

Twitter Sentiment Analysis: Comparison of Techniques

¹G.Adhithya Balaji, ²Rashmita Khilar

^{1,2}Department of Information Technology, Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences, Chennai – 602105

¹adhityag98@gmail.com, ²rashmitakhilar.sse@saveetha.com

Article Info

Volume 83

Page Number: 3223-3225

Publication Issue:

May - June 2020

Article History

Article Received: 19 August 2019

Revised: 27 November 2019

Accepted: 29 January 2020

Publication: 12 May 2020

Abstract

Analysis of twitter sentiment is a technology that offers the tools for surveying items or incidents relevant to popular emotion. Such functions are conveyed by terms of sentiment, emojis, etc. Twitter messages are brief, and often ideal for our comprehension. These classifications promptly choose whether the polarity of feeling in a tweet is negative or positive. In this paper we will compute and compare about the precision and F-score of SVM and Naïve Bayes classifiers.

Keywords: *comprehension, sentiment, emojis, SVM and Naïve Bayes classifiers.*

1. Introduction

Social media has provided a forum for users to express and share their views and perspectives on all kinds of topics and events. Twitter has nearly 700million users and more than 350 million tweets a day. And furthermore, observing their activities over brands, pieces of the overall industry and contenders turns into a gold mine for the associations. The grouping of twitter feelings likewise targets choosing a positive or negative and neutral inclination towards tweets. Twitter feeling has many specific procedures because of the nature and decent variety of the human language. This paper's contributions are primarily to calculate the accuracy and the F-score. Twitter users inventive result is likewise made on an achievement, which has additionally become a benchmark for the entirety of their alternatives which products. Furthermore, the strategies utilized in this paper additionally exhibited the precision of the two information sources and yields utilizing the SVM and Naïve Bayes classifiers. The paper's organization is as follows. Firstly, it deals with research related to this subject. Secondly, the proposed framework is established and finally it concludes with the comparison and results of both the classifiers used in the system.

2. Related Work

The researchers at [1] examined the presentation of different classifiers when applied to binary

classification issues. They have utilized diverse approval strategies, and the creators find that the SVM model doesn't perform to the degree of the Naïve Bayes when there are barely any positive tests. They are both serious in light of the fact that there is both positive and negative. In [2] the creators demonstrated that the rate precision got for an enormous informational collection by that of the size of the preparation information, the expansion in size isn't relative to that. The relationship was found not straight, and is to a greater extent a logarithmic relationship. Likewise the creators contended that there could be no purpose behind the expanded computational effort in specific circumstances. In [3] twitter sentiment investigation was utilized to decide whether they could receive answers for questions, for example, "Is it conceivable to estimate stock costs of a business from the general assessment? Is it genuine that specific organizations' stock costs will be simpler to anticipate contrasted with others? " Moreover, Naïve Bayes was used in anticipating headings of clinical subjects which cost an inordinate measure of time. In [4]The creators had the option to diminish the preparation set to a satisfactory order size that is reliable with our discoveries in the field of sentiment investigation where the Naïve Bayes seemed to meet near their best exactness by utilizing just 20% of the preparation set. Work was done in [5] by reporting the exactnesses of Naïve Bayes, SVM and KNN, while changing the extents of the preparation set [5]. It was indicated that the exhibition of the Naïve Bayes

calculation was higher when contrasted and SVM on all the various sizes of preparing sets. The discoveries stand out from what we've discovered when utilizing Naïve Bayes and SVM for sentiment investigation on our twitter information collection. Contrasted with the Naïve Bayes, the SVM classifier beat or gave a comparable accuracy on a portion of the informational collections we utilized in our examinations as we will show later in the result and analysis segment.

3. Proposed Framework

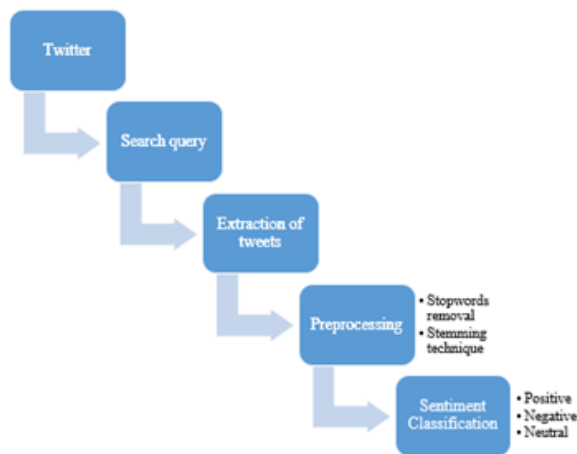


Figure - 1 Proposed Block Diagram

Search Query

In this module, client can look the application for the essential question. The search question at that point conveys the cycle and shows the relating question results.

