iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60307

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Fraud Detection and Prevention in E-commerce using Decision Tree Algorithm

Prof. K. R. Hole[1], Tejas Tidke[2], Miheer Thakur[3], Vidhan Thakur[4], Abhijit Ingole[5]
[1, 2, 3, 4, 5]*Dept.Of Computer Science & Engineering, Prof. Ram Meghe Institute of Technology & Research, Amravati, Maharashtra, India*

Abstract: ***Fraud detection is an important part of e-commerce because it helps prevent fraud such as illegal transactions, identity theft, and money laundering. Recently, there has been a lot of literature on the application of machine learning algorithms to identify e-commerce fraud. These algorithms work by learning patterns in data that indicate fraud. Pattern checking deals with discovering differences in data, such as unusual products, locations, or behavior outside the norm for certain users, through machine learning. In this project, we propose a decision tree algorithm to detect fraud in e-commerce using newly generated data from various online products on e-commerce sites. In addition to fraud detection, we also provide advice on fraud prevention. We propose a new security model that will prove the user's identity. In this security model, users are required to register their profile with some questions. Our security systems will display relevant images in response to the registration question. The user has to click on the correct answer image within the time limit. We will ask the user 3 questions in graphic format. If the user selects the correct answer, the user will be considered a real user.***
Keywords: fraud, detection, prevention, decision tree, e-commerce, transactions, graphical security.

I. INTRODUCTION

E-commerce fraud is any dishonest or fraudulent behavior carried out by people or organizations stealing financial or personal information, manipulating e-commerce systems for financial advantage, or carrying out illegal transactions. Identity theft, phishing, chargeback fraud, affiliate fraud, and deceptive advertising are a few instances of e-commerce fraud that frequently occur. In order to safeguard clients and avoid financial losses, fraud detection is an essential component of e-commerce transactions [1]. Transaction tracking, IP address geolocation, and device fingerprinting are a few methods that can be used to identify fraud in e-commerce transactions. As technology advances, machine learning algorithms can be used to examine transaction data and spot trends that point to possible fraud.

Fraud prevention and fraud detection systems are the two primary defenses against frauds and losses brought on by fraudulent activity. The proactive method of stopping fraud from happening is called fraud prevention. When criminals manage to get past fraud protection systems and initiate a fraudulent transaction, fraud detection systems are activated. It is impossible for anyone to determine whether a fraudulent transaction has gotten past the security measures. Therefore, regardless of the preventative methods in place, the aim of the fraud detection systems is to examine every transaction for the potential of being fraudulent and to promptly identify those that are once the fraudster has started to commit fraud. A summary of fraud detection is available in [2–5].

Fraudulent transactions in credit card and e-commerce systems, money laundering in financial systems, computer system intrusions, fraudulent calls or service usages in telecommunication systems, and fraudulent claims in health and auto insurance systems are the most well-known types of fraud. Due to advancements in communication channels and information technology, fraud is becoming more widespread globally, resulting in significant financial losses. Even while fraud can be committed using a variety of media, the most common ones are online ones like the Internet and mail, wire, and phone.There is a sharp increase in the number of fraudulent transactions conducted online due to the web's global accessibility and the simplicity with which individuals can conceal their identity and location when transacting. Moreover, criminals now have the opportunity to create global fraud networks through information exchange and cooperation due to increases in internetworking channel bandwidth. Due to their nature, online credit card fraud and other similar crimes end up being the most common types conducted via the Internet.

There are two categories of fraud detection techniques: Supervised strategies involve using previously identified fraud or legitimate cases to create a model that generates a suspicion score for new transactions; unsupervised techniques, on the other hand, do not have any prior sets that indicate whether a transaction is fraudulent or genuine. A succinct overview of both supervised and unsupervised methods is provided in [2]. Numerous methods, such as artificial neural networks, are applicable in both supervised and unsupervised settings.

The most widely used fraud detection approaches include logistic regression, neural networks, decision trees, SVMs (support vector machines), rule-induction techniques, and meta-heuristics like closest neighbor, k-means clustering, genetic algorithms, and genetic algorithms. These methods can be applied independently or in combination with ensemble or meta-learning techniques to construct classifiers.In this project we proposed behavioral activities based fraud detection systems for ecommerce applications. In this system we will track users' click through activities, login attempts, client side machine unique mac id to find out users behavior. The collected input will be sent to the Decision tree algorithm to find out whether the behavior is suspicious or not. If a user's behavior is found suspicious, the user will have to prove his authentication by using a graphical password. If the user provides authentication, that means the user is a genuine user otherwise the account will be automatically locked and notification will be sent on users email id.

A. Decision Tree

One of the most popular machine learning methods for applications involving regression and classification is the decision tree. It is an algorithm for supervised learning that creates a model of decisions and their potential outcomes that resembles a tree. The decision or result is represented by the nodes in the tree structure, and the potential outcomes are represented by the edges.

