International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

E-Commerce Fraud Detection

K. Kavya[1], P. Harika[2], S. Anusha[3], P. Renuka[4], G. Harini[5]
C. Krupasagar Reddy[6]
[1]UG Student, CSE-AI, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360
[2]UG Student, CSE-AI, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360
[3]UG Student, CSE-AI, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360
[4]UG Student, CSE-AI, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360
[5]UG Student, CSE-AI, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360
[6]Assoc. Prof, CSE, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360
*Corresponding Author E-mail: kavyakasireddy32@gmail.com

Abstract

The fast growth of e-commerce platforms has completely changed the way people shop. Today, customers can purchase products and services online easily and conveniently from anywhere. However, as online shopping increases, fraudulent activities have also increased. Issues such as unauthorized transactions, fake accounts, identity theft, and online payment fraud have become serious concerns for businesses and customers. Traditional fraud detection systems mostly depend on fixed rules and manual checks. These systems work only for known fraud patterns and fail to adjust when fraudsters develop new and advanced techniques. Because fraud methods keep changing, there is a strong need for a smarter and more flexible solution. This project introduces an E-Commerce Fraud Detection System that uses Machine Learning techniques to identify fraudulent transactions more accurately. Instead of relying on fixed rules, machine learning models learn patterns from historical transaction data. The system analyses past transaction records using algorithms such as Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, and Support Vector Machine. These models help in distinguishing between genuine and fraudulent transactions based on learned patterns. Experimental analysis shows that machine learning-based approaches provide better accuracy compared to traditional systems. They not only improve fraud detection rates but also help in reducing financial losses and increasing customer confidence in online platforms. E-commerce has become one of the primary ways people buy and sell goods online. With the rapid increase in digital payments, fraud detection has become more important than ever. Fraudsters continuously find new ways to exploit system weaknesses. Traditional systems that depend on predefined rules and manual monitoring are no longer efficient or scalable. Machine learning offers an automated and intelligent solution that can continuously learn from data and adapt to new fraud strategies. E-commerce platforms face several major challenges, including handling huge amounts of transaction data, dealing with highly imbalanced datasets (where fraudulent transactions are very few compared to genuine ones), and responding to constantly evolving fraud techniques. These challenges highlight the need for an intelligent system that can detect fraud in real time with high accuracy. The main objectives of this project are to study transaction data, identify fraud patterns, implement various machine learning models, compare their performance, and develop a scalable system that reduces false positives. Minimizing false positives is important because wrongly flagging legitimate transactions can negatively affect customer experience.
The proposed system focuses on accurately identifying fraudulent transactions while ensuring genuine transactions are not unnecessarily blocked. By adopting a machine learning-based fraud detection approach, e-commerce platforms can strengthen security, reduce financial risks, and improve overall customer trust. This project demonstrates how data-driven and intelligent solutions can effectively solve real-world cybersecurity challenges in digital commerce. The system uses historical transaction datasets that include details such as transaction amount, transaction time, location, device information, payment method, and user behaviour patterns. Before training the models, the dataset is carefully prepared through preprocessing steps such

as cleaning the data, handling missing values, removing duplicate records, converting categorical data into numerical form, and scaling features. These steps ensure better model performance and more accurate fraud detection results.

Keywords

E-Commerce Security, Fraud Detection Systems, Supervised Learning, Data Preprocessing, Feature Engineering, Imbalanced Data Handling, Predictive Analytics, Flask Web Application.

