Sentiment Analysis Using Machine Learning: A Comparative Study

Neha Singh and Umesh Chandra Jaiswal

Department of ITCA, MMMUT, Gorakhpur, India, 273010.
nehitca@mmmut.ac.in, ucjitca@mmmut.ac.in

KEYWORDS	ABSTRACT
sentiment analysis; machine learning; social media; logistic regression	*In recent years, sentiment analysis on social media, including Facebook, Twitter and blogs, has grown in popularity. Social media generate large amounts of information, and this has contributed to the growth of sentiment analysis as a field of research. This study demonstrates that sentiment analysis has been thoroughly researched in previous years, and numerous methods have been designed and evaluated. Nevertheless, there is still much room for improvement. This paper reviews the state of art in sentiment analysis. Various machine learning procedures for sentiment analysis are discussed, their potential to increase the level of the analysis accuracy is underscored. This paper introduces sentiment analysis types, methodologies, applications, challenges, and a comparative study of machine learning and sentiment analysis approaches. Performance evaluation parameters, for sentiment analysis, have also been tested and compared using different machine learning classifiers. Performance evaluation points to logistic regression as the model that achieves the best result. In the future, a method that is easy, versatile, and practicable, should be offered as opposed to existing machine learning methods, and more work should be put into improving the algorithms' performance.*

1. Introduction

The popularity of rapidly expanding online social networks that rely on electronic media has encouraged new researchers to continue their work on sentiment analysis (SA). Internet-based specialist organizations survey information on websites, online gatherings, reviews, tweets, and item audits. Opinion research has applications in a variety of fields, including party planning, political decision-making, medical care monitoring, sales, and mindfulness administration (J. Singh et al., 2017). Sentiment analysis provides key insights into people's opinions, aiding decision-makers in making a choice on behalf of another user (Jagdale et al., 2019). There are numerous methods for sentiment analysis, including machine learning, dictionary-based, watchword-based, and idea-based methodologies (Qazi et al., 2017).

Sentiment analysis is a "suitcase" research problem that needs to be handled by numerous natural language processing subtasks, including angle extraction, subjectivity identification, named substance,

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A Comparative Study

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

1

and mockery place. The principal objective of sentiment analysis is to separate emotions from other elements (such as items or administrations) to produce helpful data. SA is done over an immense assortment of general suppositions, which could be of various sorts. Sentiment analysis includes deciding the evaluative idea of a piece of text. For instance, an item survey can communicate positive, negative, or nonpartisan sentiments. Over the previous decade, there has been a considerable development in the publication of micro-content on blogging websites, for example, Twitter, and social media overall (Raghuvanshi & Patil, 2016). It is utilised to explain unbiased assessments, negative or positive sentiments, and assessments. Sentiment investigation is defined by utilising methods such as content mining and knowledge theory, machine learning, coding, semantic methods, and natural language processing. By utilising these methodologies, strategies, procedures, and levels, we can classify unsorted information into categories such as news stories, web journals, item surveys, and so forth into positive, negative, or nonpartisan sentiments as per the opinion presented in them (Moralwar & Deshmukh, 2015).

Sentiment analysis presents a number of difficulties. The first is a term that is believed to be an opinion word. In certain cases, what is beneficial in one setting may be judged negative in another. A second issue is that most individuals are not aware of it and always convey their thoughts in the same manner (Vinodhini & Chandrasekaran, 2012). The automatic identification of sentiments would make the handling of a lot of such obstinate information conceivable (Tchalakova et al., 2011). Opinion mining is a multidisciplinary research field that includes many areas such as semantics, computational etymology, natural language processing, data recovery, artificial intelligence, and man-made brainpower. Sentiment analysis is likewise considered a grouping issue. Accordingly, machine learning procedures and dictionary-based sentiment techniques are utilised for opinion characterization (Aydo & Akcayol, 2016). Sentiment analysis manages a few specialised difficulties, for example, object distinguishing proof, assessment direction grouping, and element extraction. Generally, sentiment analysis can be performed utilising managed resources such as neural networks, Support Vector Machines, and Naive Bayes. Among these three strategies, SVM is viewed as the most appropriate for sentiment analysis. Opinion mining can be done in three phases: feature function, sentence function, and document function (Abirami & Gayathri, 2017).

