iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.47900

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

Study on Various Sentiment Analysis Techniques

Kartik Chamoli[1], Nishant Vashist[2], Harsh Sharma[3], Dr. Manjot Kaur Bhatia[4]
[1, 2, 3, 4]*Jagan Institute of Management Studies, Rohini Sector-5*

I. INTRODUCTION

Today the scale with which social media produces data is huge. The huge volume of data generated by single social media user is massive. People nowadays share their ideas, learn from each other and share their opinions online. People use websites like Facebook, twitter, google+ etc to share their opinions with like minded people and to form a community. The data generated can used by data scientists and machine learning engineers to analyse the opinions of the people whether they are positive, negative or neutral. All of the likes, comments, popular tags on various social media websites can be used to know the general sentiment of a public towards a particular topic to gain important insights about human behaviour and thinking.

Knowing the people sentiments is know one of the most important factors for every organization to target them better and to know how they can influence the user's decision. Sentiment of a user basically consists of opinions, sentiments, appraisals, attitudes, and emotions. Analysing all these sentiments is a challenging task in itself. The major part of sentiment analysis is the retrieval of sentiments from the textual information by processing, searching and analysing the vast amount of data present. The social media algorithm works on sentiment analysis 24*7 to only display those things to the user which will keep them longer on their website.

We will focus majorly on opinion mining techniques using machine learning and lexicon-based approaches along with the evaluation metrics. We will get an overview on the already available techniques like Naïve Bayes, Max entropy and support vector machine and analyse the challenges with them on huge twitter dataset.

II. LITERATURE STUDY

Paper Title	Problem discussed/Technique	Framework setup	Result	Drawback
Vishal A. Kharde, S.S. Sonawane Sentiment Analysis of Twitter Data: A Survey of Techniques	Author research shows that how different sentiment analysis techniques give different accuracy and precision on huge twitter dataset.	Sentiment analysis methods like SVM, EWGA, CLMM , active learning and SFA are applied to various twitter datasets.	A set of synonyms are offered after semantic analysis that have similarity and show the polarization of the social media content.	The research lacks visual representation of the different datasets used for sentiment analysis.
Abdullah Alsaeedi , Mohammad Zubair Khan A Study on Sentiment Analysis Techniques of Twitter Data	Discussion and review of various sentiment analysis along with the mathematical techniques used to represent them.	Ensemble approaches and supervised machine learning	Biasness of tweets when analysed using supervised and unsupervised machine learning.	The research needs a high level of mathematical background and could not be understood by a layman.
Faizan, Twitter Sentiment Analysis	The paper focused majorly on the 5 stages of sentiment analysis.	Unigram, Bigram, N-gram, POS tagging, Subjective, objective features	Built a model for the analysis of feeling using KNN algorithm with unigram, bigram and ngram features.	the study needto improve on these limitations and applying the scope method and should give more weightage on mental and social health problems.

©IJRASET: All Rights are Reserved

Omar Y. Adwan(✉), Marwan Al-Tawil, Ammar M. Huneiti, Razan H. Al-Dibsi, Rawan A. Shahin, Abeer A. Abu Zayed Twitter Sentiment Analysis Approaches: A Survey	The paper gave an overview on the sentiment classification phases and their subcategories. The phases are pre-processing, feature-extraction, feature selection(filtering), classification.	Sentiment Features (SENF), Syntax Based Features(SYNF), Semantic Features (SEMF), Unigram Based Features(UGF),N-gram Features (NGF), Top words features (TWF).	The overview of twitter sentiment analysis was given by analysing over 40 articles and mentioning the different approaches used in each one of them.	The paper doesn't give us a deep insight on the different approaches and how they can be applied in the real world on practical huge datasets.

III. PROCESSING THE DATASET

A tweet posted by a user contains a variety of opinions about users views expressed in different ways. The dataset used for sentiment analysis is usually labelled into two classes i.e. positive and negative polarity to represent the two extreme of the human emotions. The positive and negative polarity makes it easy to perform sentiment analysis. The raw data although contain a lot of redundancy and unwanted data that can cause problems to our algorithm while calculating the sentiment. We need to do some sort of pre-processing to make out data finer before passing it to an algorithm for processing.

Pre-processing of tweet include following points.

