ENHANCING SECURITY IN DIGITALIZED HEALTH CARE SYSTEMS USING BLOCKCHAIN

**IT5712 FINAL YEAR PROJECT**

# PROJECT REPORT REVIEW 2

**Project Guide: Ms. Bala Gayathri D**

**Project Students: Pruthiev A S**

# Srevardhani R

# Yaswanth R



**DEPARTMENT OF INFORMATION TECHNOLOGY MADRAS INSTITUTE OF TECHNOLOGY CAMPUS ANNA UNIVERSITY, CHENNAI – 600044**

# AUG 2022 – DEC 2022

# Comprehensive Analysis of Related Work:

* Blockchain-Based Healthcare wherein they stored electronic data, which are shared and circulated into ETH blockchain-based systems for specific application domain in healthcare. Privacy is not achieved in this type of blockchain network thus security is breached.
* Identity-based Provable Data Possession from RSA Assumption for Secure Cloud Storage wherein the data is encrypted using RSA cryptographic algorithm which is stored in a distributed cloud environment. This centralized application suffers from data sharing and modification.
* Blockchain-Enabled Healthcare Applications, wherein it stores the medical data in a Hyperledger blockchain network using encryption techniques. This paper doesn’t provide any feedbacks to the users from their medical records.
* Provable Data Possession for Decentralized Storage wherein they Implemented blockchain based network using smart contracts. Privacy can be breached as raw data (without being encrypted) is stored.
* A general idea proposed wherein a blockchain network was implemented with provable data possession. This idea has to be extended in real life time application.
* Blockchain-based Remote Patient Monitoring in Healthcare 4.0. Implemented a blockchain network with an AI support to monitor patient health care. Privacy issues remain as the information of patients is stored in a raw plain text format.

**Problem Statement:**

In modern age, big data is essential in today's society in all industries. People can analyse, anticipate outcomes and make important decisions for various applications, like the stock market, disease prediction and provide Business Intelligence using the data obtained from data scraping, surveys, and research. These data are stored in SQL databases such as MySQL, PostgreSQL and NOSQL databases such as MongoDB, Cassandra and also in Big Data Framework such as Hadoop Distributed File System. All these data can be analysed using SQL, NoSQL, HIVE, Spark technologies.

In the modern medical industry, sensors and other medical electronic equipment such as ECG, thermometer etc that measure vitals like blood pressure, body temperature, heart beat and blood cells are used to record patient data. The information gathered helps to create a patient profile, which in turn helps doctors treat patients appropriately and helps users gain insights from their health. But the information collected is regionally specific and not worldwide. Thus, this may create problems in transparency, integrity and much more. These medical details of patients collected are stored in a simple database which appears to be centralized application which is prone to be modified or hacked by the intruders and also suffer from fault tolerance.

Our project aims to solve this problem using Blockchain and Data Possession methodologies. Blockchain helps to store multiple copies of data in decentralized fashion. Thus, a secured model could be implemented using blockchain in which the patient’s data can be securely sent to any medical professional, thus promoting transparency, security and integrity. Provable Data possession helps the users to verify that their information is intact and has not been tampered. It can be achieved using cryptography which encrypts the information of the patients stored. This project will be beneficial for the medical department and the public such that it will be beneficial in terms of transparency, scalability and security.

**Motivation of Project**

Web3 is the next iteration of the internet. It heavily relies on blockchain technology, machine learning, and artificial intelligence (AI). It aims to create a decentralized internet with open, connected, intelligent websites and web applications.

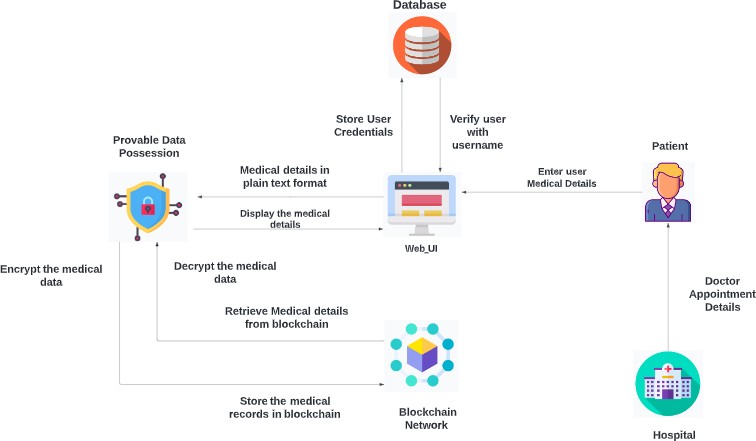
In Web 2.0, tech giants control and exploit user-generated medical data. In blockchain- powered web3, end-users take will take full ownership of the data using.

We will be able to choose what information you want to share with businesses and advertising companies and make money from it. Also, Web3 will not be controlled by a single entity. As a result, decentralized apps would not be censored nor will their access be restricted

**Scope of Project**

* Blockchain is decentralized which indicates that instead storing the entire medical data in cloud the data is broken into chunks which gets stored in the blocks.
* It is digital ledger of transactions that controls the entire blockchain medical data in a peer-to-peer network.
* Each computer or a node has a set complete medical records so if one or two nodes goes down does not result in data loss.
* The medical records that are present in blockchain are completely secured using encryption and it is possible to prove that no authorized entity has tampered the data.
* Because of its distributed nature, you can check file signatures across all the ledgers on all the nodes in the network and verify that they haven’t been changed
* Blockchains that restrict access to specific users within an organization. Thus, anyone accessing your blockchain must authenticate their identity to gain privileges and it will be restricted further for some particular transactions as well.

