STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 11, Issue 2, 2024
May–August
DOI Journal: 10.37591/JoAIRA

https://journals.stmjournals.com/joaira

Research JoAIRA

Evaluation of Machine Learning Classifiers for Sentiment Analysis

Navdeep Bohra, Ashish Kumari*

Abstract

Sentiment in social media refers to users' emotions and opinions through their posts and interactions. Sentiment analysis (SA) refers to relating and classifying the sentiments expressed as engagement and interactions between users. When analyzed, tweets frequently produce a large source of clustered data. These data help determine people's opinions about a variety of motifs. Thus, this study presents an Automated Machine Learning (ML) Sentiment Analysis Model to detect media sentiment. Due to the inclusion of important data and non-useful characters (often called noise), applying models to them gets tricky. This study uses machine literacy to analyze Twitter sentiment analysis, the sentiment of tweets delivered from the Kaggle Twitter Sentiment Analysis dataset by creating a machine learning channel that employs three different models or categories, i.e., Logistic Regression (LR), Bernoulli Naive Bayes (Bernoulli NB), and Support Vector Machine (SVM) on Python. The accuracy and F1 Scores define to measure how well these classifiers function. In the comparison of the models through the accuracy outcome, Logistic Regression gives the most efficient results with an accuracy of 95.83% followed by SVM at 95.77%, it, in turn, performs better than Bernoulli Naive Bayes at 94.90% accuracy.

Keywords: Sentiment analysis, tweets, Bernoulli Naïve Bayes, logistic regression, support vector machine

INTRODUCTION

Nowadays, people express their viewpoints online using various platforms such as social media, blogs, forums, and websites that offer product reviews. Many people follow social media channels like Instagram, Facebook, Twitter, and Google Plus to express their thoughts, emotions, and everyday experiences. Online forums provide customers with the opportunity to inform and influence others. Social media platforms such as Twitter, used for profile updates, personal posts, comments, and views generate vast amounts of sentiment-rich data that businesses can also utilize for advertising purposes. The reason being the overwhelming count of follower-generated content, sentiment analysis techniques are commonly employed to automate the analysis process. Sentiment analysis involves Natural Language Processing, which automatically mines perspectives, attitudes, and feelings from various sources like tweets, text, audio, and databases. The sentiments expressed by users can range from positive and neutral to negative and can be analyzed based on user engagement and interactions. Perspective, emotion, review, and hypothesis are often used interchangeably, but they have distinct disagreements. Subjectivity analysis, opinion mining, and assessment extraction are the main approaches to sentiment analysis [1].

- *Perspective:* A disputed conclusion (because opinions vary from person to person).
- *Sentiment:* Feeling-based opinion.
- *View:* Personal opinion.
- *Belief:* Conscious agreement.

***Author for Correspondence**
Ashish Kumari
E-mail: simplyaashi@gmail.com

Assistant Professor, Department of Computer Science and Engineering, Maharaja Surajmal Institute of Technology, New Delhi, India

Received Date: May 24, 2024
Accepted Date: July 06, 2024
Published Date: July 10, 2024

Citation: Navdeep Bohra, Ashish Kumari. Evaluation of Machine Learning Classifiers for Sentiment Analysis. Journal of Artificial Intelligence Research & Advances. 2024; 11(2): 141–154p.

Several techniques are available for sentiment analysis, such as data mining and the latest intelligence techniques. This approach has diverse applications in areas like marketing, consumer service, and public opinion surveys. Through the survey of customer sentiments, businesses can extract features into their opinions and emotions regarding new items or services of item. This information can be referred to monitor brand reputation, get the idea for improvement in particular areas, and enhance customer satisfaction. Online communities offer a platform for users to express their views and convince people using forums, tweets, status updates, blog posts, comments, and reviews, thereby producing a vast information which shows full of sentiment. All web contents are divided into categories which include document, sentence, and aspect-level sentiments, as shown in Figure 1.

Document Level

At the document level, the polarity is determined by analyzing the document as a whole. Through or with the assistance of this level, we can determine whether the opinions or feelings that are accessible to us give us a favorable sentiment or a bad sentiment. This level can get most aspects of a certain feature, but it does not get people's likes and dislikes.

Sentence Level

Every statement undergoes processing and analysis to determine its polarity, resulting in a sentiment assessment that categorizes it as positive, negative, or neutral. Identifying subjective sentences is now a practical possibility thanks to machine learning.

Feature or Aspect Level

On the other hand, sentiment analysis has some limitations when applied at the sentence level. It is also known as entity level in which the results is obtained. The aspect level is the sort of SA that goes into the highest depth, which expresses the outcome as opinion. The desired value, plus either a positive or negative outcome, are the two possible possibilities. Target opinion helps recognize the importance of this level to discover people's feelings regarding entities and the attributes such entities possess. At this level, tasks such as surveys, feedback, criticisms, and complaints, among others, are carried out [1, 2].

According to Kharde and Sonawane, public reviews have a significant impact on the opinions and decisions of 87% of internet users [1]. "Opinion Shaping" or "Emotional Machine Intelligence", as a sentiment analysis is a subfield of NLP, that has been extensively studied in previous research, particularly on social media. Identifying whether an instance of information is generally positive, negative, or neutral is a common task in the study of sentiment. This can be achieved through various methods, including supervised and unsupervised machine learning and lexicon methods based on dictionaries that match words with predetermined emotions.

