Analyzing Threat Levels from Tweets of Extremist Politicians

Divya Pandey
MT20128
divya20128@iiitd.ac.in

Waquar Shamsi
MT20073
waquar20073@iiitd.ac.in

Shubham Bhansali
MT20105
shubham20105@iiitd.ac.in

Kunal Anand
2018293
kunal18293@iiitd.ac.in

Reshan Faraz
Phd19006
reshanf@iiitd.ac.in

ABSTRACT

In the age of social media communication, it is easy to modulate the minds of users and also instigate violent actions being taken by them in some cases. There is a need to have a system that can analyze the threat level of tweets from influential users and rank their Twitter handles so that dangerous tweets can be avoided going public on Twitter before fact-checking which can hurt the sentiments of people and can take the shape of violence. The study aims to analyse and rank twitter users according to their influential power and extremism of their tweets to help prevent major protests and violent events. We scraped top trending topics and fetched tweets using those hashtags. We propose a custom ranking algorithm which considers source based and content based features along with a knowledge graph which generates the score and rank the twitter users according to the scores. Our aim with this study is to identify and rank extremist twitter users with regards to their impact and influence. We use a technique that takes into consideration both source based and content-based features of tweets to generate the ranking of the extremist twitter users having a high impact factor.

1. INTRODUCTION

The rapid growth of the internet has brought many changes in the sharing and obtaining of information with the increase in social media use by active users, and social media has become the most famous manifesto in the 21st century. In today's world, when there are several options to express ourselves or know information on a particular topic, which is an advantage if we see from one side; however from the other side, there is a disadvantage that people can easily get manipulated by reading false claims or news, and it has created opportunities to monitor the users. One of the social media is Twitter. Twitter is a micro-blogging social networking service where users can put short messages referred to as tweets that can be shared and liked. Twitter can barter not just the tweet messages but also the dynamics of political misconduct. Twitter plays a vital role in understanding the political alignment of users. For example, it helps us to know the right-wing extremist and left-wing extremist. Knowing that the politicians have progressively used Twitter to show their trust in an ongoing political affair, Twitter has recently become political advocacy in spreading political advocacy. The algorithm design and implementation help identify and rank the Twitter users who are extremists and whose threat level can impact heavily to society. For this custom ranking algorithm with the idea of a knowledge graph has been used. This considers both source-based and content-based features.

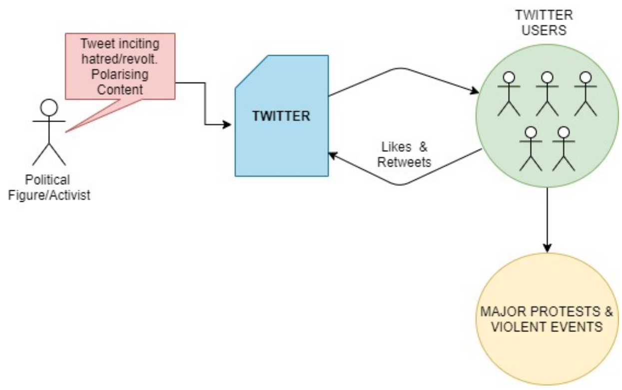

Figure 1: Interaction of users on Twitter

2. LITERATURE REVIEW

In [1], the author took 863 Twitter handles and categories into eight different types based on their role in society. Now they try to find the interaction between the accounts with the help of a directed graph where each node represents an account, and each edge represents the interaction based on the retweets, comments, and tweets related to an article that mentions some account node. The authors concluded the result by mentioned how different Twitter Handles engaged different communities based on their popularity. The result shows that tweet of an official news channel is trusted by people. Results also show that tweets involved in debate primarily engaged with the same geographical area tweet handles. However, the accounts of the official news outlets and the deception of those outlets were ignored which is a limitation.

