|  |
| --- |
| TECHNICAL REPORT |

|  |
| --- |
| **Distributed and Scalable Data Engineering**  **(DSCI-6007)** |

|  |  |
| --- | --- |
| SPRING 23 |  |



Contents

[Project Name 2](#_Toc96341549)

[Executive Summary 2](#_Toc96341550)

[Technical Report 3](#_Toc96341551)

[Highlights of Project 3](file:///C:\Users\ardia\Desktop\TEACHING\SPRING%2022\DSCI6002%20Data%20Science\Projects\Technical%20report%20Template.docx#_Toc96341552)

[Submitted on: 3](file:///C:\Users\ardia\Desktop\TEACHING\SPRING%2022\DSCI6002%20Data%20Science\Projects\Technical%20report%20Template.docx#_Toc96341553)

[Abstract 4](#_Toc96341554)

[Methodology 5](#_Toc96341555)

[Results Section 9](#_Toc96341556)

[Discussion](#_Toc96341557) 11

[Conclusion 11](#_Toc96341558)

[Contributions/References 12](#_Toc96341559)

|  |
| --- |
| AMENABILITY AIDING |

|  |  |  |
| --- | --- | --- |
| Executive Summary Any business regularly performs vulnerability evaluations and fixes its apps with security upgrades. This increases the likelihood of attackers taking advantage of the application. In order to fix this problem, we created a specially designed website that enables us to choose a vulnerability automatically in any application and directs us to all the available information about the chosen vulnerability.  Our goal is to increase the security team's efficiency while never skipping a patching or upgrading operation. | | |
| person at a table writing in a notebook with people around | | |
| **Team Members:**  **Avinash Bhavanam Reddy**  **(Application Developer)**  **Sri Harsha Linga**  **(Data Scientist)**  **Lavanyaa Murali**  **(Data Analyst)** | **Questions?**  Contact: abhav3@unh.newhaven.edu |  |

|  |
| --- |
|  |

|  |  |
| --- | --- |
| **Amenability Aiding** |  |
| Highlights of Project  * We would have multiple vulnerabilities, thus in order to improve efficiency, we will offer a single source that can be accessed. * **Since resources would be static, we structured the pipeline in such a way that it enabled the database to be updated automatically.**  Submitted on: 04/30/2023 |

## Abstract

Organizations can find and fix security flaws with the use of vulnerability scanning and patching before attackers can take advantage of them. Every firm should regularly do security updates and vulnerability scanning. One application's inefficiency allows attackers to take advantage of it.

**A scheduling approach to prevent security vulnerabilities**

Introductory Section

* Each company regularly performs vulnerability scanning and patches its apps with security upgrades.
* Patching and vulnerability scanning are two crucial security procedures that can assist organizations in locating and addressing security flaws before attackers can take advantage of them.
* While patching can help to fix those vulnerabilities and thwart attacks, vulnerability scanning can assist in identifying potential vulnerabilities.
* We might encounter n different vulnerabilities for a single application, making it ineffective to provide security upgrades.
* This increases the likelihood of attackers taking advantage of the application.
* To solve this problem, we created a specially designed website that enables us to choose a vulnerability manually in any application and directs us to all the available information about the selected vulnerability.

Data source Link

Through the following link, web scraping is being used to collect data.

<https://www.cisa.gov/known-exploited-vulnerabilities-catalog>

## 

## Methodology

The curriculum employs the CRISP DM methodology, which covers subjects including business understanding, data understanding, data preparation, modeling, evaluation, and model deployment.

Business Understanding

* As a data scientist, it is crucial to understand the type of business your company is in, the industry it operates in, how it functions, and basically everything else related to the business. Only by having a thorough understanding of the business will you be able to pinpoint its current challenges, assess them, and find a solution or solutions, as well as identify the tactics to achieve the company's objectives.
* Our goal is to increase the security team's efficiency while never skipping a patching or upgrading operation.

After using web scraping techniques to get the required data, we put it into data frames using the Pandas library and then extracted only the data, fields, and attributes that were essential to our project.

