www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Fraud Detection for Multi-Participant E-Commerce Transactions Using a Multi-Perspective Method

Mr.K.Prasanth Kumar[1]., Auvasala Swetha [2]
1 Assistant Professor, Department of CSE, Malla Reddy College of Engineering for Women., Maisammaguda., Medchal., TS, India
2, B.Tech CSE (20RG1A0565),
Malla Reddy College of Engineering for Women., Maisammaguda., Medchal., TS, 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.

Keywords: Support Vector Machine (SVM), B2C e-commerce,

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. 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. The sales volume of B2C (Business to Customer) e-commerce is expected to reach 6.5 trillion dollars by 2023. Though the growth of e-commerce and the expansion of modern technologies offer better opportunities for online businesses, new security threats have emerged over the past few years. Reportedly, the significant increase in the number

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

of online fraud cases costs billions of dollars worldwide every year. The dynamic and distributed nature of the Internet has made anti-fraud systems inevitable to ensure the security of online transactions. Existing fraud detection systems focusing on detecting abnormal user behaviors still characterize vulnerabilities when mitigating emerging security threats. An important issue in existing fraud detection systems is their lack of efficient process management during the trading process. The imperfect monitoring function is one of the key issues that need attention. The detection perspective is usually not enough due to the lack of process capture for the existing work. To this end, we propose a process-based method, where user behaviors are recorded and analyzed in real-time, and historical data is transformed into controllable data. In addition, we incorporate a multi-perspective detection of abnormal behaviors.

In the rapidly evolving realm of Ecommerce, transactions involving multiple participants present unique challenges in detecting and preventing fraud. By integrating sophisticated techniques such as user behaviour analysis, anamoly detection, and machine learning, our approach aims to provide a robust solution to enhance transactions involve a dynamic interplay among multiple participants such as buyers, sellers and intermediaries.

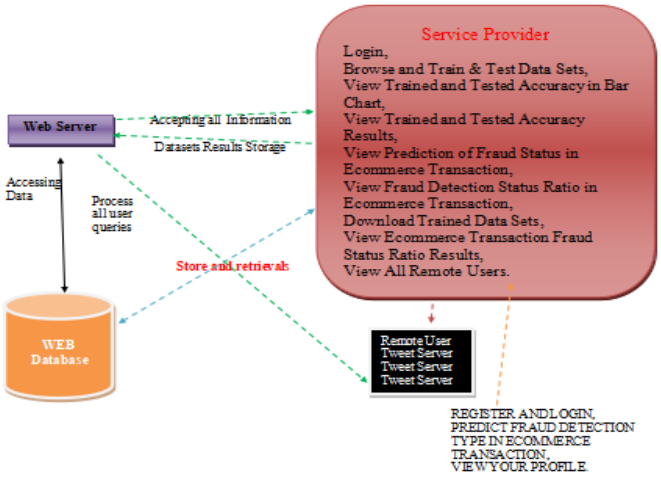


FIG ,1: SYSTEM ARCHITECTURE

II. RELATED WORK

An Analysis of the Most Used Machine Learning Algorithms for Online Fraud detection

Author: E. A. Ministering, and G. Manita

Description: The escalating complexity and transnational nature of illegal activities in online financial transactions have led to substantial financial losses for both customers and organizations. Countering this challenge, numerous techniques have been proposed for fraud prevention and detection in the online environment. However, each of these techniques exhibits distinct characteristics, advantages, and drawbacks, making it imperative to comprehensively review and analyse the existing research in fraud detection. This paper employs a systematic quantitative literature review

methodology to identify the algorithms used in fraud detection and analyses each algorithm based on specific criteria. Wang yang Yu; Yadi Wang; Lu Liu; Yusheng An; Bo Yuan; JohnPanneerselvam, A Multi perspective Fraud Detection Method for Multiparticipant E-Commerce Transactions,2021 In the persistent challenge of detecting and preventing fraudulent transactions within e-commerce platforms, traditional security systems relying on historical order information often fall short, given the elusive nature of online activities. Recognizing the limitations of existing approaches that neglect dynamic user behaviours, this article proposes an innovative fraud detection method that seamlessly integrates machine learning and process mining models for real-time monitoring. The methodology unfolds in three key stages. First, a business-to-customer (B2C) e-commerce platform is modelled, incorporating a robust framework for detecting user behaviours. This foundational process aims to better understand and adapt to the dynamic nature of user interactions within the platform. Second, the article introduces a method for analysing abnormalities, leveraging event logs to extract essential features crucial for fraud detection. This step ensures a nuanced understanding of irregular patterns indicative of potentially fraudulent activities.

