Spotting a Phony Attack by Concealing the Deception of the Web

Dr. M Venkateswara Rao[1], Anjaiah Harshitha Naik[2], Budida Vineetha[3], Chiliveru Vishnu Priya[4]

[1]Associate Professor, Dept of Computer Science and Engineering, Vignana Bharathi Institute of Technology, Aushapur, Hyderabad

[2]UG scholars, Dept of Computer Science and Engineering, Vignana Bharathi Institute of Technology, Aushapur, Hyderabad

[3]UG scholars, Dept of Computer Science and Engineering, Vignana Bharathi Institute of Technology, Aushapur, Hyderabad

[4]UG scholars, Dept of Computer Science and Engineering, Vignana Bharathi Institute of Technology, Aushapur, Hyderabad

Abstract: In the current digital environment, phishing assaults pose a serious risk since malevolent actors are always coming up with new strategies to trick people. Artificial intelligence and machine learning are examples of cutting-edge technology that can be used to identify fake URLs and evaluate web content in order to combat this. Sentiment analysis tools are useful for detecting fraudulent reviews and unethical behavior. Blockchain technology can be used to confirm the legitimacy of products and stop the sale of fake items. The fraud payment can be predicted using logistic regression and random forest. Smooth front-end and back-end system integration facilitates real-time monitoring and analysis, allowing for prompt detection and handling of fraudulent assaults within a single web page. We can more effectively identify and counteract false assaults by leveraging cutting-edge technology, blockchain, and sentiment analysis on the internet, eventually safeguarding users and promoting a safer digital ecosystem.

Keywords: Artificial intelligence, Analysis, Blockchain, fraudulent, Machine learning.

1.Introduction

The internet and social media have revolutionized our lives, but they have also made us vulnerable to fraud. To tackle this issue, we can use request-and-response API models with machine learning to detect fraudulent URLs. Sentiment analysis can help identify deceptive practices through online reviews. Blockchain technology provides a secure platform for product provenance. Additionally, implementing robust payment systems with random forest algorithms can prevent fraudulent transactions. The key is to seamlessly integrate front-end and back-end systems with the help of UI Path. These measures aim to protect consumers from fraud online.

Because using different websites and clicking different links led to the leakage of sensitive data on the internet, in today's digital age, when most sensitive and personal data is digitally kept on mobile devices, the necessity of cybersecurity measures like malicious URL detection has grown dramatically. Phishing attempts are among the most prevalent kinds of computer attacks; they trick users into giving sensitive information to hackers and attackers, resulting in significant losses of both money and data [1]. BERT uses the deep convolutional neural network method for phishing URL identification, the NLP algorithm for meaningful text features, and the extracted URL text from the Phishing Site Predict dataset [2]. Not only that but using machine learning technology with a Random Forest Classifier to detect phishing websites by evaluating authentic and phishing URLs. The goal is to identify the best algorithm based on accuracy, false positive, and false negative rates [3], [4], and [5].

As the internet grows, everything is available on the internet as a daily necessity: groceries, home appliances, electronic gadgets, medicines, food, and whatnot. It was like, "You name it, it has it." So, as things were readily available on your doorstep with one click, the route for fraudulent products was also increasing. Introducing a machine-learning-based anti-counterfeiting solution for detecting fake products. It suggests deeper research for training data, annotation, and labeling services. [6], using artificial intelligence focuses on detecting logos, both in the image and textual representation [7], and blockchain technology scans a product's QR code, allowing consumers to verify product authenticity [8], [9], [10].

Reviews are crucial for customers when making decisions about services or products. They provide authentic feedback about positive or negative services, making them credible sources of information. Misleading or inauthentic content in reviews can lead to customers making decisions generated using the Amazon Mechanical Turk (AMT) crowdsourcing tool [15]. Making use of machine learning approaches to identify fake reviews, focusing on feature extraction and reviewer behaviors [11], [14], BSTC uses pre-trained language models and convolutional neural networks to identify reviews with rich opinions and sentiment expressions [12], using a crawled Amazon China dataset, and extracts product reviews into a temporal feature vector, which gives a way for developing an isolation forest algorithm to detect outlier reviews [13].

To make payments for all the activities. Online payments have surged in popularity due to the ease of sending money from anywhere, and the pandemic has exacerbated this trend. However, the risk of online payment fraud has also increased, necessitating increased awareness for both consumers and service providers. For the sake of detection, emphasize feature selection techniques and machine learning algorithms like logistic regression and random forest, pair-wise matching, decision trees, clustering techniques, neural networks, and genetic algorithms [20]. It emphasizes the importance of accurate fraud prediction in a good fraud detection system [16], [18], such as the XG Boost-based fraud detection framework, which achieves the best results in standard classification measures for mobile payment systems, considering the financial consequences [17]. Artificial intelligence focuses on the increasing popularity of credit and debit cards and their potential for enhancing security in global economic transactions [19]. The rest of this essay is structured as follows: In Section 2, the relevant work is described and a literature survey-style summary of those works is given. The problem statement of this article with the current system and the suggested system is explained in Section 3. The study's techniques and algorithms are covered in Section 4. The final outcome is given in Section 5. The whole document and the study's future scope are concluded in Section 6.

2.Literature Survey

Asadullah Safia et al. (2023) carried out research to examine and assess phishing website detection approaches utilizing the Random Forest Classifier algorithms for the identification of fraudulent attempts. Using the Random Forest algorithm, this reached a maximum accuracy of 99.57%. Thorough analysis, excellent precision, a variety of algorithms, and a critical assessment of phishing detection methods. A comparatively small number of studies were assessed, and deep learning techniques were only partially included [1].

Muna Elsadig et al. (2022) used the CNN algorithm to enhance cybersecurity and detect phishing URLs using BERT characteristic extraction. The recommended method's effectiveness in identifying phishing URLs was 96.66%. Benefits incorporate high accuracy, BERT NLP, and better performance compared to conventional deep learning techniques. Perhaps the negative aspects include lengthy processing times, non-semantic hashing, and a reliance on feature extraction by hand [2].