Extraction of Tweets

At the point when the inquiry is perused, the live tweets are recovered from the constant utilizing the Twitter API. The parameters are set to recover at least 100 pursuit catchphrase based tweets.

Pre processing

When all tweets are taken from twitter, they will be handled and the noise data will be expelled.

Sentiment Classification

All catchphrase tweet words are sorted into positive, negative and neutral tweets. For characterization purposes the collaborative filtering algorithm is used. A positive and negative word dataset is available here. We may arrange the tweets into positive, negative and neutral tweets.

4. Result and Analysis

The following table gives the comparative results of both SVM and Naïve Bayes classifiers with their accuracy percentage and F-score with varied size of training sets.

Table 1: Statistical values of twitter analysis

Training Set	Accuracy		F-Score	
	SVM	Naïve Bayes	SVM	Naïve Bayes
20%	72.16%	72.16%	0.7023	0.6502
40%	74.19%	74.32%	0.7265	0.6967
60%	74.70%	74.58%	0.7302	0.7018
80%	75.09%	75.16%	0.7216	0.6965
100%	76.37%	74.38%	0.717	0.6862

5. Conclusion

This model for regular language examination and twitter sentiment analysis is one for the most essential and effective portrayal. With the proposed model we had the option to show a powerful correlation of two classifiers utilized in the framework. Later on, we should focus on testing about how the quantity of perspectives or angles in an expression or tweet impact the general sentiment of an expression or tweet and if there is a relationship that may assist with improving the precision of the forecast of estimation.

References

- [1] George Forman, and Ira Cohen "Learning from Little: Comparison of Classifiers Given Little Training," 15th European Conference on Machine Learning and the 8th European Conference on Principles and practice of Knowledge Discovery in Databases, 20-24 September 2004, Pisa, Italy.
- [2] Effect of Sampling Size on Data Mining Using Artificial Neural Networks. by Ashraf H. Abdelwahab, Hokey Min, Ahmed Z.Emam, Adel S Elmaghaby. in INTELLIGENT ENGINEERING SYSTEMS THROUGH ARTIFICIAL NEURAL NETWORKS;10; 449-454; Artificial neural networks in engineering conference; Smart engineering system design by ASME; 2000.
- [3] Li Bing, Keith C. C. Chan, and Carol Ou, "Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements," ICEBE '14 Proceedings of the 2014 IEEE 11th International Conference on e-Business Engineering. SemEval'2013: SemEval-2013 Task 2: Sentiment Analysis in Twitter.
- [4] Sunghwan Sohn, Won Kim, Donald C. Comeau, and W. John Wilbur, "Optimal

- Training Sets for Bayesian Prediction of MeSH® Assignment,” J Am Med Inform Assoc. 2008 Jul-Aug; 15(4): 546-553.
- [5] Yun-Nung Chen, Che-An Lu, and Chao-Yu Huang, “Anti-Spam Filter Based on Naïve Bayes, SVM, and KNN model,” 2009 AI TERMPROJECT
 - [6] Twitter sentiment analysis using Deep convolutional neural networks: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval[<https://dl.acm.org/doi/abs/10.1145/2766462.2767830>]
 - [7] Semantic sentiment analysis on Twitter.[<https://link.springer.com/conference/semweb>]
 - [8] Yu, B.: An evaluation of text classification methods for literary study. Literary and Linguistic Computing 23(3), 327– 343 (2008)
 - [9] Sentiment analysis twitters datamining 2015 International conference [<https://ieeexplore.ieee.org/xpl/conhome/7485753/proceeding>]
 - [10] A survey on sentiment analysis challenges [DoaaMoheyEI- DinMohammedHusseinin year2018 Lexicon-based method Deep sentiment analysis method]
 - [11] Comparison Research on text pre- processing Methods on Twitter Sentiment Analysis in the year 2017 [Zhao JianqiangGuiXiaolin]
 - [12] Tweet Sentiment Analysis with classifier Ensembles [Nadia F.F.da.Silva, Eduardo R.Hruschka, Estevam R. Hruschka Jr.(2014)]
 - [13] Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., Li, P.: User- level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD (2011)