# Terrorism detection from social media

## Problem Statement

The main objective of the project is to detect terrorism content in social media. The content can be in form of text to spread the terrorist sentiments or threats/messages. The content can also be of the form of terror group images or their pictures of arms and attacks. Since if such content is published on social media it can spread vast unrest and agitation in public of any city, state or nation. Hence a need is there to automatically detect such content in real time and prevent from being uploaded on internet or disabled/removed from social media site if anywhere by chance.

## Background

The terrorist group take help of social media to spread their negative sentiments and hatred messages. Through these social media text and images public is influenced and can resort to agitation and violence. Terror groups also use social media to spread fear in public and can brain wash people of certain regions to fall into such activities. In effect all the paths leading to negative terrorist activity can spread on social media like a fire in forest and can be very harmful.

[1] Ala Berzinji et al. describes number of ways to detect terrorist's groups. The most important person in a terrorist group can be identified using the graph of terrorist's connectivity. The node with highest degree of centrality is probably the person who is most important in terrorism group. The degree of centrality is given by:

Deg(node)/(|V| -1)

Where Deg(node) - degree of any node in the graph or number of other people the node is connected to.
|V| - number of vertices(people) in a graph.

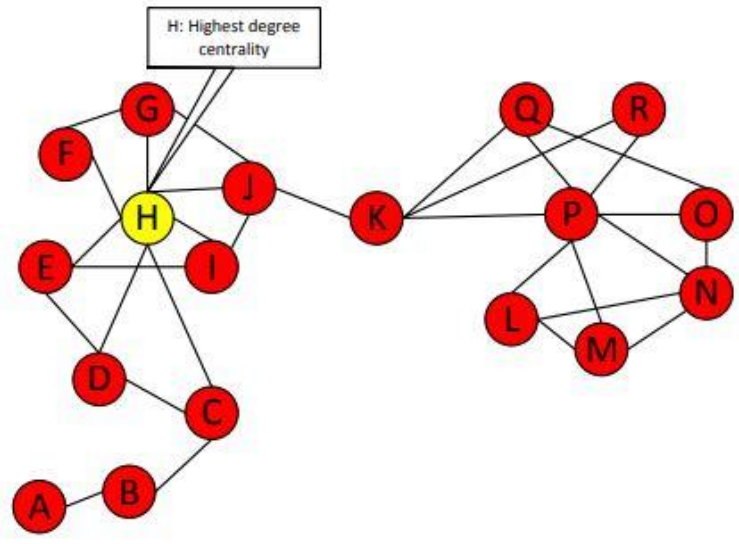


*Fig 1: Graph of terrorist network with key person in center [1]*

[2] Jacob R Scanlon and Matthew S Gerber have tried to explore the possibility of finding terrorist recruitment from internet using social media text. Authors used naive Bayes models, classification trees, logistic regression, boosting, and support vector machines (SVM) to classify the forum posts

and found that mostly the predictions for terrorism groups were matching with real data. They proposed future work for non-English languages.

[3] Robert Pelzer provides an overview on R&D of big data tools using terrorism data from social media. The tools mostly use supervised machine learning algorithms to differentiate between two classes of social media content, e.g. radical and non-radical.

[4] Ilias Gialampoukidis et al. proposes a novel framework that uses a combination of key-player identification with community detection to fetch communities of terrorism-related social media users. Authors experiments show that most of the members of retrieved key-community are mostly already suspended by Twitter, violating its terms leading to their terrorist acitivities with high degree of probability.

**Methodology**

**I** Different Machine Learning Algorithms can be applied to judge the text is having terrorism content or not. We will try two approaches: logistic regression, naïve bayes classifier.

Logistics Regression: it will need preprocessing of text by removing illegal characters and stop words. Then we will train the system by a text file of terrorism content and separate text file of simple news content. We will assign a tf-idf score to each word in a sentence and pass to Neural Network with input layers equal to maximum words in a sentence say 100. We will also give which text belongs to which category so that it can adjust the weights by stochastic gradient process to match the input to correct class: terrorism or not terrorism. When we get any new content, we will simply find its words tf-idf and give to trained Neural Network and predict whether text belongs to terrorism content or not.

