International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Survey Paper on Prediction of Fraud Detection in E-Commerce using Supervised Machine Learning

Varun Rajput
M. Tech. Scholar, Department of Computer Science and Engineering, SORT, People's University, Bhopal, India
Shekhar Nigam
Professor & HOD, Department of Computer Science and Engineering, SORT, People's University, Bhopal, India

Abstract
The rapid growth of e-commerce platforms such as Amazon and Flipkart has significantly increased the risk of fraudulent activities, leading to financial losses and reduced customer trust. This survey paper presents a comprehensive review of fraud detection techniques in e-commerce using supervised machine learning approaches. The study focuses on commonly used classification algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Networks, which are trained on labeled datasets to distinguish between legitimate and fraudulent transactions.

The paper analyzes various features used in fraud detection, including transaction amount, user behavior, purchase history, IP address, and device information. It also discusses challenges such as class imbalance, data privacy, and evolving fraud patterns. Different performance evaluation metrics, including accuracy, precision, recall, and F1-score, are compared to assess model effectiveness.

Furthermore, this survey highlights the importance of feature engineering, data preprocessing, and real-time detection systems to improve prediction accuracy. Comparative analysis from recent research studies indicates that ensemble methods, particularly Random Forest and Gradient Boosting, often achieve higher detection rates with lower false positives.

The findings of this paper provide valuable insights into selecting appropriate supervised learning techniques for fraud detection in e-commerce systems. It also suggests future research directions, including hybrid models and deep learning integration, to enhance security and reliability in online transactions.

Keywords: -Machine Learning (ML), E-Commerce, Fraud Detection

I. INTRODUCTION

The rapid advancement of internet technologies and digital payment systems has transformed the way businesses operate, giving rise to the widespread adoption of electronic commerce (e-commerce). Platforms such as Amazon and Flipkart have revolutionized retail by enabling consumers to purchase goods and services anytime and anywhere. This convenience, however, has also introduced significant security challenges, particularly in the form of fraudulent activities. E-commerce fraud includes unauthorized transactions, identity theft, account takeovers, payment fraud, and fake returns, all of which pose serious threats to both businesses and customers. As online transactions continue to increase, the need for robust and intelligent fraud detection systems has become more critical than ever.

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Traditional fraud detection methods, such as rule-based systems and manual monitoring, are no longer sufficient to handle the scale and complexity of modern e-commerce environments. These methods rely heavily on predefined rules and human expertise, making them less effective in detecting new and evolving fraud patterns. Fraudsters continuously adapt their techniques, exploiting system vulnerabilities and bypassing static detection mechanisms. As a result, there is a growing demand for advanced data-driven approaches capable of identifying suspicious activities in real time [1, 2].

In this context, supervised machine learning has emerged as a powerful tool for fraud detection in e-commerce. Supervised learning algorithms are trained on labeled datasets containing both legitimate and fraudulent transactions, enabling them to learn patterns and make accurate predictions on unseen data. Commonly used algorithms include Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and Artificial Neural Networks. These models can analyze large volumes of transactional data and identify complex relationships between various features, such as transaction amount, frequency, geographical location, device information, and user behavior [3].

One of the key advantages of supervised machine learning is its ability to continuously improve performance through training and optimization. With the availability of historical transaction data, models can be fine-tuned to achieve higher accuracy and better generalization. Additionally, ensemble learning techniques, such as Random Forest and Gradient Boosting, have shown significant improvements in fraud detection performance by combining multiple models to reduce errors and increase robustness. These approaches are particularly useful in handling the highly imbalanced nature of fraud datasets, where fraudulent transactions represent only a small fraction of the total data [4, 5].

