SHODHSAMHITA
शोधसंहिता
ISSN 2277-7067
UGC CARE Group 1

# SENTIMENT ANALYSIS IN SOCIAL MEDIA

**Geeta Bisht**

Research Scholar, Department of Science and Technology, Jayoti Vidyapeeth Women's University, Jaipur.

**Dr. Shobha Lal**

Professor of Mathematics and Computing, Department of Science and Technology, Jayoti Vidyapeeth Women's University, Jaipur, India.

**Abstract:** In this new era of social media, social networks are becoming more important sources of user-generated content on the internet. These kind of information resources, which represent a vast repository of human feelings, opinions, comments, and reviews, are regarded as strong informants for major industries, markets, and news, among other things. The importance of these platforms, together with the growing number of users creating material in reviews on social media sites. Support vector Machine (SVM) and Back-Propagation Neural Networks (BPNN), Nave Bayes, Decision Tree, and Random Forest are the five most automatic classification approaches used. When compared to other classifiers, the SVM classifier gets the greatest accuracy rate of 96.06 percent.

**Keywords:** Machine Learning, Twitter, Facebook, SVM, Naïve Bayes.

## 1. Introduction:

User-generated data on the Internet has grown exponentially in the Web 2.0 era. Facebook, Instagram, Twitter, LinkedIn, Amazon, Flip cart, and other commercial websites provide a platform for people to share their experiences, knowledge, and opinions on current events in politics, economics, and other global-critical issues Li et al. 2014 [1]. Embedding social intelligence from massive internet comments is a time-consuming task for any community or person. These issues prompted the development of Sentiment analysis, a social analytic method for automatically extracting, analysing, and summarizing user-generated data.

Sentiment analysis (SA) gathers online documents such as tweets, Facebook status updates, product reviews, blogs, and other social media platforms. Customers' attitudes, opinions, and emotions can be better understood by using online documents [2].

Machine learning (ML) is categorized as supervise learningit requires training data to be processed. SVM and the Nave Bayes model are the most widely used machine learning methods. Various machine learning models exist, but these are the most often used. When used to well-formed text corpora, Nave Bayes is successful [3], whereas support vector machine performs well on low-shape datasets.Nonetheless, ML approach performs badly on Facebook, where individuals write at random durations or make frequent spelling mistakes, & it requires a significant number of training samples to adjust the algorithm, since size and quality of output are determined by the amount of datasets [4,5]. Furthermore, machine learning analysis is time expensive, taking hours in a complicated machine learning model, especially if training is necessary [6]. With a smaller training dataset, the procedure is quicker, but the classification accuracy suffers [7].

## 2. Literature review

Sentiment analysis is a field of study that deals with a variety of complex issues. Sentiment analysis has been approached from a variety of perspectives. In general, these methods have used one of two types of machine learning techniques: supervised or unsupervised machine learning. SA approaches consume mostly remained used to Indo-European languages, particularly English.

Santidhanyaroj et al. [8] established a method that evaluates public sentiment of topics or acts influencing the general social conscience by analyzing social network data. Two methods were used: Nave Bayes and SVM classifiers. A pre-categorized data set was created from Twitter for the sentiment analysis model. In terms of analysis, SVM beat Nave Bayes classifier or produced more consistent and dependable findings.

Khasawneh et al. [9] suggested an emotional analysis for Arabic language utilizing two common social analysis tools: Social Mention or Senti-Strength. The data for this study came from Arabic-language posts on Facebook and Twitter. Tools were used to determine the separation of comments based on content of the postings. When comparing two tools' skill to estimate polarity, Nave Bayes found that the Social Mention tool had 91.83 percent accuracy and the Senti-Strength tool had 95.59 percent accuracy. Table 1 shows the summary of LOR.

