1}^{N} x_i y_i (x \cdot x_i) +$.

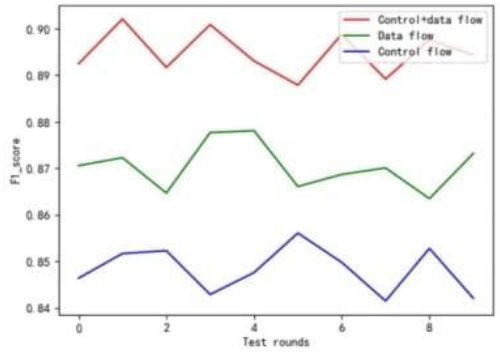
4. Decision Tree:

Decision Tree is a non-parametric supervised learning method used for both classification and regression tasks. It partitions the data into subsets based on features that best split the dataset, aiming to minimize impurity or maximize information gain at each node. Each node represents a decision point based on a feature, and each leaf node represents a class label (in classification) or a numerical value (in regression). Decision Trees are intuitive, easy to interpret, and capable of handling both numerical and categorical data. However, they are prone to overfitting noisy data and can create complex trees that generalize poorly to unseen data.

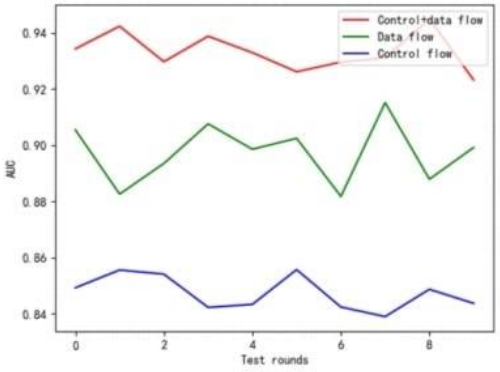
V.RESULT

FRAUD DETECTION MODEL RESULTS BASED ON SVM

Perspectives	*Precision*	*Recall*	*F1-score*	*AUC*
Control+data flow	0.946	0.852	0.895	0.935
Data flow	0.912	0.837	0.871	0.892
Control flow	0.889	0.812	0.849	0.842



(a) F1-score statistics



(b) AUC statistics

S.NO	Algorithm	Accuracy
1	Logistic Regression	52.10
2	Naive Bayes	53.63
3	SVM	51.91
4	Decision Tree	53.25

VI MODULES

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse and Train & Test Data Sets, View Trained and Tested Accuracy

in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Fraud Status in Ecommerce Transaction, View Fraud Detection Status Ratio in Ecommerce Transaction, Download Trained Data Sets, View Ecommerce Transaction Fraud Status Ratio Results, View All Remote Users

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT FRAUD DETECTION TYPE IN ECOMMERCE TRANSACTION, VIEW YOUR PROFILE.

VII.CONCLUSION

This paper proposed a hybrid method to capture fraud transactions by integrating the formal process modeling and the dynamic user behaviors. We analyzed the e-commerce transaction process under five major perspectives: control flow perspective, resource perspective, time perspective, data perspective, and user behavior patterns. This paper utilized high-level Petri nets as the basis of process modeling to model the abnormal user behaviors and created an SVM model to perform fraudulent transaction detection. Our extensive experiments showed that the proposed method can effectively capture fraudulent transactions and behaviors. The overall index of our proposed multi-perspective detection method outperformed the single-perspective detection method. As our future work, related deep learning [38-42] and model checking methods [43-45] would be incorporated in the proposed framework for higher accuracy. Additionally, it's also a future work to incorporate more time features to the behavior patterns so as to make the risk identification more accurate. Furthermore, we will conduct research on constructing a standard fraud mode library, and apply the proposed methodology to other malicious behavior areas by coordinating the models.

VIII.REFERENCES

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

[2] M. Abdelrhim, and A. Elsayed, "The Effect of COVID-19 Spread on the e-commerce market: The case of the 5 largest e-commerce companies in the world." *Available at SSRN 3621166*, 2020, doi: 10.2139/ssrn.3621166.

[3] P. Rao et al., "The e-commerce supply chain and environmental sustainability: An empirical investigation on the online retail sector." *Cogent. Bus. Manag.*, vol. 8, no. 1, pp. 1938377, 2021.

[4] S. D. Dhobe, K. K. Tighare, and S. S. Dake, "A review on prevention of fraud in electronic payment gateway using secret code," *Int. J. Res. Eng. Sci. Manag.*, vol. 3, no. 1, pp. 602-606, Jun. 2020.

[5] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *J. Netw. Comput. Appl.*, vol. 68, pp. 90-113, Apr. 2016.

[6] E. A. Minastireanu, and G. Mesnita, "An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection," *Info. Econ.*, vol. 23, no. 1, 2019.