Data Understanding

Data Preparation

Knowing what data, you already have, where to get more data, how much data is available, and what tools to employ to get it are all necessary during this phase. Your data science project will make more sense if you comprehend your data from the beginning

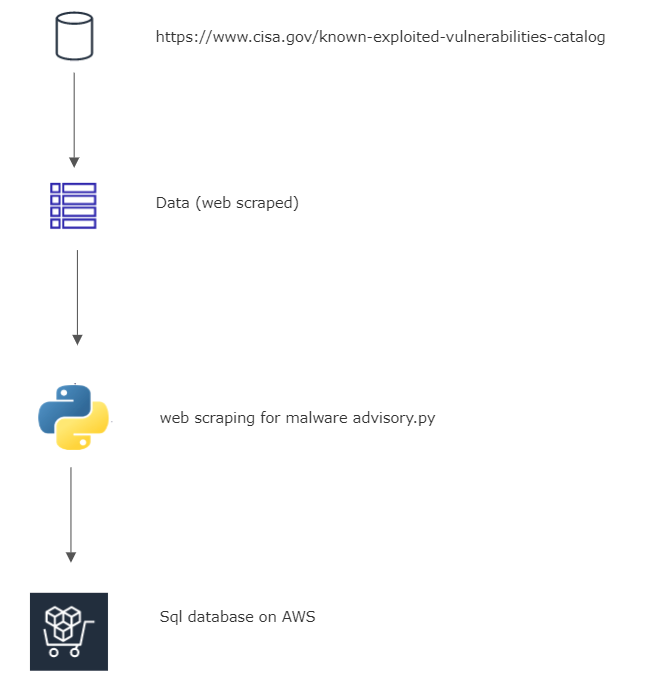
In our Project,

* Data like CVE IDs, network topologies, access controls, user rights, and known exploits or threats must be gathered for this project.
* Additionally, security incident logs, penetration test findings, and vulnerability scan reports can all offer insightful information about potential flaws that need to be fixed.
* We have used web scraping methods to collect this data.

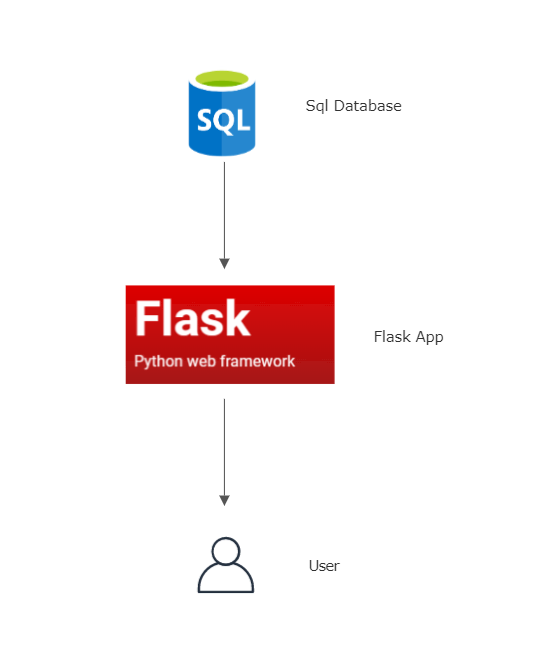
**Modeling and Evaluation**

* Evaluation results are obtained by assessing the degree to which the model meets the objectives of the business and testing the models on test applications if time and budget permit.
* Although we are not now using modeling or assessment methodologies in our research, we want to in the future in order to increase the accuracy and validity of our conclusions.