Machine learning approaches were employed by P.Amba Bhavani et al. (2022) to detect and block phishing websites. whereby XGBoost as CNN LSTM, CNN Bi-LSTM, Random Forest, and Logistic Regression are a few of the ML techniques employed. The accuracy values of the models used in the paper range from 56.36% to 92%. Advantages include superior phishing website detection accuracy through the application of advanced machine learning techniques. Among the disadvantages are potentially false positive and negative results in phishing website recognition [3].

Abad Shayan et al. (2023) made use of machine-learning models such as decision trees (DT), random forests (RF), K-nearest neighbors (KNNs), and support vector machines (SVM) that can classify and block dangerous URLs. The machine-learning models in the research were trained using 85% of the dataset. increased security while using the internet. An effective model training method involves a number of components, including essential feature highlighting, computational inefficiency, sensitivity to data selection, performance variability, model appropriateness, and training time [4].

In order to detect bogus URLs, Mansi Mehndiratta et al. (2023) created a model that makes use of the K-Nearest Neighbors (K-NN) technique. By utilizing the KNN technique, the model achieved 90.61% accuracy, whereas using logistic regression, it reached 75% accuracy. Instantaneous identification, robotic learning, preparation of data, pattern visualization, and dataset reliance are other features [5].

Using machine learning approaches, Eduard Daoud et al. (2020) focused on textual and behavioral elements to identify bogus reviews. SVM, Naive Bayes, Random Forest, K-Nearest Neighbors, and Logistic Regression were the methods they employed. The reviewer behavioral qualities caused CNN to attain the maximum f-score of 82.40% and increase it by 3.80%. This method incorporates reviewer behavioral characteristics and provides increased accuracy and application possibilities. Its shortcomings include bias potential and a tiny dataset, though [6].

A model named BSTC was created by Ms. Reema Anne Roy et al. (2021) and employs pre-trained language models and convolutional neural networks to identify fake reviews. Across several datasets, the model's accuracy ranged from 90.94% to 93.44%, indicating strong performance. The F1 score, precision, and recall of the BSTC model are all enhanced above baseline techniques. However, it might use new pre-trained language models and lacks sentiment analysis tools [7].

An isolation forest technique is proposed by Vaishnavi Hedaoo et al. (2022) to identify fake reviews based on temporal features of product evaluation records. The program outperforms current temporal outlier detection techniques by extracting outlier scores using the isolation forest technique. Additionally, the method's running duration is significantly lower than that of previous systems, indicating that the training and evaluation phases may have been shortened. Better efficiency, higher accuracy, and effective false review detection are some of the advantages. To find the optimal moment to find fake reviews, extra investigation is necessary [8].

Machine learning algorithms were employed by Swaroop Jambhulkar et al. (2022) to identify fraudulent social media reviews. They employed Artificial Neural Network, Random Forest, and Support Vector Machine methods. The Random Forest algorithm achieved a maximum recall of 95.27% and a competency rate of 92.55%, outperforming the other algorithms in accuracy. The tiny dataset and lack of interest in reviewing the time dimension are two of the study's shortcomings [9].

A technique to identify fraudulent reviews was created by Nafisa Anjum et al. (2022) utilizing information-theoretic criteria and supervised learning. They identified fake store reviews using graph-based techniques, and by applying n-gram features to authentic Yelp data, they were able to identify false reviews with 67.8% accuracy. The use of AMT fictional evaluations in the study has certain limitations because it might not be representative of real-world false reviews, which could have an impact on how the results are applied [10].

In order to empower consumers and identify counterfeit goods, Ahmed M. Elmogy et al. (2021) offer a machine learning technique. They recognize brands or certification marks in product images using transfer learning and pre-trained deep learning object recognition models. The accuracy range of the model is 21 to 35, providing excellent precision, affordable implementation, and consumer empowerment. The drawbacks, however, include the scarcity of training data, the arduous annotation process, and the inefficiency of real-time applications [11].

An artificial intelligence-based technique focused on logo detection was proposed by Junwen Lu et al. (2023) to help consumers who are not tech-savvy recognize counterfeit items. This study uses the Naïve Bayes Classifier for text and picture classification and the Deep Learning algorithm (CNN) for logo identification and classification. While the research does not provide an exact result %, the suggested solution using the CNN algorithm and Naïve Bayes classifier can effectively recognize fake product logos and boost testing phase efficiency. The Naïve Bayes classifier and CNN algorithm are AI-based, portable, and easy to use. Cons: unclear metrics, unstated boundaries, inadequate execution strategy, and unpublished experimental results [12].

In order to protect supply chains from counterfeiting and ensure product authenticity, Jingsha He et al. (2019) presented a blockchain-based method for product authentication. Though no specific algorithms are provided, the research generates QR codes using the SHA-256 algorithm. The paper suggests an open, safe system, but it doesn't address implementation or scalability [13].

A blockchain-based solution for supply chain counterfeit goods detection and prevention is what Huy Le et al. (2020) hope to offer. The paper discusses blockchain technology and QR codes, but it doesn't outline any specific methods. It does not state an accuracy % because the document is a system proposal rather than an implementation report. Benefits include increased consumer confidence and improved product authenticity. Cons: Implementation challenges and potential QR code scanning issues [14].

A blockchain-based method for locating and halting counterfeit goods in the supply chain is proposed by Arjun Mukherjee et al. (2013). They employ blockchain technology, QR codes for product verification, and Algorithm 2 for product delivery. The technology provides effective counterfeit identification, cost savings, and enhanced transparency [15].

Using machine learning approaches, Taranjyot Singh Chawla et al. (2022) were able to identify online payment fraud with an accuracy rate of 89.9% using logistic regression and 87% using the Random Forest classifier. Notwithstanding the difficulty in attaining zero false positives and negatives, the study emphasizes the significance of feature reduction in machine learning techniques for enhanced performance, fraud detection, and model efficiency [16].

With an accuracy of 0.9963, Petr Hajek et al. created an XGBoost-based fraud detection framework for mobile payment systems in 2022. The framework provides quick capabilities, class imbalance control, cost reductions, and teamwork tactics; nonetheless, it necessitates further analysis of machine learning techniques and real-world data evaluation [17].

