Mini Project
Phase-3

~ Sejal Gurkhe (2019130017)

~ Ishita More (2019130039)

~ Under the Guidance of Prof. Kiran Gawande

"Fraud Web Page Detection using Machine Learning"

Problem Statement

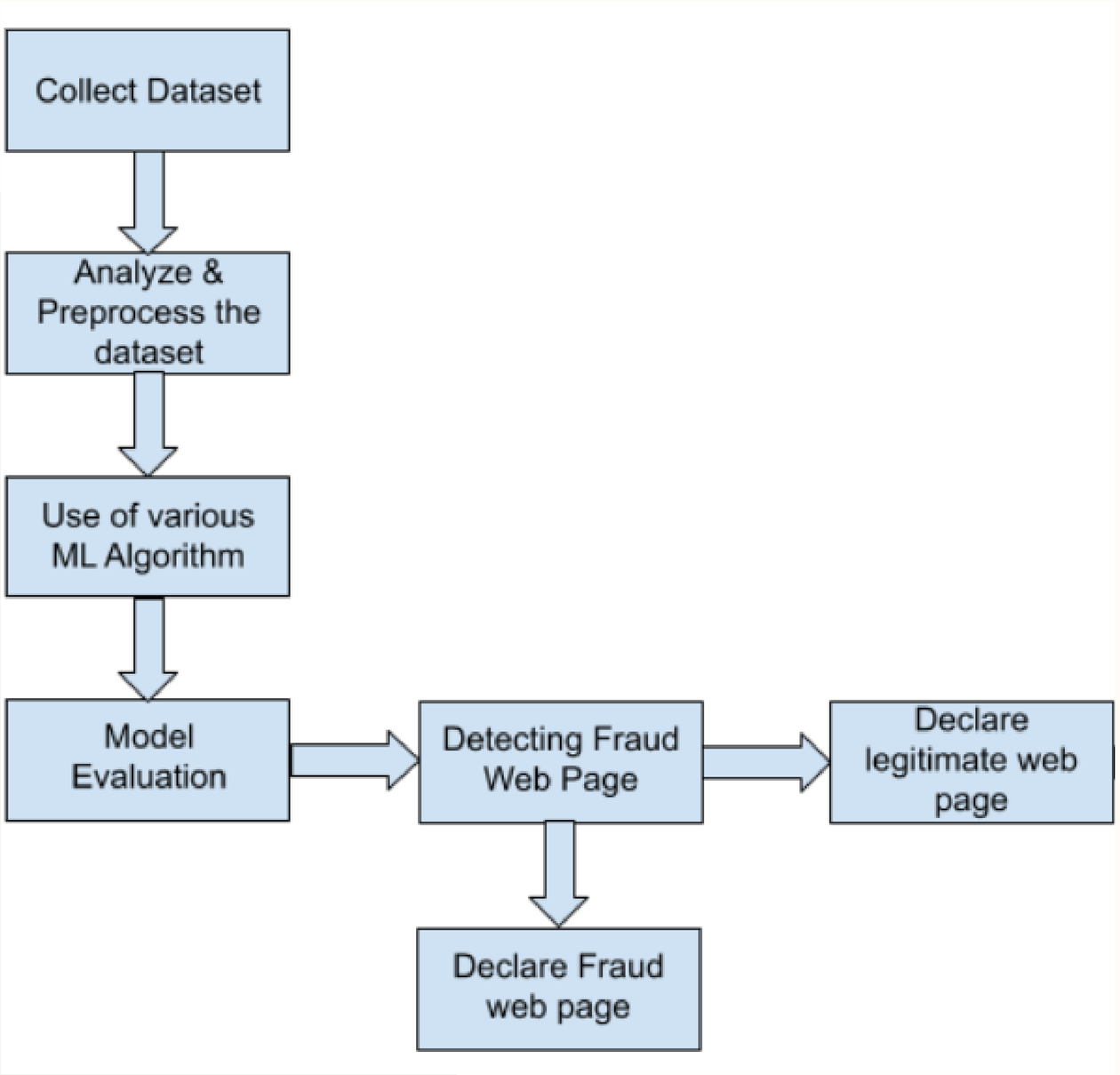
- Phishing Websites are duplicate Web pages created to mimic real Websites in-order to deceive people to get their personal information.
- Because of the adaptability of their tactics with little cost Detecting and identifying.
- The purpose of the project is to detect fraud websites that might encounter.
- The task is to determine if a user still can be trusted or if it should be flagged for potential fraud through that website.

Objectives:

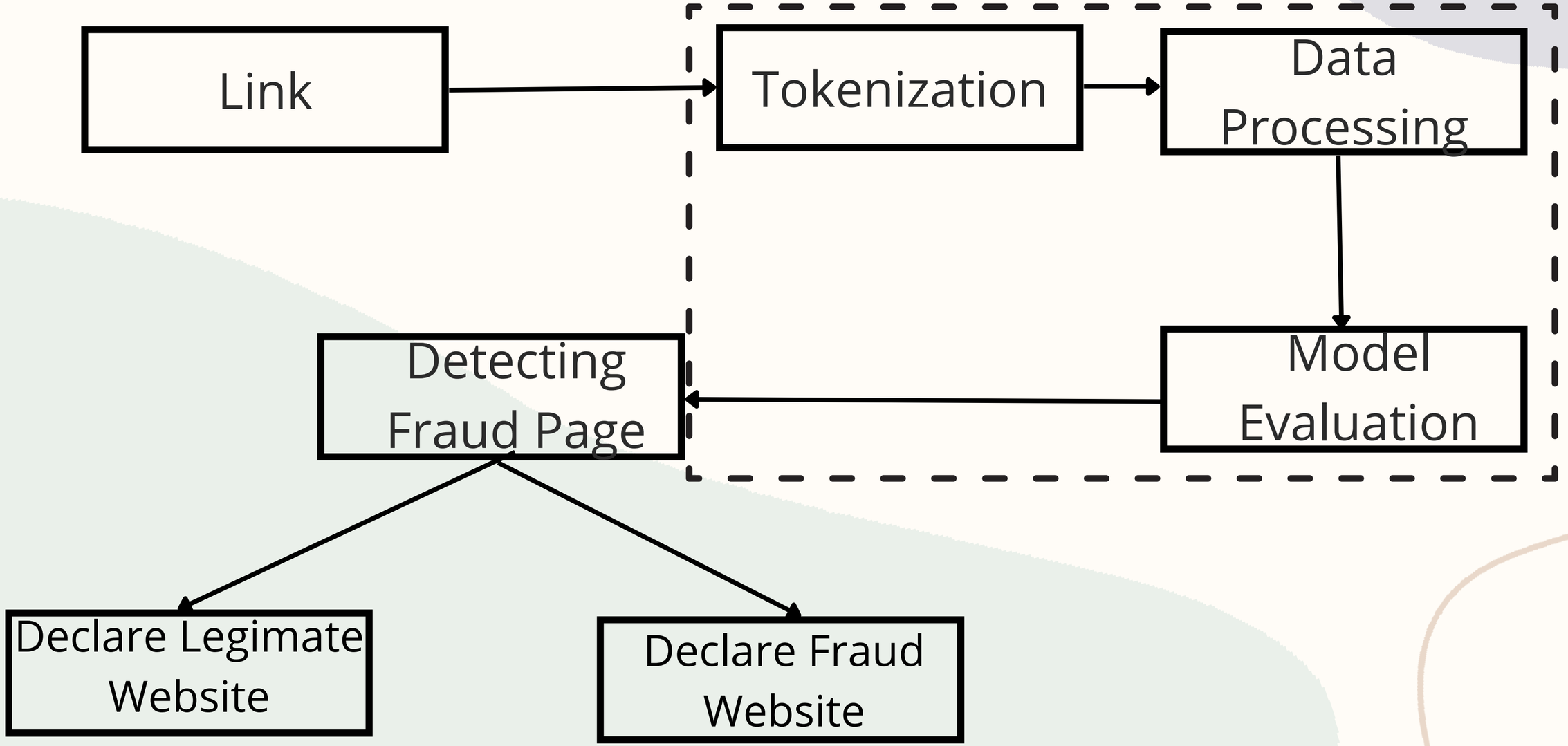
The objective of this project is

- To train machine learning models on the dataset created to predict fraud web pages.
- To use the dataset of fraud websites data and from that data use it to detect Fraud Web pages.
- The performance level of each algorithm is measured and compared.
- The objective of phishing website URLs is to purloin the personal information like user names, passwords and etc.

Design and Methodology



Methodology



Algorithms/techniques used

- Logistic Regression

```
Training Accuracy : 0.9783572688946115
Testing Accuracy : 0.963833489882551 7

CLASSIFICATION REPORT

              precision    recall  f1-score   support

         Bad       0.91      0.96      0.93     36442
        Good       0.99      0.96      0.98    100895

    accuracy                           0.96    137337
   macro avg       0.95      0.96      0.95    137337
weighted avg       0.97      0.96      0.96    137337
```

Algorithms/techniques used

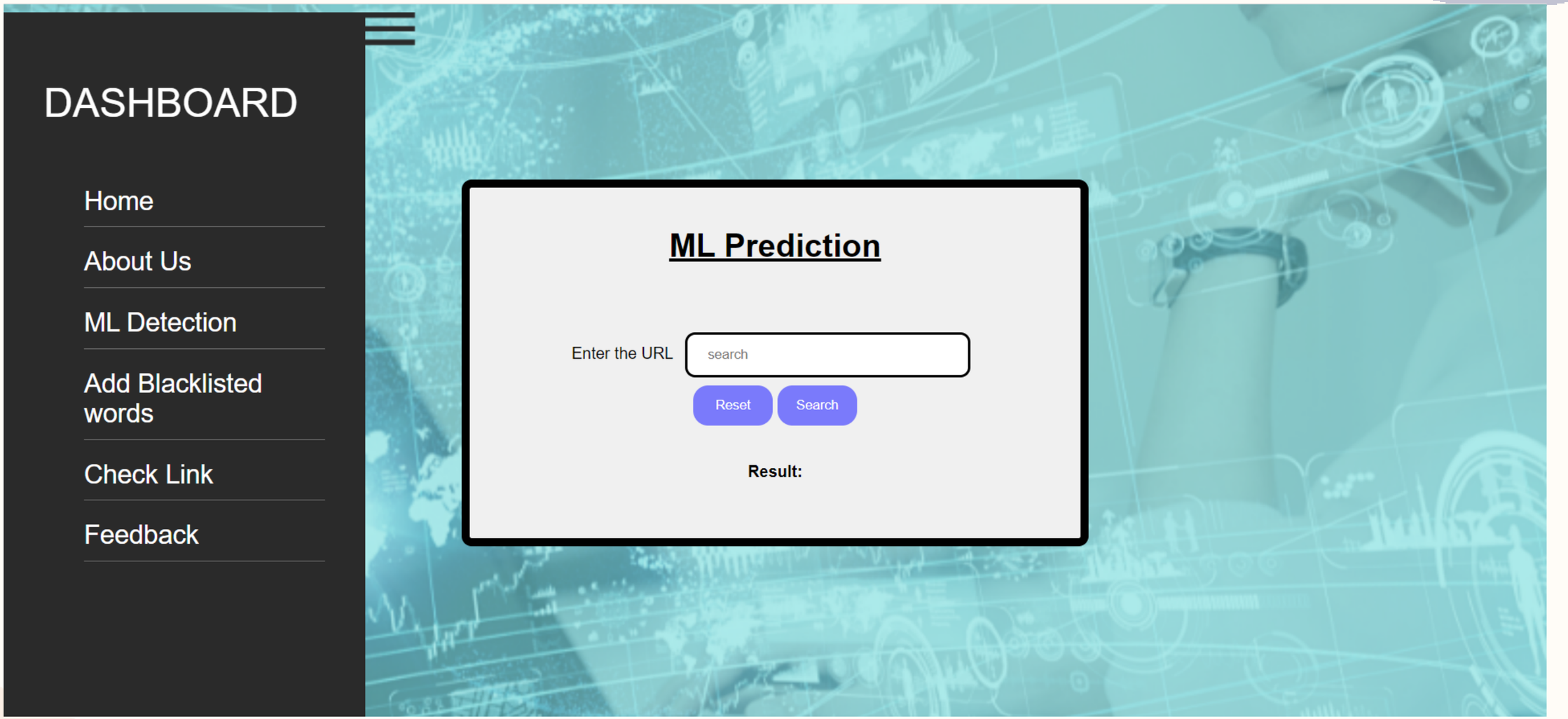
- Multinomial Naive Bayes classifier object using MultinomialNB() function.