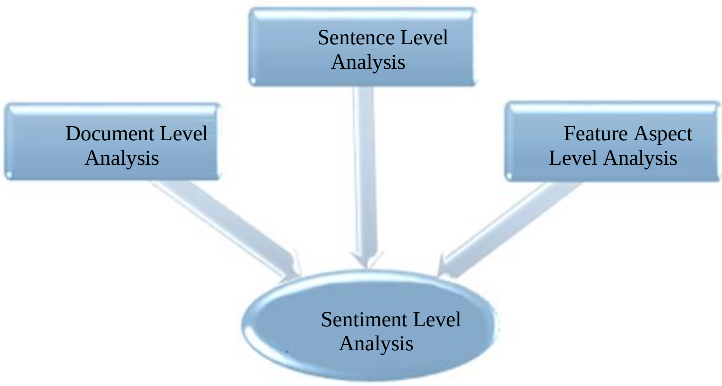


Figure 1. Types of sentiment analysis on which machine learning works.

Another common task is identifying the specific topic or aspect of a text that is being discussed. In addition, researchers have also explored issues related to handling sarcasm and irony in social media texts, as well as emoticons and other forms of nonverbal communication. Overall, social networking sentiment analysis is a thriving field of study with multiple applications in fields as diverse as advertising, customer service, and even political science. Sentiment analysis includes the following features:

- *Social media monitoring:* Companies can use sentiment analysis to track and respond to mentions of their brand on social networks, which identify patterns and trends in consumer opinions.
- *Market research:* Using sentiment evaluation, businesses may sift through mountains of customer feedback to learn about their customers' biggest concerns, interests, and degrees of satisfaction.
- *Content moderation:* Sentiment analysis can be used to automatically flag content that is offensive or violates community guidelines.
- *Political monitoring:* One can utilize sentiment analysis to monitor and scrutinize the political inclination expressed on social media platforms, news outlets, and various other sources. The Twitter website responds to billions of API calls daily [3]. More than one million outside websites also use Twitter material. According to this massive advancement, recently, Twitter has recently come under a lot of attention, as client sentiment on contentious matters is regularly expressed in tweets and is used to run agendas over the internet [4]. Text-based data-gathering strategies concentrate primarily on collecting, detecting, and interpreting the existing real information. There is an objective dimension to facts, yet there are additional textual components that represent subjective qualities. Such contents consist mostly of views, sentiments, assessments, behaviors, and feelings, which are the foundation of Sentiment Analysis. It provides several demanding options for application development, Selecting the most effective classifier is of the utmost importance because it will be applied in the classifying of the feelings expressed by Twitter users. Evaluating a person's state of mind ought to be carried out in the most cautious and perceptive manner possible. The machine learning algorithm being deployed must have a very high accuracy rating.

Main Contribution

The major goal of this study is to create a robust model for sentiment analysis using Twitter data, which will automatically categorize tweets as either positive or negative depending on the author's attitude. Using an inquiry word model, Twitter posts must be labeled as positive or negative. The goal is to appropriately categorize the tone of tweets using machine learning techniques. The present state of the art provides numerous deep learning algorithms for forecasting the popularity of content shared on social media platforms.

Paper Organization

The article is combined in follows manner: Next Section presents a concise summary of past work on sentiment analysis. The Section after that defines the methodology accepted for supervised learning which contains the evaluation of results using three classification techniques. The Section following that presents results obtained from the machine learning classification models. And the last Section provides a summary of the study and points the way for future study.

RELATED WORK

Natural language processing research has centered on sentiment analysis. One approach to this problem involves creating classifiers that can determine whether the material is either subjective or objective and, if subjective, to what extent the language used can be categorized as positive or negative. In contrast, other researchers have examined three-way classification systems that include a neutral category in addition to positive and negative [5]. The tweet Classification model uses Twitter API to gather tweets and automatically emoticons can be used to create a corpus. N-gram and POS tags were taken from that database and fed into a multinomial NB sentiment classifier. As it only contained tweets with emoticons, their training set was inefficient [5, 6]. Distant supervision analyzes tweet sentiment based on emoticons as noisy labels. SVMs, MaxEnt, and Naive Bayes are employed. Unigrams, bigrams, and part-of-speech (POS) were their features. SVMs beat other models, and unigram features were better [7]. A multi-faceted automated sentiment analysis technique is for tweet classification in which Tweets were first categorized as objective or subjective, then as good or bad comments. Word polarity and position in the sentence (POS) were used alongside replies, hashtags, hyperlinks, punctuation marks, and quotation marks [8]. After a few times, a 3-way sentiment classification model tried unigram, feature, and tree kernel models. The tree kernel model's tweets took the form of a tree. The unigram structure has over 10,000 features and around a 100 feature-based models. They discovered that characteristics that incorporate the symbol's former polarity using POS labels are more pertinent and important for categorization. Tree kernel models outperformed the other two [9]. Twitter API collects data and Training data has three categories i.e., mobile, camera, and movie. Positive, negative, and non-opinion data are labelled. Filtered opinion tweets. Unigram Naive Bayes model and simplified independence assumption were used. Mutual Information and Chi-square feature extraction reduced unneeded features. Predicting tweet orientation concludes. Positive or negative outperformed the other two models [10].

The article discusses the design and assessment of classifiers for sentiment analysis, with a focus on achieving higher accuracy using ensemble methods and lexicons like bag-of-words attributes [11]. Various hybrid algorithms have been examined, and the results have been compared to lexicon-based methods [11]. The majority-voting ensemble method has been proposed as a more accurate approach, with the highest accuracy score of 81.06% achieved using ensemble methods for the STS dataset. The research concentrates on positive/negative classification, and punctuation, phrases, n-grams, and sequences are only a few of the many feature types used in object instances to enable sentiment analysis. The NB classifier has been evaluated for the Handwritten Character Recognition dataset, achieving an F1 score of 68.15. Social platforms like Facebook and Twitter have allowed the thoughts of millions of people to be shared, making sentiment analysis a useful tool in many industries [12, 13].

