International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Identifying Hate Speech and Abusive Content in Social Media Data Based On Machine Learning Based Techniques

Rahul Ringe[1], Prof. Pankaj Raghuwanshi[2]

Abstract- Due to the emergence of social media becoming a common platform for sharing of views and day to day activities, more researches are aimed at extracting data out of social media data mining, one subset of which is text mining of social media data such as twitter, facebook, Whatsapp etc. With the advent of social media, the exchange of data has become extremely easy. However, this sometimes may lead to easy share of radical content which needs to filtered out very quickly to avoid losses, damages and possible scenarios of violence. Machine learning based approaches are indispensable for the detection and classification of the radical and possible terrorist activity conducive content due to the size and complexity of data being bombarded on social media platforms. Due the enormous amount of data available with the websites, they become a natural choice for text mining and/or opinion mining. This paper presents the necessity for text cum opinion mining based Sentiment analysis and its various associated techniques to filter out radical content. It is expected that the paper will pave a path for future researchers to carry forward their research in a direction which best suits their application

Keywords- Opinion Mining, Text Mining, Opinion Mining, Clustering, Artificial Intelligence (AI), Machine Learning (ML)

I. INTRODUCTION

Since Tim Berners Lee proposed the Web for the very first time in 1989, it has been continually evolving towards the current Social Web or Web 2.0. In the Social Web, users have become both content consumers and producers. These contents, either inherited from static Web 1.0 pages or user-generated in the Web 2.0, may range from merely plain text documents to more complex Web resources which present some kind of structure. While traditional Web pages were static interlinked hypertext documents which basically contained raw text or multimedia files, new Web documents are substantively different from prior Web ones, allowing all users to freely contribute. This new scenario presents some interesting points:

- Dynamic and permanently updated content enriches the user experience.
- Information flows bidirectional between site owners and site users by means of evaluation, review, and commenting.
- Site users may add content for others to see, comment, modify and improve (crowd sourcing).
- Web 2.0 sites develop APIs to allow automated usage (e.g. by individual apps or by sites that gather data from different resources and build an aggregated mash up).
- One of the most important key features of the Web 2.0 is that users have the ability to collectively classify information by means of tags, developing free taxonomies of information called folksonomies.

Data mining is an application of data processing in which expert patterns and information is extracted. This extracted information is consumed using applications and actual time programs for making choices. The net is a rich domain of knowledge, potential and leisure. Tremendous amount of consumer's entry internet, in between they are at all times still connected via their buddies making use of these offerings. Often small human grasping nature over internet invites the hidden risks. Analyzing information with all its elements (e.g. temporal and geographical) will cause info which may create easier following and keeping check of your voters in elections. Moreover, the gathered info will cause you to a lot of palm in influencing them and knowing what info, influencing youth, however gaining info on different elements of population also. Should the information lack context, analysis becomes a tricky downside. Statements may become disrespectful, or lose their wit. Straight forward sentences might have their which means inverted. It is, therefore, important to understand the context of knowledge creation, publication and consumption. The basic sentiment analysis techniques are shown in the following figure.