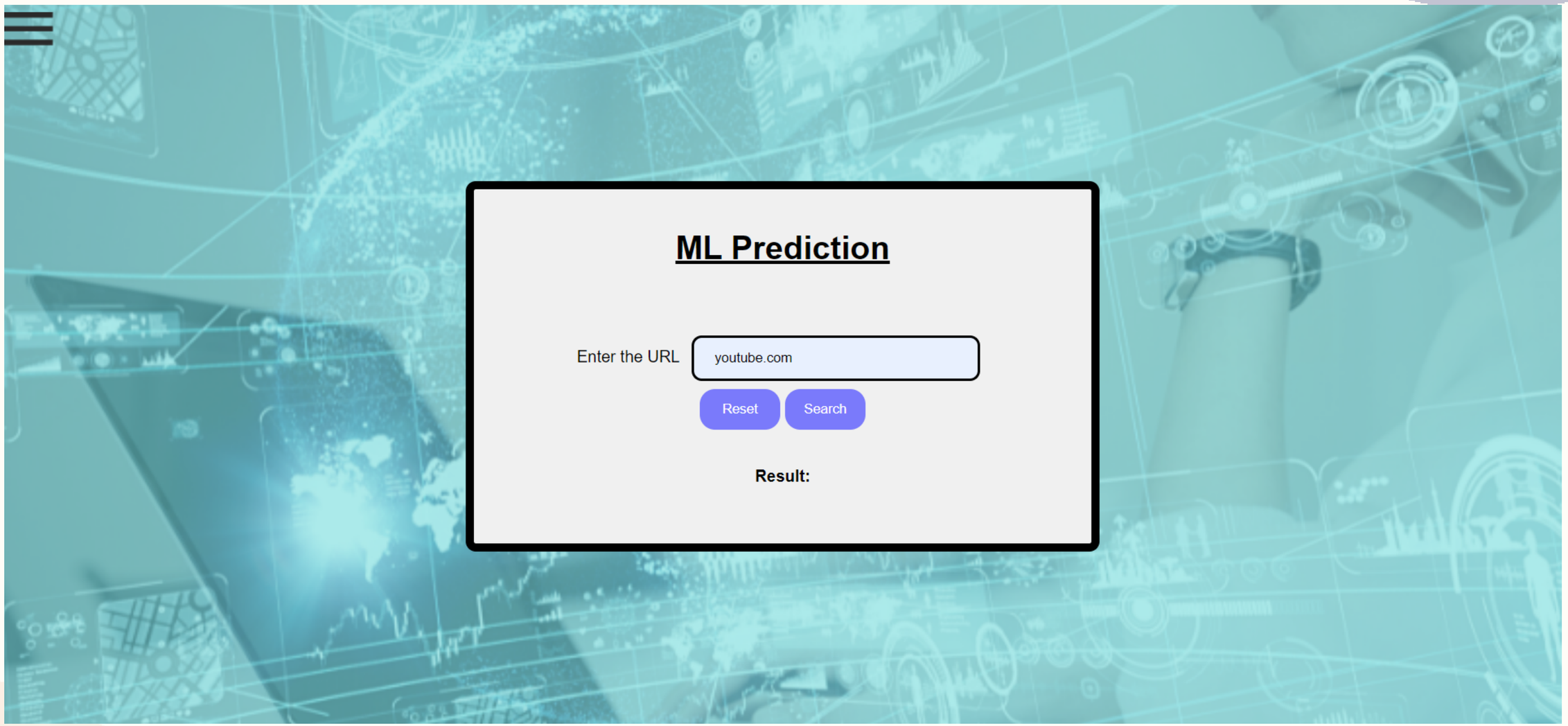
```
Training Accuracy : 0.9739204726110352
Testing Accuracy : 0.9574477380458288

CLASSIFICATION REPORT

              precision    recall  f1-score   support

         Bad       0.91      0.93      0.92     38035
        Good       0.97      0.97      0.97     99302

    accuracy                           0.96    137337
   macro avg       0.94      0.95      0.95    137337
weighted avg       0.96      0.96      0.96    137337
```

