2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-3701-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/Confluence52989.2022.9734201

A QUANTITATIVE PERFORMANCE EVALUATION OF MACHINE LEARNING ALGORITHMS FOR ANALYSING SENTIMENTS OF EMOTICONS

1st Aditi G
Department of Computer Science and Engineering,
Amity University,
Noida, Uttar Pradesh, India
aditi.ganapathi.2000@gmail.com

2nd Utsav Sharma
Department of Computer Science and Engineering,
Amity University,
Noida, Uttar Pradesh, India
utsavsharma1006@gmail.com

3rd Sumit Kumar
Department of Computer Science and Engineering,
Amity University,
Noida, Uttar Pradesh, India
sumitkumarbsr19@gmail.com

4th Jitendra Singh Jadon
Department of Electronics and Telecommunication,
Amity University,
Noida, Uttar Pradesh, India
jsjadon@amity.edu

***Abstract* - Twitter is one among the most popular platforms for people to express their opinions on a certain issue or product. As it consists of a multitude of users, Twitter is a significant source of data for companies to analyze product reviews, or general public opinion. On Twitter, a tweet can consist of 280 characters. Due to this limitation, users have resorted to using emoticons in a more efficient fashion. This has been achieved by replacing multiple words that indicate an emotion, with a single emoji. Traditional sentiment analysis algorithms and classifiers are highly efficient for usage on text data that does not contain any emojis. This paper evaluates the existing sentiment analysis algorithms in their capability to analyse emoticons in text.**

Keywords – ***Natural Language Processing, Sentiment Analysis, Twitter Sentiment Analysis***

I. INTRODUCTION

Business models of today are more customer-driven than profit-driven. With multiple organizations arising to offer similar goods and services, the victor of the contemporary business race is the one whose products and trade ethics most suit the clients. Online reviews make an impact on the mind of potential customers; many good reviews for a product urge people to try it out. Studies show that consumers are more likely to spend 31% more money on a product with positive feedback [1]. It has been found that online recommendation reviews form the basis of 79% of the customers' confidence [2]. Online reviews can also make a gargantuan impact on the conversion rates of a company – a massive gain of 37% was found to be the result of 100 positive reviews, while 200 of them can raise this gain to 44% [1].

Twitter is one of the most popular social media platforms, with 353 million monthly active users [3]. 94% of account-holders use Twitter on their mobile phones. Hence, it is undeniable that the introduction of real-time information networking platforms like Twitter has resulted in the formation of an unrivalled public collection of views on any worldwide entity of interest [4]. A sentiment has been defined as "a personal belief or judgment that is not founded on proof or certainty" [5]. Text data has always been the status quo, but due to the stringent character limit on Twitter, users have begun to use emoticons to convey sentiments. On an average, about 6 million tweets within a span of 2 hours on a given day contain emoticons [6].

Sentiment Analysis focuses on the analysis of subjective emotions. Fig. 1 shows the sentiment analysis classifiers used today.

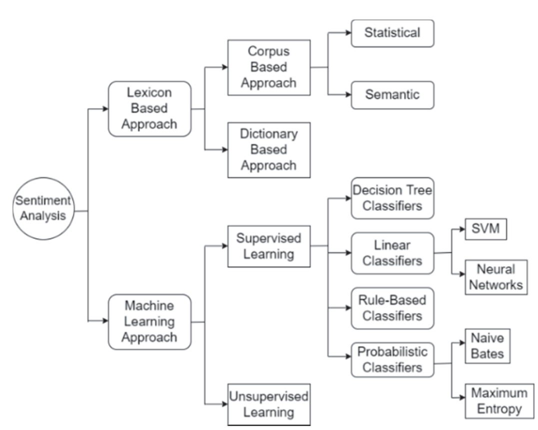

Fig. 1. The various approaches to solving a sentiment analysis problem

This paper aims to study the difference in accuracy by processing the same tweets with emojis, and after removing emojis to test the capability of the existing algorithms to process emoticons.

The Literature Review (Section II) examines the trends in sentiment analysis over the years, starting from enhancing lexicon-based models, to exploring emoticon sentiment analysis. The process of the experiment has been outlined in Proposed Methodology (Section III), while Results and Discussion (Section IV) discusses the outcomes of the experiment. Conclusion and Future Work (Section V) includes the final comments as well as the future scope of this project, and ways in which it can be expanded.

