Journal of
Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776576

A NEW APPROACH: SENTIMENT ANALYSIS OF TWITTER DATA USING A NOVEL DATA MODEL ALGORITHM

Dr. Rupesh Sendre

(Dept. of International Institute of Professional Studies, Devi Ahilya University, Indore, Madhya Pradesh, India, e-mail: sendre.rupesh@gmail.com)

Abstract:

With the continuous growth of web technology, there has been a significant increase in the volume of data available on the internet. This vast amount of data is generated by internet users who utilize online platforms for learning, exchanging ideas, and sharing opinions. Social networking sites such as Twitter, Facebook, and Google+ have gained immense popularity due to their ability to facilitate global discussions, enable expression of viewpoints, and allow individuals to post messages worldwide. Sentiment analysis, a subfield of text mining, focuses on the computational examination of people's opinions, attitudes, and emotions towards a particular subject or entity. This analysis, also known as opinion mining, aims to categorize articles based on their contributions to various sentiment analysis techniques. The objective of this article is to provide a comprehensive overview of sentiment analysis techniques and their related fields with concise explanations. In this study, sentiment analysis is applied to a dataset, which is then divided into positive and negative clusters based on the sentiments expressed. The paper introduces a novel hybrid algorithm for sentiment analysis, designed to enhance accuracy compared to previous methods. Overall, this research contributes to the understanding of sentiment analysis techniques, providing insights into their applications and proposing an improved algorithm for sentiment analysis.

Keywords:

Sentiment Analysis, Twitter Data, Sentiment Classification, Opinion Mining, Hybrid Approach, Novel Data Model algorithm.

I. Introduction:

Data Mining is a subfield of computer science that involves extracting valuable information from large datasets using artificial intelligence, machine learning, statistics, and algorithms. Sentiment analysis, on the other hand, focuses on categorizing expressions as positive, negative, or neutral. It falls under the umbrella of Natural Language Processing (NLP) and is employed to gauge public sentiment towards specific topics. Moreover, sentiment analysis has given rise to the field of opinion mining, which entails the collection and examination of customer reviews regarding particular products or subjects.

The emergence of Social Networking [1] has had a profound impact on our lives. Microblogging services, for instance, enable users to send concise messages. Twitter, a popular microblogging platform, allows users to post and read short 280-character messages known as tweets. On Twitter, individuals share their complaints, suggestions, and reviews on a wide range of products, expressing both positive and negative sentiments. From a business perspective, this information is invaluable. Manufacturing companies often analyze user reactions on microblogs and respond accordingly. By analyzing these sentiments, businesses can increase sales, enhance profits, and encourage positive feedback.

The sentiment found within comments or feedback can be categorized by polarity i.e. either negative or positive.

1. *Positive Sentiment:* "I love the Harry Potter Novel" This sentence expressed positive sentiment about the novel and the sentiment threshold value can be decided by loved word.
2. *Negative Sentiment:* "The Three Robbers is a flop movie" Defined sentence in a negative way and the sentiment threshold value of the sentence is the threshold value of the word flop.

Analyzing sentiments poses a significant challenge due to the variability in how people express their emotions. The same word can have different connotations depending on the context [9]. For instance, the sentence "the book disappoints" conveys a negative sentiment, while "The book does not disappoint" expresses a positive sentiment. Another complication arises from the fact that tweets often combine multiple sentiments within a single message, making it difficult for systems to interpret. People employ diverse emotions, hashtags, and acronyms to express their feelings, and these tweets can be written in any language.

This paper proposes a new algorithm or tool to classify Twitter tweets as either positive or negative, specifically focusing on English-language tweets. While a few researchers have classified sentiments in languages such as Arabic, Italian, Chinese, and Thai, our tool concentrates on English sentiment analysis. Sentiment analysis can be conducted at three levels [10].

Document Level: The entire document is treated as a single entity, and sentiment analysis is performed on the document as a whole, categorizing it as positive, negative, or neutral. However, this approach may not always yield accurate results.

Sentence Level: approach treats each sentence as an individual entity, applying analysis techniques to each sentence and then summarizing the results to provide an overall sentiment for the document.

Aspect Level: approach focuses on discussing specific entities. This type of system identifies the key aspects related to an entity and estimates the average sentiment associated with those aspects.

The proposed a novel algorithm or tool specifically focuses on English-language tweets and sentiment analysis can be performed at the Document Level, Sentence Level, or Aspect Level, depending on the desired granularity of analysis.