Sentiment analysis begins with the preprocessing of review datasets and continues through sentiment categorization and opinion mining using different machine learning approaches or other dictionary-based methods (see Figure 1).
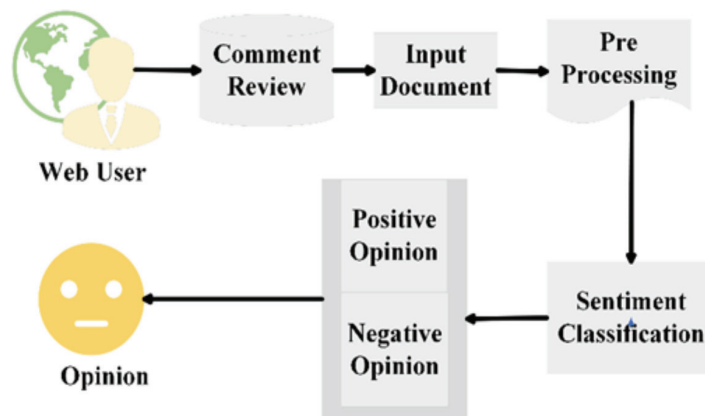

Figure 1. Sentiment Analysis Concepts (Abirami & Gayathri, 2017)

1.1. Sentiment Analysis Applications

Social media monitoring, customer support management and user review analysis are some of the most common applications of sentiment analysis (*8 Applications of Sentiment Analysis*, n.d.). Figure 2 shows sentiment analysis applications.

- *Social Media Observation:* It is specifically used in social networking examinations and learns how users of the website feel about a specific category (*Sentiment Analysis Guide*, n.d.).

- *Brand Observation*: Brand observation provides a plethora of information from online conversations about a given brand. Analyses of news stories, blogs, forums, and other sources provide insight into sentiment and, where applicable, many focus on specific locations.

- *Voice of Customer (VoC)*: All consumer input, including online, customer analysis, chats, single points, and emails, should be associated and calculated. Sentiment analysis helps to categorize and organise these records with the aim of identifying tendencies, recurring problems and concerns (*8 Applications of Sentiment Analysis*, n.d.).

- *Customer Service*: Examine customer service encounters to check that personnel are following proper procedures. Improve service efficiency so clients do not have to wait for help. Attrition rates should be reduced; after all, keeping clients is easier than finding new ones.

- *Market Research*: All types of market research and competition analysis are aided by sentiment analysis. Sentiment analysis can help in breaking into a new market, predict future trends, or gain a competitive advantage (*Sentiment Analysis Guide*, n.d.).

1.2. Challenges of Sentiment Analysis

Sentiment analysis addresses different issues. Sentiment analysis may be challenging in natural language processing because robots must be educated to evaluate opinions in a similar way to the human brain. Sentiment analysis technologies will be able to better solve these challenges as data science progresses (*Sentiment Analysis Challenges: Everything You Need to Know*, n.d.).
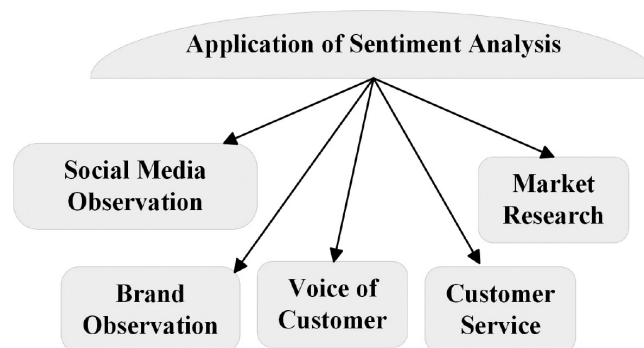
Figure 2. Applications of Sentiment Analysis

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A Comparative Study

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

3

- *Subjectivity and Tone*: These are the two kinds of contexts. Subjective contexts present opinions, while objective ones do not. For instance, the opinions of the subsequent two contexts could be analyzed and compared: The packaging is attractive, and red is the color of the box (*Sentiment Analysis Guide*, n.d.).

- *Context and Polarity*: All speech is created at several points in time, in diverse places, by and for a group of users. All contexts are given in the text. Analyzing sentiment without context gets quite difficult. Machines cannot learn from texts if they are not mentioned explicitly.

- *Irony and Sarcasm*: The public may convey their argumentative attitudes using irony and sarcasm, which are difficult for robots to recognise without a detailed interpretation of the service in which emotions are present.

- *Emojis:* Emojis are classified into two types. Modern emojis and eastern emojis. Emojis play a significant role in conveying emotion in messages, especially in tweets.

- *Defining Neutral*: Another issue to overcome in order to undertake effective sentiment analysis is explaining the meaning of neutral. One of the most crucial components of any classification challenge is identifying types, which is done by defining the neutral tag.

- *Human Annotator Accuracy:* Even for humans, sentiment analysis is a difficult task. On average, inter-annotator agreement is pretty low when it comes to sentiment analysis (*Sentiment Analysis Guide*, n.d.).

1.3. Contributions

The following are potential outcomes from a comparison study on sentiment analysis with machine learning:

- *Evaluation of different algorithms:* The research contrasts different machine learning techniques, including Naïve Bayes, multilayer perceptron, and logistic regression. Researchers can identify which algorithms are best suited for certain sorts of data or domains by analysing how well they perform on sentiment analysis tasks.

- *Comparative performance analysis:* The study offers a quantitative comparison of how well various models perform in terms of different parameter measures. Researchers may determine which models excel in sentiment analysis tasks by comparing these metrics, and they can also gain knowledge about the models' advantages and disadvantages.

