WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Social Media Analysis Data Using Data Mining

[1]Mrs. Pratibha Mishra, [2]Divya Dabdi, [3]Uma BV, [4]T Omsree, [5]Vaishnavi HK

[1]Asst. Professor, Department of CSE, Ballari Institute of Technology & Management, Ballari. prathibhamishra@bitm.edu.in
[2,3,4,5] Department of CSE, Ballari Institute of Technology & Management, Ballari
[2]divyadabdiprakash2002@gmail.com, [3]bvuma842597@gmail.com, [4]omsreesurendrababu@gmail.com, [5]vaishnavihk48@gmail.com

ABSTRACT:

Social media networks widespread use has completely changed how content and information are shared across various communities across the globe. While this connectivity has brought numerous benefits, it has also exposed significant challenges, notably the widespread dissemination of offensive and trolling content. Effective online moderation is necessary to address these issues and maintain a healthy social media environments. The project "Analyzing Social Media Data Through Data Mining" shows an innovative approach to tackle this issue. The project focuses on extracting text from memes to identify offensive and trolling behaviour specifically in English-language content. The method employs a sophisticated weighted ensemble of machine learning classifiers, aiming to enhance accuracy in detecting harmful content. This project not only presents a detailed methodology but also highlights its findings and implications for social media monitoring and content moderation. By leveraging advanced techniques in data mining and machine learning, it aims to contribute significantly to creating safer and more constructive online spaces.

Keywords: Social media platforms, Diverse communities, Offensive and trolling content, Extracting text from memes, Identifying offensive and trolling behaviour, English-language content, Improved accuracy, harmful content detection, Social media monitoring, Data mining methods Safer online spaces.

INTRODUCTION:

The social media networks have completely changed the landscape of global information dissemination, fostering unprecedented connectivity among diverse communities. This transformative power, however, comes with inherent challenges, prominently marked by the rampant spread of offensive and trolling content. Specifically focusing on extracting textual data from memes, the project aims to pinpoint and address offensive and trolling behaviour within English-language content. This project not only delves into intricate methodologies but also sheds light on consequential findings and implications crucial for advancing social media monitoring and content moderation practices.

LITERATURE SURVEY:

In this paper [1], "Social Network Data Analytics" by Liu, J., & Tang, J. (2021) explores various techniques and approaches for social network data analysis, focusing on data mining methods. It discusses the challenges in handling diverse social media data and highlights the significance of extracting valuable insights for applications such as recommendation systems and user behaviour analysis. In essence, the review explores the methods and challenges of analyzing social media data, highlighting the significance of this analysis for applications like recommendation systems understanding user behaviour.

In the paper [2], "A Survey of Data Mining methods in Social Media" by Almasri, N. A., et al. (2020), the paper provides a comprehensive survey of methods applied to social media data. It categorizes and reviews methods for sentiment analysis, topic modelling, and user profiling. The article also discusses emerging trends and challenges in this field. This paper discusses different data mining methods applied to study social media information. It categorizes and assesses methods for sentiment analysis, topic modelling, and user profiling while also highlighting emerging trends and challenges within this domain.

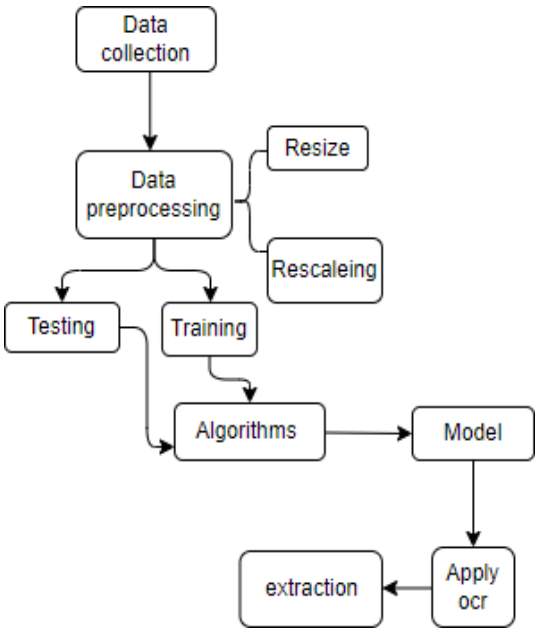
In the paper [3], "A Review of Social Media Analytics: Challenges and Opportunities" by Meng et al. (2019) provides an in-depth examination of the landscape of analytics for social media networks. It delves into key areas such as sentiment analysis, event detection, and user behaviour modelling. Emphasizing the significance of real-time data mining techniques, the paper highlights the necessity of integrating analytics for social networks with business intelligence for enhanced decision-making capabilities. Overall, Meng and colleagues' review highlights the difficulties and possibilities involved in using social media information for actionable insights.

