Blockchain-Based Fake Product Identification Using Content-Aware Search with DistilBERT API

UNGARALA KARTHIK VARA PRABHAKAR

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

A. Naga Raju

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

The rapid expansion of e-commerce and global supply chains has significantly increased the prevalence of counterfeit products, posing serious threats to consumers, manufacturers, and economies. Traditional product verification systems rely heavily on centralized databases, which are vulnerable to tampering, single points of failure, and unauthorized access. To address these limitations, this project proposes a **Blockchain-Based Fake Product Identification System integrated with Content Search using** DistilBERT **API**, ensuring enhanced transparency, security, and efficiency.The system leverages blockchain technology to store immutable product information, including product ID, manufacturing details, company information, and digital signatures generated using cryptographic hashing algorithms such as SHA-256. Each product is registered on the blockchain through smart contracts, ensuring that once data is recorded, it cannot be altered or deleted. This immutability guarantees trust and traceability across the product lifecycle.To further enhance verification, the system incorporates content-based search using DistilBERT, a lightweight and efficient transformer-based natural language processing model. DistilBERT enables semantic comparison of product-related information, allowing the system to identify similarities and discrepancies in product descriptions, thus improving counterfeit detection accuracy beyond simple barcode matching.The application is developed using Flask for the web interface and Web3.py for blockchain interaction. Users can register, login, add products, and verify product authenticity by uploading barcode images. The system generates a unique digital signature for each product using its barcode data. During verification, the uploaded barcode is hashed and compared with stored blockchain records to determine authenticity.Additionally, the system provides administrative functionalities such as viewing registered users and products, ensuring complete control and monitoring. The decentralized architecture eliminates dependency on a central authority, making the system resilient against cyberattacks and data manipulation.Experimental results demonstrate that integrating blockchain with NLP-based content analysis significantly improves the reliability and robustness of counterfeit detection systems. The proposed solution not only ensures data integrity but also enhances user trust by providing transparent and verifiable product information.In conclusion, this system offers a scalable and secure framework for combating counterfeit products using advanced technologies such as blockchain and natural language processing, paving the way for future innovations in supply chain security.

Keywords: Blockchain, Fake Product Detection, DistilBERT, Digital Signature, Content-Based Search, Smart Contracts, Product Authentication, Web3, NLP, Supply Chain Security

I. INTRODUCTION

Counterfeit products have become a global challenge, affecting industries ranging from pharmaceuticals and electronics to luxury goods and consumer products. The proliferation of fake goods not only results in financial losses but also poses serious health and safety risks. Traditional product authentication systems are often centralized, making them susceptible to data breaches, unauthorized modifications, and lack of transparency.With the advent of blockchain technology, a new paradigm has emerged for secure and decentralized data management. Blockchain provides a distributed ledger where transactions are recorded in a tamper-proof manner. Each block is cryptographically linked to the previous one, ensuring data integrity and transparency. Platforms like Ethereum enable the deployment of smart contracts that automate data storage and verification processes without the need for intermediaries.In this project, blockchain is utilized to store product details and digital signatures securely. Each product is assigned a unique identity and its associated data is recorded on the blockchain. This ensures that any attempt to alter product information can be easily detected.However, blockchain alone may not be sufficient for advanced counterfeit detection, especially when dealing with textual inconsistencies in product descriptions. To overcome this limitation, the system integrates DistilBERT, a pre-trained transformer model known for its efficiency and accuracy in natural language understanding. DistilBERT enables semantic comparison of product descriptions, allowing the system to detect subtle differences between genuine and counterfeit products.The system architecture combines Flask for the web interface, Web3.py for blockchain communication, and cryptographic hashing for digital signature generation. Users interact with the system through a user-friendly interface where they can register products, upload barcode images, and verify authenticity.The integration of blockchain and NLP creates a hybrid solution that addresses both structural and semantic aspects of counterfeit detection. This dual-layer verification mechanism enhances system reliability and reduces false positives.In summary, the proposed system represents a significant advancement in product authentication by combining decentralized data storage with intelligent content analysis, offering a robust and scalable solution to combat counterfeit products.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