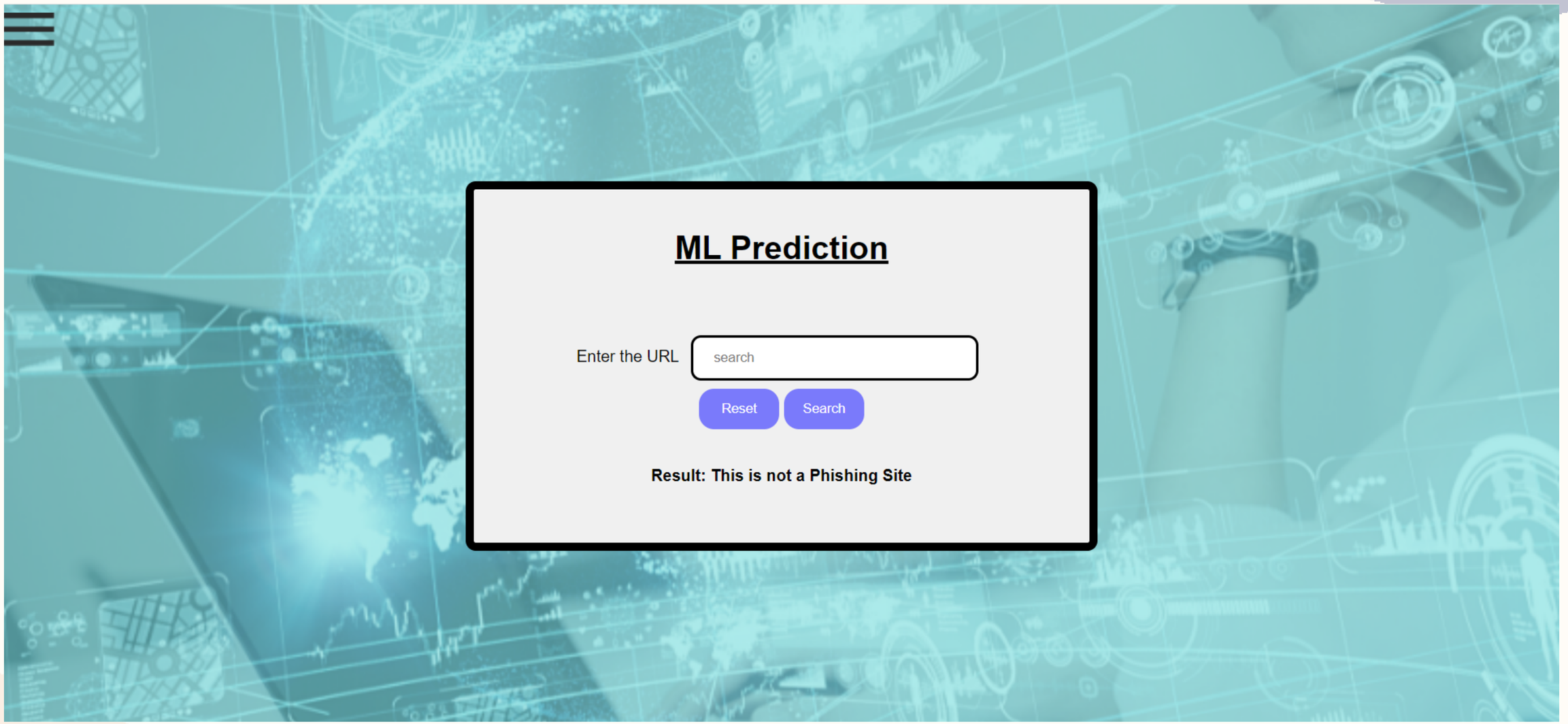
Tech Stack Used

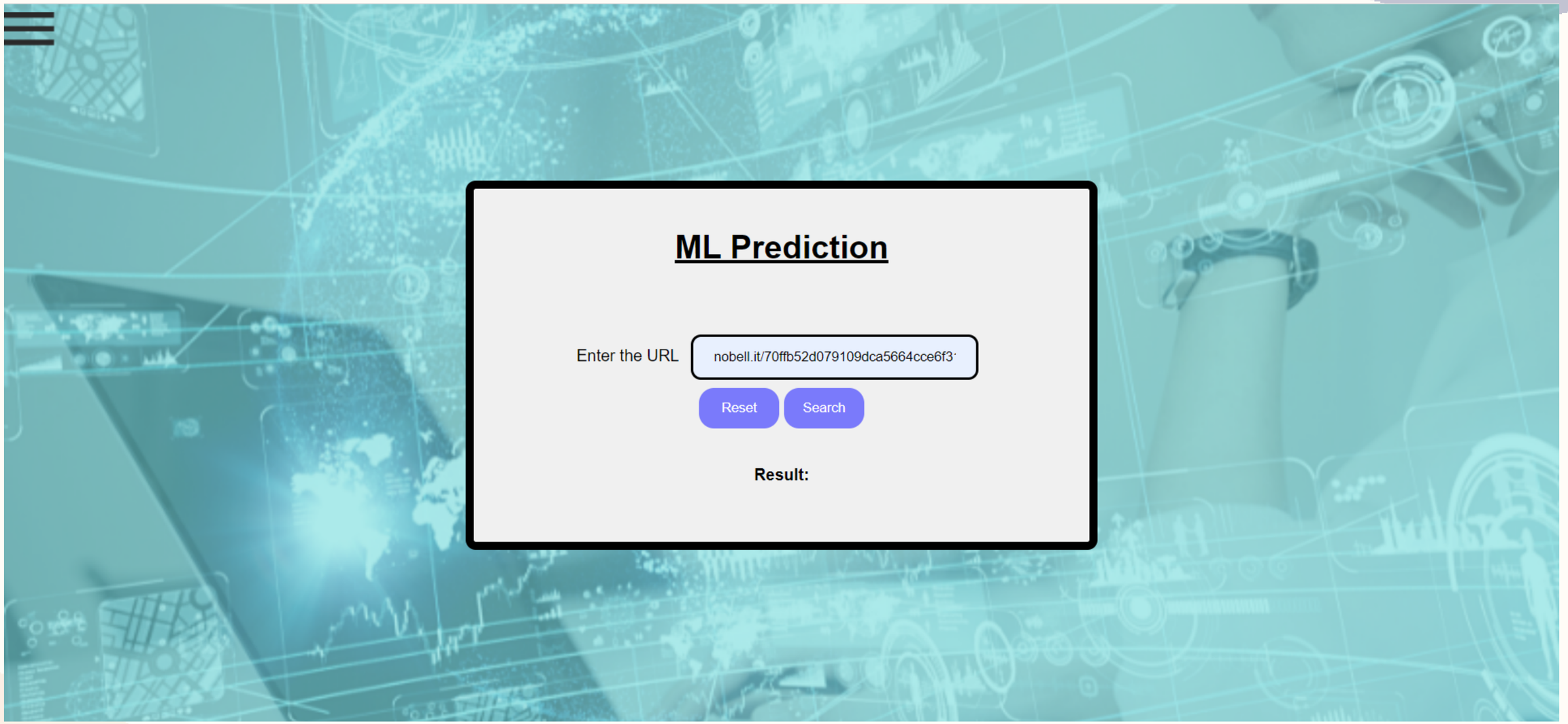
Kaggle(CSV File)
Google Colab (Jupyter)
Python
HTML / CSS
Flask

