JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Web Mining to Detect Online Spread of Terrorism

Vinay Mahajan, Prayag Adhikari, Omkar Molawade, Damu Kharat

Student, Student, Student, Student

Computer Engineering,

Abstract : Terrorism has grown its roots quite deep in certain parts of the world. With increasing terrorist activities it has become important to curb terrorism and stop. Its spread before a certain time. So as identified internet is a major source of spreading terrorism through speeches and videos. Terrorist organizations use internet to brain wash individuals and also promote terrorist activities through provocative web pages that inspire helpless people to join terrorist organizations. So here we propose an efficient web data mining system to detect such web. Properties and flag them automatically for human review. Data mining is a technique used to mine out patterns of useful data from large data sets and make the most use of obtained results. Data mining algorithms are used to manage organized data sets and web mining algorithms can be helpful in mining and extracting from unstructured web pages and text data that is available across the web. Websites built in different platforms have varying data structures and that makes it quite difficult to read for a single algorithm.

Keywords - Machine learning, Data Mining, Python, Visual Studio, Tkinter, Requests, BeautifulSoup4, Pillow, Ttkthemes.

I. INTRODUCTION

Since the late 1980s, the Internet has proven to be a highly dynamic means of communication, reaching an ever-growing audience worldwide. The development of increasingly sophisticated technologies has created a network with a truly global reach, and relatively low barriers to entry. Internet technology makes it easy for an individual to communicate with relative anonymity, quickly and effectively across borders, to an almost limitless audience. The benefits of Internet technology are numerous, starting with its unique suitability for sharing information and ideas, which is recognized as a fundamental human right.2 It must also be recognized, however, that the same technology that facilitates such communication can also be exploited for the purposes of terrorism. The use of the Internet for terrorist purposes creates both challenges and opportunities in the fight against terrorism.
Terrorist organizations are using the internet to spread their propaganda and radicalize youth online and encourage them to commit terrorist activities. In order to minimize the online presence of such harmful websites we need to devise a system which detects specific keywords in a particular website. The website should be flagged inappropriate if the keywords are found for efficient system development. Data mining consists of text mining methods that help us to scan and extract useful content from unstructured data. Text mining helps us to detect keywords, patterns and important information from unstructured texts.

II. ALGORITHMS

We use web mining algorithms to mine textual information on web pages and detect their relevancy to terrorism. Websites created in different platform can be tracked using this application. This system will check web pages whether a webpage is promoting terrorism. This system will classify the web pages into various categories and sort them appropriately. There are two features used in this system that is data mining and web mining. Data mining is a technique used to mine out patterns of useful data from large data sets and make the most use of obtained results. Web mining also consists of text mining methodologies that allow us to scan and extract useful content from unstructured data. This System are used only by the government officials who work for country security. System will help the cops to easily track the susceptible community who are held in terrorism. Website will have following characteristics:
Load Balancing: Since the system will be available only the admin logs in the amount of load on server will be limited to time period of admin access.
Easy Accessibility: Records can be easily accessed and store and other information respectively.
User Friendly: The Website will be giving a very user-friendly approach for all user.
Efficient and reliable: Maintaining the all secured and database on the server which will be accessible according the user requirement without any maintenance cost will be a very efficient as compared to storing all the customer data on the spreadsheet or in physically in the record books.
Easy maintenance: Web Data Mining for Terrorism Analysis website is design as easy way. So maintenance is also easy.

A. MACHINE LEARNING ALGORITHMS

- **Random Forest**:

Random forest algorithm, like its name implies, consists of a large number of individual decision trees that operate together. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. [Deziel, M. et al.]

- **Decision Tree**:

A decision tree is a tree-like graph with nodes representing the place where we pick an attribute and ask a question; edges represent the answers to the question; and the leaves represent the actual output or class label. They are used in non-linear decision making with a simple linear decision surface.

- **Naïve Bayes**:

Naive Bayes classifiers are a collection of classification algorithms based on Bayes Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

- **Logistic Regression**:

The logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

- **K-nearest Neighbors**:

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique.

B. DATA CHAINING ALGORITHM

Input: Log Table (LT)

Output: Summarized Log Table (SLT)

'*' = access pages consist of embedded objects

(i.e. .jpg, .gif, etc.)

'**' =successful status codes and requested methods (i.e. 200, GET etc)

Begin

Step 1: Read records in LT

Step 2: For each record in LT

Step 3: Read fields (Status code, method)

Step 4: If Status code='**'and method= '**'