Despite its advantages, the application of supervised machine learning in fraud detection faces several challenges. One major issue is class imbalance, which can lead to biased models that favor the majority class (legitimate transactions) while failing to detect fraudulent ones. Another challenge is the dynamic nature of fraud, where patterns change over time, requiring models to be regularly updated and retrained. Data privacy and security concerns also limit access to high-quality datasets, which are essential for training effective models. Furthermore, achieving a balance between detection accuracy and false positives is crucial, as excessive false alarms can negatively impact customer experience [6].

This survey paper aims to provide a comprehensive overview of supervised machine learning techniques used for fraud detection in e-commerce. It reviews existing literature, compares different algorithms, and analyzes their performance based on various evaluation metrics. The study also explores feature selection methods, data preprocessing techniques, and emerging trends in fraud detection. By highlighting current challenges and potential solutions, this paper contributes to the development of more efficient and reliable fraud detection systems, ultimately enhancing the security and trustworthiness of e-commerce platforms [7, 8].

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

II. E-COMMERCE FRAUD

E-commerce fraud refers to any illegal or deceptive activity conducted during online transactions with the intent of gaining financial or personal benefits. With the rapid expansion of digital marketplaces such as Amazon and Flipkart, fraud has become a major concern affecting businesses, financial institutions, and consumers alike. The anonymity, speed, and global reach of online platforms make them attractive targets for cybercriminals.

E-commerce fraud can take various forms, depending on the methods used by attackers. One of the most common types is payment fraud, where stolen credit or debit card information is used to make unauthorized purchases. Another prevalent form is identity theft, in which fraudsters use personal information to create fake accounts or gain access to legitimate user accounts. Account takeover (ATO) attacks involve unauthorized access to a user's account, allowing criminals to make purchases or steal sensitive data. Additionally, refund fraud and chargeback fraud occur when customers falsely claim refunds or dispute legitimate transactions to obtain money dishonestly.

Fraudsters often exploit weaknesses in authentication systems, payment gateways, and user verification processes. Techniques such as phishing, malware attacks, and social engineering are commonly used to collect sensitive information. Furthermore, the use of proxy servers, VPNs, and botnets enables attackers to mask their identities and bypass traditional security measures. This makes detecting fraudulent behavior increasingly complex and challenging.

One of the key characteristics of e-commerce fraud is its dynamic and evolving nature. Fraud patterns continuously change as attackers develop new strategies to evade detection systems. This creates significant challenges for traditional rule-based detection methods, which rely on static rules and predefined patterns. As a result, many fraudulent transactions go undetected or are identified too late, leading to financial losses and reputational damage.

To address these challenges, advanced techniques such as supervised machine learning have been widely adopted. These approaches analyze historical transaction data to identify hidden patterns and anomalies associated with fraudulent activities. Features such as transaction frequency, purchase amount, IP address, device type, and user behavior are used to train models that can accurately classify transactions as legitimate or fraudulent.

In summary, e-commerce fraud represents a critical issue in the digital economy, requiring continuous monitoring and the implementation of intelligent detection systems. Understanding the types, techniques, and challenges associated with fraud is essential for developing effective prevention strategies and ensuring secure online transactions.

Fig. 1: E-Commerce Fraud

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

III. LITERATURE REVIEW

Srinivas et al. (2025) presented a comprehensive study on fraud detection and prevention in e-commerce using machine learning approaches to enhance transaction security. The authors discussed the growing complexity of online fraud and emphasized the limitations of traditional rule-based systems in handling evolving fraud patterns. Various supervised machine learning algorithms were analyzed for their effectiveness in identifying fraudulent transactions. The study highlighted the importance of feature engineering, data preprocessing, and model evaluation to achieve improved accuracy and reduced false positives. The authors concluded that machine learning-based approaches significantly enhance fraud detection performance and provide a scalable solution for modern e-commerce platforms.

Mutemi and Bacao (2025) conducted a systematic literature review focusing on machine learning techniques applied to e-commerce fraud detection. The study reviewed a wide range of research articles published over recent years and categorized them based on algorithms, datasets, evaluation metrics, and application domains. The authors identified supervised, unsupervised, and hybrid machine learning models as dominant approaches, with ensemble methods showing superior performance. The review emphasized challenges such as data imbalance, concept drift, and real-time processing requirements. The authors suggested that future fraud detection systems should focus on scalable architectures and adaptive learning mechanisms.

