Available online at www.garmian.edu.krd

Passer Journal

Passer 4 (2022) 71-77

http://passer.garmian.edu.krd/

Machine Learning Approach to Sentiment Analysis in Data Mining

Dalshad Jaafar Hussein[1], Mustafa Nihad Rashad[2], Kosar Ibrahim Mirza[1], Dana Lateef Hussein[3*]

[1]*Database Department, Computer Science Institute, Sulaimani Polytechnic University, Kurdistan Region, Iraq*
[2]*Information Technology Department, Chamchamal Technical Institute, Sulaimani Polytechnic University, Kurdistan Region, Iraq*
[3]*Information Technology Department, Computer Science Institute, Sulaimani Polytechnic University, Kurdistan Region, Iraq*

Received 30 October 2021; revised 22 January 2022;
accepted 14 March 2022; available online 31 March 2022

doi:10.24271/psr.43

ABSTRACT

Widespread internet use and the web have brought about new ways of expressing individual sentiments. A sentiment is defined as an individual's view in which feelings, attitudes, and thoughts can be represented. When it comes to analysing and extracting Sentiment analysis and opinion mining are two of the most prominent disciplines of research. They derive insights using text data through numerous sources like Facebook and Twitter. Sentiment analysis frequently elicits information on how people feel about various events, brands, products, or businesses. Researchers collect and improvise replies from the general public to conduct evaluations. This paper looks into sentiment analysis for classifying Twitter subscriber tweets. This approach can help analysing the information gathered and stored in positive, neutral and negative opinions. This information is first pre-processed before creating feature vectors. On the basis of machine learning, classification methods were used. The study's algorithms are used Maximum Entropy, Naive Bayes and Support Vector Machine; they are used to categorize documents as positive or negative. The dataset for this paper are obtained from Twitter and includes subscribed tweets by using the API. Following pre-processing, machine learning methods are used to determine whether the tweets are positive or negative.

Keywords: Machine Learning, Twitter, Sentiment analysis, Opinion Mining, Social Medi.

1. Introduction

Sentiment analysis has emerged as an interesting and popular research area in recent years. It is used to review and analyse the opinions of a large number of people[1]. These tweets can be about an event, a brand, a person, or a product. Previously, people's opinions were expressed through magazines, newspapers, and other media. However, as technology has advanced, people express their emotions on various social networking and microblogging sites such as Twitter, Facebook etc.[2-3].

Tweets are short messages that users post to express their sentiment and attitude toward a particular subject[4-5]. Individuals' opinions have been extracted, studied, and then evaluated constructively by researchers[6]. In recent years, Twitter has surpassed all other microblogging platforms in popularity. It can be regarded as a reliable indicator of people's feelings[6-7]. Several media organizations have developed various methods to mine Twitter information[8]. This is because it is easy to conduct training, testing, and analysis since tweets are collected datasets using Twitter API[9-10]. Also, Twitter users post messages on any topic they can think of[9-11]. This differs from other microblogging sites in that only a specific topic and purpose are discussed. Another factor has led to an increase in harnessing Twitter data's real-time nature[11]. This paper seeks to apply Machine learning techniques in performing sentimental analysis on customer reviews data extracted from Twitter[12].

* Corresponding author
E-mail address: dana.hussein@spu.edu.iq (Instructor).
Peer-reviewed under the responsibility of the University of Garmian.

Sentiment analysis,is often known as opinion mining,it is a technique for determining the general public's opinion on a specific subject, such as positive or negative[13-14]. Using natural language processing skills, is feasible to develop a model that predicts sentiment's polarity, whether it is positive or negative, for each tweet[5,15-17]. Machine learning algorithms like support vector machine, Maximum Entropy and Naive Bayes are used to classify documents as negative or positive. In this study, a model is developed of neural networks to do two-class categorize on each tweet[8-9,17-19]. All tweets are tokenized in the pre-processing stage and then presented each tweet as a fixed-length vector using the Word Embedding for text analysis techniques[5,20]. The paper aims to develop a different system capable of classifying each input tweet as positive or negative sentiment using recently developed Algorithms.

2. Theoretical Background

In (Figure 1), we illustrated our proposed framework for sentiment analysis. This diagram illustrates the data collection process, which involves the use of Twitter's Application Programming Interface (API) and the pre-processing of the gathered twitter data[18]. In (Figure 2), we illustrated our sentiment analysis diagram.

