JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

IDENTIFYING FRAUDLENT ACTIVITIES DETECTION IN E-COMMERCE WEBSITES

Dr. P V KUMAR - Professor, Department of IT, Anurag University,

Mr. V. SAI GANESH- Student, Department of Computer Science, Anurag University,

Mr. V. NAGARAJU - Student, Department of Computer Science, Anurag University,

Mr. CH.VENKATESWARA RAO- Student, Department of Computer Science, Anurag University,

Abstract

The rapid growth of internet users has led to a corresponding increase in e-commerce transactions, but this surge in online activity also brings a rise in fraudulent activities. Fraud prevention on e-commerce platforms is an escalating concern, requiring advanced solutions to maintain customer trust and safeguard financial transactions. This paper focuses on using machine learning to detect and prevent fraud in e-commerce. It examines a variety of machine learning algorithms, including Decision Trees, Naive Bayes, Random Forest, and ensembled approaches, to identify which is most effective in combating online fraud. E-commerce platforms handle significant amounts of money daily, making them prime targets for fraudulent activities such as credit card theft, identity theft, and money laundering. Fraudulent users often exploit these platforms to conduct illegal activities, posing a risk to both businesses and customers. Effective fraud prevention mechanisms are crucial to mitigate these risks, and machine learning provides a powerful tool for identifying suspicious patterns and behaviors. This study analyzes the performance of the aforementioned algorithms in detecting e-commerce fraud, focusing on their accuracy, precision, and adaptability to changing fraud trends. By leveraging these algorithms, e-commerce platforms can automatically flag potentially fraudulent transactions for further review, reducing the impact of fraud on businesses and consumers.The aim of this research is to identify the most suitable machine learning approach for e-commerce fraud detection, providing insights that can guide platform developers and security professionals in implementing robust fraud prevention systems.

1. INTRODUCTION

A report by eMarketer, a leading digital marketing and commerce research company, estimates that global e-commerce sales in 2020 would reach $4 trillion, representing about 14.6% of the estimated total retail sales of $27 trillion. This reflects a significant increase compared to 2016, when e-commerce sales accounted for only 8.7% of global retail sales, totaling $1.915 trillion out of $22.049 trillion. The substantial growth in e-commerce is driven by shifting consumer behavior, where shoppers prefer the convenience of comparing and purchasing products from home or work. Retailers find the e-commerce model more cost-effective than traditional retail, enabling them to offer products at lower prices, further boosting online sales. Additionally, the rise of online marketplaces, where multiple vendors use a common platform, has contributed to this trend. Despite the growth of e-commerce, the industry has seen a corresponding increase in fraud. Since 1993, e-commerce-related fraud has escalated, with a 2013 report indicating that 5.65 cents are lost due to fraud for every $100 in e-commerce transactions. Fraud detection is crucial to prevent financial losses and maintain consumer trust. Common types of e-commerce fraud include stolen credit or debit card information and fraudulent product returns. Researchers have proposed various strategies for detecting card-related fraud, such as Artificial Immune Systems, Inductive Learning, and

Neural Data Mining. Another emerging concern is merchant fraud in online marketplaces, which directly impacts customer satisfaction and marketplace credibility. This makes it critical for marketplace owners to identify fraudulent sellers. With the advent of big data and machine learning, it is now possible to analyze historical data to identify potential fraudulent behavior in e-commerce. Machine learning techniques like Decision Trees, Random Forests, and Neural Networks can be used to predict fraudulent activities and proactively prevent them. This framework aims to leverage these technologies to enhance e-commerce fraud detection and maintain a safe online marketplace.

1.2 Scope of the Project:

This project aims to explore and evaluate various machine learning algorithms for detecting fraud in e-commerce transactions. Specifically, it will analyze the performance of Decision Trees, Naive Bayes, Random Forest, and ensembled approaches in identifying fraudulent activities. By comparing their accuracy, precision, and adaptability, the project seeks to determine the most effective algorithm for e-commerce fraud detection. The findings will guide e-commerce platform developers and security professionals in implementing robust fraud prevention systems to safeguard customer data and financial transactions.

2. LITERATURE SURVEY

Srivastava et al. (2018) [1] proposed a method using a Hidden Markov Model (HMM) to detect fraud in credit card transactions. This model is trained on the normal behavior of cardholders, and any transaction not fitting the expected pattern is flagged as potentially fraudulent. The paper shows detailed experimental results demonstrating the effectiveness of this approach in detecting fraud.