- *Recommendations and guidelines:* The study's recommendations and rules for choosing the best machine learning methods for sentiment analysis work can be based on comparative analysis. Making wise selections when creating sentiment analysis systems can be aided by this for practitioners and scholars.

Overall, a comparison of sentiment analysis methods utilising machine learning sheds light on the merits of various strategies, advancing the field, and enhancing the precision and dependability of sentiment analysis systems.

The review paper consists of a series of sections, namely: Section 2 presents a literature review. Section 3 defines the types of sentiment analysis. Section 4 defines approaches to sentiment analysis and presents a comparison of sentiment analysis and machine learning classification. Section 5 evaluates the performance of sentiment analysis, and tests and compares performance evaluation parameters. Section 6 concludes the entire paper.

2. Literature Review

Sentiment analysis is an interesting range of research in the present era of business promotion. Different kinds of research have been conducted by various researchers, using the lexicon-based technique, the rule-based technique, the hybrid technique, and many more.

Mitra used sentiment analysis in various companies for item analysis and social media opinions and a small part of the study was dedicated to determining whether a text was subjective or objective (2020). The authors proposed rule-based methods in this experiment, which expressed a set of rules and programs. Opinion mining was accomplished using a lexicon technique, machine learning methods, and a hybrid method. The authors used Python NLTK features. Brar and Sharma defined a feature-based sentiment mining and supervised learning tool to examine opinions in a movie survey (2018). The important goal of this analysis is to reduce the polarity of the survey by employing nouns, adjectives, and verbs as sentiment texts. This document analysis also provided some data on information categorization based on polarity. Smetanin and Komarov used convolutional neural networks and proposed an approach for the thorough examination of product reviews in Russian (2019). As neural network inputs, the authors used Word2Vec pre-trained vectors. In this case, the training data file was built from user-ranked scores on top- ranked commodities on a prominent e-commerce site. In this experiment, the archive accuracy F-measure was 75.45%. Pujari et al., developed a new framework using Python's built-in packages to mine all people's thoughts about an item and aggregate them according to their opinions, assisting potential purchasers in forming a capitalised record of the product (2018). In this study, the performance of three different classification methods was examined. Goularas and Kamis used CNN and RNN; the authors compared Word2Vec and GloVe models as well as other word embedding systems (2019). To evaluate those strategies, they used information from the international workshop on semantic calculation. This research contributed to the area of sentiment analysis by comparing the performance, merits, and demerits of the various methodologies using a single testing method. El Rahman et al., demonstrated a technique for sentiment analysis in Twitter reviews (2019). It is critical to analyse Twitter reviews since they are so unstructured. While earlier research has focused on supervised and unsupervised machine learning techniques, the authors' proposed approach is a combination of both. They used a large range of machine learning methods. Cross-validation and f-scores were employed to evaluate these models' performance. When it comes to mining Twitter for text, the model showed superiority over other models in terms of performance. Batanović et al., proposed a data file balancing technique that reduced pattern selection bias by removing organised disparities among opinion classes that are not important (2016). The authors demonstrated its superiority to random sampling, and they utilised it to generate the SerbMR data file, the first balanced and topically uniform sentiment analysis data file. They presented an incremental method for determining the best merger of simple content processing options and tool learning methods. Shivaprasad and Shetty presented the taxonomy of several sentiment analysis approaches (2017). In comparison to Naive Bayes and maximum entropy approaches, this paper indicated that support vector machines provide greater accuracy. It has been discovered that sentiment analysis plays a crucial role in deciding on a product or service. Many advanced methods are described in the literature study that characterized sentiment analysis in terms of various elements. Mishra et al., analyzed Modi Ji's Digital India Campaign comments on Twitter (2016). According to the identified sentiments the authors have classified these opinions and their polarity. Options included positive, negative, and neutral. To acquire Twitter data for analysis, the Twitter API was used. The most popular technique for sentiment analysis was the dictionary-based approach, which the authors employed to analyse data from a variety of people. Finally, the data was

categorised according to polarity. Samaras et al., aimed to determine how well Twitter messages can be used to gauge emotional reactions in connection to instances that are occurring in real time (2023). Y\ild\ir\im et al., examined the attitudes people had towards distant learning during the pandemic (2023). Mendi presented an LSTM architecture combined with a blockchain layer (2022). The suggested approach was used on a Twitter dataset that included negative, neutral, and positive emotions. P. Singh identified six key data-centric AI characteristics that are currently being used by academics to raise the bar for AI systems (2023). Additionally, the authors emphasised the current data-centric methods used by ML practitioners.

The review of the above-described surveys outlines the numerous sentiment classification techniques, algorithms, and enhancement techniques. Sentiment analysis is used to obtain meaningful information from a large amount of data. In the survey carried out in this paper, the researcher evaluated mean absolute error using various algorithms such as Naive Bayes, k-nearest neighbor, random forest, support vector machine, and regression etc. The surveys mentioned above also provide a description of user reviews, accuracy labels, a comparison of different sentiment analysis techniques, and a generalisation of the suggested approach to address various problem categories.