[7] X. Niu, L. Wang, and X. Yang, "A comparison study of credit card fraud detection: Supervised versus unsupervised," *arXiv preprint arXiv*: vol. 1904, no. 10604, 2019, doi: 10.48550/arXiv.1904.10604. [8] L. Zheng et al., "Transaction Fraud Detection Based on Total Order Relation and Behavior Diversity," *IEEE Trans. Computat. Social Syst.*, vol. 5, no. 3, pp. 796-806, 2018.

[9] Z. Li, G. Liu, and C. Jiang, "Deep Representation Learning With Full Center Loss for Credit Card Fraud Detection," *IEEE Trans. Computat. Social Syst.*, vol. 7, no. 2, pp. 569-579, 2020.

[10] I. M. Mary, and M. Priyadharsini, "Online Transaction Fraud Detection System," in *2021 Int. Conf. Adv. C. Inno. Tech. Engr.*(*ICACITE*), 2021, pp. 14-16.

[11] D. Choi, and K. Lee, "Machine learning based approach to financial fraud detection process in mobile payment system," *IT Conv. P*. (*INPRA*), vol. 5, no. 4, pp. 12-24, 2017.

[12] R. Sarno et al., "Hybrid Association Rule Learning and Process Mining for Fraud Detection," *IAENG Int. J. C. Sci.*, vol. 42, no. 2, 2015.

[13] J. J. Stoop, "Process mining and fraud detection-A case study on the theoretical and practical value of using process mining for the detection of fraudulent behavior in the procurement process," M.S. thesis, Netherlands, ENS: University of Twente, 2012.

[14] M. Jans et al., "A business process mining application for internal transaction fraud

mitigation," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13351-13359, 2011.

[15] C. Rinner et al., "Process mining and conformance checking of long running processes in the context of melanoma surveillance," *Int. J. Env. Res. Pub. He*, vol. 15, no. 12, pp. 2809, 2018.

[16] E. Asare, L. Wang, and X. Fang, "Conformance Checking: Workflow of Hospitals and Workflow of Open-Source EMRs," *IEEE Access*, vol. 8, pp. 139546-139566, 2020.

[17] W. Chomyat and W. Premchaiswadi, "Process mining on medical treatment history using conformance checking," in *2016 14th Int. Conf. ICT K. Eng.* (*ICT&KE*), 2016, pp. 77-83.

[18] M. D. Leoni, W. M. Van Der Aalst, and B. F. V. Dongen, "Data-and resource-aware conformance checking of business processes," in *Int. Conf. Bus. Info. Sys.*, Springer, Berlin, Heidelberg, 2012. pp. 48-59.

[19] S. M. Najem, and S. M. Kadeem, "A survey on fraud detection techniques in ecommerce," *Tech-Knowledge*, vol. 1, no. 1, pp. 33-47, 2021.

[20] K. Böhmer, and S. Rinderle-Ma, "Anomaly detection in business process runtime behavior--challenges and limitations," *arXiv preprint arXiv*, 2017,doi: 10.48550/arXiv.1705.06659.

[21] K. D. Febriyanti, R. Sarno and Y. Effendi, "Fraud detection on event logs using fuzzy association rule learning," in *2017 11th Int. Conf. Info. Comm. Tech. Sys.*, Surabaya, Indonesia, 2017, pp. 149-154.

[22] T. Chiu, Y. Wang and M. Vasarhelyi, "A framework of applying process mining for fraud scheme detection," *SSRN Electronic Journal*, 2017, doi:10.2139/ssrn.2995286.

[23] W.Yang et al., "Show Me the Money!Finding Flawed Implementations of Third-party In-app Payment in Android Apps," in *Proc. NDSS*, Shanghai, China, 2017.

[24] W. Rui, S. Chen, X. Wang and S.Qadeer, "How to Shop for Free Online--Security Analysis of Cashier-as-a-Service Based Web Stores," in *Proc. SSP*, Oakland, CA, USA, 2011, pp. 465-480.

[25] E. Ramezani, D. Fahland and W. Aalst, "Where did I misbehave? Diagnostic information in compliance checking," in *BPM.*, Berlin，Germany, Springer, 2012, pp. 262-278.

AUTHOR:

[1] Mrs. Chepuri. Deepti, currently working as an Assistant Professor in the Department of Computer Science and Engineering, QIS

College of Engineering And Technology, Ongole, Andhra Pradesh. She did her BTech from Uttar Pradesh Technical University, Lucknow, M.Tech from JNTUK, Kakinada. Her area of interest is Machine Learning, Artificial intelligence, Cloud Computing and Programming Languages.

[2] Ms.Polisetti Ramya, currently pursuing Master of Computer Applications at QIS College of engineering and Technology (Autonomous) Ongole in Andhra Pradesh. She Completed B.S.C in Statistics from Sri Harshini Degree College Martur Andhra Pradesh. Her areas of interest are Machine learning &Pytho