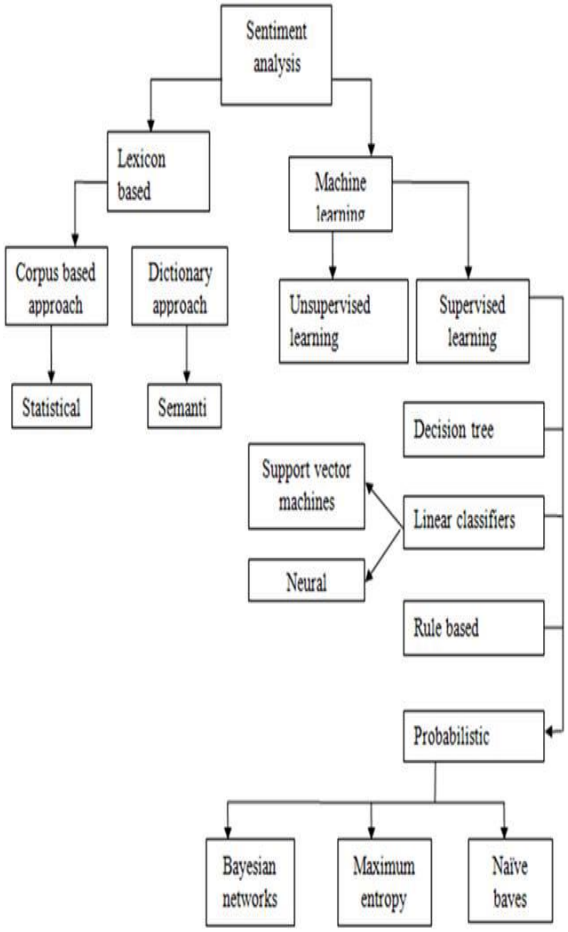

Fig.1 Basic Sentiment Analysis Techniques.

II. OPINION MINING

Opinion Mining has emerged as one of the domains of web mining or data mining that has influenced several domains of day to day life [1]. Some of the common examples are

1) Opinion polling
2) Marketing
3) Advertizing
4) Education
5) Politics
6) Finance and Business Predictive Modelling

The sentiment extraction of users from large and complex data sets is however daunting. This is to be ensured that the context (semantics) is to be taken into account prior to reaching conclusions and implicit meaning has to be inferred correctly. Moreover accurate data pre-processing needs to be imposed in order to segregate the useful information from the raw data. Since user sentiments have a critical impact on several parameters and domains, hence sentiment analysis is critically important [2].

III. PREVIOUS WORK

The various techniques used for text and opinion mining based sentiment analysis are explained in brief below.

In [1], Kapitonov et al. studied about the malicious messages that the terrorist communities used to share through the use of social media. The messages on propaganda are circulated on these instant messaging handles and the social network. The only method to combat such messages is to block the handles and such spread of messages. To carry this out, the researchers require processing significantly huge amounts of data. In this paper, the authors present a method based on artificial intelligence and machine learning that could automate the classification process of separating the radical content. The only limitation of this approach is that it was more based on the conventional statistical implementation.

In [2], Lopez et al. proposed a unique approach and framework for automated monitoring of radical information in the occurring in the social network Twitter. It mainly focussed on two main aspects; detection of the majority of the users who had radical content propagation agenda and then supervision of the interaction of the radical content between users that were involved in such things. The authors also did a case study. The main problem with this approach was that the model couldn't specify the exact demarcating boundary between the content. Many of the datasets overlapped with each other and higher precision was needed.

In [3], Bobashev et al. presented and researched on a case of 1995 Paris metro Attack. Where, the incidents of the classes were traced to find the messages of radical content belonging to them. This led to the origination of the group that did such activities. Dynamic visualization and analysis were performed for the tracing of the formation of the terrorist group. This was the subtle use of the Natural Language processing. The major limitation here analysed is that the NLP used was in its very initial stage and better NLP modelling and methods could fetch better results.

In [4], Sun et al. put forth that method of prediction of the terrorist attacks by groups. They studied that though it is a very important issue around intelligence and security analytics, it is also a hard task to perform it. Conventional schemes and approaches are not enough for classifying such complex datasets. So the there is a need to build a robust and very advanced system to classifying such information. The amount of data that is there on the digital platforms is enormous. Feature extraction is a very important part of machine learning. Better the features extracted, better is the neural network trained. This is one of the areas which could have been improved to increase the accuracy and precision.