3. Types of Sentiment Analyses

Models of sentiment analyses center around extremity (positive, negative, or neutral), but they also extend beyond polarity to identify certain moods and emotions (furious, upbeat, miserable), earnestness (pressing, not dire), and even expectations (intrigued vs. not intrigued) (*Sentiment Analysis Guide*, n.d.). Figure 3 shows different types of sentiment analysis models.

- *Fine-grained sentiment analysis:* This study can help comprehend the input received from a client. In terms of the polarity of the input, we can achieve exact results. If extremity accuracy is basic to the owner of a business, adding the accompanying extremity classes should be considered: very positive, positive, neutral, negative, and very negative. Fine-grained sentiment analysis is the term for this sort of investigation.

- *Emotion detection:* Sentiment investigation of this sort tries to distinguish sentiments such as satisfaction, disappointment, fury, pity, etc. Many emotion recognition frameworks rely on vocabulary (word arrangements and the emotions they evoke) or refined machine learning strategies.

Figure 3. Types of Sentiment Analyses

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A Comparative Study

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND
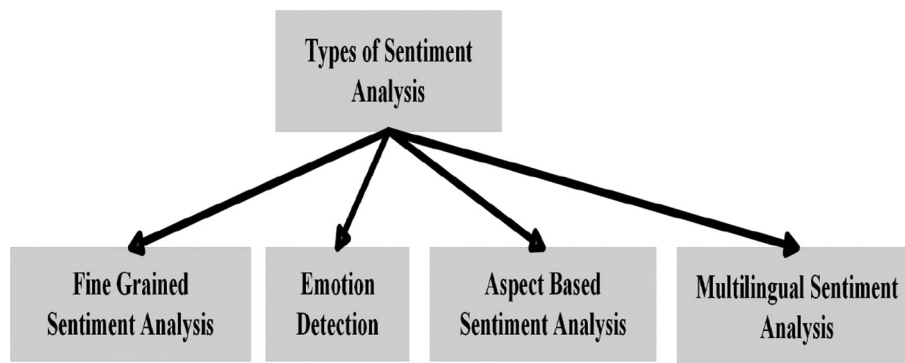
6

- *Aspect Based Sentiment Analysis:* While assessing opinions in messages, such as item reviews, we ordinarily need to find which explicit parts or individual elements are being referred to in a positive, neutral, or negative light. A classifier which is primarily perspective-based may be able to affirm that the following sentence suggests a poor opinion regarding battery life: "The battery life of this camera is excessively short".
- *Multilingual Sentiment Analysis:* Sentiment analysis may be done in different dialects. It requires a lot of preprocessing and assets. Most of these resources are available on the web (for example, opinion dictionaries), while others should be created (for example, interpreted corpora or clamoring strategies), but the user must know how to code to use them (*Sentiment Analysis Guide*, n.d.).

4. Methodologies Used for Sentiment Analysis

Sentiment analysis methodologies are categorized into three types: machine learning, rule-based, and lexicon-based approaches. Figure 4 shows various sentiment analysis approaches.

4.1. Machine Learning Approach

It is a part of artificial thinking that enables machines to work, and what is more, improve, without being unequivocally adjusted. For opinion mining, AI estimations have been comprehensively used. Machine learning methodologies have made data crunching faster. The computation adjusts naturally to the given data. This collaboration is known as the planning time of the estimation, followed by a testing stage. Here some test data is added to the estimation to get the expected result. It is requested by the machine learning-based approach using gathering methodologies (Srivastava et al., 2020). Machine learning utilizes different techniques, namely:
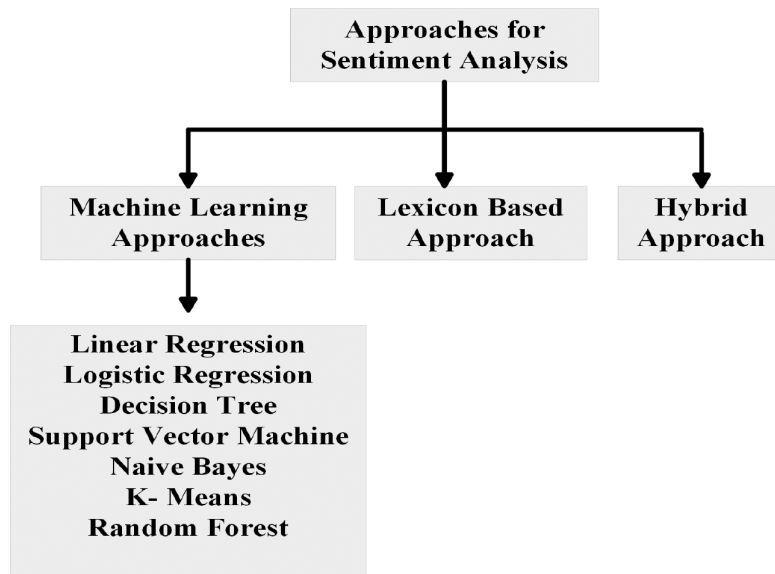
Figure 4. Sentiment Analysis Approaches