II. Literature Review:

The field of sentiment analysis encompasses various aspects, including sentiment classification, opinion summarization, and feature-based sentiment classification. Research in this area has been conducted at different levels. Initially, Turney (2002) [2] and Pang & Lee (2004) [3] focused on document-level classification. Later, Kim & Hovy (2004) [4] classified each sentence within a review, and sentiment analysis expanded to the phrase level with the use of tools like Opinion Finder lexicon. Wilson et al. (2005) [5] emphasized that the inclusion of domain-specific acronyms and emoticons in sentiment analysis algorithms could lead to degraded performance when applied to different domains. Aue & Gamon (2005) [6] conducted a survey of four approaches to customizing sentiment classification systems for new target domains in the absence of large labeled datasets. They performed experiments on data from different domains, highlighting the limitations and benefits of each approach. The results revealed that naïve cross-domain classification yielded poor accuracy compared to the four approaches discussed. In another study, Go et al. (2009) [7] presented the outcomes of using machine learning algorithms for classifying sentiment in Twitter messages. Their training data consisted of Twitter messages with emoticons, which served as noisy labels. This approach of utilizing tweets with emoticons for distant supervised learning was a significant contribution of their research.

Several authors have conducted reviews of both early and recent results in sentiment analysis, highlighting different approaches and techniques. Bermingham & Smeaton (2010) [8] explored the use of a native Bayes classifier and compared the performance of unigram, bigram, and trigram models. Additionally, they introduced a classification method based on emoticons, where tweets ending with :) or :-) were considered positive, while those ending with :(or :-(were classified as negative. Support vector machines were also utilized in their models, with both approaches relying on n-gram models. Agarwal et al. (2011) [10] took a hybrid approach by combining various models to create new hybrid models, comparing their performance against established methods. They also developed features to capture the expression of writing style in sentiment analysis. Furthermore, Benhardus & Kalita (2013) [12] conducted sentiment analysis using different tools for various purposes and proposed various approaches for real-time monitoring of Twitter data.

Jandail (2014) [13] conducted a study that utilized feature selection to identify sentiments in blog posts related to recent product policies and service reviews. The research aimed to demonstrate the feasibility of clustering and classifying opinion mining through analyzing a vast amount of user-generated data across geographical, demographic, and cultural boundaries. The study proposed the idea that even sentences without explicit opinions or sentiment-specific words can express sentiments. Sonachalam (2015) [15] focused on predicting the future performance of industries in the stock market by using sentiment analysis of tweets. The study explored the relationship between sentiment expressed in tweets and the subsequent gains or losses in the stock market. Kharde and Sonawane (2016) [18] provided a comprehensive survey and comparative analysis of existing techniques for opinion mining, including machine learning and lexicon-based approaches. They also evaluated different machine learning algorithms such as Naive Bayes, Max Entropy, and Support Vector Machine, specifically examining their application to sentiment analysis on Twitter data streams. Liao et al. (2017) [19] proposed an approach based on deep learning techniques for sentiment analysis of Twitter data to understand real-world situations. Their method utilized Convolutional Neural Networks (CNN) and aimed to predict user satisfaction with products, emotional responses to specific environments, and post-disaster situations. The study highlighted the effectiveness of deep learning, particularly CNN, in image analysis and classification. Patel et al. (2017) [20] discussed an approach that involved preprocessing and classifying streams of tweets from the Twitter micro-blogging site based on their emotional content as positive, negative, or neutral. The study analyzed the performance of an unsupervised algorithm and compared it with existing systems, presenting various applications and potential implications of the research. Ahuja and Dubey (2017) [21] conducted a study on clustering techniques in sentiment analysis, demonstrating that clustering can efficiently differentiate tweets based on their sentiment scores. The research focused on identifying strongly positive or negative tweets on a weekly basis using different dictionaries. The paper surveyed various clustering approaches and presented a methodology for establishing relationships between tweets based on polarity and subjectivity.