Then,

Step 5: Get IP address and URL link

Step 6: If suffix. URL Link= {*.gif,*.jpg,*.css}

Then

Step 7: Remove suffix. URL link

Step 8: Save and URL Link

End if

Else

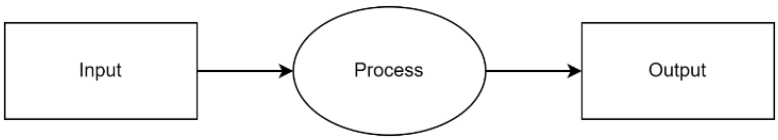
Step 9: Next record

End if

End

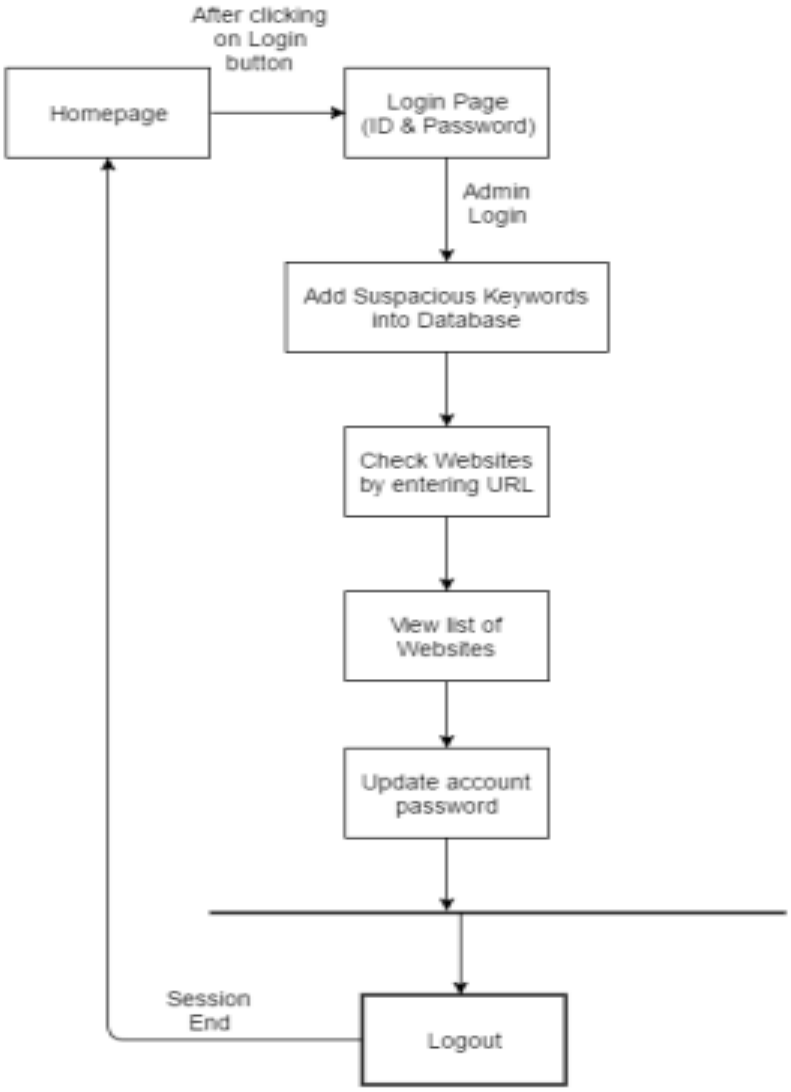
III. Flow Diagram

A. Data Flow Diagram Level-0

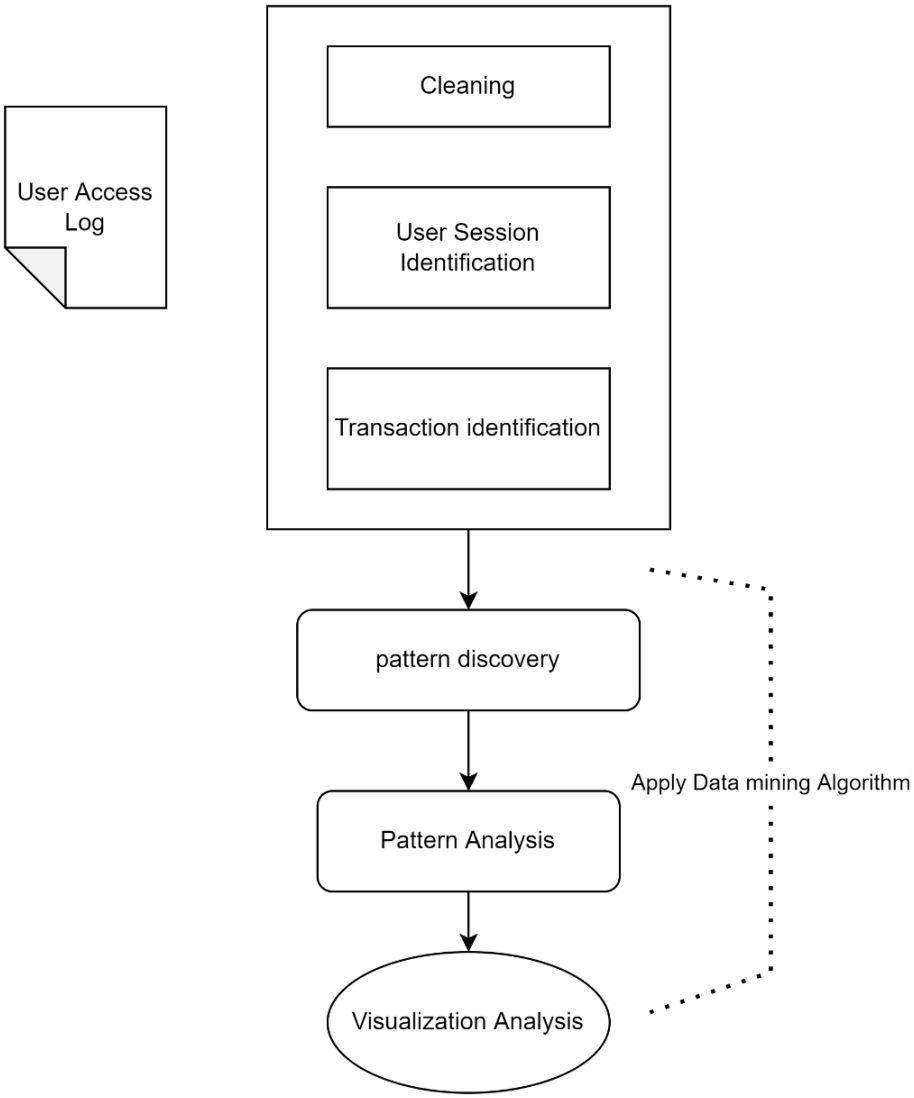


System will track web pages that are more susceptible to terrorism and will report IP Address to the user who is using the system.