Current research concentrates on positive/negative classification. Classifying tweets with different sentiments and emotions requires a clear conclusion. Punctuation, features, and word bag classification are then incorporated into an object instance and coordinates enabling sentiment analysis, using Blogger hashtags to characterize sentiment kind, etc. [14]. The NB classifier had an F score of nearly 70% for the Handwritten Character Recognition dataset. In contrast to ensemble methods, social platforms attract millions of people who are open with their thoughts on many different subjects. As a result, these platforms have multiple applications across various fields of study, including business and social science [15]. The last 2 years have seen the provision of almost 90% of today's data, and it is complex to gain insight from such a vast amount of data [16, 17]. The suggested approach uses semantic closeness measurements to understand the emotions of opinions [18]. One of the most difficult parts was discovering the best approach to using Twitter data to recognize emotions, as evaluating various techniques is difficult without defined standards [19]. The best machine learning model for analysis in many regional languages will be used by the system. The output of the sentiment analysis algorithms will be used to calculate accuracy, f1 score, precision, and recall; and the performance of the machine learning models may be measured by their accuracy at identifying emotional states. Once the most effective model is identified, it will be applied to perform sentiment analysis in all three languages, and the analysis results will be reported using accuracy percentage, f1 score, and other classification report matrices [20].

METHODOLOGY

The sentiment or attitude exhibited in a piece of text can be analyzed and classified using the sentiment analysis architecture shown in Figure 2. This architecture can be broadly categorized into two types: traditional and deep learning-based approaches. Traditional sentiment analysis architecture involves several steps, including text preprocessing, feature extraction, model training, and model evaluation. During preprocessing, the text is cleaned and tokenized, and stop words are removed. Feature extraction transforms the data into a numerical representation that can be fed into a Jupyter

Notebook, including the extracted text from a tweet. This process may involve the use of TF-IDF Vectorizer and vector representation for extracted data. Extraction techniques such as Unigram, N-gram, and external lexicon features are employed [21]. The bag-of-words method and the term frequency and inverted document frequency (TF-IDF) are also employed. Over the last decade, sentiment analysis has made significant progress, with some systems achieving efficiency levels of up to 85–90%.

However, there are still challenges related to data diversity and internet shorthand and acronyms as used on social media. An ML model, like a support vector machine (SVM) or a random forest, is put through its paces during the training phase and is trained on the preprocessed and extracted features.

Model evaluation involves testing the efficacy of the trained (supervised) model on an unrelated data source. The experiment's flow for each is discussed in detail below. These steps are essential in performing sentiment analysis on Twitter data.

- Data Collection.
- Data Preprocessing.
- Data Visualization.
- Model Selection.
- Model Evaluation.

Data Collection

Data collection is always the first stair for the analysis, so, first of all, collected data for analysis. For the experiments, here Python programming language is used as a tool. Stated that there are two ways to acquire data for this particular analysis. The first method is to get pre-organized data from websites like Kaggle and Google Analytics. Another option is to extract the data from Twitter API and label each query manually into classifiers to get the test dataset before training the model. Data collection in sentiment analysis typically involves gathering a text dataset, like media posts, reviews, or customer feedback. Then the text is labelled with a sentiment score, such as positive, negative, or neutral. The present study analyzes the positive and negative aspects of experiments. This annotated data is then used to educate a data-driven algorithm that can identify the emotional tone of the novel text. The quality of the dataset used to prepare the classifier will have a substantial effect on the result formation.

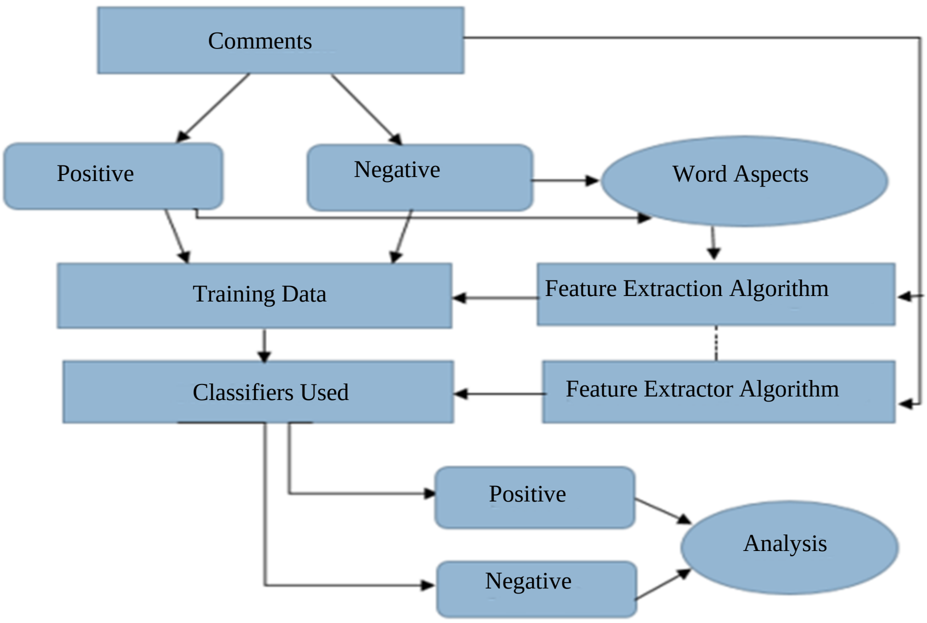


Figure 2. Sentiment analysis architecture.


```
df.sample(5)
```

	id	label	tweet
29530	29531	0	happy day! #altwaystoheal #healthy
16022	16023	0	#dwd #wetterwarnung starkes #gewitter in #schw...
3130	3131	0	@user <jk b-dayð > jungkookday ì êµ...
3007	3008	0	@user @user yes, which is very . open primari...
25595	25596	0	hard to believe this amazing voice contender @...

Figure 3. Sample data with label.

There are several methods for collecting data for sentiment analysis. This study provides a training data set containing Twitter posts and tags, wherein label '1' indicates an unfavorable/negative tweet and label '0' indicates a favorable/positive tweet [22]. It contains three fields: id, tweet, and label, with approximately 32000 entries. The sample is shown in Figure 3.

