International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Terrorism Analysis through Social Media using Data Mining

Shamika Karnik[1], Sonal Singh[2], Sahithi Shetty[3], Sujata Kullur[4]

[1,2,3] *Student, Dept. of Information Technology, Usha Mittal Institute of Technology, Maharashtra, India*
[4] *Professor, Dept. of Information Technology, Usha Mittal Institute of Technology, Maharashtra, India*

Abstract - *There has been a tremendous increase in internet usage in the following decade which leads to many great ideas and technologies emerging. But as they say "with good comes bad". So, a lot of malicious activity is also noticed which has been done by unknown groups. Groups can be labelled as fraudster group, terror groups and many other unethical groups. Groups use this social media platform for extracting information, and this piece of information is used by them to expand terrorist activities. This has resulted as a growing activity over the past decades. The basic idea behind this is to put an end to such kind of activities or curb this activity through technology. This paper introduces various deep learning approaches applied to the task of classifying terrorism through toxicity in online comments. Out of which the impact of Deep Neural Network(DNN), Convolutional Neural Networks (CNN), together with word and character-level embeddings, on identifying toxicity in text is implemented. The aim is to provide both the models DNN and CNN because of their capability to provide appropriate results with the help of technical analysis of the data set. The objective is to compare both the models and use the best one suited when it comes to a particular company data set. The results obtained show that the individual models work well when the data provided suits the model and appropriate parameter values are set. The accuracy of the models for each attribute is over 90%. Both the models can prove to be an asset to social media users. Thus, the paper presented provides models which can predict the percentage of a particular tweet and classify it into 6 labels: toxicity, severe toxicity, obscenity, threat, insult or identity hate.*

***Key Words*: Social media, Terrorist activities, Deep Neural Network, Convolutional Neural Network, Deep Learning, GUI.**

1. INTRODUCTION

Terror teams use social media as a medium to extract data and recruit's violence within the society. Social media operate through world platforms with open access that leave digital signatures for enforcement agencies to track them down. Successful operations to curb malicious activities and track those accountable are performed with the assistance of open- source data. By acting smoothly on this obtainable data, it conjointly becomes attainable to prevent incidents before they occur. Terrorist organisations operate in such a way that the social networks ought to be able to give the primary pur- pose of contact with potential candidates for accomplishment. Once initial contact is established additional communication is shifted to associate encrypted channels wherever tracking becomes nearly impossible, which is further complicated by privacy and protection laws prevalent in various countries around the world. Conversational toxicity is a difficulty that may lead people both to prevent genuinely expressing them- selves and to prevent seeking others' opinions out of fear of abuse/harassment. The goal of this project is to use deep learning to spot terrorism through toxicity in text, which can be accustomed to help deter users from posting potentially hurtful messages, craft more civil arguments when engaging in discourse with others, and to determine the toxicity of other users' comments.

This project aims to implement various deep learning mod- els - specifically multilayer perceptron (MLP),Deep Neural Networks(DNN) and Convolutional Neural Networks (CNN) - to tackle the above task, assessing these models' performances on both binary and multi-label classification tasks. Our project also studies the applications of those models at both word-level and character-level granularities.

2. REVIEW OF LITERATURE

The basis of study idea: The fastest growing area within the research is web data processing. It consists of web data processing, web structure mining, and website mining. Web data mining is that the process of discovering the users accessing pattern from the online data logs. the net structure mining is helpful in discovering the knowledge from the structure of the hyperlinks. The Paper presented by George Kalpakis, Theodora Tsikrika, Stefanos Vrochidis and Ioannis Kompat- siaris their have examine a social media network consisting of terrorism related Twitter accounts. Our data were collected through asocial media discovery tool executing queries on the Twitter API (https://dev.twitter.com/) employing a set of