In [2], the researchers have provided numerous methods to understand the right and left-leaning of political alignment and recognize websites frequently visited and tweeted by Twitter users. They collected tweets of three months during the 2010 U.S. mid-term elections and by making use of different approaches identification of political hashtags was done. First, the researchers did two kinds of feature analysis: content-based feature and network-based feature. For network analysis, they build two networks based on mention and retweet from Twitter. A force-directed algorithm was used. They did direct examination by combining information which is topological and content data. For content-based features, they made use of TF-IDF vectors. After that, they applied linear SVM to restrict the left and right-leaning users. People who utilized left leaned hashtags appeared with negative, and right leaned appeared with the positive weight of hashtag coefficient. The researcher culminated that high comprehension hashtags are highly fertile in providing political leaning and dictating about the websites which are the most tagged and frequently look in on by users on Twitter. The ranking list is generated according to domain popularity tells the websites most frequently visited by the users of Twitter. The researchers have figured out that the tweet data generated on Twitter plays a vital role in shaping political opinions and impacting the user. However, the proposed idea lacks and limits to some of the aspects like generalizability of the proposed approaches to the international level of political consultation and systems for multiple parties.

In [3], the authors observed that the features based on content were as important as the features based on the source on Twitter with respect to credibility; their work remains limited by human annotation to obtain ground truth.

In [4], authors show that identifying polarizing content on Twitter based only on the content may generate false positives; instead, they proposed identifying the radical content using behavioral and psychological properties as well. The authors used TF-IDF scores of uni-gram, bi-grams, and tri-grams and used the word2vec model for word embedding generation. For extracting the radical language, the authors calculated TF-IDF scores for every gram and used word embedding for capturing semantic meanings. However, their work remains limited by different evasion techniques.

In [5], the group of researchers created 2 kinds of fake news datasets: the first one is Fake-news Dataset which comprises six different categories of news and the second one is Celebrities Dataset. Then the authors developed a classification model based on several linguistic features like readability, punctuation, complete LIWC, etc., and then trained the model and analyzed the performance of the model on different sets of linguistic features. When trained the model on their respective dataset, the model worked fairly, getting an accuracy of above 0.5 on most of the features. The best performing classifier for the FakeNewsAMT dataset was derived from the Readability features and for the Celebrity dataset, the best performing was derived using the Punctuation features. But, one of the limitation was that when did the cross-domain analysis on these two datasets, there is a significant loss in accuracy if compared with the "within-domain" results.

In [6], the researchers have proposed an approach to semantic categorization of multimedia where they have performed Entity Linking for text content and extracted Semantic Concepts for visual contents. Eventually, this labeling process made the analysis easier for them. According to the researchers, the graph-based approaches can find the unseen relations among the multimedia contents, and therefore they have investigated the capability of Graph convolutional networks (GCNs) for the same contents. They have provided a graph containing encoded values of blog posts as nodes and encoded multimodal relations as edges in the form of an adjacency matrix as the input to the GCNs. After training, the researchers compared a list of GCNs to Multi-Layer Perceptron (MLP) with the report containing Precision, Accuracy, Recall, and F1 metrics and in all the metrics, the proposed approach of GCNs exceeded the baseline by a significant margin in the domain of performing the qualitative analysis of extreme information related to politics. However, multiple GCNs were compared to MLP, and among the list of GCNs, GCN-TU outperformed other GCNs as well as MLP in Accuracy, Recall, and F1 while GCN-TECU outperformed other GCNs and MLP in Precision.

3. DATASET

Election Data was scrapped from twitter using Tweepy API. 84561 tweets were collected. Firstly, manual annotation of 1068 tweets during phase 1 of project later annotation of remaining 83493 tweets were done using the classifier model i.e by making use of Multilevel perceptron during phase 2 and 3 as it performed best.

4. PLAN OF WORK

The project aims to propose a system that can analyze the impact of extremists on twitter and rank them by making use of knowledge graph and a custom ranking algorithm which considers different factors.