Data Preprocessing

Pre-processing data involves converting information from its original form into a structure that can be read by a classifier to increase the classifier's accuracy. For example, the raw data being pulled from Twitter needs to be more structured and accurate because it appears to contain many characters, i.e. the clean tweets, as shown in Figure 4.

- Remove user handles (@user).
- Remove all unrelated characters, numeric, and grammar details.
- Removing short words (len<3).
- Tokenize individual words.
- Stem the words in simplest form.
- Combine everything into a single sentence.
- Lowercase string.

Data Visualization

Data sampling methods facilitate the visibility and share insights into trends and vulnerabilities, and an intuitive way [23, 24]. Before training the model, it is essential to analyze and visualize the data and its aspects to accurately develop a model for sentimental analyses. Some key visualizations from our data frame are shown in Figure 5. It explains the top hashtags used in tweets and refers to the most frequently used hashtags on the social media platform Twitter. These hashtags are typically trending and are associated with current events, popular culture, or viral memes. Individuals, organizations, and businesses can use them to join in on a conversation, increase visibility, and reach a larger audience. Positive and negative word visualization is shown in Figure 6(a) and (b) respectively. A word cloud of positive words shows the loving words used on social media, i.e., ‘love’, ‘happi’, ‘beauty’, etc. A word cloud of negative words represents ‘hate’, ‘racist’, etc.

The data set is scattered and lies more on the negative side of the polarity in the analysis. Table 1 gives a count of true sentiment and wrong sentimental words in the dataset. This is true in the common scenario where most users post negative and offending comments on social media. This dataset is quite imbalanced, a common problem in sentiment analysis where many negative instances compared to positive examples can hinder a machine learning model's ability to reliably predict positive sentiment. Various metrics can handle this for the evaluation of F1-score, precision, and recall.

Model Selection

There are several approaches to analyzing the emotion of social media, i.e., AI-based Approach, and the Charlatanical Approach (Lexicon is a set of pre-compiled sentiment expressions, statements, and

Journal of Artificial Intelligence Research & Advances
Volume 11, Issue 2
ISSN: 2395-6720

analogies covering traditional learning categories). An approach grounded on machine learning employs a classification strategy to divide the text into several categories. Two main categories of machine learning methods exist:

- Unsupervised Learning is based on a clustering approach since it has no classification and fails to supply the right targets.
- In Supervised Learning, the model receives labels from a labelled dataset. Labelled datasets are trained to produce relevant decision-making responses.

Both Supervised and unsupervised learning approaches rely on both selecting and extracting sentiment-detecting features. Sentiment analysis uses supervised categorization machine learning, which is based on training data and test data. Many machine learning methods classify Twitter posts. Various ML methods are very famous in this field. The selected model for emotion analysis of social media uses Logistic Regression, and Naïve Bayes, and supports vector machines.

Table 1. Count of words label wise.

Label	Count
Negative	29720
Positive	2242

```python
# combine everything into single sentence
for i in range(len(tokenized_tweet)):
    tokenized_tweet[i] = " ".join(tokenized_tweet[i])

df['clean_tweet'] = tokenized_tweet
df.head()
```

	id	label	tweet	clean_tweet
0	1	0	@user when a father is dysfunctional and is s...	when father dysfunct selfish drag kid into dys...
1	2	0	@user @user thanks for #lyft credit i can't us...	thank #lyft credit caus they offer wheelchair ...
2	3	0	bihday your majesty	bihday your majesti
3	4	0	#model i love u take with u all the time in ...	#model love take with time
4	5	0	factsguide: society now #motivation	factsguid societi #motiv

Figure 4. Clean tweets after preprocessing.

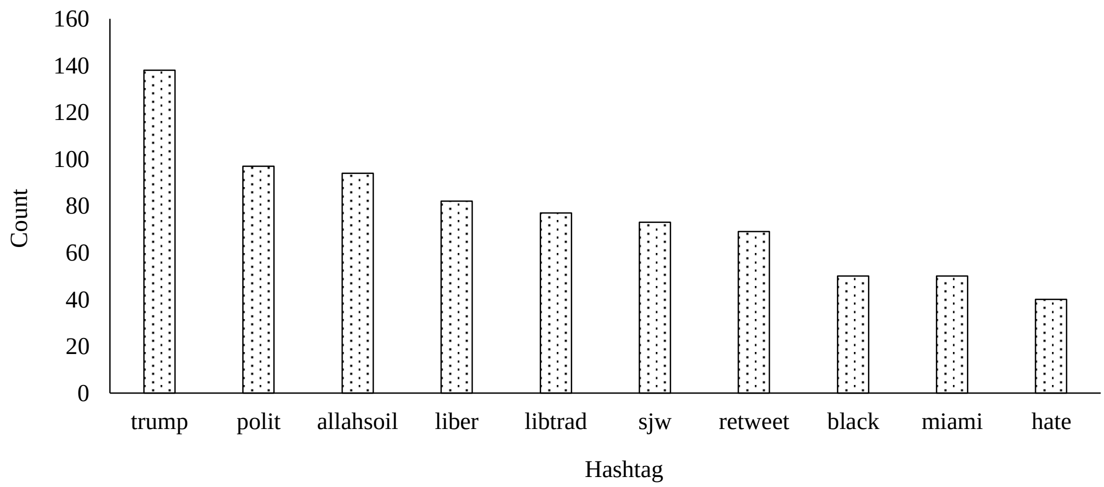


Figure 5. Top 10 hashtags used in tweets.