Naïve Bayes Classifier: We can calculate using probabilities of a text belonging to a category using the sample equation below:

P(We will attack the city | Terrorism ) = P(We | Terrorism ) * P(will | Terrorism ) * P(attack | Terrorism ) * P(the | Terrorism ) * P(city | Terrorism )

Now we can calculate individual probabilities using formula:

P(attack | Terrorism ) = (frequency of attack in Terrorism samples) + 1)/(total words in Terrorism samples + total distinct words in both sample)

If we get P (Sentence | Terrorism) > P(Sentence | Not Terrorism) then we can say that text belongs to terrorism category.

**II** To classify the images into Terrorism or not terrorism we can use CNN and train with several images of terrorists with arms etc. in class 1 and simple images in class 2. After training when we input new image then it will automatically tell which class it belongs to.

Terrorism data can be collected from links such as:

https://data.world/carlvlewis/terrorism-cases-2001-2016

**Experimental Design**

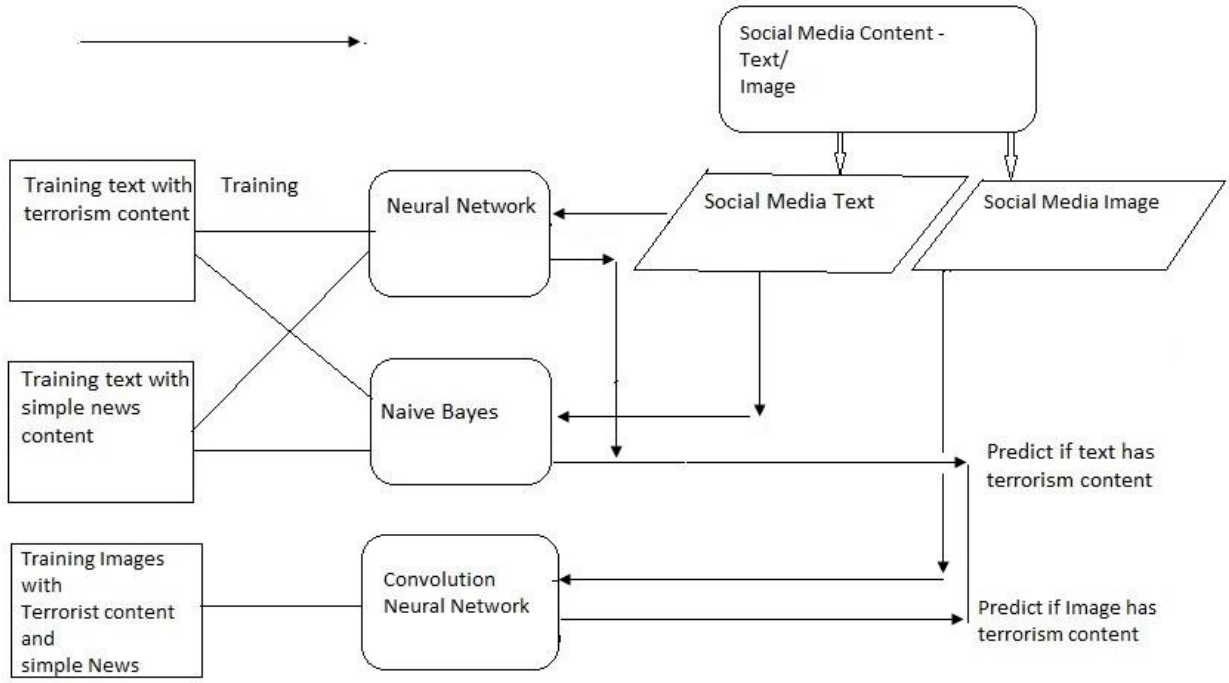


*Fig 2: Architecture diagram for detecting social media data has terrorism content in it*