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A Comparative Study

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

7

- *Linear Regression:* It is used to forecast unique data by utilising persistent factors. The relapse line is the best fit line and is communicated by the direct condition Z = X + Y + C.
- *Logistic Regression*: From a series of free factors, strategic relapse is utilised to anticipate different qualities, for example, double worth, such as 1 or 0, yes or no, and valid or invalid (*Top 10 Machine Learning Algorithms With Python & R Code*, n.d.-a).
- *Decision Tree*: It is employed to handle problems with order. It works for each kind of variable: express and non-stop established variables.
- *Support Vector Machine*: It is based on a sort of grouping function. In this, each given value is plotted as a given mark on an n-dimensional space vector.
- *Naive Bayes*: This model is particularly helpful for enormous informational collections.
- *K-Means*: It tackles the issue of grouping. Its interaction is the fundamental and straightforward standard for describing a given informational collection.
- *Random Forest*: In this, there are a huge number of decision trees, which are also called "forests" in a random forest. Every tree assigns a distinct division to a current object based on its attributes (*Top 10 Machine Learning Algorithms With Python & R Code*, n.d.-a).

4.2. Rule-Based Method

The rule-based technique is used to obtain an opinion by creating numerous rules that are formed by tokenizing each text. In every document, the presence of each token, or phrase, is checked. If the term exists and has a favourable connotation, it receives a +1 rating for sentiment. Every post starts with a zero-point neutral point and is deemed positive. If the total point was less than zero, then the polarity rate was bigger than zero. Then, it verifies the output or inquires whether the output is accurate or not using a rule-based approach (Devika et al., 2016).

4.3. Lexical Based Approach

Lexical assembly tasks based on the theory that the total farthest point of a context or file is the measure of the polarities of the specific articulations or contexts. In each area, word references were refreshed with examination expressions of the appropriate preparation assortment that had the most noteworthy weight, as determined by the relevance frequency strategy (Devika et al., 2016). Table 1 shows a comparison of various sentiment analysis approaches.

Table 2 shows comparative analysis of various types of machine learning classifications.

Table 1. Comparison of Sentiment Analysis Approaches (Devika et al., 2016)

Approach	Advantages	Disadvantages
Machine Learning Approach	In this approach, vocabulary, is not important and its classification is highly precise.	The classifier mostly reads text in a single domain.
Rule Based Approach	When compared to the word function, the sentence function performs better in terms of opinion classification.	The defining law is used to calculate accuracy.
Lexicon Based Approach	Labelled information and a learning method are not needed.	It requires a lot of language proficiency resources, which are not always available.

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A Comparative Study

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

8

Table 2. Comparative Analysis of Machine Learning Classifications

Machine Learning Classification	Merits	Demerits
Linear Regression	It has less complexity.	It does not give a detailed explanation of the relationship between variables.
Logistic Regression	It classifies unidentified data very quickly.	It may lead to overfitting.
Decision Tree	It requires less effort for data preparation during pre-processing than other methods.	A minor change in a value can cause a huge change in the format.
Support Vector Machine	In high-dimensional spaces, it is more effective and works well.	It is not effective for huge datasets.
Naïve Bayes	It is convenient for solving multi-class feature issues and also works faster.	Its probability outputs are sometimes taken seriously because its estimations can be wrong in some instances.
K- Means	It gives the best computational cost and also enhances accuracy.	It cannot handle noisy data.
Random Forest	It is a robust model that gives accurate results and is also used for both regression and classification.	It is more difficult to use than the other models.

5. Performance Evaluation in Sentiment Analysis

The performance of sentiment analysis can be measured using four types of indexes generated using the various equations: accuracy, precision, recall, and f1-score (Yang & Chen, 2017).

$$Accuracy = \frac{TP+TN}{P+N} \qquad (1)$$

$$= \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

$$Recall = \frac{TP}{TP+FP} = 1 - \frac{FN}{T} \qquad (3)$$

$$F1 = \frac{Precision * Recall}{Precision + Recall} \qquad (4)$$

Where TP = true positives, FP = false positives, TN = true negative, and FN = false negative.

Furthermore, relative error and relative error percentages are used to calculate the value of an error or error rate. Any classification, grouping, or prediction procedure can have these flaws:

$$Relative\ Error = \frac{|T - P|}{TP} \qquad (5)$$

$$Relative\ Error\ Percentage = \frac{|P - TP|}{TP * e} \quad (6)$$

where P is the predicated value.

5.1. Performance Evaluation Parameters Testing

We performed tests on the performance evaluation parameter for sentiment analysis using various machine learning classifiers. Figure 5 shows the proposed model.

In this proposed model, the first collected dataset has been taken from Kaggel [kaggle datasets download-d spscientist/students-performance-in-exams]. In this experiment, the student performance in exams 1000 dataset was used. As a result, the data must first be preprocessed before the model can be applied, and the sentiment category must then be classified using the students' grades. The sentiment categorization criteria for the datasets are shown in Table 3.