Lakshmi et al. (2019) [2] investigated the use of logistic regression, decision trees, and random forests for credit card fraud detection. Working with an imbalanced dataset from Kaggle, they applied oversampling to achieve a 60% fraud and 40% non-fraud distribution. The study shows that among the three techniques, random forest achieved the highest accuracy rate, outperforming logistic regression and decision trees. This finding aligns with previous work on machine learning in credit card fraud detection.

Aljarah et al. (2019) [3] proposed a new training algorithm based on the Whale Optimization Algorithm (WOA) to optimize connection weights in neural networks. The paper examines WOA's effectiveness across 20 datasets, demonstrating improved local optima avoidance and convergence speed compared to traditional back-propagation and other evolutionary algorithms. This work underscores the potential of neural networks in improving fraud detection.

Xuan et al. (2018) [4] introduced a novel approach to refine the voting mechanism in random forests. Instead of treating all base classifiers equally, they proposed weighted voting to improve decision accuracy. Their experiments on public datasets showed better performance than traditional random forests, offering a more nuanced method for fraud detection.

Zhao et al. (2016) [5] developed a framework to analyze implicit online user behavior to detect fraudulent activities in e-commerce. Their approach achieved an average true positive detection rate of 83% with a low false alarm rate of 2.4%, demonstrating its effectiveness in identifying fraudulent transactions. The study's emphasis on behavioral evidence provides a broader perspective on fraud detection.

Supporting References

The findings from these studies can be related to the broader context of e-commerce growth and fraud challenges, as discussed in reports by the Asosiasi Penyelenggara Jasa Internet Indonesia [6][7], e-commerce business insights from Laudon and Traver [8], and retail e-commerce revenue forecasts from Statista [9]. These sources indicate that fraud detection is critical as e-commerce continues to grow, with various strategies being explored, such as those mentioned in Renjith's work on Support Vector Machines [10], Roy's deep learning-based fraud detection [11], and Zhao's studies on implicit behavioral evidence [12][13].

3. OVERVIEW OF THE SYSTEM

3.1 Existing System

Fraud detection in e-commerce has been an area of growing concern due to the increasing volume of online transactions and the corresponding rise in fraudulent activities. While fraud detection in credit card transactions has seen significant development using machine learning algorithms such as decision trees, naïve Bayes, neural networks, and random forests, the focus on fraud detection in e-commerce transactions has been relatively limited. Research in this area has primarily focused on identifying features or attributes that can distinguish between fraudulent and non-fraudulent transactions. However, comprehensive studies and robust systems for fraud detection specifically tailored for e-commerce transactions are lacking. Despite the potential of machine learning algorithms in addressing this challenge, the existing system remains underdeveloped, leaving e-commerce platforms vulnerable to fraudulent activities such as the use of stolen credit cards and money laundering. As such, there is a pressing need for further research and development in fraud detection systems specifically designed for e-commerce transactions to mitigate financial losses and safeguard online transactions for both businesses and consumers.

3.1.1 Disadvantages of Existing System

The existing systems exhibit several shortcomings that can impact their overall effectiveness and security:

Limited Development in E-commerce Fraud Detection: While fraud detection in credit card transactions has seen significant advancements using machine learning algorithms, the focus on e-commerce transactions remains relatively limited. This lack of development in e-commerce-specific fraud detection systems may leave online platforms vulnerable to emerging fraudulent activities unique to e-commerce environments.

Incomplete Coverage of Fraudulent Activities: The existing systems primarily focus on identifying features or attributes to distinguish between fraudulent and non-fraudulent transactions. However, this approach may not adequately cover the diverse range of fraudulent activities that occur in e-commerce, such as account takeover, identity theft, fake reviews, and transaction laundering. Consequently, the existing systems may overlook certain types of fraud, leaving e-commerce platforms susceptible to exploitation.

Potential Inaccuracy and False Positives: The reliance on traditional machine learning algorithms like decision trees, naïve Bayes, and random forests may lead to inaccuracies and false positives in fraud detection. These algorithms may struggle to adapt to the dynamic and evolving nature of fraud patterns in e-commerce, resulting in misclassifications and erroneous fraud alerts. This could lead to inconvenience for legitimate customers and undermine the effectiveness of the fraud detection system.

Scalability and Performance Challenges: As the volume of online transactions continues to increase, existing fraud detection systems may face scalability and performance challenges. Traditional machine learning algorithms may struggle to process large datasets efficiently, leading to delays in fraud detection and response times. Additionally, maintaining the accuracy and reliability of the system as transaction volumes grow can be a daunting task.