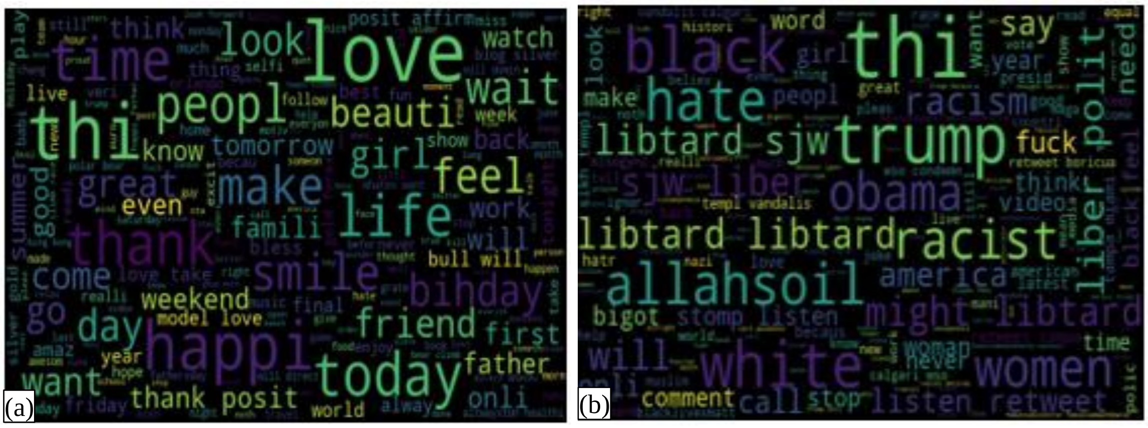


Figure 6. The most frequently used positive and negative words cloud in (a) and (b) respectively.

Logistic Regression

It is a statistical approach commonly defined to solve problems that involve classification and prediction. It involves creating a model that predicts the likelihood of something happening given certain information. LR estimates the probability of event occurrence which provides a dataset of independent variables. To find the repetition or use of a single independent variable, linear regression works, but when multiple independent variables want to compute simultaneously then logistic regression executes.

For several independent variables, the fundamental linear regression (LR) equation is:

$$\hat{Y} = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots\ldots\ldots.. + \beta iXi \tag{1}$$

Here Eq. (1) represents:

- $\hat{Y}$ used for the expected continuous result;
- $\beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta iXi$ is used for the independent variable in LR:
- $\beta 0$ is the intersection point; and
- $\beta 1X1 + \beta 2X2 + \ldots\ldots\ldots.. + \beta iXi$ define the Xi weighted with $\boldsymbol{\beta i}$.

Logistic regression, like linear regression, can have one or more independent variables; however, looking at many variables at once is more instructive since it displays the significant difference from every element after controlling every other variable.

To search independent variables from the dataset, logistic regression can be:

$$\text{Probability of outcome } (\hat{Yi}) = \frac{e^{\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta iXi}}{1+e^{\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta iXi}} \tag{2}$$

Eq. (2) shows logistic regression which includes $\boldsymbol{\widehat{Yi}}$ and $\boldsymbol{\beta 0}$ to find the required result in binary form using a linear equation. In this $\boldsymbol{\widehat{Yi}}$ provides the probability for binary result for the category i. And $e^{\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta iXi}$ is used in the logistic scale to show the independent variables.

The domain of the dependent variable is between 0 and 1. In logistic regression, individual chances are changed using the logit formula, which equals the success probability divided by the chance it will fail. If this inconsistency is not fixed, the regression model's predictions may be beyond the valid range. This issue can be resolved by transformation from linear to logit function by using $\hat{Y}$ versus $(1–\hat{Y})$, in Eq. (3).

$$\ln(\hat{Y}/1-\hat{Y}) = \beta0 + \beta1X1 + \beta2X2 + \cdots + \beta iXi \quad (3)$$

By the iterative use of this logit function, the outcome of independent variables may produce the strongest according to requirement. To ensure that logistic regression develops an appropriate model, it is crucial to consider the identification for independent variables and the approach for model construction.

Bernoulli Naive Bayes

The significant algorithm that gives the standard for the statistical machine learning technique called Naive Bayes, is utilized to solve a range of classification issues. Naive Bayes predicts class probabilities using many attributes. It shows event probability. Two principles underlie the NB classifier:

- Conditional Independence (Independent Components): One component does not influence the other's performance.
- Feature Significance (Equal Priority components): To anticipate well and achieve reliable outcomes, you must know all the features.

Multinomial, Bernoulli, and Gaussian Naive Bayes are the main types of NB. From all, Bernoulli NB is good at handling Boolean/binary attributes. Give 1 is search data present in the document otherwise 0. It performs basic NB training as well as classification methods for multivariate Bernoulli distribution data. Each feature is supposed to be a binary-valued or Boolean variable [25]. Like the multinomial model, this approach is well-liked for tasks involving document classification when binary term occurrence features, that is, whether a comment occurs inside a document or not, are used instead of word frequencies.

The basic NB theory first generates the prior components through Eq. (4). The mixture model generated by the disjoint set Θ of the components $\boldsymbol{c_j}$ contains C={c1, c2, …….. $c_{|C|}$}. Then the sum of products of components represents P($\boldsymbol{d_i}$ | Θ) to choose the components priori.

$$P(d_i \mid \Theta)=\sum_{j=1}^{|C|} P(c_j \mid \Theta)\, P(d_i \mid c_j; \Theta) \quad (4)$$

In Eq. (4), c_j defines both jth for a class and for jth mixture model. The Bernoulli Model Eq. (5) indicates:

- Vocabulary is Z.
- Distance of space given as t.
- Where t $belongs\ to$ $\{1, \ldots\ldots, |Z|\}$ corresponding to the word $\boldsymbol{w_t}$ from available vocabulary. The document $\boldsymbol{d_i}$ is represented as $\boldsymbol{A_{it}}$ which can be 0 or 1 to represent the $\boldsymbol{w_t}$ word that occurs for once in the document.