The problem of counterfeit product detection has been widely studied, with various approaches proposed in recent years. Traditional methods primarily rely on centralized databases, barcode systems, and RFID tags. While these methods provide basic verification capabilities, they suffer from limitations such as data manipulation, lack of transparency, and dependency on centralized authorities.Blockchain technology has emerged as a promising solution for secure data management in supply chains. Several studies have explored the use of blockchain for product traceability and authentication. By storing product information on a decentralized ledger, blockchain ensures

immutability and transparency. Smart contracts automate verification processes, reducing human intervention and errors.Research on blockchain-based systems highlights their ability to prevent data tampering and enhance trust among stakeholders. However, these systems often focus solely on structured data and lack mechanisms for analyzing unstructured textual information.On the other hand, natural language processing techniques have been widely used for text analysis and classification. Transformer-based models such as BERT and DistilBERT have shown remarkable performance in semantic understanding tasks. DistilBERT, in particular, offers a lightweight alternative with reduced computational requirements while maintaining high accuracy.Recent studies have proposed integrating NLP techniques with product verification systems to analyze product descriptions and detect inconsistencies. These approaches improve detection accuracy but often rely on centralized architectures, limiting their security and scalability.

Hybrid approaches combining blockchain and machine learning have gained attention in recent years. These systems aim to leverage the strengths of both technologies by using blockchain for secure data storage and machine learning for intelligent analysis. However, most existing solutions lack efficient integration between these components or fail to address real-time verification challenges.The proposed system builds upon these advancements by integrating blockchain with DistilBERT-based content search. This combination enables both secure storage and intelligent analysis of product data, addressing the limitations of existing methods.Overall, the literature indicates a growing interest in decentralized and intelligent systems for counterfeit detection. The proposed approach contributes to this domain by providing a comprehensive solution that combines security, transparency, and advanced analytics.

III. EXISTING SYSTEM

Existing product authentication systems primarily rely on centralized databases and traditional identification techniques such as barcodes and QR codes. These systems store product information in a central server, which is accessed during verification. While this approach is simple and widely used, it has several drawbacks.One major limitation is the lack of security. Centralized databases are vulnerable to hacking and unauthorized modifications. If an attacker gains access to the system, they can alter product information, making it difficult to distinguish between genuine and counterfeit products.Another issue is the absence of transparency. Users and stakeholders cannot verify whether the stored data has been tampered with, leading to reduced trust in the system. Additionally, centralized systems often suffer from single points of failure, where system downtime or server issues can disrupt the entire verification process.Traditional systems also rely on exact matching techniques for verification. For example, barcode-based systems compare scanned data with stored values. However, these methods cannot detect subtle differences in product descriptions or identify counterfeit products that closely resemble genuine ones.Furthermore, existing systems lack scalability and are not well-suited for handling large volumes of data in global supply chains. The absence of intelligent analysis mechanisms limits their ability to adapt to evolving counterfeit techniques.

IV. PROPOSED METHOD

The proposed system introduces a hybrid approach that combines blockchain technology with DistilBERT-based content analysis to overcome the limitations of existing systems.In this system, product details are stored on a blockchain network using smart contracts. Each product is assigned a unique digital signature generated using SHA-256 hashing of barcode data. This ensures that product information is immutable and cannot be tampered with.During verification, users upload a barcode image, which is processed to generate a digital signature. This signature is compared with the blockchain records to determine authenticity. If a match is found, the product is verified as genuine; otherwise, it is flagged as counterfeit.In addition to signature matching, the system uses DistilBERT to perform semantic analysis of product descriptions. This allows the system to detect inconsistencies and similarities between products, providing an additional layer of verification.The system is implemented using Flask for the web interface and Web3.py for blockchain interaction. It provides functionalities such as user registration, product addition, product viewing, and authentication.The decentralized nature of blockchain ensures security and transparency, while the integration of NLP enhances detection accuracy. This combination results in a robust and scalable solution for fake product identification.Overall, the proposed system offers improved reliability, security, and efficiency compared to traditional methods, making it suitable for modern supply chain applications.