- First, to predict the tweets belonging to different categories, i.e., Neutral, Moderate, and Extreme. The following steps show the way to generate the classes for each tweet.
  - Downloading of Election Tweets.
  - Pre-processing of Tweets to get cleaned Tweets.
  - Labelling and Manual annotation of 1068 distinct Tweets on three categories Neutral, Moderate, and Extreme.
  - Using Word2Vec as feature vector representation then applying of Baseline Models and training of model on those Tweets.
  - To predict the labels of new Tweets through models applied and these labels would be further used for ranking of Twitter handles.
  - Obtaining Training and Testing Accuracy.

  We use this feature in the Figure 6: Prototype Model - 2 for ranking the twitter users. To rank the twitter users following two methods are developed

- Considering the two criteria Content-based and Source-based, ranking of twitter users is done. The whole plan of work and system working is shown in figure 2 : Plan of work: Architecture

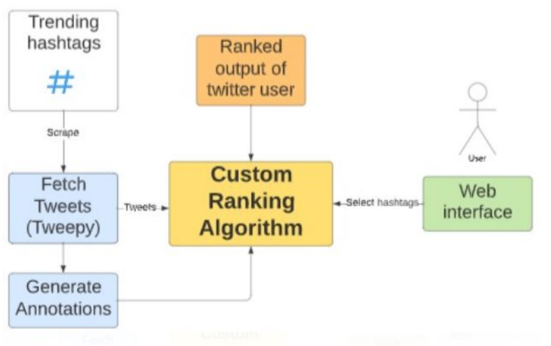

Figure 2: Plan of Work : Architecture

5. BASELINE RESULTS

Machine Learning models were applied to predict the labels after the model being trained on 67% of the data set and tested on 33% of new tweets. Predicted labels show three categories of Tweets. The Result Table 1 gives the baseline results.

5.1 DECISION TREE CLASSIFIER

By making use of the attribute selection measure, it chooses the best attribute to divide the records. A further attribute is made as a decision node and breaks the data set into smaller subsets. This process is followed recursively until the halting condition comes and a tree is built completely. The project model is trained on 67% of the data, and for the rest, 33% testing is being performed. The achieved training and the testing accuracy is shown in the Result Table 1.

5.2 MULTINOMIAL NAIVE BAYES

It is a probabilistic classifier that assumes that the features it uses are conditionally independent of each other. Multinomial Naive Bayes makes us understand that each $p(f_i|c)$ is a multinomial distribution. Given some class **c** to find the probability of features, let's say from f_1 to f_n, then Naive Bayes holds the following:

$$p(f_1, \ldots, f_n|c) = \prod_{i=1}^{n} p(f_i|c)$$

The project model is trained on 67% of the data, and for the rest, 33% testing is being performed. The achieved training and the testing accuracy are shown in the Result Table 1.

5.3 SUPPORT VECTOR MACHINE

It is one of the most robust supervised learning algorithm used for regression as well as classification problems both, as it creates the best decision boundary. It works in two ways, both linear and non-linear. The two-dimensional linearly separable data can be separated by a line $ax_1 - x_2 + b = 0$. The project model is trained on 67% of the data, and for the rest, 33% testing is being performed. The achieved training and the testing accuracy is shown in the Result Table 1.

5.4 RANDOM FOREST CLASSIFIER

It is a supervised learning algorithm. Random forests generate decision trees on randomly chosen data samples. Then it tries to obtain predictions from each tree. After that, it selects the best possible solution using polling or voting. The project model is trained on 67% of the data, and for the rest, 33% testing is being performed. The achieved training, and the testing accuracy are shown in the Result Table 1.

