Anomaly Recognition in Digital Commerce for Unethical Activity Prevention Using Advanced Learning Models

Siri MP[1], Prof. Vijayakumara Y M[2]

[1]Department of MCA, Akash Institute of Engineering and Technology, Devanahalli, Bangalore, Karnataka, India.
[2]Assistant Professor, Data science in CSE, Akash Institute of Engineering and Technology, Devanahalli, Bangalore, Karnataka, India

ABSTRACT - This research introduces a system titled "Adaptive Illicit Transaction Identification in Online Marketplaces via Computational Intelligence", which focuses on the detection of fraudulent activities within e-commerce platforms. The system architecture is developed using Python for backend processing and employs HTML, CSS, and JavaScript on the frontend, integrated seamlessly through the Flask web framework to ensure a dynamic and user-friendly interface. To achieve high detection accuracy, the system incorporates two sophisticated machine learning models: a Stacking Classifier and an XGBoost (XGB) Classifier. The Stacking Classifier recorded a perfect training accuracy of 100% and a test accuracy of 99%, while the XGB Classifier attained 96% training and 95% test accuracy. These outcomes highlight the models' robustness in effectively distinguishing between legitimate and fraudulent transactions. A synthetic dataset of 23,634 records was generated using the Faker library, enhanced with custom logic to realistically simulate transaction behavior and illicit patterns. The dataset features 16 key attributes—including transaction amount, payment method, and fraud indicators that collectively support the accurate modeling of transactional anomalies. The results confirm the potential of computational intelligence and ensemble learning approaches in strengthening fraud detection mechanisms, thereby enhancing transactional security and user trust in online marketplaces.

Index Terms— Illicit Transaction Detection, E-Commerce Fraud, Machine Learning, Stacking Classifier, XGBoost, Ensemble Learning, Computational Intelligence, Synthetic Data Generation, Online Marketplace Security.

I. Introduction

The rise of E-Commerce has revolutionized the global marketplace, enabling consumers and businesses to engage in seamless digital transactions across geographic and economic boundaries. From retail purchases to subscription services, millions of online transactions occur daily, supported by the convenience, speed, and accessibility offered by digital platforms. However, this rapid expansion of online commerce has also led to an alarming surge in fraudulent activities, posing a serious threat to both financial security and consumer trust.

Illicit transactions in online marketplaces have become increasingly sophisticated, leveraging advanced techniques such as identity spoofing, account takeovers, synthetic fraud, and automated bots. Traditional rule-based fraud detection mechanisms, while once effective, have grown increasingly inadequate in addressing these evolving threats. These static systems typically depend on hard-coded patterns or predefined thresholds and lack the flexibility to adapt to new fraud behaviors that do not conform to historical trends. As fraudsters continuously refine their methods, the need for dynamic, intelligent, and adaptive detection systems has become more urgent than ever.

In response to these challenges, the integration of computational intelligence—particularly machine learning (ML) and data-driven approaches—has emerged as a powerful solution in the domain of fraud detection. Machine learning algorithms possess the capability to uncover hidden patterns, detect anomalies, and make predictive decisions in real time based on large and complex datasets. Among the various techniques, ensemble methods such as Stacking and Extreme Gradient Boosting (XGBoost) have shown exceptional performance in classification tasks by combining the predictive strengths of multiple base learners.

This research presents an intelligent fraud detection framework that applies these ensemble learning techniques to identify illicit transactions in online marketplaces with high accuracy and efficiency. The system is built using Python and employs the Flask framework to facilitate integration between the backend processing and a responsive frontend interface developed using HTML, CSS, and JavaScript. The design is not only scalable and adaptable to different e-commerce environments but also optimized for real-time detection capabilities. To train and evaluate the proposed models, a synthetic dataset comprising 23,634 records was generated using Python's Faker library. This dataset was carefully constructed with custom logic to simulate realistic transaction patterns and a variety of fraudulent scenarios. It includes 16 well-defined features such as transaction amount, customer ID, payment method, and binary fraud indicators each contributing critical information to support accurate classification.

Initial experiments with the system demonstrated strong predictive performance. The Stacking Classifier achieved a training accuracy of 100% and a test accuracy of 99%, while the XGBoost Classifier yielded 96% training and 95% test accuracy. These results underscore the potential of the system to detect illicit activities reliably and efficiently, while maintaining scalability and user accessibility. In an era where e-commerce fraud continues to evolve in complexity and scale, there is an urgent demand for adaptive systems that not only react to known threats but proactively learn from emerging patterns. This paper explores the implementation and evaluation of such a system, contributing to the growing body of knowledge in the intersection of machine learning and e-commerce security. Through the development of this platform, we aim to enhance digital trust, reduce financial losses, and support the secure growth of online marketplaces.