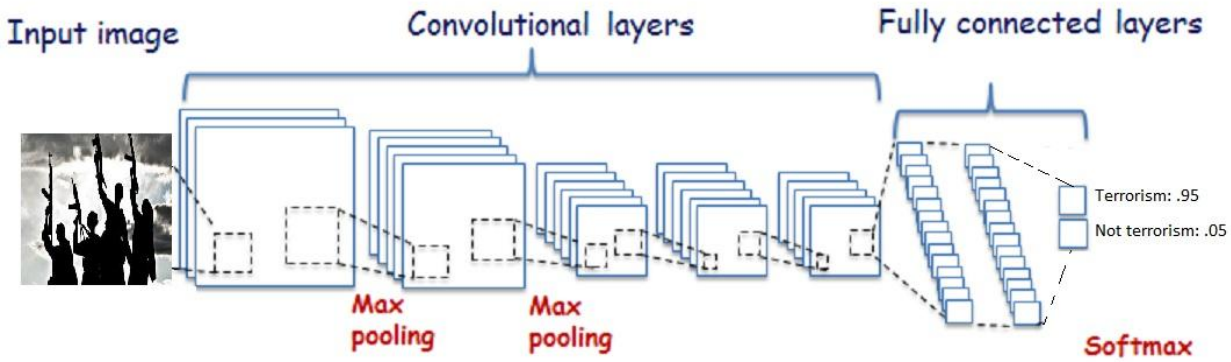


*Fig 3: CNN for predicting if image has terrorism content or not*

**I Text category recognition**

Step 1: Two separate text files are created with terrorism content and one with simple news.

Step 2: Neural Network is trained using tf-idf of these two documents to classify text belongs to which class.

Step 3: A new text on social media is given to pre trained NN and predicted whether it belongs to terrorism content or not.

Step 4: Above 3 steps can also be used for Naïve Bayes classifier.

**II Image category recognition**

Step 1: The terrorism photos are collected from news agency and previous terrorism videos and footages.

Step 2: CNN is trained using python keras and tensor flow to create a weights files from training dataset.

Step 3: Testing image is given to trained CNN and checked which class it is predicting the image to be (terrorism or not).

*Evaluation Measures:*

The text falls into correct category and with good degree of accuracy it can check if text or image has terrorism content in it.

*Software and Hardware Requirements:*

Anaconda with tensorflow can be used to make the logistic regression model and it can easily predict the new text class after training. Simple python logic can be used to train system using Naïve Bayes classifier whether a text is terrorism related or not.

For Image detection Anaconda with spyder is used for CNN which uses python libraries of keras and tensorflow. The hardware needed will be of multi core fast processor or a GPU machine to train on large dataset with epochs more than 40. This will take training time nearly equal to 1 hour. After saving these weights we get a trained model and this is used to predict new image class.

The CNN can be multi layer with 3-4 hidden layers and 3 classes or categories with Relu (Rectified Linear Unit) activation function. The loss function used will be adams optimizer and categorial cross entropy.

**References:**

[1] Detecting Key Players in Terrorist Networks, Ala Berzinji, Lisa Kaati and Ahmed Rezine, 2012 European Intelligence and Security Informatics Conference

[2] Automatic detection of cyber-recruitment by violent extremists, Jacob R Scanlon and Matthew S Gerber, Scanlon and Gerber Security Informatics 2014, 3:5

[3] Policing of Terrorism Using Data from Social Media, Robert Pelzer, Springer International Publishing AG, part of Springer Nature 2018

[4] Detection of Terrorism-related Twitter Communities using Centrality Scores, Ilias Gialampoukidis et al., Proceedings of the 2nd International Workshop on Multimedia Forensics and Security, Pages 21-25, Bucharest, Romania — June 06 - 06, 2017