Arabic keywords associated with terrorists' propaganda. and next their examine which centrality measure is in a position to detect the foremost influential Twitter user(s) and limit the communication effectively within the terrorist network.(1) The paper presented by Taiwo Kolajo, Olawande Daramola is to present how big data will be accustomed prevent and combat terrorism in developing countries. First, we present a bird-eye view of the analytical procedure for handling Big Data for terrorism detection, second, the architectural framework, thereafter, an outline of key components of the architecture and lastly, a use case scenario. during this paper, a use case of Social Media Analysis for Combating Terrorism (SMACT) Model was also presented so as to demonstrate the plausibility of its implemen- tation to resolve real security problems that relates to terrorism.(2) The paper presented by Ali Alzahrani, Khalid Bashir Bajwa, Turki Alghamdi and Hanaa Aldahawi their presented the methods utilized by terrorists to spread their messages using social media. it's understood that containing terrorism- related material on social media is critical. We analysed the associated problems and proposed strategies towards an answer for containment of terrorism-related activities. The proposed strategy includes the employment of linguistic communication processing to create and expand a dataset by searching for terrorist related data on social networks, the reduction of the info using data sampling techniques, the employment of information mining methods on reduced data to spot the patterns and extract useful information and therefore the use of social network analysis to uncover the associations and relationship between individuals and terrorist groups, their structure and their modes of operations.(3) The Journal written by Aakash Negandhi, Soham Gawas, Prem Bhatt, Priya Porwal during this the system detects patterns and relevant information in texts in websites using data processing. The web site will have these characteristics: Load balancing, easy accessibility, user friendly , efficient and reliable and simple maintenance. This technique has used basic data processing moreover as a web mining concept . They used Visual Studio and Sql Server 2008 to make and record the project.(4) The article written by J. Kiruba, P. Sumitha, K. Monisha, S. Vaishnavi during this the system delivers event notification which is employed to watch the activities and delivers notification consistent with the investigation knowledge. The monitored population is unbalanced so the entire population of internet users is really near low, that's zero. it's not expected that the terrorists activities on the net to be a stable either-in. This study demonstrates how to use search engines to categorise websites into the categories of terrorism and anti-terrorism.This study results the concepts and methods to detect terrorists and their behaviour and their supporters who supports for terror related activities in society using an online access.(5) The journal written by Mrs. T. Suvarna Kumari, Mr. Narsaiah Putta during this framework will check website pages whether a page is advancing fear based oppression. This framework will characterize the web site pages into different classes and kind them fittingly. There are two highlights utilized during this framework that's information mining and web mining.

During this paper, the info pre-handling could be a significant undertaking of TTUM application. Accordingly web mining procedure will be utilized for distinguishing and maintaining a strategic distance from fear dangers led to by psychological oppressors everywhere throughout the planet.(6)

3. METHODOLOGY

This paper aims to make the simplest model to predict and analyze the epochs values for the Terrorism Analysis. The paper presents two models CNN and DNN for the analysis which can help the concerned officials to create a more informed decision. The structure of the model presented is as shown in figure below. The software framework is made using python for the 2 algorithms of CNN and DNN.

3.1 Dataset

Cleaning and filtering of the data set is done to remove duplicate records, normalize the values, accounting for missing data and removing irrelevant data items. The training data set is provided to the classifier as input. This classified data is als used for the purpose of testing. The system will operate mainly in two stages:

- Training phase
- Testing phase

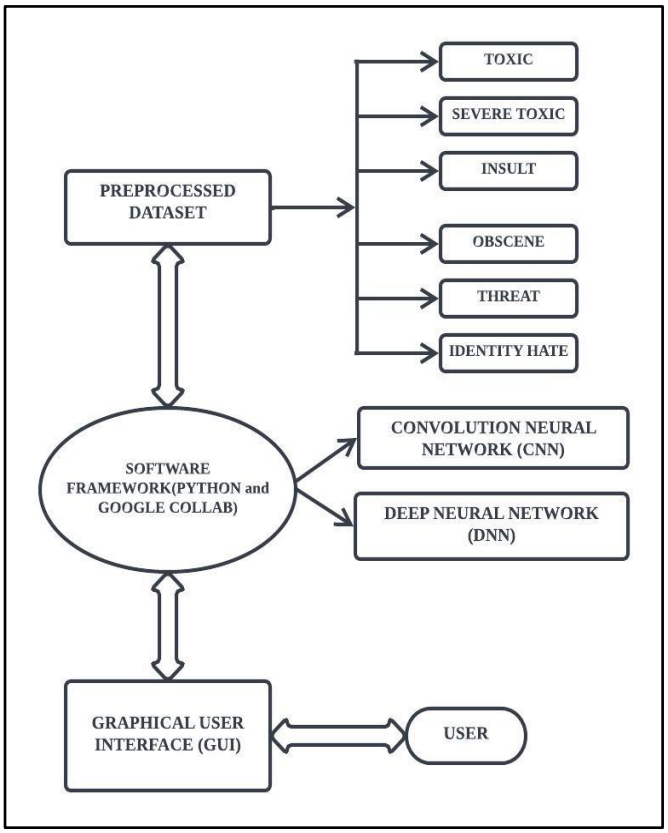

Fig -1: Model Architecture

3.2 Training Phase

Training a model simply means learning (determining) good values for all the weights and the bias from labeled examples. The goal of training a model is to find a set of weights and biases that have low loss, on average, across all examples. Classification assumes labeled data: we know how many classes there are, and for each class we have examples (labeled data).