The paper [4], "OffensEval" by Kumar, Aggarwal, and Singh (2018) outlines a competition called Offensively shared task aimed at identifying offensive language in social media. Offensive language encompasses hate speech, cyberbullying, and harassment, making automatic detection crucial for fostering a safe online environment. The authors create the Offense Eval dataset by collecting labelled tweets, including subtask-specific labels for various kinds of offensive language, and provide annotation guidelines for consistency.

In a paper [5], Waseem and Hovy's (2016) paper addresses hate speech detection on Twitter via machine learning. They define hate speech as attacks based on attributes like race or religion. The study collects labelled tweets and extracts features and sentiment scores. Through experimentation, they evaluate different feature combinations and classifiers using precision, recall, and F1 score metrics. The goal is to identify hate speech on social media platforms, encompassing various forms such as direct attacks and coded language.

PROPOSED METHODOLOGY:

1. **Data Collection and Pre-processing:** The first step involves setting up accounts and access permissions to enable secure data collection. Subsequently, the collected data undergoes pre-processing, where cleaning and standardization procedures are applied, particularly focusing on extracting text from memes to ensure data uniformity and accuracy.
2. **Training Machine Learning Models:** Select the right algorithms for sentiment analysis, particularly for identifying whether English language content is neutral, positive or negative. Train the chosen algorithms using pre-processed data. Utilize a DataStore to securely manage and store user data for training purposes.
3. **Prediction and Analysis :** Trained models are implemented to predict offensive and trolling behaviour specifically in English-language content. Simultaneously, the dataset is analysed to comprehend the offensive memes across various platforms. The accuracy and effectiveness of the prediction models are evaluated to assess their performance in identifying harmful content accurately.
4. **Generate Result:** The methodology detects negative sentiment related to offensive and trolling content. It's significance for social media moderation, emphasizing the possibility of advanced techniques to foster safer online interactions.



SIGNIFICANCE OF "SOCIAL MEDIA DATA MINING":

The significance of this study lies in addressing the challenges posed by offensive and trolling content on these platforms. By leveraging advanced techniques, such as extracting text from memes and employing machine learning classifiers, this approach aims to accurately identify harmful content, thereby contributing to the creation of safer and healthier online environments. Through its innovative methodology, this project underscores the vital role of data mining in social media monitoring and content moderation, ultimately striving to foster more constructive online interactions.

MOTIVATION FOR PROJECT:

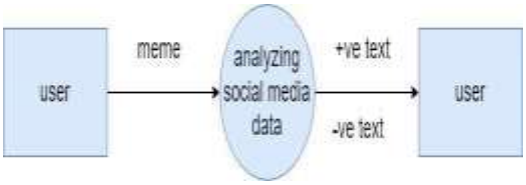
This is motivated by the prevalence of offensive and trolling content on social media platform, highlighting the need for effective online moderation. It analyses social media information and aims to enhance the precision of detecting harmful content, ultimately contributing to provide safer online environments and raise positive interactions within diverse communities.

OBJECTIVES :

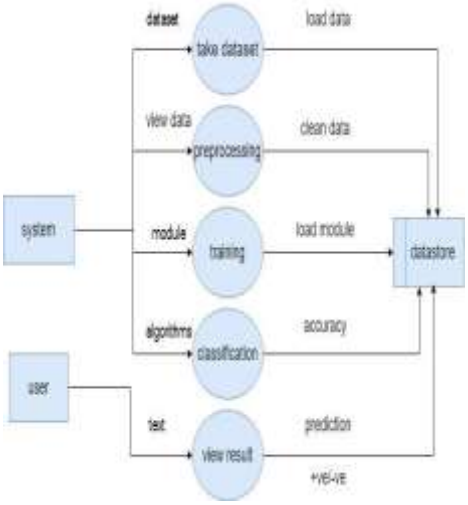
1. To develop an innovative approach for examining social media data using data mining methods.
2. To develop a methodology for accurately identifying harmful content, particularly in English-language memes.
3. To filter the harmful contents and enhance the precision of detecting offensive behaviour.