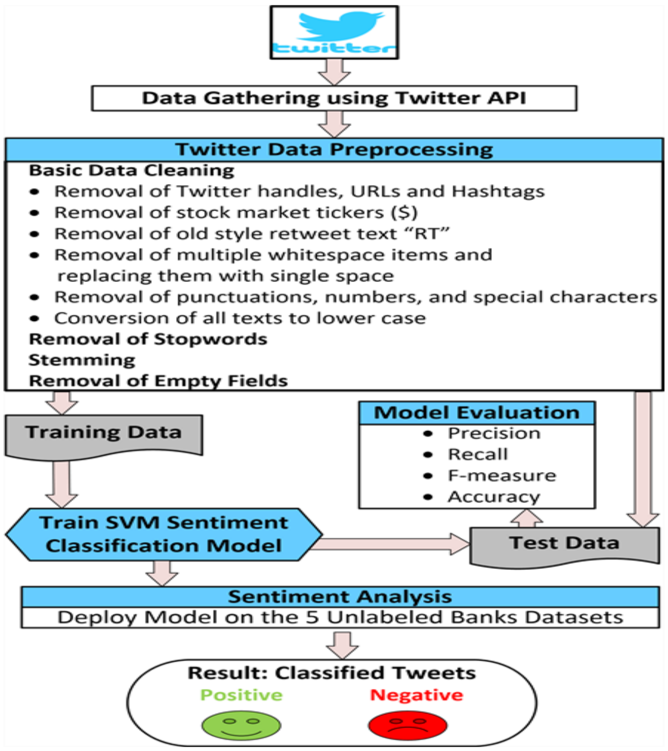

Figure 1: Framework for sentiment analysis that has been proposed.

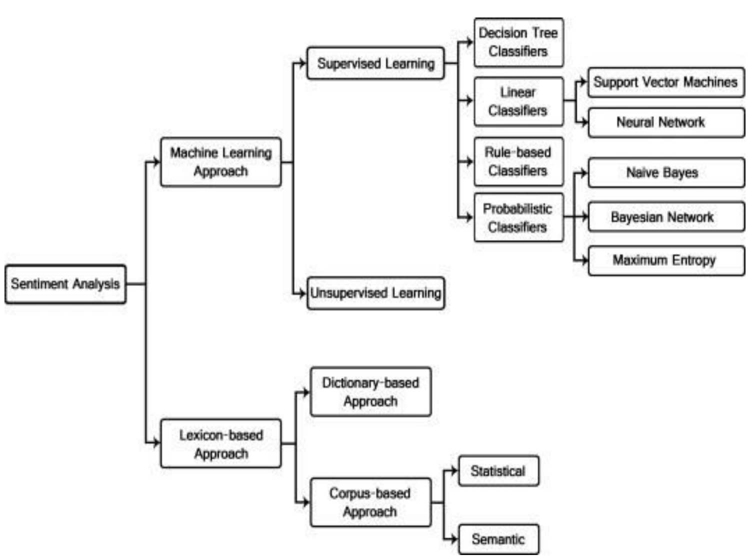

Figure 22: Techniques for Sentiment Analysis diagram.

3. Related Work

Today, many topics are talking about how to figure out people feeling based on the text as well as many researchers have been developing algorithms for extracting and analysing massive amounts of user-generated data from Facebook and Twitter. For text mining purposes, two primary methods have been used to determine how well a review is positive or negative, it is important to know that. Techniques such as semantic orientation and machine learning are employed two of these methods[1]. The review dataset uses techniques such as Naive Bayes. The scientists suggest an app for assessing Arabic sentiment in Twitter data. 2000 tweets (One thousand negative tweets and one thousand positive tweets) are evaluated for polarity detection using Support Vector Machine, Decision Tree and Naïve Bayes. The accuracy of classifiers is improved by applying a feature vector technique. Among the issues which are raised by researchers on the training, data are multiple tweet occurrences, spammed opinions and dual viewpoint tweets[7]. Performance evaluation is done using the False Negative, False Positive, True Negative and True Positive, values from results[18]. The study's primary goal is to ascertain the distributions of negative, positive, and neutral polarity. Numerous approaches and classifiers are employed, including the lexicon-based style, the Support Vector Machines and Naive Bayes classifier algorithm[19]. In another study, sentiments are determined using Natural Language Processing (NLP) techniques and called Sentiment Analyzer, which extracts sentiments automatically and can be utilized for efficiently discover all references on a given subject[21]. Many scholars have attempted to use data mining techniques to extract sentiments from text data. Some studies which are chosen for this research are listed below. Some datasets are classified using the classifier and Multilayer Perceptron (MLP)[22]. In a recent article, the authors propose a strategy for analysing movie and hotel review datasets using IMDb and OpinRank dataset, respectively. Each dataset consisted of 5000 positive and 5000 negative reviews. They obtains the most significant results when the training dataset had 4500 reviews. The accuracy values for Nave Bayes and K Nearest Neighbor (KNN) 82.43 per cent and 69.81 per cent, respectively, in the movie reviews dataset, whereas the values for the hotel reviews dataset were 55.09 per cent and 52.14 per cent, respectively[23].