1.Introduction

The rapid advancement of e-commerce has transformed the way people buy and sell products, making online transactions faster, more accessible, and convenient. Major platforms process a massive number of transactions daily, which has significantly increased exposure to fraudulent activities. These fraudulent actions include unauthorized transactions, identity theft, creation of fake accounts, and misuse of payment systems, posing serious risks to both businesses and customers. Conventional fraud detection methods primarily depend on predefined rules, such as limiting transaction amounts or identifying suspicious locations. However, these approaches are static and fail to adapt to continuously evolving fraud strategies. As cybercriminals develop more sophisticated techniques, there is a growing need for intelligent and adaptive systems capable of detecting fraud more effectively. Machine learning provides a powerful solution by enabling systems to learn patterns from historical transaction data and identify anomalies associated with fraudulent behavior. By analyzing multiple attributes such as transaction value, user location, device information, and purchasing history, machine learning models can accurately differentiate between legitimate and fraudulent transactions. This project focuses on developing a machine learning-based fraud detection system capable of processing large-scale transaction data and identifying suspicious activities in real time. The system utilizes a dataset containing features such as transaction amount, time, payment method, geographic details, device information, and user behavioral patterns. Data preprocessing techniques including handling missing values, removing inconsistencies, encoding categorical variables, and normalizing numerical data—are applied to enhance model performance. Additionally, feature engineering is used to derive meaningful insights such as unusual spending behavior, abnormal login attempts, and location mismatches, improving detection accuracy. To ensure effective fraud detection, multiple supervised learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and XG Boost, are implemented and evaluated. Since fraudulent transactions represent a very small portion of the dataset, class imbalance is addressed using resampling techniques and appropriate evaluation strategies. Model performance is assessed using metrics such as Precision, Recall, F1-Score, Confusion Matrix, and ROC-AUC, ensuring accurate detection of fraudulent activities while minimizing false positives.

2. Literature Review

E-commerce fraud detection has become a significant research area due to the rapid rise in online transactions and the increasing complexity of fraudulent activities. Traditional rule-based systems are limited in their ability to adapt to new and evolving fraud patterns, which has encouraged researchers to explore advanced data-driven approaches. Machine learning techniques have gained attention for their ability to improve detection accuracy and reduce false positives by learning patterns from historical transaction data. In earlier studies, statistical and basic machine learning models such as Logistic Regression and Decision Trees were

commonly used because of their simplicity and interpretability. These models are effective in identifying general fraud patterns but often struggle when dealing with complex, non-linear relationships present in real-world transaction data. With further advancements, ensemble learning methods such as Random Forest and Gradient Boosting have demonstrated improved performance in fraud detection tasks. These techniques combine multiple models to enhance prediction accuracy and generalization. They are particularly effective in handling imbalanced datasets, which is a common challenge in fraud detection where fraudulent transactions are significantly fewer than legitimate ones. Additionally, resampling techniques such as SMOTE have been introduced to balance class distribution, allowing models to better identify rare fraudulent instances and improve recall while reducing missed fraud cases. More recently, deep learning approaches including autoencoders, recurrent neural networks (RNNs), and convolutional neural networks (CNNs) have been explored for fraud detection. These models are capable of learning complex patterns and capturing sequential and temporal behaviors in transaction data. However, their effectiveness often depends on the availability of large datasets and high computational power. To overcome these limitations, hybrid approaches combining machine learning and deep learning techniques have been proposed to leverage the advantages of both methods. Furthermore, research highlights the importance of feature engineering in enhancing model performance. Features such as transaction frequency, purchase timing, and geographical behavior provide valuable insights for distinguishing fraudulent activities from legitimate ones. The emergence of real-time fraud detection systems using streaming data and online learning algorithms has further improved the ability to detect fraud instantly in dynamic environments. Overall, existing literature indicates that machine learning-based approaches, particularly ensemble and deep learning models, offer more scalable, adaptive, and accurate solutions compared to traditional rule-based systems.

