International Journal of Applied Mathematics

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

MULTI-ASPECT SENTIMENT ANALYSIS OF SOCIAL MEDIA TEXT USING LEXICON BASED FASTEXT

Dr. S.Punitha, S Abarna

Assistant Professor, Annamalai University, Tamil Nadu, India

spddeprof@gmail.com

Research Scholar

Annamalai University

Tamil Nadu, India

abarnaabirami@gmail.com

Abstract

Social media platforms serve as vast repositories of diverse user-generated content expressing opinions, sentiments, and experiences across multiple facets of life. Traditional Sentiment Analysis (SA) methods often fall short in capturing the multifaceted nature of opinions present in social media text, as they tend to oversimplify the analysis by focusing solely on overall sentiment polarity. This paper introduces a novel approach to overcome this constraint by the usage of Multi-Aspect SA utilizing a Lexicon-Based FasText framework. By leveraging FasText, an efficient word embedding technique, enriched with lexicon-based SA, this research aims to discern sentiments across various aspects or dimensions within social media texts. Our methodology involves the creation and utilization of domain-specific sentiment lexicons to capture sentiments related to different aspects such as product features, service quality, user experience, and more. The FasText framework is employed to generate context-aware word representations, enhancing SA accuracy across diverse aspects within social media texts. The experimental evaluation demonstrates how effective the recommended approach is in extracting nuanced sentiments, providing insights into multidimensional user opinions present in social media content. This research contributes to advancing SA methodologies by enabling a more granular and comprehensive understanding of sentiment expressions within the dynamic landscape of social media communication.

Keywords: Social Media; Sentiment Analysis; Lexicon based FasText; Domain Specific; Communication

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1. Introduction

As the Internet has expanded, a substantial amount of information has become readily available to the public in the past few years. Individuals worldwide are digitizing their varied opinions and experiences, resulting in a plethora of information available in digital form. The advancement of the internet is progressing towards a phase in which customer thoughts not only forecast but also apply influence, shaping both products and services [1]. The processing of these messages and the extraction of valuable information involve the use of SA methods. SA has garnered significant attention over the last decade and has attained the status of an emerging research direction [2]. The term "sentiment" or "opinion" refers to an individual's perspective, understanding, or experience about a particular entity. SA finds wide applications in various aspects of life, and expressed opinions hold prime importance in this field, especially for businesses where opinions make or decide the future. SA finds major applications in areas such as Brand Monitoring, Market Intelligence, Stock Market Analysis, Reputation Management, Flame Detection, and more [3]. Opinion-Oriented Words (OOWs) play a crucial role in conveying an individual's sentiments, also a main tool for identifying inferred sentiment in textual data. A lexicon is produced by combining words or phrases that express opinions. A polarity lexicon is a repository that lists terms of opinion together with their scores or polarity orientation. When doing SA, this lexicon is an essential tool, as it minimizes the need for recreating efforts and facilitates faster processing [4].

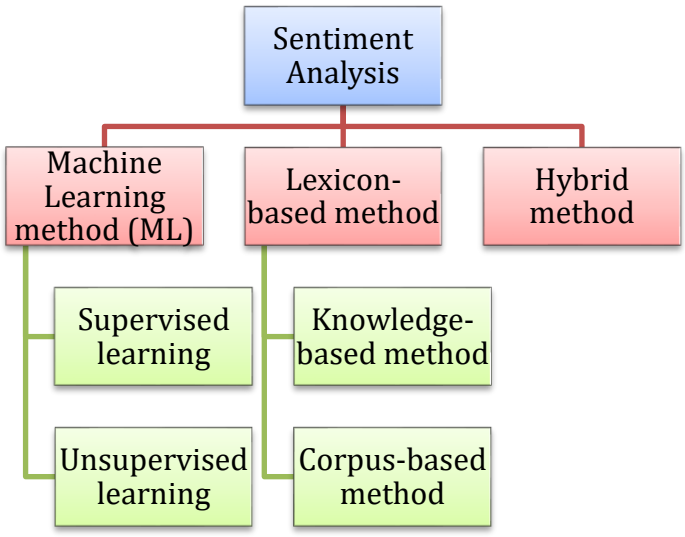


Figure 1: General classification of SA approaches

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

SA approaches are commonly classified into two main types: machine learning-based approaches and lexicon-based approaches. Here this combination of two is the hybrid approach. The general classification of SA approaches is depicted in Figure 1. Machine learning methods address SA by treating it as a conventional text classification challenge, employing various algorithms [5]. These approaches leverage both syntactic and linguistic features, incorporating fundamental elements like part-of-speech patterns, pruning, n-grams, and deep features. SA is a computational examination of individuals' emotions and opinions but also combine sentiments, attitudes as expressed in texts concerning a specific entity [6]. SA serves various objectives, encompassing the identification of public sentiment toward political events, gauging customer satisfaction levels, market intelligence and predicting film sales. With the surge in e-commerce interest, social media websites and networks enable users to contribute their thoughts, ideas, and reviews [7].

The wealth of opinionated data serves as a valuable source for expressing and analysing sentiments. Customer and user opinions take a critical part in enhancing the quality and standards of the products. Reviews posted on social media networks and e-commerce platforms such as IMDb, Snap deal, eBay, Amazon, Flipkart and epinions.com can significantly influence customer decisions when purchasing products and subscribing to services [8].