B. Data Flow Diagram Level-1

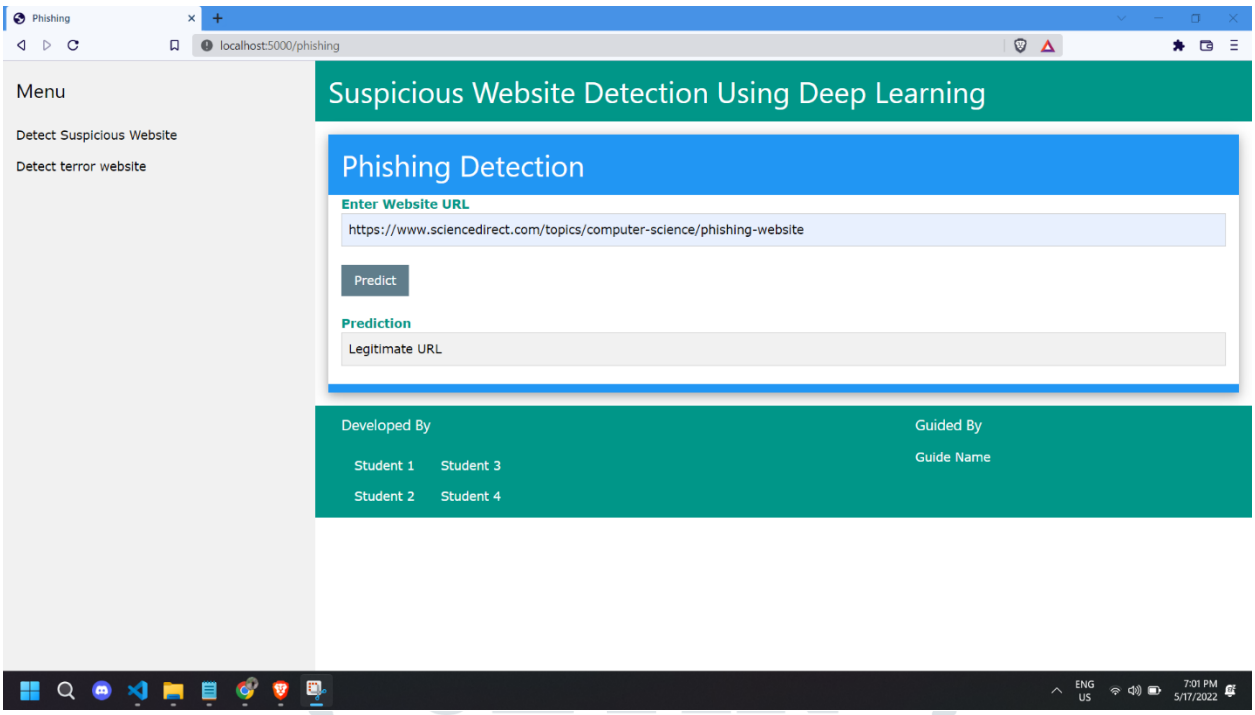


C. Data Flow Diagram Level-2

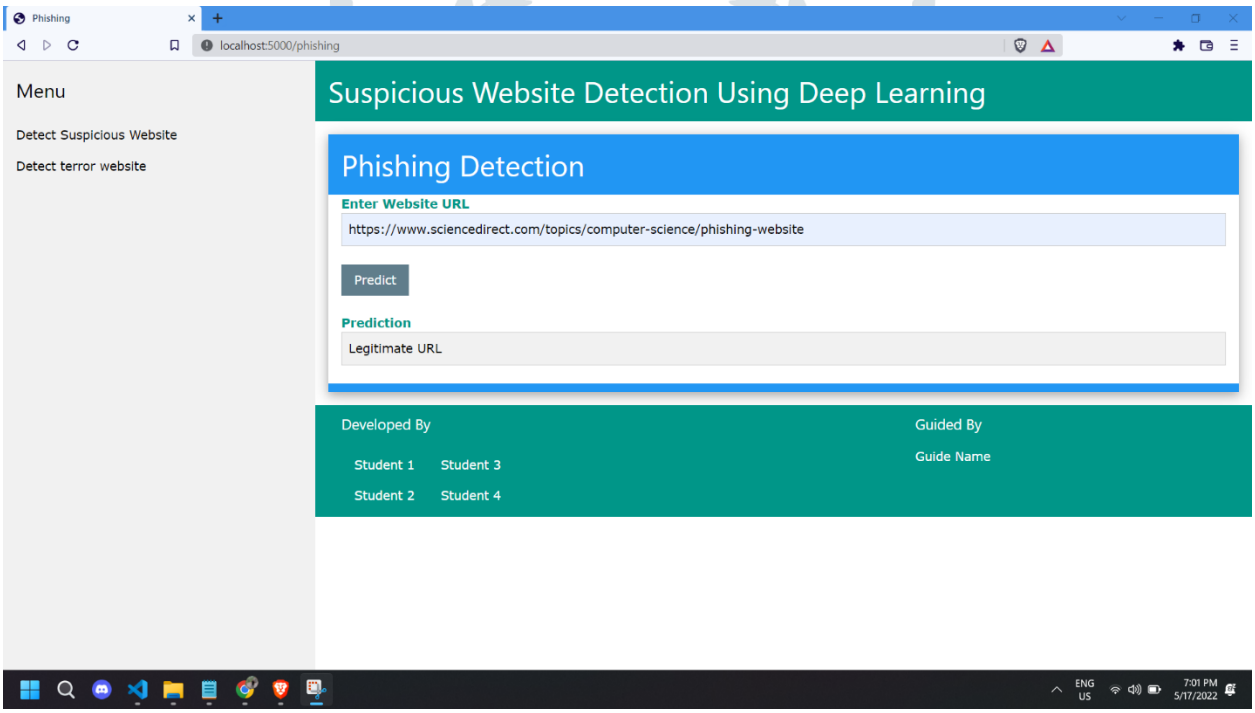


IV. Demo (With Snapshots):-

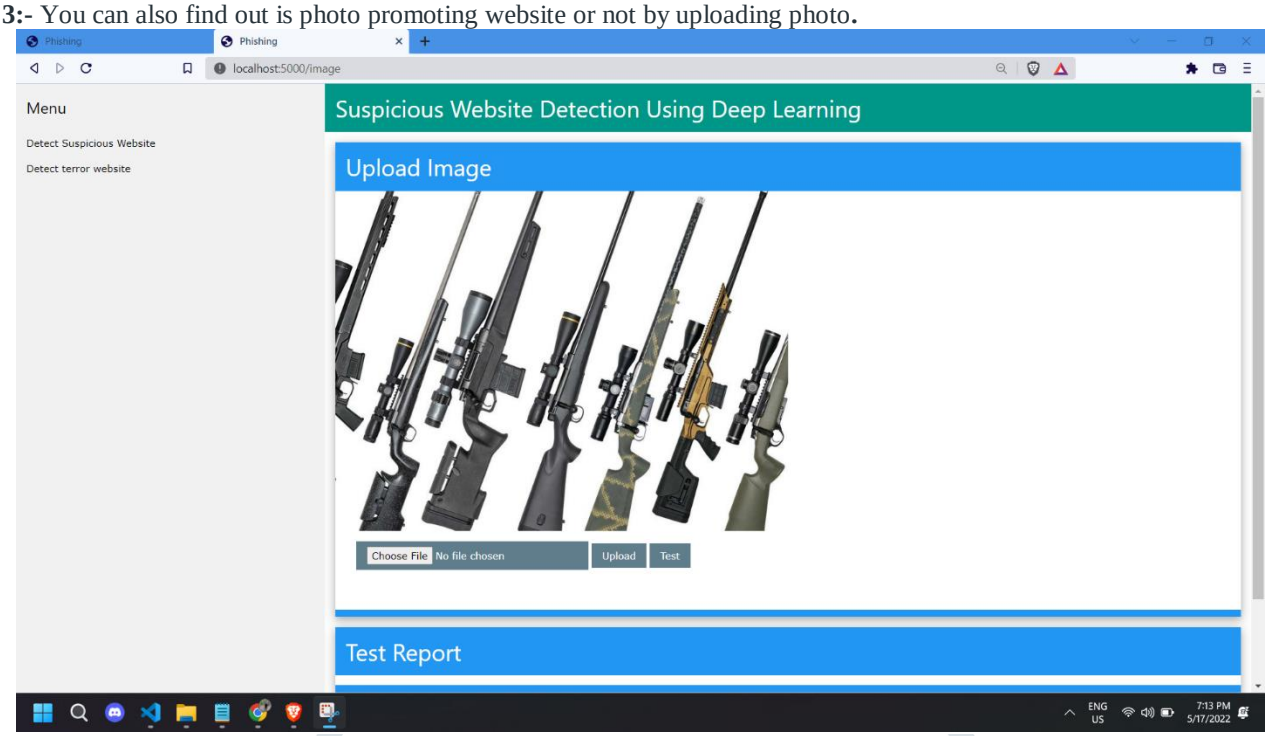
Step 1:- Choose module



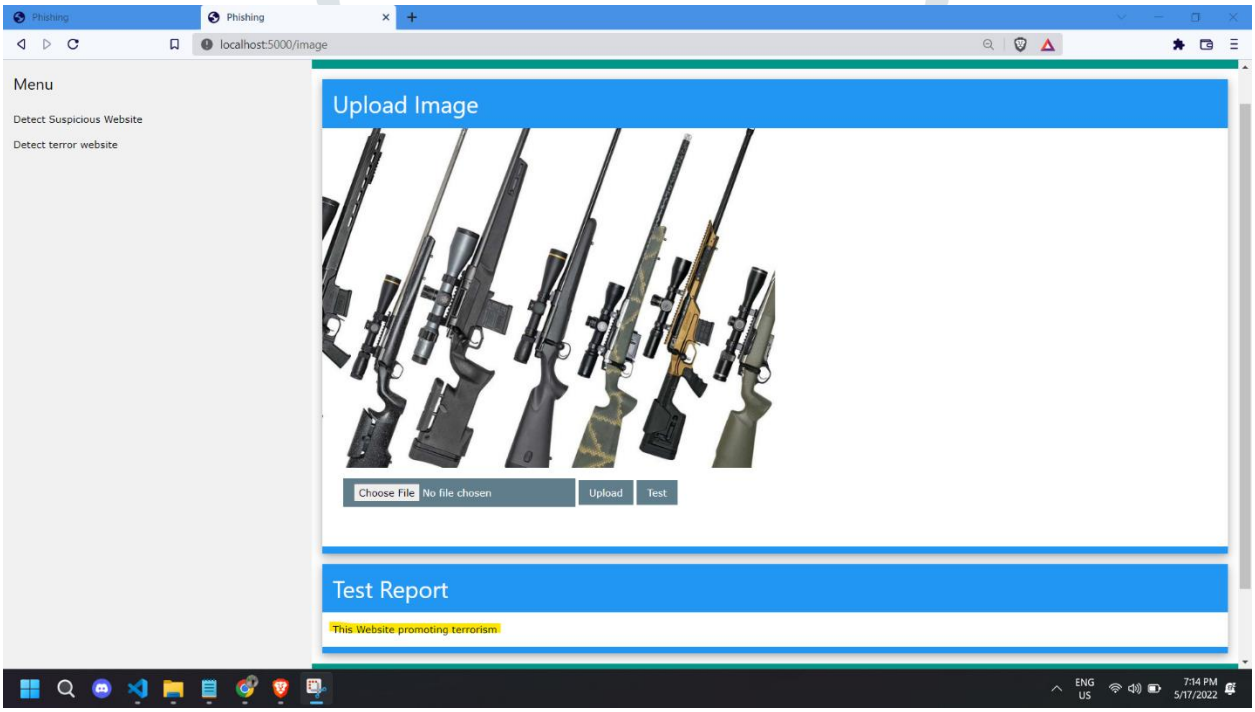
Step 2:- You can detect phishing site by copy pasting site in column.You can see below it shows result.