5.5 MULTI-LEVEL PERCEPTRON

It is a kind of feed forward ANN. It uses multiple level of perceptron layers. It has activation function which is considered as threshold. When a particular threshold is reached the perceptron gets fired. It consists of three layers mainly input, hidden and output layer. Every node is a perceptron and gets fired depending on whether it cross the threshold or not. If yes perceptron it gets fired else doesn't gets fired.It makes use of loss function gradient to update parameters as shown below: This deep learning architecture provided the highest accuracy after the baseline and thus was used to further analysis. The achieved accuracy is shown in the Result Table 2. Below formula tells to update the parameters by making use of loss function gradient.

$$w \leftarrow w - \eta(\alpha \frac{\partial R(w)}{\partial w} + \frac{\partial Loss}{\partial w})$$

5.6 LOGISTIC REGRESSION

It is a supervised statistical learning technique. It uses the concept of probability. It uses sigmoid as a cost function instead of the linear function. This sigmoid function is also called a logistic function. This logistic function maps the real value to another value between 0 and 1. The project model is trained on 67% of the data, and for the rest, 33% testing is being performed. The achieved training and the testing accuracy are shown in the Result Table 1.

5.7 CUSTOM RANKING ALGORITHM

The algorithm used takes source based and content based features as a parameter to generate a score for each user according to which they are ranked. Normalization of all features is essential to prevent dominance of subset of features in determining the score. The algorithm gave the following ranking on the dataset used as depicted in the Figure 6: Prototype Model-2.

6. PROPOSED METHOD

In this project, we propose to develop a system that could do ranking of the twitter users based on various algorithms taking different factor into consideration. Custom ranking algorithm is applied to generate ranking of most extremist users by taking the basis and cognizance from knowledge graph. In this we use some content based and source based features of tweets to rank the twitter users for their extremism and impact. The content based features include the polarity of sentiment and a categorical feature which classifies any tweet as 'Normal', 'Moderate' or 'Extremist'. For this we use supervised algorithms where we trained on the manually annotated dataset. This helps in understanding the nature of the tweet. We assign labels as Label 0 for Neutral, Label 1 for Moderate, and Label 2 for Extreme. The source based feature include number of followers of the twitter user and the number of tweets the user made, which can provide insight on reachability of the user's content. The degree of centrality and degree of betweeness are the two major parameters taken into consideration from the knowledge graph.In this project, we have used Twitter as the source of data collection, and this dataset is prepared by first scraping the top trends from a website using beautiful soup after which theses trends are used to fetch tweets using tweepy library to create the entire dataset. It is very important to have a good representation and quality of data before applying the model for analysis. Thus Pre-processing is an important step that improves the generalizability of a model.

6.1 KNOWLEDGE GRAPH

Knowledge graph contains set of vertices and edges where data points are interlinked and represents relation and provides meaningful information. A graph G(V,E) where V represents vertices i.e Twitter users, E represents the edges i.e represents the relation based on the mentions of the tweets by user. The degree of centrality and betweeness centrality are the major parameters used. Betweeness centrality algorithm gauges the shortest path between each and every pair of nodes. The graph made is directed. This made use of (Breadth first search) BFS algorithm internally. Score is calculated for every score.This score is dependent on the shortest path that goes via the nodes. Higher betweeness centrality score is obtained when the nodes comes in between the shortest path. Degree of betweeness is calculated using:

$$\frac{1}{(N-1)(N-2)} \sum_{i \neq j, i \neq k, j \neq k} \frac{\mathrm{sp}_{ij}(k)}{\mathrm{sp}_{ij}}$$

Figure 3: Betweeness Centrality

Centrality measures highly stressed numbers of paths.Centrality attempts to discover the most important nodes in a graph. centrality is calculated using:

$$\frac{\sum_{j \neq i} \mathrm{dist}(i,j)}{N-1}$$

Figure 4: Centrality

6.2 PREPROCESSING

- **Cleaning of Tweets**: Removal of URLs, Special Symbols, Links, Hashtags, Usernames. Twitter original data contains a lot of HTML entities thus, it is important to get rid of them.
- **Removal of Punctuation**: Removal of unnecessary commas and symbols is very necessary in order to get cleaned Tweets. “.”, “,”,”?” are necessary punctuation and thus required to be retained while others have to be removed.
- **Removal of Duplicates**: It is very necessary to remove duplicate tweets as the data-set downloaded from Twitter contains the same retweets multiple times. It is a very important step to remove duplicate tweets from the Twitter data-set for further processing.
- **Word Tokenization**: Before applying any of the models, it is very important to tokenize the dataset so that we get the words that make processing easier by deep learning models.
- **Removal of Stop Words**: Stop words are commonly used words such as a, an, in which occurs many times in a sentence, and such words do not add much meaning to sentences, so it is important to remove such words before applying deep learning algorithms in further stages.
- **Lemmatization**: Morphological analysis is done by doing lemmatization by removing inflectional endings. This is done so that grouping of inflected items can be analyzed as a single item. For this, Vocabulary is used in a better way.

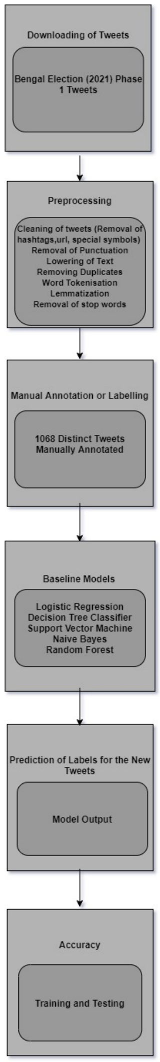

Figure 5: Prototype Model - 1

Starting from duplicates 2467 tweets, we were left with 1068 distinct tweets after processing. These tweets were manually annotated into three categories namely **Neutral**, **Moderate**, and **Extreme** describing the level of extremism of the tweet.Then Word2Vec feature vector representation is applied. In this project, we applied numerous Machine Learning algorithms such as Multinomial Naive Bayes, Logistic Regression, Support Vector Machine, Decision Tree Classifier, Random Forest to predict the level of extremism in any tweet. The training and testing accuracy are depicted in the below Result Table 1. Label prediction is made by the models, and later these predicted labels are used in the ranking algorithm. This feature is then used in the Figure 5: Prototype Model - 3 for generating the ranking of twitter users.

6.3 RANKING GENERATION

For ranking custom based ranking algorithm is used as shown in figure 5: Prototype Model-2

It generates scores for each user available in the dataset, depening upon **source-based features** i.e., the user and **content-based features** i.e., the tweets by the user. The source based features consist of number of followers The content based features include polarity of sentiment of tweet and the level of extremism that the tweet belongs to.

The features are normalized so that one feature does not dominate over others. Knowledge graph is being constructed on those tweets dataset. It takes two major parameters into consideration i.e. degree of centrality and degree of betweeness. Another major factor taken into consideration is Tweet Count. Thereafter final score is generated using all these features and then the ranking is done on the basis on the decreasing value of score as shown in Figure 5: Prototype Model - 2. Highest score implies top rank. The baseline results are shown in Figure 2.