Shehar and Huang (2018) [22] conducted sentiment analysis on Twitter data related to donations, fundraising, and charities. The study focused on exploring techniques and approaches to capture the polarity of sentiments expressed by individuals toward donating for various causes. Exploratory data analysis was utilized, and the Natural Language Processing Toolkit (NLTK) was employed to determine whether a tweet exhibited a neutral, positive, or negative polarity. The collected data on donation-related tweets served as training data and aimed to attract potential clients for future endeavors. Rahman et al. (2019) [23] proposed a novel model that combined supervised and unsupervised machine learning algorithms for sentiment analysis. The process involved extracting tweets directly from the Twitter API, performing data cleaning and discovery, and feeding the data into multiple models for training purposes. Each tweet was classified based on its sentiment, whether positive, negative, or neutral. The study collected data on two subjects, McDonald's and KFC, to compare their popularity. Various machine learning algorithms were employed, and the results were evaluated using testing metrics such as cross-validation and f-score. The proposed model demonstrated strong performance in mining texts extracted directly from Twitter. In a recent review, Wang et al. (2022) [24] examined the most recent developments in the field of sentiment analysis. The review explored a wide range of newly proposed algorithms and applications. The publications were categorized based on their significance to specific types of Text Sentiment Analysis (TSA) methods. The survey aimed to provide a concise and comprehensive overview of TSA techniques and related fields. The primary contributions of the survey were the detailed classification of numerous recent articles and the depiction of the current research direction in the field of TSA.

III. Significance of Study:

The primary focus of sentiment analysis is to accurately assign the appropriate emotion to a tweet based on the user's intended expression. However, several challenges arise due to the unstructured and ungrammatical nature of the text. Since tweets are limited to 280 characters, users often resort to abbreviations, slangs, and emoticons to condense their messages. Additionally, typing errors can occur when users are typing quickly, potentially altering the meaning of an entire sentence. Another challenge lies in the presence of ambiguity, such as when the word "orange" can refer to both a color and a fruit.

While existing approaches typically address sentiment analysis at the sentence level, document level, or feature level, our proposed approach will cater to all of these levels. We aim to tackle these challenges by splitting the document into sentences and performing sentiment analysis on each individual sentence. The polarity of each sentence will then contribute to determining the overall sentiment of the document.

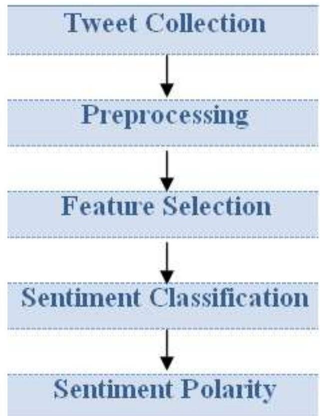

Figure – 1: Sentiment Analysis Process

1. ***Tweets:*** Twitter is a microblogging service that allows users to post real time messages, called tweets. Tweets are short messages restricted to 280 characters in length. There is a small terminology associated with tweets.
   1.1 *Emoticons:* a representation of a facial expression such as a smile formed by various combinations of keyboard characters and used in electronic communications to convey the author's feelings or intended tone. It is also known as smileys.
   1.2 *Target:* twitter uses the "@" symbol to refer to the other users on the twitter. Referring to other users notify them, also called as tag.
   1.3 *Hashtags:* users normally use the hashtags to mark the topics. It is mainly done to increase the visibility count of their tweet.

Large numbers of tweets are collected about a particular product form www.twitter.com (Official twitter website). For collection of tweets twitter API or manual process can be used. Twitter API requires some technical skills plus a difficult task but once a twitter API established then it is very fast than the manual process which involves selection of individual tweet from the twitter.

2. ***Preprocessing:*** the tweets are collected from a commercial source. These tweets can be in the foreign languages. So, Google translate can be used to convert the tweets into English before the annotation process. Each tweet is labeled by a human annotator as positive, negative, neutral or junk. The "junk" label means tweet can't be understood by a human annotator. So, these junk tweets should be eliminated from our data. Now, an emoticons dictionary is used with their emoticons state which is used to classify the emoticons as positive and negative. For example :=) is labeled as positive whereas :=(labeled as negative. A dictionary can be prepared with all such emoticons with the labels as: Extremely Positive, Positive, Neutral, Negative, and Extremely Negative. The list is already available at wikipedia. Since, the tweets are very short so a people use slangs instead of the whole word so an acronym dictionary is used which will classify all the slangs into their original meaning. For example – *lol* is translated into the laughing out loud, *rotf* is translated into rolling on the floor. An acronym dictionary is used which contain many acronyms.