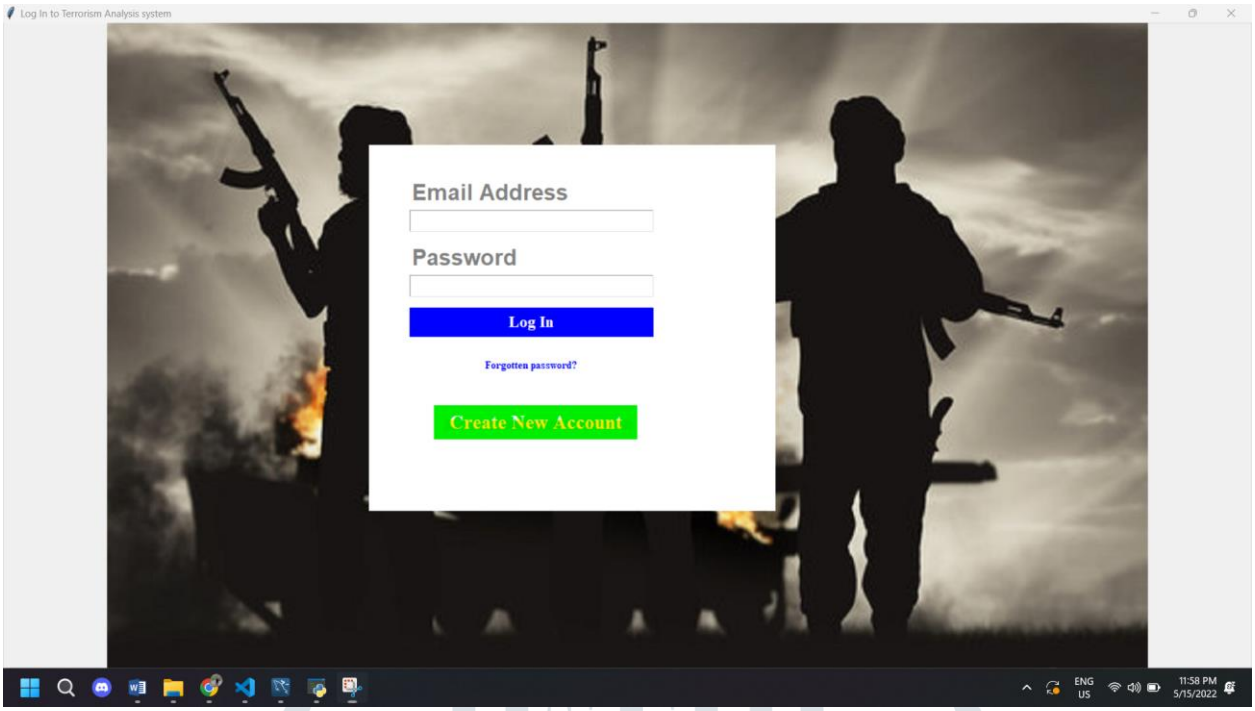
Step 3:- You can also find out is photo promoting website or not by uploading photo**.**



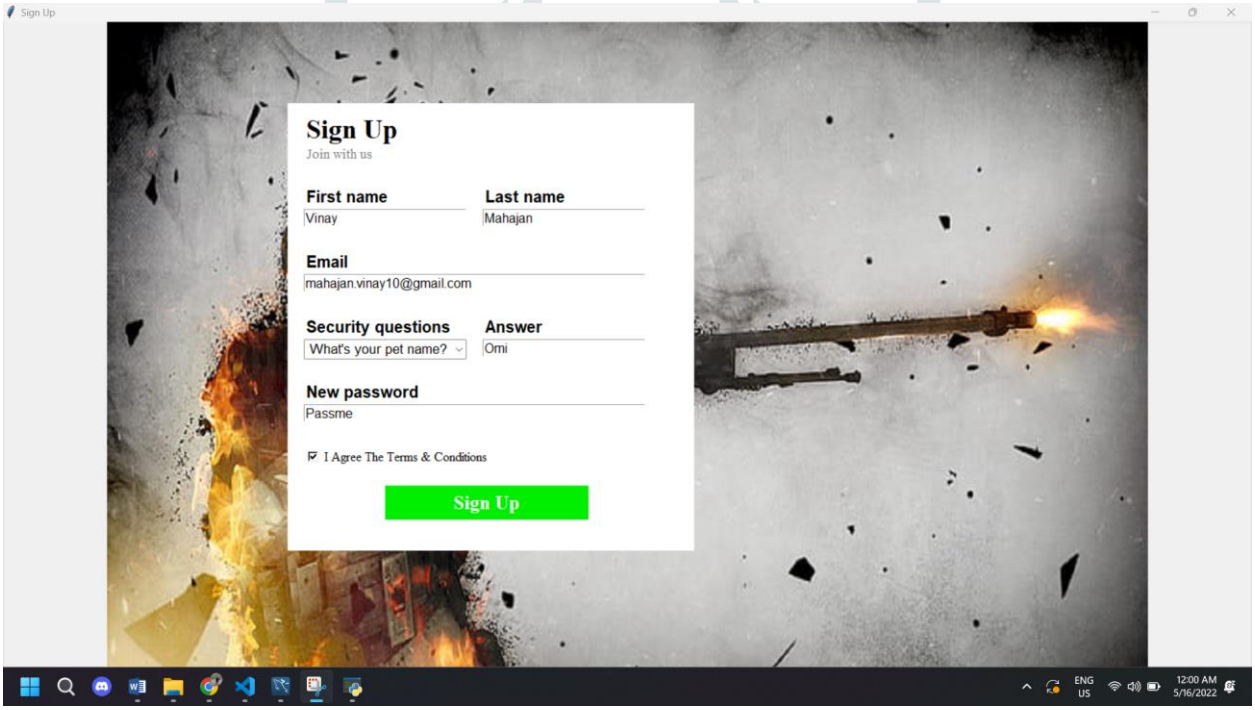
Step 4:- After hitting test button it show will show you result.