II. LITERATURE REVIEW

Sentiment analysis has been subjected to extensive research in terms of the accuracy of models, the different scenarios in which they can be used, features, vectorizers, etc. There used to be a glossary of words in the lexicon-based approach of sentiment analysis, where the dictionary had no records of contexts to the words used. Hassan et al. proposed SentiCircle, a variant of the lexicon-based approach that provided a situational definition to the words [7]. While this is an enhancement of the existing lexicon-based models, it is still expensive and has little scalability.

Pang et al. suggested the removal of objective sentences or facts, and process only text that has a subjective value [8]. The popular Naive Bayes approach based on the Bayes Theorem is used in the machine learning-based model. It is simple to use, but it is incorrect in that it presupposes each feature's strong independence. Kang et al. proposed an improved NB classifier to address the problem of positive classification accuracy being around 10% higher than negative classification accuracy [9]. Whipple found that for text data, Support Vector Machine (SVM) outperformed Naïve Bayes when used with the TfIdfVectorizer [10]. Vohra et al. theorize that a combination of various sentiment classifiers is effective depending on the domain of the purpose [11]. Devlin et al. put forward BERT, which used Transformer, an attention mechanism that learns contextual relationships between words in a text. Transformer has two different components in its basic form: an encoder that analyzes the input text and a decoder that generates a task prediction [12].

The above methods, however, only deal with data that is purely textual in nature.

Emoticons have always been considered noisy labels, or tags that are classified incorrectly at the time of data training due to the drawbacks in the currently used sentiment analysis models. It has been noticed that if the emoticons are not removed from the training data, there is a negative effect on the Support Vector Machine (SVM) and the Maximum Entropy classifiers; however, the impact this has on the Naïve Bayes model is found to be minimal [13]. Most emoji-based models use manually classified scores for emojis such as Plutchik's Eight Emotions and Ekman's Six Emotions [14]. Human classification is bound to have misinterpretations and a certain amount of discretion, and can be time taking for a large amount of data. Various approaches have been suggested during the wide research on implementing emojis in sentiment analysis. Yang et al. propose the creation of an emoticon lexicon that relies on textual data which contains emojis [15]. Another lexicon-based proposal was given by Castellucci et al., where a Twitter-specific lexicon was suggested using neural networks that analyzed tweets on their positivity, negativity and neutrality [16]. Go et al. and Boia et al. have approached this issue using distant supervision on emotionally tagged data to generate a polarity lexicon [13][17]. However, as per Aoki et al., the existence and usage of an emoticon lexicon on a classifier does not imply improved performance in comparison to text without emoticons [18].

Another problem that arises with the usage of emoticons is that a single emotion or sentiment has multiple emoticons associated with it. It becomes difficult for the classifier to track these variations. Tyagi et al. proposed the formation of an emotion vector for each emoticon [19]. Each element of an emoticon's vector represents the importance of a single emotion in defining that emoticon. Emoticon vectors were created by analyzing the emotive words that appeared beside each emoticon.

III. PROPOSED METHODOLOGY

Accumulation of Tweets with Emoticons (A) explains the process and criteria for scraping tweets from Twitter. Data Cleaning (B) discusses the procedure for removing all noise. Data Normalization (C) describes the process of making the dataset uniform and basic. Exploratory Data Analysis (D) describes the features of the dataset and its further simplification. Application of Classifiers (E) delineates the implementation of vectorizers and classifiers to the dataset to achieve results. The steps have been outlined in Fig. 2.

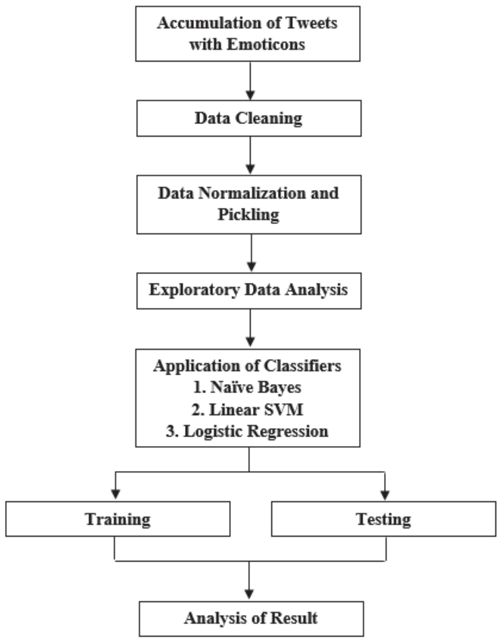

Fig. 2. The approach adopted to perform sentiment analysis on the dataset

A. Accumulation of Tweets with Emoticons:

The chosen dataset contains 1.6 million English language tweets that were extracted through the Twitter API. A total of 359 tweets were scraped, with about 73 search query terms used to collect positive and negative tweets [11]. Table I shows the number of tweets scraped for each category of query terms, and also details the percentage of their presence in the dataset.