4. Methodology

4.1 Data Extraction

The data for this research are gathered from people's tweets posted on Twitter. The tweets are extracted using the Twitter API. The public API's " Tweepy library" version is used and implemented in Python[24]. "Tweepy library is an open-source" platform Python package that simplifies this use of the Twitter API by providing its own set of classes and functions[24]. Python's Tweepy framework is used in conjunction with Twitter's streaming API. In Figure 3, a sample of code using an API can be run directly on web servers or local hosts, and only a few parameters are considered for the query. During the extraction of tweets from Twitter, a large number of filtering parameters are specified so that they can fit any precise criteria. After the query has been generated, an API is utilized to keep it running. This query will return all the required twitter data[25-26].

```
import tweepy,panda as pd
import sys
import jsonpickle
import os,random

# Authenticate to Twitter
auth =
tweepy.OAuthHandler("xOCDelyewVjVLvqUhVPOFnis
D", "sd6YM3RScvq8qz9yG0P9GmZBuPNG195Z4bLjV")
auth.set_access_token("ACCESS_TOKEN","ACCESS_
TOKEN_SECRET")api = tweepy.API(auth)# test
authentication
try:
    api.verify_credentials()
    print("Authentication OK")
except:
    print("Error during authentication")
```

Figure 3: Tweepy library API code syntax.

4.2 Data Pre-processing

Within the data extracted from Twitter, there exists irrelevant data. Any random characters must be filtered from the tweet data[27-28]. This useless data is filtered out using the Natural Language Processing tool[28]. This NLP tool outputs any grammatical relationship between the words of a sentence. These relationships are utilized to find tweets with relevant information. Facilitating filtering and adding more relationships has little effect on the outcomes. Before extracting the features, another pre-processing step is undertaken to filter out slang words and misspellings. During pre-processing, stop word removal, stemming, punctuation mark removal and tokenization have all been performed. It has been transformed into a bag of words.

4.3 Dataset

The full dataset contains the 1.6 million tweets retrieved via the Twitter API and stored in the CSV file. The Sentiment 140 dataset is utilized in this study to train and verify the model. The sentiment label of each tweet has been annotated 0 for negative or 1 for positive, respectively. In Table 1, the researchers illustrated their corresponding sentiment example. This dataset includes the two-class label. Sentiment140 provides 335,650 unique words[4]. After pre-processing, each tweet had, on average, 60 characters or 11 words. The researchers chose 128 million tweets at random as training data and 32 million tweets as validation data. Utilizing a word embedding with 400,000 words and 200 dimensions.

Table 1: An Illustration of some of Tweets in 1.6 million tweets, each with a different sentiment positive and negative labled.

id	sentiment	text
1	1	@stellargirl I loooooooovvvvvveee my Kindle2. Not that the DX is cool, but the 2 is fantastic in its own right.
2	1	Reading my kindle2... Love it... Lee childs is good read.
3	1	Ok, first assesment of the #kindle2 ...it fucking rocks!!!
4	1	@kenburbary You'll love your Kindle2. I've had mine for a few months and never looked back. The new big one is huge! No need for remorse! :)
5	1	@mikefish Fair enough. But i have the Kindle2 and I think it's perfect :)
6	1	@richardebaker no. it is too big. I'm quite happy with the Kindle2.
7	0	Hate this economy. I hate aig and their non loan given asses.