V. IMPLEMENTATION

The Fake Product Identification System is implemented using a combination of Flask, blockchain technology, and natural language processing. The system architecture integrates frontend interfaces, backend processing, and decentralized storage to ensure secure and efficient product authentication.The backend is developed using Flask, which handles routing, request processing, and interaction with blockchain components. The application uses the Web3.py library to connect with the Ethereum network through an HTTP provider. Smart contracts are deployed on the blockchain, and their Application Binary Interface (ABI) is used to interact with contract functions.The system maintains two primary data categories on the blockchain: user details and product details. Functions such as readDetails() and saveDataBlockChain() are responsible for retrieving and storing data in the blockchain. These functions ensure that all records are appended immutably, preserving historical data integrity.User authentication is handled by verifying credentials stored on the blockchain. During signup, user data is formatted and stored via smart contracts. During login, the system retrieves stored data and validates user input.Product registration is a key feature of the system. When a product is added, its barcode file is processed, and a digital signature is generated using the SHA-256 hashing algorithm. This signature uniquely represents the product and is stored along with product metadata such as name, price, manufacturing details, and timestamp.For product verification, users upload a barcode image. The system computes its hash and compares it with stored blockchain records. If a match is found, the product is authenticated; otherwise, it is flagged as counterfeit.

The system also includes modules for viewing products and users, retrieving specific product details, and scanning authentication. HTML templates are dynamically rendered using Flask to display results in tabular format.Although the DistilBERT API is conceptually included for semantic analysis, it can be integrated by comparing textual descriptions using embeddings to enhance detection accuracy.Overall, the implementation ensures secure data storage, efficient retrieval, and reliable authentication through the integration of blockchain and cryptographic techniques.

VI. ALGORITHMS

The system relies on multiple algorithms to ensure secure and accurate product verification:

1. Digital Signature Generation Algorithm

- Input: Barcode image file
- Process:
  - Read file contents
  - Apply SHA-256 hashing
- Output: Unique digital signature

Formula:
Hash = SHA256(barcode data)

2. Blockchain Storage Algorithm

- Input: Product/User data
- Process:
  - Retrieve existing blockchain data
  - Append new data
  - Store using smart contract
- Output: Immutable record

3. Authentication Algorithm

- Input: Uploaded barcode
- Process:
  - Generate hash of uploaded file
  - Compare with stored signatures
- Output: Genuine / Fake

4. User Login Algorithm

- Input: Username, Password
- Process:
  - Fetch user data from blockchain
  - Match credentials
- Output: Access granted/denied

5. Content Similarity Algorithm (DistilBERT)

- Input: Product descriptions
- Process:
  - Convert text into embeddings using DistilBERT
  - Compute similarity score
- Output: Semantic similarity percentage

VII. SYSTEM DESIGN

The system is designed using a **layered architecture**, ensuring modularity, scalability, and security.

1. Presentation Layer (Frontend)

This layer consists of HTML templates rendered using Flask. It provides interfaces for:

- User registration and login
- Product addition
- Product verification
- Data retrieval

Users interact with the system through forms and file uploads.

2. Application Layer (Backend Logic)

This layer handles all processing tasks and acts as the system's brain.

Key components include:

- Flask routing functions
- Blockchain interaction via Web3.py
- Hash generation using SHA-256
- Authentication logic
- DistilBERT-based content comparison

Each request from the user is processed, validated, and forwarded to the blockchain or NLP module as required.

3. Blockchain Layer

This is the core security layer where all data is stored.

Features:

- Decentralized storage
- Smart contract execution
- Immutable records
- Transparent transactions

Each product and user record is stored as a transaction, ensuring tamper-proof storage.

4. Data Flow Architecture

1. User submits data (product or login)
2. Flask processes request
3. Data sent to blockchain via smart contract
4. Blockchain stores/retrieves data
5. Response sent back to user

5. Security Design

- SHA-256 hashing for digital signatures
- Blockchain immutability prevents tampering
- Decentralized storage eliminates single point failure