TABLE I. QUERY TERMS THAT DETERMINED DATA COLLECTION

Category	Number of Tweets	Percent Presence in Dataset
Company	119	33.15%
Miscellaneous	67	18.66%
Person	65	18.11%
Product	63	17.55%
Movie	19	5.29%
Location	18	5.01%
Event	8	2.23%
Grand Total	**359**	**100.00%**

For a certain emotion, different people may use different emoticons. Hence, emojis were mapped with an emotion and grouped together to give better results. The mapping has been detailed in Table II.

TABLE II. EMOTICONS AND THEIR MAPPING

Base Emoticon	Mapped Emoticon(s)	Emotion
:)	:)	Happiness Joy Humor
	:-)	
	:)	
	=)	
	:D	
:(	:(	Sadness Boredom
	:-(	
	: (	

Apart from these features, the dataset has been manually given sentiment scores, where negative tweets have been marked with 0 and positive tweets have been marked with 4.

B. Data Cleaning:

Data cleaning was done in two different ways. By removing all the punctuation marks including semi-colons, colons, parentheses, dashes and the equals signs, it was ensured the removal of the emojis that contain these punctuation marks in the first data cleaning algorithm, hereafter labelled **'without emoticon'**. The second data cleaning algorithm, henceforth labelled the '**with emoticon'** set, was such that it retained the semi-colons, colons, parentheses, dashes and the equals signs in the tweets so as to preserve the emoticons in them. Apart from that, common to both the datasets, all hashtags, URLs and @ mentions, as well as retweets, whitespaces, and repeated characters in words were removed.

C. Data Normalization and Pickling:

Stopwords such as 'the', 'a', 'there', etc., were removed using the inbuilt stopwords provision in the NLTK module. The clean data was then tokenized. Using the inbuilt NLTK PortStemmer, stemming was performed on the tweets. The stemmed data was also lemmatized using the inbuilt NLTK WordNetLemmatizer. The data, now void of any stopwords, and fully stemmed and lemmatized, was pickled for use in later steps.

D. Exploratory Data Analysis:

The number of positive and negative tweets were counted and plotted on a bar graph. As they were equal in number, no balancing was required. The original dataset contains positive and negative tweets together. They were separated into two sub-datasets, one containing the positive tweets, and the other containing the negative tweets. Wordclouds were generated using the most frequently appearing terms in the sub-datasets.

E. Application of Classifiers:

The pickled file was loaded for sentiment analysis and was passed through the TfIdfVectorizer. The count returned by the vectorizer was stored as an array. The 'target' value, or the actual sentiment score allocated to each tweet in the form of 'neg' for negative tweets, and 'pos' for positive tweets, was stored as another array. Train-test split was authenticated, with the size of the test set being set to 20% of the whole data. Multinomial Naïve Bayes Classifier, Linear SVM, and Logistic Regression were instantiated, and the train data for both features was fit into them.

IV. RESULTS AND DISCUSSION

The dataset consisted of 1.6 million tweets; the number of positive tweets and negative tweets was equal at 0.8 million each, ensuring an equal distribution for train data. This has been represented in the Fig. 3 shown below.

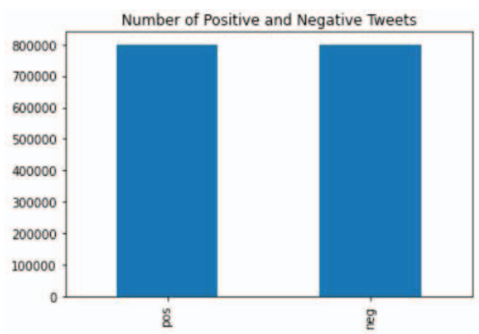

Fig. 3. The distribution of positive and negative tweets

Fig. 4(a) and Fig. 4(b) show the wordcloud representations of the 40 most common words in the respective categories.