LITERATURE SURVEY

S. Monteith, M. Bauer highlight that since the onset of the COVID-19 pandemic, the global increase in online activities ranging from work and education to healthcare and shopping has been paralleled by a significant rise in cybercrime. Their review explores how this digital shift has intensified vulnerabilities, particularly among individuals with mental health conditions who may lack awareness of online risks and safety practices. The authors call on psychiatrists to recognize these threats, understand their potential impact on mental well-being, and help protect patients by recommending trusted cybersecurity resources to promote safer digital behavior.

S. Kodate, R. Chiba investigated the detection of fraudulent behavior in consumer to consumer e-commerce platforms by focusing on network-based features rather than traditional user information, which can be easily manipulated by malicious actors. In their study, they modeled the marketplace as a directed graph where users were connected through transactions, constructing egocentric networks for both fraudulent and legitimate users. From these networks, they derived twelve features based on eight local connectivity indices and used them to train random forest classifiers. The results showed that the models effectively distinguished fraudulent users from legitimate ones across various types of problematic transactions, demonstrating the robustness of network-based detection methods irrespective of the specific fraud type.

R. Samani and G. Davis in the McAfee Mobile Threat Report (2019) highlighted the growing risks associated with mobile-based cyber threats, emphasizing how the increasing dependence on smartphones has created new avenues for attackers. The report discusses the surge in mobile malware, phishing attacks, and malicious applications that exploit user data, often bypassing traditional security measures. It underscores the urgent need for improved mobile security awareness and the implementation of advanced threat detection technologies to safeguard users in an increasingly mobile-first digital environment.

E. W. T. Ngai, Y. Hu, reviewed 49 studies on data mining techniques used in financial fraud detection, classifying them by fraud type and analytical method. They found most research focused on insurance and credit card fraud, with less attention to areas like money laundering and mortgage fraud. Common techniques included neural networks, decision trees, Bayesian networks, and logistic regression. The study highlights research gaps and urges further exploration of underrepresented fraud types to meet industry needs.

Sam Smith and Juniper Research forecasted that global losses from online payment fraud will exceed $362 billion between 2022 and 2027, driven by increasingly sophisticated attack methods and rapid growth in digital transactions. The report identifies emerging threats across various segments, including e-commerce, digital banking, and mobile payments, emphasizing the urgent need for advanced fraud prevention strategies. It highlights the importance of AI-driven detection systems and stronger regulatory measures to combat the escalating risks in the online payment landscape.

A. Abdallah, M. A. Maarof, and A. Zainal conducted a detailed survey on fraud detection systems, focusing on various techniques, models, and challenges across multiple domains. The study examines the evolution of fraud detection methods, highlighting the transition from rule-based systems to intelligent approaches such as machine learning, data mining, and hybrid models. It emphasizes the importance of real-time detection, adaptive learning, and handling imbalanced data, while also addressing key challenges like data privacy, scalability, and false positive rates. The survey serves as a foundational reference for researchers aiming to develop more robust and efficient fraud detection frameworks.

R. J. Bolton and D. J. Hand provided a comprehensive review of statistical methods used in fraud detection, emphasizing their application across various industries. The paper outlines key challenges such as the rarity of fraud cases, evolving fraud patterns, and the lack of labeled data. It discusses a range of statistical techniques including supervised and unsupervised learning, anomaly detection, and classification models. The authors highlight the effectiveness of combining multiple approaches to improve accuracy and adaptability, making the study a foundational reference in the field of data-driven fraud detection.

L. Akoglu, H. Tong, and D. Koutra presented a comprehensive survey on graph-based anomaly detection, addressing the growing need to identify unusual patterns in structured data across domains like finance, security, and healthcare. As graph data becomes increasingly common, the authors review state-of-the-art methods designed for detecting anomalies in both static and dynamic, as well as attributed and plain graphs. They propose a unified framework covering unsupervised and semi-supervised approaches, emphasizing key aspects such as scalability, robustness, and interpretability. The survey also explores techniques for explaining detected anomalies and discusses real-world applications in areas like social networks, online auctions, and network traffic, while identifying open challenges for future research.

II. Existing System

The existing approach to understanding e-commerce fraud detection primarily involved an in-depth literature review that drew insights from a wide range of scholarly sources. By employing the PRISMA framework, the research ensured a thorough and systematic selection of relevant publications, allowing for a balanced and structured overview of prevailing fraud detection methodologies. The content was then synthesized and organized based on the techniques and models discussed in the selected works. The review concentrated on grouping these studies according to the specific machine learning and data mining algorithms they utilized. Notably, Artificial Neural Networks (ANNs) and Random Forests emerged as dominant models across the literature. Their frequent appearance underlines the trust placed in their capabilities by the research community, particularly in the context of identifying fraudulent behavior on e-commerce platforms. The investigation was guided by a set of core research questions, which helped shape the overall scope and direction of the study. These questions laid the foundation for exploring how different algorithms functioned, their effectiveness, and the emerging trends in the field of fraud detection. A recurring theme identified in the literature was a strong focus on credit card fraud, reflecting its critical impact on the e-commerce sector. This trend has driven much of the technological innovation and research efforts in fraud detection. Among the algorithms explored, ANNs were the most widely used due to their ability to model complex, non-linear patterns in transactional data, making them particularly effective in identifying subtle anomalies. Random Forests were also heavily utilized for their robustness, scalability, and accuracy in handling large datasets with multiple features. In addition to these mainstream methods, the review also acknowledged a subset of studies using unclustered or alternative data mining techniques. Although less frequently explored, these methods offered different perspectives and insights into fraud detection, contributing to a broader understanding of potential detection strategies.