Table: 1 Summarizes the LOR

| Author | Method | Application | Context |
|---|---|---|---|
| Yuliyanti, Djatna & Sukoco. (2017) | Lexicon-bases and machine learning | The program's level of success in community development | Twitter |
| Martin-Domingo, Martin, & Mandsberg. (2019) | Machine learning | Examine the quality of airport services. | Twitter account |
| Mansour. (2018) | Lexicon based | ISIS is viewed as a danger by the majority of users. | Twitter |
| Saragih & Girsang. (2017) | Lexicon based | Examine the performance of internet transportation as a company. | Facebook and twitter comments |
| Shayaa, Wai, Chung, Sulaiman, Jaafar & Zakaria. (2017) | Lexicon based | In terms of employment, there is a negative sentiment score. | Social media with several channels |
| Poecze, Ebster, Strauss & Christine (2018) | Machine learning | Improve brand communication or gain a better grasp of customer feedback. | YouTube Gamers' |

| | | | |
|---|---|---|---|
| | | | Facebook page |
| Ramanathan & Meyyappan (2019) | Lexicon based | Oman tourism reviews | Twitter |

3. **Sentiment analysis approaches**

### 3.1 Linear discriminant analysis (LDA)

Li et al. [10] discovered method for reducing the dimensionality of feature matrices using Principle Component Analysis (PCA). LDA looks for data with big variations all directions and then projects them to progressively decrease. Fisher LDA addresses the covariance matrix J's minimization (w).

$$J(w) = \frac{WT.SB\,(w)}{WT.SW(w)} \qquad (1)$$

SB stands for 'between groups scatter matrix' or 'within classes scatter matrix,' respectively.

Scatter matrices are defined as follows if x is the general mean of all cases, $U_c$is  mean of class c:

$$SB = \sum(U_C - x)(U_C - x)^T \qquad (2)$$

$$SW = \sum\sum(x_i - U_C)(x_i - U_C)^T \qquad (3)$$

### 3.2 Support vector analysis (SVM)

In 2015, Bhadane et al. [11] published a paper in Science Direct that used SVM with 0.78 precision to provide a novel technique to sentiment analysis and opinion mining. SVM stands for supervised learning with decision planes and decision bounds. These decision boundaries divide classes based on the membership of several objects. The SVM training algorithm considers or classifies the greed of a statement or tweet retrieved from Twitter as good, bad, or neutral. There are four SVM models to choose fromare used for classification, whereas the other two are used for regression SVM. These SVM approaches may be trained to reduce error function:

$$E = \frac{1}{2} W^T E + C \sum \beta_i \qquad (4)$$

C is a constant, while $\beta$ is a parameter. W is the coefficient of a vector. It uses a kernel function to translate input feature vectors into an image area wherever categories are linearly divisible. It employs a polynomial kernel whose size, coefficient, tolerance, and numFolds all have a significant impact.The coefficients of nouns, adjectivesadverbs, and verbs are obtained from prescribed values of their feelings to create vectors.

### 3.3 Back propagation neural network (BPNN)

Vindhoni [12] utilized BPN to classify emotions in online reviews.BPN is a type of adaptive learning technology that can classify feelings from social data. Inputs are useful to network for each training pattern. Neurons in the first layer's nodes are first trained with weights within a specific range. The hidden layer then considers various groupings of bigrams or tri-grams that exist in sentences. Weights stay modified as a result of the applications of these various combinations when they reach the output layer. Weight adjustments are computed using error function signals. The no. of neurons in the 3 layers fluctuates depending on the amount ofadjectives, nouns or verbs in a phrase.

### 3.4 Probalistic neural network (PNN)

PNN was used by Savchenko [13] to recognize discrete patterns from sets. PNN, the Gaussion Kernel function is employed neural network's hidden layer. The average operation of outputs for each review class was performed on the third layer. The final class belongingness is determined in the fourth layer by picking the biggest value of class.