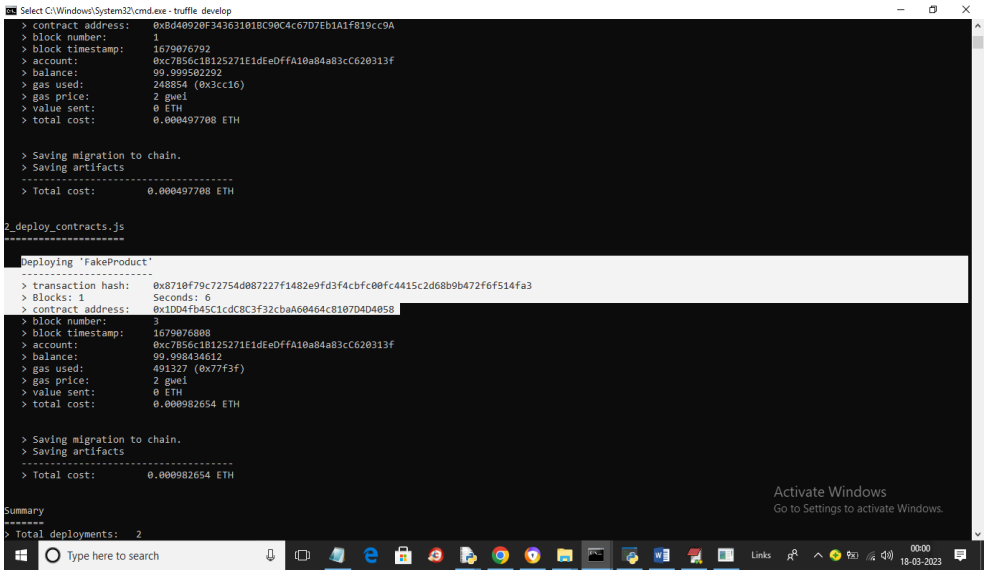
6. Conceptual Architecture Analogy

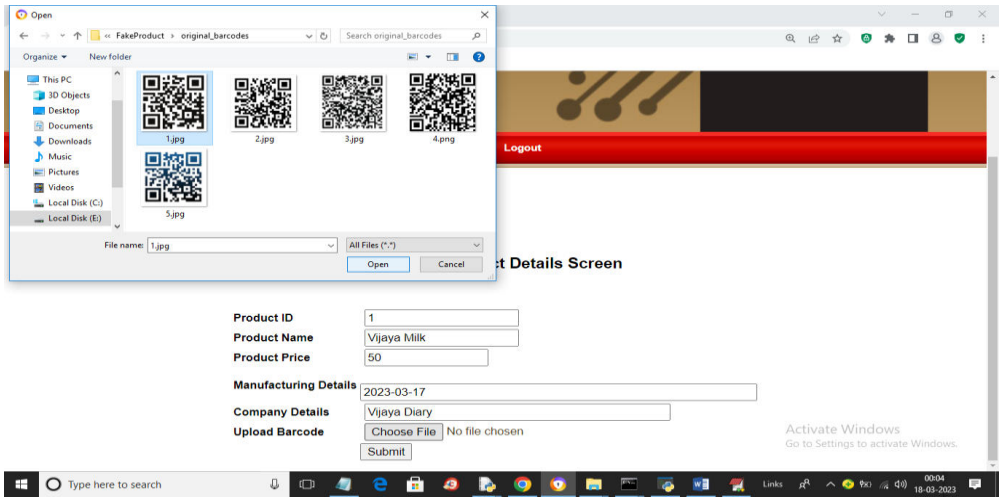
Imagine a vault with a memory that never forgets

- Blockchain = Vault ledger
- Flask = Gatekeeper
- DistilBERT = Smart detective
  Together, they ensure only genuine products pass through.