Disadvantage of existing system

The existing system, while thorough in reviewing literature, has several limitations. It places heavy emphasis on commonly used algorithms like Artificial Neural Networks and Random Forests, overlooking potentially effective techniques such as SVMs and ensemble methods. The focus is predominantly on credit card fraud, neglecting other critical fraud types like identity theft and account takeovers. Additionally, the reliance on synthetic or academic datasets reduces real-world applicability, and the lack of real-time analysis limits adaptability to evolving fraud tactics. Selection bias through the PRISMA framework and limited insights into practical deployment further constrain the system’s effectiveness in diverse e-commerce environments.

III. Proposed System

The proposed e-commerce fraud detection system is designed to meet the growing demand for accurate and efficient identification of fraudulent transactions. Built with Python at its core, the backend is seamlessly connected to a user-friendly frontend developed using HTML, CSS, and JavaScript, all integrated through the Flask web framework. This combination ensures a smooth and responsive interface for end users. To enhance the system's predictive capabilities, two powerful machine learning models—Stacking Classifier and XGBoost (XGB) Classifier—are employed. The Stacking Classifier excels by combining the strengths of multiple base models, while the XGB Classifier is known for its high performance in complex classification tasks. For training and evaluation, a synthetic dataset consisting of 23,634 records was generated using Python's Faker library, supplemented with custom logic to reflect real-world transaction patterns, both legitimate and fraudulent. The dataset includes 16 carefully chosen features such as transaction ID, customer ID, transaction amount, payment method, and a binary fraud indicator. These attributes provide a solid foundation for identifying behavioral anomalies associated with fraudulent activity. The system is built to be both scalable and adaptable, making it suitable for a wide variety of e-commerce environments, regardless of size or transaction volume. By utilizing advanced machine learning methods and a realistic dataset, the proposed solution aims to deliver real-time fraud detection, reinforcing the security and credibility of digital marketplaces.

Advantages

The proposed system offers high accuracy in fraud detection by utilizing powerful machine learning models—Stacking and XGBoost achieving up to 99% test accuracy. Its use of a realistic synthetic dataset ensures strong generalization to various transaction patterns. Designed for scalability and real-time detection, the system is adaptable to different e-commerce platforms and capable of immediate fraud identification. A user-friendly interface and easy integration through the Flask framework make it practical for both technical and non-technical users. Additionally, the system is cost-effective, customizable, and built using open-source tools, making it accessible for businesses of all sizes.

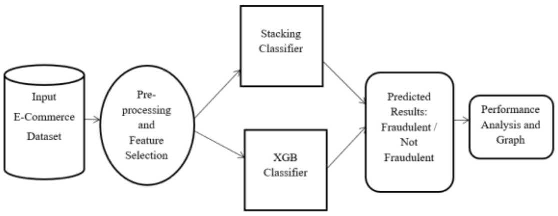

Fig: Architecture Digram

IV. IMPLEMENTATION

The proposed fraud detection system is composed of several integrated modules that collectively enable the identification of fraudulent transactions in a user-friendly and efficient manner.

The **Login Module** serves as the entry point for users, offering a static authentication interface. Although it does not support dynamic user management, it ensures controlled access to the application and restricts usage to authorized personnel. This approach maintains system integrity during demonstration or testing phases.

The **Upload Dataset Module** allows users to import transactional data into the system. In this implementation, a synthetic dataset generated via the Faker library is used, which includes 23,634 records representing both legitimate and fraudulent transactions. This module accepts CSV files and performs initial validation before processing.

Once the dataset is uploaded, the **Prediction Module** becomes active. This module utilizes trained machine learning models—Stacking Classifier and XGBoost—to classify each transaction as either fraudulent or non-fraudulent. It processes the input data in real time and delivers immediate classification results, showcasing the system's practical applicability in active e-commerce environments.

The **Preview Module** provides users with a detailed view of the uploaded data before and after prediction. This helps users verify the structure and contents of the dataset, inspect prediction outcomes, and confirm the model's output.