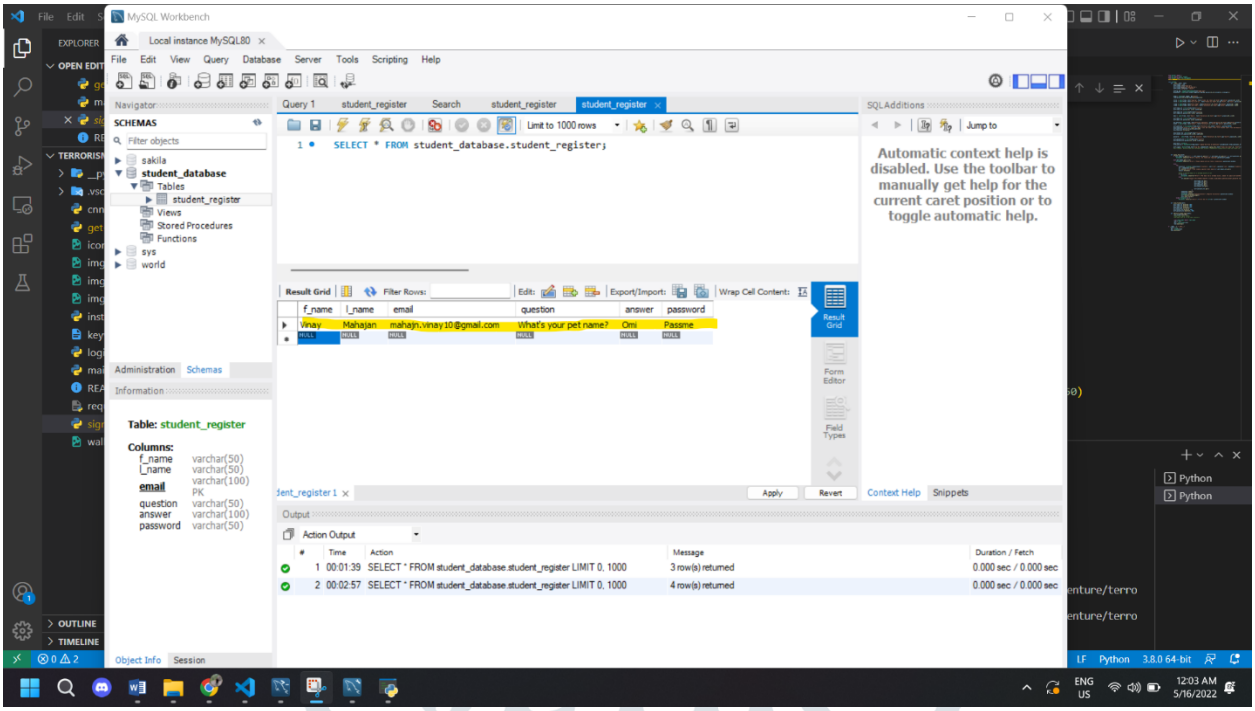
Step 5:- Open login_page.py



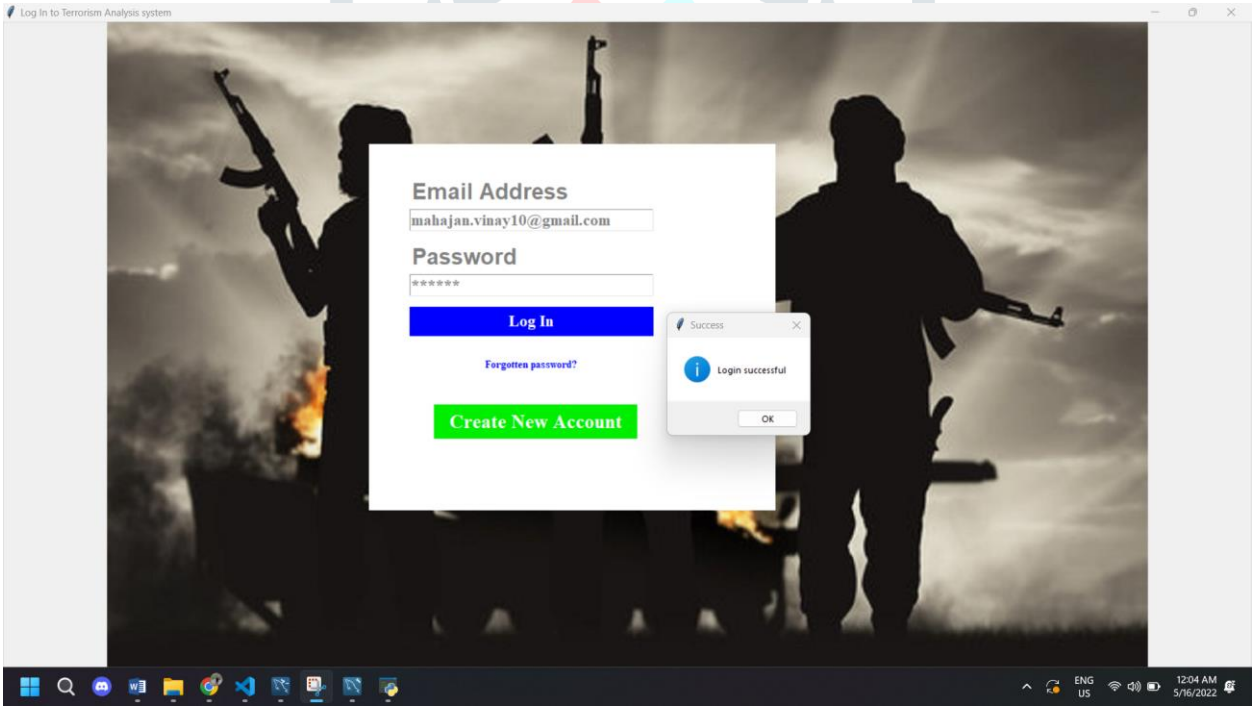
Step 6:- Click on "Create New Account".



Step 7:- Now Sign-up page is open, Fill Details to create new account and click on Sign Up Button.