**Architecture Diagram**



* Here we propose a Blockchain based solution that stores the medical details of the patients which enhances security and integrity of the data. Provable Data Possession is implemented on top of blockchain. Provable Data Possession can be achieved using strong authentication, authorization and encryption techniques. Strong authentication and authorization can be achieved using Java Web Token (JWT) and two factor authentications mechanisms.
* Architecture diagram depicts how the user details gets stored in the Blockchain network. The hospital’s doctor and appointment details are collected, the user enters his/her medical details via a Web Application. The application checks the user credentials and then encrypts the data using his/her key and then stores the details in the blockchain network.
* When the user wants to retrieve his/her data, the data is retrieved from blockchain is decrypted to get the plain human readable text and is displayed.

**Baseline Requirements**

The basic requirement is indeed to collect medical details of patients such as,

1. Phone Number
2. Age
3. Blood Group
4. Blood Pressure
5. Appointment Date
6. Appointment Time
7. Address - State
8. Address – City
9. Interplanar File Storage (IPFS) Image hash

* These details can act as a contract structure variable.
* A constructor can be initialized to assign the values taken from the user.
* Getter and Setter functions can be implemented to store and retrieve the values from the blockchain.
* Most of the papers analyzed so far have stored only numerical and text data.
* But we implemented smart contracts in such a way it can store the image hash.
* The Image uploaded by the user is stored in the **Pinata** Cloud which is Interplanar File System (IPFS).
* The image can be stored and retrieved from **Pinata** Cloud using their API using the Image IPFS hash.
* The Image hash of the Image is stored in Blockchain.
* Each user’s health information is stored in blocks which has a contract address.
* When the user’s information is required, we access the details using his/her contract address.

**Interplanar File System**

* The current default way to exchange data across the Internet is HTTP, but it fails in some cases. Large files cannot be transferred using HTTP, data is not permanent on HTTP, HTTP mainly uses a Client-Server protocol which leads to low latency and makes it difficult to establish a peer-to-peer connection, also real-time media streaming is difficult on HTTP. All of these failures are overcome using IPFS.
* Unlike HTTP which is IP addressed, an IPFS network is content addressed. Which means, when any data is uploaded on an IPFS network, it returns a Hash and the data is then requested using that hash. Anyone can provide storage on the IPFS network and everyone is incentivized with crypto tokens. Data is distributed and replicated throughout the network which leads to data permanence. While requesting data it searches for the nearest copy of that data which leads to high latency and overcomes any bottleneck points. As the data is completely distributed, it has no scope for the centralization of data.
* Every node on the network is identified using a Node\_ID which is nothing but the hash of its public key. Everyone on the network can store files on their local storage and they are incentivized to do so. Each node maintains a DHT which is used to find out Ids of other peers on the network and what data those peers can serve.
* Storing Images in Blockchain is costly operation as it takes a lot of ethers and gas price and inefficient as well.
* So, in our application we have stored the images in IPFS and each Image is identified using an Image Hash.

**SQL Injection Attacks:**

SQL injection is a technique used to exploit user data through web page inputs by injecting SQL commands as statements. Basically, these statements can be used to manipulate the application’s web server by malicious users.

* SQL injection is a code injection technique that might destroy your database.
* SQL injection is one of the most common web hacking techniques.
* SQL injection is the placement of malicious code in SQL statements, via web page input.

Web servers communicate with database servers anytime they need to retrieve or store user data. SQL statements by the attacker are designed so that they can be executed while the web-server is fetching content from the application server. It compromises the security of a web application.

**Impact of SQL Injection**

The hacker can retrieve all the user-data present in the database such as user details, credit card information, social security numbers and can also gain access to protected areas like the administrator portal. It is also possible to delete the user data from the tables.

Nowadays, all online shopping applications, bank transactions use back-end database servers. So, in-case the hacker is able to exploit SQL injection, the entire server is compromised.

**First Order SQL Injection Attack:**

Let’s take this user query triggered from the Web UI

SELECT \* from STUDENT where STUDENT-ID == 12222345 or 1 = 1

Now this 1=1 will return all records for which this holds true. So basically, all the student data is compromised. Now the malicious user can also delete the student records in a similar fashion.

**Second Order Injection Attack:**

Let’s take this query

SELECT \* FROM Users WHERE UserId = 105; DROP TABLE Patients;

When this query is triggered even if there isn’t exist a User Id with 105 the entire table will be dropped thus the data can be lost which results in the loss of integrity.

**Impact of SQL Injection Attacks**

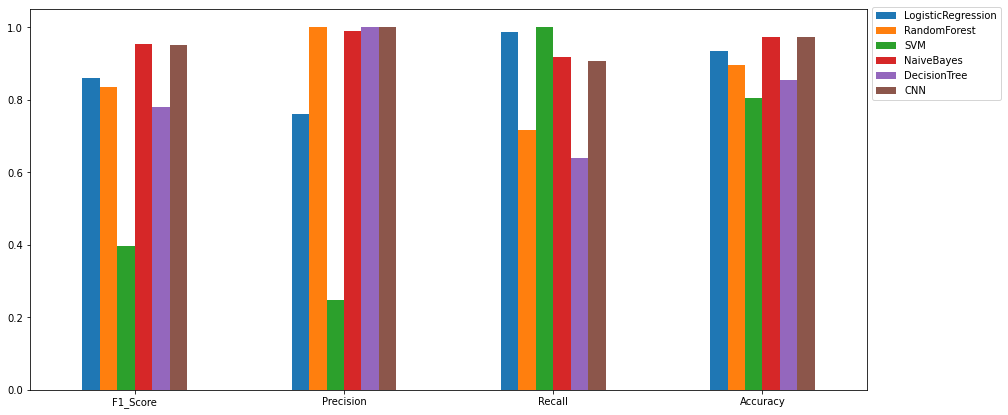
* If SQL Injection Attacks are triggered in our application the intruder can bypass the authentication using 1st order attack thus can change or modify the patient details.
* Using 2nd order Attacks, the intruder can either change or update patient’s details thus integrity is lost.