B. Fraud Detection Attributes

1) Client side machine authentication
2) Login attempts
3) Click through activities
4) Shopping behavior

II. LITERATURE SURVEY

Fraud detection, which has developed very rapidly, is fraud on credit cards. Many studies discuss the fraud method. An auto-encoder and constrained Boltzmann machine study is one that uses deep learning [9]. Deep learning is used to build a fraud detection model that runs like a human neural network, where data will be made in several layers that are tailored for the process, starting from the encoder at layer 1 and the and the hinge decoder at layer 4. The researcher compares the deep learning method with other algorithms, such as the Hidden Markov Model (HMM) [10].

Machine learning was employed in credit card fraud detection studies as well [11], including neural networks, random forests, naïve Bayes, and decision tree algorithms among others. Because decision trees are so simple to use, they are one method that is frequently employed in fraud detection. A decision tree is a prediction model that makes use of hierarchical or tree structures. Because Naïve Bayes is a classification using statistical methods and probability, it is utilized in fraud detection credit cards. In real-world scenarios, Naïve Bayes is incredibly quick yet has a significant degree of error. Genetic algorithms are used by neural networks on fraud detection credit cards to ascertain the number of hidden layer structures [12]. By using genetic algorithms, the neural network generates the maximum number of hidden layers that is ideal [13]. Additionally, random forest is used in credit card fraud detection [14].

To create a single model, Random Forest combines all of the best tree combinations. According to Random Forest, every decision tree has a maximum depth and is based on a random vector value with the same distribution across all trees [13]. Up until now, not much has been studied about e-commerce fraud detection. Only the identification of characteristics or traits that will be utilized to distinguish between fraudulent and non-fraudulent transactions is covered by research on fraud detection in e-commerce [14]. The behavior in e-commerce transactions is determined by an extraction attribute/feature technique that is described in the paper. In online business, this characteristic is used to detect fraud. The parameters for transactions are established by this attribute. Another study on e-commerce fraud detection examines reasoned transactions that rely on the characteristics of e-commerce transactions. Features of the transaction, such as invalid rating, confirmation interval, average stay duration on commodities, and buyer features, such as genuine name, positive rating ratio, and transaction frequency, are the features/attributes that are employed. Inadequate categorization outcomes arise from data imbalance.

The dataset in the article has 151,112 records in total, of which 14,151 records are categorized as fraudulent. The percentage of fraud data is 0.093 percent. One technique for balancing data is the Synthetic Minority Oversampling Technique (SMOTE) [15]. This oversampling technique works by raising the number of positive classes through random data replication, ensuring that the amount of positive data is equal to the amount of negative data. Replicating data in a small class is one technique to use synthetic data.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

In order to build duplicate synthetic data as much as the needed percentage between randomly and positively picked k classes, the SMOTE algorithm first locates the k nearest neighbor for a positive class. A recent fraud detection research restricted its scope to feature or attribute determination. Machine learning is applied in e-commerce to improve fraud detection. Neural networks, Random Forests, Decision Trees, and Naïve Bayes are the machine learning techniques used.

III. PROPOSED SYSTEM

A. Proposed Concepts

In this project we proposed an e-commerce application in which users will register personal details, some security questions and answers. In this application we will use the amazon product dataset. When a user logged in into the system, our tracking model will track user activities, and shopping behavior and when a user tries to place an order, our fraud detection model will check all the activities. If suspicious behavior is predicted, the user has to prove his identity using a security model and if the user is proved to be authentic, the order will be placed and transaction will be considered as normal otherwise the account will be blocked and notification will be sent to the user.

B. System Architecture/Design

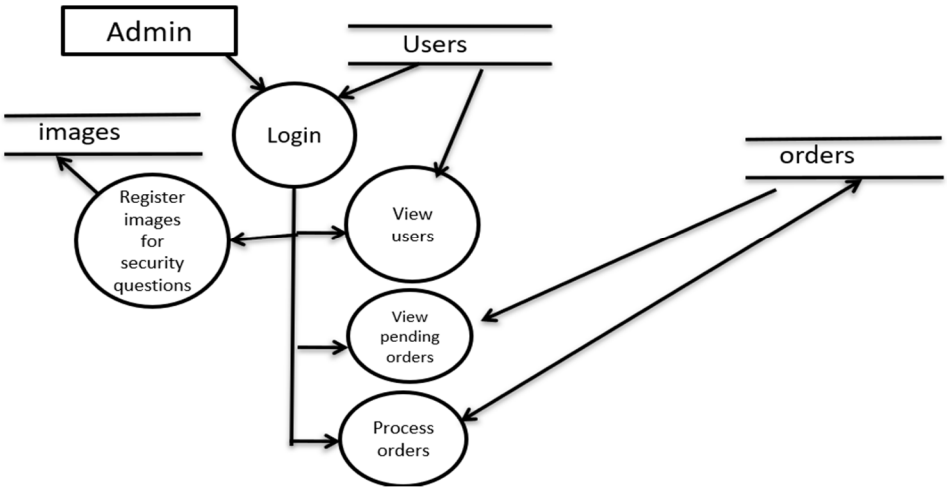


Fig 1. Admin Panel

1) Admin Panel

The admin panel is a secure web interface designed for administrators to manage and oversee the e-commerce platform. It provides access to various functionalities and tools necessary for monitoring, analyzing, and controlling the website's operations. In the context of fraud detection and prevention, the admin panel may include features such as viewing user data, managing orders, accessing fraud detection reports, and configuring fraud prevention settings.