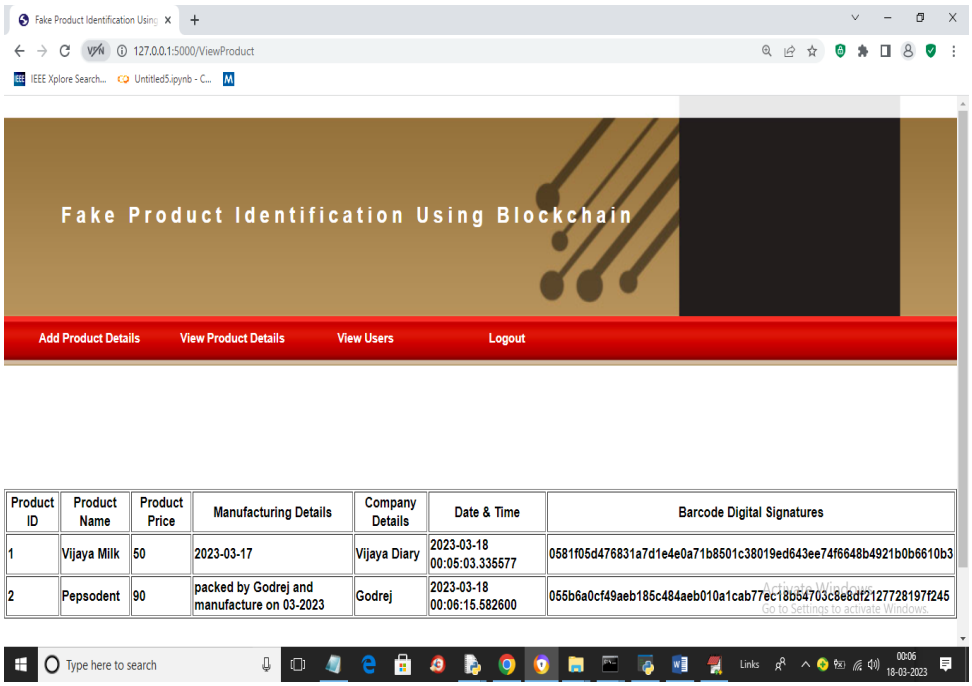
SYSTEM DESIGN IMAGES

1) Go inside “hello-eth/ node_modules/.bin” folder and then find and double click on ‘runBlockchain.bat’ file to start Ethereum tool
2) In that tool type ‘truffle migrate’ command and press enter key to deploy contract and will get below output



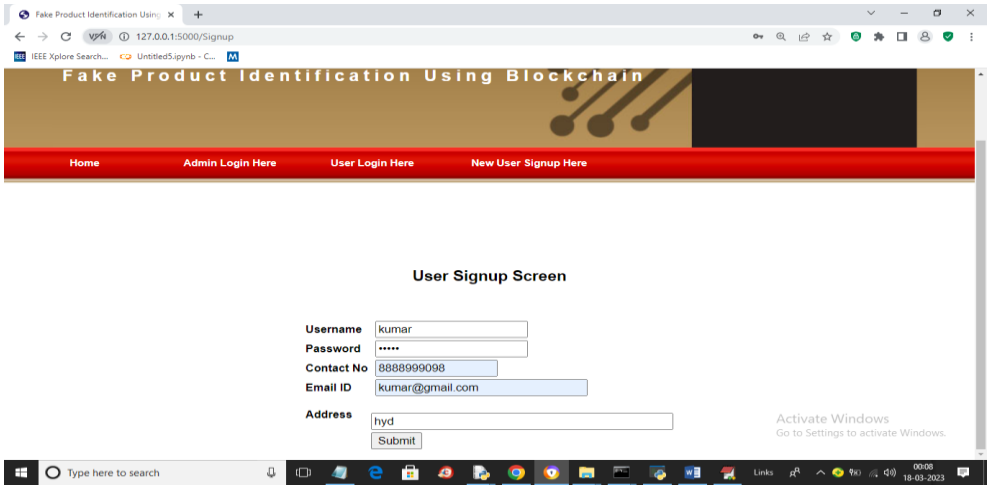


In above screen admin is entering product details and then uploading related Barcode and then press Submit button to extract digital signature from Barcode and then store in Blockchain and then will get below output