   After emoticons and acronyms another challenge is removal of urls. Url or Uniform Resource Locator is the global address of the resource on the world wide web. It will not help us into the sentiment detection. So, urls should be removed from the tweets. The urls can be replaced with a ||U|| letter because direct removal of urls can change the POS tagging described later in paper. Similarly, the tags like @RupeshSendre should be replaced by the ||R|| letter.

Now, stemming will be performed. Stemming defines a technique that is used to find the root or stem of a word. The filtered token undergoes stemming to reduce the length of works until a minimum length is reached. This resulted in reducing the different grammatical forms of a word to a single term. The basic stemming process is shown here: Removing the endings is also important. The general rules for dropping the endings from words include:

2.1 If a word ends in ‘es’ drop the s.
2.2 If a word ends in ‘ing’ delete the ing.
2.3 If a word ends in a consonant, other than s, followed by s then delete s.

The words can be transformed to some other grammatical form using a set of defined rules. For example – if the word ends with ‘ies’ but not ‘eies’ and ‘aies’ then the ‘ies’ ban be replaced with a ‘y’ such as ‘Butterflies’ can be replaced with ‘Butterfly’.

Word	Stem
User, used, using, users	Use
Engineering, engineer, engineered	Engineer
Architectural, architectural, architecturally	Architecture

Figure 2: Different Grammatical forms of a word and the corresponding stem.

Now, we replace all negations (e.g. not, no, never, n’t, cannot, etc) by tag “NOT”. Other than these replacing a sequence of repeated characters by three characters, for e.g. Convert cooooooooool to cool.

Acronym	English expansion
Gr8, gr8t	Great
Lol	Laughing out loud
Rotf	Rolling on the floor
Bff	Best friend forever
Idk	I don’t know

Figure 3: Example acronym and their expansion in the acronym dictionary.

Emoticon	Polarity
:-) :) :) :] :c :c)	Positive
:D C:	Extremely-Positive
:-(:(:c :[	Negative
D8 D: D= DX v.v	Extremely-Negative
:|	Neutral

Figure 4: Part of the dictionary of emoticons.

POS Tagging: Collected reviews are sent to the POS tagger that tags all the words of the document to their appropriate part of speech tag. POS tagging is necessary to determine the opinion words. It can be done manually or with the help of POS tagger. POS tagger is used here to tag the entire document. For example Figure 5 shows an example of POS tagging.

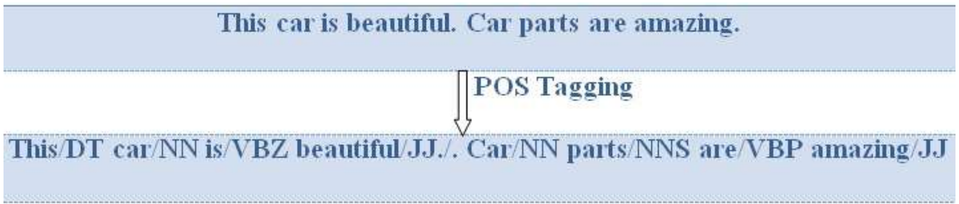

Figure 5: Example of POS Tagging.

3. ***Feature Selection:*** The majority of the approaches for SA involve a two-step process:
   3.1 Identify the parts of the document that will likely contribute to positive or negative sentiments.
   3.2 Combine these parts of the document in ways that increase the odds of the document falling into one of these two polar categories.

Feature selection methods reduce the original feature set by removing irrelevant features for text sentiment classification to improve classification accuracy and decrease the running time of learning algorithms. We have investigated performance of five commonly used feature selection methods in data mining research, i.e., DF, IG, CHI, GR and Relief-F. All these feature selection methods compute a score for each individual feature and then select top ranked features as per that score.

Document Frequency: Document Frequency (DF) measures the frequency with which a feature appears across the documents in a dataset. This method removes features that have a document frequency below or above a predefined threshold frequency. By selecting frequent features, the likelihood increases that these features will also be present in prospective future test cases. The underlying assumption is that both rare and common features do not contribute significantly to sentiment category prediction, or they may not have a substantial impact. This approach is considered simple, scalable, and effective for text classification.