a) *Login:* The login functionality allows administrators to securely access the admin panel by entering their credentials (username and password). It is essential for maintaining the security and confidentiality of the admin panel, ensuring that only authorized personnel can access sensitive information and perform administrative tasks.

b) *Register Images for Security Questions:* This feature enables administrators to register the images for security purposes, which can be helpful for fraud detection and prevention. Administrators can check and compare the selected images by users as answers for the registered images as questions to identify the suspicious or fraudulent accounts.

c) *View users:* This feature enables administrators to view a list of registered users on the e-commerce website. It provides valuable insights into user demographics, behaviors, and activities, which can be helpful for fraud detection and prevention purposes. Administrators can use this information to identify suspicious or fraudulent accounts and take appropriate actions to mitigate risks.

d) *View Pending Orders:* The view pending orders feature allows administrators to see a list of orders that are awaiting processing or approval. It provides visibility into the current status of orders, helping administrators to prioritize their workflow and ensure timely fulfillment. In the context of fraud detection, this feature can be used to identify suspicious orders that require further investigation.

e) Process Orders: The process orders functionality enables administrators to manage and fulfill customer orders. It includes actions such as updating order status, tracking shipment, and generating invoices. In the context of fraud prevention, administrators can use this functionality to verify orders, flag suspicious transactions, and take preventive measures to protect against fraudulent activities.

2) User Management

This feature is crucial for tracking user behavior and identifying patterns that may indicate fraudulent activity. Managing user accounts also helps in keeping track of user activities and detecting any suspicious behavior.

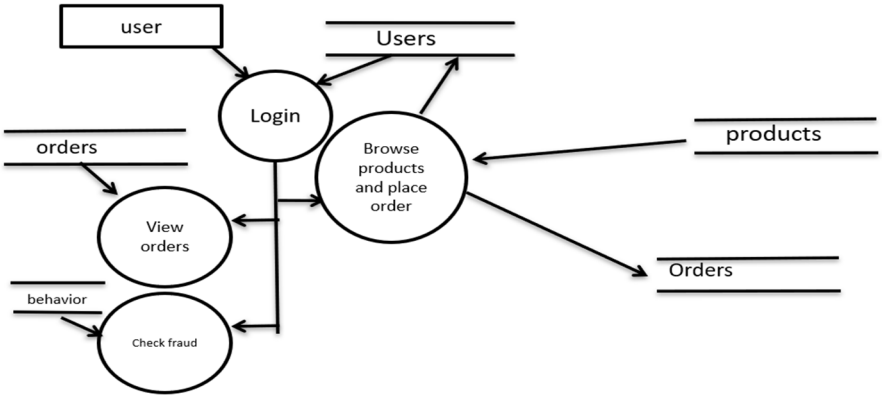


Fig 2. User Panel

a) Login: The login feature allows you to track user activity and behavior, which is important for detecting any unusual or suspicious login patterns that may indicate fraud.

b) Browse Products: While browsing products, users may exhibit certain behaviors that can indicate fraudulent activity, such as repeatedly viewing high-value items without making a purchase.

c) Place Order: Monitoring the order placement process can help in detecting fraudulent transactions, such as orders placed using stolen credit card information.

d) View own Orders: Allowing users to view their own orders helps in detecting any unauthorized transactions or suspicious activity associated with their account.

e) Check Fraud: By analyzing users' activities, such as browsing behavior, order history, and transaction patterns, you can detect any unusual or suspicious behavior that may indicate fraud.

C. Working of Proposed System

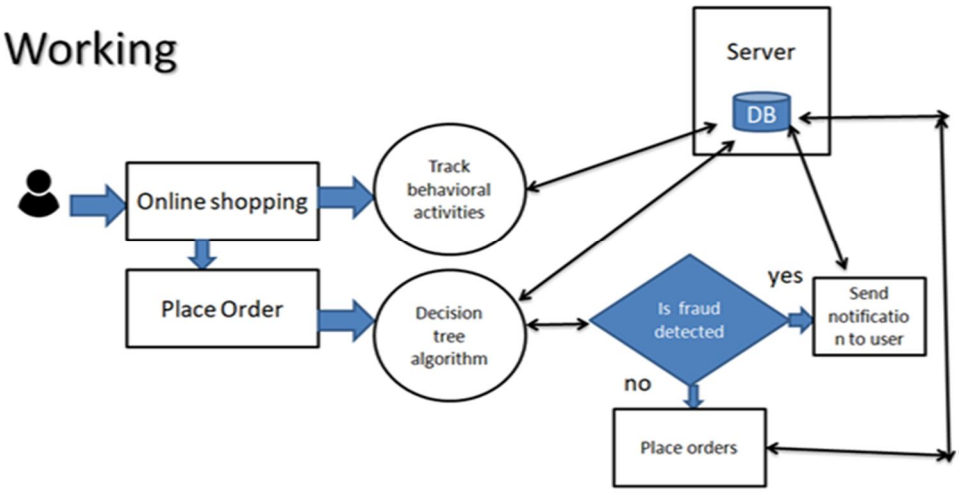


Fig 3. Working of System

1) Fraud Detection

The system will be designed to detect fraudulent activities on e-commerce websites. This includes analyzing user behavior and transaction patterns to identify potential fraud.