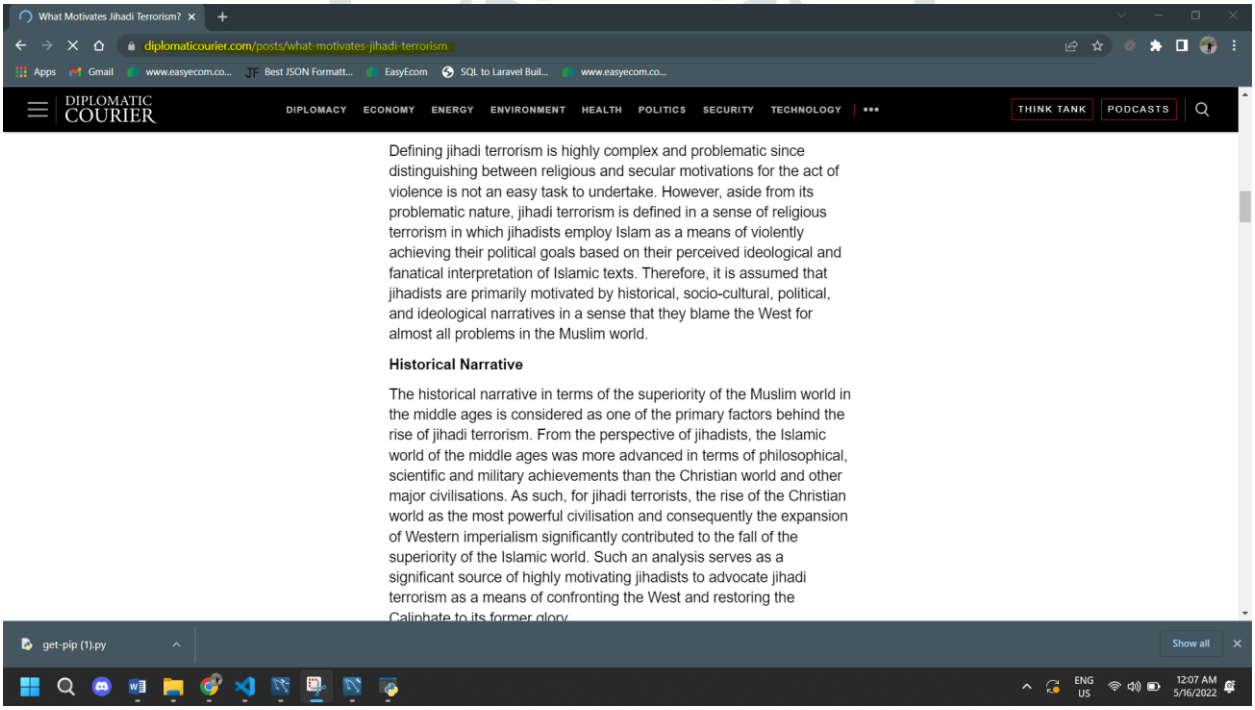


Step 8:- You can see new account is created in DB in MySQL.

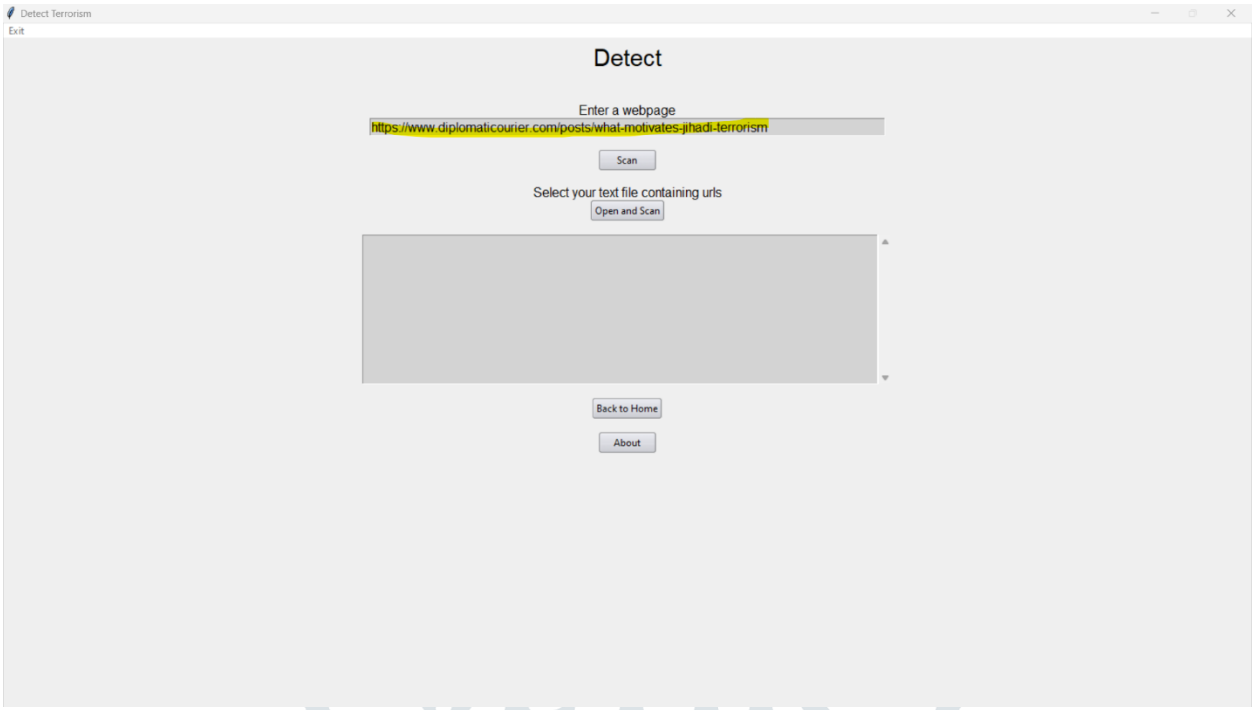


Step 9:-You can see login page is opened. Fill Login Credentials. After that you can see there is pop-up message saying "Login successful".

Step 10:- Click on Detect Button.

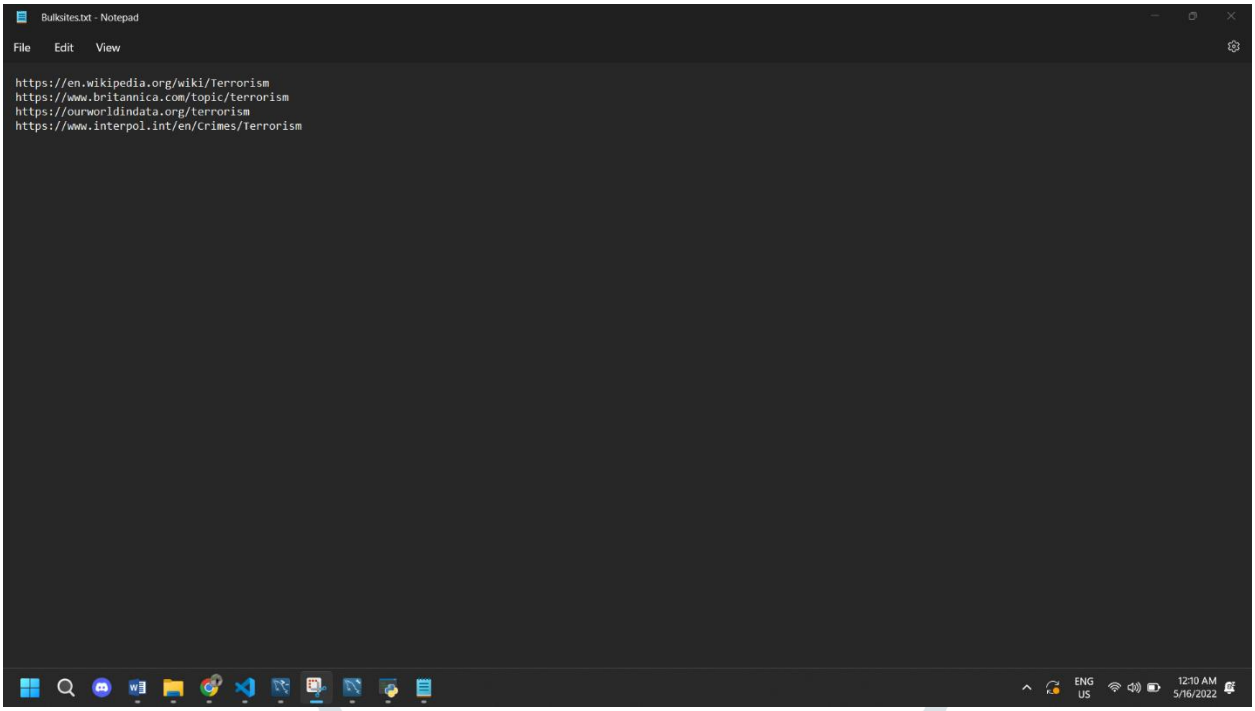


Step 11:- Copy Websites URL which you want to evaluate.