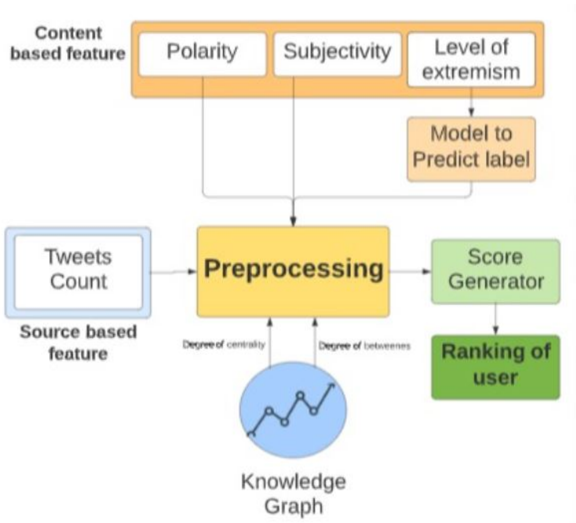

Figure 6: Prototype Model - 2

7. CONCLUSION

The project helped in identifying the most extremist twitter users with the ranking generated. The ranking got improved with the cognizance and output from the Knowledge graph and given to the custom ranking algorithm. The increased accuracy and precision of ranking was due to the extra number of parameters included in the algorithm. The idea of knowledge graph found to be very effective as it helped in identifying the importance of the node. Between centrality and degree centrality was major parameter in the analysis of knowledge graph.The deep learning architecture was proved to be better in analysis in comparison to the baselines models. Class prediction using model MLP helped in identifying the categories which provided us the information about which tweets were moderate, neutral or highly extremist. The above algorithm is thus effective in doing analysis of threat level of twitter users during the high impact event. The UI shows the top ranked extremist twitter users on the basis of hashtags fetching and its tweets collected, by applying the custom ranking algorithm. This indicates that these users are highly responsible for generating the threat level during elections and have high impact during these times.

8. RESULTS

	Names
2420	Anita Pal
2358	Kamlesh Bansal
1595	উজ্জ্বল গোস্বামী(ফুচু)
1860	#BanglaNijerMeyekeiChay
1885	Bjp4Dankuni Mandal (Serampore Org District)
...	...
2192	SURAJ KUMAR
2164	Gulistan News
2180	Feeler
1028	Citizen Durga Prasad Tudu
227	NDTV

Figure 7: Ranked Result

9. EVALUATION

The baseline results for five different models are shown below with their training and testing accuracy.

Baseline Models		
Models	Training Accuracy	Testing Accuracy
Naive Bayes	86.29	73.37
Support Vector Machine	86.29	75.92
Logistic Regression	98.04	77.34
Decision Tree Classifier	98.74	76.49
Random Forest	98.74	76.2

Table 1: Result Table

Application of Deep Learning model Multi-Level perceptron with improved training and testing accuracy.

Deep Learning Model		
Models	Training Accuracy	Testing Accuracy
Multi-Level Perceptron	99.5	82.4

Table 2: New Result Table

10. FUTURE WORK

- Knowledge graph provides a base which is used to strengthen its search results. Improve the speed of generating analysis from Knowledge graph,because decreased speed may hamper user experience. As the knowledge graph is applied on big dataset so it consumes a lot amount of time there is need to improve and speed up so that it can be prevented from hampering the performance.
- Add more analysis on tweets to produce better results. More analysis can be done by considering the retweet just like mentions. Further many different computations can be done on Knowledge graph like Eigen vector analysis.

11. REFERENCES

[1] M. Cinelli, S. Cresci, A. Galeazzi, W. Quattrociocchi, and M. Tesconi. The limited reach of fake news on twitter during 2019 european elections. *PloS one*, 15(6):e0234689, 2020.

[2] M. D. Conover, B. Goncalves, J. Ratkiewicz, A. Flammini, and F. Menczer. Predicting the Political Alignment of Twitter Users. In *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, pages 192–199, Boston, MA, USA, Oct. 2011. IEEE.

[3] A. Gupta and P. Kumaraguru. Credibility ranking of tweets during high impact events. In *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media - PSOSM '12*, pages 2–8, Lyon, France, 2012. ACM Press.

[4] M. Nouh, J. R. C. Nurse, and M. Goldsmith. Understanding the Radical Mind: Identifying Signals to Detect Extremist Content on Twitter. *arXiv:1905.08067 [cs, stat]*, May 2019. arXiv: 1905.08067.

[5] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea. Automatic Detection of Fake News. *arXiv:1708.07104 [cs]*, Aug. 2017.

[6] S. Rudinac, I. Gornishka, and M. Worring. Multimodal Classification of Violent Online Political Extremism Content with Graph Convolutional Networks. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017 - Thematic Workshops '17*, pages 245–252, Mountain View, California, USA, 2017. ACM Press.