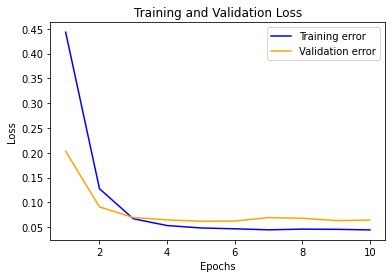
**Mitigation**

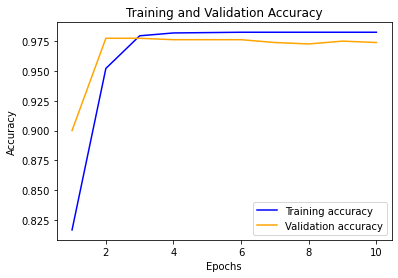
* We have proposed a machine learning based solution which can identify the user queries triggered and thus classify whether it is a valid or malicious query.
* The user query dataset consisting of valid and invalid user queries are trained on the machine Learning Algorithms and results obtained helps in mitigating this attack.

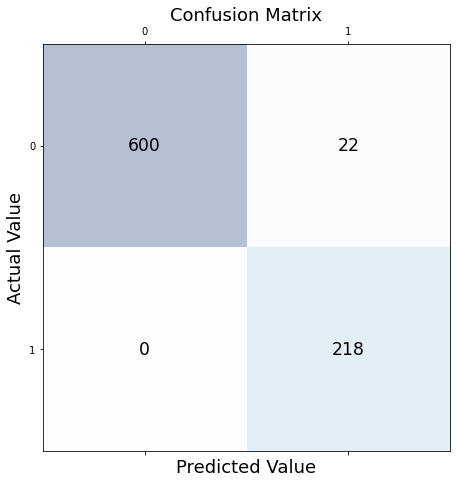
**Machine Learning:**

The user query dataset is taken from Kaggle and Machine Learning algorithms such as Logistic Regression, Random Forest, Support Vector Machine, Naïve Bayes, and ANN.









* The ANN has an accuracy of 97% which indicates that most of the user queries are classified correctly.
* This model is saved, whenever a user query is triggered this ML model can classify whether it is valid or malicious query.
* By this way the SQL Injection attacks are detected and mitigated from our website.
* This ML model can able to classify both the 1st order and 2nd order SQL Injection Attacks.