Limited Integration of Deep Learning: While there has been rapid development in fraud detection using deep learning for credit card transactions, the integration of deep learning techniques into e-commerce fraud detection systems remains limited. Deep learning approaches have the potential to uncover complex patterns and anomalies in transaction data, but their adoption in e-commerce fraud detection research is still in its infancy.

Addressing these disadvantages will be crucial for advancing fraud detection capabilities in e-commerce and effectively mitigating the risks associated with online transactions.

3.2 Proposed System

The proposed system aims to leverage machine learning techniques for detecting fraudulent activity in e-commerce, focusing specifically on predicting the likelihood of a new user's first transaction being fraudulent. Company XYZ, an e-commerce site specializing in handmade clothing, requires a model to identify users who may engage in illegal activities based solely on their initial transaction. The system will analyze transaction data to perform binary classification (fraud or no fraud) for each user. Various machine learning algorithms will be employed and compared to identify the most accurate model for fraud detection in e-commerce transactions. By implementing this system, online shopping websites, merchants, and customers can benefit from enhanced security measures and reduced risk of fraudulent transactions, thereby fostering trust and confidence in online transactions.

3.2.1 Advantages of Proposed System

Improved Security Measures: By leveraging machine learning techniques to detect fraudulent activity, the proposed system enhances the security measures of e-commerce platforms. It helps identify potentially fraudulent transactions, thereby reducing the risk of financial losses and protecting both merchants and customers from fraudulent activities.

Early Detection of Fraudulent Behavior: The system focuses on predicting the likelihood of fraud in a user's first transaction, allowing for early detection of potentially fraudulent behavior. This proactive approach enables e-commerce platforms to take preventive measures to mitigate the impact of fraudulent transactions before they occur.

Customized Fraud Detection Model: The system enables the customization of fraud detection models based on the specific needs and characteristics of the e-commerce platform. By comparing different machine learning algorithms, the system identifies the most accurate model tailored to the transaction dataset of the e-commerce platform, thereby improving the effectiveness of fraud detection.

Enhanced Accuracy and Efficiency: By employing advanced machine learning algorithms, the proposed system enhances the accuracy and efficiency of fraud detection in e-commerce transactions. It analyzes transaction data to identify patterns and anomalies associated with fraudulent behavior, enabling more accurate classification of transactions as either fraudulent or legitimate.

Cost Reduction and Resource Optimization: By automating the fraud detection process through machine learning techniques, the proposed system reduces the need for manual intervention and human resources dedicated to fraud prevention. This leads to cost savings for e-commerce platforms and optimizes resource allocation for other business operations.

Increased Trust and Confidence: Implementing an effective fraud detection system instills trust and confidence among customers, encouraging them to engage in online transactions without fear of fraudulent activities. By protecting the integrity of the e-commerce platform, the proposed system strengthens the relationship between merchants and customers, fostering long-term loyalty and satisfaction.

Overall, the proposed system offers several advantages, including improved security measures, early detection of fraudulent behavior, customized fraud detection models, enhanced accuracy and efficiency, cost reduction, resource optimization, and increased trust and confidence among customers. These advantages contribute to the overall success and sustainability of e-commerce platforms in today's digital landscape.

3.3 Proposed System Design

Sure, here's the proposed system design for fraud prevention in e-commerce using machine learning algorithms such as Decision Tree, Naive Bayes, Random Forest, and ensemble methods:

1. Data Gathering

Objective: Collect relevant data related to e-commerce transactions, including user profiles, transaction history, and other relevant metadata.

Functions:

Gather transaction data from e-commerce platforms, including user information, purchase details, payment methods, and transaction timestamps.Acquire additional data sources such as IP addresses, device information, and geographical locations for enhanced fraud detection.Aggregate historical transaction data and label fraudulent transactions for supervised learning.

2. Data Preprocessing & Enhancement

Objective: Clean, preprocess, and enhance the collected data to prepare it for machine learning model training.

Functions:

Remove missing values, outliers, and irrelevant features from the dataset.Standardize or normalize numerical features to ensure consistency.Encode categorical variables using techniques like one-hot encoding or label encoding.Perform feature engineering to create new features or extract relevant information from existing ones.Split the dataset into training, validation, and testing sets for model evaluation.

3. Model Training & Evaluation

Objective: Train machine learning models using various algorithms including Decision Tree, Naive Bayes, Random Forest, and ensemble methods to detect fraudulent transactions.