Word occurrence indicates the independent variable in NB, by the productivity of variable probability on all character attributes by using the following NB equation:

$$P(d_i \mid c_j, \Theta)= \prod_{t=1}^{|Z|} P(w_t \mid c_j; \Theta) + (1 - B_{it})(1 - P(w_t \mid c_j; \Theta))) \quad (5)$$

Bernoulli NB represents a group of several independent experiments. On the basis of probability, an experiment was performed for each word in a given vocabulary (Z), a word expressed by component, $P(w_t \mid c_j; \Theta)$, which shows the Bayesian network (BN) distribution based on document. In this binary representation, result is dependent on the class for present or absent.

The training data named with D={$d_1, \ldots, d_{|D|}$}, is used for the classification model to determine the word probabilities. The Mixture model is represented using $\Theta w_i | c_j$=P($d_i | c_j, \Theta$), where $\Theta w_i | c_j$ lies between 0 and 1. Here Eq. (6) can calculate the probabilities for the optimal solution by counting the events. So, the Laplace theorem is used to avoid the probability of 0 or 1 by counting the words by 1. Explain P($c_i | d_i$)∈(0,1) through a class label. Now calculate the probability of both $\mathbf{w_t}$ and $\mathbf{c_j}$ using:

$$\widehat{\Theta}\, w_t | c_j = P(w_t | c_j ; \Theta) = \frac{\mathbf{1 + \sum_{i=1}^{D} B_{it}\, P(c_j | d_j)}}{\mathbf{2 + \sum_{i=1}^{|D|} P(c_j | d_i)}} \tag{6}$$

Finally, class pre-fix parameters (Eq. (7)), $\Theta \boldsymbol{c_j}$ fixed with the maximum estimated value:

$$\widehat{\Theta}\, c_j = P\left(c_j | \widehat{\Theta}\right) = \frac{\mathbf{\sum_{i=1}^{|D|} P(c_j | d_i)}}{\mathbf{|D|}} \tag{7}$$

Here, this model counts the presence or absence of words as 1 or 0 respectively. And avoids counting the number of words that occur in the documents.

Support Vector Machine

Indeed, this method of machine learning relies on categorization, outlier detection, and regression things [24]. Imagine you have a group of colored dots on the paper. Some of the dots are red and some are blue. A support vector machine is like a line that you can draw on paper to separate the red dots from the blue dots. The line is called a "boundary". This algorithm's objective is to locate the most suitable line for dividing the points. It does this by finding the line that is the farthest distance away from the red dots and blue dots. This line is the best boundary because it gives the most space between the dots and the line.

We can then use this boundary to classify new dots we might see. If we see a new dot on one side of the boundary, we know it is probably a red dot. If we see a new dot on the other side of the boundary, we know it is probably a blue dot. So, the mathematical representation for Support Vector Classifiers (SVC) trains the vector $x_i \in R^p$, i=1, 2, n for different classes, as well as vector almost correct for the document.

Eq. (8) solves the primary problem by using:

$$\min_{w,b,\varsigma} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \varsigma i \text{ subject to } yi(w^T \oint (x)\text{+b})\geq 1\text{-}\varsigma_i\text{, where}$$

$$\varsigma_i \geq 0\text{, and, i=1, 2, ….., n} \tag{8}$$

Here,

- $w^t w$ is used to increase the margin.
- yi($w^T \oint (x)$+b)>=1, would be same for whole document.
- ς_i allows sample data at boundary.
- C is used to control the strength.

The Dual problem for the primal sample is:

$$Min_\alpha \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \text{ subject to } y^T \alpha = 0$$

$$\text{Where } 0\leq \alpha_i \leq \text{C, and, i=1, 2, ….., n} \tag{9}$$

Here e stands for vector, and Q is equal to n for n positive matrix. And α_i is uses for dual coefficients, upper bounded trough C. Eq. (9) is used for the optimization. Now decision function for sample data x will be:

$$\sum_{i \in SZ} y_i \alpha_i \, K(x_i, x) + b \tag{10}$$

After that sum is required over the support vectors because dual coefficients α_i are zero for other sample data through Eq. (10).

Model Selection

Various Machine Learning algorithms were trained and deployed to get efficient results using artificial intelligence techniques. Further, each model was calculated for different metrics and then performance was evaluated and compared. Measures of success included accuracy, precision, recall, and the f1 score, among others. This quartet of measurements would reveal whether or not a dataset was being trained properly and how accurately the system was making predictions. It was also helpful in choosing the most accurate predictive model for tweet classification [26]. For Eq. (1), accuracy was calculated as the percentage of tweets that were correctly categorized. The quantity of misclassified unfavorable tweets was quantified in Eq. (2). Precision calculated the ratio of negative tweets from all negative forecasts in Eq. (3). The F1-score in Eq. (4) is a harmonic mean of precision and recall, taking into account both metrics. The result of these three methods is represented in Figure 7.

A confusion matrix is an approach to determining how well a machine learning model classifies things. It contains:

- *Accuracy:* Accuracy in machine learning refers to how reliably a model makes the intended classification of fresh data. So, if the model was correct 80% of the time, its accuracy would be 80%.
- *Accuracy* = Total correct Predictions/Total Predictions
- *Recall:* In machine learning (ML), to measure the efficacy of a binary classification model, we can look at its recall, which is the fraction of actual positive occurrences for which the model made a correct prediction. Essentially, it provides a numerical representation of the proportion of positive examples accurately identified by the model.

$$\text{Recall}=\frac{\text{True P}}{(\text{True P+False N})}$$

Precision: Precision in machine learning (ML) indicates the proportion of correct positive predictions to all positive predictions made by the model. It counts how often the model's optimistic forecasts come true.