1) Remove all URLs (e.g. www.yzy.com), hash tags (e.g. #topic), targets (@username)
2) Try to correct the spellings in the tweet especially if there are any repeated characters.
3) Try to replace all the emojis and emoticons with their sentiment.
4) Remove all the punctuations, symbols and numbers from the tweet as they have no significance
5) Remove Stop Words
6) Try expanding the acronyms used in the tweets by using an acronym dictionary as the machine learning algorithm may not understand acronyms.
7) Remove non English tweets as they can change the correctness and preciseness of the algorithms and can act as bad input.

IV. FEATURE EXTRACTION

The pre-processed dataset has numerous distinctive properties. In the point birth system, we prize the aspects from the reused dataset. These aspects are used to decipher the positive and negative opposition in a judgment which is useful for determining the opinion of the individualities using models like unigram, bigram(18).

Machine learning is used for representing the major features of texts and documents. These are known as feature vectors and these features .Some exemplifications features that have been reported in literature are

A. Words And Their frequentness

Unigrams, bigrams and n- gram models with their frequency counts are considered as features. There has been further exploration on using word presence rather than frequentness to better describe this point. Pangetal showed better results by using presence rather of frequentness.

B. Speech Tags

Speech like adjectives, adverbs and some groups of verbs and nouns are good pointers of subjectivity and sentiment. We can induce syntactic reliance patterns by parsing or reliance trees.

C. Opinion Words And Expressions

Piecemeal from specific words, some expressions and expressions which convey sentiments can be used as features.
Eg: cost someone an arm and leg.

D. Position Of Terms

The position of a term with in a textbook can affect on how important the term makes difference in overall sentiment of the textbook.

E. Negation

Negation is an important but delicate point to interpret. The presence of a negation generally changes the opposition of the opinion.

V. MODEL EVALUATION

One of the most common and appropriate technique used for evaluation of a classifier is through confusion matrix.

The confusion matrix is given in a generalized form below.

	Predicted class1	Predicted class 2
Actual class 1	True positive(tp)	False negative(fn)
Actual class 2	False positive(fp)	True negative(tn)

By applying this technique we can derive the generalized evaluation parameters. These parameters include:

1) Accuracy: The classifier accuracy indicates how accurately the classifier can predict the result. It can be calculated using the below formula accurately the classifier has predicted the result. It can be calculated using the formula:

$accur$(a) = $tp + tn/tp + tn + fp + f$n

2) Precision: It represents how often the prediction by the classifier when it indicated true is correct. The formula for precision is represented below.

$precision(p) = tp/\ tp + fp$

3) Recall: Recall represent the true positive rate of the classifier.

Formula for recall is:

$recall(r) = tp/tp + fp$

4) F1 Score: It indicates the weighted average of recall and precision. Formula for recall is:

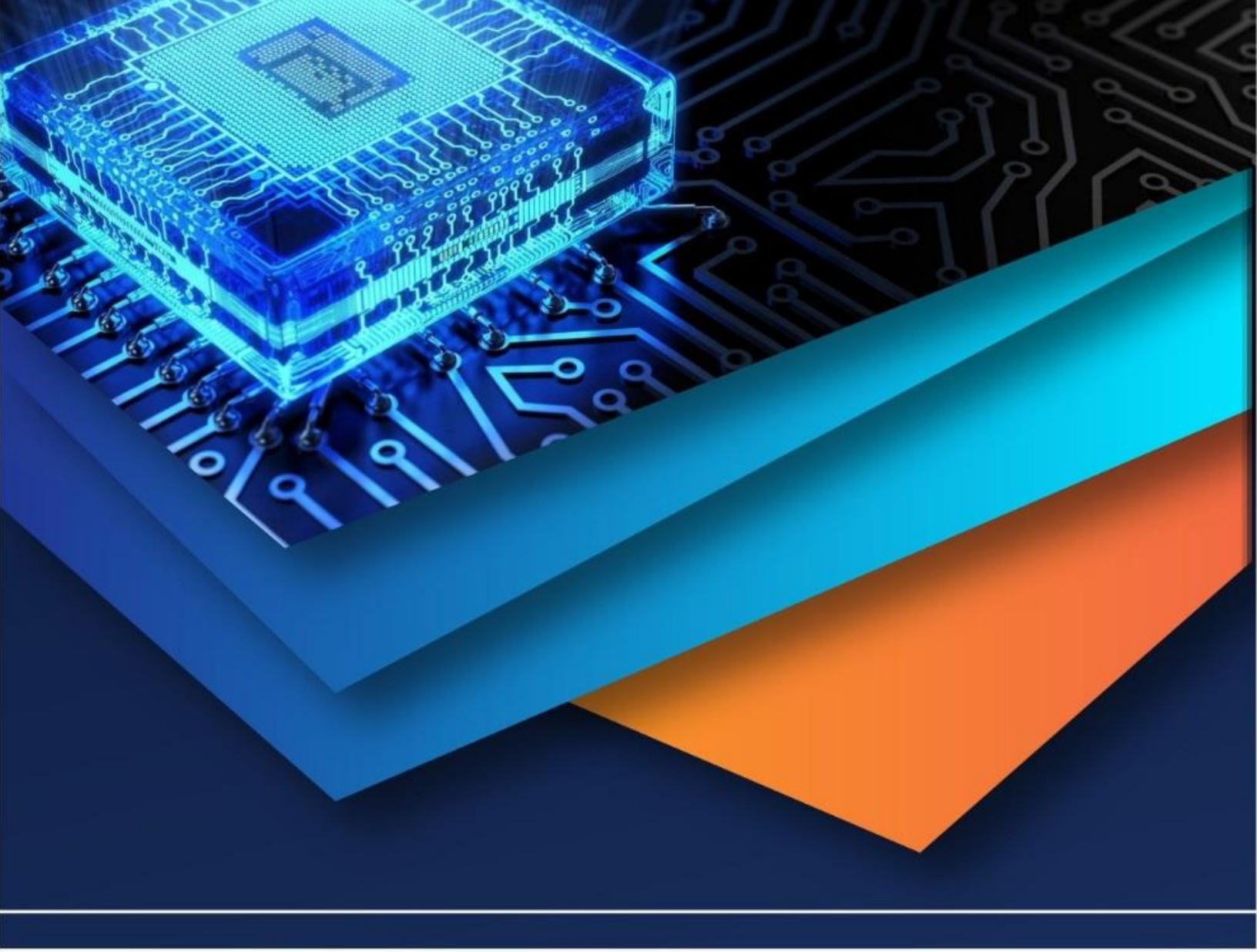
$F1\ score = 2p.\ r$/p-r

VI. CONCLUSION

In this paper, we provide a survey and comparative study of existing techniques for opinion mining including machine learning and lexicon-based approaches, together with cross domain and cross-lingual methods and some evaluation metrics. An attempted was made to compare the different techniques and outcomes of algorithms performance. Research results show that machine learning methods, such as SVM and naive Bayes have the highest accuracy and can be regarded as the baseline learning methods, while lexicon-based methods are very effective in some cases, which require few effort in human-labelled document .We also studied the effects of various features on classifier. We can conclude that more the cleaner data, more accurate results can be obtained. We can focus on the study of combining machine learning method into opinion lexicon method in order to improve the accuracy of sentiment classification and adaptive capacity to variety of domains and different languages.

REFERENCES

[1] "Arun k, Sinagesh a & Ramesh m, "twitter sentiment analysis on demonetization tweets in India using r language", international journal of computer engineering in research trends,vol.4, no.6, (2017), pp.252-258"
[2] "Aliza sarlan, chayanit nadam, shuib basri ," twitter sentiment analysis 2014 international conference on information technology and multimedia (ICIMU), november 18 – 20, 2014, putrajaya, malaysia 978-1-4799-5423-0/14/$31.00 ©2014 ieee 212"
[3] "Mandava geetha bhargava , duvvada rajeswara rao ," sentiment analysis on social media data using R" , international journal of engineering & technology, 7 (2.31) (2018) 80-84"
[4] "Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. "Sentimentanalysis of twitter data." In Proceedings of the ACL 2011,Workshop on Languages in Social Media, pp. 30–38, 2011.
[5] "IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012 ISSN (Online): 1694-0814"
[6] "International Journal of Innovations & Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 6, Issue 7 July 2017 "
[7] J. Ranganathan and A. Tzacheva, "Emotion mining in social media data," *Procedia Comput. Sci.*, vol. 159, pp. 58–66, 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)