IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 12, May 2023

Cyber Bullying and Hate Speech Detection

Prof. Ravindra Chilbule[1], Kasifraza Siddique[2], Sangharsh Moon[3], Nirmal Zade[4], Aditya Fusate[5]
Professor, Computer Science Engineering Department1
Students, Computer Science Engineering Department[2,3,4,5]
Rajiv Gandhi College of Engineering, Research and Technology, Chandrapur, Maharashtra, India

Abstract: *Online platforms frequently have problems with hate speech, which harms people, discriminates against people, and polarises society. The fast expansion of social media networks and online groups has increased the spread of hate speech, necessitating the creation of reliable detection systems. With the capacity of computational algorithms to automatically identify and report instances of hate speech, machine learning approaches have emerged as possible solutions to this issue.*

The identification of hate speech using machine learning techniques is thoroughly reviewed in this work. The goal is to give a broad overview of the many methods used, the difficulties faced, and the developments in this area. The review discusses the advantages and disadvantages of modern deep learning models as well as conventional machine learning techniques.

The significance of hate speech identification and its effects on online groups and society at large are covered in the first section of the study. It then gives a general review of the various varieties of hate speech as well as the difficulties involved in classifying and detecting it. It then explores the various tools and information sources frequently used for detecting hate speech, such as text-based tools, user profiles, and contextual data.

The paper examines a variety of machine learning methods, including supervised, unsupervised, and semi-supervised learning, that are used in the identification of hate speech. It addresses how to effectively capture patterns of hate speech using feature engineering techniques like n-grams, word embeddings, and topic modelling. Additionally, it explores how ensemble techniques and transfer learning might enhance detection performance.

In addition, the research discusses difficulties in detecting hate speech, including class disparity, context sensitivity, and changing linguistic trends. It covers methods for overcoming these difficulties, including as sampling approaches, data augmentation, and model adaption.

Keywords: Hate speech, Natural Language processing , Social network ,Text mining

REFERENCES

[1]Hern, A., Facebook, YouTube, Twitter, and Microsoft sign the EU hate speech code. The Guardian, 2016. 31.
[2] Rosa, J., and Y. Bonilla, Deprovincializing Trump, decolonizing diversity, and unsettling anthropology. American Ethnologist, 2017. 44(2): p. 201-208.
[3] Travis, A., Anti-Muslim hate crime surges after Manchester and London Bridge attacks. The Guardian, 2017.
[4] MacAvaney, S., et al., Hate speech detection: Challenges and solutions. PloS one, 2019. 14(8): p. e0221152.
[5] Fortuna, P. and S. Nunes, A survey on automatic detection of hate speech in text. ACM Computing Surveys (CSUR), 2018. 51(4): p. 85.
[6] Mujtaba, G., et al., Prediction of cause of death from forensic autopsy reports using text classification techniques: A comparative study. Journal of forensic and legal medicine, 2018. 57: p. 41-50.
[7] Cavnar, W.B. and J.M. Trenkle. N-gram-based text categorization. in Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval. 1994. Citeseer.
[8] Ramos, J. Using tf-idf to determine word relevance in document queries. in Proceedings of the first instructional conference on machine learning. 2003. Piscataway, NJ.
[9] Mikolov, T., et al. Distributed representations of words and phrases and their compositionality. in Advances in neural information processing systems. 2013.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 12, May 2023

[10] Le, Q. and T. Mikolov. Distributed representations of sentences and documents. in International conference on machine learning. 2014.
[11] Kotsiantis, S.B., I.D. Zaharakis, and P.E. Pintelas, Machine learning: a review of classification and combining techniques. Artificial Intelligence Review, 2006. 26(3): p. 159-190.
[12] Lewis, D.D. Naive (Bayes) at forty: The independence assumption in information retrieval. in European conference on machine learning. 1998. Springer.
[13] Xu, B., et al., An Improved Random Forest Classifier for Text Categorization. JCP, 2012. 7(12): p. 2913-2920.
[14] Joachims, T. Text categorization with support vector machines Learning with many relevant features. in European conference on machine learning. 1998. Springer.
[15] Zhang, M.-L. and Z.-H. Zhou, A k-nearest neighbor based algorithm for multi-label classification. GrC, 2005. 5: p. 718-721.