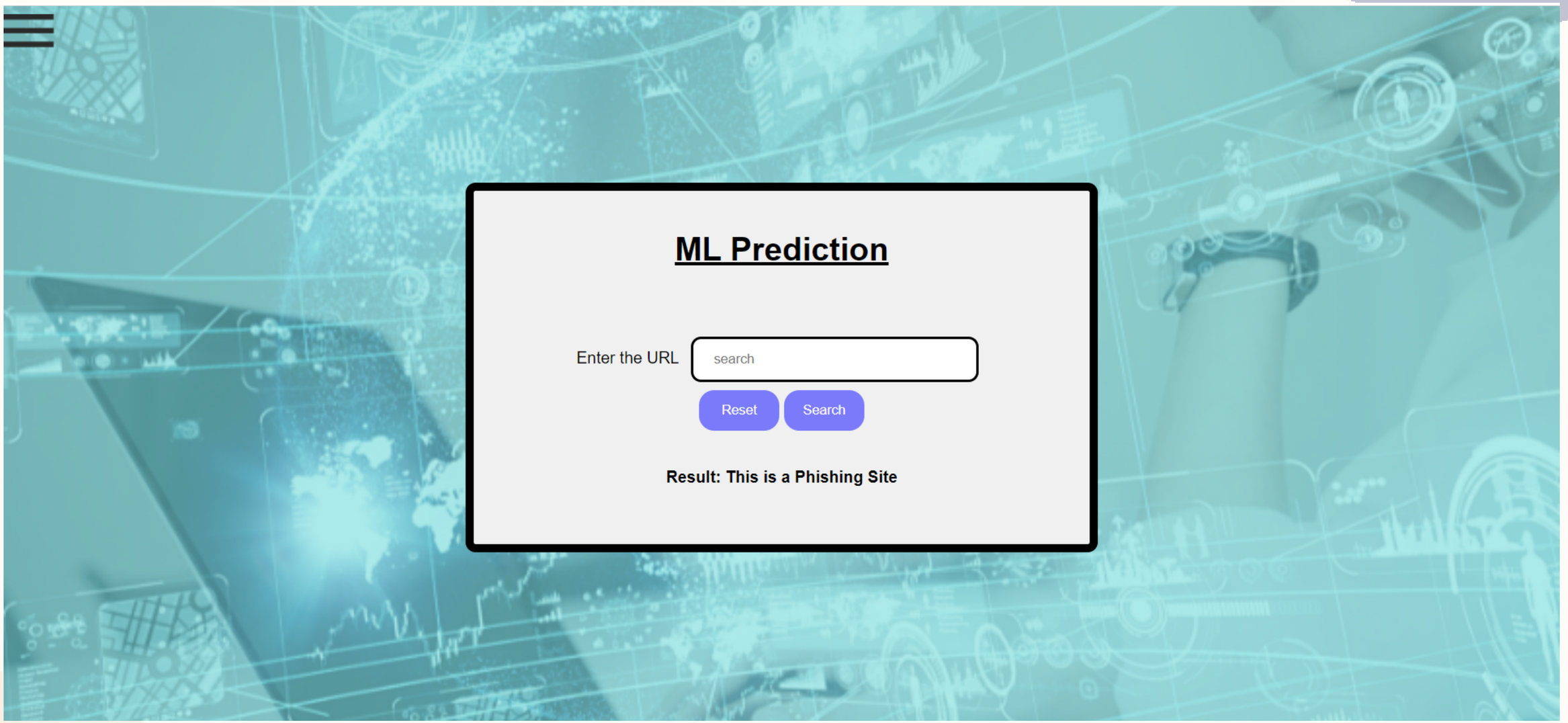
Module-wise Implementation Screenshots

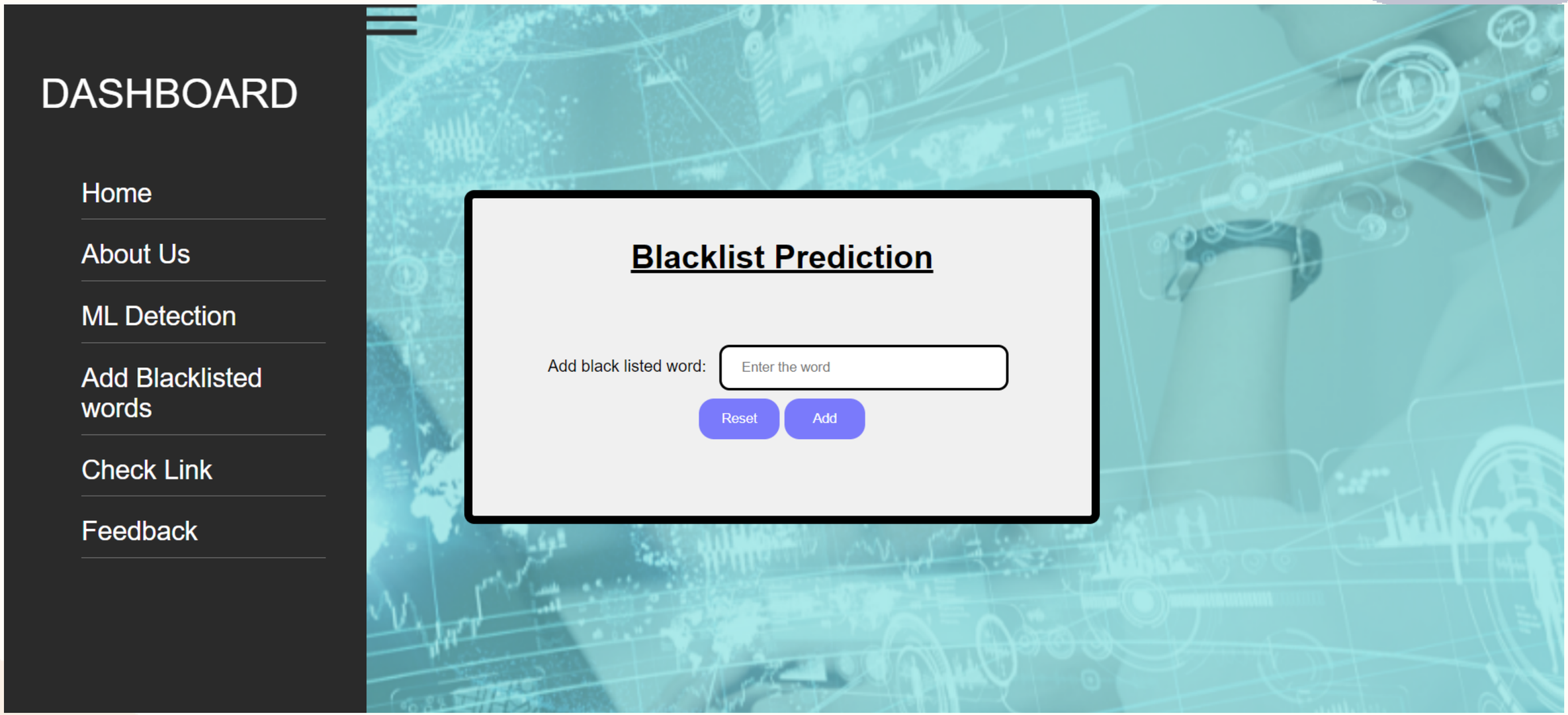

DASHBOARD
Home
About Us
ML Detection
Add Blacklisted words