In [5], Tundis et al. worked on an approach to state that using a high end computer associated approach could yield viable results. With every positive aspect, there are also chances of

using the social media in a malicious way. In some of the latest research works in this context, it has been found that several radical content are being spread using the medium of social networks. Many illegal community groups who choose to remain anonymous are misusing the platform to spread radical content and misinformation. This approach used text analysis methods by considering multiple language aspects. But one of the major issues that surfaced was that it was hard to classify based on one token of word taken at a time. Better multi token word analysis could be evaluated for the same purpose.

In [6], Zevairi et al. studied the traits and researched on the Islamic terrorist activities and involvement. They specifically studied their characteristics and motives that differentiated from other sets of radicals. Supervised machine learning approach was used to implement this that was a combination of 4. Using the different combination of learning algorithms helped but the understated use of pre processing posed as a major problem. Pre-processing of data is a major necessity for the machine learning methods which helps in improved classification of data.

In [7], Johnston et al. explored about the Sunni extremists groups that were involved in the jihadist radical content propaganda. Dark net is usually a well targeted place of engagement in such illegal and radical activities. Invoking terrorist activities on the cyberspace and spreading communal posts has seen a recent increase. Terrorism on cyberspace is equally dangerous and can inflict harm. This method combines the neural network with deep learning approach. While the results were good and the neural network worked well, but the use of a single neural network could not process the multitude of training datasets. Hybrid or ensemble approach could be used to help with this approach for better outcomes and accuracy.

.In [8], Ishitaki et al. studied about the Tor application that was being utilized for propagating radical content and information. The spread was attributed to the anonymity of the users that helped in such rapid and enormous circulation of the messages. This was the tool for rolling out malicious content on the digital landscape. Conventional schemes and approaches are not enough for classifying such complex datasets. So the there is a need to build a robust and very advanced system to classifying such information. The amount of data that is there on the digital platforms is enormous. It's very important to classify such information to stop it from being broadcasted. They used the deep learning on tor web server. The system could yield better accuracy if probabilistic approach was used for such complex set of data.

In [9], Lourentzou et al. studied on the use of the deep neural networks based on geographical locations. They studied the use on the geographical prediction and tried to figure out the radical content based on locations. Without any specific and clear boundary that can demarcate the radical and non radical data, the classification of such content is also an uphill task. The warfare of recent times is just not confined to the battleground alone. The cyberspace has also become a well targeted medium. Invoking terrorist activities on the cyberspace and spreading communal posts has seen a recent increase. Deep Neural method has been implemented but there need more variety of data sets for proper training and better performance.

In [10], Lara-Cabrera et al. provided insight into the works of some extremists groups that use the medium of social media to propagate malicious content the social network. Terrorism on cyberspace is equally dangerous and can inflict harm. This method combines the neural network with deep learning approach. While the results were good and the neural network worked well, but the use of a single neural network could not process the multitude of training datasets. The indicators of radical content were researched upon. Better pre-processing mechanisms could be enforced for better classification of the data sets used for the analysis of the proposed system.

Salient Takeaways:

The most critical aspects which have been the salient takeaways of the previous works can be summarized as:

1) The approaches used s far typically work on un-pre-processed data which often leads to under-fitted training of machine learning tools.

2) The machine learning architectures do not comprehend the textual semantics of the web mining data and the translation of the textual semantics into the numerical counterparts is often challenging and non-optimized.

3) The weighted parameter approach for couplets of data are not utilized often leading to large divergences in the data.

The above mentioned points lay the foundation of the problem identification.

IV. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING USED FOR OPINION MINING ANALYSIS

Although different mechanisms can be utilized for text mining and subsequent sentiment analysis, the most effective technique that is coming up front is the use of artificial intelligence and machine learning for the classification of sentiment based data yielding to sentiment analysis. Artificial Intelligence is formally defined as the development of computer systems that would perform tasks generally needing human intervention. Such a system generally exhibits the following attributes:

1) Accepting data in a parallel manner.

2) Analyzing data

3) Finding regularities or patterns in the data

4) Producing an output based on the above steps.