a) *Tracking User Activities After Login:* The system will track user activities after they log in to the e-commerce website. This data will be used to analyze user behavior and detect any suspicious activity.

b) *Applying Decision Tree Algorithm:* The decision tree algorithm will be applied to the current users' activities at the time of order placing. This algorithm will help in making decisions about whether a particular user's activity is fraudulent or not.

c) *Notification System:* If fraud is detected, the system will send a notification to the user's email ID. This notification will alert the user about the suspicious activity and prompt them to verify their identity.

d) *Verification of Identity:* If the user verifies their identity and it is proven that they are an authorized user, then the system will allow the payment 5to proceed. This step ensures that genuine users are not inconvenienced by false fraud alerts.

e) *Automatic Account Lock:* If the user fails to verify their identity and it is determined that the activity is fraudulent, the system will automatically lock the user's account. This is done to prevent any further unauthorized transactions.

f) *Unlocking Account:* Genuine users who have had their accounts locked can unlock them by approving their identity. This step ensures that legitimate users can regain access to their accounts after resolving any security issues.

2) Authentication Verification

This is the process of confirming the identity of a user before allowing access to the e-commerce website. It involves validating the user's credentials, such as username and password, to ensure they are legitimate.

a) *Graphical Password Authentication:* In case fraud is detected, the system prompts the user to prove their authentication using graphical passwords. Instead of traditional text-based passwords, graphical passwords use images or patterns selected by the user to authenticate their identity.

b) *Profile Questions and Answers:* The graphical password authentication is based on profile questions and answers that the user has previously registered. These questions could be related to personal information, preferences, or any other details that are unique to the user.

c) *Image Selection:* The system presents a set of images to the user, and the user must click on the image that corresponds to their registered answer for the profile question. This adds an additional layer of security by requiring the user to recall and select the correct image among several options.

D. Block Diagram of System

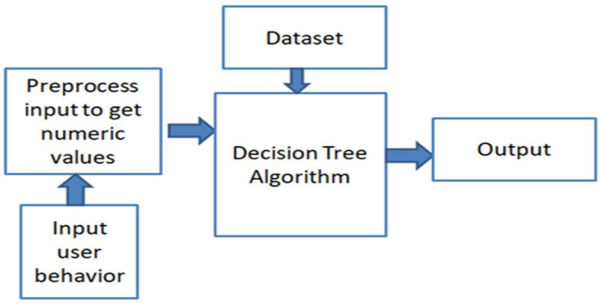


Fig 4. Block Diagram of System

1) Input - User Behavior: This refers to the data about a new transaction that you want to classify as fraudulent or legitimate. The decision tree will use this data to navigate the tree from the root node to a leaf node, based on the splitting conditions.

2) Preprocess: Before using the data to train the decision tree, it needs to be preprocessed. This may involve cleaning the data (e.g., removing missing values, formatting inconsistencies), and transforming it into a format suitable for the algorithm (e.g., converting categorical data into numerical data).

3) Dataset: This refers to the collection of data used to train the decision tree. In your case, the data would likely include information about past transactions, both fraudulent and legitimate.

4) *Decision Tree Algorithm:* Once the data is prepared, it is fed into the decision tree algorithm. The algorithm works by splitting the data into subsets based on specific conditions. The final product of the decision tree algorithm is a tree-like structure with decision nodes and leaf nodes. The decision nodes represent the splitting conditions, and the leaf nodes represent the final classification (fraudulent or legitimate).
5) *Output:* The output of the decision tree is a classification of the new transaction as either fraudulent or legitimate.

IV. ALGORITHMS USED

A. *Decision Tree Algorithm*

1) Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
2) In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
3) The decisions or the test are performed on the basis of features of the given dataset.
4) It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
5) It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
6) In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.
7) A decision tree simply asks a question, and based on the answer (Yes/No), it further splits the tree into subtrees.
8) Below diagram explains the general structure of a decision tree:

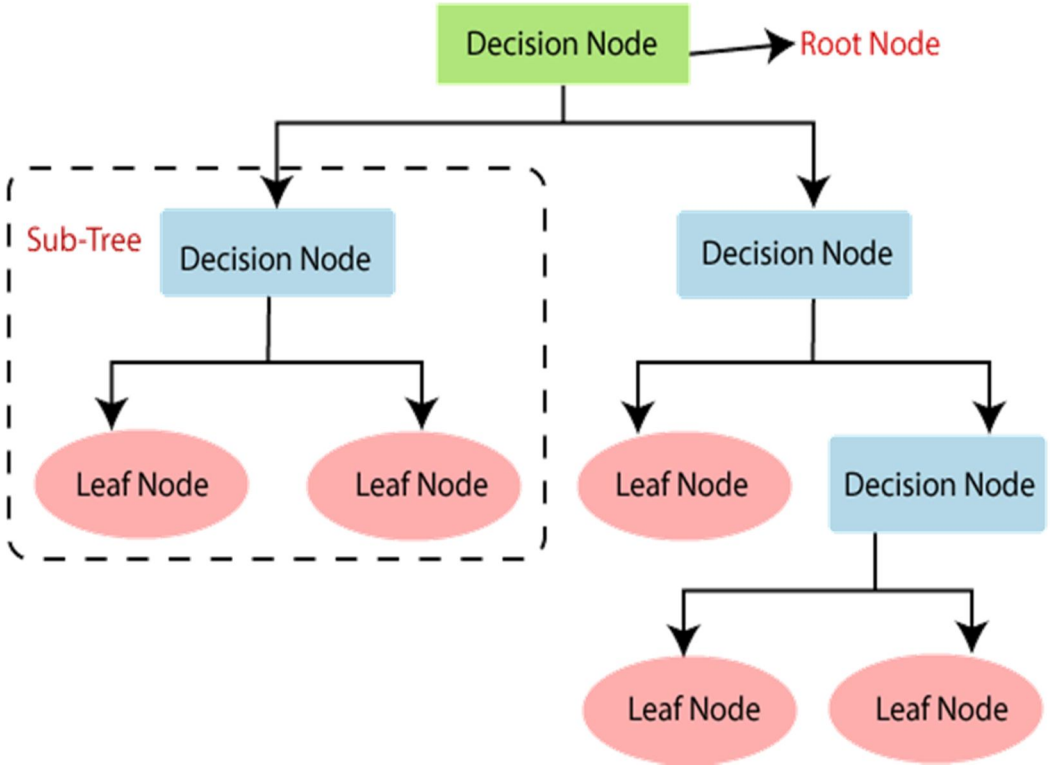


Fig 4. Decision Tree Algorithm

B. *Why use Decision Trees?*

There are many algorithms in machine learning; so choosing the best algorithm for a given data and problem is an important thing to keep in mind when designing machine learning. There are two reasons to use decision trees:

1) Decision trees generally help people think when making decisions, so they are easy to understand.
2) The logic behind the decision tree is easy to understand as it displays a tree-like structure.

C. Decision Tree Terminologies

1) *Root Node:* The root is where the decision tree begins. It represents all data layers that are isomorphically split into two or more.
2) *Leaf Node:* Leaf nodes are the final output nodes, and the tree cannot be segregated further after getting a leaf node.
3) *Splitting:* Splitting is the process of splitting the decision node/root node into child nodes under certain conditions.
4) *Branch/Sub-Tree:* A tree formed by splitting the tree.
5) *Pruning:* Pruning is the process of removing the unwanted branches from a tree.
6) *Parent/Child node:* The root node of the tree is called the parent node, and the other nodes are called the child nodes.

D. How does the Decision Tree algorithm work?

In a decision tree, the algorithm starts from that tree to predict the class of the given data. This algorithm compares the base value with the data (real data) feature and jumps to the next branches based on the comparison.

For the next node, the algorithm again compares the attribute value with the implicit (real data) attribute value of the other child nodes and continues. He continued the process until he reached a leaf on the tree. The entire process can be better understood using the following algorithm:

1) *Step 1:* Begin the tree with the root node, says S, which contains the complete dataset.
2) *Step 2:* Find the best attribute in the dataset using Attribute Selection Measure (ASM).
3) *Step 3:* Divide the S into subsets that contain possible values for the best attributes.
4) *Step 4:* Generate the decision tree node, which contains the best attribute.
5) *Step 5:* Recursively make new decision trees using the subsets of the dataset created in step-3. Continue this process until a stage is reached where you cannot further classify the nodes and call the final node as a leaf node.