Li et al. (2025) proposed an unsupervised fraud detection framework using contrastive learning for identifying fraudulent transactions in e-commerce environments. The study addressed the challenge of limited labeled data by leveraging representation learning techniques to distinguish between normal and abnormal transaction behaviors. The proposed approach demonstrated strong performance in detecting previously unseen fraud patterns, making it suitable for dynamic e-commerce scenarios. The authors highlighted that unsupervised methods can complement supervised learning to improve robustness and adaptability in fraud detection systems.

Zeng et al. (2025) introduced NNEnsLeG, a novel ensemble learning framework that combines neural networks with ensemble strategies for e-commerce payment fraud detection. The proposed model integrated multiple neural network classifiers to capture diverse fraud patterns and improve generalization. Experimental results showed that the ensemble-based approach outperformed individual models in terms of accuracy, precision, and recall. The study emphasized that ensemble learning enhances detection reliability and reduces the risk of model bias in large-scale e-commerce applications.

Islam et al. (2025) explored fraud detection in financial networks using a layer-weighted Graph Convolutional Network (GCN). The study modeled transaction data as a graph to capture complex relationships between users, merchants, and transactions. By assigning

adaptive weights to different network layers, the proposed method effectively detected various fraud patterns. The authors demonstrated that graph-based learning approaches significantly outperform traditional feature-based models, especially in identifying organized and coordinated fraud activities within financial ecosystems.

Sha et al. (2025) proposed a heterogeneous graph neural network with graph attention mechanisms for detecting credit card fraud. The study incorporated multiple types of nodes and edges to represent real-world transaction relationships more accurately. The attention mechanism enabled the model to focus on critical features and interactions contributing to fraudulent behavior. Experimental results indicated improved detection accuracy and robustness, highlighting the potential of graph-based deep learning models in complex fraud detection scenarios.

Luo, Wang, and Zhu (2025) introduced an advanced fraud detection and risk assessment framework that integrates large language models (LLMs) with graph convolutional networks for e-commerce payment platforms. The framework leveraged LLMs for contextual understanding of transaction metadata and GCNs for relational pattern analysis. The authors demonstrated that combining semantic and structural learning significantly improves fraud detection performance. The study highlighted the future potential of hybrid AI architectures for intelligent and explainable fraud detection systems.

Lakkaraju (2025) investigated the application of machine learning techniques to combat e-commerce fraud, focusing on real-world implementation challenges. The study analyzed supervised learning models such as decision trees, random forests, and gradient boosting for transaction fraud detection. The author emphasized the importance of continuous model updating to handle concept drift and evolving fraud strategies. The findings suggested that machine learning-driven fraud detection systems can significantly reduce financial losses and enhance customer trust when properly deployed.

IV. MACHINE LEARNING ALGORITHM

Machine Learning is a subset of Artificial Intelligence concerned with “teaching” computers how to act without being explicitly programmed for every possible scenario. The central concept in Machine Learning is developing algorithms that can self-learn by training on a massive number of inputs. Machine learning algorithms are used in various applications, such as email filtering and computer vision, where it is difficult or infeasible to develop conventional algorithms to perform the needed tasks [4]. Machine learning enables the analysis of vast amounts of information. While it usually delivers faster, more precise results to identify profitable prospects or dangerous risks, it may also require additional time and assets to train it appropriately. Merging machine learning with AI and perceptive technologies can make it even more effective in processing vast volumes of information. Machine learning is closely associated with computational statistics, which focuses on making predictions using computers. Machine learning approaches are conventionally divided into three broad categories, namely Supervised Learning, Unsupervised Learning &

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Semi-supervised Learning, depending on the nature of the "signal" or "feedback" available to the learning system.