Check Link
Feedback
ML Prediction
Enter the URL
search
Reset
Search
Result:

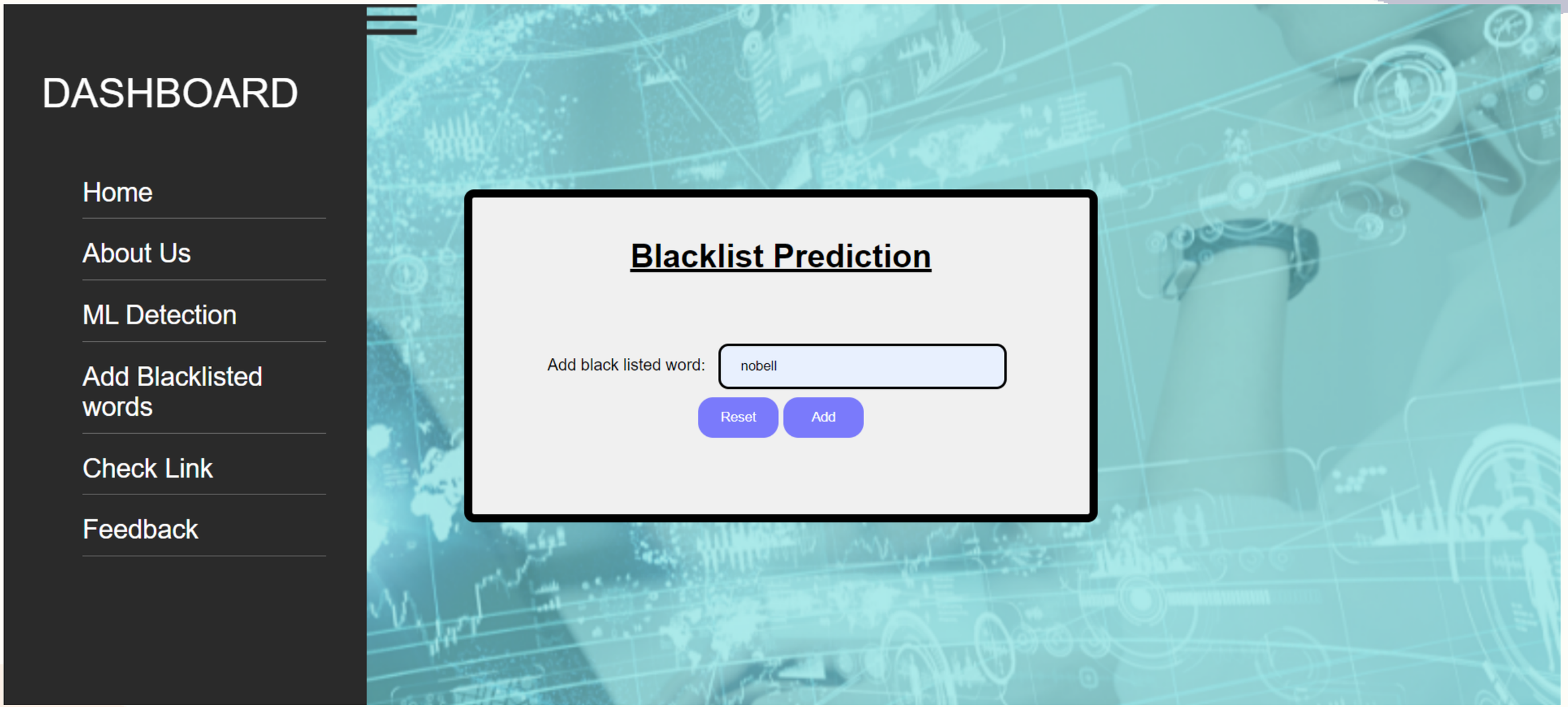

ML Prediction
Enter the URL
youtube.com
Reset
Search
Result:

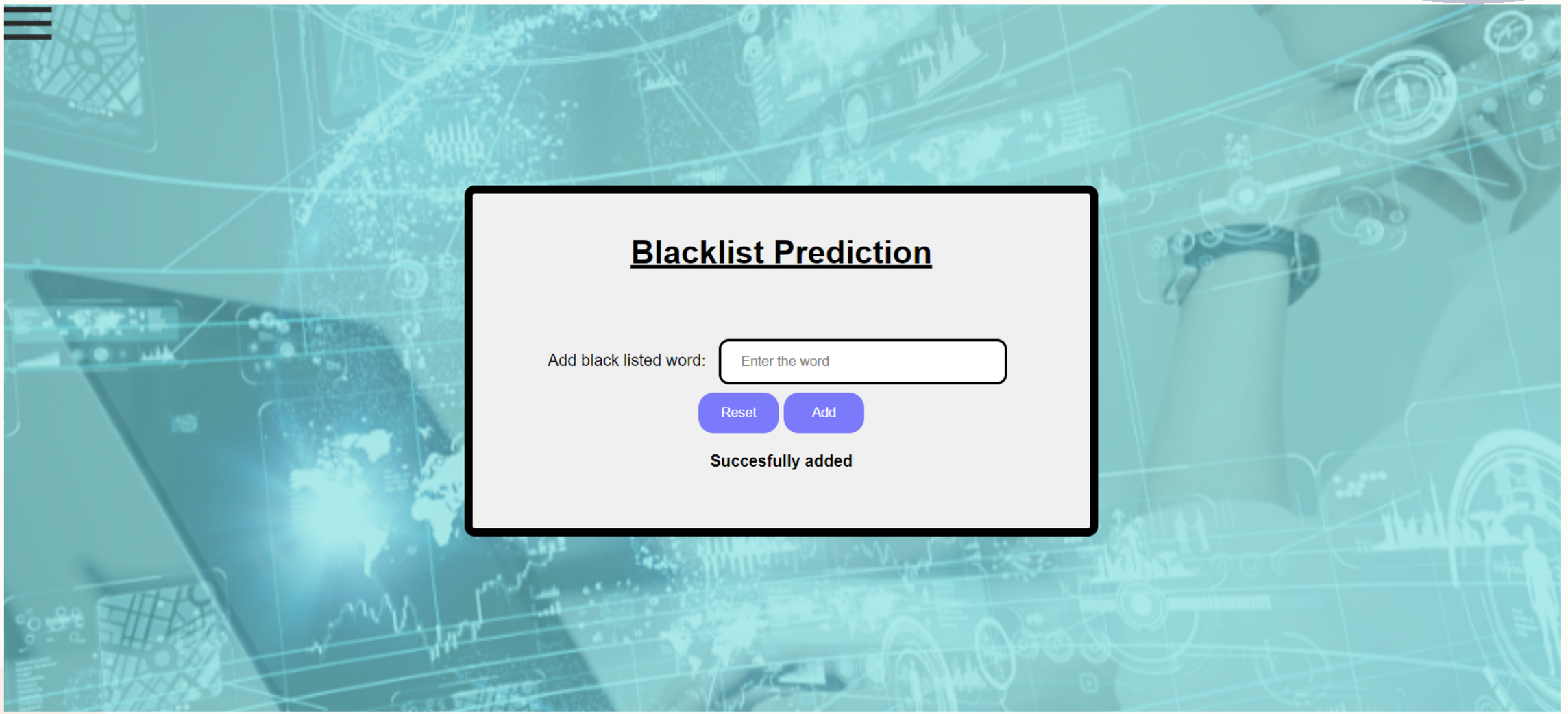

ML Prediction
Enter the URL
search
Reset
Search
Result: This is not a Phishing Site

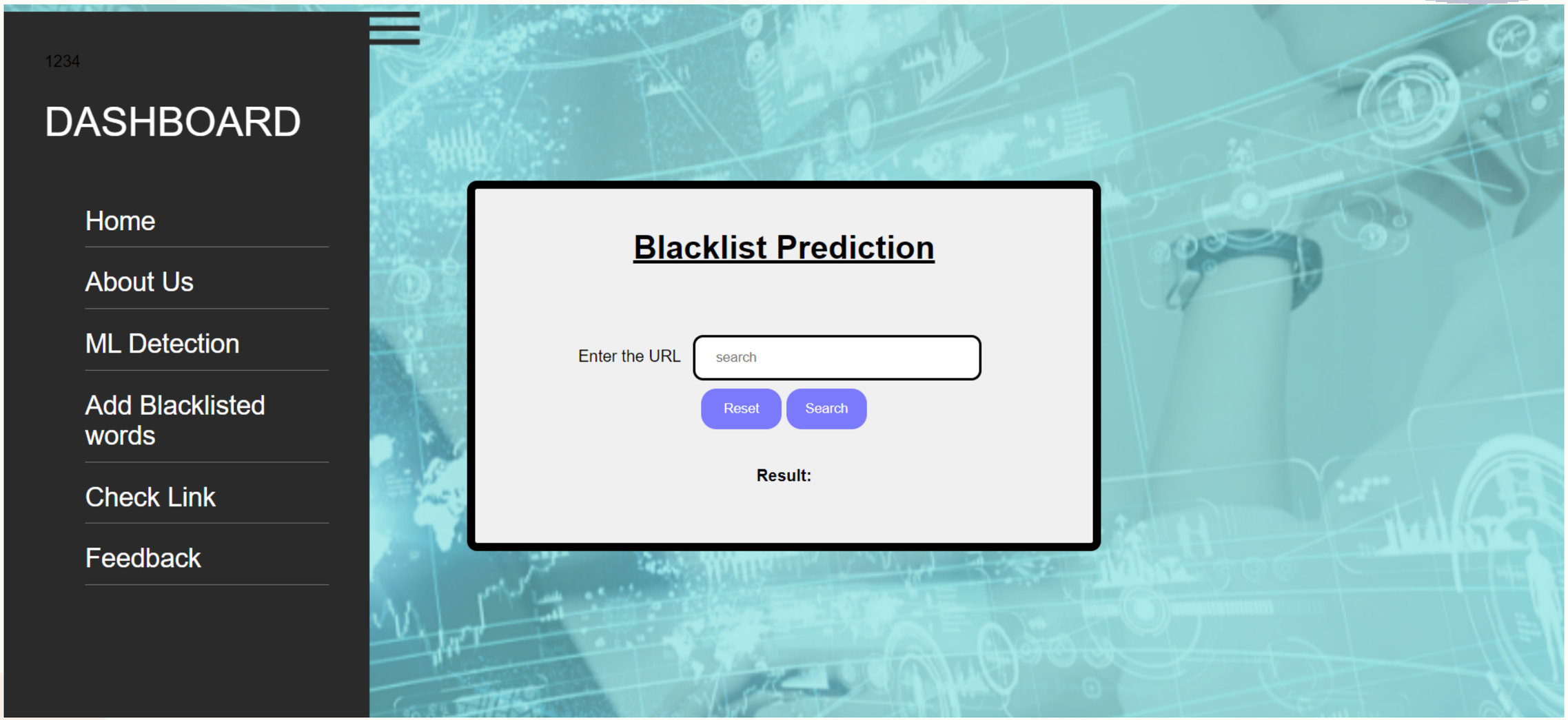

ML Prediction
Enter the URL
nobell.it/70ffb52d079109dca5664cce6f3
Reset
Search
Result:

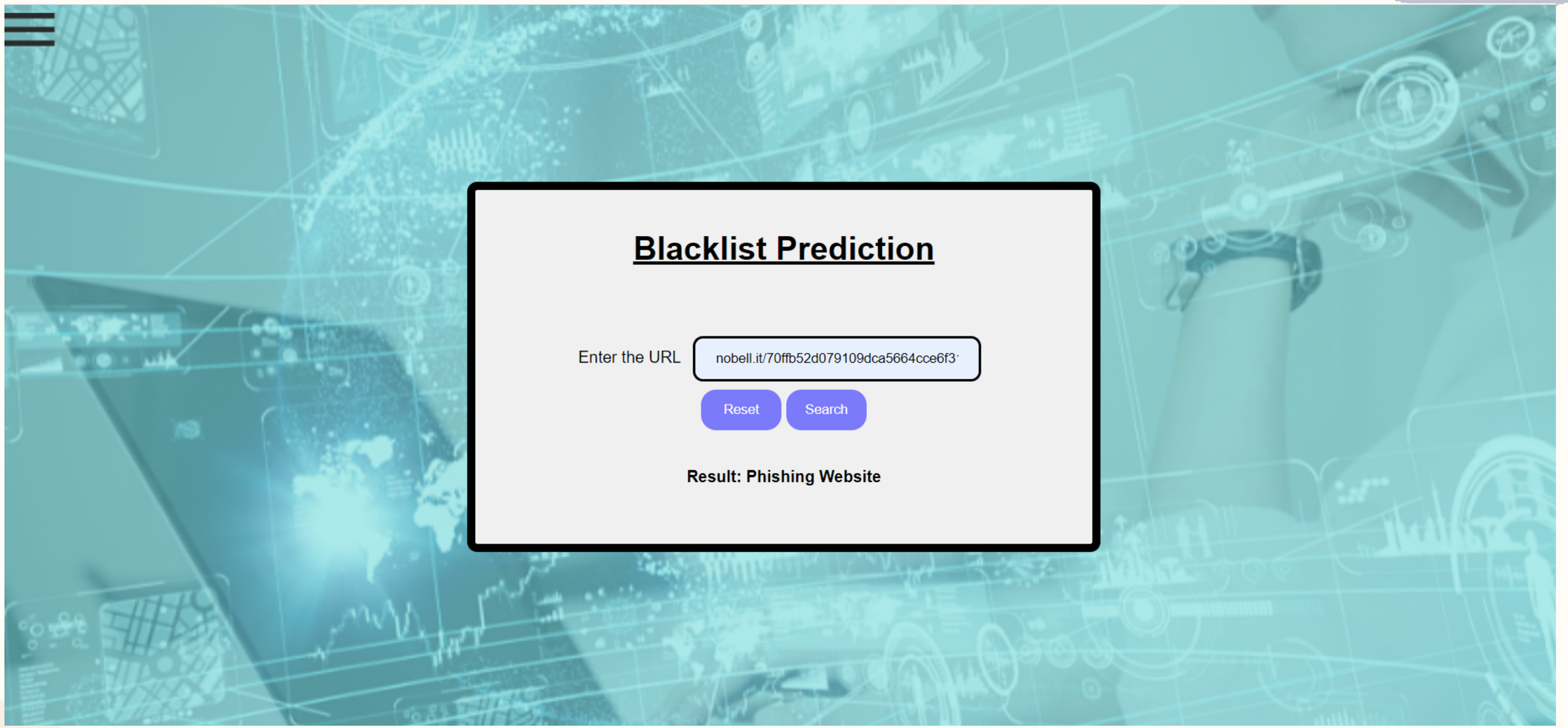

ML Prediction
Enter the URL
search
Reset
Search
Result: This is a Phishing Site


DASHBOARD
Home
About Us
ML Detection
Add Blacklisted words
Check Link
Feedback
Blacklist Prediction
Add black listed word:
Enter the word
Reset
Add


DASHBOARD
Home
About Us
ML Detection
Add Blacklisted words
Check Link
Feedback
Blacklist Prediction
Add black listed word:
nobell
Reset
Add


Blacklist Prediction
Add black listed word:
Enter the word
Reset
Add
Succesfully added


1234
DASHBOARD
Home
About Us
ML Detection
Add Blacklisted words
Check Link
Feedback
Blacklist Prediction
Enter the URL
search
Reset
Search
Result:


Blacklist Prediction
Enter the URL
nobell.it/70ffb52d079109dca5664cce6f3
Reset
Search
Result: Phishing Website

Future Work:

- For future enrichment, we aim to form the phishing detection system as an expandable web service that will integrate with online learning.
- So that new phishing attack patterns can easily be learned and enhance the accuracy of our models with better feature extraction.

Conclusion:

- URL phishing analysis is very useful in determining whether a certain URL is a legitimate URL or not and whether it should be visited or not.
- Thus, it prevents them from revealing their sensitive information to unknown or illegitimate sources.
- Since, the Logistic Regression algorithm gave better accuracy as compared to that of the Multinomial Naive Bayes algorithm,

Reference:

- Weiheng Bai; "Phishing Website Detection Based on Machine Learning Algorithm" https://ieeexplore.ieee.org/document/9275957
- Junaid Rashid; Toqeer Mahmood; Muhammad Wasif Nisar; Tahira Nazir; "Phishing Detection Using Machine Learning Technique" https://ieeexplore.ieee.org/document/9283771
- Mohammed Hazim Alkawaz; Stephanie Joanne Steven; Asif Iqbal Hajamydeen; "Detecting Phishing Website Using Machine Learning" https://ieeexplore.ieee.org/document/9068728

THANK YOU