4.4 Pre-processing of Tweets

The first step is to convert the tweets to lower case. As a result, the researchers may obtain the words from each tweet in the same case (i.e., lower case). Then, all URLs are removed and replaced with plain text in the following phase. Then, substituting the generic term AT USER for "@username." The following step eliminates punctuation from the beginning and conclusion of tweets and replaces additional white spaces with a single white space. The #hashtag is replaced with the same term, sans the hash.

In Figure 4 the researchers are illustrated a simple dataset diagram.

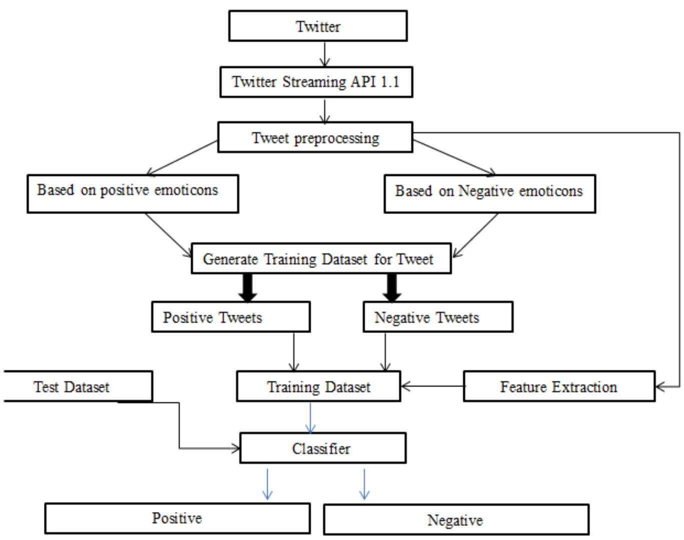

Figure 4: Dataset process diagram collect data and extract with analyses to positive and negative

4.5 Modelling of Feature Vector

To begin, remove any stop words from tweets. Then, using the two characters, replace the character that appears more than twice in the given term, i.e., trim the character that appears more than once. For example, replace “Testttt” with “Test” etc. The following section contains examples of feature words retrieved from sample tweets. See Table 2.

Table 2: An Illustration of Tweets and Featured Words with positive and negative example

Positive Tweets	Feature Words
Bajrangi Bhaijann The film is exceptionally positive .Celebrate Humanity. Doesn’t take any religion or country’s side.	‘positive’, ’Humanity’, ’religion’, ’country’s’, ’side’

Negative Tweets	Feature Words
AT_USER disappointed. Watched a movie. It is a waste of time.	‘disappointed’,watched’,’movie’,’waste’,’time’
I miss my mom and dad. I hate this life.	‘miss’, ‘hate’

4.6 Classification Techniques

Classification algorithms are used to classify the text using a machine learning-based approach. These machine learning approaches can be classified into two techniques:

• Unsupervised learning techniques where there is no category involved and they do not supply any targets. As a result, clustering is a critical factor in this case.

• Supervised learning uses of labelled datasets, the labels are given to the model when the classification strategy is being developed. These labelled datasets are trained in order to obtain significant outputs when making decisions. Both of these learning techniques have been successful in determining and extracting specific sets of features that can be used to detect sentiments. Semi supervised and unsupervised algorithms are used when it is not easy to have labelled opinions for training the different classifiers[29-30].

4.6.1 Classification Techniques

This classifier makes use of the large number of characteristics in the feature vector[29-31]. Because these characteristics are equally independent, it is critical to examine them separately. The mathematical representation of the Naive Bayes conditional probability is given by the formula, in (Figure 5), NB formula is lustrated[19,32]. It is an easy for probabilistic classifier Bayes' theorem that is well suited for high-dimensional inputs. Text classification is the process by which a given document is allocated a class. It is simple, quick and easy to forecast test dataset. Additionally, NB excels at multiclass prediction[33].