## Data Pipeline Architecture



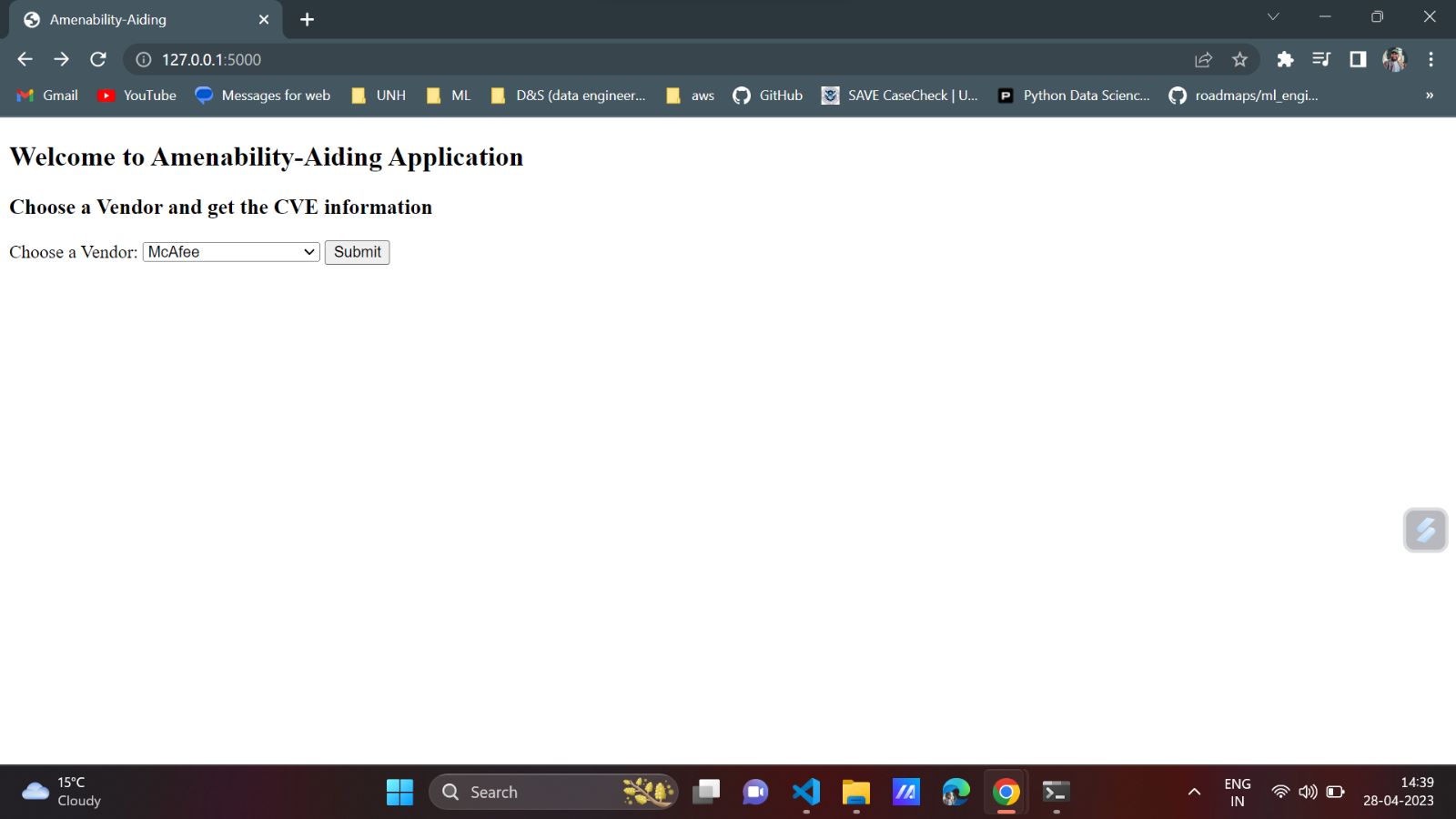
Application Architecture

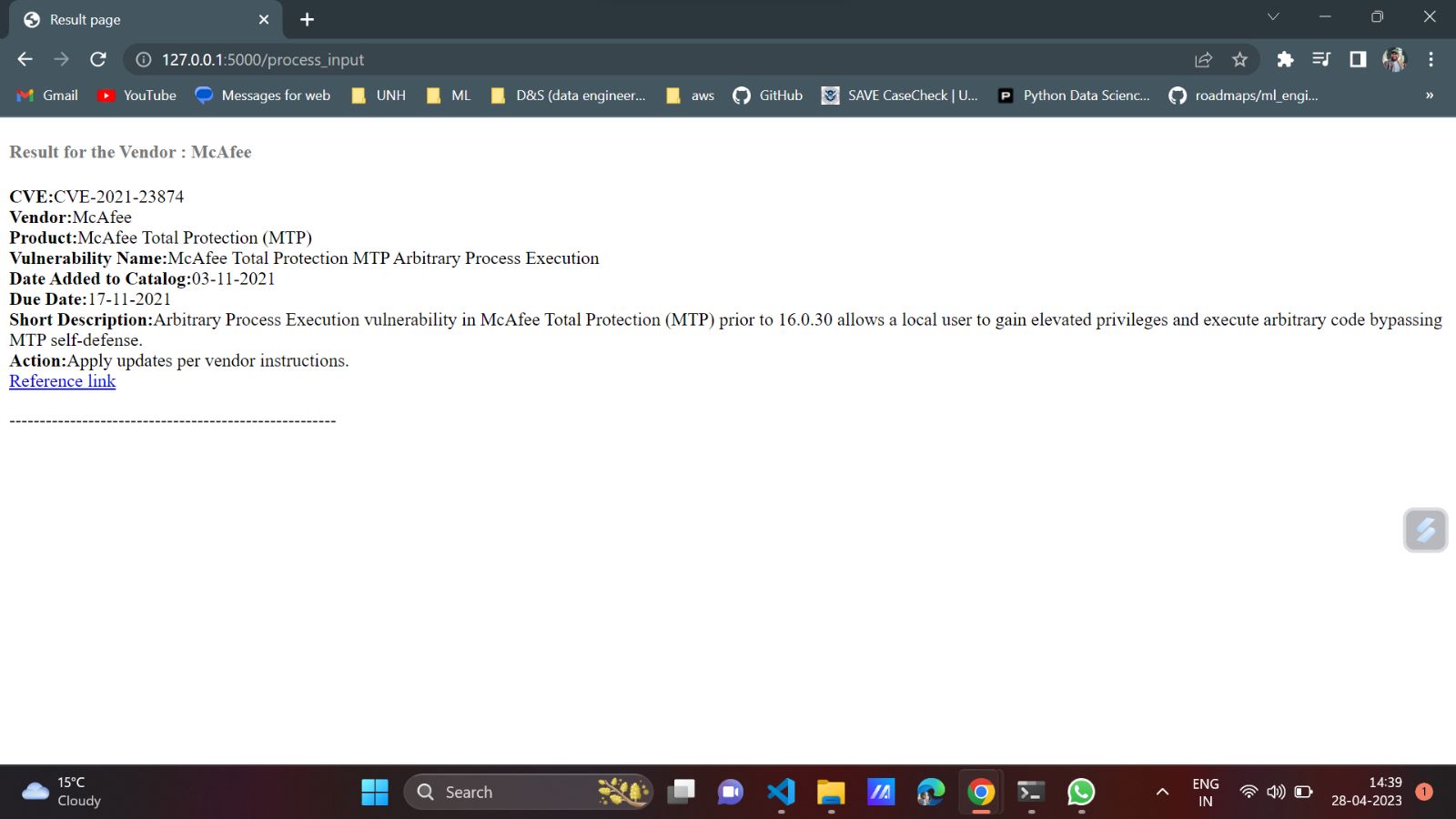


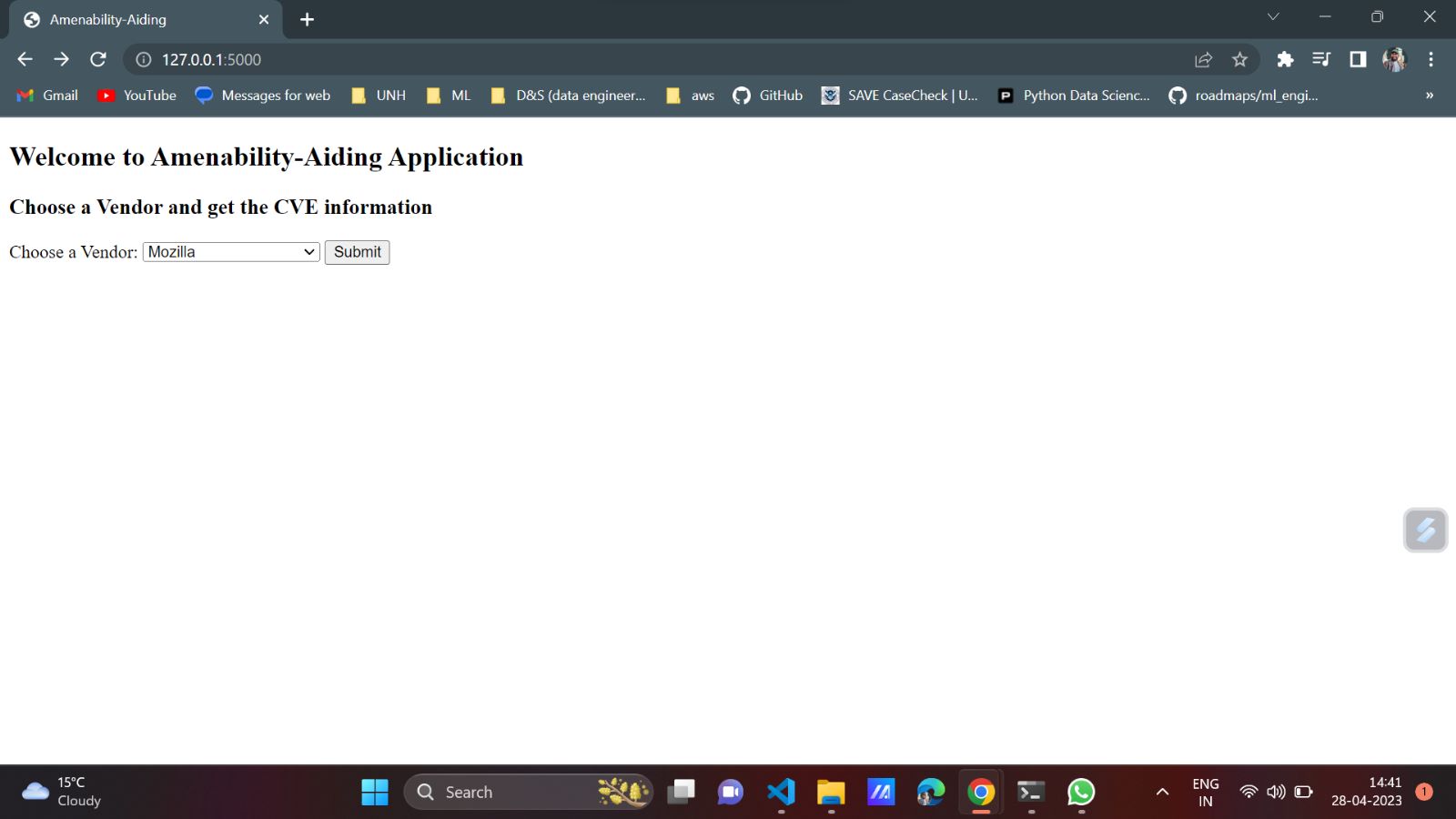
## Results Section

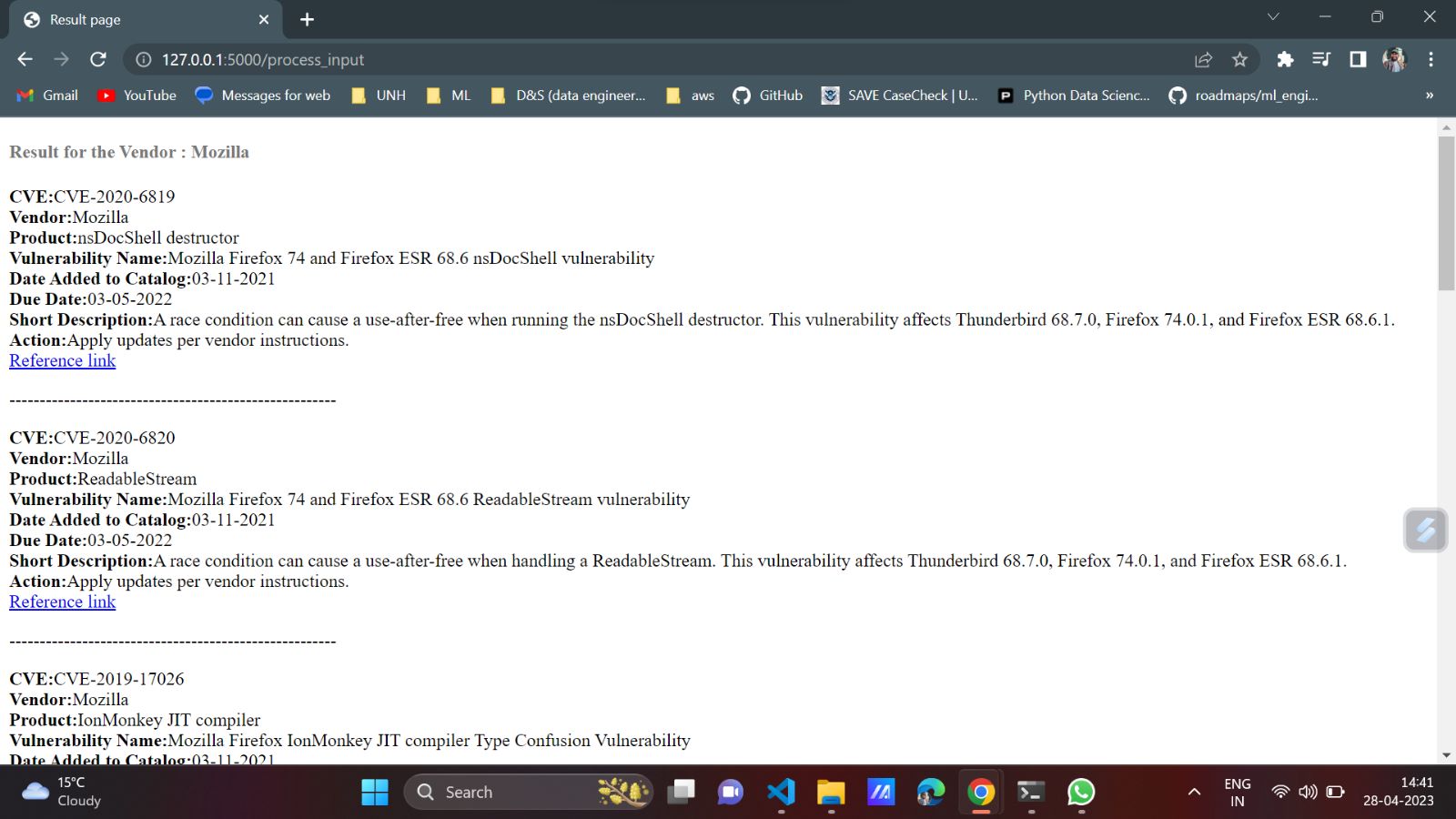
The output is deployed

<http://ec2-34-202-51-65.compute-1.amazonaws.com>









## Discussion

First, we developed an architecture to address the issue, but this was a technical failure in many aspects. Finally, we concluded that this appeared to be the most effective answer to the problem.

## Conclusion

With the use of this information, machine learning models may be developed and used in real-world situations. The data will automatically be updated as a result of the web scraping technique we utilized, and the Machine Learning model will also be updated appropriately.

## 

## Contributions/References

<https://www.cisa.gov/known-exploited-vulnerabilities-catalog>

<https://medium.com/techfront/step-by-step-visual-guide-on-deploying-a-flask-application-on-aws-ec2-8e3e8b82c4f7>

<https://aws.amazon.com/getting-started/hands-on/create-microsoft-sql-db/>