Fig. 4(a). Positive Wordcloud

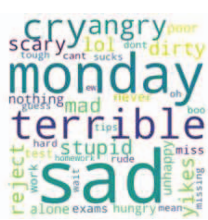

Fig. 4(b). Negative Wordcloud

An interesting thing to note from the above wordclouds is the presence of the word 'lol' – short for 'laughing out loud' in both the positive and the negative contexts. This might be because 'lol' is an abbreviation very popularly used in sarcastic or moody situations as well, apart from its usual usage as a reaction to something humorous.

The classification reports for the dataset using the three classifiers have been tabulated in Table III and Table IV.

TABLE III. WITHOUT EMOTICON CLASSIFICATION REPORT

	Multinomial Naïve Bayes	**Linear SVM**	**Logistic Regression**
Accuracy	0.790	0.790	0.80
Avg Precision	0.785	0.785	0.80
Avg Recall	0.785	0.79	0.80
Avg f1-score	0.785	0.785	0.80

TABLE IV. WITH EMOTICON CLASSIFICATION REPORT

	Multinomial Naïve Bayes	**Linear SVM**	**Logistic Regression**
Accuracy	0.780	0.780	0.790
Avg Precision	0.785	0.780	0.790
Avg Recall	0.785	0.780	0.795
Avg f1-score	0.785	0.780	0.790

Their ROC curves have been plotted in Fig. 5(a) and Fig. 5(b).

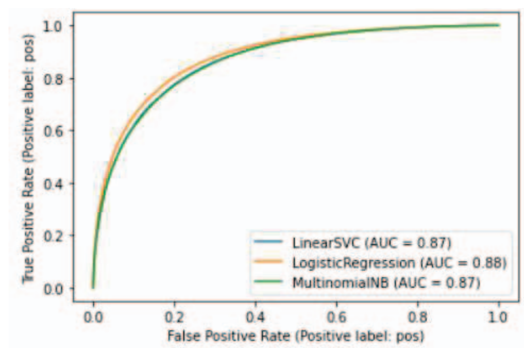

Fig. 5(a). Train and test accuracy for the without emoticon set

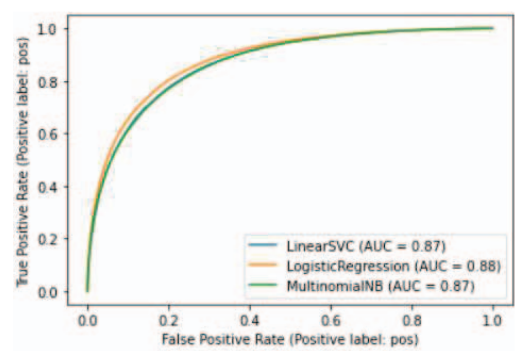

Fig. 5(b). Train and test accuracy for the with emoticon set

In the ROC curves, one observes that both SVM and Naïve Bayes have an AUC of 0.87, which results in overlapping curves.

There is little difference between the with emoticon and without emoticon results using the Naïve Bayes classifier. This might be due to the fundamental mathematical logic behind its functioning. Naïve Bayes works differently as compared to SVM and the Maximum Entropy models with regards to feature weight selection. While tweets with emojis do have an extra feature, i.e., the emoji, from which sentiment can be analyzed, more weight is given to the other words present in the tweet.

The classification report shows a minor decrease in accuracy of the Linear SVM as well as the Logistic Regression models in the without emoticon and with emoticon sets. This is possibly due to the fact that Naïve Bayes treats every feature of the tweet as an independent entity, and therefore, the emoticon is a separate feature of the data which it might singularly rely on. Logistic Regression, on the other hand, sets the linear decision function low for negative classes, and high for positive classes, hence labelled a discriminative model.

However, if two features are correlated, Naïve Bayes allots to them high weights, therefore counting their effect on the data twice. This leads to a breach in the conditional independence feature on which the Naïve Bayes classifier functions. In contrast, when correlated features occur in Logistic Regression, it weighs the repeated feature lower. Therefore, Logistic Regression works better in this case. Hence, of the three classifiers, Logistic Regression shows the best accuracy.