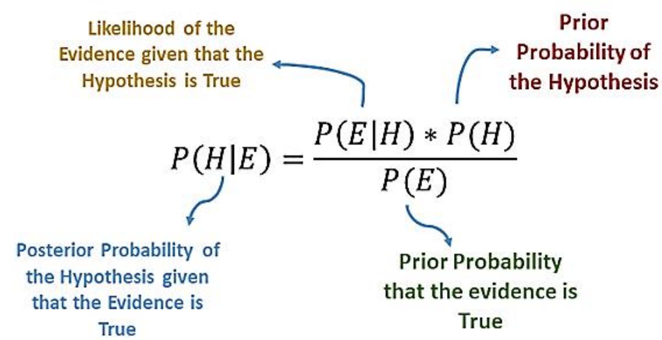

Figure 5: Naive Bayes Theorem

4.6.2 Support Vector Machine (SVM) Classifier

SVM introduces support vector machines, which essentially seek the optimal surface for separating training samples as a negative and positive. It is classification techniques that rely on supervised learning[19,34]. It is developed by[35]:

4.6.3 Maximum Entropy Classifier

In the relationship between features, the maximum entropy classifier makes no assumptions. This classifier estimates the distribution of class label by maximizing the system's entropy[2,29,31]. The following is a mathematical depiction of its conditional distribution in Maximum Entropy Theorem equation1:

$$\boldsymbol{P_{ME(c\backslash d,\lambda)}} = \frac{\mathbf{exp}\ [\sum_i \lambda_i \boldsymbol{f}_i\,(\boldsymbol{c},\boldsymbol{d})]}{\sum_c \mathbf{exp}\ [\sum_i \lambda_i \boldsymbol{f}_i\,(\boldsymbol{c},\boldsymbol{d})]} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{1})$$

4.7 Word2Vec Features

Word embeddings are a method for encoding words as vectors in the current era. The purpose of word embeddings is to reduce the high-dimensional properties of words to low-dimensional feature vectors while keeping the corpus's contextual similarity [18,36].

4.8 Model Evaluation Parameters

It is necessary to evaluate the classifier's performance following classification using SVM.

This is accomplished through the use of testing dataset in this research. Precision, Recall, F-measure, and Accuracy are all metrics used to evaluate the performance of text categorization systems.

4.8.1 Performance evaluation

Various metrics is used to determine the performance of a ML algorithm based on a contingency matrix. The true positive (TP). False positives (FP), true negative (TN) and false negative (FN) see Accuracy formula [2].

$$\boldsymbol{Accuracy} = \frac{\mathbf{TP + TN}}{\boldsymbol{TP + TN + FP + FN}} \quad \dots\dots\dots\dots\dots\dots\dots.(2)$$

While the data have been balanced, it would be beneficial to examine the Specificity metric for comparison, particularly for Specificity datasets. The following syntax could be used to express the procedure see sensitivity formula[3] and Specificity formula[4]:

$$\boldsymbol{Sensitivity} = \frac{\mathbf{TP}}{\boldsymbol{TP + FN}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. \ (3)$$

$$\boldsymbol{Specificity} = \frac{\mathbf{TN}}{\boldsymbol{TN + FP}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots..(4)$$

4.8.2 Precision

Classifier's precision with respect to each class. It is said as follows see Precision formula [5]:

$$\boldsymbol{Precision} = \frac{\mathbf{TP}}{\boldsymbol{TP + FP}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots..(5)$$

4.8.3 Recall

Classifier's completeness with respect to each class. It is stated as follows see Recall formula [6]:

$$\boldsymbol{Recall} = \frac{\mathbf{TP}}{\boldsymbol{TP + FN}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.\dots..(6)$$

4.8.4 F-Measure

Is the acoustical equivalent of precision and recall. It is stated as follows see F-measure formula [7]:

$$\boldsymbol{F - measure} = \frac{\mathbf{2 * Precision * Recall}}{\mathbf{Precision + Recall}} \quad \dots.\dots\dots\dots(7)$$

4.8.5 Kappa Statistics

Cohen's Kappa coefficient quantifies the classifier's performance in comparison to guessing with a random classifier. It accomplishes this by comparing the measured accuracy to the expected accuracy (random chance) see Kappa Statistics formula[8]:

$$\boldsymbol{Kappa\ Statistics} = \frac{\boldsymbol{P_0 - P_e}}{\boldsymbol{1 + P_e}} \quad \dots\dots\dots\dots\dots\dots\dots.\dots..(8)$$

where $P_0 = \frac{\mathrm{TP+TN}}{TP+TN+FP+FN}$ Is the observed accuracy and $Pe = expected$