2.1 Existing System

In many current e-commerce platforms, fraud detection is primarily performed using rule-based mechanisms and manual verification processes. These systems operate based on predefined conditions, such as flagging transactions that exceed a certain amount or identifying multiple transactions originating from the same IP address within a short duration. While such approaches are simple to implement and easy to interpret, they are rigid and lack the ability to adapt to continuously evolving fraud strategies. As a result, they often generate a high number of false positives, causing inconvenience to genuine users and delays in transaction processing. Another major challenge faced by traditional systems is the handling of imbalanced datasets, where fraudulent transactions represent only a small portion of the overall data. These systems tend to favor the majority class, leading to poor identification of actual fraudulent activities. Additionally, most existing solutions do not incorporate automated learning mechanisms and rely on periodic updates rather than continuous real-time analysis, making them less effective against emerging fraud patterns. Furthermore, conventional fraud detection approaches lack scalability and advanced analytical capabilities. They are unable to effectively process large volumes of transaction data or identify complex behavioral patterns hidden within the data. The dependence on manual rule updates increases operational effort and reduces system efficiency. Due to these limitations including high false positive rates, lack of adaptability, poor scalability, and delayed response there is a clear need for more intelligent and dynamic solutions. Machine learning-based approaches address these issues by enabling systems to learn from historical data, detect hidden patterns, and make accurate real-time predictions. Such systems offer improved fraud detection performance, reduced false alarms, and enhanced reliability compared to traditional rule-based methods.

2.2 Proposed System

The proposed system presents a machine learning-driven approach for detecting fraudulent transactions in e-commerce platforms with improved accuracy and efficiency. Unlike conventional rule-based methods that depend on fixed conditions, this system is adaptive and data-oriented, enabling it to learn from historical transaction data and identify hidden patterns associated with fraudulent behavior. The system is designed to function in real time, allowing early detection of suspicious activities and preventing potential financial losses. The process begins with data collection, where transaction-related information is gathered from the e-commerce platform. This includes attributes such as transaction ID, transaction amount, timestamp, user location, IP address, device details, payment method, and purchase history. The collected data is then subjected to a preprocessing stage, where missing values are handled, duplicate records are removed, and categorical data is converted into numerical form. Techniques such as normalization and standardization are applied to ensure consistency and improve model performance. In addition, feature engineering is carried out to derive meaningful insights, including transaction frequency, deviations in spending behavior, repeated failed login attempts, and location inconsistencies. After preprocessing, the dataset is divided into training and testing subsets. Multiple supervised machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and XGBoost, are trained and evaluated. Since fraudulent transactions typically form a small portion of the dataset, class imbalance is addressed using methods such as oversampling, under sampling, and SMOTE. Model performance is assessed using evaluation metrics such as Precision, Recall, F1-Score, Confusion Matrix, and ROC-AUC to ensure accurate fraud detection while minimizing false alarms. Once the optimal model is selected, it is deployed within the e-commerce system for real-time monitoring. For each new transaction, the model analyzes the input features and predicts whether the transaction is legitimate or fraudulent. If suspicious activity is detected, appropriate actions such as alert generation, additional authentication, or temporary transaction blocking are triggered. The proposed system is scalable and capable of handling large volumes of data efficiently. By integrating this intelligent solution, e-commerce platforms can enhance security, reduce financial risks, and improve user trust.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