$$\text{Precision}=\frac{\text{True P}}{(\text{True P+ False P})}$$

F1 Score: The F1 score quantifies a model's accuracy at detecting true positives while simultaneously reducing the number of false positives it generates. The greater the F1 score, the more efficient the model is.

$$\text{F1 Score}=2\times\frac{\text{Precision} * \text{Recall}}{\text{Precision}+\text{Recall}}$$

RESULTS

The comparison results for the various machine learning techniques (LR, NB, and SVM) consider all of the metric parameters displayed in Table 2. As for the experiment, Logistic Regression has the best result from other experiments. It performs the best with an accuracy of 95.83% for the sentiment analysis and offers the greatest prediction accuracy and F1 Score, followed by the Support Vector Machine which in turn performs better accuracy at 95.77% as compared to Bernoulli Naive Bayes (accuracy of 94.90%), as shown in Figure 8. Logistic regression achieved Precision (0.97), Recall (0.96), and F1 Score (0.96), which is higher than Naïve Bayes and SVM algorithms.

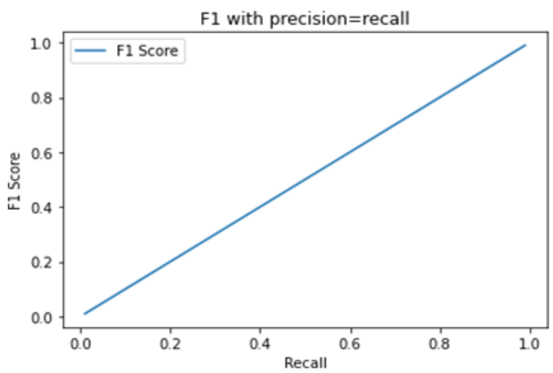
The generated F1-Scores for each class specifically are as follows:

- *class 0:* Accuracy of (Bernoulli NB (0.9583)<SVM (0.9490)<LR (0.9577))
- *class 1:* Accuracy of (Bernoulli NB (0.9583)<SVM (0.9490)<LR (0.9577))

Many factors can influence the performance of a classification algorithm on a particular dataset, such as the complexity of the task, the quality and nature of the features, and the amount and balance of training data. According to the performance scores, the sentiment analysis model did a good job of correctly categorizing positive and negative tweets. The results indicated that logical regression was the most effective machine-learning technique, compared to the other model. It possessed the highest levels of accuracy, recall, and F1 score, and was capable of identifying tweets that may include all types of emotions of the commenter. The classification report generated for all the algorithms, used in this article and their respective scores are shown in Figure 9(a–c).

Table 2. Final Results and Accuracy.

	Precision	Recall	F1 Score	Accuracy
Logistic Regression	0.97	0.96	0.96	95.83
Naive Bayes	0.96	0.95	0.95	94.90
Support Vector Machine	0.96	0.94	0.96	95.77



	Precision	Recall	F1-Score	Difference
0	0.1	0.1	0.1	0.0
1	0.2	0.2	0.2	0.0
2	0.3	0.3	0.3	0.0
3	0.4	0.4	0.4	0.0
4	0.5	0.5	0.5	0.0
5	0.6	0.6	0.6	0.0
6	0.7	0.7	0.7	0.0
7	0.8	0.8	0.8	0.0
8	0.9	0.9	0.9	0.0
9	1.0	1.0	1.0	0.0

Figure 7. Precision, Recall and F1 Score Relation.

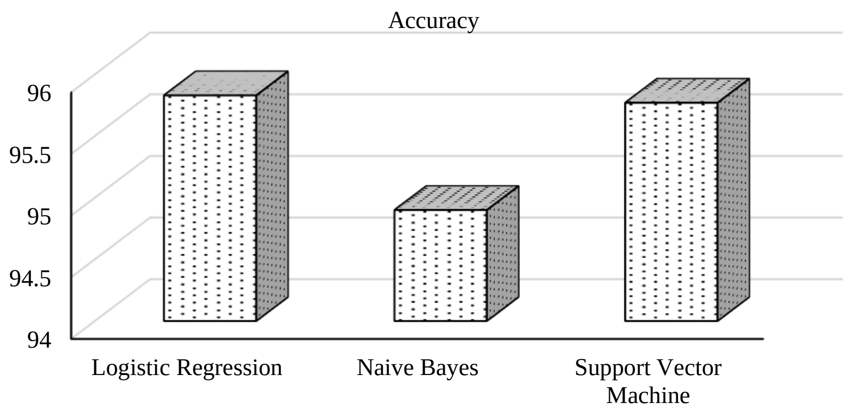


Figure 8. Algorithm’s Accuracy.

(a)

```
              precision    recall  f1-score   support

           0       0.99      0.96      0.98      7643
           1       0.51      0.82      0.63       348

    accuracy                           0.96      7991
   macro avg       0.75      0.89      0.81      7991
weighted avg       0.97      0.96      0.96      7991

[[7371  272]
 [  61  287]]
0.9583281191340258
```

(b)

```
              precision    recall  f1-score   support

           0       0.98      0.97      0.97      7551
           1       0.53      0.67      0.59       440

    accuracy                           0.95      7991
   macro avg       0.76      0.82      0.78      7991
weighted avg       0.96      0.95      0.95      7991

[[7288  263]
 [ 144  296]]
0.9490677011638093
```

(c)

```
              precision    recall  f1-score   support

           0       0.98      0.97      0.98      7498
           1       0.64      0.72      0.68       493

    accuracy                           0.96      7991
   macro avg       0.81      0.85      0.83      7991
weighted avg       0.96      0.96      0.96      7991

[[7296  202]
 [ 136  357]]
0.9577024152171193
```

Figure 9. Classification report (a) Logistic regression classification, (b) Bernoulli classification, (c) SVM classification.