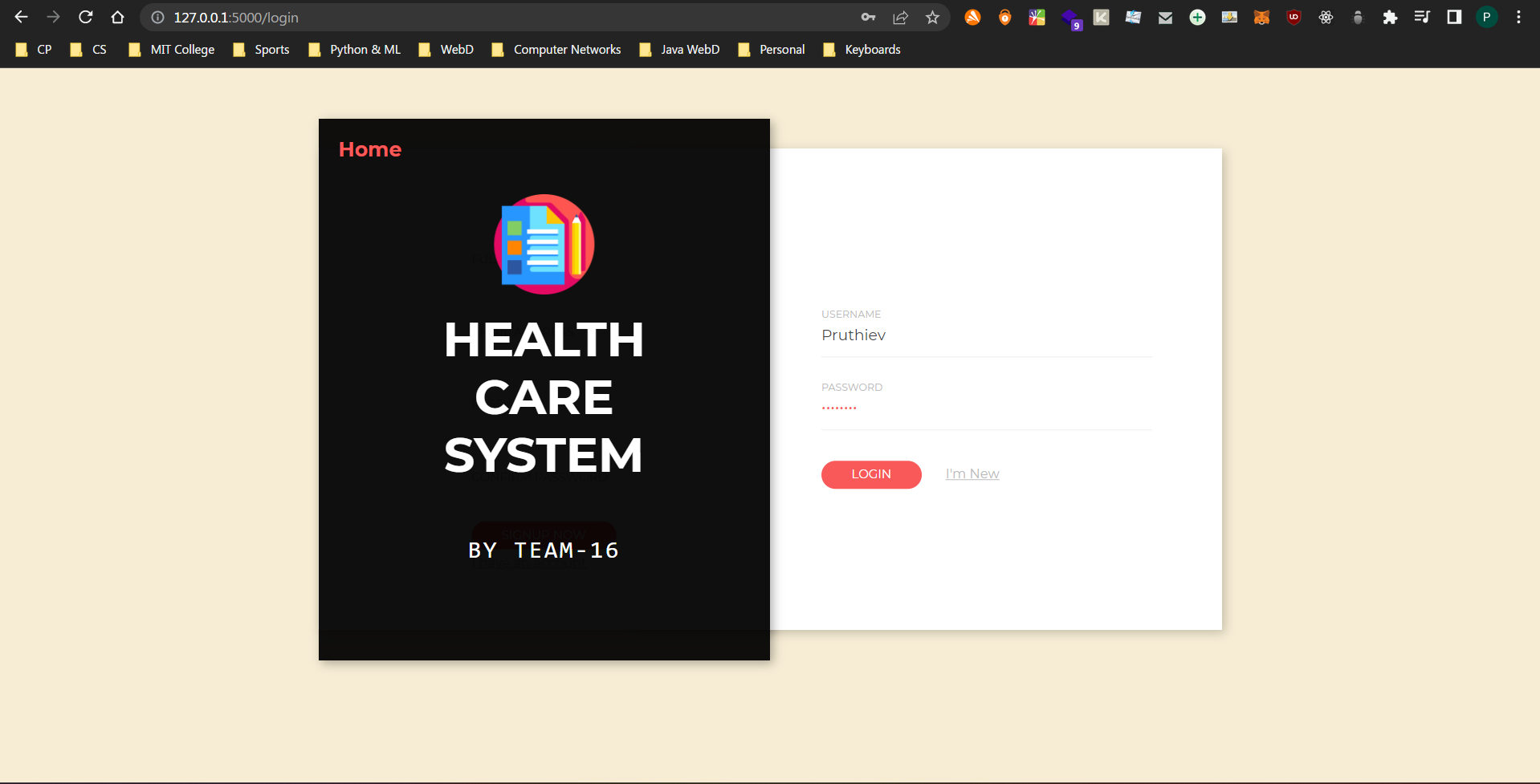
**Results:**

Enter correct username and password

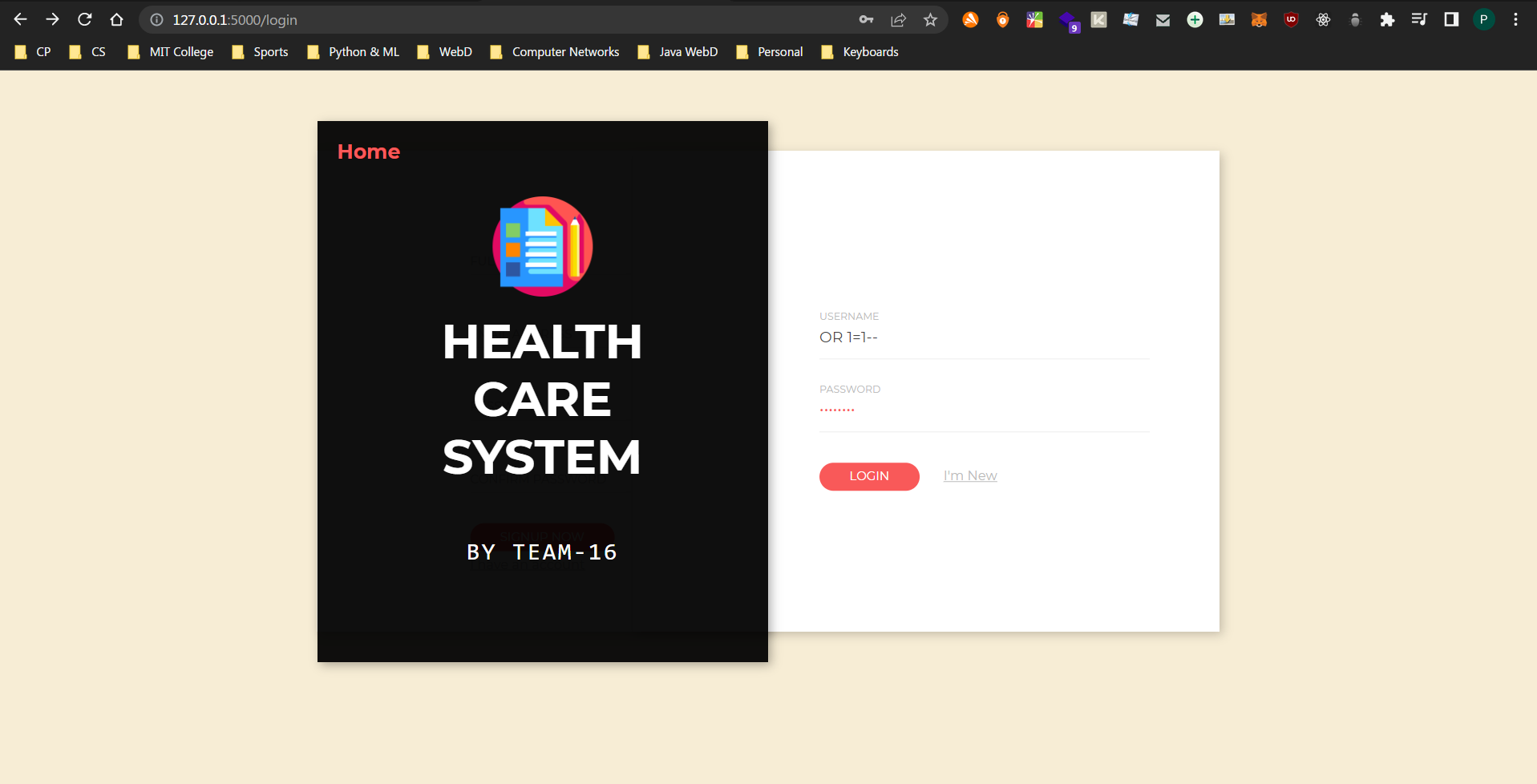
The SQL query would be

**SELECT \* FROM PATIENT WHERE username = “Pruthiev” AND Password = “\*\*\*\*\*\*\*\*”**