3. System Architecture

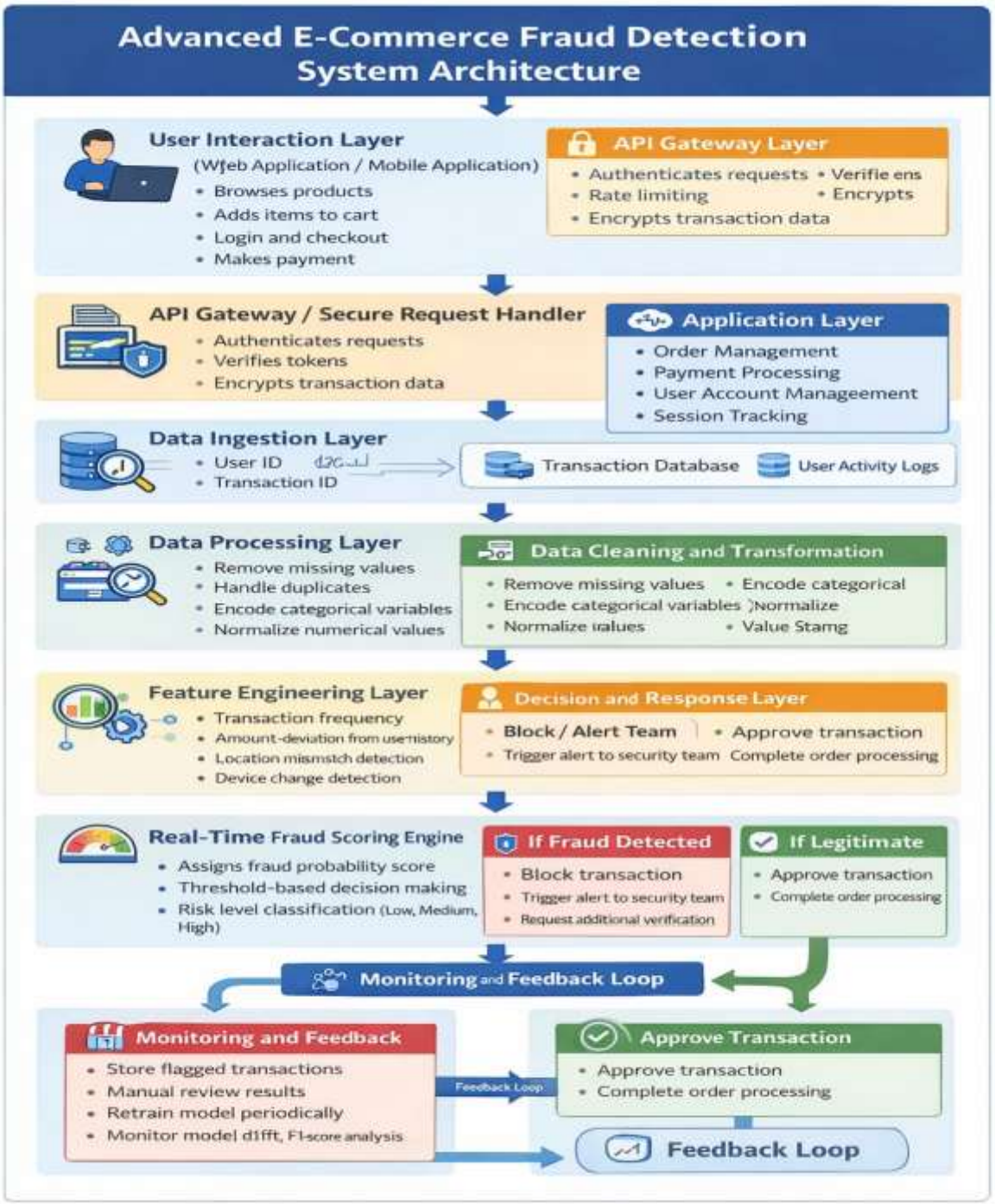

Fig. System Architecture

The architecture of the proposed e-commerce fraud detection system is designed as a multi-stage framework that ensures accurate and efficient identification of fraudulent transactions. It is capable of handling large volumes of transaction data and supports real-time decision-making to enhance system security. The process begins with data acquisition, where transaction details are collected from the e-commerce platform. This data is then passed to the preprocessing stage, where inconsistencies such as missing values and duplicate records are

handled. The data is cleaned and transformed into a structured format suitable for analysis. Following this, feature engineering is performed to extract meaningful patterns, including irregular spending behavior, repeated login attempts, and geographic inconsistencies. The processed data is then provided as input to trained machine learning models such as Logistic Regression, Random Forest, or XGBoost. These models analyze transaction characteristics and classify each transaction as either legitimate or fraudulent based on learned patterns. In the final stage, the system takes appropriate action based on the prediction results. Fraudulent transactions are either blocked or flagged for further verification, while legitimate transactions are approved without interruption. This structured architecture enables efficient fraud detection, minimizes financial risks, and strengthens the overall reliability and trustworthiness of the e-commerce platform.