V. CONCLUSION AND FUTURE WORK

The Naïve Bayes classifier does not discriminate between tweets containing emoticons, and tweets that do not contain them. However, Linear SVM and Logistic Regression display similar results when tested with the with emoticon and without emoticon datasets. This is evident from the accuracy scores obtained during the course of the experiment; the without emoticon dataset gives a 79% test accuracy with Naïve Bayes and Linear SVM classifiers, and 80% with Logistic Regression. Naïve Bayes and Linear SVM see 78% accuracy in the with emoticon dataset, while Logistic Regression again yielding a higher accuracy score of 79%. From these findings, we can conclude that Logistic Regression performs the best under both the conditions.

This paper only considered emoticons which were formed out of punctuation marks. However, in the future, tweets with emojis, or image emoticons (emojis), may be studied similarly. This study can also be extended to implement and analyse the performance of more advanced classifiers, like neural networks, to compare their difference in analyzing emoticons and emojis.

REFERENCES

[1] Nick Galov, "Staggering online review statistics that will help you improve your business in 2021", Hosting Tribunal, 2021.

[2] B. Thomas, "What consumers think about brands on social media, and what businesses need to do about it" Keep Social Honest, Report, 2013.

[3] Brian Dean, "How many people use twitter in 2021?", Backlinko, February 2021.

[4] Huseyin Bicen and Nadire Cavus, "Twitter usage habits of undergraduate students", Procedia - Social and Behavioral Sciences, Volume 46, pp. 335-339, 2012.

[5] Dmitry Davidov, Oren Tsur, and Ari Rappoport, "Enhanced sentiment learning using hashtags and smileys", COLING '10: Proceedings of the International Conference on Computational Linguistics: Posters, pp. 241-249, August 2010.

[6] M. Li, E. Chang, A.Y.L Chong , and S. See, "An empirical analysis of emoji usage on Twitter", Industrial Management & Data Systems, Vol. 119 No. 8, pp. 1748-1763, September 2019.

[7] Hassan Saif, Yulan He, Miriam Fernandez, and Harith Alani, "SentiCircles for contextual semantics for semantic analysis of Twitter", ScienceDirect., Information Processing Management, Volume 52, Issue 1, pp. 83-98, January 2016.

[8] Pang, B. and Lee, L., "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts", Association of Computational Linguistics (ACL), pp. 271-278, 2004.

[9] Hanhoon Kang, Seong Joon Yoo, and Dongil Han, "Senti-lexicon and improved naïve bayes algorithms for sentiment analysis of restaurant reviews", Expert Systems with Applications, vol. 39., no. 5, pp. 6000-6010, 2008.

[10] Adam Lane Whipple, "Comparison of algorithms for Twitter sentiment analysis", The University of Texas at Austin, 2017.

[11] Saifee Vohra and Jay Teraiya, "Applications and challenges for sentiment analysis: a survey", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 2, pp. 907-910, 2013.

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: pre-training of deep bidirectional transformers for language, understanding", arXiv, 2019.

[13] Alec Go, Richa Bhayani, and Lei Huang, "Twitter sentiment classification using distant supervision", CS224N Project Report, Stanford University.

[14] Bjarke Felbo, Alan Mislove, Anders Sogaard, Iyad Rahwan, and Sune Lehmann, "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm", Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1615-1625, August 2017.

[15] Yang, C., Lin, K. H.-Y., Chen, H.-H, "Building emotion lexicon from weblog corpora", In Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, Association for Computational Linguistics, pp. 133–136, 2007.

[16] G. Castellucci, D. Croce, and R. Basili, "Acquiring a large-scale polarity lexicon through unsupervised distributional methods", In Natural Language Processing and Information Systems, Springer, pp. 73–86, 2015.

[17] M. Boia, B. Faltings, C.-C Musat, and P. Pu, "A :) is worth a thousand words: how people attach sentiment to emoticons and words in Tweets", In Social Computing (SocialCom), International Conference on, IEEE, pp. 345–350, 2013.

[18] S. Aoki, and O. Uchida, "A method for automatically generating the emotional vectors of emoticons using weblog articles", In Proc. 10th

WSEAS International Conference on Applied Computer and Applied Computational Science, Stevens Point, Wisconsin, USA, pp. 132–136, 2011.

[19] Priyanka Tyagi, and R.C. Tripathi, "A review towards the sentiment analysis techniques for the analysis of twitter data", Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE), February 2019.

[20] Siti Sendari, Ilham Zaeni, Dian Lestari, and Hanny Hariyadi, "Opinion analysis for emotional classification on emoji tweets using the naïve bayes algorithm", Knowledge, Engineering and Data Science (KEDS), Vol. 3, No. 1, pp. 50-59, July 2020.