The user is authenticated and redirected to Skin Consult Form Page



**Injecting 1st Order SQL Attack – Bypassing the Authentication**

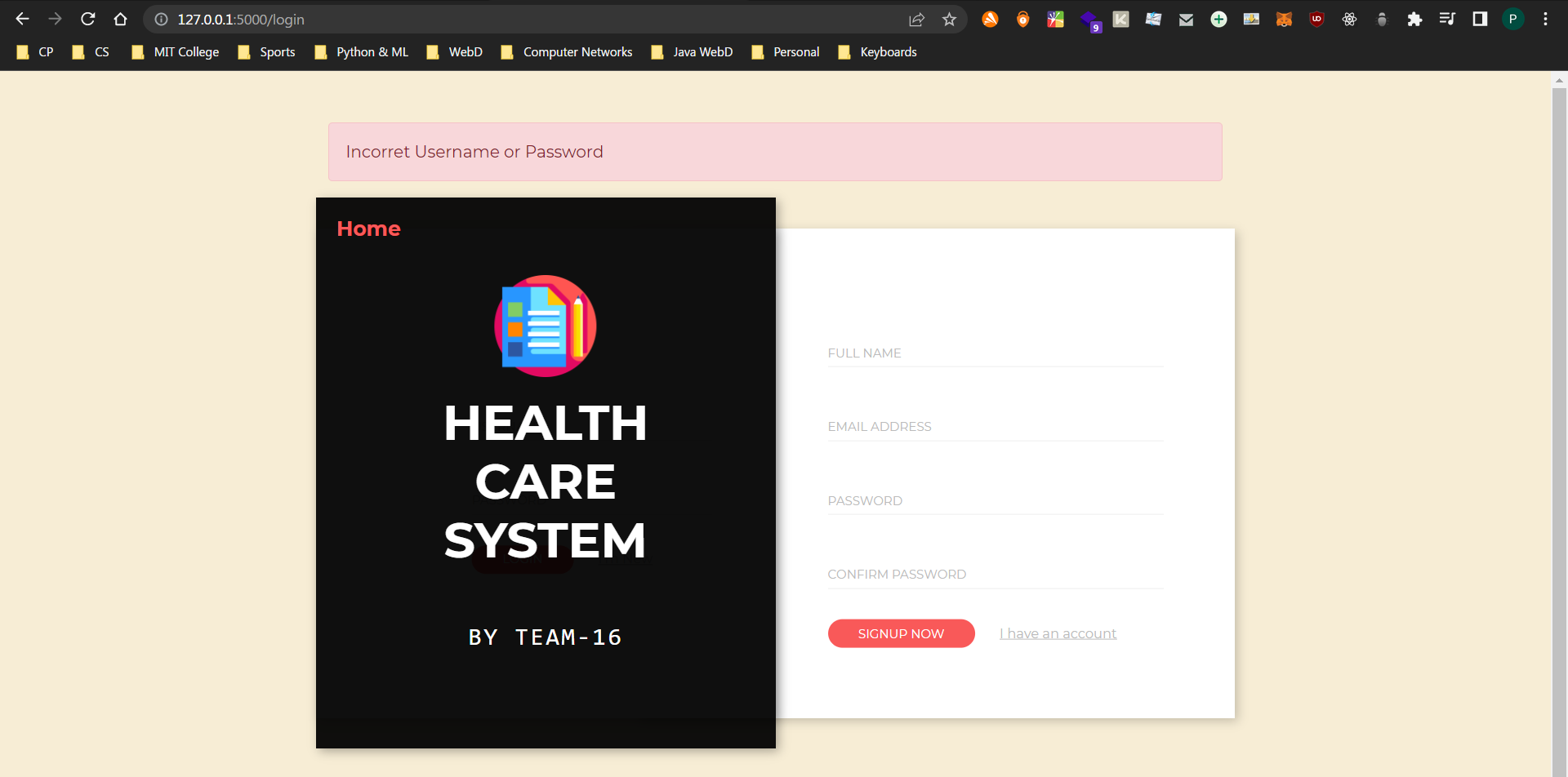


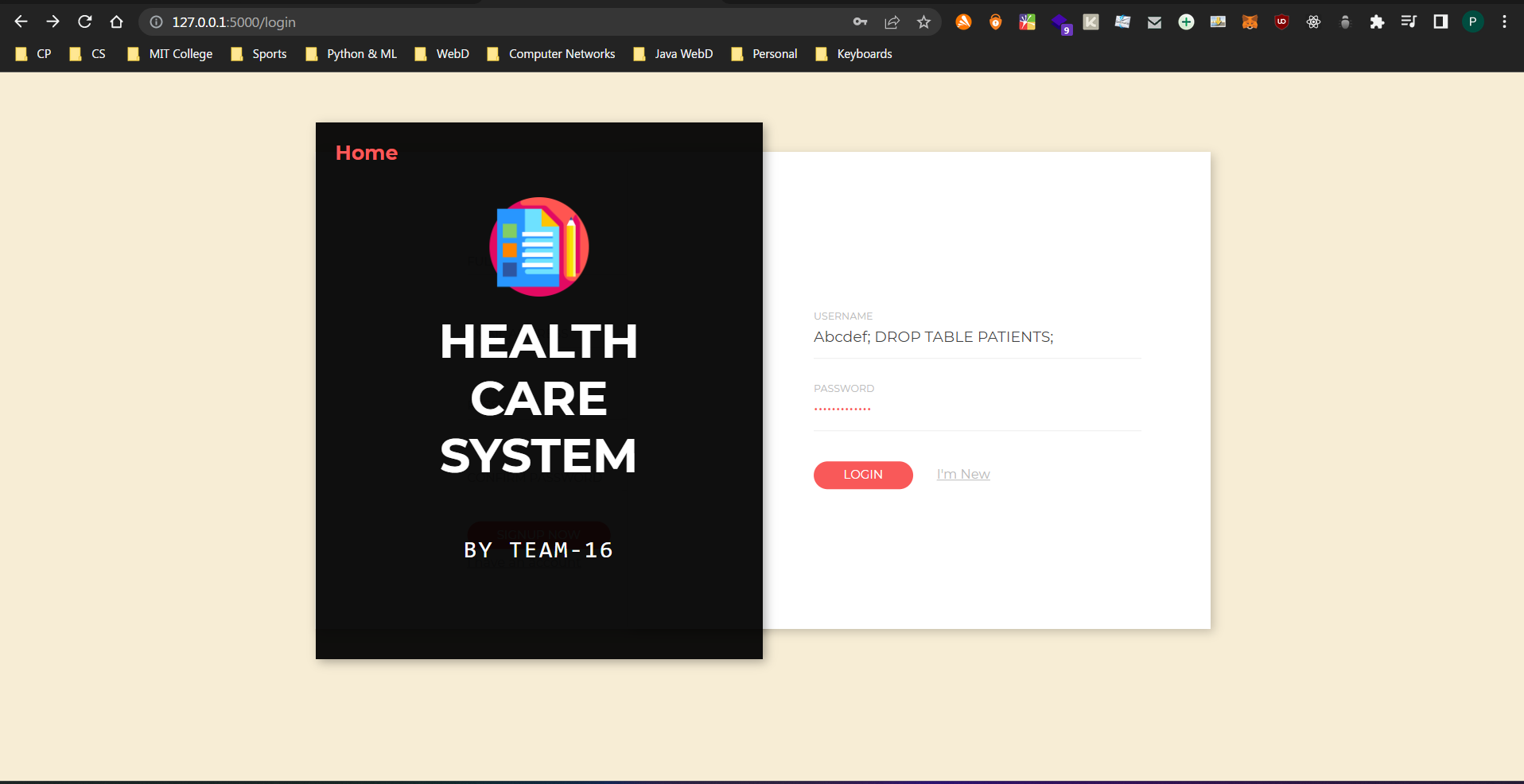
The query would be

**SELECT \* FROM PATIENTS WHERE username = OR 1=1 AND Password = “\*\*\*\*\*\*”**

This query would return the details and would have authenticated the intruder.

Since we have used ML to mitigate the attack it classifies it as malicious query and returns