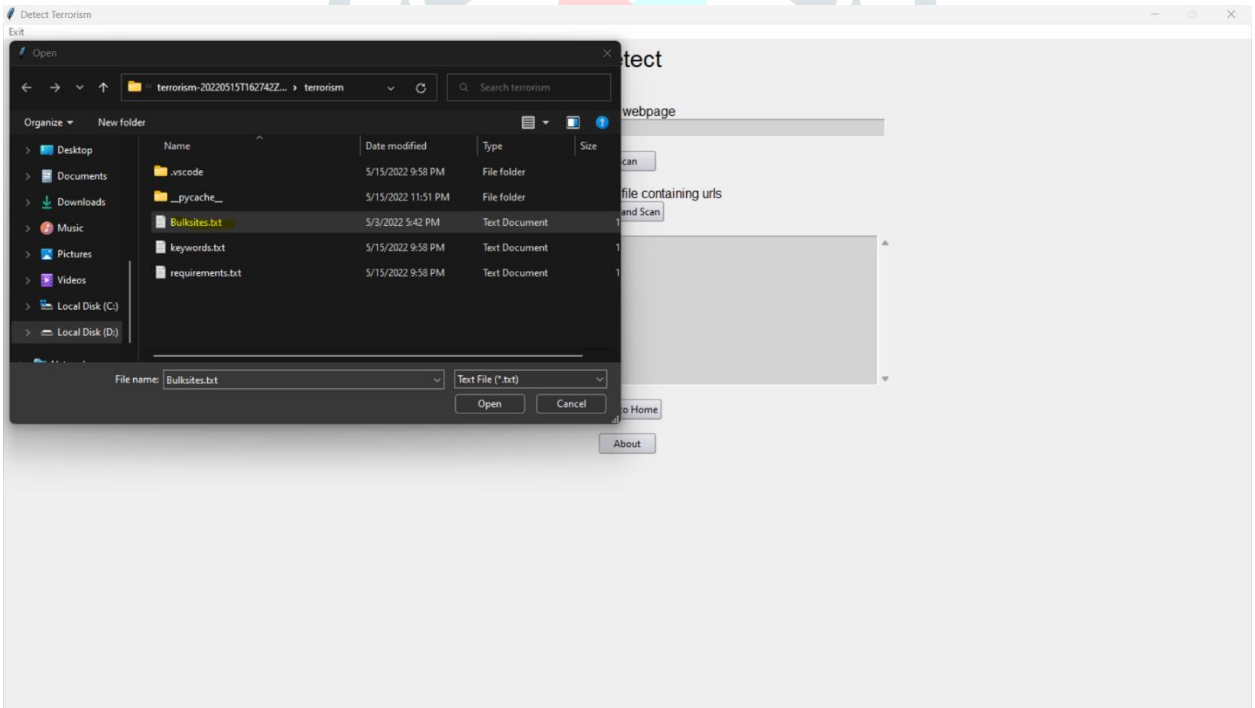


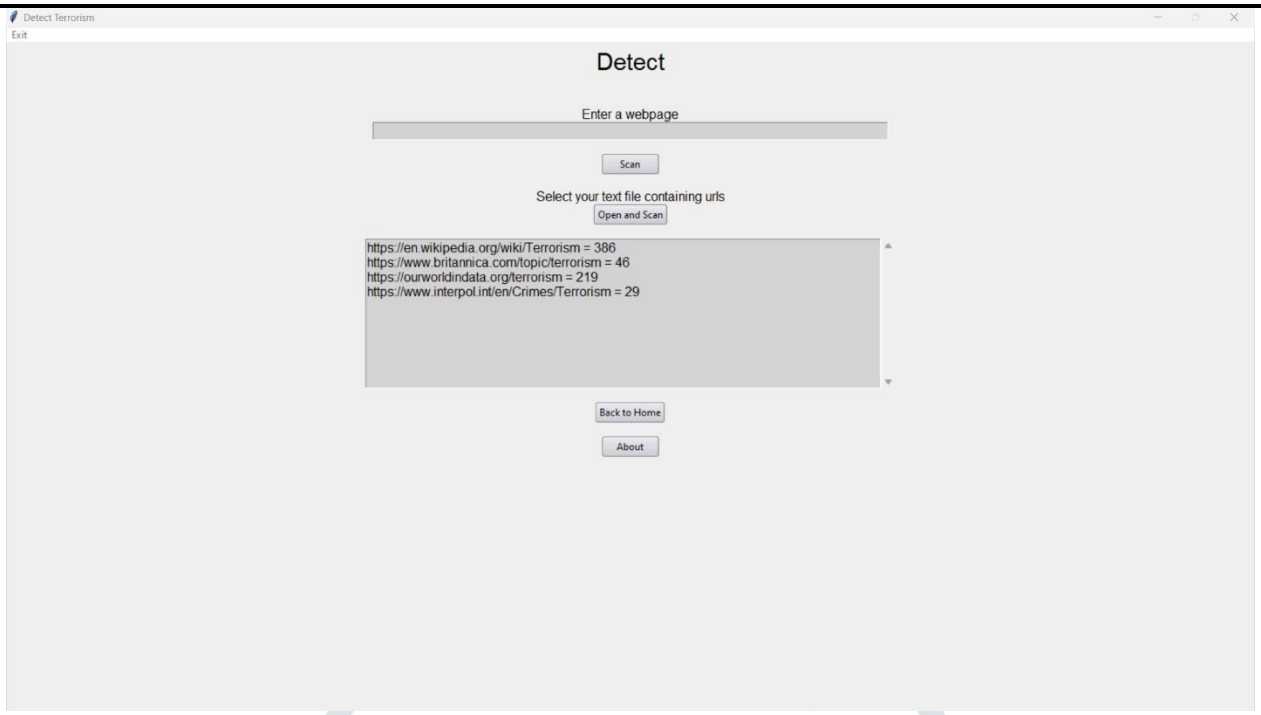
Step 12:- Paste that URL in Enter Webpage Column And click on scan Button.

Step 13:- Program will show you The result. You can also add bulk sites by clicking Open and Scan button.