3.3 Convolutional Neural Network

Convolutional Neural Network is a class of deep neural networks which is used in Computer Vision. A CNN generally comprises three layers: a convolutional layer, a pooling layer, and a totally connected layer.

3.3.1) Convolution Layer: The convolution layer is the core building block of the CNN.It carries the most portion of the network's computational load. This layer conducts a real between two matrices, one of which is a set of learnable parameters referred to as a kernel, and the other is the restricted region of the receptive field. The kernel is smaller in size than a picture, but it has more depth. A stride is the size of the kernel when it slides. If we consider an input of size W x W x D and Dout number of kernels with a spatial size of F with stride S and amount of padding P, then the scale of output volume will be calculated by the formula:Formula for Convolution Layer:

$$Wout = ((W - F + 2P)/S) + 1 \quad (1)$$

It will give an output volume of size Wout x Wout x Dout.

3.3.2) Pooling Layer: The pooling layer is responsible for replacing the network's output at specific points by calculating a summary statistic of neighboring outputs. This minimizes the representation's spatial size, which reduces the amount of computation and weights required. If we have an activation map of size W x W x D, a pooling kernel of spatial size F, and stride S, then the scale of output volume will be calculated by the subsequent formula: Formula for Padding Layer:

$$Wout = ((W - F)/S) + 1 \quad (2)$$

It will give an output volume of size Wout x Wout x D.

3.3.3) Fully Connected Layer: In this layer, neurons have full connectivity with all other neurons within the preceding and succeeding layer as seen in regular FCNN. A matrix operation followed by a bias effect can be used to calculate it. The FC layer aids in the mapping of representations between the input and output.

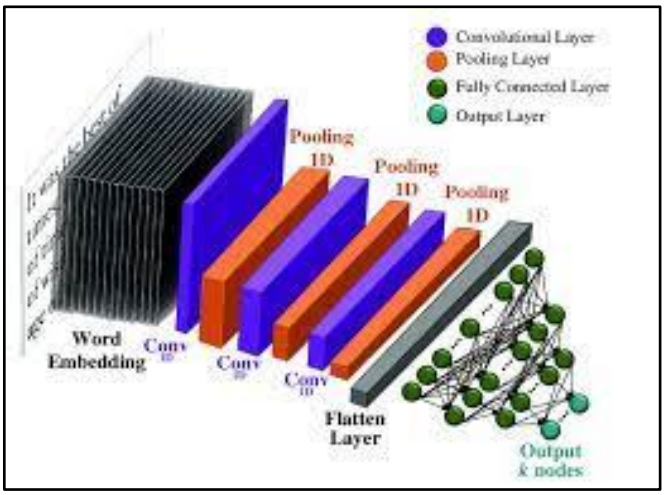

Fig -2: CNN architecture for text classification,by researchgate.net

3.4 Deep Neural Network

A deep neural network (DNN) is a synthetic neural network with multiple layers between the input and output layers.It is a set of neurons organized in a sequence of multiple layers. Here,the neurons receive input from the neuron activations from the previous layer, and perform

an easy computation for example we use a weighted sum of the input followed by a nonlinear activation. The neurons of this network jointly implement a posh nonlinear mapping from the input to the output. This mapping is learned from the information by adapt- ing the weights of every neuron with the assistance of a way called error backpropagation. A deep neural network (DNN) will be considered as stacked neural networks, i.e., networks composed of several layers. FF-DNN, also called multilayer perceptrons (MLP), are because the name suggests DNNs where there's quite one hidden layer and therefore the network moves in barely forward direction (no loopback). These neural networks are well-suited to both classification and prediction. For spoken LID, we use the classification approach. When the FF-DNN is employed as a classifier, the input and output nodes will match the input features and output classes. The most important concepts in an exceedingly FF-DNN are weights, biases, nonlinear activation and backpropagation. The input layer has four elements which are I-1, I-2, I-3, and I-4. They define an input, I = . What we want is to seek out one or more patterns from the entities of the input, so those patterns are accustomed to classify one output from the opposite. so as to try and do that, we devise a variety of hidden units with activation functions. Nonlinear activation functions like Sigmoid, tanh, ReLU are accustomed to form a pattern of active hidden units.