Using the Local Outlier Factor technique and ensemble learning, Paolo Vaninil et al. (2023) aim to develop machine learning models for the detection of online payment fraud. The study reports detection rates ranging from 18% to 45%, effectively optimizing detection rates and minimizing fraud losses by over 60%. The study's limitations include the use of synthetic data and the restricted literature comparison [18].

Eishvak Sengupta et al. (2011) investigated machine learning and artificial intelligence methods for credit card fraud detection. The article discusses algorithms like neural networks and Naive Bayes. The report makes no explicit mention of results or percentages. This article provides a thorough analysis of AI and ML techniques for credit card fraud detection. The report lacks accurate result percentages and may not incorporate all available fraud detection techniques [19].

According to Lahari Madabhattula et al. (2021), fraudulent internet transactions can be detected and prevented by employing behavior and location analysis. A combination of the Hidden Markov Model (HMM) and Behavior and Location Analysis (BLA). The increased use of BLA and HMM for secure online transactions, user behavior analysis, and fraud detection does not have a specific outcome %. Inadequate field testing, imprecise implementation guidelines, and potential scalability issues [20].

3. Methodology

Spotting a phony attack by concealing the deception of the web was not the primary goal...the major purpose of this study was to determine challenging task of identifying fake online attacks where the attackers cleverly hide their true intentions within the website itself. APIs analyse request-response to sniff out phony URL disguise. The result of fake product journey tracks using block chain exposing fakes at a glance. Even fake reviews have been detected using sentimental analysis where AI sniffs out words to expose hidden review algorithms shift data, guarding against fraudulent transactions.

The overall view of this paper is, spotting online fraud involves highly skilled and smart tools, URLs flagged for plagiarism, reviews examined for suspicious tone, products tracked on block chains, and payments detected using effective algorithms.

3.1 FAKE URL'S

This study looked at and detailed a system that uses N-gram models, request-response APIs, and Support Vector Machines (SVMs) to detect malicious URLs. The complete procedure is shown in Figure 1.

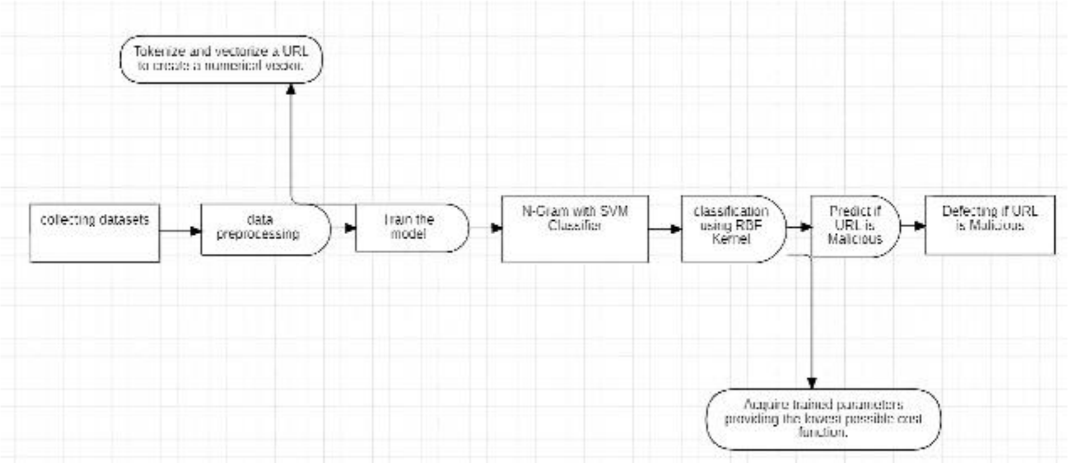

Tokenize and vectorize a URL to create a numerical vector.
collecting datasets
data preprocessing
Train the model
N-Gram with SVM Classifier
classification using RBF Kernel
Predict if URL is Malicious
Detecting if URL is Malicious
Acquire trained parameters providing the lowest possible cost function.

Fig 1. Fake URL architecture

3.1.1 Motivation:

Malicious URLs pose a significant threat to cybersecurity, enabling malware attacks and data breaches. Early detection of these URLs is crucial for user protection.

3.1.2 Proposed Approach:

Utilize N-gram models to extract features from the URL structure, capturing patterns and sequences of characters.

Employ SVMs, a machine learning algorithm, to classify URLs as malicious or benign based on the extracted features.

3.1.3 Justification For Using Svms:

SVMs offer high accuracy for large datasets, suitable for handling extensive URL lists. They excel at text classification tasks like identifying malicious patterns. Malicious URL Detection with N-gram and svm.

3.1.4 SVM:

Focuses on analysing the URL structure itself (characters, sequences) using N-grams. Employ's machine learning (SVM) to classify URLs as malicious or benign based on extracted features. Doesn't directly involve API requests or responses.

3.1.5 Spotting Deceptive Web Attacks:

Often targets APIs as entry points for injecting malicious code or data. Involves analysing API request-response patterns, data payloads, and potential vulnerabilities. Doesn't directly utilize N-gram or SVM techniques. While both areas aim to enhance cybersecurity, they address different aspects of web threats.

3.1.6 Implementation Steps:

Pre-Processing:

Tokenize URLs into individual words or characters. Extract N-gram features (e.g., consecutive pairs or triplets of characters).

Training:

Train the SVM model on a labelled dataset of known malicious and benign URLs. The model learns to distinguish between the two classes based on the extracted features.

Prediction:

Present new URLs to the trained model. The model predicts whether each URL is likely to be malicious or not.

Results:

The authors achieved the best results using the N-gram and SVM approach compared to other methods.

3.2 Fake Product

Here's the next module for this paper: Ganache has introduced a smart contract for Ethereum, enabling secure Dapp development, launch, and testing, while a MetaMask cryptocurrency wallet is used for transactions. Dapp users can't log in without first authenticating. Manufacturers join the DApp with ID by enrolling products with price, QR code, and serial number. Customers use a QR code or serial number to verify items, sellers submit a seller code, and data is stored on a blockchain network and its architecture in fig 2.