DATA FLOW DIAGRAM:

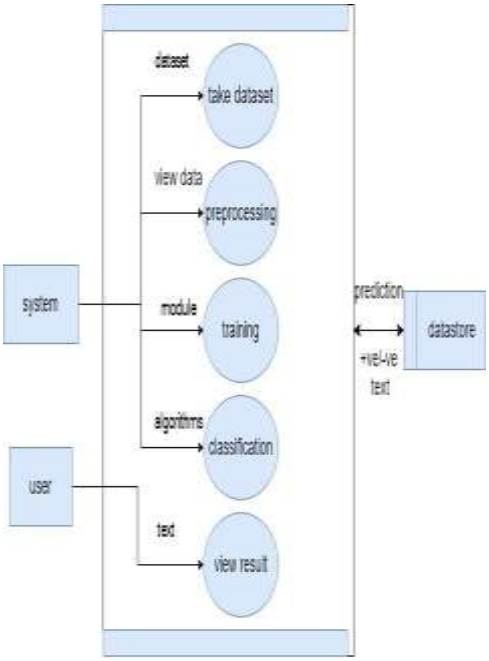
LEVEL 0 DIAGRAM:



LEVEL 1 DIAGRAM:



LEVEL 2 DIAGRAM:



DISCUSSION:

The results demonstrate the effectiveness of a data mining approach to address offensive content on social networks, focusing on text extraction from memes. It aims to enhance accuracy with machine learning classifiers and emphasizes the significance of effective online moderation for safer social media environments.

CONCLUSION:

This highlights development of a user friendly application called a examining social media information using methods such as Decision Tree, Random Forest, Naive Bias, NLP to address offensive content on social media which aims to enhance detecting accuracy and gives the estimation of the text as Positive, Negative or Neutral.

REFERECES:

[1] Levy, M. (2016). Playing with Twitter Data. [Blog] R-bloggers. Available at: https://www.rbloggers.com/playing-with-twitter-data/ [Accessed 7 Feb. 2018].

[2] Popularity Analysis for Saudi Telecom Companies Based on Twitter Data. (2013). Research Journal of Applied Sciences, Engineering and Technology. [online] Available at: http://maxwellsci.com/print/rjaset/v6-4676-4680.pdf [Accessed 1 Feb. 2018].

[3] Zhao, Y. (2016). Twitter Data Analysis with R –Text Mining and Social Network Analysis. [online] University of Canberra, p.40. Available at:https://paulvanderlaken.files.wordpress.com/201 7/08/rdataminingslides-twitteranalysis.pdf [Accessed 7 Feb. 2018].

[4] Alrubaiee, H., Qiu, R., Alomar, K. and Li, D. (2016). Sentiment Analysis of Arabic Tweets in eLearning. Journal of Computer Science. [online] Available at: http://thescipub.com/PDF/jcssp.2016.553.563.pdf [Accessed 7 Feb. 2018].

[5] Qamar, A., Alsuhibany, S. and Ahmed, S. (2017). Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies. (IJACSA) International Journal of Advanced Computer Science and Applications, [online] 8. Available https://thesai.org/Downloads/Volume8No1/Paper_50Sentiment_Classification_of_Twitter_Data_Belonging.pdf [Accessed 1 Feb. 2018].

[6] R. M. Duwairi and I.Qarqaz, "A framework for Arabic sentiment analysis using supervised classification" , Int. J. Data Mining, Modelling and Management, Vol. 8, No. 4, pp.369-381 , 2016.

[7] Hossam S. Ibrahim, Sherif M. Abdou, Mervat Gheith, "Sentiment Analysis For Modern Standard Arabic And Colloquial", International Journal on Natural Language Computing (IJNLC), Vol. 4, No.2, pp. 95-109, April 2015.

[8] Assiri, A., Emam, A. and Al-Dossari, H. (2016). Saudi Twitter Corpus for Sentiment Analysis. International Journal of Computer and Information

Engineering, [online] 10. Available at: http://waset.org/publications/10003483/sauditwitter-corpus-forsentiment-analysis[Accessed 1 Mar. 2018].

[9] Khadar, A. A., Hayath, T. M., & Nagraj, M. S. (2017, August). Design and implementation of ZigBee based smart grid system for power management. In 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon) (pp. 1339-1343). IEEE.

[10] L. Wasser and C. Farmer, "Sentiment Analysis of Colorado Flood Tweets in R", Earth Lab, 2018. [Online]. Available:

https://earthdatascience.org/courses/earthanalytics/get-data-usingapis/sentiment-analysisoftwitter-data-r/. [Accessed: 01- Mar- 2018].

[11] D. Robinson, "Text analysis of Trump's tweets confirms he writes only the (angrier) Android half", Variance explained, 2016.