E. Decision Tree Dataset

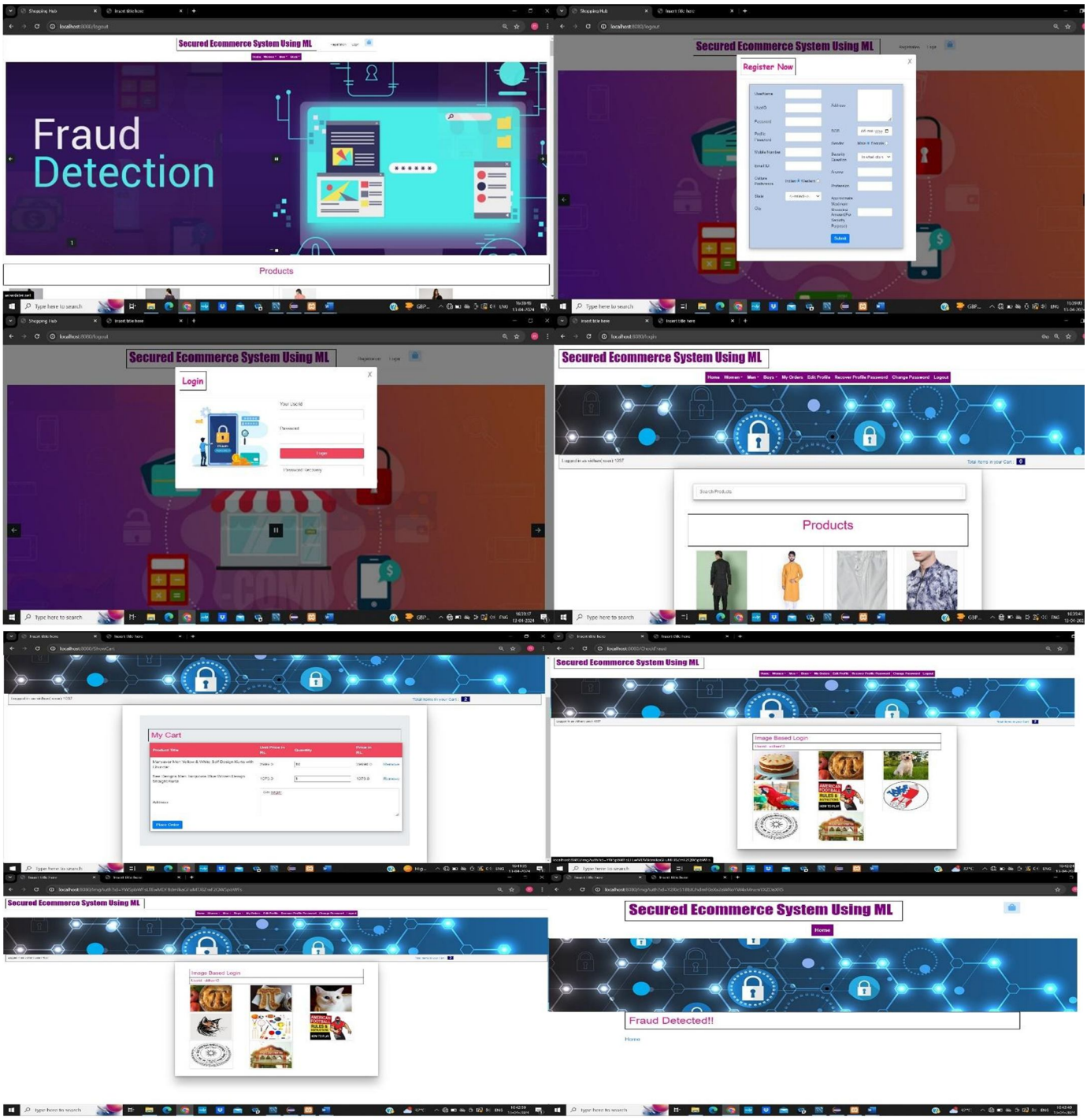
login_attempts	login_time	editProfile	ProfileAuth	ProfileAuthTime	PassRecovery	product_range	amt	addrchanged	email_changed	shopamt	label
1	1	1	0	1	0	0	1	1	0	0	1
0	0	1	0	1	0	1	1	1	0	1	1
0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	1	0
0	0	1	1	1	1	0	0	1	0	1	1
1	1	1	1	1	0	0	0	1	1	0	1
0	0	0	0	0	0	1	1	0	0	0	1
0	0	0	0	0	0	0	0	1	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	1	1	1	1	1	1	1	1
0	1	0	1	1	1	1	0	0	0	0	1
0	0	1	1	1	1	1	1	1	1	0	1
0	0	0	0	0	0	1	0	0	0	0	1
0	0	0	0	0	0	0	1	0	0	0	1
0	0	1	0	0	0	0	0	0	0	1	0
0	0	1	1	0	1	0	0	0	0	0	1
1	1	1	0	1	0	0	1	1	0	0	1

F. Dataset Attributes

1) login_attempts
2) login_time
3) editProfile
4) ProfileAuth
5) ProfileAuthTime
6) PassRecovery
7) product_range
8) amt
9) addrchanged
10) email_changed
11) shopamt

V. RESULT ANALYSIS

Fraud Detection and Prevention in e-commerce using decision tree algorithm has been shown to be effective in combating fraud in online stores. Using decision tree algorithms, the system effectively detects suspicious patterns and anomalies in data changes, thus ensuring timely intervention and fraud prevention. Analysis of the results showed that fraudulent transactions decreased, thus increasing trust and security for merchants and consumers. Additionally, the decision tree model proves its effectiveness in protecting the business ecosystem from fraud by providing performance metrics that include high performance and accuracy. Overall, the project highlights the importance of using advanced analytics to combat fraud and maintain integrity in online businesses.