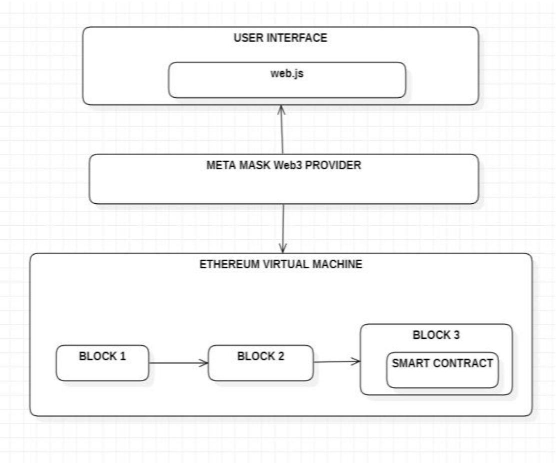


Fig 2. Fake Product architecture

3.2.1 Block Chain

Supply chain management can benefit greatly from blockchain technology's ability to store data on nodes, which allows for the tracking, sharing, and security of accounts, payments, orders, and product prices. Some features of this technology include:

1.Security And Privacy:

Blockchain ensures users have a public and private key combination for transaction validation, guaranteeing immutable and permanent transactions. Blockchain uses public key encryption to secure data.

2.Decentralization:

Blockchain is a distributed ledger technology that functions independently of outside parties or a centralized authority.

3.Transparency:

Smart contracts control the public access to blockchain data. The suggested solution involves Manufacturer, Seller, and Consumer stakeholders and makes use of Ganache Ethereum's and the MetaMask wallet.

3.2.2ethereum Dapp Architecture

The user interface of the system is based on NodeJS, and Web3.js establishes a connection with MetaMask, a provider, to generate and certify transactions using the user's private key. This ensures safe network communication without storing the private keys of users.

The blockchain's distributed ledgers are transforming the ownership tracking system. The business and e-commerce industries are changing quickly, which is having an impact on supply chain trends. The Dapp created here may be trusted to be used in e-commerce and guarantees more openness in supply chain management. We have exclusively examined medical products in this study.

3.3 Fake Review

The present study investigates the efficacy of machine learning to detect fraudulent e-commerce evaluations by concentrating on linguistic and affective factors. The objective is to improve transparency by identifying characteristics of fraudulent reviews, using an Amazon dataset encompassing 21,000 reviews that have been equally divided into real and fraudulent reviews and its architecture in fig 3.

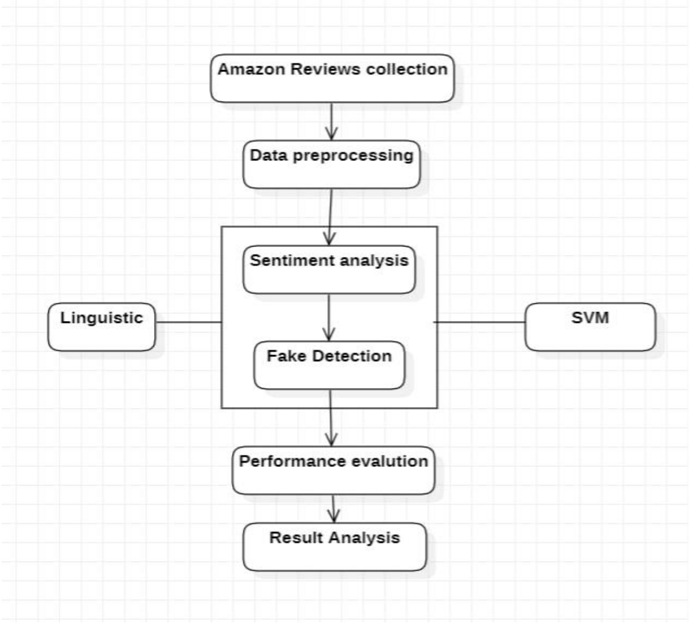


Fig 3. Fake Review architecture

3.3.1.Data Preparation

Further steps in the implementation and data preparation process might be necessary, such as controlling punctuation and special characters and applying cutting-edge methods to improve the sentiment analysis model's performance.

3.3.2. Feature Extraction

Feature extraction is a method that produces better results than directly applying machine learning to raw data because it transforms unprocessed input into numerical features.

3.3.3. Feature Selection Code Using The Rfe

Recursive Feature Elimination (RFE) is a machine learning technique that ranks features based on relevance, removing the least important ones until the required number of features is achieved.

3.3.4. Sentiment Classification Algorithms.

Several well-known classifiers will be used in this study:

1) Naïve Bayes (Nb)

A simple probabilistic classifier based on the Bayes theorem is the NB classifier. A set of probabilities is computed by the NB using combinations of values from a given dataset.

Additionally, the NB classifier makes decisions quickly.

2) Svm, Or Support Vector Machine

Despite the development of new techniques, SVM is a popular supervised learning model in machine learning, evaluating data to uncover patterns for regression analysis and classification.

3) Logistic Regression

Although the name suggests otherwise, this technique is applied to classification problems. It's a straightforward yet effective sentiment analysis method.

4) Random Forest

To increase accuracy and performance, this ensemble learning technique mixes many decision trees.

5. Detection Process And Result

Following training, the model's output on the testing dataset is predicted, and a confusion matrix is created to categorize the reviews as positive or negative.

This study emphasizes the need for greater awareness and comprehension of online reviews by examining the use of textual traits, mood, and linguistics to identify and forecast fraudulent reviews.

3.4 Fake Payment

The most widely used form of payment in the world today is online. But as the number of online payments rises, so does the incidence of payment fraud. This study aims to distinguish between fraudulent and non-fraudulent payments and its architecture in fig 4.