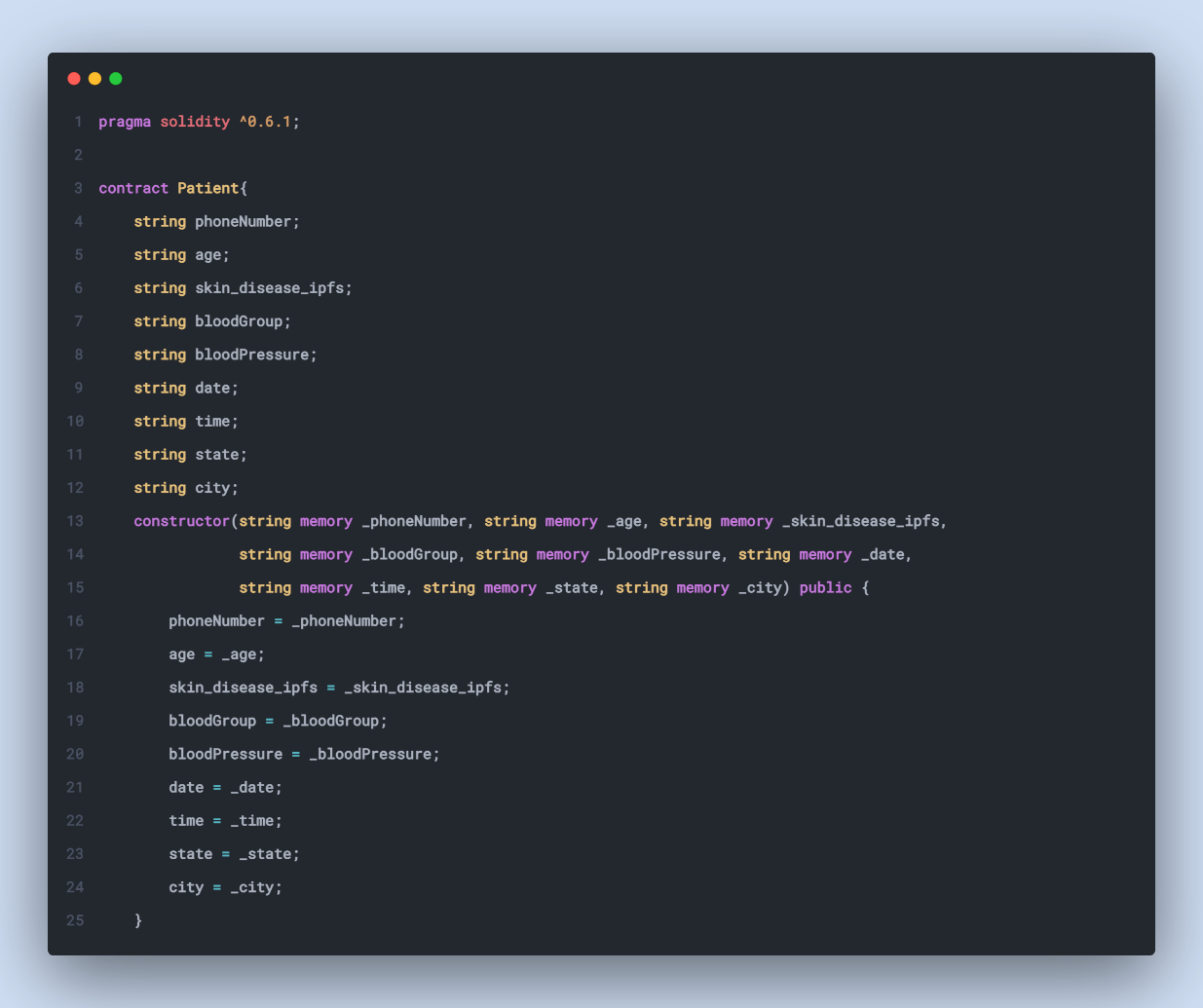


**Injecting 2nd Order SQL Attack – Destroying the Integrity. **

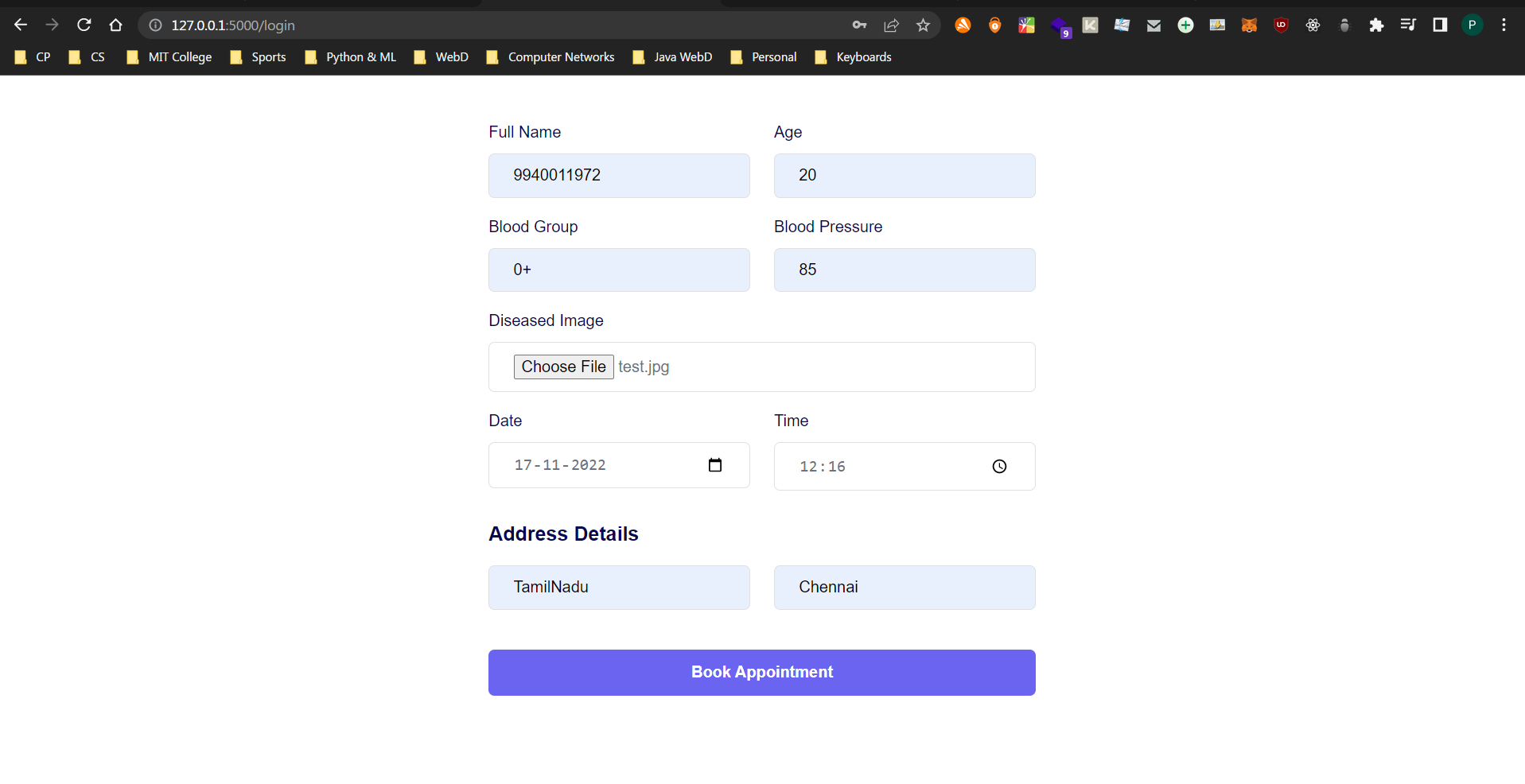
SQL: **SELECT \* FROM PATIENTS WHERE username = “Abcdef”; DROP TABLE PATIENTS; AND Password = “\*\*\*\*\*\*\*\*\*\*\*”.**

Even if the username or password is incorrect this can destroy the database by deleting all the information from the users. Thus, all their blockchain contract details are lost.