### 3.5 Gaussian mixture model

In the Neurocomping Elsevier Journal, Abdel Fatteh [14] introduced "Multiple Classifiers for Sentiment Analysis."The query data points are assigned to multivariate normal components., GMMs are used to cluster data. Hard clustering is the process of assigning data points to clusters. Because it employs soft clustering approaches, GMM clustering has a lot of power. For each cluster, they provide the assigning of a score to each data point.

### 3.6 Naïve bayes (NB)

In 2013, Dr. Y. S Kumarswamy or J. Jotheeswaran published an article on opinion utilizing the NB classifier with data from the Manhattan Hierarchical Cluster in the journal of Theoretical or Applied Information Technology. TheNB classifier really a focused graph with nodes representing edges or variables representing conditional dependencies. NB is shown to be quite costly in text organization for SA.

### 3.7 Lexicon based opinion classifier (LEX)

KG. Nanda provided study in which they explored how to integrate lex and KNN classification algorithms with ME (maximum entropy). Polarity estimatesor product characteristics are identified with entity ranking of lexemes in LEX-based classifier. T. Christopher discovered that the accuracy of lex alone is 50.08 percent, but when it is combined with ME and KNN, it improves to 80.21 percent. Table 2 shows the approaches of SA.

Table: 2 Approaches of SA

| SA | Features | Advantages | Limitations |
|---|---|---|---|
| LDA | Matrix of scattering between and | Covariance minimization | Compositeor time- |

| | | | |
|---|---|---|---|
| | within classes | | consuming |
| Support vector machines (SVM) | Words or phrases that are negative, favorable, or neutral | Error function is minimized. | It's not a good idea to use non-separable terms. |
| Back propagation neural network (BPN) | With combinations, auto-updating opinion terms is possible. | Precision (0.78) | Iterative aspector slow. |
| Probabilistic neural network (PNN) | Occurrence of terms or components of speech | Adaptability to complicated texts | Pertinence to fresh data is limited. |
| Naıve Bayes (NB) | Frequency, greatest entropy | Capable of extracting aspectsthe context level | Only castoff specified purposes |
| K-Nearest neighbour (KNN) | Minkowski's distance from Manhattan | Accuracy (0.81) | Sensitive to the number of words in a sentence |

## 4. Level of analysis

The goal of sentiment analysis, as previously stated, is to "design automatic algorithms capable of extracting subjective information from texts in natural language." When doing sentiment analysis, the first step is to define what text (i.e., the examined object) signifies in the context of research.

In general, sentiment analysis in social networks may be studied at three levels, which are graphically depicted. Fig.1 shows the analysis of levels.

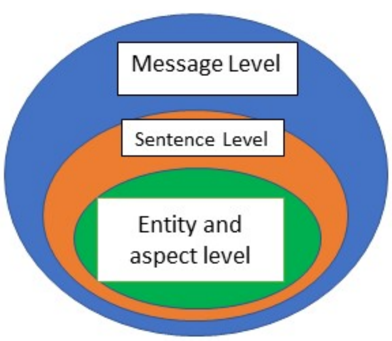


Fig.1. Analysis of levels

5. **Experimental analysis** The most common performance measures used in trials, like precision (P), F-measure (F) and recall (R). Precision (P) or recall (R) measurements are definite by: Given a trial set of reviews S1 labelled to polarity k, or prediction set S2 that labelled with polarity k by ML method,

$$R = \frac{S1 \cap S2}{S1} \quad (1)$$

$$P = \frac{S1 \cap S2}{S2} \quad (2)$$

similarly [15], utilizing F-measure, which combines recall or accuracy, to comparison different methods withsingle rate.F-measure, which we employed in our research, is defined as:

$$F - measure = \frac{2RP}{R+P} \quad (3)$$

The data is been collected from different social media like Facebook, YouTube, and twitter. Reviews of data can be positive, negative or neutral. Table 3 shows the results of the tests utilizing BPNN, Nave Bayes, SVM, random forestand Decision Tree classification algorithms to apply sentiment analysis to the suggested data set in terms of Recall, F-measure or Precision. Table 2 shows that SVM has greatest accuracy of 96.06 percent, followed by random forest with an accuracy of 95 percent orNBwith an accuracy of 85.82 percent. With an average exactness of 69.77 percent, BPNN was the least accurate. Furthermore, SVM has the best accuracy and recall. Fig.2 and Fig.3 shows the graphs of F-measure and precision with the different classifiers. The best result is shown by the SVM.