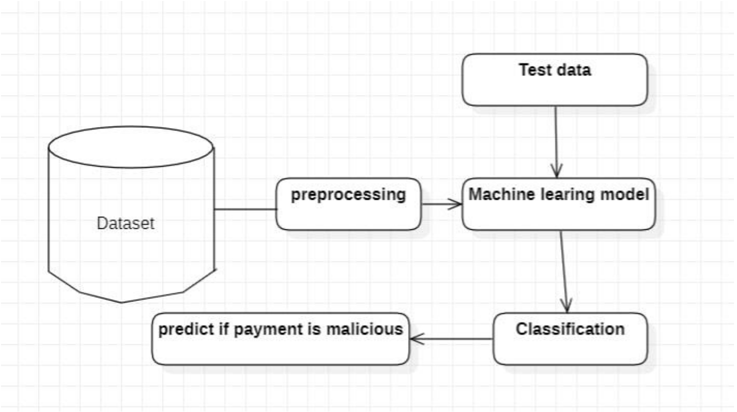


Fig 4. Fake Payment architecture

3.4.1 Data Preprocessing:

Data loading into a pandas Data Frame, handling missing values, fraud detection feature engineering, categorical variable encoding, and data splitting into groups for testing and training is every phase in the process

3.4.2 Exploratory Data Analysis (Eda):

To begin comprehend variable distribution, this task involves creating descriptive statistics. Data distribution and fraud patterns then become apparent through the use of visualizations.

3.4.3 Feature Selection:

Correlation analysis is used in the process to find significant predictors, and feature importance techniques were performed to select relevant behaviors.

3.4.4 Model Building:

The Random Forest algorithm is used to detect fraud. Performance metrics are determined, hyperparameters change, and a dataset is employed to train the model.

3.4.5 Model Interpretation:

In an effort to discover areas for improvement, the research comprises studying the model's predictions, establishing the factors influencing its evaluation, and looking at misclassifications.

3.4.6 Deployment:

The trained model is to be involved into the present setup, watched and retrained on frequently, and a feedback loop should be set up to collect information pertaining to fraud cases that have been identified.

3.4.7 Continuous Improvement:

The necessity of remaining up with the most recent fraud detection tools and incorporating feedback from ongoing monitoring and user interactions is repeatedly highlighted during the text.

4.Results

The system is divided into four primary sections that make use of HTML, CSS, and JS to determine the fake URL, fake product, fake review, and fake payment as shown in fig 5. UI-Path was used to connect the frontend and backend, which helps detect fraud.

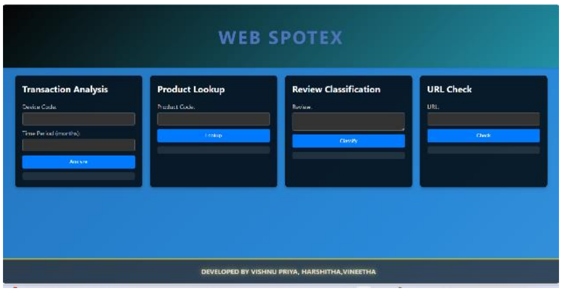
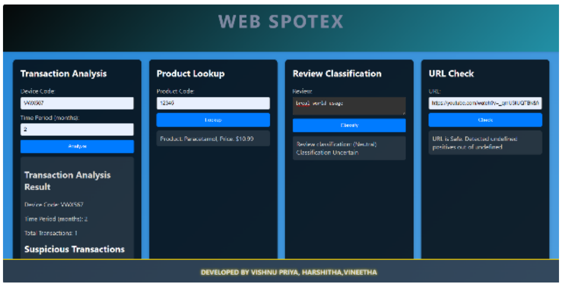


Fig 5. Website for recognizing deception

4.1. Fake URL

This real-time system uses a request response paradigm to determine whether the URL it has received is malicious or legal. It does this by connecting to another system through an API and confirming whether the website is malicious or not.It checks a lot of reliable websites, takes the bulk of the sites' replies, and decides if a website is malicious in fig 6 and 7.



Device Code: VWX567

Time Period (months): 2

Transaction Analysis Result

Device Code: VWX567

Time Period (months): 2

Total Transactions: 1

Suspicious Transactions

Product Code: 12345

Product: Paracetamol, Price: $10.99

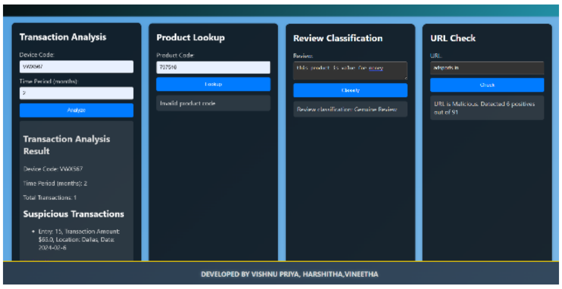
Review: [illegible]

Review classification: (Neutral) Classification Uncertain

URL: https://youtube.com/watch?v=_gnUStUQTBk8/

URL is Safe. Detected undefined positives out of undefined

Fig 6. Safe URL Checker



Device Code: VWX567

Time Period (months): 2

Transaction Analysis Result

Device Code: VWX567

Time Period (months): 2

Total Transactions: 1

Suspicious Transactions

- Entry: 15, Transaction Amount: $65.0, Location: Dallas, Date: 2024-02-6

Product Code: 707510

Invalid product code

Review: this product is value for money

Review classification: Genuine Review

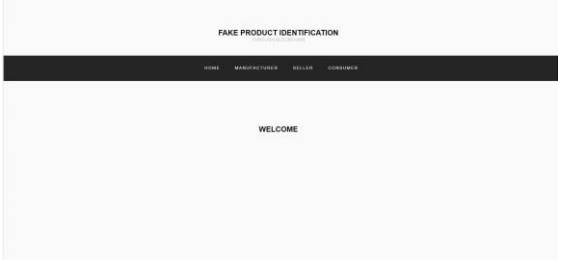
URL: adsports.in

URL is Malicious. Detected 6 positives out of 91

Fig 7. Unsafe URL

4.2. Fake Product

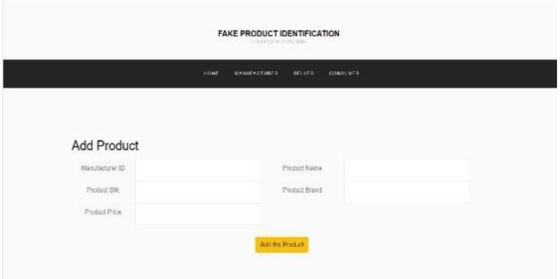
The dashboard shown in figure 8 is where this blockchain-based system, which checks if a product is authentic or fake, is located. The system's three main parts are consumers, sales, and manufacturing. Figure 9 illustrates the location of the product's manufacturing area. The product is consumed in the consumer field and sold in the seller field (fig. 10). We can only confirm if the product is genuine or fake in the consumer field (fig. 11). By scanning a product's QR code, it can determine whether it is authentic or fake.



