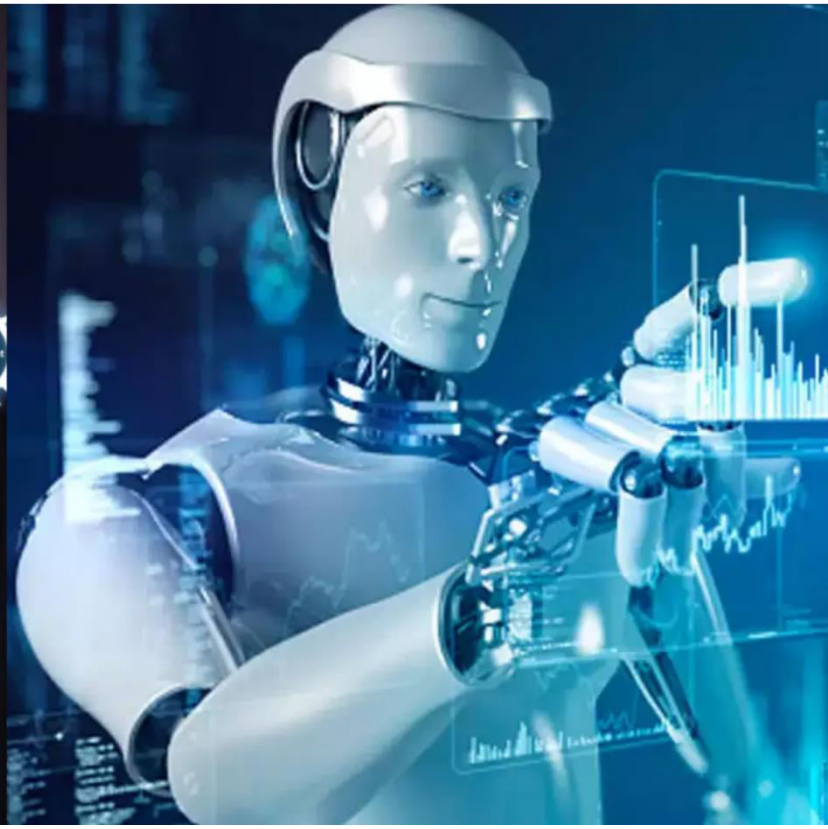
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405350|

E-Commerce Fraud Detection based on Machine Learning

Dr.Veena M, Charan K R ,Chethan H R, Bharath K J, Chethan K U

Associate Professor, Department of Computer Science and Engineering, P.E.S. College of Engineering, Mandya, Karnataka, India

U.G. Student, Department of Computer Science and Engineering, P.E.S. College of Engineering, Mandya, Karnataka, India

U.G. Student, Department of Computer Science and Engineering, P.E.S. College of Engineering, Mandya, Karnataka, India

U.G. Student, Department of Computer Science and Engineering, P.E.S. College of Engineering, Mandya, Karnataka, India

U.G. Student, Department of Computer Science and Engineering, P.E.S. College of Engineering, Mandya, Karnataka, India

ABSTRACT: The project titled "E-Commerce Fraud Detection Based on Machine Learning" addresses the critical challenge of identifying fraudulent transactions in e-commerce platforms. Developed using Python for backend processing and HTML, CSS, and JavaScript for the frontend interface, the system is integrated within the Flask web framework to deliver a responsive and interactive user experience. The project leverages two advanced machine learning models: a Stacking Classifier and an XGB Classifier. The Stacking Classifier achieves an impressive train accuracy of 100% and a test accuracy of 99%, while the XGB Classifier attains a train accuracy of 96% and a test accuracy of 95%. These high performance metrics underscore the models' effectiveness in distinguishing between legitimate and fraudulent transactions. The dataset utilized comprises 23,634 synthetic records generated through Python's Faker library, enriched with custom logic to emulate realistic transaction patterns and fraudulent scenarios. The dataset includes 16 features such as Transaction ID, Customer ID, Transaction Amount, Payment Method, and a binary indicator for fraudulent activity, among others. These features collectively capture the intricacies of transaction behaviors and customer profiles, enabling robust fraud detection. The project's results demonstrate the potential of machine learning techniques in enhancing security and trust in e-commerce environments, providing a powerful tool for preventing financial loss due to fraudulent activities.

KEYWORDS: E-Commerce, Fraud Detection, Machine Learning, Stacking Classifier, XGBoost, Transaction Analysis, Anomaly Detection, Classification Algorithms, Data Preprocessing, Predictive Modeling, Real-Time Detection, Financial Fraud, Flask Web Application.