5. Results and Discussion

The purpose of this research is to identify which of the SVM, Naive Bayes, and Maximum Entropy machine learning algorithms provides better results at text classification machine learning methods are used to determine whether the tweets are positive or negative. This is performed by utilizing the dataset of Twitter tweets. Classifiers are tested by comparing their accuracies across a variety of experimental conditions. For analysis, TP, FP, TN and FN are utilized to compute performance metrics like accuracy, f-score, recall and precision. The table 3 shows the performance of the classifiers, performance statistics of several classifiers. Naïve Bayes achieves the highest accuracy compared to the other classifiers. Its accuracy of 88.50%, Maximum entropy has the lowest at 60.67 while SVM was 87.33% accurate. Naïve Bayes has high F-score and precision values as compared to the rest. However, SVM has the highest Recall. It quantifies the number of positive predictions correctly made out of all positive predictions in the dataset. See Figure 6.

Table 3: An Illustration of Performance statistics comparation

Method	Recall (%)	Precision (%)	F-score (%)	Accuracy (%)
Naive Bayes	83.33	92.94	87.87	88.50
SVM	89.33	85.90	87.58	87.33
Maximum Entropy	84.67	57.21	68.28	60.67

All the classifiers in this paper are sensitive to parameter optimization. In as much as Naïve Bayes classifier achieved better results on the subscribers' tweets data, SVM achieves the best results especially with a smaller dataset. The high recall value in SVM indicates that SVM returned the highest number of relevant results.

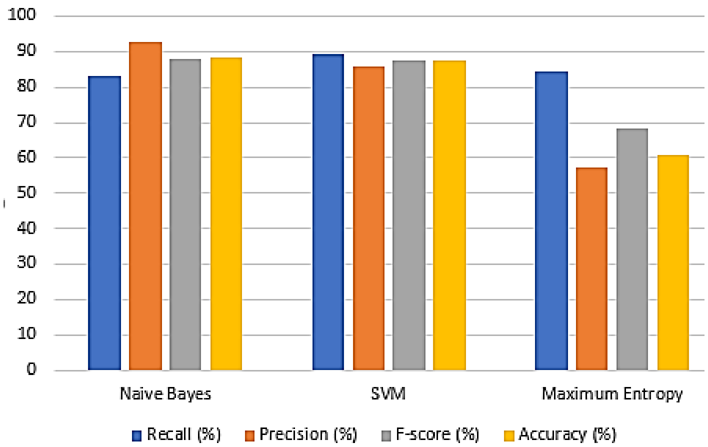

Figure 6: Performance of different classifiers result

6. Conclusion

It has been concluded that machine learning is a much simpler and more efficient technique than symbolic ones. These techniques are easily applicable to sentiment analysis on Twitter. Following that, features are retrieved from the tweet, which is simply plain text devoid of hash tags or slang terms. And then these retrieved features are combined to create the feature vector. Different machine learning classifiers are used to categorize tweets. Sentiment analysis is critical in understanding the feelings expressed about anything, including tweets, posts, products, social media, and so on. Machine Learning techniques can be used to perform sentiment analysis. Machine learning on the other hand is simpler and more efficient but requires labelled data. This research applies a machine learning strategy in classifying polarity in subscribers' tweets. First, a dataset is gathered from Twitter using Twitter API. The data is then pre-processed using the Natural Language Processing tool (NLP) before being divided into two sets: the train and test sets. Next feature vectors are created after which the train set is trained using Machine Learning classifiers. Once the data has been trained, the test set is used to evaluate how good the classifiers performed. By Using a confusion matrix, the performance of the classifiers has been measured. From this experiment, Naive Bayes recorded an accuracy of 88.50% and a Precision of 92.94%, which is the highest. This means that Naive Bayes is the best classifier for returning the highest number of relevant results, whereas Naïve Bayes is the lowest number in the recall. However, SVM has the highest recall.

Conflict of interests

None.

References

1 M. Rezapour, "Sentiment classification of skewed shoppers' reviews using machine learning techniques, examining the textual features," Engineering Reports, vol. 3, no. 1, 2021.

2 H. A. Alatabi and A. R. Abbas, "Sentiment analysis in social media using machine learning techniques," Iraqi J. Sci., pp. 193–201, 2020.

3 M. W. Habib and Z. N. Sultani, "Twitter sentiment Analysis using different machine learning and feature extraction techniques," Al-Nahrain Journal of Science, vol. 24, no. 3, pp. 50–54, 2021.

4 M. Ghiassi and S. Lee, "A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach," Expert Syst. Appl., vol. 106, pp. 197–216, 2018.