The process is often referred to as machine learning wherein the machine is not explicitly programmed for a certain task due to the complexity of the data to be handled and the non-predictive nature of the data to be handled. Such systems mimic the human nature of performing the tasks at the outset with a randomly chosen mechanism and gradually adapting to the changes that yield lesser errors in the output. There are several challenges in sentiment analysis from textual data. The challenges can be stated as follows:

Contextual Analysis

It is often difficult to estimate the context in which the statements are made. Words in textual data such as tweets can be used in different contexts leading to completely divergent meanings.

Frequency Analysis

Often words in textual data (for example tweets) are repeated such as
##I feel so so so happy today!!
In this case, the repetition of the word is used to emphasize upon the importance of the word. In other words, it increases to its weight. However, such rules are not explicit and do not follow any regular mathematical formulation because of which it is often difficult to get to the actuality of the tweet.

Converting textual data into numerically weighted data
The biggest challenge in using an ANN based classifier is the fact that the any ANN structure with a training algorithm doesn't work upon textual data directly to find some pattern. It needs to be fed with numerical substitutes. Hence it becomes mandatory to replace the textual information with numerical information so as to facilitate the learning process of the neural network.

Challenges in Existing Systems
1) Data mining of relevant data which is exhaustive in nature so as to cover most of the cognitive parameters of tweets [2].
2) Making data suitable for analysis by requisite and effective pre-processing.
3) Extracting semantic parts from whole tweets which would in turn reduce the dimensionality of the enormous data size for training based on tokenization [8].
4) Attaining low time complexity in training an exhaustive set of data.
5) Designing some training mechanism for the extracted data which would yield high accuracy of classification.First and foremost, the machine or artificial intelligence system requires training for the given categories [10]. Subsequently, the neural network model needs to act as an effective classifier. The major challenges here the fact that sentiment relevant data vary significantly in their parameter values due to the fact that the parameters for each building is different and hence it becomes extremely difficult for the designed neural network to find a relation among such highly fluctuating parameters. Generally, the Artificial Neural Networks model's accuracy depends on the training phase to solve new problems, since the Artificial Neural Networks is an information processing paradigm that learns from its environment to adjust its weights through an iterative process [19]. The main challenge or shortcoming is to design the Artificial Neural Networks structure using a training algorithm that is:
a) Stable: The inference is the fact that using such an algorithm, the errors should monotonically decrease.
b) Fast: The algorithm should not have excess time complexity. To overcome this shortcoming, evolutionary algorithms are to be used to adopt the search algorithm to evolve the Artificial Neural Networks (ANN) connection weights, learning rules, architectures or the input features [10]. Moreover accurate feature extraction and structuring is necessary to train the Artificial Neural Networks (ANN) accurately.

V. PERFORMANCE PARAMETERS

The performance parameters are often chosen as Accuracy and Complexity of the algorithm. While the accuracy may increase if the complexity of the algorithm increases, but it may adversely affect the applicability of the final system. Often the Mean Square Error (MSE) is computed as the performance metric for the system. It is defined as

$$\frac{\sum_{i=1}^{i=N} e_i^2}{N} = MSE$$

Also the execution time is a critical aspect in the performance of the system.

VI. CONCLUSION

It can be concluded that sentiment analysis has emerged as a field which can have diverse applications in various fields such as banking, stock pricing, politics, social media, advertising, academics etc. While several techniques are available for social media text mining, but Artificial Intelligence and Machine Learning have emerged as the most effective techniques for sentiment analysis of social media data. Also the performance metrics should be kept in mind while designing of any sentiment classification technique.

References

[1] Andrey I. Kapitanov, Ilona I. Kapitanova, Vladimir M. Troyanovskiy, Vladimir F. Shangin, Nikolay O. Krylikov, "Approach to Automatic Identification of Terrorist and Radical Content in Social Networks Message". IEEE 2021
[2] D López-Sáncez, J Revuelta, F de la Prieta, "Towards the Automatic Identification and Monitoring of Radicalization Activities in Twitter," IEEE 2020
[3] G Bobashev, M Sageman, AL Evans, "Turning Narrative Descriptions of Individual Behavior into Network Visualization and Analysis: Example of Terrorist Group Dynamics, IEEE 2019.
[4] Z Li, D Sun, B Li, Z Li, A Li, "Terrorist group behavior prediction by wavelet transform-based pattern recognition", hindawi 2018.
[5] A Tundis, G Bhatia, A Jain, "Supporting the identification and the assessment of suspicious users on Twitter social media", IEEE 2018.
[6] M Al-Zewairi, G Naymat, "Spotting the Islamist Radical within: Religious Extremists Profiling in the United State", Elsevier 2017
[7] AH Johnston, GM Weiss, "Identifying sunni extremist propaganda with deep learning", IEEE 2017
[8] T Ishitaki, R Obukata, T Oda, "Application of deep recurrent neural networks for prediction of user behavior in tor networks", IEEE 2017
[9] I Lourentzou, A Morales, CX Zhai "Text-based geolocation prediction of social media users with neural networks", IEEE 2017
[10] R Lara-Cabrera, A Gonzalez-Pardo, "Extracting radicalisation behavioural patterns from social network data" IEEE 2017
[11] L Ball, "Automating social network analysis: A power tool for counter-terrorism", Springer 2016.
[12] T Ishitaki, T Oda, L Barolli, "A neural network based user identification for Tor networks: Data analysis using Friedman test", IEEE 2016
[13] T Oda, R Obukata, M Yamada, "A Neural Network Based User Identification for Tor Networks: Comparison Analysis of Different Activation Functions Using Friedman Test", IEEE 2016
[14] T Sabbah, A Selamat, MH Selamat, R Ibrahim, H Fujita, "Hybridized term-weighting method for dark web classification", Elsevier 2016
[15] R Scrivens, R Frank, "Sentiment-based Classification of Radical Text on the Web", IEEE 2016
[16] T Sabbah, A Selamat, "Hybridized Feature Set for Accurate Arabic Dark Web Pages Classification", Springer 2015
[17] T Ishitaki, T Oda, L Barolli, "Application of Neural Networks and Friedman Test for User Identification in Tor Networks" IEEE 2015
[18] R Frank, M Bouchard, G Davies, J Mei, "Spreading the message digitally: A look into extremist organizations' use of the internet", Springer 2015
[19] T Sabbah, A Selamat, "Hybridized Feature Set for Accurate Content Arabic Dark Web Pages Classification", researchgate, 2015
[20] Basant Agarwal ,Soujanya Poria,Namita Mittal,Alexander Gelbukh,Amir Hussain, "Concept-Level Sentiment Analysis with Dependency-Based Semantic Parsing: A Novel Approach", Springer 2015
[16] Bing Liu ,Lei Zhang, "A Survey of Opinion Mining and Sentiment Analysis", SPRINGER 2012

[17] Alena Neviarouskaya , Helmut Prendinger , Mitsuru Ishizuka, "Secure SentiFul: A Lexicon for Sentiment Analysis", IEEE 2011

[18] Jorge Carrillo de Albornoz, Laura Plaza, Pablo Gervás, Alberto Díaz, "A Joint Model of Feature Mining and Sentiment Analysis for Product Review Ratings", SPRINGER 2011

[19] Gang Li, Fei Liu, "A clustering-based approach on sentiment analysis", IEEE 2010

[20] Wei Wang, "Sentiment analysis of online product reviews with Semi-supervised topic sentiment mixture model", IEEE 2010

[21] Erik Boiy, Marie-Francine Moens, "A machine learning approach to sentiment analysis in multilingual Web texts", SPRINGER 2009

[22] Songbo Tan, Xueqi Cheng, Yuefen Wang, Hongbo Xu, "Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis", SPRINGER 2009