I. INTRODUCTION

Deep learning, a subset of machine learning, utilizes multi-layered neural networks to learn hierarchical data representations. Unlike conventional methods that rely heavily on manual feature engineering, deep learning models automatically extract patterns from raw data using stacked nonlinear transformations. This approach has revolutionized fields such as image recognition, natural language processing, speech recognition, and increasingly, cybersecurity and fraud detection.

The advent of deep neural networks, including Convolutional Neural Networks (CNNs) for image and video processing and Recurrent Neural Networks (RNNs) for sequential data like speech and text, has enabled machines to learn complex abstractions. Innovations such as Long Short-Term Memory (LSTM) units and memory-augmented models (e.g., Neural Turing Machines, Memory Networks) address the limitations of traditional RNNs, particularly in

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405350|

capturing long-term dependencies and reasoning over sequences.

In the domain of fraud detection, deep learning offers a significant advantage due to its ability to analyze vast volumes of transactional data in real time. With the rise of E-commerce platforms, particularly during the COVID-19 pandemic, the frequency and complexity of online fraud have surged. Traditional rule-based systems often fail to adapt to evolving fraud tactics. Deep learning models, however, can dynamically learn fraudulent patterns by analyzing behavioral signals, transaction sequences, and contextual metadata.

Furthermore, the availability of large datasets and advances in GPU computing have catalyzed the training of deeper and more accurate fraud detection models. Techniques such as autoencoders, LSTMs, and hybrid architectures (e.g., stacked classifiers) have shown great promise in classifying anomalies and fraudulent behavior in financial transactions.

As cybercriminals continuously evolve their strategies, adaptive models powered by deep learning provide a scalable and intelligent solution for proactive fraud prevention across digital platforms.

II. LITERATURE SURVEY

E-commerce platforms have seen tremendous growth in recent years, especially after the global pandemic, which increased online transactions. However, this expansion has also led to a significant rise in fraudulent activities, making fraud detection a critical area of research.

Monteith et al. (2021) highlighted the surge in cybercrime during the pandemic, especially targeting individuals using online services from home. Their study emphasized the psychological and financial consequences of fraud, underscoring the need for robust detection systems.

Kodate et al. (2020) introduced a network-based approach for detecting problematic transactions in consumer-to-consumer marketplaces. By analyzing the connections between users and applying local network indices, they effectively differentiated between normal and fraudulent users.

Phua et al. (2010) presented a comprehensive survey on data mining techniques used in fraud detection. They categorized the fraud types, datasets, and machine learning models used, emphasizing the importance of selecting appropriate algorithms and data features to improve detection rates.

Ngai et al. (2011) reviewed the use of data mining in financial fraud detection, offering a classification framework. They concluded that hybrid and ensemble models outperform single algorithms in complex fraud scenarios.

In recent research, XGBoost and Stacking Classifiers have shown remarkable performance in fraud detection due to their high accuracy and ability to handle imbalanced datasets. The use of synthetic datasets has also been explored to simulate realistic transaction environments while maintaining privacy and control over data characteristics.

These studies collectively indicate that a combination of data preprocessing, feature engineering, and advanced machine learning models can significantly improve fraud detection accuracy. They also highlight the growing need for real-time and scalable solutions that can adapt to evolving fraud tactics.

III. METHODOLOGY

The methodology adopted for this project involves multiple phases aimed at developing a robust fraud detection system using machine learning techniques. The steps include data preparation, model building, evaluation, and deployment.

1. Data Collection:A synthetic dataset consisting of 23,634 records was generated using Python's Faker library. The dataset simulates real-world E-commerce transactions, containing 16 features such as Transaction ID, Customer ID, Transaction Amount, Location, Payment Method, and a binary target indicating fraud status.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405350|

2. Data Preprocessing:Preprocessing tasks were performed to clean and normalize the data:
Removal of null and duplicate entries
Feature selection to retain the most relevant attributes
Encoding of categorical variables
Scaling of numerical features
3. Model Selection and Training:Two machine learning models were implemented:
Stacking Classifier: Combines the output of base models like Random Forest and Decision Tree using a Logistic Regression meta-classifier to enhance accuracy.
XGBoost Classifier: A powerful gradient boosting algorithm optimized for performance and accuracy.
The dataset was split into training (80%) and testing (20%) sets. Both models were trained and fine-tuned to classify transactions as fraudulent or legitimate.