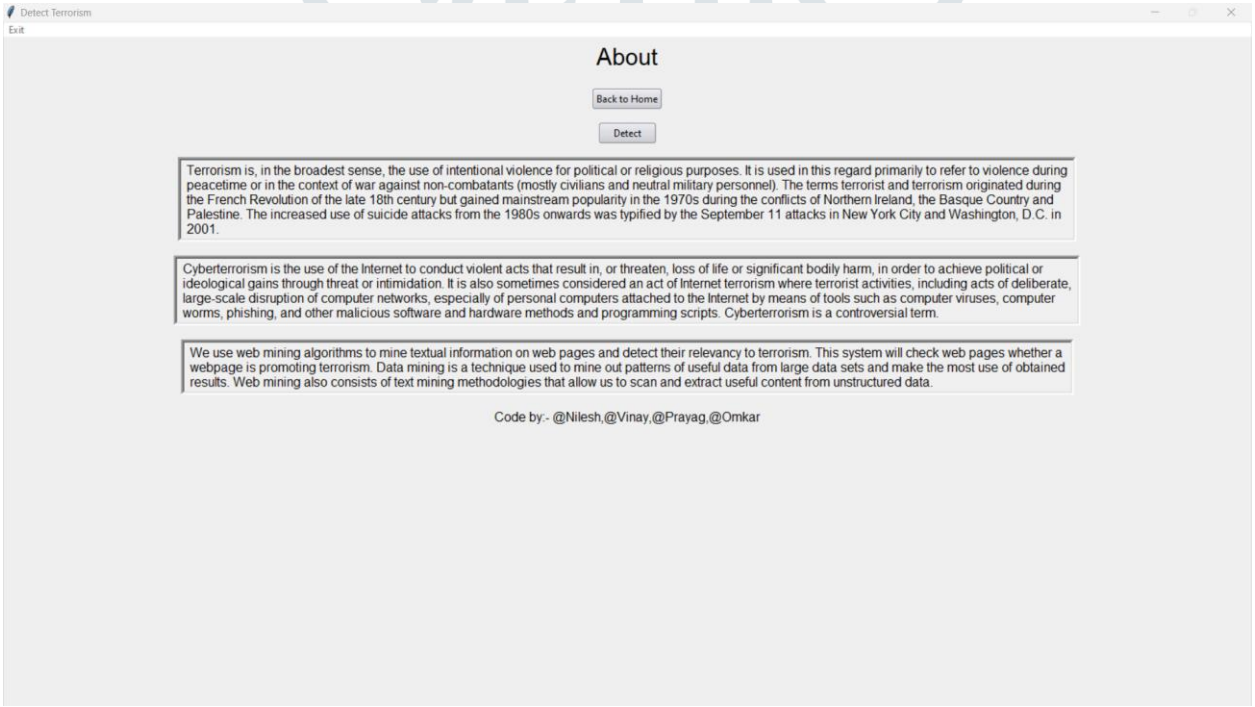


Step 14:- You have to add website line by line in notepad and then select and upload that file.





Step 15:- Now you can see the all scanned websites result.



Step 16:- You can know more about project by clicking About Section.

V. Advantages:-

- Very High accuracy in image recognition problems.
- Automatically detects the important features without any human supervision.
- Weight sharing.

VI. Disadvantages:-

- CNN do not encode the position and orientation of object.
- Lack of ability to be spatially invariant to the input data.
- Lots of training data is required.

VII. Rule-of-law considerations:-

32. Respect for human rights and the rule of law is an integral part of the fight against terrorism. Due care must be taken to respect international human rights standards in all phases of counter-terrorism initiatives, from preventive intelligence gathering to ensuring due process in the prosecution of suspects. This requires the development of national counter-terrorism legislation and practices that promote and protect fundamental human rights and the rule of law.24

33. States have both a right and a duty to take effective measures to counter the destructive impact of terrorism on human rights, in particular the rights to life, liberty and physical integrity of individuals and the territorial integrity and security of States. Effective counter-terrorism measures and the protection of human rights are complementary and mutually reinforcing objectives which must be pursued together.25 Counterterrorism initiatives relating to Internet use may have an impact on the enjoyment of a range of human rights, including the rights to freedom of speech, freedom of association, privacy and a fair trial. While a comprehensive analysis of human rights issues is beyond the scope of the present publication, it is important to highlight key areas for consideration.

34. As noted in subsection B.1*(b)* above, the proscription of incitement to terrorism may involve restrictions on freedom of expression. Freedom of expression is not an absolute right. It may be restricted, subject to satisfaction of strictly construed tests of legality, necessity, proportionality and non-discrimination, when that freedom is used to incite discrimination, hostility or violence. A key difficulty in cases of glorification or incitement to terrorism is identifying where the line of acceptability lies,

VIII. Conclusion and Future Scope:-

To curb the menace of terrorism and to destroy the online presence of dangerous terrorist organizations like ISIS and other radicalization websites. We need a proper system to detect and terminate websites which are spreading harmful content used to radicalizing youth and helpless people. We analyzed the usage of Online Social Networks (OSNs) in the event of a terrorist attack. We used different metrics like number of tweets, whether users in developing countries tended to tweet, re-tweet or reply, demographics, geo-location and we defined new metrics (reach and impression of the tweet) and presented their models. While the developing countries are faced by many limitations in using OSNs such as unreliable power and poor Internet connection, still the study finding challenges the traditional media of reporting during disasters like terrorist's attacks. We recommend centers globally to make full use of the OSNs for crisis communication in order to save more lives during such.

IX. References:-

[1] Aakash Negandhi, Soham Gawas, Prem Bhatt , Priya Porwal "Detect Online Spread of Terrorism Using Data Mining".IOSR Journal of Engineering Volume 13,17 April 2019. So here they propose an efficient web data mining system to detect such web properties and flag them automatically for human review. Keywords: Anti-Terrorism, Data Mining, Online, Terrorism,World

[2] Avishag Gordon "The spread of terrorism publications: A database analysis",Terrorism and Political Violence journal publishedin Dec 2007.This research note focuses on the spread of terrorism publications from 1988 to 1995 compared to their frequency of appearance from 1996 to 1998. It also identifies the core journals of this research field.

[3] A.Sai Hanuman, G.Charles Babu , P.Vara Prasad Rao, P.S.V.Srinivasa Rao ,B.Sankara Babu "A Schematic Approach on Web Data Mining In Online Spread Detection of Terrorism",International Journal of Recent Technology and Engineering Volume-8, Issue-1, May 2019.So here they have propose a compelling web data mining structure to recognize such web properties and standard them thusly for human review. Index Terms: web data mining, terrorism, web structure mining, dread monger affiliations.

[4] Counter Terrorism on Online Social Networks Using Web Mining Techniques Fawad Ali, Farhan Hassan Khan, Saba Bashir, and Uzair Ahmad, Department of Computer Science, Federal Urdu University of Arts, Science and Technology (FUUAST), Islamabad, Pakistan.In this paper some major web mining techniques have been discussed which can be helpful to identify such people and terrorism may be countered from OSN. Each technique is discussed thoroughly, and effectiveness along with its pros and cons are also presented.

[5] Chen, H.. "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums." ACM Transactions on Information Systems, forthcoming,June 2008.In this study the use of sentiment analysis methodologies is proposed for classification of Web forum opinions in multiple languages. The utility of stylistic and syntactic features is evaluated for sentiment classification of English and Arabic content.

[6] J. Kiruba, P. Sumitha, K. Monisha, S. Vaishnavi "Enhanced Content Detection Method to Detect Online Spread of Terrorism",International Journal of Engineering and Advanced Technology Volume-8, Issue-6S3, September 2019.They proposed a system delivery event notification which is used to monitor the activities and delivers notification according to the investigation knowledge. Alert reporting system is developed that takes earthquakes from websites and a message is sent the registered user.

[7] Michael Grenieri, Anthony Estrada "Down Converter Characterization in a Synthetic Instrument Context" 2006 IEEE Autotestcon.This paper provides an overview of the need for a common set of specification parameters to characterize a down converter in a synthetic instrument (SI).The paper then provides an in-depth technical discussion of two of the less understood down converter related intermediate frequency (IF) output parameters: group delay and phase linearity.

[8] Naseema Begum "Detection of online spread of terrorism using web data mining" A. Institute of Engineering and Technology, Coimbatore, Tamil Nadu, International Journal of Advance Research, Ideas and Innovations in Technology- Volume 5, Issue 1. The basic idea of this project is to reduce or stop spreading of terrorism and to remove all these accounts