5 B. Noori, "Classification of customer reviews using machine learning algorithms," Appl. Artif. Intell., vol. 35, no. 8, pp. 567–588, 2021.

6 A. Mitra, "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)," September 2020, vol. 2, no. 3, pp. 145–152, 2020.

7 M. M. Altawaier and S. Tiun, "Comparison of machine learning approaches on Arabic twitter sentiment analysis," Int. J. Adv. Sci. Eng. Inf. Technol., vol. 6, no. 6, p. 1067, 2016.

8 A. Kumar and A. Jaiswal, "Systematic literature review of sentiment analysis on Twitter using soft computing techniques," Concurr. Comput., vol. 32, no. 1, p. e5107, 2020.

9 F. C. Onwuegbuche, J. M. Wafula, and J. K. Mung'atu, "Support vector machine for sentiment analysis of Nigerian banks financial tweets," J. data anal. inf. process., vol. 07, no. 04, pp. 153–173, 2019.

10 A. M. Ramadhani and H. S. Goo, "Twitter sentiment analysis using deep learning methods," in 2017 7th International Annual Engineering Seminar (InAES), 2017.

11 R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment analysis on product reviews using machine learning techniques," in Cognitive Informatics and Soft Computing, Singapore: Springer Singapore, 2019, pp. 639–647.

12 A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentimental reviews using machine learning techniques," Procedia Comput. Sci., vol. 57, pp. 821–829, 2015.

13 M. Krommyda, A. Rigos, K. Bouklas, and A. Amditis, "An experimental analysis of data annotation methodologies for emotion detection in short text posted on social media," Informatics (MDPI), vol. 8, no. 1, p. 19, 2021.

14 A. S. Rathor, A. Agarwal, and P. Dimri, "Comparative study of machine learning approaches for Amazon reviews," Procedia Comput. Sci., vol. 132, pp. 1552–1561, 2018.

15 C.-Y. Huang, H. Tong, J. He, and R. Maciejewski, "Location prediction for tweets," Front. Big Data, vol. 2, p. 5, 2019.

16 R. Motta Goulart, C. C. Hasenclever Borges, and R. Fonseca Neto, "An evolutionary algorithm for large margin classification," Soft Comput., vol. 25, no. 11, pp. 7593–7607, 2021.

17 A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts," Math. Comput. Appl., vol. 23, no. 1, p. 11, 2018.

18 M. W.Habib, Computer Science Department, College of Science, Al-Nahrain University, Baghdad-Iraq, Z. N. Sultani, and Computer Science Department, College of Science, Al-Nahrain University, Baghdad-Iraq, "Twitter sentiment Analysis using different machine learning and feature extraction techniques," Al-Nahrain Journal of Science, vol. 24, no. 3, pp. 50–54, 2021.

19 K. Korovkinas and P. Danėnas, "SVM and Naïve Bayes Classification Ensemble Method for Sentiment Analysis," Balt. J. Mod. Comput., vol. 5, no. 4, 2017.

20 A. Kanavos, N. Nodarakis, S. Sioutas, A. Tsakalidis, D. Tsolis, and G. Tzimas, “Large scale implementations for twitter sentiment classification,” Algorithms, vol. 10, no. 1, p. 33, 2017.

21 R. Joshi and R. Tekchandani, “Comparative analysis of Twitter data using supervised classifiers,” in 2016 International Conference on Inventive Computation Technologies (ICICT), 2016.

22 S. N. Jayesh Chaudhary1, “Twitter Sentiment Analysis using Tweepy,” Twitter Sentiment Analysis using Tweepy, vol. 08, no. 04, p. 4512, Apr. 2021.

23 L. Dey, Department of Computer Science & Engineering, Heritage Institute of Technology, Kolkata, India, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, “Sentiment Analysis of Review Datasets Using Naïve Bayes‘ and K-NN Classifier,” Int. j. inf. eng. electron. bus., vol. 8, no. 4, pp. 54–62, 2016.

24 P. Baid, A. Gupta, and N. Chaplot, “Sentiment analysis of movie reviews using machine learning techniques,” Int. J. Comput. Appl., vol. 179, no. 7, pp. 45–49, 2017.

25 A. Rasool, R. Tao, K. Marjan, and T. Naveed, “Twitter sentiment analysis: A case study for apparel brands,” J. Phys. Conf. Ser., vol. 1176, p. 022015, 2019.