In this dataset, the sentiment category -excellent- belongs to 81-100, -very good- is at 71-80, -good- is at 51-70, and -average- ranges from 31-50, and while -fail- is at 0-30. Furthermore, we applied machine learning classification algorithms for the analysis of the mentioned performance parameter. Naïve Bayes (NB), multilayer perceptron (MP), and logistic regression (LR) algorithms were utilized in the study to build classification models. Finally, evaluation of metric performance involved the observation of the results of the experiments on the dataset.
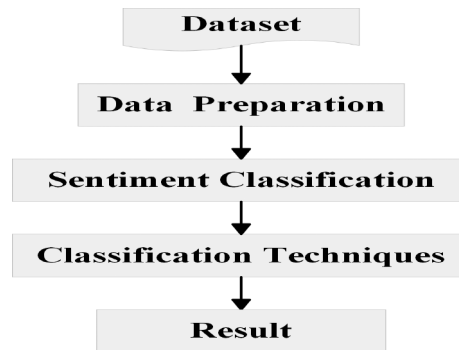
Figure 5. Proposed Model

Table 3. Dataset classification

Sentiment	Grade Classification
Excellent	81-99
Very Good	71-80
Good	51-70
Average	31-50
Fail	1-30

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A
Comparative Study

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

10

5.1.1. Experimental Results

The findings of the proposed model are discussed in this section, as well as a comparison of the system's performance measurement with that of other classifiers.

Figure 6 shows the classification of sentiments into categories. In this classification of sentiment categories, the good sentiment category is 459 students, the very good sentiment category is 215 students, the excellent category is 176 students, the average category is 134 students, and the fail category is 16 students. Table 4 shows the proposed model results for several categorization classifiers. The performance of the classifier is examined using six parameters.

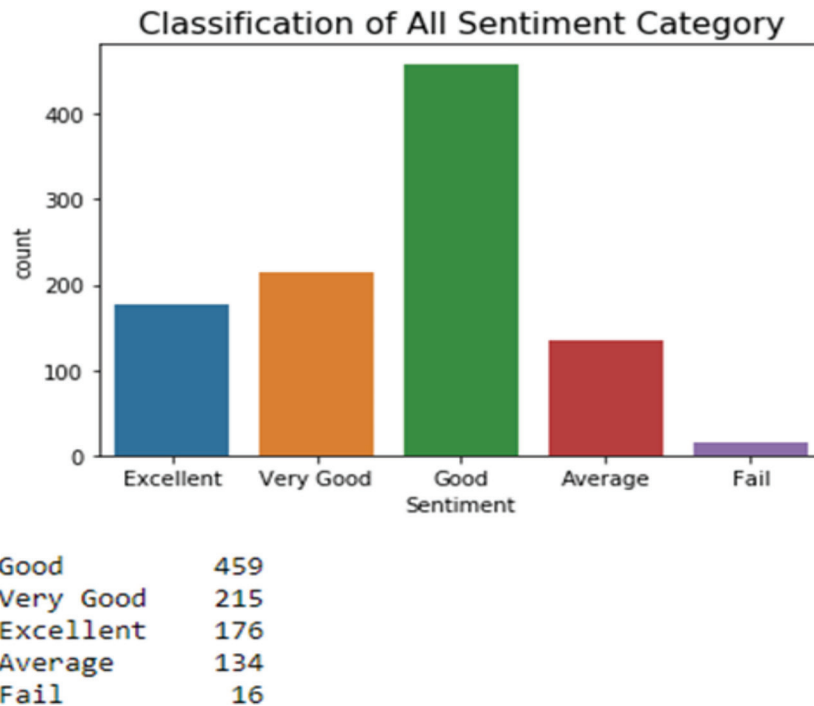- Note: In the above table, the values in bold indicate the best result in all parameters.


```
Good        459
Very Good   215
Excellent   176
Average     134
Fail         16
```

Figure 6. Display classification of the whole dataset into sentiment categories

Table 4. Overall performance parameter comparison of proposed model

Parameters	Naïve Bayes	Multilayer Perceptron	Logistic Regression
Accuracy	89.4	98	**99.9**
Precision	0.898	0.981	**0.999**
Recall	0.894	0.980	**0.999**
F1	0.895	0.976	**0.999**
Relative Error	20.2624	4.077	**0.183**
Relative Error %	46.9476	16.8462	**5.5536**

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A Comparative Study

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

11

5.1.2. Statistical Results

- In terms of accuracy, logistic regression performance shows a 10.446% increment with Naïve Bayes while multilayer perceptron achieves an accuracy of 1.9%.
- In terms of precision, logistic regression performance provides an increment of 0.101%, which is better than Naïve Bayes. A precision of 0.018% was achieved with multilayer perceptron.
- In the case of recall, logistic regression performance was 0.105% better than Naïve Bayes. In comparison, multilayer perceptron achieved 0.019%.
- In the case of F1, logistic regression performance was better by 0.104% than that of Naïve Bayes and provided a 0.023% increment with multilayer perceptron.
- In the case of relative error, logistic regression performance provided an increment of 20.0794% with Naïve Bayes and 3.894% with multilayer perceptron.
- In terms of relative error percentage, logistic regression performance showed a 41.394% better result with Naïve Bayes as well as an 11.29266% increment with multilayer perceptron.