CONCLUSION

Without a doubt, Twitter is the primary platform people use to express their views and feelings. Having a reliable method for predicting the sentiment of tweets is essential. This study presents a sentiment analysis approach that employs Logistic Regression, Bernoulli Naive Bayes, and Support Vector Machine models using Python which evaluate the popularity of tweets or comments. The data, collected from Kaggle, was processed with qualifiers. The study introduces a straightforward supervised algorithm that can tell the difference between a good and bad review. A total of five steps are used for this analysis:

- Collection of data,
- Processing of raw data,
- Visualize the data to see and understand,
- Model used for analysis (Logistic regression, Bernoulli NB, and Support vector machine), and
- Data evaluation through used methods.

Therefore, the efficacy of these classifiers is measured by their Accuracy and F1 Scores. Comparing the accuracy of the models, Logistic Regression came out on top with an accuracy of 95.93%, followed by SVM at 95.77% and Bernoulli Naive Bayes at 94.90%. The further research scope areas in sentiment analysis are:

- Improving accuracy and precision.
- Mixture models create more sophisticated ensemble methods which can produce better results.
- Emojis and emoticons are popularly used which may affect the polarity and need to be handled along.

REFERENCES

1. Kharde V, Sonawane P. Sentiment analysis of twitter data: a survey of techniques. arXiv preprint arXiv:1601.06971. 2016.
2. Mehta P, Pandya S. A review on sentiment analysis methodologies, practices and applications. International Journal of Scientific and Technology Research (IJSTR). 2020; 9(2): 601–609.
3. Brauwers G, Frasincar F. A survey on aspect-based sentiment classification. ACM Comput Surv. 2022; 55(4): 1–37.

4. Bohra N, Bhatnagar V. Group level social media popularity prediction by MRGB and Adam optimization. J Comb Optim. 2021; 41(2): 328–347.
5. Saberi B, Saad S. Sentiment analysis or opinion mining: A review. Int J Adv Sci Eng Inf Technol. 2017; 7(5): 1660–1666.
6. Bohra N, Bhatnagar V, Choudhary A, Ahlawat S, Sheoran D, Kumari A. Popularity Prediction of Social Media Post Using Tensor Factorization. Intell Autom Soft Comput. 2023; 36(1): 205–221.
7. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. CS224N project report, Stanford. 2009; 1(12).
8. Barbosa L, Feng J. Robust sentiment detection on twitter from biased and noisy data. In Coling 2010: Posters. 2010 Aug; 36–44.
9. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau RJ. Sentiment analysis of twitter data. In Proceedings of the workshop on language in social media (LSM 2011). 2011 Jun; 30–38.
10. Liang PW, Dai BR. Opinion mining on social media data. In 2013 IEEE 14th international conference on mobile data management. 2013 Jun; 2: 91–96.
11. Ren B, Cheng L. Research of classification system based on naive bayes and metaclass. In 2009 IEEE 2nd International Conference on Information and Computing Science. 2009 May; 3: 154–156.
12. Alsaeedi A, Khan MZ. A study on sentiment analysis techniques of Twitter data. Int J Adv Comput Sci Appl. 2019; 10(2): 361–374.
13. Kewsuwun N, Kajornkasirat S. A sentiment analysis model of agritech startup on Facebook comments using naive Bayes classifier. Int J Electr Comput Eng (2088-8708). 2022; 12(3): 2829–2838.
14. Patil RS, Kolhe SR. Supervised classifiers with TF-IDF features for sentiment analysis of Marathi tweets. Soc Netw Anal Min. 2022; 12(1): 51.
15. Bagheri H, Torkamani MA, Ghaffari Z. Multi-agent approach for facing challenges in ultra-large scale systems. Int J Electr Comput Eng. 2014; 4(2): 151–154.
16. Ortigosa A, Martín JM, Carro RM. Sentiment analysis in Facebook and its application to e-learning. Comput Hum Behav. 2014; 31: 527–541.
17. Burigo M, Schultheis H. The effects of direction and orientation of located objects on spatial language comprehension. Lang Cogn. 2018; 10(2): 298–328.
18. Kersbergen B, Schelter S. Learnings from a Retail Recommendation System on Billions of Interactions at bol. com. In 2021 IEEE 37th International Conference on Data Engineering (ICDE). 2021 Apr; 2447–2452.
19. Azzouza N, Akli-Astouati K, Oussalah A, Bachir SA. A real-time Twitter sentiment analysis using an unsupervised method. In proceedings of the 7th international conference on web intelligence, mining and semantics. 2017 Jun; 1–10.
20. Aliman G, Nivera TFS, Olazo JCA, Ramos DJP, Sanchez CDB, Amado TM, Valenzuela IC, et al. Sentiment analysis using logistic regression. J Comput Innov Eng Appl. 2022; 7(1): 35–40.
21. Fernández-Gavilanes M, Álvarez-López T, Juncal-Martínez J, Costa-Montenegro E, González-Castaño FJ. Unsupervised method for sentiment analysis in online texts. Expert Syst Appl. 2016; 58: 57–75.
22. Kaur VD. Sentimental analysis of book reviews using unsupervised semantic orientation and supervised machine learning approaches. In 2018 IEEE 2nd International conference on green computing and internet of things (ICGCIoT). 2018 Aug; 519–524.
23. Santos MLBD. The “so-called” UGC: an updated definition of user-generated content in the age of social media. Online Inf Rev. 2022; 46(1): 95–113.
24. Bohra N, Bhatnagar V. Social media popularity prediction by TTM clustered Bayesian factorization. Advances and Applications in Mathematical Sciences (AAMS). 2022; 21(6): 3293–3302.
25. Ramadhan WP, Novianty SA, Setianingsih SC. Sentiment analysis using multinomial logistic regression. In: IEEE International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC). 2017 Sep; 46–49.
26. Luo Z, Osborne M, Wang T. An effective approach to tweets opinion retrieval. World Wide Web. 2015; 18(3): 545–566.