Face anti-spoofing (FAS) has lately attracted increasing attention due to its vital role in securing face recognition systems from presentation attacks (PAs). As more and more realistic PAs with novel types spring up, traditional FAS methods based on handcrafted features become unreliable due to their limited representation capacity. With the emergence of large-scale academic datasets in the recent decade, machinelearning based FAS achieve remarkable performance and dominate this area.

Supervised Learning

A model is trained through a process of learning in which predictions must be made and corrected if those predictions are wrong. The training process continues until a desired degree of accuracy is reached on the training data. Input data is called training data and has a known spam / not-spam label or result at one time.

Unsupervised Learning

By deducting the structures present in the input data, a model is prepared. This may be for general rules to be extracted. It may be through a mathematical process that redundancy can be systematically reduced, or similar data can be organized. There is no labeling of input data, and there is no known result.

Semi-Supervised Learning

Semi-supervised learning fell between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data). There is a desired problem of prediction, but the model needs to learn the structures and make predictions to organize the data. Input data is a combination of instances that are marked and unlabeled.

V. SIMULATION PARAMETER

Accuracy gives a proportion of how precise your model is in anticipating the real up-sides out of the absolute up-sides anticipated by your framework. Review gives the quantity of real up-sides caught by our model by grouping these as obvious positive. F-measure can give a harmony among accuracy and review, and it is liked over precision where information is uneven. The findings demonstrate that supervised machine learning methods can effectively enhance the security and reliability of e-commerce platforms by enabling early detection and prevention of fraudulent activities.

Accordingly, F-measure was used in this review as a presentation metric to give a decent and fair measure utilizing the equation.

$$\text{Precision} = \text{TP TP} + \text{FP} \times 100$$

$$\text{Recall} = \text{TP TP} + \text{FN} \times 100$$

$$\text{F} - \text{measure} = 2 * \text{Precision} * \text{Recall Precision} + \text{Recall} \times 100$$

Where,

TP—True Positive, FP—False Positive, FN—False Negative

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

VI. CONCLUSION

E-commerce fraud has emerged as a significant challenge in the digital era due to the rapid expansion of online platforms such as Amazon and Flipkart. The increasing volume of online transactions has created more opportunities for fraudulent activities, making traditional rule-based detection systems insufficient. This survey highlights the importance of adopting advanced techniques, particularly supervised machine learning, to effectively identify and prevent fraud in e-commerce environments.

Various supervised learning algorithms, including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and Neural Networks, have demonstrated strong capabilities in detecting fraudulent transactions by analyzing historical data patterns. Among these, ensemble methods such as Random Forest and Gradient Boosting have shown superior performance in handling complex and imbalanced datasets. The study also emphasizes the role of feature engineering, data preprocessing, and proper evaluation metrics in improving model accuracy and reliability.

However, challenges such as class imbalance, evolving fraud patterns, and data privacy concerns still persist. Continuous model updating, integration of real-time detection systems, and the use of hybrid approaches combining machine learning and deep learning can further enhance fraud detection efficiency.

In conclusion, supervised machine learning provides a robust and scalable solution for fraud detection in e-commerce. Future research should focus on developing adaptive, real-time, and more accurate models to ensure secure online transactions and maintain user trust in digital commerce systems.

REFERENCES

[1] M. Srinivas, M. Kilaru, D. Jain, and K. S. Sidhu, “Fraud Detection and Prevention in E-Commerce: Machine Learning Approaches to Secure Transactions,” in *Proc. 2025 First Int. Conf. Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)*, 2025, pp. 1416–1420, doi:10.1109/CE2CT64011.2025.10939681.

[2] A. Mutemi and F. Bacao, “E-Commerce Fraud Detection Based on Machine Learning Techniques: Systematic Literature Review,” *IEEE Xplore*, 2025.