Figure 7 provides an analysis of student performance in the exam dataset. In the overall comparison of performance parameters of classifiers, logistic regression gives the best result when compared to Naïve Bayes and multilayer perceptron.

Figure 7. Comparison of results for selected parameters

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A Comparative Study

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

12

6. Conclusion

Sentiment analysis is an emerging area of research. Sentiment analysis uses information links such as micro-blogs, forums, and news etc. The data contained in these links plays a vital part in presenting a user's emotions and opinions about a specific problem. This paper has provided details of previous work and has explained the types, applications, challenges, and comparisons of machine learning and sentiment analysis approaches. This paper has also tested and compared performance evaluation parameters on sentiment analysis using various machine learning classifiers. In the performance parameter comparison, logistic regression has given the best result when compared to Naïve Bayes and multilayer perceptron. Sentim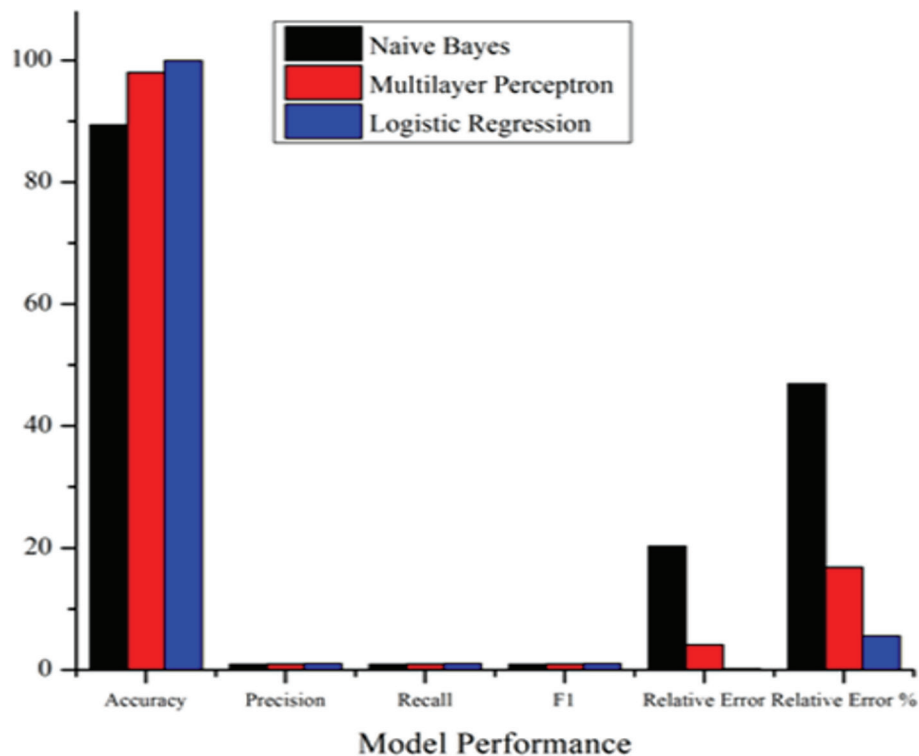ent analysis has become more challenging due to the extensive use of deep learning. In this survey, we find that recent methods are still unable to work well in different domains as well as handle difficult problems, which is also a challenging issue. Therefore, there is a need to further explore and test different domains and datasets with state-of-the-art emerging techniques to increase effectiveness and usability of produced results in real-life applications.

7. References

8 Applications of Sentiment Analysis. (n.d.). Retrieved August 1, 2023, from https://monkeylearn.com/blog/sentiment-analysis-applications/

Abirami, A. M., & Gayathri, V. (2017). A survey on sentiment analysis methods and approach. *2016 Eighth International Conference on Advanced Computing (ICoAC)*, 72–76. https://doi.org/10.1109/ICoAC.2017.7951748

Aydo, E., & Akcayol, M. A. (2016). *Aydogan2016.Pdf*.

Batanović, V., Nikolić, B., & Milosavljević, M. (2016). Reliable baselines for sentiment analysis in resource-limited languages: The serbian movie review dataset. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2688–2696.

Brar, G. S., & Sharma, A. (2018). Sentiment analysis of movie review using supervised machine learning techniques. *International Journal of Applied Engineering Research*, *13*(16), 12788–12791.