FAKE PRODUCT IDENTIFICATION

WELCOME

Fig 8. Welcome page



FAKE PRODUCT IDENTIFICATION

Add Product

Manufacturer ID

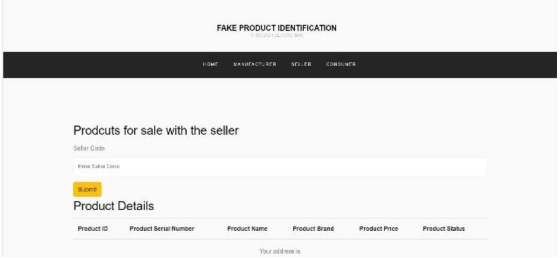
Product Name

Product SN:

Product Brand

Product Price

Fig 9. Manufacturer page



FAKE PRODUCT IDENTIFICATION

Prodcuts for sale with the seller

Seller Code

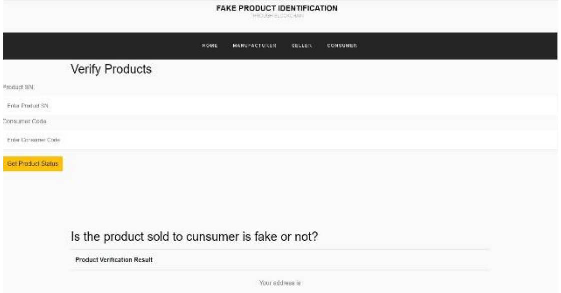
Enter Seller Code

Product Details

Product ID	Product Serial Number	Product Name	Product Brand	Product Price	Product Status

Your address is

Fig 10. Seller Page



FAKE PRODUCT IDENTIFICATION

Verify Products

Product SN:

Enter Product SN

Consumer Code

Enter Consumer Code

Is the product sold to cunsumer is fake or not?

Product Verification Result

Your address is

Fig 11. Product detection page

4.3. Fake Review

The design of this system makes use of sentimental analysis to determine whether the review is authentic or artificial. Use the Gaussian Navie Bayes algorithm and the Random Forest algorithm to find it to ascertain whether or not the review was fraudulent. Figures 12 and 1 illustrate how to apply the results in a confusing way. Matrix, accuracy for 41 features using random forest, with results displayed in the form of confusion in Figure 13 and Table 2. Matrix, Gaussian accuracy for 35 features Navie Bayes ambiguity. The remaining findings are displayed as graphs in Figures 14, 15, and 16.

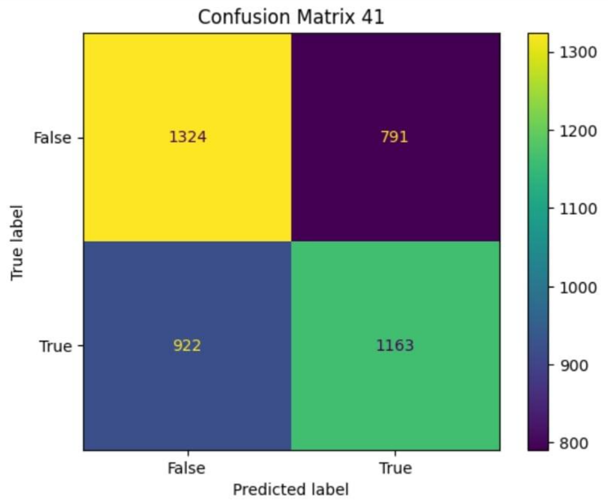


Fig 12. Random Forest confusion Matrix

Table 1. Random Forest method included 41 features.

Accuracy	59.11
Precision	59.75
Recall	54.05
F1 Score	56.76

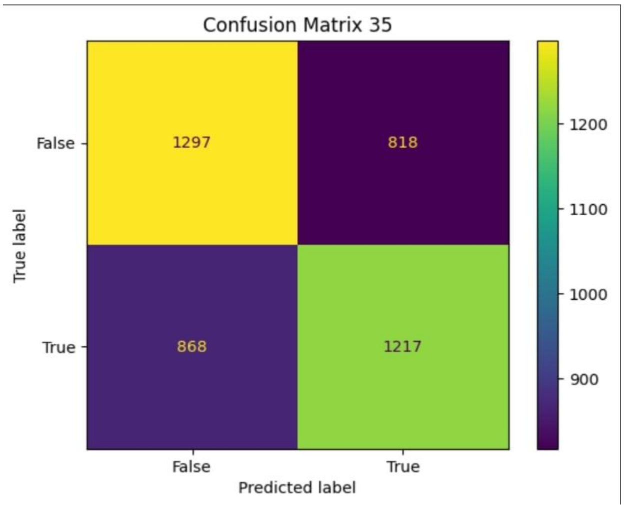


Fig 13. Gaussian Navie Bayes confusion matrix

Table.2. Gaussian Navie Bayes included 35 features.