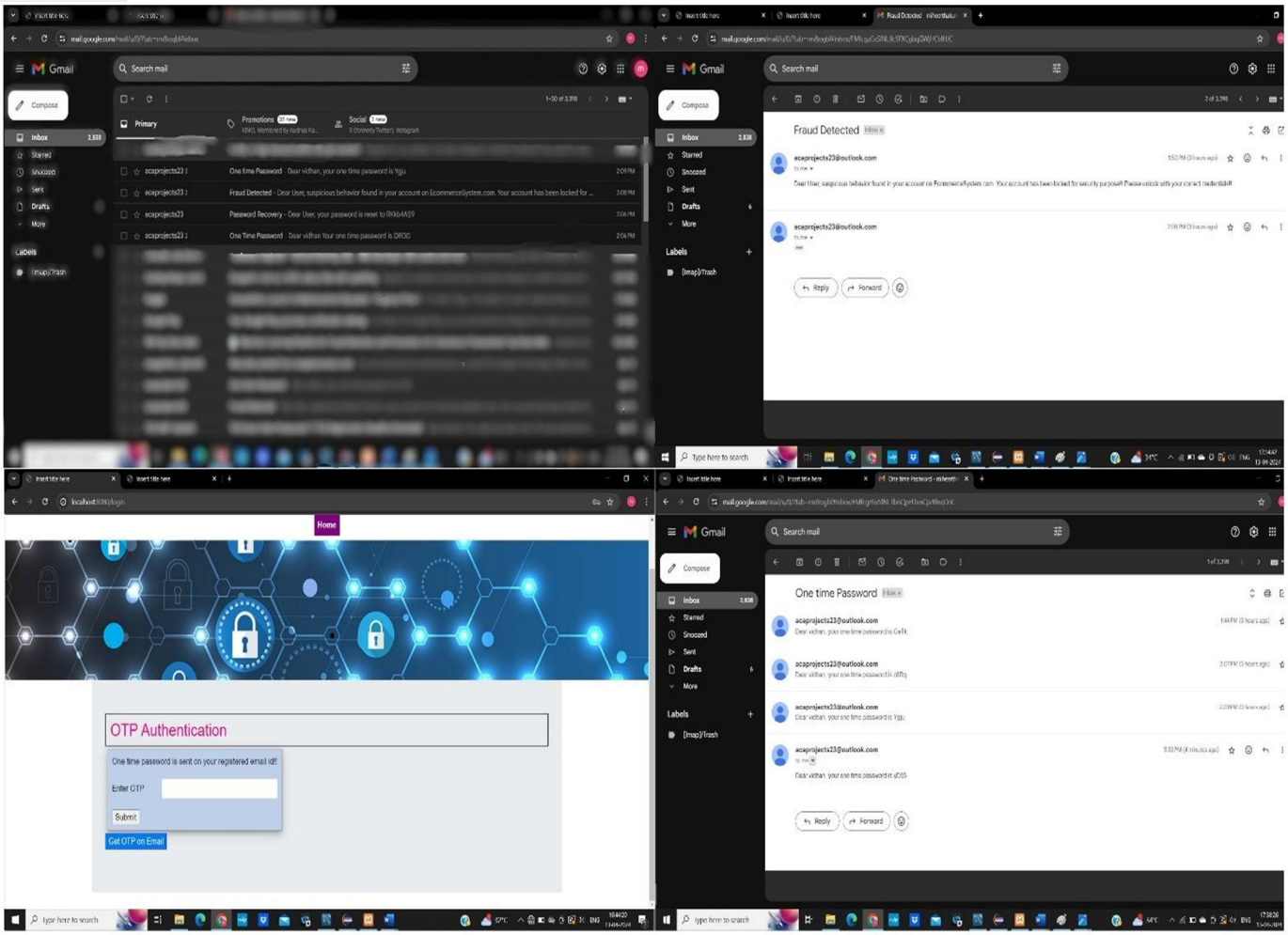


Fig:- Results for Fraud Detection

VI. CONCLUSION

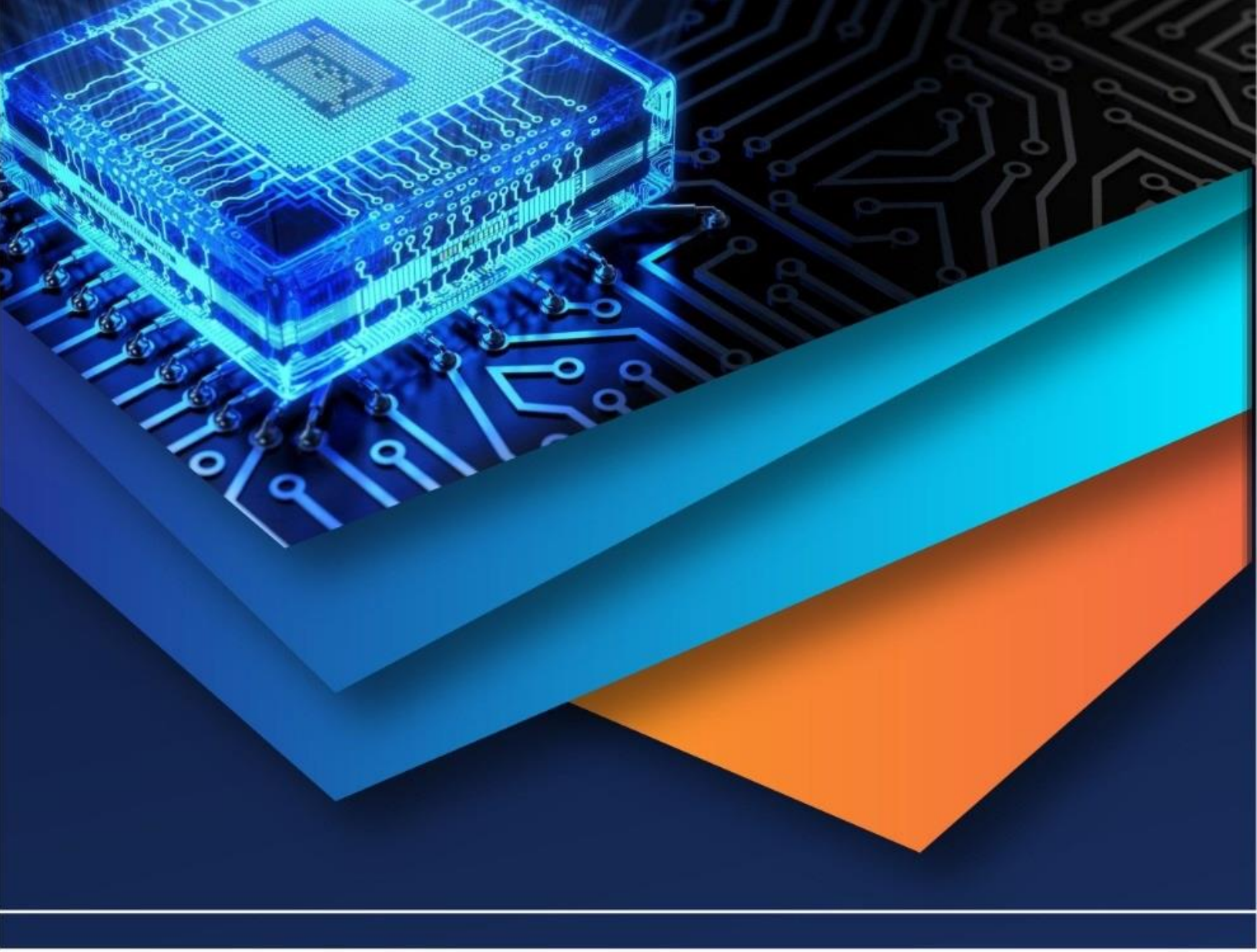
The decision tree algorithm has demonstrated encouraging outcomes when used for fraud prevention and detection on e-commerce platforms. Users are given a safe and convenient way to authenticate themselves with the help of graphical password authentication based on profile questions and answers. In order to identify and stop fraudulent activity, the decision tree algorithm analyzes user behavior and transaction patterns effectively. All things considered, this research shows how crucial it is to use machine learning algorithms to improve the security of e-commerce platforms. Future research could concentrate on enhancing the fraud detection system's precision and effectiveness even more and investigating additional machine learning methods.

VII. FUTURE SCOPE

The future of Fraud Detection and Prevention in e-commerce using decision tree algorithms is promising and versatile. First, the project can leverage the integration of additional machine learning algorithms and advanced analytics to improve the accuracy and efficiency of fraud detection. Search methods such as random forest or gradient boosting can provide better predictive capabilities. Additionally, the integration of real-time data streaming and processing capabilities allows the system to quickly adapt to changing fraud patterns and trends. Additionally, the integration of blockchain technology for immutable data transfer can provide additional security and transparency. Collaborating with financial institutions and regulators can help improve fraud detection processes and industry best practices. Finally, expanding the solution beyond e-commerce to other industries such as banking, insurance, and healthcare can create new opportunities to create current and relevant solutions. Overall, the future of the project is broad and will offer opportunities for continued development and expansion towards the changing anti-fraud of the digital age.