**Smart Contract:**

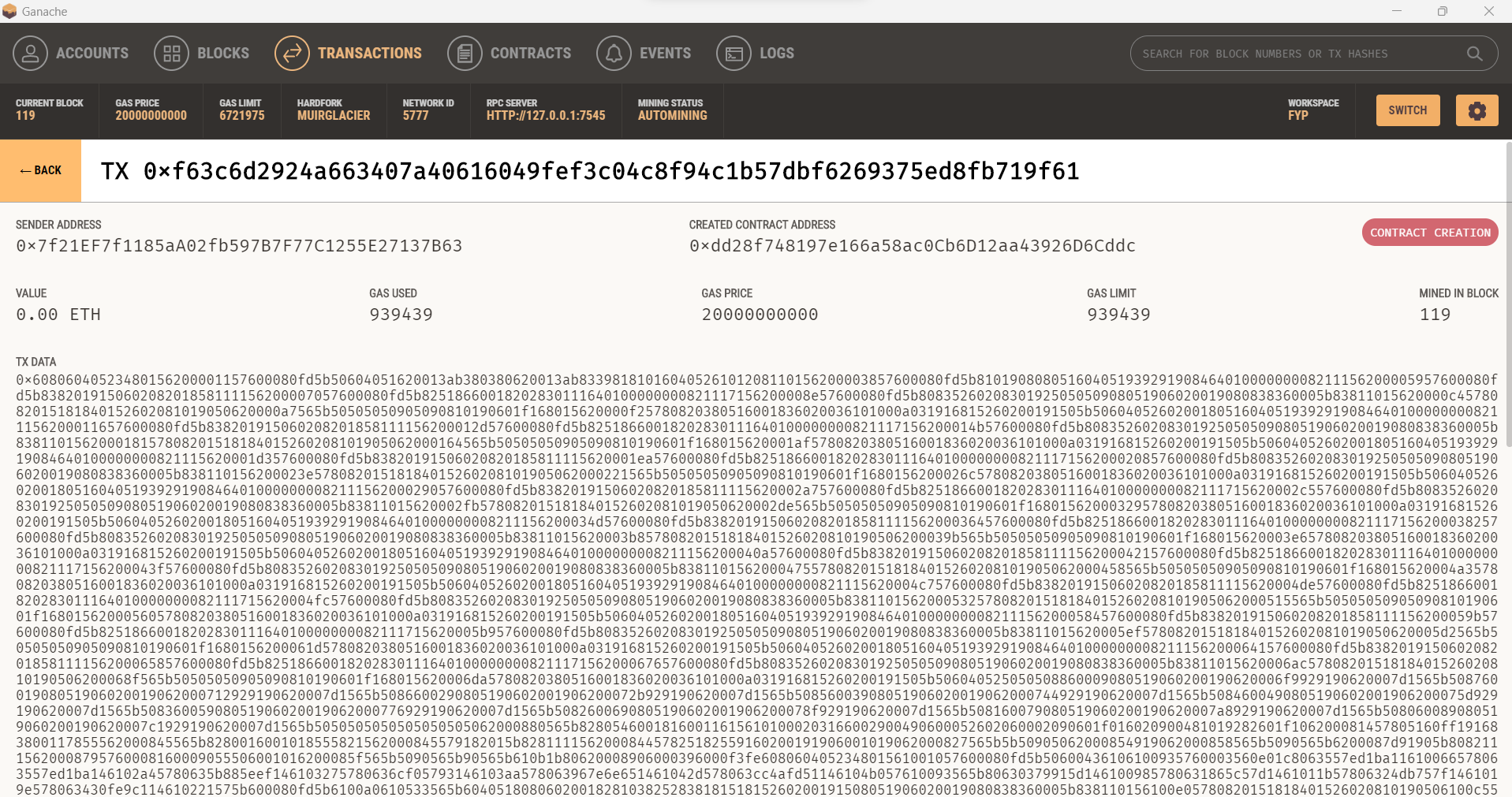


**Storing Images in IPFS:**

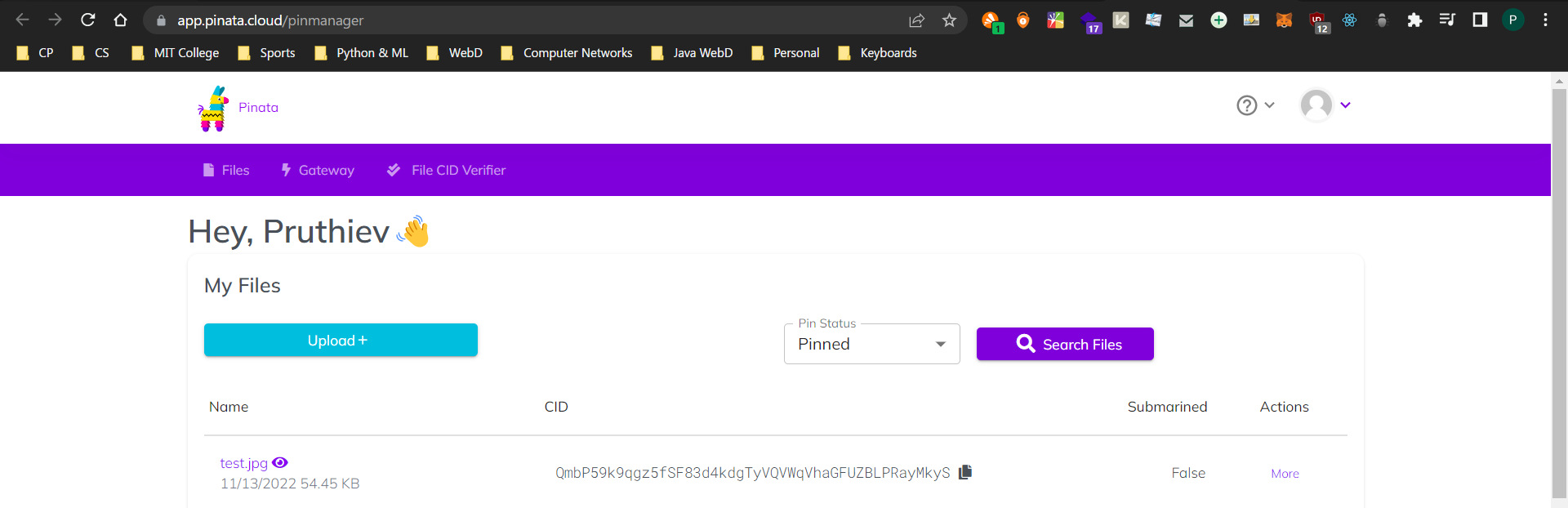


* This form collects the user details from the patients and stores the details except image in Ethereum Blockchain Network.
* The Image is stored in Pinata IPFS Cloud and the Image IPFS hash indeed is stored in blockchain.

The user details are stored in the blockchain and at 119th block.



The Image **test.png** is stored in Pinata IPFS Cloud.



**Future Plan:**

* The Image stored in IPFS Pinata Cloud is to be secured using strong cryptographic

Algorithm.

* Build Frontend pages for Heart and Cancer Patients by storing their medical records in blockchain and IPFS.

**REFERENCES:**

* Huaqun Wang, Qihua Wang, Debiao He, ‘Blockchain-Based Private Provable Data Possession”, IEEE Transactions on Dependable and Secure Computing, 2022
* Leili Soltanisehat, Reza Alizadeh, Haijing Hao, and Kim-Kwang Raymond Choo, “Technical, Temporal, and Spatial Research Challenges and Opportunities in Blockchain-Based Healthcare: A Systematic Literature Review”, Ieee Transactions on Engineering Management, 2021
* Jigna Hathaliya, Priyanka Sharma, Sudeep Tanwar et all “Blockchain-based Remote Patient Monitoring in Healthcare 4.0”, IEEE International Advance Computing Conference, IACC, 2019
* Jianbing Ni, Kuan Zhang, Tingting Yang, “Identity-based Provable Data Possession from RSA Assumption for Secure Cloud Storage”.
* Yannan Li et. all “Provable Data Possession for Decentralized Storage”, IEEE Journal on selected areas in communications, 2020