A Review on Data Driven Approaches for Identifying Potentially Radical Content on Social Media Platforms

Krishna[1], Prof. Ashish Tiwari[2]

Abstract- With advancements in technology and cyber warfare, terrorist and radical content is being used as a tools to spread violence and social unrest. Due to the complexity and the largeness of the data size, it is humanly infeasible to analyze the data manually or use statistical techniques. Thus machine learning based approaches are indispensable for the detection and classification of the radical and possible terrorist activity based content. In this approach, the use of social media messages in the form of tweets and messages have been considered. The paper presents the need for radical content filtration along with the basic models based on opinion mining for identifying potentially radical content. A literature review and theoretical background for identifying radical content in presented in this paper.

Keywords- Radical Content, Opinion Mining, Sentiment Analysis, Machine Learning

I. INTRODUCTION

Opinion mining, also known as sentiment analysis, is the process of identifying and extracting subjective information from text data, particularly in social media platforms. Social media has become a crucial source of information for businesses, governments, and individuals to understand public opinion on various topics. In this report, we will discuss the importance of opinion mining from social media, its applications, and some of the challenges involved in the process. Importance of Opinion Mining from Social Media:

Fig.1. Common Social Media Platforms

Social media platforms, such as Facebook, Twitter, and Instagram, have millions of users, making them a vast source of data for understanding public opinion. Opinion mining from social media can help individuals, businesses, and governments make informed decisions. For example, businesses can use opinion mining to analyze customer feedback, identify issues, and improve their products or services. Governments can use opinion mining to understand public opinion on policies and make decisions that align with the people's views. Applications of Opinion Mining from Social Media:

There are several applications of opinion mining from social media, including:

1. Customer feedback analysis: Opinion mining can be used to analyze customer feedback on social media platforms, identify issues with products or services, and improve customer satisfaction.
2. Political analysis: Opinion mining can help political parties and governments understand public opinion on policies, political events, and election outcomes.
3. Brand monitoring: Opinion mining can be used to monitor brand reputation on social media platforms and identify potential issues before they become bigger problems.
4. Crisis management: Opinion mining can be used to monitor social media during a crisis, such as a natural disaster or a public health emergency, to understand public sentiment and respond accordingly.

Challenges in Opinion Mining from Social Media:

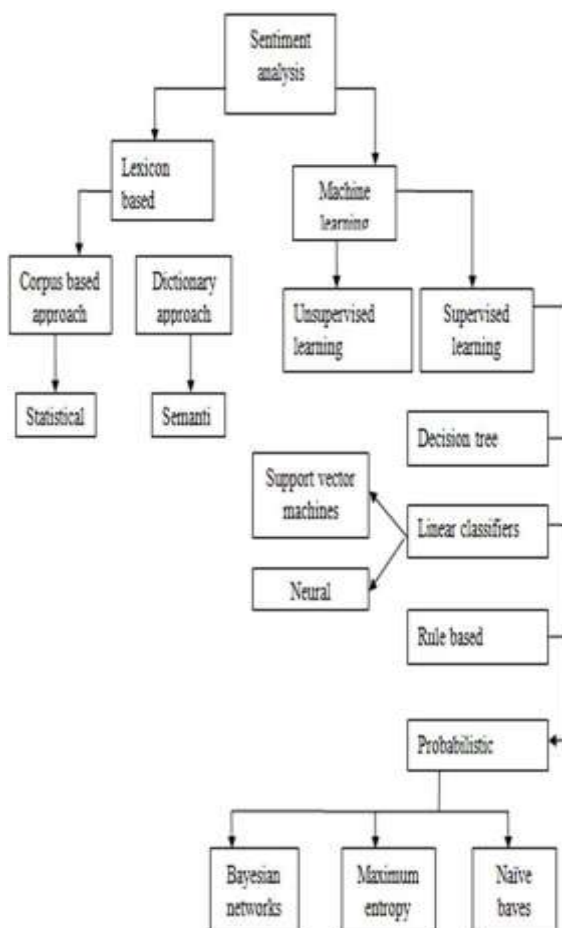Opinion mining from social media is not without its challenges. Some of the challenges include:

Data quality: Social media platforms have a lot of noise, such as spam and irrelevant content, which can affect the accuracy of opinion mining.

Language: Social media platforms have users from different countries and cultures, making it difficult to analyze sentiment accurately in different languages.

Context: The context in which a message is posted can greatly affect its sentiment. Opinion mining algorithms need to be able to understand the context of a message to accurately analyze sentiment.

Bias: Opinion mining algorithms can be biased based on the data used to train them. It's essential to ensure that the training data is diverse and representative of different viewpoints. Thus, opinion mining from social media is an essential tool for businesses, governments, and individuals to understand public opinion. It can help improve products and services, inform policy decisions, and monitor brand reputation. However, opinion mining from social media has its challenges, such as data quality, language, context, and bias. Addressing these challenges can help improve the accuracy of opinion mining and ensure that the insights gained are actionable.

The basic sentiment analysis techniques are shown in the following figure.

Fig.2 Opinion Mining Techniques.

II. OPINION MINING

Sentiment analysis can be a useful tool in identifying potentially radical content on social media platforms. Radical content can refer to messages that promote extremist or violent ideologies, which can pose a threat to public safety. By analyzing the sentiment of social media messages, we can identify potentially radical content and take appropriate action to mitigate any potential harm.

Here are some ways sentiment analysis can be used to identify potentially radical content:

Identifying negative sentiment: Sentiment analysis can identify negative sentiment in social media messages, such as anger, fear, or hate speech. These negative sentiments can be a sign of potentially radical content.

Identifying language patterns: Radical content often includes specific language patterns, such as extremist rhetoric or violent language. Sentiment analysis algorithms can be trained to recognize these language patterns and flag messages that contain them.

Monitoring sentiment over time: By monitoring sentiment over time, we can identify any sudden shifts in sentiment that may indicate the spread of radical content.

Identifying influencers: Social media influencers can have a significant impact on the spread of radical content. Sentiment analysis can be used to identify influencers who are promoting potentially radical messages and take appropriate action.

It's important to note that sentiment analysis should not be used in isolation to identify potentially radical content. Other techniques, such as network analysis, content analysis, and manual review, should also be used to ensure accuracy and prevent false positives. Additionally, any actions taken based on the results of sentiment analysis should be subject to legal and ethical considerations to protect free speech and prevent discrimination.

III. PREVIOUS WORK

The highlights of some of the key findings and trends of baseline techniques have been presented in this section.

Katipanov et al. proposed a deep learning-based approach to detect radical content on social media platforms. The study utilized a Naïve Bayes Network to identify extremist language patterns and achieve high accuracy in identifying radical content.

Sanchez et al. proposed an approach using graph convolutional networks to detect radical content on social media. The study showed that the approach achieved better results than traditional methods in identifying radical content.

Bohashev et al. proposed a novel approach to identify radical content on social media platforms using a multi-dimensional dynamic Bayesian network. The approach achieved high accuracy in detecting radical content by considering the context of the message and the sentiment of the user.

Li et al. in 2021 proposed a method to identify radical content on social media platforms using feature selection and machine learning algorithms. The study showed that the proposed method achieved high accuracy in identifying radical content while reducing false positives.

Tundis et al. proposed a model using word embeddings and a convolutional neural network to identify radical content on social media platforms. The study showed that the model achieved high accuracy in identifying radical content while reducing false positives.

Zevari et al. in 2019 proposed an approach using graph-based semi-supervised learning to identify radical content on social media. The study showed that the approach achieved high accuracy in identifying radical content by considering the network structure of social media platforms.

Johnston et al. proposed a hybrid approach to identify radical content on social media platforms. The approach combined machine learning algorithms with user behavior analysis to detect radical content.

Ihsitaki et al. proposed a method to identify radical content on social media using lexical features and a support vector machine. The study showed that the proposed method achieved high accuracy in identifying radical content.

The analysis shows that the combination of machine learning algorithms and user behavior analysis can achieve high accuracy in identifying radical content while reducing false positives.

IV. CHALLENGES WITH EXISTING SYSTEMS

Machine learning models have been used extensively for identifying radical content on social media platforms. These models can analyze text data to identify patterns and classify messages as potentially radical or non-radical. Here are some common machine learning models that have been used for this purpose:Support Vector Machines (SVMs): SVMs are a popular machine learning algorithm for text classification tasks. They work by finding the best boundary between classes of data points in a high-dimensional space. SVMs have been used to identify radical content on social media platforms by analyzing features such as word frequency, sentence structure, and sentiment.

Convolutional Neural Networks (CNNs): CNNs are deep learning models that have been used for image and text classification tasks. They work by extracting features from input data and learning patterns in these features to make predictions. CNNs have been used to identify radical content on social media platforms by analyzing features such as n-grams, word embeddings, and sentence structure.

Recurrent Neural Networks (RNNs): RNNs are another type of deep learning model that have been used for text classification tasks. They work by processing sequential data, such as text, and using feedback loops to incorporate context into predictions. RNNs have been used to identify radical content on social media platforms by analyzing features such as word embeddings, sentence structure, and sentiment.

Random Forests: Random Forests are an ensemble learning algorithm that combines multiple decision trees to make predictions. They have been used for text classification tasks by analyzing features such as word frequency, sentence structure, and sentiment. Random Forests can be particularly effective for identifying rare or unusual patterns in data.

Logistic Regression: Logistic Regression is a popular machine learning algorithm for binary classification

tasks. It works by modeling the probability of a data point belonging to a particular class based on input features. Logistic Regression has been used to identify radical content on social media platforms by analyzing features such as word frequency, sentence structure, and sentiment.

The major challenges in the identification of Radical Content are:

Typically the data size is extremely large and complex with overlapping data tags which makes classification extremely difficult. For example: Kill the President and Kill the Snake have completely different meanings.

Often the messages and tweets have redundant data and hidden emotions which are difficult to be converted to numerical counterparts since neural architectures work best on numerical data. Previous works generally did not consider couplets or group of words which can have a specific meaning. For example: Unite against law may NOT have any radical words like attack or bomb but is possibly radical.

It is generally extremely challenging to get high accuracy for such datasets.

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 Accuracy is computed as the performance metric for the system. It is defined as

$$Ac = \frac{TP + TN}{TP + TN + FP + FN}$$

Here,

TP: True positive

TN: True Negative

FN: False Positive

FN: False Negative

VI. CONCLUSION

It can be concluded that machine learning models can be highly effective for identifying radical content on social media platforms. However, it is important to note that these models are not perfect and can produce false positives or false negatives. Additionally, the use of machine learning models for identifying radical content raises important ethical considerations related to free speech, privacy, and bias. It is important to carefully evaluate and address these considerations when using machine learning models for this purpose. This paper presents a literature review of the recent machine learning techniques for the purpose.

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 2023

[2] D López-Sáncez, J Revuelta, F de la Prieta, "Towards the Automatic Identification and Monitoring of Radicalization Activities in Twitter," IEEE 2023

[3] G Bobashev, M Sageman, AL Evans, "Turning Narrative Descriptions of Individual Behavior into Network Visualization and Analysis: Example of Terrorist Group Dynamics, IEEE 2022

[4] Z Li, D Sun, B Li, Z Li, A Li, "Terrorist group behavior prediction by wavelet transform-based pattern recognition", hindawi 2020.

[5] A Tundis, G Bhatia, A Jain, "Supporting the identification and the assessment of suspicious users on Twitter social media", IEEE 2019.

[6] M Al-Zewairi, G Naymat, "Spotting the Islamist Radical within: Religious Extremists Profiling in the United State", Elsevier 2018

[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

[21] B Akhgar, F Tabatabayi, PS Bayerl, "Investigating Radicalized Individual Profiles through Fuzzy Cognitive Maps", Elsevier 2014