Let's assume H-13, H-14, H-17, H-22, H-23, H-26,H-32, H-33, H-36, and H-37 are needed for active output O-1. To achieve this, we need to adjust the weights and biases in such a way that these hidden nodes are activated by the activation function. Firstly, the weights and biases are randomly initialized. Later, we train this network with tens of thousands of inputs. We make use of backpropagation of error to adjust weights and biases, so that the suitable values among them will activate the hidden neurons. The set of weights and biases are called feature sets or kernels., which will be used to uniquely identify a particular output.(8)

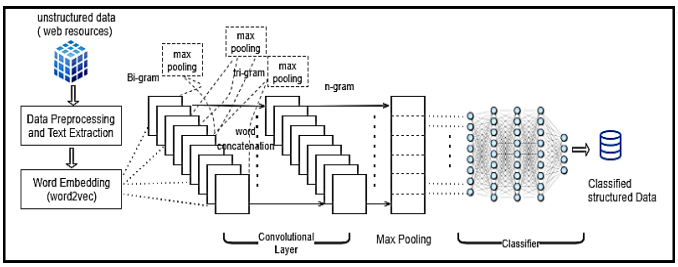

Fig -3: Text classification using a multilayer deep neural network architecture, via researchgate.net

3.5 Comparison between the two models

The two models were compared and supported the methods employed for pre-processing, and processing the info for predictions, the following table shows the comparison between them.

Comparison Between CNN and DNN		
Sr No.	CNN	DNN
1.	Convolutional neural networks (CNN) uses a vari- ation of multilayer percep- trons and contains one or more convolutional layers that may be either entirely connected or pooled.	Deep Neural Network (DNN) is comparable to artificial neural networks and aims to mimic the data processing of the brain.
2.	In CNN rather than weights we use kernels that result in weight sharing.	In DNN there's no Weight sharing.
3.	CNN uses a convolution operation which represents a selected filter.	DNN focuses more on data from input represented by a bunch of non- linear functions.
4.	Consists of multiple layers which are convolutional layer, pooling layer, fully connected layer, and nor- malising layer.	Consists of more than 1 hidden layer.
5.	When data flows from input layer to output layer, CNN remembers the input topology.	When data flows from input layer to output layer,DNN forgets the input topology.
6.	Automatically detects the important features with no human supervision.	Operating of neural networks should be trained under human supervision.

Table -1: Comparison between the models

3.6 Testing Phase

Testing phase involves the prediction of unknown data sample. The test set is a set of observations used to evaluate the performance of the model using some performance metric. It is important that no observations from the training set are included in the test set.

4. RESULTS

Testing phase involves the prediction of unknown data sample. The test set is a set of observations used to evaluate the performance of the model using some performance metric. It is important that no observations from the training set are included in the test set.

4.1 DNN Model Results

It was observed that the number of epochs highly impacted the accuracy of the model. Also, the model performed well when a large dataset was included.

The accuracy using the DNN model was 97.21% .

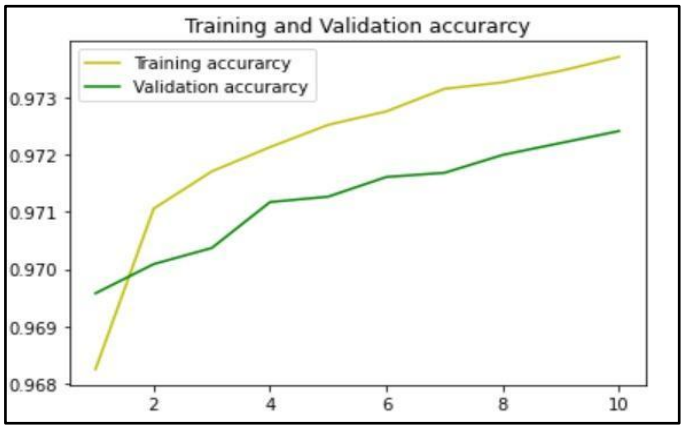

Fig -4: Training and Validation accuracy by DNN model

Fig -5: Training and Validation loss by DNN model

4.2 CNN Model Results

The model was trained using 1D Convolutional Layer and tested. The accuracy using CNN model was 97.64% .