26 R. Jansher, “Sentimental analysis of Amazon product reviews using machine learning approach,” Unpublished, 2020.

27 M. H. Abd El-Jawad, R. Hodhod, and Y. M. K. Omar, “Sentiment analysis of social media networks using machine learning,” in 2018 14th International Computer Engineering Conference (ICENCO), 2018.

28 A. Onan, “Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach,” Comput. Appl. Eng. Educ., vol. 29, no. 3, pp. 572–589, 2021.

29 A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, “Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression,” in 2017 International Conference on Information and Communication Technology Convergence (ICTC), 2017.

30 M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, “Sentiment analysis of tweets using machine learning approach,” in 2018 Eleventh International Conference on Contemporary Computing (IC3), 2018.

31 A. Rahman and M. S. Hossen, “Sentiment analysis on movie review data using machine learning approach,” in 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), 2019.

32 A. Goel, J. Gautam, and S. Kumar, “Real time sentiment analysis of tweets using Naive Bayes,” in 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), 2016.

33 S. Ray, “Learn Naive Bayes Algorithm,” Analytics Vidhya, 11-Sep-2017. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/. [Accessed: 08-May-2021].

34 R. N. B. Kajaree Das, “A Survey on Machine Learning: Concept, Algorithms and Applications,” A Survey on Machine Learning: Concept, Algorithms and Applications, vol. 5, no. 2, p. 1301, Feb. 2017.

35 V. N. Vapnik, The nature of statistical learning theory. New York, NY: Springer New York, 1995.

36 N. Mukhtar and M. A. Khan, “Urdu Sentiment Analysis using supervised machine learning approach,” Intern. J. Pattern Recognit. Artif. Intell., vol. 32, no. 02, p. 1851001, 2018.

37 N. Azzouza, K. Akli-Astouati, and R. Ibrahim, “TwitterBERT: Framework for twitter sentiment analysis based on pre-trained language model representations,” in Advances in Intelligent Systems and Computing, Cham: Springer International Publishing, 2020, pp. 428–437.

38 U. Rhohmawati, I. Slamet, and H. Pratiwi, “Sentiment analysis using maximum entropy on application reviews (study case: Shopee on Google play),” Jurnal Ilmiah Teknik Elektro Komputer dan Informatika, vol. 5, no. 1, 2019.

39 D. R. Kawade and D. K. S. Oza, “Sentiment analysis: Machine learning approach,” Int. J. Eng. Technol., vol. 9, no. 3, pp. 2183–2186, 2017.

40 M. Ahmad and S. Aftab, “Analyzing the performance of SVM for polarity detection with different datasets,” Int. j. mod. educ. comput. sci., vol. 9, no. 10, pp. 29–36, 2017.

41 A. Al-Hashedi et al., “Ensemble classifiers for Arabic sentiment analysis of social network (Twitter data) towards COVID-19-related conspiracy theories,” Appl. Comput. Intell. Soft Comput., vol. 2022, pp. 1–10, 2022.

42 D. Sudarsa, S. Kumar.P, and L. Jagajeevan Rao, “Sentiment analysis for social networks using machine learning techniques,” Int. j. eng. technol., vol. 7, no. 2.32, p. 473, 2018.

43 Soumya and Pramod, “Sentiment analysis of malayalam tweets using machine learning techniques,” ICT Express, 2020.

44 B. Agarwal and N. Mittal, “Machine learning approach for sentiment analysis,” in Socio-Affective Computing, Cham: Springer International Publishing, 2016, pp. 21–45.

45 J. Singh, G. Singh, and R. Singh, “Optimization of sentiment analysis using machine learning classifiers,” Hum.-centric comput. inf. sci., vol. 7, no. 1, 2017.

46 V. Uma Ramya and K. Thirupathi Rao, “Sentiment analysis of movie review using machine learning techniques,” Int. j. eng. technol., vol. 7, no. 2.7, p. 676, 2018.

47 B. Gaye, D. Zhang, and A. Wulamu, “A tweet sentiment classification approach using a hybrid stacked ensemble technique,” Information (Basel), vol. 12, no. 9, p. 374, 2021.

48 V. Tyagi, A. Kumar, and S. Das, “Sentiment analysis on twitter data using deep learning approach,” in 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020.