Information Gain: Information Gain (IG) is utilized as a feature (term) goodness criterion in machine learning based classification. It measures information obtained (in bits) for class prediction of an arbitrary text document by evaluating the presence or absence of a feature in that text document. Information Gain is calculated by the feature's contribution on decreasing overall entropy. The expected information needed to classify an instance (tuple) for partition D or identify the class label of an instance in *D* is known entropy and is given by:

$$Info(D) = - \Sigma (P_i) \log_2 (P_i)$$

Where m represents the number of classes ($m = 2$ for binary classification) and P_i donotes probability that a random instance in partition *D* belongs to class *Ci* estimated as $|Ci , D| / |D|$ (i.e. proportion of instances of each class or category). A log function to the base 2 justifies the fact that we encode information in bits

Gain Ratio: Gain Ratio (GR) enhances information gain as it offers a normalized score of a feature's contribution to an optimal information gain based classification decision. GR is utilized as an iterative process where we select smaller sets of features in incremental fashion. These iterations terminate when there is only predefined number of features remaining. GR is used as one of disparity measures and the high gain ration for selected feature implies that the feature will be useful for classification. GR was firstly used in decision tree, and applies normalization to information gain score by utilizing a split information value.

4. ***Sentiment Classification:*** Sentiment classification techniques can be roughly divided into machine learning approach, lexicon based approach and hybrid approach. The Machine Learning (ML) approach applies the famous ML algorithm and uses linguistic features. The Lexicon-based approach relies on a sentiment lexicon, a collection of known and precompiled sentiment terms. It is divided into dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity. The Hybrid approach combines both approaches and is very common with sentiment lexicons playing a key role in the majority of methods.

4.1 *Lexicon – Based Classifier:* the lexicon-based classifier is based on the idea that the polarity of a text can be given by the sum of the individual polarity values of each word or phrase present in the text. For this, a sentiment lexicon identifies polarity words and assigns polarity values to them (known as semantic orientations). In our algorithm, the semantic orientations of each individual word in the text are added up. In this approach, the algorithm searches for each word in the lexicon and only the words that were found are returned. We associate the value +1 to the positive words, and -1 to the negative words. If a polarity word is negated, its value is inverted. This lexicon-based classifier assumes the signal of the final score as the sentiment class (positive or negative) and the score zero as neutral.

4.2 *Machine Learning Classifier:* the features used by the classifier are:

- Unigrams, bigrams and trigrams.
- The presence of negation.
- The presence of three or more characters in the words.
- The sequence of three or more punctuation marks.
- The number of words with all letters in uppercase.
- The total number of each tag present in the text.
- The number of positive words computed by the lexicon-based method.
- The number of negative words computed by the lexicon-based method.

There are so many open-source text-analytics tools used for natural language processing such as information extraction and classification can also be applied for sentiment analysis.

a. *NTLK:* The natural language toolkit is a tool for text processing, classification, tokenization, stemming, tagging, parsing, etc. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as Word Net, along with a suite of http://www.nltk.org/

b. *GATE:* Useful if you want to develop a pipeline. Language analysis modules for various languages are contributed by developers are available to be used plugged in your pipeline.

c. *OpenNLP:* Perform the most common NLP tasks, such as POS tagging, named entity extraction, chunking and co-reference resolution. Htt://opennlp.apache.org/

d. *Opinion Finder:* It aims to identify subjective sentences and to mark various aspects of subjective in these sentences, including the opinion holder of the subjectivity and words that are included in phrases expressing positive or negative sentiments. http://code.google.com/p/opinionfinder/

e. *Ling Pipe:* Ling Pipe is used for linguistic processing of text including, clustering classification and entity extraction etc. http://alias-i.com/lingpipe/

5. ***Results:*** We are able to analyses the total text in its positive and negative polarity for the particular feature.

S (W +ve) = Set of Positive Sentiment words
S (W -ve) = Set of Negative Sentiment words
For N^{th} Feature –
$S (W +) = (W_1 + W_2 + W_3 + \ldots\ldots\ldots.. W_n)$
Set of Positive Sentiment words
$S (W -) = (W_1 + W_2 + W_3 + \ldots\ldots\ldots.. W_n)$
Set of Negative Sentiment words

First of all we calculate the negative and positive polarity every sentiment by calculating its probability. Then we shall calculate the mutual information of its positive or negative sentiment. After calculating the overall polarity of the individual feature we shall calculate the mutual information for every feature for its final product review analysis.