The **Performance Analysis** Module generates a comprehensive evaluation of the system's accuracy. It presents model performance metrics such as training and testing accuracy, which reached 100% and 99% for the Stacking Classifier, and 96% and 95% for the XGBoost model, respectively. To enhance interpretability, the Pie Chart Module visualizes the distribution of predicted fraudulent versus legitimate transactions. This graphical representation offers a quick overview of the system's classification results, helping users to better understand the scale of detected fraud.

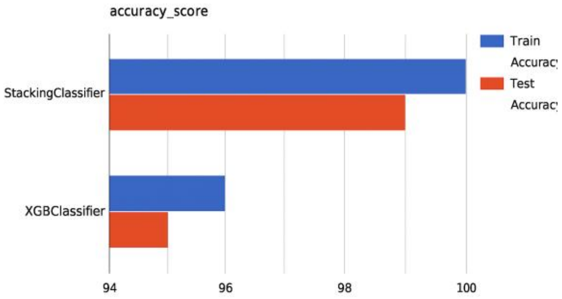

Fig 2: Train & Test Accuracy Score

The bar chart compares the training and testing accuracy of Stacking and XGBoost classifiers. StackingClassifier achieved 100% train and 99% test accuracy, while XGBClassifier achieved 96% and 95% respectively.

Each module plays a vital role in streamlining the fraud detection workflow, from data input to performance evaluation, while ensuring a smooth and intuitive user experience.

V. RESULT

The performance of the proposed fraud detection system was evaluated using a synthetically generated dataset containing 23,634 transaction records, each annotated as either fraudulent or legitimate. The dataset featured 16 variables that reflect real-world e-commerce behavior, such as transaction amount, customer ID, payment method, and a binary fraud indicator.

Two machine learning models were trained and tested: the Stacking Classifier and the XGBoost Classifier. The Stacking Classifier achieved a training accuracy of 100% and a testing accuracy of 99%, indicating excellent generalization with minimal overfitting. The XGBoost Classifier also performed strongly, achieving 96% accuracy on the training set and 95% on the test set, highlighting its robustness and efficiency in classification tasks.

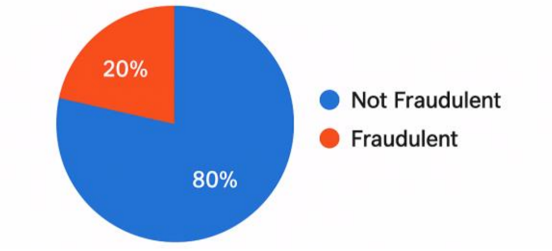

Fig 3: Pie Chart

The system's frontend includes modules for data upload, real-time prediction, and result visualization. Users can preview predictions and evaluate system performance via charts and tables. A pie chart visualization shows the ratio of fraudulent to non-fraudulent predictions, giving users a quick insight into fraud prevalence in the uploaded data. Additionally, accuracy graphs help compare model performance and validate system effectiveness.

The results confirm the system's ability to accurately and efficiently detect fraudulent transactions in real time. High accuracy, fast response, and ease of use make it a suitable tool for practical deployment in online marketplaces.

VI. References

1] S. Monteith, M. Bauer, M. AIda, J. Geddes, P. C. Whybrow, and T. Glenn, Increasing cybercrime since the pandemic: Concerns for psychiatry, Curr. Psychiatry Rep., vol. 23, no. 4, p. 18, 2021.

[2] S. Kodate, R. Chiba, S. Kimura, and N. Masuda, Detecting problematic transactions in a consumer-to-consumer e-commerce network, Appl. Netw. Sci., vol. 5, no. 1, p. 90, 2020.

[3] R. Samani and G. Davis, McAfee mobile threat report, https://www.mcafee.com/enterprise/en-us/assets/reports/rp-mobile-threat-report-2019.pdf, 2019.

[4] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature, Decis. Support Syst., vol. 50, no. 3, pp. 559–569, 2011.

[5] Sam Smith and Juniper Research, Online payment fraud: Market forecasts, emerging threats & segment analysis 2022–2027,https://www.juniperresearch.com/press/losses-online-payment-fraud-exceed-362-billion/, 2024.

[6] A. Abdallah, M. A. Maarof, and A. Zainal, Fraud detection system: A survey, J. Netw. Comput. Appl., vol. 68, pp. 90–113, 2016.

[7] R. J. Bolton and D. J. Hand, Statistical fraud detection: A review, Statistical Science, vol. 17, no. 3, pp. 235–255, 2002.

[8] C. Phua, V. Lee, K. Smith, and R. Gayler, A comprehensive survey of data mining-based fraud detection research, arXiv preprint arXiv: 1009.6119, 2010.

[9] L. Akoglu, H. Tong, and D. Koutra, Graph based anomaly detection and description: A survey, Data Min. Knowl. Discov., vol. 29, no. 3, pp. 626–688, 2015.