A Survey on Fraud Detection Techniques in Ecommerce

Author: khyati chaudary, B. Mallick

Description: This paper shows the approaches used in fraud detection in e-commerce and suggests ways to design potent and efficient fraud detection algorithms for reducing the losses in transaction. Commerce becomes popular for online shopping, banking, financial institution and government. Fraudulent activity exits in many areas of businesses and our daily life. Such activities are most prevalent in telecommunication, credit card fraud detection, network intrusion, finance and insurance and scientific applications, Billions of dollars are loss every day due to the increase in the number of credit card transactions by using online as well as offline. To design potent and efficient fraud detection algorithms is the key for reducing the losses in transaction. There were numerous approaches have been implemented for the fraud detection. This paper shows the approaches used in fraud detection in e-commerce.

Machine Learning-Powered Fraud Detection & Prevention: A Comprehensive Implementation

Author: B. Kumar, Shivam KumarGuota, Manorama Patnaik

Description: This study focuses on the training of a comprehensive fraud detection as well as prevention system employing leading-edge

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

machine learning ways, with a particular emphasis on Python as the primary programming language.Fraud detection as well as prevention represent paramount challenges in today's digital scenario, where the threat landscape continually evolves. This research study focuses on the training of a comprehensive fraud detection as well as prevention system employing leading-edge machine learning ways, with a particular emphasis on Python as the primary programming language. The scope of this research study encompasses a multitude of domains, ranging from financial institutions combatting credit card fraud to e-commerce platforms shielding against payment fraud, and even healthcare providers preventing insurance fraud. The proposed approach encompasses data collection, preprocessing, as well as feature engineering, followed by model selection, training, as well as evaluation. Through threshold adaptations and real- time integration, the system seeks to triumph a harmonious balance among precision as well as recall, thereby minimizing false positives and negatives. Legal and ethical considerations, as well as user education, are integral aspects of the project. This attempt not only serves as a robust defence against evolving fraud tactics but also highlights the versatility of machine learning in addressing complex challenges across various industries. With fraudsters perpetually devising new methods, this project signifies the continuous commitment to innovative and adaptable fraud detection and prevention systems.

III. SYSTEM ANALYSIS

System analysis is the process of examining and understanding complex systems in order to identify their components, interactions, and behaviors. It involves studying a system's structure, function, and objectives to determine how it operates and how it can be improved.

During system analysis, the analyst collects and analyzes information about the system's requirements, constraints, and goals. This typically involves conducting interviews, reviewing documents, and observing the system in action. The purpose is to gain a comprehensive understanding of the system's current state, including its strengths, weaknesses, and areas for improvement.

The next step in system analysis is to identify potential solutions or changes that can address the system's issues or enhance its performance. This may involve proposing modifications to the system's design, processes, or technologies

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. Xuetong Niu et al. conducted a comparative study on credit card fraud detection

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

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

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 has proved that the machine learning method can efficiently capture fraudulent transactions in credit card applications. Despite existing progress in fraud detection, it is still necessary to develop hybrid learning methods to improve the accuracy of detection. 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.

EXISTING SYSTEM DISADVANTAGES:

Complexity: Implementing a multi-perspective approach can be complex and require advanced technological infrastructure.

Resource Intensive: It may require significant resources in terms of data processing and analysis.

Integration Challenges: Integrating multiple viewpoints and data sources can be challenging and may require seamless coordination.

Potential False Positives: The method might generate false positives due to the complexity of analyzing multiple perspectives simultaneously.