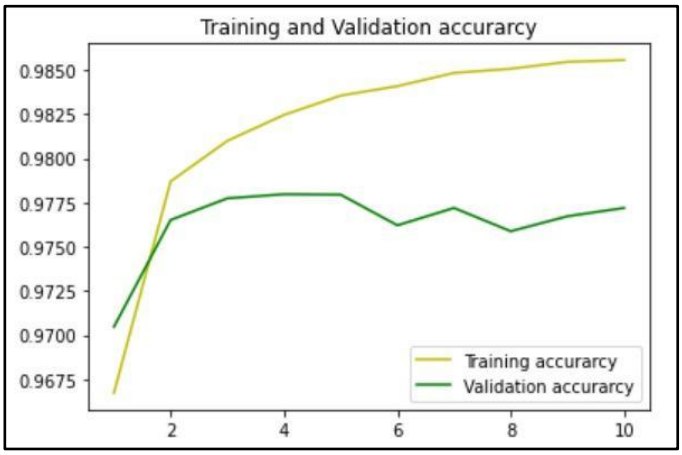

Fig -6: Training and Validation accuracy by CNN model

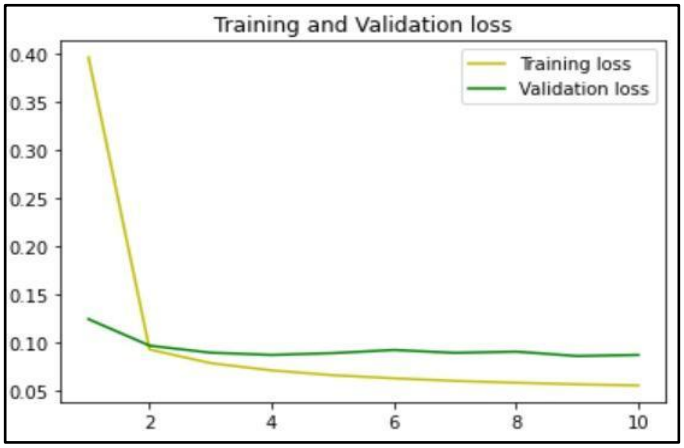

Fig -7: Training and Validation loss by CNN model

5. CONCLUSIONS

Social media is said to be a trend nowadays which has integrated technology, social interactions, construction of words, pictures, videos and audios. It helps to induce access to friends, tweets and user credentials. The method of Extraction and Analysis is the challenging attribute in Social Media because the data is dynamic and data is unstructured in numerous sites. The project is principally for the scholars to assist them in understanding the steps of extracting content from different sites and are available up with new advancement in developing applications by using the active data. The info Mining practices and algorithms is to develop different applications and help in improving the status of the marketing field. This is often our initial study which supplies basics about a way to start

with the applying development. Further studies are added with improved algorithms and process of application development. The proposed model is currently offline in future work; it is extended for real-time Twitter data streaming to predict further crimes. More crime classes will be added to make the system efficient and robust.

ACKNOWLEDGEMENT

We have a great pleasure to express our gratitude to all those who have contributed and motivated us during our project work. We would like to express my special thanks to our guide Prof. Sujata Kullur as well as our principal Dr. Shikha Nema who gave gave us the golden opportunity to do this wonderful project on the topic: Terrorism analysis through social media using data mining, which also helped us in doing a lot of Research and we came to know about so many new things. We would also like to thank Dr. Sanjay Shitole and Prof. Mohan Bonde for their valuable inputs in completion of our project.

We are really thankful to them. Secondly we would also like to thank our parents and friends who helped us a lot in finalizing this project within the limited time frame.

REFERENCES

[1] Ilias Gialampoukidis, I., Kalpakis, G., Tsikrika, T., Vrochidis, S. \& Kompatsiaris, I. (2016, August). Key player identification in terrorism-related social media networks using centrality measures. In 2016 European Intelligence and Security Informatics Conference (EISIC) (pp. 112-115).

[2] Taiwo Kolajo and Olawande Daramola. "Leveraging big data to combat terrorism in developing countries." In 2017 Conference on Information Communication Technology and Society (ICTAS), pp. 1-6. IEEE, 2017.

[3] Alzahrani, Ali, Khalid Bashir Bajwa, Turki Alghamdi, and Hanaa Aldahawi. "Countering Terrorism on Social Media Using Big Data." (2018): 35-42.

[4] Negandhi, Aakash, Soham Gawas, Prem Bhatt, and Priya Porwal. "Detect Online Spread of Terrorism Using Data Mining." IOSR Journal of Engineering (IOSRJEN) 17 (2016): 19.

[5] J. Kiruba, P. Sumitha, K. Monisha, S. Vaishnavi(2019), “Enhanced Content Detection Method to Detect Online Spread of Terrorism”.

[6] Mrs. T. Suvarna Kumari, Mr. Narsaiah Putta (2019),” Web Data Mining to Detect Online Spread of Terrorism”.

[7] Sanskar Soni, Satyendra Singh Chouhan and Santosh Singh Rathore (March 2022),”TextConvoNet:A Convolutional Neural Network based Architecture for Text Classification”.

[8] Shikha Jain, Ajit Kumar Jain and Shashi Pal Singh (Jan 2021),”A Multilayer Deep Learning Framework for Auto-content Tagging”.