IV. Proposed Algorithm:

The proposed algorithm opens up several avenues for future development and expansion. Here are some potential areas of future scope:

1. *Enhanced Data Collection:* As the algorithm relies on tweet collection, future work can explore more efficient and comprehensive methods for collecting tweets about specific products or topics. This could involve utilizing advanced data mining techniques, leveraging social media APIs, or implementing automated processes to gather a larger and more diverse dataset.
2. *Advanced Preprocessing Techniques:* The preprocessing step can be further enhanced by incorporating more sophisticated techniques. This may involve utilizing advanced natural language processing algorithms, sentiment-specific lexicons, or machine learning models to improve the accuracy of preprocessing tasks such as emoticon grouping, misspelling correction, and part-of-speech tagging.
3. *Robust Feature Selection:* The feature selection process can be optimized by exploring additional schemes or algorithms. Future work can investigate the performance of novel feature selection methods, consider domain-specific features, or explore the incorporation of word embeddings or contextual representations to capture more nuanced sentiment information. The schemes used for word vector creation includes, Term Occurrence, Binary term occurrence, Tern frequency and TF-IDF (Term Frequency-Inverse Document Frequency). These are based on the following values – *fij*: total occurrences of the term *i* in the document *j*. *fdj*: total number of terms occurring in document *j*. *ft1*: total number of documents in which the term *i* occurs.
   Term occurrence: defines the absolute number of occurrences of a term.
   Term occurrence = *fij*
   Term frequency: defines the relative frequency of a term in the document.
   Term frequency = *fij / fdj*
   Binary term occurrence: term occurrence is defined as the binary value.

Binary Term occurrence = 1 for *fij*>0 and = 0 otherwise.

TF-IDF: it describes how important a word is for a document. It consists of two parts – term frequency (TF) and invert document frequency (IDF).

TF-IDF = (*fij* / *fdj*)*log*(1/*fii*)

4. *Integration of Ensemble Classifiers:* The sentiment classification pipeline can be extended to incorporate ensemble classifiers. Ensemble methods, such as stacking or boosting, can be explored to leverage the strengths of multiple classifiers and improve overall sentiment classification accuracy. Additionally, the confidence thresholds for each classifier can be fine-tuned dynamically based on the specific requirements of the system.
5. *Evaluation Metrics and Analysis:* Future research can focus on utilizing various evaluation metrics to assess the performance of the sentiment classification system. Metrics such as precision, recall, F1-score, and accuracy can be employed to provide a comprehensive evaluation of the system's performance. Additionally, conducting in-depth analysis of misclassified tweets and exploring strategies for error analysis can further enhance the system's accuracy and identify areas for improvement.

   By addressing these future directions, the proposed algorithm can be refined, optimized, and adapted to handle diverse scenarios and achieve more accurate sentiment classification results.
6. *Results:* Moreover, the behaviors of the twitter users have been collected from the twitter application using twitter streaming API. Now, tweets are classified into positive, negative and neutral. The dataset used for the experiments was divided into two classes positive and negative. For a given classifier and a document there are four possible outcomes: true positive, false positive, true negative and false negative. If the document is labeled positive and is classified as positive it is counted as true positive else if it is classified as negative it is counted false negative. Similarly, if a document is labeled negative and is classified as negative it is counted as true negative else if it is classified as positive it is counted as false positive.
   - Accuracy = (tp + tn) / (P + N)
   - Precision = tp / (tp + fn)
   - Recall/true positive rate = tp / P
   - F-measure = 2 / ((1/precision) + (1/recall))
   - False alarm rate/false positive rate = fn / N
   - Specificity = tn / (fp + tn) = (1 – fp rate)

V. Experimental Study:

The initial results obtained by Pedro et al. (2013) [11] are presented in Table 1. The precision and recall values for the positive and negative classes are provided. These results serve as the baseline for comparison.

Class	Precision	Recall
Positive	0.6935	0.6145
Negative	0.5614	0.4110

Table – 1: Results for twitter data set.

Following the implementation of the new algorithm, an improvement in performance is expected. The improved results, after the new algorithm's implementation, are shown in Table 2.

Class	Precision	Recall
Positive	0.7041	0.6900
Negative	0.6869	0.6800

Table – 2: Improved Results for twitter data set.

Comparing the results from Table 1 and Table 2, it is evident that the new algorithm has led to improvements in both precision and recall for both positive and negative classes. The precision for positive class increased from 0.6935 to 0.7041, while the recall increased from 0.6145 to 0.6900. Similarly, the precision for the negative class improved from 0.5614 to 0.6869, with a recall increase from 0.4110 to 0.6800.