4. Model Evaluation:Performance metrics such as accuracy, precision, recall, and F1-score were used to evaluate the models:
Stacking Classifier: 100% train accuracy, 99% test accuracy
XGBoost: 96% train accuracy, 95% test accuracy
Confusion matrices and heatmaps were generated to visualize classification performance.

5. System Implementation:The backend was developed using Python and Flask, while the frontend used HTML, CSS, and JavaScript. Users can submit transaction details through a web interface, and the system provides immediate feedback on whether the transaction is fraudulent.
6. Model Deployment:The trained models were saved using pickle for deployment.

IV. EXPERIMENTAL RESULTS

The effectiveness of the proposed E-commerce fraud detection system was evaluated using a synthetically generated dataset comprising 23,634 transaction records. Each record included sixteen distinct features representing various aspects of a transaction, such as transaction amount, payment method, customer identification, and location data. These features were selected to mimic real-world E-commerce activity, with the binary class label indicating whether a transaction was fraudulent or legitimate. For model development and evaluation, the dataset was divided into training and testing sets in an 80:20 ratio. Two machine learning models were implemented and assessed: the XGBoost Classifier and a Stacking Classifier composed of Random Forest and Decision Tree as base learners, with Logistic Regression serving as the meta-classifier. The models were trained to recognize complex patterns that distinguish between fraudulent and non-fraudulent transactions. To ensure a thorough evaluation, standard performance metrics were employed, including accuracy, precision, recall, and F1-score. The XGBoost Classifier demonstrated solid performance, achieving a training accuracy of 96 percent and a testing accuracy of 95 percent. Its precision and recall values reflected reliable fraud detection with minimal misclassification. However, the Stacking Classifier outperformed XGBoost across all metrics, achieving a perfect training accuracy of 100 percent and a near-perfect testing accuracy of 99 percent. This superior performance was consistent across other metrics as well, with the model displaying high precision and recall, suggesting its strong ability to detect fraudulent transactions without generating excessive false positives. The models were tested in a controlled environment using a system with modest hardware specifications, specifically an Intel i3 processor with 8GB of RAM running on Windows 11. Despite these limited resources, the application maintained an efficient response time, classifying transactions in under two seconds, thus demonstrating its potential for real-time deployment.Overall, the results highlight the effectiveness and reliability of ensemble-based approaches such as the Stacking Classifier in fraud detection tasks. The experimental evaluation confirms that integrating multiple learners can significantly enhance classification performance, especially in domains where the cost of false negatives is high, such as financial fraud prevention.

V. CONCLUSION

The "E-Commerce Fraud Detection Based on Machine Learning" project successfully addresses the critical need for effective fraud detection mechanisms in the rapidly growing e-commerce sector. By employing advanced machine learning models like the Stacking Classifier and XGB Classifier, the system achieves high accuracy in identifying fraudulent transactions, significantly reducing the risk of financial losses and enhancing the security of online

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405350|

transactions. The use of a synthetic dataset, carefully designed to mimic real-world transaction scenarios, enables the system to generalize well across different types of fraud, ensuring its applicability in diverse e-commerce environments. The integration of this system into a user-friendly web interface, built using HTML, CSS, JavaScript, and the Flask framework, ensures that it can be easily utilized by businesses of varying sizes and technical capabilities. Overall, the project demonstrates the power and potential of machine learning in combating e-commerce fraud, providing a robust and scalable solution that enhances the trust and reliability of online marketplaces. By offering real-time detection and a comprehensive analysis of transaction data, the system stands as a valuable tool in the ongoing effort to safeguard digital commerce.

ACKNOWLEDGEMENT

We sincerely thank our project guide, Dr. Veena M, for their guidance and support throughout the development of this project. We also express our gratitude to the faculty of PES College of Engineering, Mandya, for providing the resources and encouragement needed. Lastly, we thank our friends and family for their constant support.

REFERENCES

[1] S. Monteith, M. Bauer, M. Alda, 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] 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.
[4] 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.
[5] A. Abdallah, M. A. Maarof, and A. Zainal, Fraud detection system: A survey, J. Netw. Comput. Appl., vol. 68, pp. 90–113, 2016.
[6] R. J. Bolton and D. J. Hand, Statistical fraud detection: A review, Statistical Science, vol. 17, no. 3, pp. 235–255, 2002.
[7] 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.
[8] 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

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details