Abstract

Sentiment analysis is a natural language processing tool that is useful for monitoring and accessing information from Web applications, as it is a treasure of public opinion about numerous issues without requiring to find and assess authentication of the document. With the rapid growth of social networks and microblogging websites, communication between people from different countries, cultural and psychological and physical backgrounds became more direct, resulting in more and more conflicts about thoughts and speech used between these people. The meaning of Hate speech can be explained and put into context as the use of aggressive, violent or offensive language, targeting a specific group of people sharing a common property, whether this property is their gender (i.e., sexism), their ethnic group or race (i.e., racism) or beliefs and religion, political parties they resemble. As in India a lot of hate speech generated posts are placed on social networking sites. So to block and catch the hate speech generated posts to avoid necessary conflicts. So we thought of introducing a technique using Twitter as our data source. As twitter is generally and widely used by millions of Indians it is great source of raw data that can be monitored and made sharable without conflict. The idea
came up of creating and analyzing datasets that can be used for machine Learning. By using political leaders as the study topics e.g. Narendra Modi, Rahul Gandhi etc. It is used to analyze the negative and positive hate speech tweets generated on them. Machine Learning algorithms are used to analyze the tweets and find the correct meaning behind it whether it is offensive or not. In this project we are various techniques such as Stop Words, Lexicon Analysis, Datasets and Machine Learning are used to analyze tweets and find out the sentiments behind it. Apache Spark based parallel processing technique is also used to access only the latest tweets and not the old ones which are already being analyzed.

References

3. Hajime Watanabe, Mondher Bouazizi and Tomoaki Ohtsuki, Hate Speech on Twitter, IEEE, 2018.
12. J. Zhao, L. Dong, J. Wu, and K. Xu, Moodlens: An emoticon-based sentiment analysis system for Chinese is used.
13. A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, Learning word vectors for sentiment analysis, in Proc. ACL, 2011.

Index Terms

Computer Science
Information Sciences
Keywords

Twitter, Sentiment Analysis, Stop words, Lexicon Analysis, Machine Learning, Apache Spark.