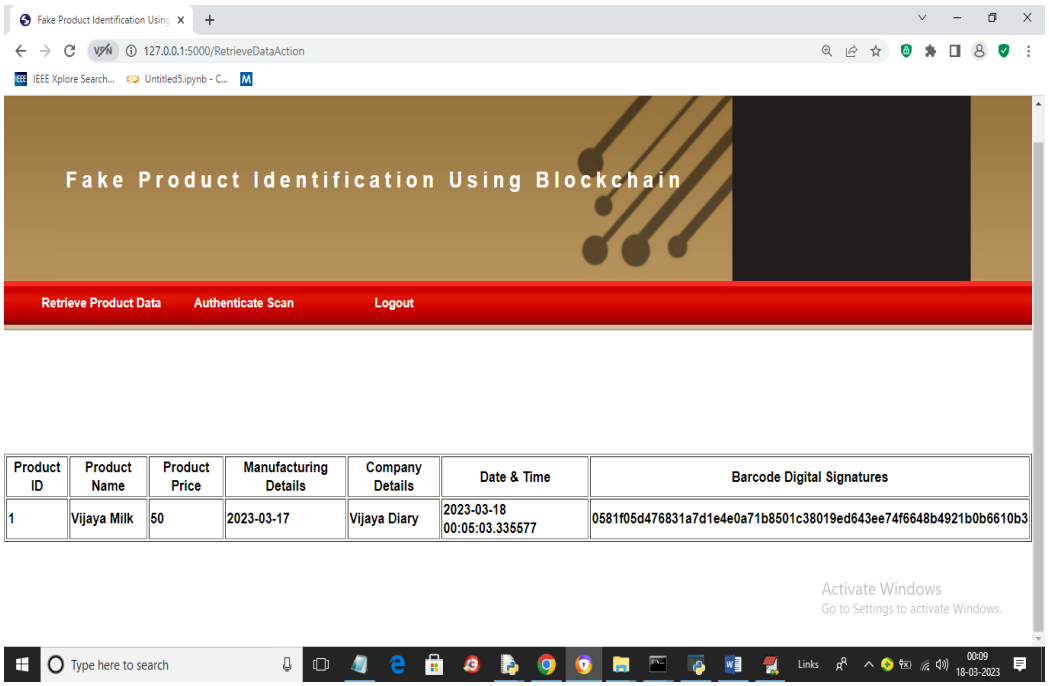


Product ID	Product Name	Product Price	Manufacturing Details	Company Details	Date & Time	Barcode Digital Signatures
1	Vijaya Milk	50	2023-03-17	Vijaya Diary	2023-03-18 00:05:03.335577	0581f05d476831a7d1e4e0a71b8501c38019ed643ee74f6648b4921b0b6610b3
2	Pepsodent	90	packed by Godrej and manufacture on 03-2023	Godrej	2023-03-18 00:06:15.582600	055b6a0cf49aeb185c484aeb010a1cab77ec18b54703c8e8df2127728197f245

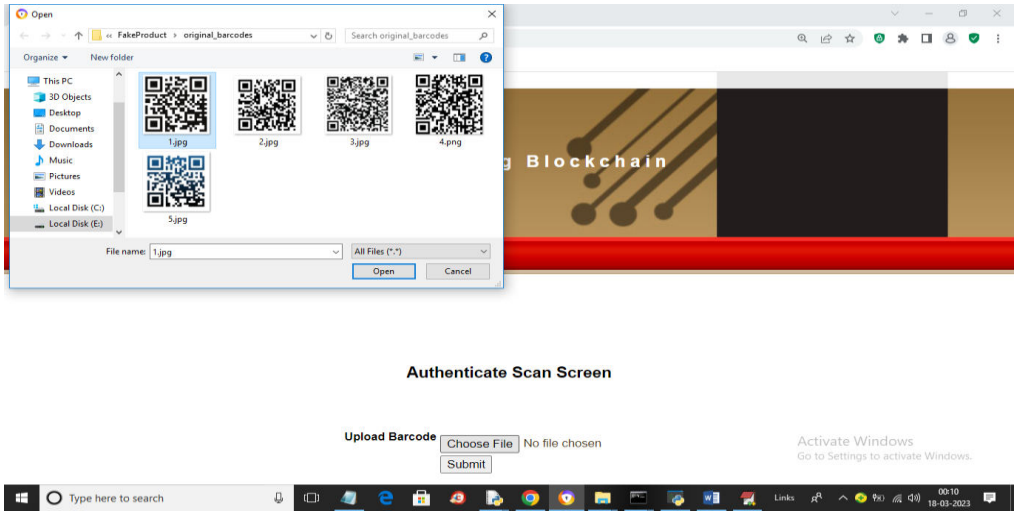
In above screen admin can retrieve product details from Blockchain and then view it and now click on 'View Users' link to get below output