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 integritas era 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-revenueforecast-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, Jie, et al. "Extracting and reasoning about implicit behavioral evidence for detecting fraudulent online transactions in e-Commerce." Decision support systems 86 (2016): 109-121.
[8] Pumsirirat, Apapan, and Liu Yan. "Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine." International Journal of advanced computer science and applications 9.1 (2018): 18-25.
[9] Srivastava, Abhinav, et al. "Credit card fraud detection using hidden Markov model." IEEE Transactions on dependable and secure computing 5.1 (2008): 37-48.
[10] Lakshmi, S. V. S. S., and S. D. Kavilla. "Machine Learning For Credit Card Fraud Detection System." International Journal of Applied Engineering Research 13.24 (2018): 16819-16824.
[11] Aljarah, Ibrahim, Hossam Faris, and Seyedali Mirjalili. "Optimizing connection weights in neural networks using the whale optimization algorithm." Soft Computing 22.1 (2018): 1-15.
[12] Bouktif, Salah, et al. "Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches." Energies 11.7 (2018): 1636.
[13] Xuan, Shiyang, Guanjun Liu, and Zhenchuan Li. "Refined weighted random forest and its application to credit card fraud detection." International Conference on Computational Social Networks. Springer, Cham, 2018.
[14] Hong, Haoyuan, et al. "Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China)." Catena 163 (2018): 399-413.
[15] Sharma, Shiven, et al. "Synthetic oversampling with the majority class: A new perspective on handling extreme imbalance." 2018 IEEE International Conference on Data Mining (ICDM). IEEE, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY