

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

TEDIA

International Journal of Innovative Research in SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 7, July 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462

6381 907 438

 \bigcirc

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)



e-ISSN: 2319-8753, p-ISSN: 2320-6710 <u>www.ijirset.com</u> | Impact Factor: 8.118

Volume 11, Issue 7, July 2022

| DOI:10.15680/LJIRSET.2022.1107104 |

Twitter Data Preprocessing in Sentimental Analysis using NLP

Ramya J Katti¹, Lakshmi B², Harshavardhan Gowda ³, Madhusudhan R⁴

U.G. Student, Department of Information Science and Engineering BIET Engineering College, Davangere,

Karnataka, India¹

U.G. Student, Department of Information Science and Engineering BIET Engineering College, Davangere,

Karnataka, India²

U.G. Student, Department of Information Science and Engineering BIET Engineering College, Davangere,

Karnataka, India³

U.G. Student, Department of Information Science and Engineering BIET Engineering College, Davangere,

Karnataka, India⁴

ABSTRACT: In today's world, Social Networking website like Twitter, Facebook, Tumbler, etc. plays a very significant role. Twitter is a micro-blogging platform which provides a tremendous amount of data which can be used for various applications of Sentiment Analysis like predictions, reviews, elections, marketing, etc. Sentiment Analysis is a process of extracting information from large amount of data, and classifies them into different classes called as sentiments.Python is simple yet powerful, interpreted, high-level and dynamic programming language, which is well known for its functionality of processing natural language data by using NLTK (Natural Language Toolkit). NLTK is a library of python, which provides the base for building programs and classification of the data. NLTK also provide graphical demonstration for representing various results or trends and it also provide sample data to train and test various classifiers respectively.

KEYWORDS: Text detection, Inpainting, Morphological operations, Connected component labelling.

I. INTRODUCTION

Twitter is an American online news and social networking service on which users post and interact with messages known as "tweets". Tweets were originally restricted to 140 characters, but on November 7, 2017, this limit was doubled to 280 for all languages except Chinese, Japanese, and Korean. Registered users can post, like, and retweet tweets, but unregistered users can only read them. Users access Twitter through its website interface, through Short Message Service (SMS) or its mobile-device application software ("app"). Twitter, Inc. is based in San Francisco, California, and has more than 25 offices around the world.

Twitter was created in March 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams and launched in July of that year. The service rapidly gained worldwide popularity. In 2012, more than 100 million users posted 340 million tweets a day, and the service handled an average of 1.6 billion search queries per day. In 2013, it was one of the ten most-visited websites and has been described as "the SMS of the Internet". As of 2018, Twitter had more than 321 million monthly active users. Since 2015 Twitter has been a hotbed of debates and news covering politics of the United States. During the 2016 U.S. presidential election, Twitter was the largest source of breaking news on the day, with 40 million election-related tweets sent by 10:00 p.m. (Eastern Time) that day.

II. RELATED WORK

Twitter Sentiment Analysis System International Journal of Computer Applications (0975 - 8887) Volume 180 – No.47, June 2018. Social media is increasingly used by humans to express their feelings and opinions in the form of short text messages. Detecting sentiments in text has a wide range of applications including identifying

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)



e-ISSN: 2319-8753, p-ISSN: 2320-6710 <u>www.ijirset.com</u> | Impact Factor: 8.118

Volume 11, Issue 7, July 2022

| DOI:10.15680/IJIRSET.2022.1107104 |

anxiety or depression of individuals and measuring well-being or mood of a community. Sentiments can be expressed in many ways that can be seen such as facial expression and gestures, speech and by written text. Sentiment Analysis in text documents is essentially a content – based classification problem involving concepts from the domains of Natural Language Processing as well as Machine Learning. In this paper, sentiment recognition based on textual data and the techniques used in sentiment analysis are discussed.

Approaches, Tools and Applications for Sentiment Analysis Implementation. International Journal of Computer Applications (0975 - 8887) Volume 125 - No.3, September 2015. The paper gives an overview of the different sentiment classification approaches and tools used for sentiment analysis. Starting from this overview the paper provides a classification of (i) approaches with respect to features/techniques and advantages/limitations and (ii) tools with respect to the different techniques used for sentiment analysis. Different application fields of application of sentiment analysis such as: business, politic, public actions and finance are also discussed in the paper.

Sentimental Analysis with NLP on Twitter Data. Conference Paper · July 2019. Every social networking sites like facebook, twitter, instagram etc become one of the key sources of information. It is found that by extracting and analyzing data from social networking sites, a business entity can be benefited in their product marketing. Twitter is one of the most popular sites where people used to express their feelings and reviews for a particular product. In our work, we use twitter data to analyze public views towards a product. Firstly, we have developed a natural language processing (NLP) based pre-processed data framework to filter tweets. Secondly, we incorporate Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) model concept to analyze sentiment. This is an initiative to use BoW and TFIDF are used together to precisely classify positive and negative tweets. We have found that by exploiting TF-IDF vectorizer, the accuracy of sentiment analysis can be substantially improved and simulation results show the efficiency of our proposed system. We achieved 85.25% accuracy in sentiment analysis using NLP technique.

III. METHODOLOGY

At this stage, each sentence of the review and opinion is examined for subjectivity. Sentences with subjective expressions are retained and that which conveys objective expressions are discarded. Sentiment analysis is done at different levels using common computational techniques like Unigrams, lemmas, negation and so on.

Sentiments can be broadly classified into two groups, positive and negative. At this stage of sentiment analysis methodology, each subjective sentence detected is classified into groups-positive, negative, good, bad, like, dislike.



Fig 1:Sign-Up Page

IV. EXPERIMENTAL RESULTS





International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)



e-ISSN: 2319-8753, p-ISSN: 2320-6710 <u>www.ijirset.com</u> | Impact Factor: 8.118

Volume 11, Issue 7, July 2022

| DOI:10.15680/IJIRSET.2022.1107104 |





Fig 5: Graph Analysis

Fig 6:Pie Chart

V. CONCLUSION

The task of sentiment analysis, especially in the domain of micro-blogging, is still in the developing stage and far from complete. So, we propose a couple of ideas which we feel are worth exploring in the future and may result in further improved performance. The closer the negation word is to the unigram word whose prior polarity is to be calculated, the more it should affect the polarity.

REFERENCES

- [1] A.Pak and P. Paroubek. "Twitter as a Corpus for Sentiment Analysis and OpinionMining". InProceedingsoftheSeventhConferenceon InternationalLanguageResourcesandEvaluation, 2010, pp.1320-1326
- R. Parikh and M. Movassate, "Sentiment Analysis of User- Generated TwitterUpdatesusingVariousClassi_cation Techniques", CS224N Final Report,2009
- [3] Go, R. Bhayani, L. Huang. "Twitter Sentiment Classification Using DistantSupervision". Stanford University, TechnicalPaper, 2009
- [4] L.Barbosa, J.Feng. "RobustSentimentDetectiononTwitter fromBiasedandNoisyData". COLING2010: Poster Volume, pp. 36-44.
- [5] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M.."Lexicon based methods forsentiment analysis".Computational linguistics, 2011:37(2),267-307.
- [6] V. M. K. Peddinti and P. Chintalapoodi, "Domain adaptation in sentiment analysis oftwitter,"in AnalyzingMicrotext Workshop, AAAI, 2011





Impact Factor: 8.118





INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com