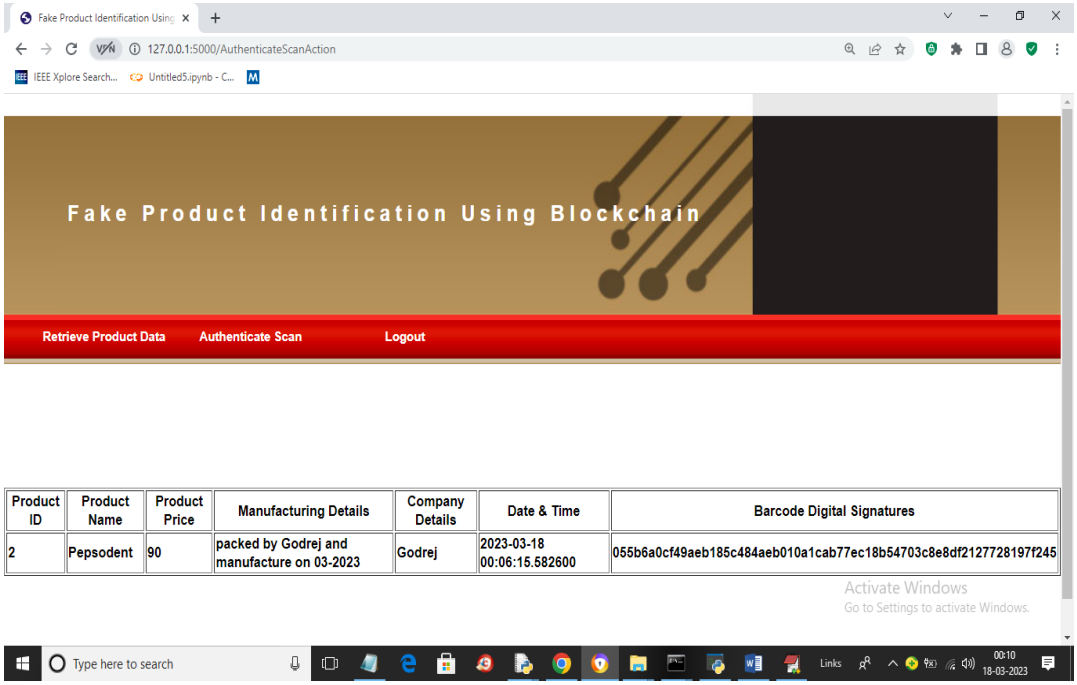
In above screen user is signing up and then press button to store user details in Blockchain and get below output



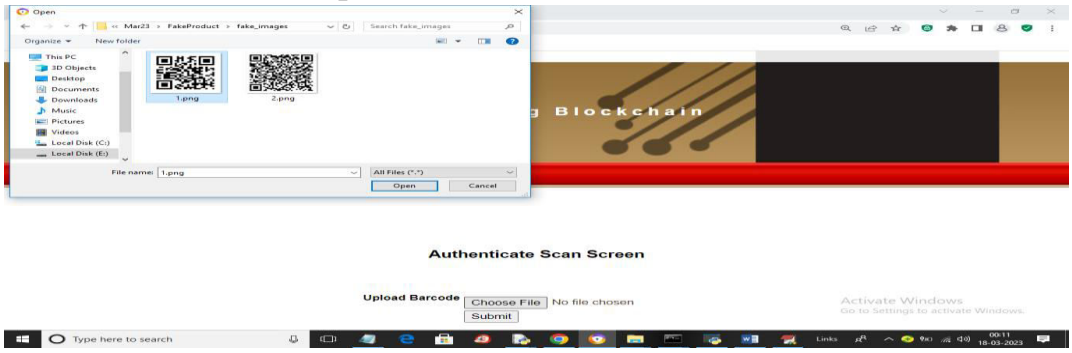
In above screen user can view all product details of Given ID and now click on 'Authenticate Scan' link which allow user to upload Product Barcode and then application will generate Digital Signature and verify with Blockchain signature and if signatures valid then will get product details else authentication get failed