Accuracy	55.23
Precision	69.74
Recall	17.36
F1 Score	27.80

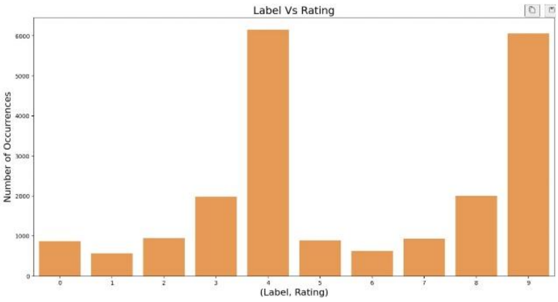


Fig:14 Graph of label vs rating with number of occurrences

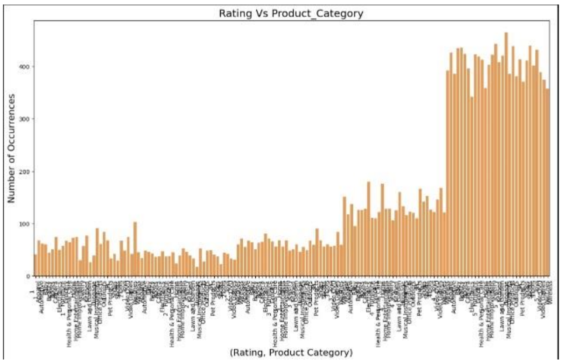


Fig:15 Rating vs Product Category

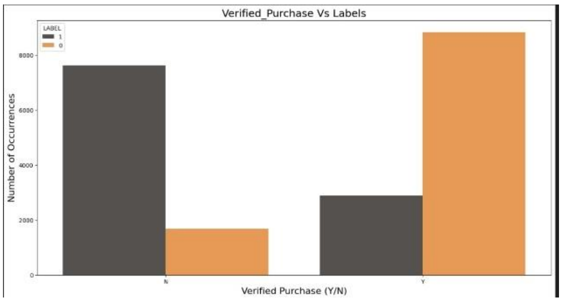


Fig:16 verified purchase vs category

4.4. Fake Payment

For training and testing, the dataset is divided 80:20. To determine if the payment is fraudulent or not, the system utilized k-fold cross validation. To identify if a payment is fraudulent or genuine, logistic regression and random forest modelling are used. Table 3 and Fig. 17 present the accuracy using random forest and the result in the form of a confusion matrix. The remaining findings are displayed as graphs in Figures 18, 19, and 20 and table 4.

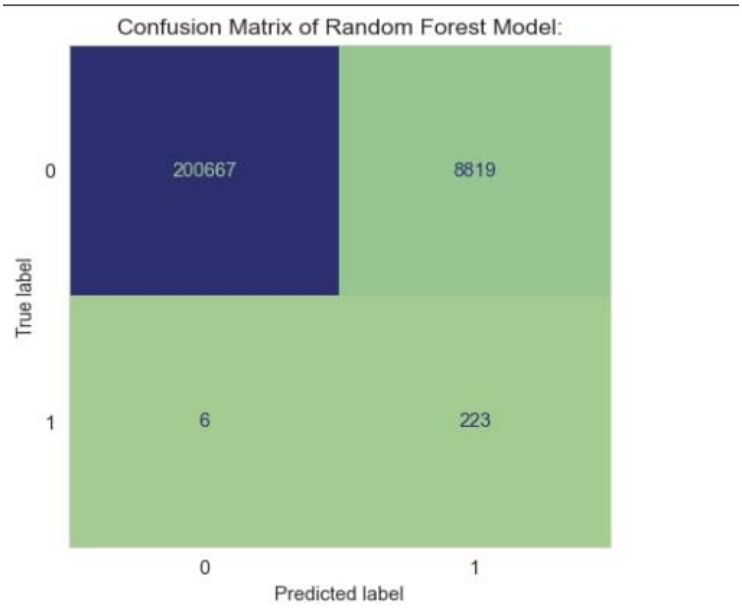


Fig 17. Confusion Matrix of Random Forest

Table 3. Random Forest Result

Mean accuracy	0.985 (0.003)
Mean precision	0.975 (0.006)
Mean recall	0.996 (0.002)
Mean f1 score	0.985 (0.003)
Mean roc_auc score	0.998 (0.000)