Functions:

- Implement and train individual machine learning models using libraries like scikit-learn or TensorFlow.
- Evaluate model performance using metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).
- Tune hyperparameters of each model using techniques like grid search or random search to optimize performance.
- Compare the performance of different algorithms and select the best-performing model(s) for deployment.

4. Model Deployment & Integration

Objective: Deploy trained machine learning models into production environments for real-time fraud detection in e-commerce transactions.

Functions:

- Integrate deployed models with e-commerce platforms' backend systems to analyze transactions in real-time.
- Implement API endpoints or microservices for seamless communication between the models and the e-commerce platform.
- Monitor model performance and recalibrate as necessary to adapt to evolving fraud patterns.
- Implement feedback mechanisms to continuously update models with new data and improve detection accuracy over time.

5. Reporting & Alerts

Objective: Generate reports and alerts to notify stakeholders of potential fraudulent activities detected by the deployed models.

Functions:

- Generate daily, weekly, or monthly reports summarizing fraudulent transactions detected, false positives, and model performance metrics.
- Send real-time alerts to e-commerce platform administrators or users in case of suspicious activities.
- Implement escalation procedures for handling high-risk transactions and coordinating with law enforcement agencies if necessary.

3.4 Architecture

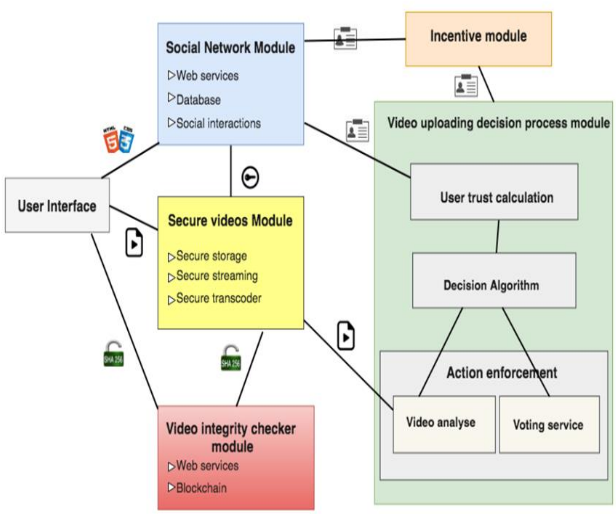

Fig 1: System Architecture

4. RESULT SCREENSHOTS

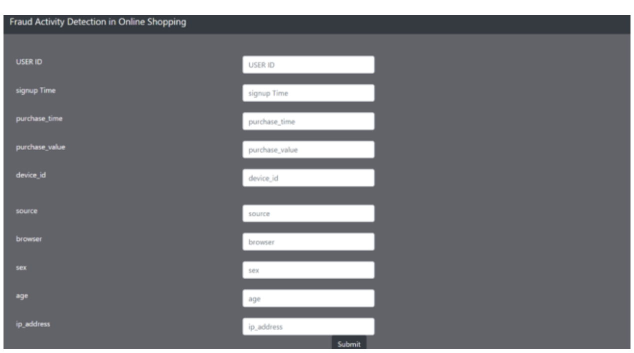

Fig 1 : Upload Fraud e commerce details

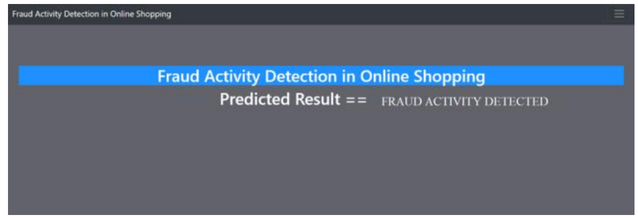

Fig 2 : Fraud activity detected

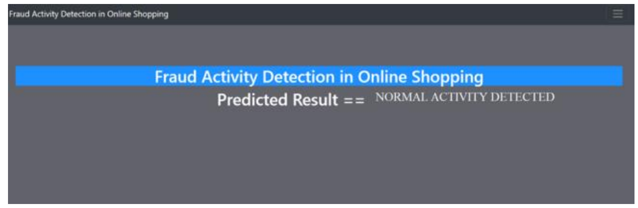

Fig 3 : Normal activity detected

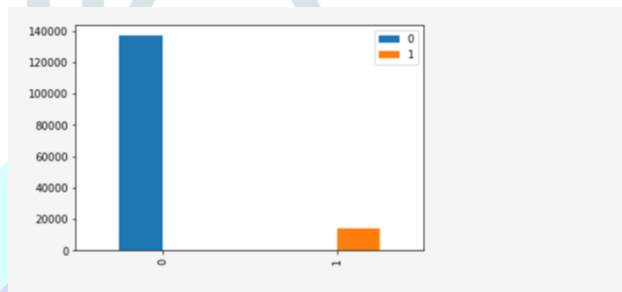

Fig 4 :Labels Analysis

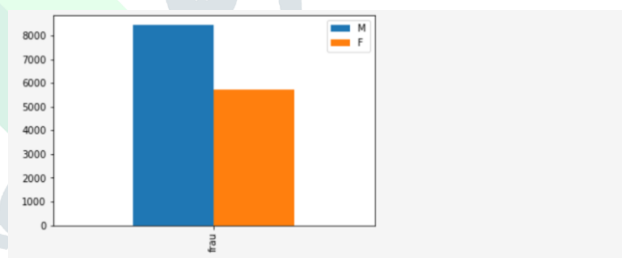

Fig 5: Male Female fraud

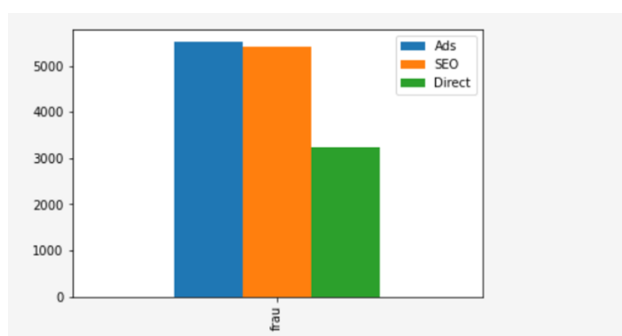

Fig 6 : Type of Ads

Fig 7 : Analysis Of the Output

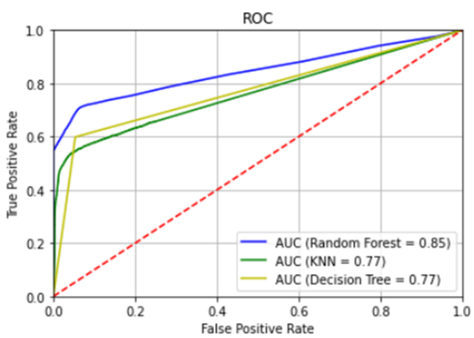

Fig 8 : True Positive and False Positive rate

5. CONCLUSION