Table: 3 Results by different classifiers

| Classifier | Precision | Recall | F-measure |
|---|---|---|---|
| SVM | 95.8% | 96.4% | 96.06% |
| BPNN | 72.14% | 67.61% | 69.77% |
| NB | 92.62% | 85% | 87.60% |
| )ecision Tree | 83.67% | 86.72% | 84.23% |
| andom forest | 94.63% | 94.57% | 95.11% |

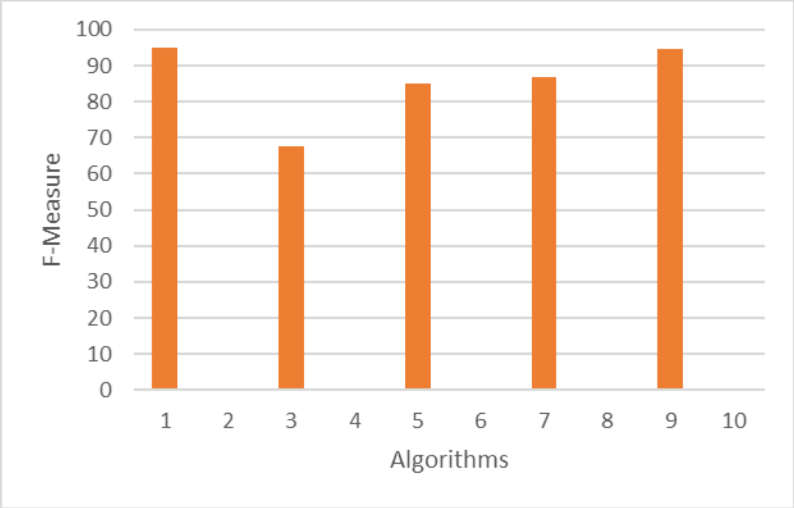


Fig.2. Graph of F-measure with different classifiers

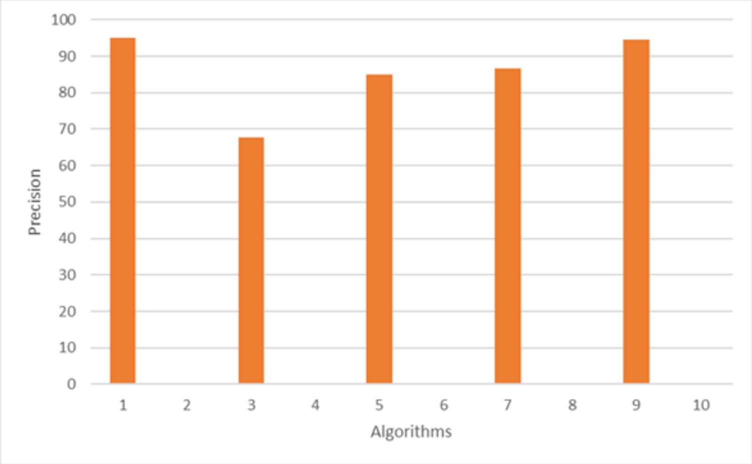


Fig.3. Graph of precision with different classifiers

## 6. Sentimental analysis applications

Sentiment analysis is used in a variety of fields, including business marketing, politics, health, and public policy. Sentiment analysis isn't only for one use; it can be used in a variety of ways to help people make better decisions. Sentiment analysis may be used to analyse world events such as a sporting event, a sporting activity, or a natural catastrophe [16, 17]. Some of the examples include a research that compared how individuals in western and eastern nations see ISIS. The outcome demonstrates how two parts of the world regard ISIS as a terrorist [18]. Sentiment analysis may also be used to raise awareness about data security or dangers of data breaches. Serves as a model for how businesses should respond to security breaches in order to influence public opinion [19]. In addition, social media sentiment research was done on employment sentiment score or unemployment rate[20].