VI. Conclusion:

In conclusion, this study demonstrated the results of sentiment analysis conducted on Twitter data. Sentiment analysis is a complex technology that requires careful execution; however, the potential benefits are significant. By utilizing this new method or tool for sentiment analysis on Twitter data, it is expected that the accuracy of the analysis would surpass that of human processing. One potential future direction for this research involves enhancing the performance of each individual classifier. In our study, we employed simple methods for each classifier used, which suggests that the application of hybrid classification techniques could yield even better outcomes. These findings support the hypothesis that hybrid techniques have the potential to surpass the current state-of-the-art in sentiment analysis.

References

[1] https://www.wikiwand.com/en/Social_networking_service, a free encyclopedia.

[2] Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting on ACL'02, pp. 417-424.

[3] Pang, B. & Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. Proceedings of the 42nd Annual Meeting of ACL-04, pp. 271–278.

[4] Kim, S. M. & Hovy, E. (2004).Determining the sentiment of opinions. Proceedings of the International Conference on Computational Linguistic. pp. 271–278.

[5] Wilson, T. et al. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 347–354.

[6] Aue, A. & Gamon, M. (2005). Customizing sentiment classifiers to new domains: A case study. Proceedings of the 5th International Conference on Recent Advances in Natural Language Processing.

[7] Go, A. et al. (2009). Twitter sentiment classification using distant supervision.

[8] Bermingham, A. & Smeaton, A. (2010). Classifying sentiment in microblogs: Is brevity an advantage. Proceedings of the 19th ACM, CIKM – 10.

[9] Vinodhini G. & Chandrasekaran R.M. (2012). Sentiment analysis and opinion mining: A survey. IJARCSSE, Vol. 02, Issue 06, pp. 282-292.

[10] Agarwal, A. et al. (2011). Sentiment analysis of Twitter data. Proceedings of the Workshop on Language in Social Media (LSM 2011), pp. 30-38.

[11] Pedro, P. et al. (2013). A hybrid system for sentiment analysis in Twitter messages. Second Joint Conference on Lexical and Computational Semantics, Vol. 2: Seventh International Workshop on Semantic Evaluation, pp. 568-572.

[12] Benhardus, J. & Kalita, J. (2013). Streaming trend detection in Twitter. International Journal of Web Based Communities, Vol. 09, No. 01, pp. 122-140.

[13] Jandail, R. R. S. (2014). A Proposed Novel approach for sentiment analysis and opinion mining. IJU, Vol. 05, No. 1/2. pp. 1-10.

[14] Chen, R. & Lazer, M. (2011). Sentiment analysis of Twitter feeds for the prediction of stock market movement. CS 229, pp.1-15.

[15] Sonachalam, D. (2015). Using Twitter to predict stock market returns. International Journal of Scientific & Engineering Research, Vol. 06, Issue 10, pp. 76-77.

[16] Rao, T. & Srivastava, S. (2012). Analyzing stock market movement using Twitter sentiment analysis. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM-2012), pp. 119-123.

[17] Nahar, V. et al. (2012). Sentiment analysis for effective detection of cyber bullying. Springer, Fourteenth Asia-Pacific Web Conference, pp. 767-774.

[18] Kharde V. A. & Sonawane, S. S. (2016). Sentiment analysis of twitter data: A survey of techniques. IJCA, Vol. 139, No. 11, pp. 5-15.

[19] Liao, S. et al. (2017). CNN for situations understanding based on sentiment analysis of twitter data. Elsevier, Procedia Computer Science 111, pp. 376-381.

[20] Patel, A. P. et al. (2017). Literature survey on sentiment analysis of twitter data using machine learning approaches. IJIRST, Vol.03, Issue 10, pp. 19-21.

[21] Ahuja, S. & Dubey, G. (2017). Clustering and sentiment analysis on Twitter data. Proceedings of the IEEE Second International Conference on Telecommunication and Networks.

[22] Shehar, A. & Huang, C. Y. (2018). Sentiment analysis of Twitter data. Proceedings of the IEEE International Conference on Computational Science & Computational Intelligence.

[23] Rahman, A. EI. S. et al. (2019). Sentiment analysis of Twitter data. Proceedings of the IEEE Xplore, International Conference on Computer and Information Sciences.

[24] Wang, Y. et al. (2022). Sentiment analysis of twitter data. Applied Sciences, 2022, 12, pp. 1-14.