In conclusion, our project proposes a comprehensive framework that addresses the pressing challenge of ensuring the secure delivery of trusted video content over Social Media Networks (SMNs). By leveraging collaboration between humans and machines, we aim to enhance trust levels, automate content evaluation processes, and safeguard the integrity and security of video content within SMNs. The key components of our framework include assigning trust levels to users based on their history, deploying intelligent agents to determine content publication, and implementing video integrity checks during streaming. Through these efforts, we seek to increase the trustworthiness of SMNs while also promoting efficiency in capital expenditure. Our vision aligns with the broader goal of fostering certainty, authenticity, and security in data exchanges across social network nodes.

REFERENCES

[1] Asosiasi Penyelenggara Jasa Internet Indonesia," Magazine APJI(Asosiasi Penyelenggara Jasa Internet Indonesia)" (2019): 23 April 2018.

[2] Asosiasi Penyelenggara Jasa Internet Indonesia, "Mengawali integritasera digital 2019 - Magazine APJI(Asosiasi Penyelenggara Jasa Internet Indonesia)" (2019).

[3] Laudon, Kenneth C., and Carol Guercio Traver. E-commerce: business. technology, society. 2016.

[4] statista.com. retail e-commerce revenue forecast from 2017 to 2023 (in billion U.S. dollars). (2018). Retrieved April 2018, from Indonesia:: https://www.statista.com/statistics/280925/e-commerce-revenue- forecast-in-indonesia/.

[5] Renjith, S. Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine Approach. International Journal of Engineering Trends and Technology (2018).

[6] Roy, Abhimanyu, et al. "Deep learning detecting fraud in credit card transactions." 2018 Systems and Information Engineering Design Symposium (SIEDS). IEEE, 2018.

[7] Zhao, Jic, et al. "Extracting and reasoning about implicit behavioral evidences for detecting fraudulent online transactions in e-Commerce." Decision support systems 86 (2016): 109-121.

[8] Zhao, Jic, et al. "Extracting and reasoning about implicit behavioral evidences for detecting fraudulent online transactions in e-Commerce."

Decision support systems 86 (2016): 109-121.

[9] Pumsirirat, Apapan, and Liu Yan. "Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine." International Joumal of advanced computer science and applications 9.1 (2018): 18-25 Focus