Sentiment analysis is used in healthcare and where the research employs it. A methodology for SA as service is provided, which uses spatio-temporal features to detect disease epidemic areas [21]. Furthermore, sentiment analysis may be used to assess people's emotional needs during a crisis and plan a suitable rescue response [22]. Furthermore, evaluating emotions from text, SA may be used to determine a person's state of sadness [23].

SA may be used to forecast political elections, with results indicating that data studied from twitter more reliable as platform, with a 94 percent connection to polling data, and has the ability to challenge advanced polling methodologies [24].

Finally, consumer input is critical in the use of sentiment analysis, as it may help businesses and organizations take suitable action to enhance their products or services, as their business strategy.

In research it concludes social media users' perceptions and experiences with medicine and cosmetic products [25]. Sentiment analysis also enables for the detection of areas in which airport service quality may be improved, as the implementation of appropriate remedial steps, like paying attention to passenger criticism on social media [26]. SA may be used to study patterns or features of people's eating habits, which is beneficial to businesses when developing product and marketing strategies [27].

Sentiment analysis benefits business owners by allowing them to determinemeasure the efficacy and competence of corporate social media or brand communication [28] analyses their stock price movement through social media [29], as well as measure their popularity among consumers and they feel about their product or service.

## 7. Conclusion

Results of the performed systematic literature review include research on SA in social media.Following are contributions made by the paper. First,go through the approach for assessing social media sentiment. Although various approaches have been presented by researchers, SVM shows the better result among all. The data itself determines whether type of sentiment analysis is acceptable. Both strategies yielded comparable results in terms of accuracy. Structure of text, as the amount or time of data, must all be considered. If data structure is jumbled, there is a little amount of data, oronly have a short amount of time to analyse, the lexicon-based strategy is advised. Machine learning-based methods are better suited to larger data since they take more time and data to train. Combining lexical and machine learning methods to increase the quality and accuracy of the outcome is recommended.

## 8. References

**[1].** Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, *69*, 14-23.

**[2].** Das, S., & Chen, M. (2001, July). Yahoo! for Amazon: Extracting market sentiment from stock message boards. In *Proceedings of the Asia Pacific finance association annual conference (APFA)* (Vol. 35, p. 43).

**[3].** Hassan, A. U., Hussain, J., Hussain, M., Sadiq, M., & Lee, S. (2017, October). 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)* (pp. 138-140). IEEE.

**[4].** Mahtab, S. A., Islam, N., & Rahaman, M. M. (2018, September). Sentiment analysis on bangladesh cricket with support vector machine. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)* (pp. 1-4). IEEE.

**[5].** Chekima, K., & Alfred, R. (2017, November). Sentiment analysis of Malay social media text. In *International Conference on Computational Science and Technology* (pp. 205-219). Springer, Singapore.

**[6].** Dhaoui, C., Webster, C. M., & Tan, L. P. (2017). Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing*.

**[7].** Santidhanyaroj, P., Khan, T. A., Gelowitz, C. M., & Benedicenti, L. (2014, May). A sentiment analysis prototype system for social network data. In *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 1-5). IEEE.

**[8].** Khasawneh, R. T., Wahsheh, H. A., Al-Kabi, M. N., & Alsmadi, I. M. (2013, December). Sentiment analysis of arabic social media content: a comparative study. In *8th International Conference for Internet Technology and Secured Transactions (ICITST-2013)* (pp. 101-106). IEEE.

**[9].** Li Tao, Zhu Shenghuo, Ogihara Mitsunori (2008) Text categorizationviageneralized discriminant analysis. Inf Process Manage44:1684–1697

**[10].** Bhadane C, Dalal H, Doshi H (2015) Sentiment analysis: measuringopinions. Procedia Comput Sci 45:808–814. doi:10.1016/j.procs.2015.03.159.