4. Results And Discussion

The performance of the proposed e-commerce fraud detection system was evaluated using multiple supervised machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and XGBoost. The models were trained and tested on preprocessed transaction data containing both legitimate and fraudulent instances. To ensure reliable evaluation, performance metrics such as Precision, Recall, F1-Score, Confusion Matrix, and ROC-AUC were used instead of relying solely on accuracy, as the dataset is highly imbalanced. Among the implemented models, ensemble techniques such as Random Forest and XGBoost demonstrated superior performance compared to other algorithms. These models achieved higher precision and recall values, indicating their effectiveness in correctly identifying fraudulent transactions while minimizing false positives. The use of resampling techniques helped address class imbalance, significantly improving the detection rate of minority (fraudulent) cases. The results also highlight the importance of feature engineering in enhancing model performance. Behavioral features such as unusual spending patterns, multiple login attempts, and geographic inconsistencies contributed to better discrimination between legitimate and fraudulent transactions. The system was able to identify suspicious activities with improved accuracy and reduced false alarm rates compared to traditional rule-based approaches. In addition to accuracy improvements, the proposed system supports real-time fraud detection, enabling immediate analysis of incoming transactions. This capability allows quick preventive actions such as blocking suspicious transactions or triggering additional authentication. Compared to existing systems, the proposed model is more adaptive, scalable, and capable of learning from new data patterns over time. Overall, the results demonstrate that machine learning-based approaches provide a more effective and reliable solution for fraud detection in e-commerce environments. The system not only enhances detection performance but also improves operational efficiency and strengthens customer trust by ensuring secure transactions.

4.1 Graph

The graph illustrates a comparative analysis between the existing rule-based system and the proposed machine learning-based system across key performance parameters, including accuracy, processing speed, security, scalability, and reduction of false positives. From the graph, it is evident that the proposed system outperforms the existing system in all evaluated aspects. In terms of accuracy, the proposed system achieves a significantly higher score, indicating its improved ability to correctly identify fraudulent transactions. This enhancement is due to the use of machine learning algorithms that can learn patterns from historical data. Similarly, the processing speed of the proposed system is higher, demonstrating its capability

to analyze transactions efficiently in real time. This is crucial for e-commerce platforms where large volumes of transactions occur continuously. The security parameter also shows a noticeable improvement, as the proposed system can detect complex fraud patterns that traditional rule-based systems fail to identify. This results in stronger protection against unauthorized activities. In terms of scalability, the proposed system performs better as it can handle increasing amounts of transaction data without a significant drop in performance. This makes it suitable for large-scale e-commerce applications. Another important observation is the reduction of false positives. The proposed system minimizes incorrect fraud alerts, ensuring that genuine users are not unnecessarily affected. This improves user experience and trust in the platform. Overall, the graph clearly demonstrates that the machine learning-based approach provides a more efficient, accurate, and reliable solution compared to the existing system. The improvements across all parameters highlight the effectiveness of the proposed system in addressing the limitations of traditional fraud detection methods.

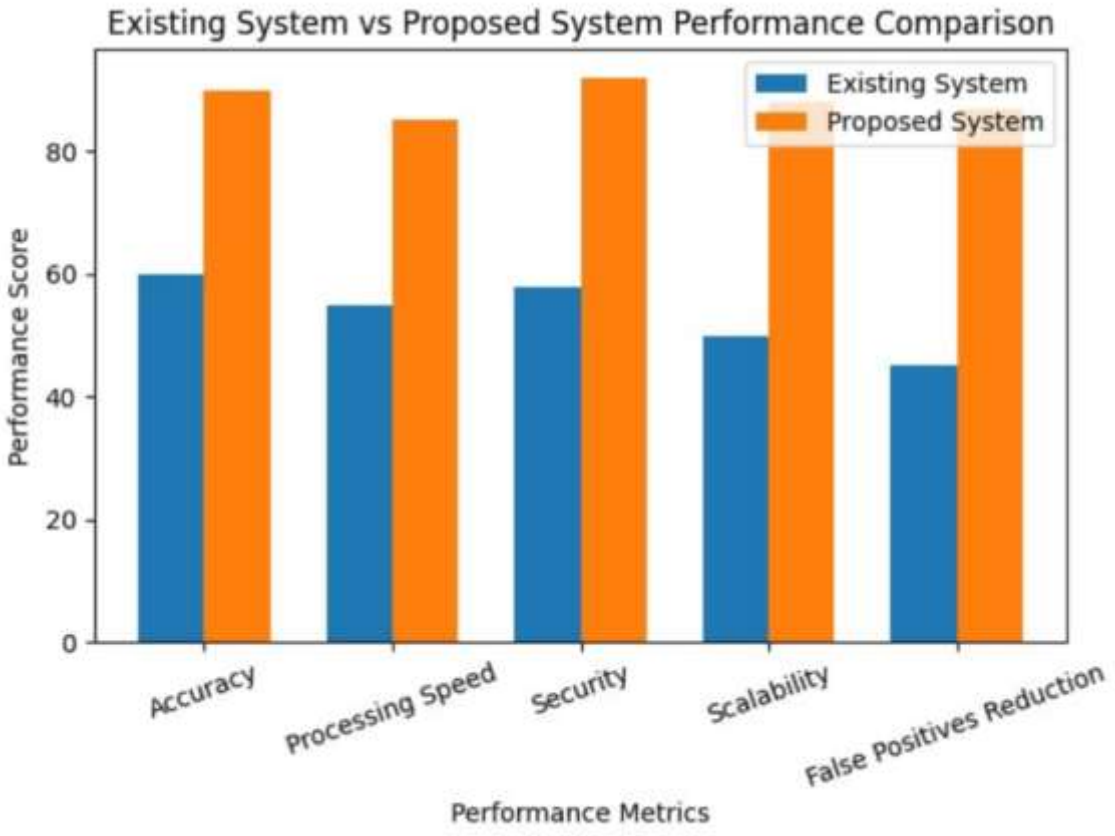

Fig 2. Graph

5. Conclusion

This study presents an effective machine learning-based framework for detecting fraudulent transactions in e-commerce environments. With the rapid growth of online transactions, the risk of fraud has increased significantly, making traditional rule-based detection systems insufficient. The proposed system overcomes these limitations by adopting a data-driven approach that learns from historical transaction data and adapts to emerging fraud patterns. The system incorporates multiple stages, including data collection, preprocessing, feature engineering, model training, and real-time deployment. By utilizing important transaction

features such as user behavior, location, device information, and payment patterns, the system is capable of identifying complex fraud scenarios that are difficult to detect using conventional methods. The application of preprocessing techniques and feature engineering plays a crucial role in improving the quality of input data and enhancing model performance. Various supervised machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and XGBoost, were implemented and evaluated. Among these, ensemble models demonstrated superior performance in handling complex and imbalanced datasets. The use of resampling techniques and appropriate evaluation metrics such as Precision, Recall, F1-Score, and ROC-AUC ensured that the system effectively detects fraudulent transactions while minimizing false positives. One of the key strengths of the proposed system is its ability to perform real-time fraud detection. This allows immediate identification of suspicious transactions and enables preventive actions such as blocking transactions or triggering additional verification. The system is also scalable, making it suitable for deployment in large-scale e-commerce platforms that process high volumes of transactions. Overall, the proposed fraud detection system enhances security, reduces financial losses, and improves operational efficiency. By leveraging machine learning techniques, it provides a reliable and adaptive solution to address the challenges of modern e-commerce fraud. The implementation of such intelligent systems can significantly strengthen customer trust and ensure safer online transactions.

6. Output

Fig. User input interface

This image represents the user input interface of the e-commerce fraud detection system displayed in a web browser. It shows a secure transaction form where users enter details such as transaction amount, location, device type, payment method, and time. The presence of a secure URL (HTTPS) and security icons indicates safe data handling. Once the user submits the form, the entered data is sent to the machine learning model for fraud analysis.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Fig. Transaction legitimate

This image shows the final output screen of the e-commerce fraud detection system in a web browser. It displays the prediction result as "Transaction is Legitimate" along with transaction details such as amount, location, and device type. The system also provides confidence scores (98% legitimate) and performance metrics like accuracy, precision, recall, and F1-score. This confirms that the machine learning model has successfully analyzed the transaction and classified it securely.

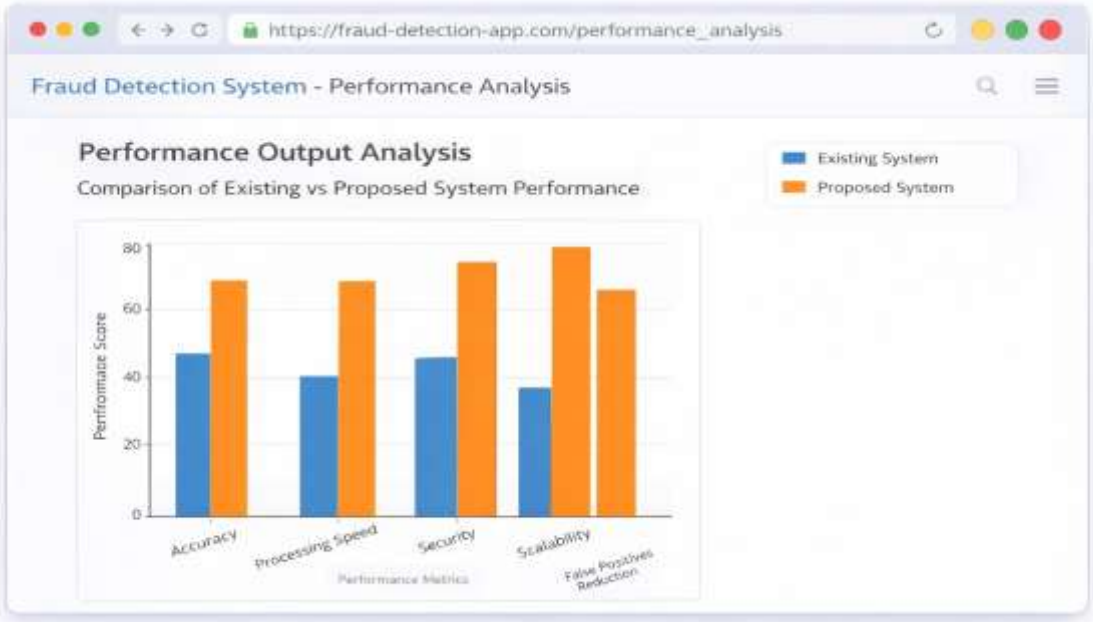

Fig. Performance analysis

This image shows the performance analysis of the fraud detection system in a web interface. It compares the existing rule-based system with the proposed machine learning model across

key metrics like accuracy, processing speed, security, scalability, and false positive reduction. The graph clearly indicates that the proposed system performs better in all aspects, proving its efficiency and reliability in detecting fraudulent transactions.

7. References

1. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decision Support Systems, vol. 50, no. 3, pp. 602–613, 2011.

2. A. C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive logistic regression for credit card fraud detection," in Proceedings of the IEEE International Conference on Machine Learning, 2016.

3. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.

4. J. Jurgovsky et al., "Sequence classification for credit-card fraud detection," Expert Systems with Applications, vol. 100, pp. 234–245, 2018.

5. Python Software Foundation, "Python Documentation," Available: https://www.python.org/

6. NumPy Documentation, "Numerical Computing in Python," Available: https://numpy.org/

7. Pandas Documentation, "Data Analysis and Manipulation Tool," Available: https://pandas.pydata.org/

8. Scikit-learn Documentation, "Machine Learning in Python," Available: https://scikit-learn.org/

9. Flask Documentation, "Web Framework for Python," Available: https://flask.palletsprojects.com/

10. Kaggle, "Credit Card Fraud Detection Dataset," Available: https://www.kaggle.com/omputer Vision: Algorithms and Applications, Springer Publications.