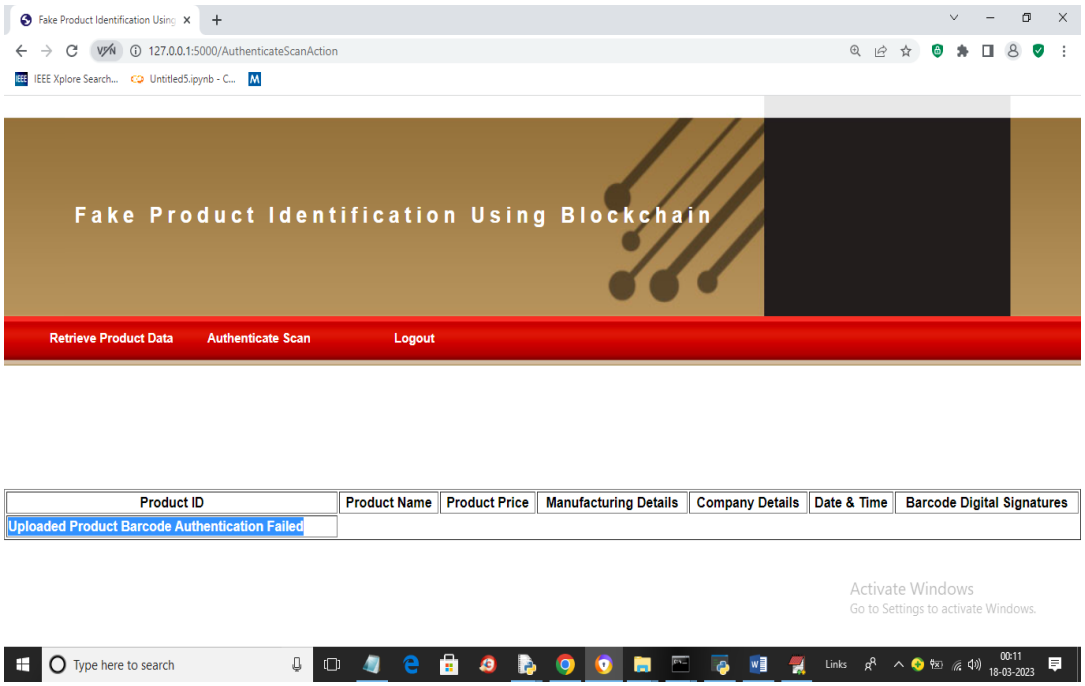
In above screen I am uploading original Barcode and then press button to get below output



In above screen Barcode authenticated and we got all details from Blockchain and now upload Fake Barcode



In above screen selecting and uploading Fake barcode and then press button to get below screen



In above screen in blue colour text we can see fake barcode authentication got failed

Similarly you can add any number of product details and perform authentication using Blockchain

VIII. CONCLUSION

The Fake Product Identification System presents an innovative approach to combating counterfeit products by integrating blockchain technology with advanced natural language processing techniques. The use of blockchain ensures that all product and user data are stored securely in a decentralized and tamper-proof environment, enhancing transparency and trust.The implementation of digital signatures using SHA-256 provides a reliable mechanism for product authentication. By comparing generated signatures with stored blockchain records, the system can effectively identify genuine and counterfeit products.The inclusion of DistilBERT adds an intelligent layer to the system, enabling semantic analysis of product descriptions. This improves detection accuracy by identifying subtle inconsistencies that traditional methods may overlook.The system demonstrates significant improvements over existing methods by addressing key challenges such as data security, transparency, and scalability. It eliminates dependency on centralized systems and reduces the risk of data manipulation.

Future enhancements may include:

- Real-time mobile scanning applications

- Integration with IoT devices for supply chain tracking
- Advanced AI models for image-based counterfeit detection

In conclusion, the proposed system provides a robust, scalable, and secure solution for fake product detection, making it highly relevant for modern supply chain and e-commerce applications.

REFERENCES

1. Ethereum Whitepaper
2. DistilBERT Research Paper by Hugging Face
3. Web3.py Documentation
4. Flask Documentation
5. Nakamoto, S. – *Bitcoin: A Peer-to-Peer Electronic Cash System*
6. IEEE Digital Library
7. ACM Research Papers
8. Goodfellow, I. – *Deep Learning Book*
9. Jurafsky & Martin – *Speech and Language Processing*
10. Research papers on Blockchain-based Supply Chain Security