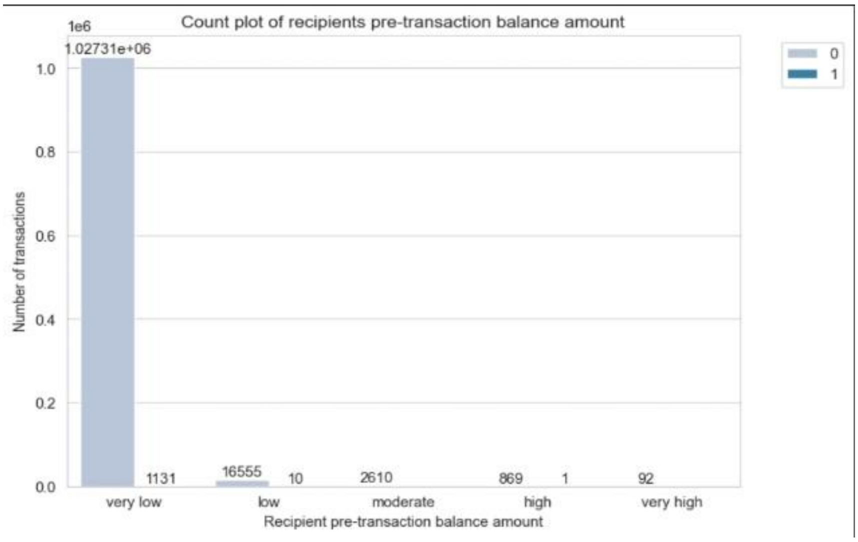


Fig 18.Recipient pre-transaction balance amount

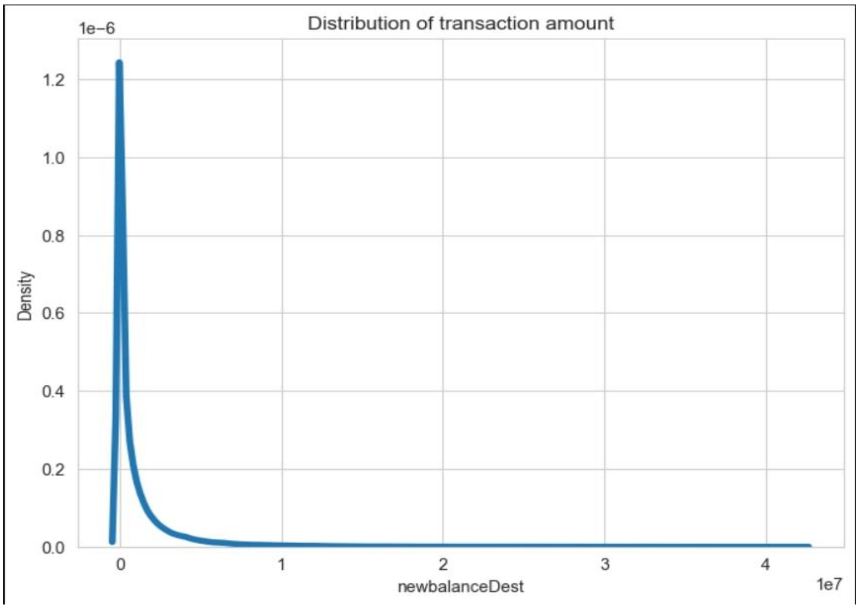


Fig 19. Distribution of transaction amount

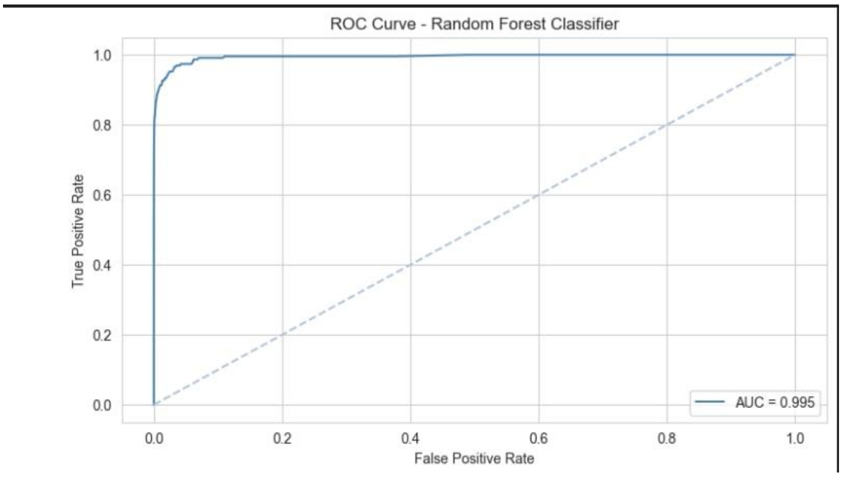


Fig 20. ROC Curve

From the confusion matrix, 1,239,155 were correctly classified as non-fraudulent payments, and 31,726 people were misclassified as non-fraudulent payments. According to the confusion matrix, 1,639 payments were incorrectly labelled as fraud while 4 payments were correctly identified as fraud.

Table 4. Logistic Regression Results

Mean accuracy	0.848 (0.007)
Mean precision	0.843 (0.008)
Mean recall	0.856 (0.005)
Mean f1 score	0.849 (0.006)
Mean roc_auc score	0.927 (0.004)

5. Conclusion And Future Scope

This paper by reducing search time, the detection of fraudulent URLs, products, payments, and reviews on a single website contributes to the development of a more secure and reliable online community. The rise in internet usage has coincided with a rise in fraud. It promotes moral corporate practices, protects customers from online risks, and helps to build confidence in the virtual economy. To stay ahead of developing hazards in the ever-changing world of online interactions, this paper's organizations, security systems, and regulatory authorities continually adapt these detection procedures.

The value of more research on this approach is evident in the reality that it can be implemented on a single page and that multiple technologies have been used with various APIs and numerous other algorithms. Some of such tasks include finding fake IDs for Instagram, Facebook, and Twitter as well as videos, photos, outlets, endorsements, paperwork, job profiles, screenings, and more.

6. References

[1] A systematic literature review on phishing website detection techniques by Asadullah Safia,Satwinder Singh in 2023

[2]Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction by Muna Elsadig , Ashraf Osman Ibrahim in 2022

[3] Phishing Websites Detection Using Machine Learning by P. Amba Bhavani ASST.PROFESSOR in 2022

[4] Classification of Malicious URLs Using Machine Learning by Shayan Abad, Hassan Gholamy and Mohammad Aslani in 2023

[5] Malicious URL: Analysis and Detection using Machine Learning by Mansi Mehndiratta in 2023

[6]Enhancing Fake Product Detection Using Deep Learning Object Detection Models by Eduard Daoud, Dang Vu, Hung Nguyen and Martin Gaedke in 2020

[7] Fake Product Monitoring System using Artificial Intelligence by Ms. Reema Anne Roy in 2021

[8] Review Of Blockchain- Fake Product Identification by Ms. Vaishnavi Hedaoo, Ms. Sakshi Sawarkar, Ms. Mayuri Kosare, Ms. Pragati Gawande, Mr. Swapnil Wahokar in 2022

[9]Blockchain Based Fake Product Identification System by Swaroop Jambhulkar , Harsh Bhoyar, Shantanu Dhore, Arpita Bidkar, Prema Desai in 2022

[10] Identifying Counterfeit Products using Blockchain Technology in Supply Chain System by Nafisa Anjum in 2022

[11] Fake Reviews Detection using Supervised Machine Learning by Ahmed M. Elmogy,Usman Tariq, Atef Ibrahim in 2021

[12] BSTC: A Fake Review Detection Model Based on a Pre-Trained Language Model and Convolutional Neural Network by Junwen Lu in 2023

[13] A Method for the Detection of Fake Reviews based on Temporal Features of Reviews and Comments by Wenqian Liu a, Jingsha He, Song Han, Nafei Zhu in 2019

[14] Detection Of Fake Reviews On Social Media Using Machine Learning Algorithms by Huy Le, Seattle in 2020

[15] Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews by Arjun Mukherjee , Vivek Venkataraman, Bing Liu, Natalie Glance in 2013

[16] Online Payment Fraud Detection using Machine Learning Techniques by Taranjyot Singh Chawla in 2022

[17] Fraud Detection in Mobile Payment Systems using an XGBoost-based Framework by Petr Hajek, Mohammad Zoynul Abedin, Uthayasankar Sivarajah in 2022

[18] Online payment fraud: from anomaly detection to risk management by Paolo Vanini1, Sebastiano Rossi, Ermin Zvizdicand Thomas Domenig in 2023

[19] A Review of Payment Card Fraud Detection Methods Using Artificial Intelligence by Eishvak Sengupta, Naman Jain, Dhruv Garg, Tanupriya Choudhury in 2018

[20] Online Transaction Fraud Detection by Lahari Madabhattula, Maridu Manikanta, Pradeep Kumar in 2021