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.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

PROPOSED SYSTEM 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. IMPLEMENTATION

1.Decision tree classifiers

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

2.Gradient boosting

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest. A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

3.K-Nearest Neighbors (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning

- Does not "learn" until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

4.Logistic regression Classifiers

Logistic Regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name logistic regression is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name multinomial logistic regression is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables.

It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

5.Naive Bayes

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .

Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).

While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique.

6.Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random . Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.

7.SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an independent and identically distributed (iid) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point x and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

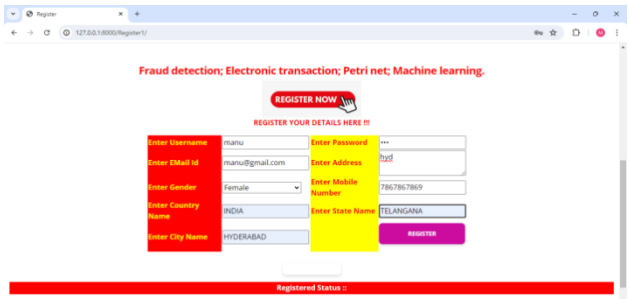
SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to genetic algorithms (GAs) or perceptrons, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

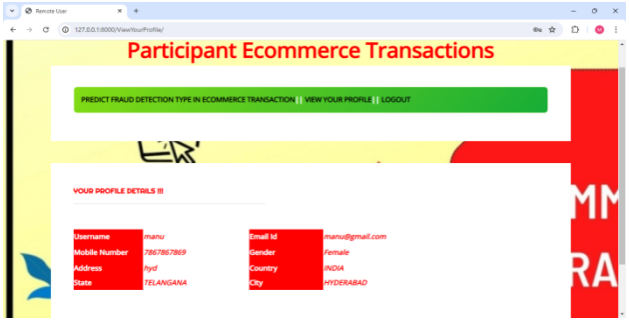
uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

V. RESULTS AND DISCUSSION

Fig,1: Home Page



Fig, 2: User Register Page



Fig, 3: Profile Details

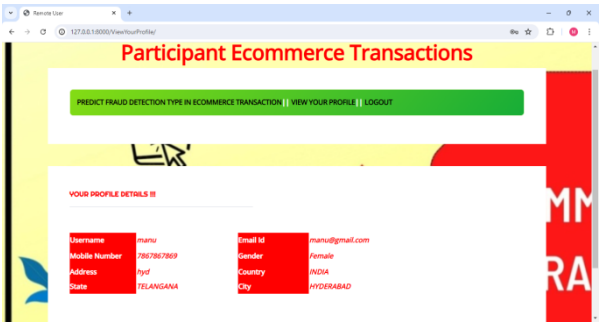
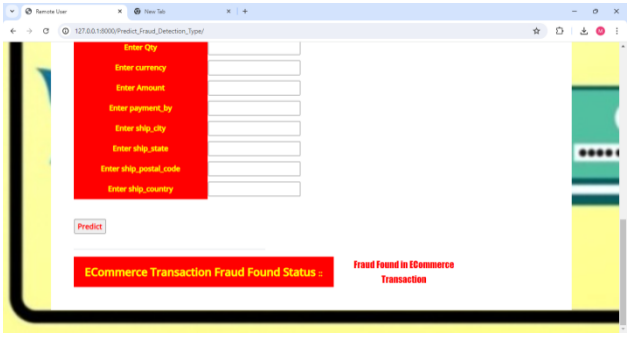


Fig 4: Checking for Fraud



5:

Fig 5: Status of Fraud

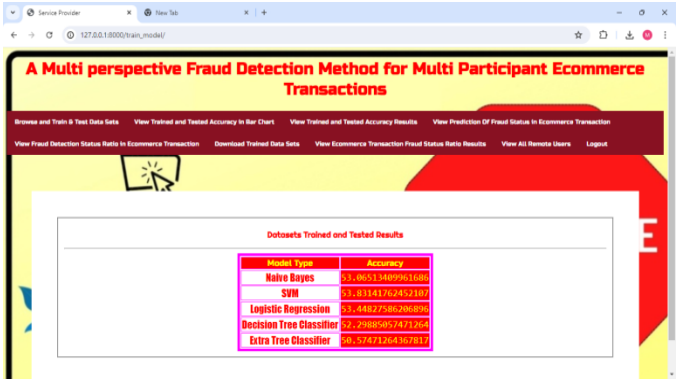


Fig 6:Browse and Test Datasets

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

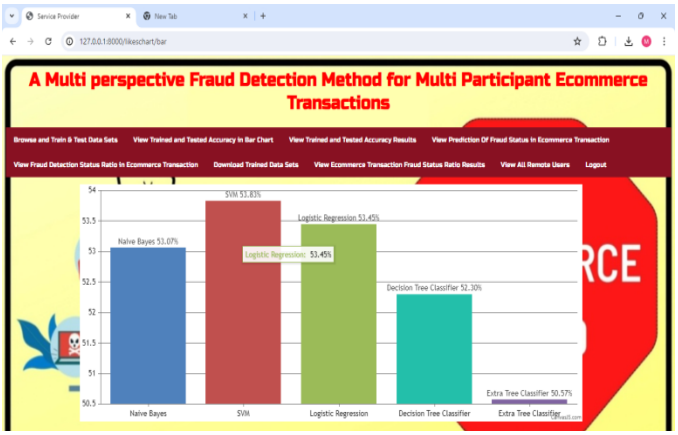
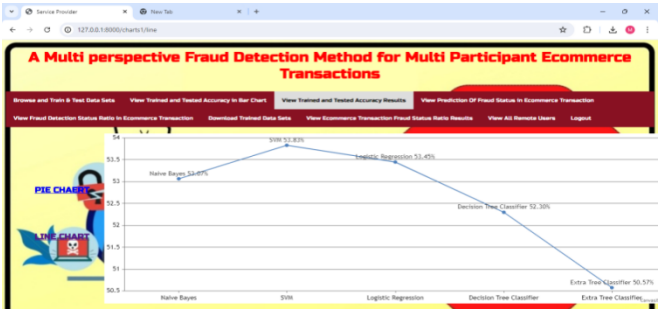
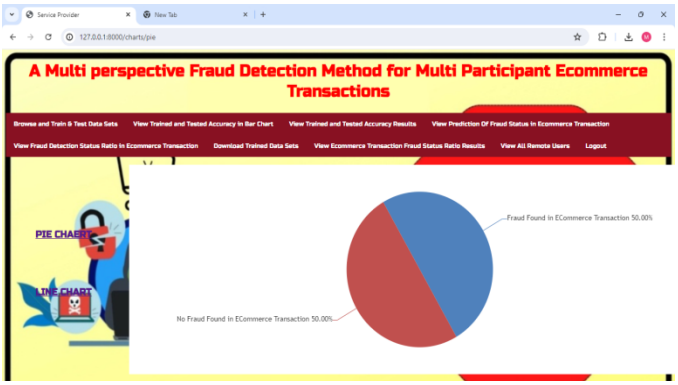


Fig 7: View Tested Datasets Accuracy in Bar chart



Fig,8:View trained and test datasets Accuracy results



Fig,9,View prediction of Fraud status Ratio

VI. 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.

FUTURE ENHANCEMENTS:

The multi-perspective fraud detection system can be further enhanced by integrating real-time processing, advanced deep learning architectures, and explainability techniques. Additionally, incorporating active learning, graph-based analysis, and multimodal data

fusion will improve detection accuracy. Continuous updating, human-in-the-loop review, and cloud-based deployment will ensure the system remains effective and scalable. Finally, integration with other e-commerce systems will streamline fraud detection and prevention, providing a comprehensive security solution. Future enhancements to the fraud detection system will focus on integrating emerging technologies such as artificial intelligence, machine learning, and blockchain to improve accuracy and efficiency. Advanced analytics and predictive modeling will enable real-time detection and prevention of fraudulent activities. Additionally, the system will be enhanced with natural language processing capabilities to detect and prevent phishing and social engineering attacks. Furthermore, the system will be integrated with Internet of Things (IoT) devices to detect and prevent fraud in real-time transactions. These enhancements will ensure the system remains ahead of emerging fraud threats and provides robust security for e-commerce transactions.

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.