Devika, M. D., & Sunitha Cªand Ganesh, A. (2016). Sentiment analysis: a comparative study on different approaches. *Procedia Computer Science*, *87*, 44–49. https://doi.org/10.1016/j.procs.2016.05.124

El Rahman, S. A., AlOtaibi, F. A., & AlShehri, W. A. (2019). Sentiment analysis of twitter data. *2019 International Conference on Computer and Information Sciences (ICCIS)*, 1–4. https://doi.org/10.1109/ICCISci.2019.8716464

Goularas, D., & Kamis, S. (2019). Evaluation of deep learning techniques in sentiment analysis from twitter data. *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, 12–17. https://doi.org/10.1109/Deep-ML.2019.00011

Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. (2019). Sentiment analysis on product reviews using machine learning techniques. *Advances in Intelligent Systems and Computing*, *768*, 639–647. https://doi.org/10.1007/978-981-13-0617-4_61 https://doi.org/10.1007/978-981-13-0617-4_61

Mendi, A. F. (2022). A Sentiment Analysis Method Based on a Blockchain-Supported Long Short-Term Memory Deep Network. *Sensors*, *22*(12), 4419. https://doi.org/10.3390/s22124419

Mishra, P., Rajnish, R., & Kumar, P. (2016). Sentiment analysis of Twitter data: Case study on digital India. *2016 International Conference on Information Technology (InCITe)-The Next Generation IT*

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A
Comparative Study

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

13

Summit on the Theme-Internet of Things: Connect Your Worlds, 148–153. https://doi.org/10.1109/INCITE.2016.7857607

Mitra, A. (2020). Sentiment analysis using machine learning approaches (Lexicon based on movie review dataset). *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, *2*(03), 145–152. https://doi.org/10.36548/jucct.2020.3.004

Moralwar, S. B., & Deshmukh, S. N. (2015). Different approaches of sentiment analysis. *Int. J. Comput. Sci. Eng*, *3*(3), 160–165.

Pujari, C., Aiswarya, & Shetty, N. P. (2018). Comparison of classification techniques for feature oriented sentiment analysis of product review data. *Data Engineering and Intelligent Computing: Proceedings of IC3T 2016*, 149–158. https://doi.org/10.1007/978-981-10-3223-3_14

Qazi, A., Raj, R. G., Hardaker, G., & Standing, C. (2017). A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges. *Internet Research*. https://doi.org/10.1108/IntR-04-2016-0086

Raghuvanshi, N., & Patil, J. M. (2016). A brief review on sentiment analysis. *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2827–2831. https://doi.org/10.1109/ICEEOT.2016.7755213

Samaras, L., García-Barriocanal, E., & Sicilia, M.-A. (2023). Sentiment analysis of COVID-19 cases in Greece using Twitter data. *Expert Systems with Applications*, 120577. https://doi.org/10.1016/j.eswa.2023.120577

Sentiment Analysis Challenges: Everything You Need to Know. (n.d.). Retrieved August 1, 2023, from https://www.repustate.com/blog/sentiment-analysis-challenges-with-solutions/

Sentiment Analysis Guide. (n.d.). Retrieved August 1, 2023, from https://monkeylearn.com/sentiment-analysis/

Shivaprasad, T. K., & Shetty, J. (2017). Sentiment analysis of product reviews: a review. *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 298–301. https://doi.org/10.1109/ICICCT.2017.7975207

Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Human-Centric Computing and Information Sciences*, *7*, 1–12. https://doi.org/10.1186/s13673-017-0116-3

Singh, P. (2023). Systematic review of data-centric approaches in artificial intelligence and machine learning. *Data Science and Management*. https://doi.org/10.1016/j.dsm.2023.06.001

Smetanin, S., & Komarov, M. (2019). Sentiment analysis of product reviews in Russian using convolutional neural networks. *2019 IEEE 21st Conference on Business Informatics (CBI)*, *1*, 482–486. https://doi.org/10.1109/CBI.2019.00062

Srivastava, S., Nagpal, A., & Bagwari, A. (2020). Various Approaches in Sentiment Analysis. *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, 92–96. https://doi.org/10.1109/CICN49253.2020.9242618

Tchalakova, M., Gerdemann, D., & Meurers, D. (2011). Automatic sentiment classification of product reviews using maximal phrases based analysis. *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, 111–117.

Top 10 Machine Learning Algorithms With Python & R Code. (n.d.-a). Retrieved August 1, 2023, from https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A Comparative Study

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

14

Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. *International Journal*, *2*(6), 282–292.

Y\ild\ir\im, E., Yazgan, H., Özbek, O., Günay, A. C., Kocaç\inar, B., \cSengel, Ö., & Akbulut, F. P. (2023). Sentiment Analysis of Tweets on Online Education during COVID-19. *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 240–251. https://doi.org/10.1007/978-3-031-34111-3_21

Yang, P., & Chen, Y. (2017). A survey on sentiment analysis by using machine learning methods. *2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 117–121. https://doi.org/10.1109/ITNEC.2017.8284920

Neha Singh and Umesh Chandra Jaiswal

Sentiment Analysis Using Machine Learning: A Comparative Study

ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e26785
eISSN: 2255-2863 - https://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY-NC-ND

15