**[11].** Vinodhini G, Chandrasekaran R (2016) A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. J King Saud Univ—Comput Inf Sci 28(1):2–12. doi:10.1016/j.jksuci.2014.03.024.

**[12].** Savchenko AV (2013) Probabilistic neural network with homogeneity testing in recognition of discrete patterns set. Neural Netw 46:227–241.

**[13].** Su Y, Zhang Y, Ji D, Wang Y, Wu H (2013) Ensemble learning for sentiment classification, Chinese lexical semantics. Springer, Berlin, pp 84–93

**[14].** Abdel Fattah M (2015) New term weighting schemes with combination of multiple classifiers for sentiment analysis. Neurocomputing 167:434–442. doi:10.1016/j.neucom.2015.04.051.

**[15].** Rijsbergen, C.J.V.: Information Retrieval: Butterworth-Heinemann (1979).

**[16].** Mahtab, S. A., Islam, N., & Rahaman, M. M. (2018, September). Sentiment analysis on bangladesh cricket with support vector machine. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)* (pp. 1-4). IEEE.

**[17].** Karamollaoğlu, H., Doğru, İ. A., Dörterler, M., Utku, A., & Yıldız, O. (2018, September). Sentiment analysis on Turkish social media shares through lexicon based approach. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)* (pp. 45-49). IEEE.

**[18].** Mansour, S. (2018). Social media analysis of user's responses to terrorism using sentiment analysis and text mining. *Procedia Computer Science*, *140*, 95-103.

**[19].** Hao, J., & Dai, H. (2016). Social media content and sentiment analysis on consumer security breaches. *Journal of Financial Crime*.

**[20].** Shayaa, S., Wai, P. S., Chung, Y. W., Sulaiman, A., Jaafar, N. I., & Zakaria, S. B. (2017). Social media sentiment analysis on employment in Malaysia. In *the Proceedings of 8th Global Business and Finance Research Conference, Taipei, Taiwan*.

**[21].** Ali, K., Dong, H., Bouguettaya, A., Erradi, A., & Hadjidj, R. (2017, June). Sentiment analysis as a service: a social media based sentiment analysis framework. In *2017 IEEE International Conference on Web Services (ICWS)* (pp. 660-667). IEEE.

**[22].** Ragini, J. R., Anand, P. R., & Bhaskar, V. (2018). Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management*, *42*, 13-24.

**[23].** Hassan, A. U., Hussain, J., Hussain, M., Sadiq, M., & Lee, S. (2017, October). 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)* (pp. 138-140). IEEE.

**[24].** Joyce, B., & Deng, J. (2017, November). Sentiment analysis of tweets for the 2016 US presidential election. In *2017 IEEE MIT Undergraduate Research Technology Conference (URTC)* (pp. 1-4). IEEE.

**[25].** Isah, H., Trundle, P., & Neagu, D. (2014, September). Social media analysis for product safety using text mining and sentiment analysis. In *2014 14th UK workshop on computational intelligence (UKCI)* (pp. 1-7). IEEE.

**[26].** Martin-Domingo, L., Martín, J. C., & Mandsberg, G. (2019). Social media as a resource for sentiment analysis of Airport Service Quality (ASQ). *Journal of Air Transport Management*, *78*, 106-115.

**[27].** Akter, S., & Aziz, M. T. (2016, September). Sentiment analysis on facebook group using lexicon based approach. In *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)* (pp. 1-4). IEEE.

**[28].** Poecze, F., Ebster, C., & Strauss, C. (2018). Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. *Procedia computer science*, *130*, 660-666.

**[29].** Suman, N., Gupta, P. K., & Sharma, P. (2017, December). Analysis of Stock Price Flow Based on Social Media Sentiments. In *2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)* (pp. 54-57). IEEE.