# Survey on Data Mining

**Guide Name:- Prof. A K Gaikwad**

**1] First Author Name:-** Vinay Mahajan

**Collage Name:-**Sinhgad Academy of Engineering Pune

**Email: -** mahajan.vinay10@gmail.com

**2] Second Author Name:-** Prayag Adhikari

**Collage Name: -** Sinhgad Academy of Engineering Pune.

**Email: -** adhikariprayag@gmail.com

**3] Third Author Name:-** Omkar Molawade

**Collage Name: -** Sinhgad Academy of Engineering Pune

**Email: -** omkarmolawade00@gmail.com

**4] Fourth Author Name:-** Damu Kharat

**Collage Name: -** Sinhgad Academy of Engineering Pune

**Email: -** nilesh.kharat321@gmail.com

**Keyword:-**

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

## 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 terorist 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.

## Introduction: -

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.

## Literature Survey: -

Applied various machine learning algorithms in "Detect Online Spread of Terrorism Using Data Mining" to mine textual information on web pages and detect their relevancy to terrorism.Used the features of sentiment analysis to segregate the words of a web page, classify them and assert a score to each word in "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums."Studied various methods by which graphical data can be fetched and scanned and executed them to counter Terrorism on Online Social Networks using web mining techniques.classified the web pages into various categories and sorted them appropriately.

## 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.

- 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

## CNN Architecture (5 Layers)

The CNN architecture consists of several kinds of layers; Convolutional layer, pooling layer, fully connected input layer, fully connected layer and fully connected output layer.

Convolutional layer: Convolutional layer is the backbone of any CNN working model. This layer is the one where pixel by pixel scanning takes place of the images and creates a feature map to define future classifications.

Pooling layer: Pooling is also known as the down-sampling of the data by bringing the overall dimensions of the images. The information of each feature from each convolutional layer is limited down to only containing the most necessary data. The process of creating convolutional layers and applying pooling is continuous, may take several times.

Fully connected input layer: This is also known as the flattening of the images. The outputs gained from the last layer are flattened into a single vector so that it can be used as the input data from the upcoming layer.

Fully connected layer: After the feature analysis has been done and it's time for computation, this layer assigns random weights to the inputs and predicts a suitable label.
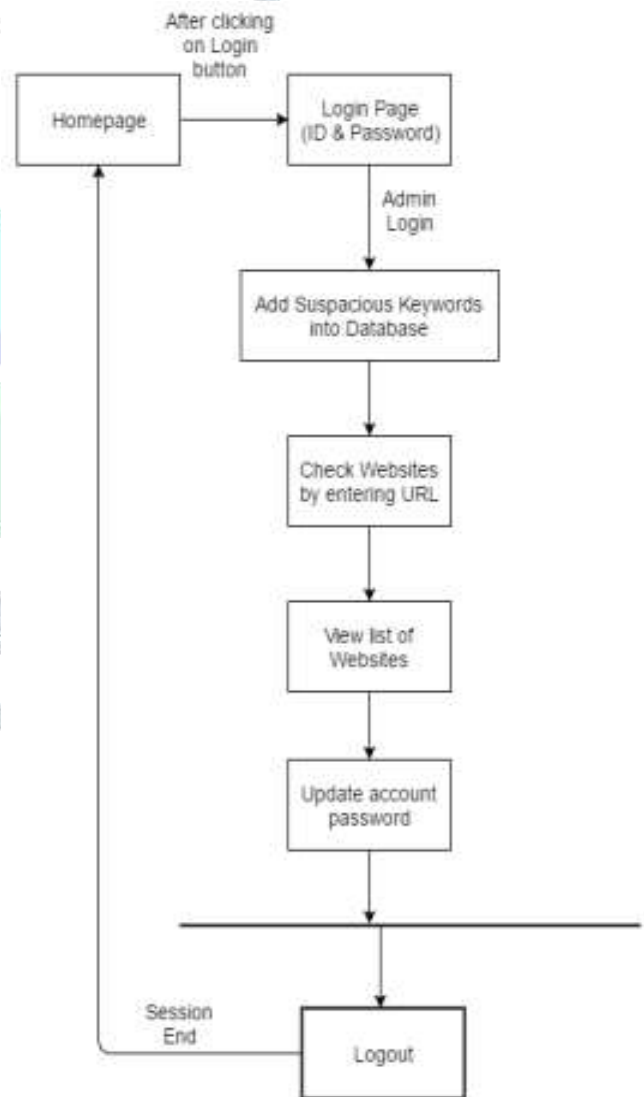
Fully connected Output layer: This is the final layer of the CNN model which contains the results of the labels determined for the classification and assigns a class to the images.
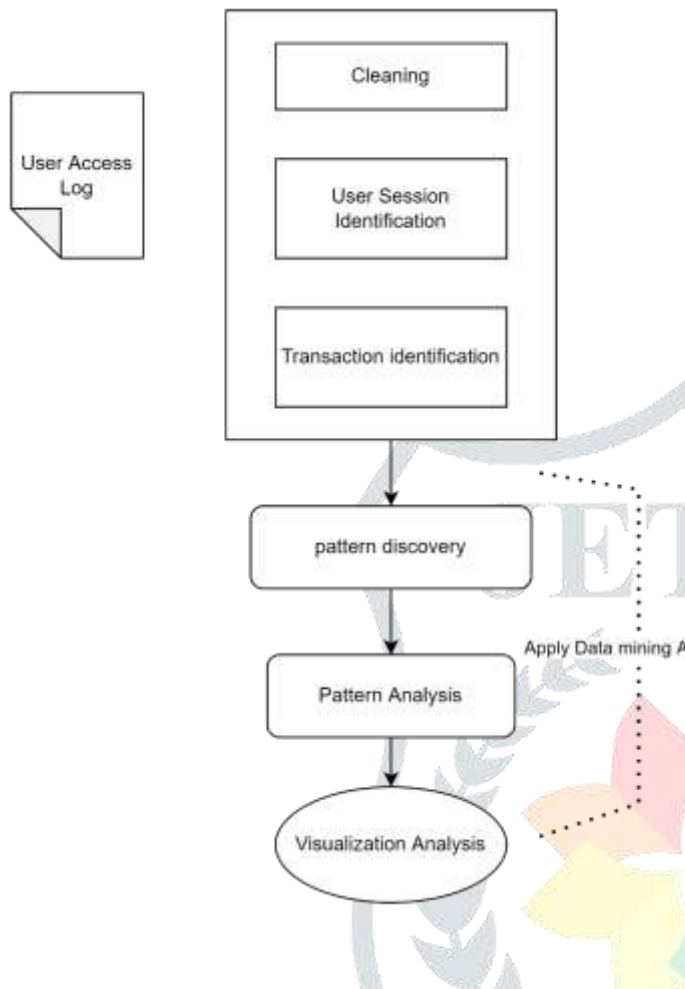
## Data Flow Diagram Level-0



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

## Data Flow Diagram Level-1

## Data Flow Diagram Level-2



## Conclusion:-

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 analysed 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 centres globally to make full use of the OSNs for crisis communication in order to save more lives during such.

## Advantages:-

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

## 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.

## 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