[3] X. Li et al., “Unsupervised Detection of Fraudulent Transactions in E-commerce Using Contrastive Learning,” *arXiv preprint*, Mar. 2025.

[4] Q. Zeng et al., “NNEnsLeG: A Novel Approach for E-Commerce Payment Fraud Detection Using Ensemble Learning and Neural Networks,” *Information Processing & Management*, vol. 62, 2025.

[5] S. Islam, G. Raj Gupta, A. Chakraborty et al., “Detecting Fraudulent Transactions for Different Patterns in Financial Networks Using Layer Weighted GCN,” *Human-Centric Intelligent Systems*, vol. 5, pp. 181–195, 2025.

[6] X. Sha et al., “Detecting Credit Card Fraud via Heterogeneous Graph Neural Networks with Graph Attention,” *arXiv preprint*, Apr. 2025.

[7] R. Luo, N. Wang and X. Zhu, “Fraud Detection and Risk Assessment of Online Payment Transactions on E-Commerce Platforms Based on LLM and GCN Frameworks,” *arXiv preprint*, Sep. 2025.

[8] S. Lakkaraju, "Using Machine Learning to Combat E-Commerce Fraud," *Intl. Journal of Information Technology and Management Information Systems*, vol. 16, no. 1, pp. 844–859, Jan.–Feb. 2025.
[9] Bhanu Pratap Singh, Shekhar Nigam, Detection of SQL Injection Attack using Machine Learning Techniques: A Review, *International Journal of Innovative Research in Technology (IJIRT),* ISSN: 2349-6002, Volume:11, Issue 11, April: 2025
[10] Pinky Mishra, Shekhar Nigam, Android Malware Detection using Convolutional Neural Networks and Genetic Algorithm: A Review, *Journal of Emerging Technologies and Innovative Research (JETIR),* ISSN: 2349-5162 Volume:12, Issue 4, April: 2025
[11] A. Mutemi and F. Bacao, "E-Commerce Fraud Detection Based on Machine Learning Techniques: Systematic Literature Review," *Big Data Mining and Analytics*, vol. 7, no. 2, pp. 419–444, Jun. 2024.
[12] A. S. Yussiff et al., "The Best Machine Learning Model for Fraud Detection on E-Platforms: A Systematic Literature Review," *Computer Science and Information Technologies*, vol. 5, no. 2, pp. 195–204, Jul. 2024.
[13] P. Jeyachandran et al., "Leveraging Machine Learning for Real-Time Fraud Detection in Digital Payments," *Integrated Journal for Research in Arts and Humanities*, vol. 4, no. 6, pp. 70–94, Nov. 2024.
[14] M. I. Ismail and M. A. Haq, "Enhancing Enterprise Financial Fraud Detection Using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 14854–14861, Aug. 2024.
[15] S. Hashemi, S. L. Mirtaheri, and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," *IEEE Access*, vol. 11, pp. 3034–3043, 2023.
[16] N. Verma, K. Uboveja, and M. K. Singh, "Machine Learning Based Fraud Detection for E-Commerce," *Intl. Journal of Futuristic Innovation in Engineering, Science and Technology*, vol. 2, no. 1, 2023.
[17] 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.
[18] H. Zhou, G. Sun, S. Fu, W. Jiang and J. Xue, "A Scalable Approach for Fraud Detection in Online E-Commerce Transactions With Big Data Analytics," *Computers*, 2020.
[19] R. Banerjee, G. Bourla, S. Chen, M. Kashyap, and S. Purohit, Comparative analysis of machine learning algorithms through credit card fraud detection, in *Proc. IEEE MIT Undergraduate Research Technology Conf.*, Cambridge, MA, USA, 2018, pp. 1–4.
[20] A. O. Adewumi and A. A. Akinyelu, A survey of machine-learning and nature-inspired based credit card fraud detection techniques, *International Journal of System Assurance Engineering and Management*, vol. 8, no. 2, pp. 937–953, 2017.