In the 21st century, social media stands out as an important technology for information exchange. Individuals of all ages utilize social media sites like Facebook and Twitter to post films, images, and messages describing their daily activities and experiences. These platforms offer efficient and accessible means of public communication and information sharing. Subsequently, social medias contribution towards criminal investigations and also in criminal prevention is expanding rapidly [9]. Recently the Social media platform is swiftly transforming into a vital information source for early cautioning systems ensuring the general public is safe. According to a recent LexisNexis survey, four out of every five law enforcement specialists make use of these kind of social media extensively for their investigative tasks. As per the provided statistics, approximately 41% utilize social media for anticipating and preventing criminal activities, predominantly 69% of individuals use these sites as a tool to collect information about criminal activities [10].

SA has already found application in various non-security areas for the purpose of tracking and forecasting public opinion. Customer reviews of hotels were categorized into five-star ratings

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

in one instance using a domain-specific language. An attempt was made to forecast movie box office receipts using SA on Twitter [11]. The findings revealed a correlation between the frequency of tweets related to a movie and its actual box-office performance in the real world. A study by the authors shown a correlation between online debates on IMDb, box office performance, and Academy Award nominations. This is an additional similar use of social media analysis. By introducing a framework that uses Twitter to identify the subjects and segments of an event, the researchers created a novel model for event analytics [12].

In the end, the paper focuses on the analysis of tweets, which are messages posted on Twitter. Twitter is a popular social media and microblogging platform, serves as a platform for transferring text-based messages to a range of 140 characters. SA on tweets is conducted differently compared to longer plain text messages. The limitation starts from abbreviations that are typed, the shortness of tweets, the incorporation of slangs and the inadequate sentence structure often present in tweets. These factors collectively contribute to the increased difficulty in analysing the text. The process of automatically extracting emotions from the source text is achieved through two main approaches [13]. However, for this paper, we use the lexicon-based approach. Establishing a space for sharing ideas and promoting open discussions, the online service encourages individuals to actively participate in group conversations. This interactive forum allows for swift feedback and responses from people on a wide range of subjects and content types, such as textual posts, news, images, and videos [14]. Therefore, these entities can be used to analyse people's opinions, providing insights into consumer attitudes and market trends. Research has indicated an important effect of blog sentiments on the financial performance of enterprises [15].

1.1 Problem Statement

Social media data consists of audios, videos, images, URLs, word variations, emoticons, hashtags, contextual texts, short text like acronyms etc. The combination of various data rigorously affects the polarity of words. The acronyms, emoticons, and negations are major issues, which affect the accuracy of lexicon based approach. To extract the sentiments from the dataset, selecting the relevant features and applying the appropriate classification technique are major tasks in that approach.

In recent times, SA has emerged as a crucial research area, experiencing rapid growth. The development of a human-computer interaction system that can discover feelings in data

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

gathered from several sources is urgently needed, especially in perspective of the advances made in artificial intelligence. Social media's long evolution has produced a tonne of multimodal data that can be recognized for SA, such as text, audio, and image/video. However, in SA a significant portion of current research concentrates solely on a particular modality, making it challenging to ascertain the precise sentiments of individuals. Unimodal systems exhibit limitations in terms of accuracy, reliability, and robustness. Therefore, to improve SA system performance, it is necessary to look into the integration of several modalities. The main goal is to improve SA performance overall by creating a multimodal system that recognizes inputs from multiple sources. Our primary focus is on offering a detailed account covering datasets, techniques for feature extraction, fusion methods, classification approaches, and the challenges linked to multimodal SA.

1.2 Motivation of the Research

Among the three SA approaches, lexicon-based approaches are useful for an immediate task as well are valuable due to building learned resources, which is useful for future tasks. Major research works from literature are based on pre-learned and existing lexicons to improve performance including recent neural network approaches. Henceforth a lexicon with a regular sentiment should be the constructed first. SA approaches yield optimal results when utilizing sentiment lexicons specific to the domain. Applying a sentiment lexicon learned from one domain directly to other domains tends to result in sub-optimal performance.

The primary factor contributing to suboptimal results is the variability in the polarity and intensity of sentiment words across different domains. Sentiment lexicons learned from specific domains maintain the domain-specific orientation of sentiment words, and they demonstrate superior performance compared to general-purpose sentiment lexicons. The fluctuation in sentiment word polarities underscores the necessity of introducing domain-based sentiment lexicons. The observations above highlight that accuracy in SA tasks is considerably better when employing a sentiment lexicon that is more domain specific. Nevertheless, due to the limited availability of sentiment lexicons tailored to specific domains, there is a crucial requirement to create lexicons for numerous and varied domains. This insight has underscored the necessity for an automated model for generating sentiment lexicons across various other domains.

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Furthermore, the process of labeling data necessitates a considerable investment of time and manual effort. Consequently, expanding supervised approaches to cover multiple domains poses a challenge. Recent semi-supervised methods have addressed the scarcity of labelled data by incorporating techniques such as combining labelled and unlabelled data, further unsupervised approaches include sentiment topic-model, self-training, and co-training. Sentiment lexicon knowledge graph propagation is another method for tackling the lack of labelled data. Meanwhile the researchers, while exploring multiple domains and transfer learning approaches, noted that the transfer of knowledge is more successful between domains that share their similarity than the transfer amidst the dissimilar ones. Domains with similar word distribution exhibit more similarity compared to domains with dissimilar word distribution.

1.3 Scope of the Research

Exploration and investigation helped to set a research outline and primary research goals. This research work aims to investigate and explore the observed multiple challenges. The most important challenge is to curb the labelled input and investigate alternate approaches such as exploring the seed word-based approach. The second challenge is to develop a mechanism for learning sentiment lexicon using a process that can be scaled to multiple domains. In-domain sentiment lexicon building approaches restrict the scope to domain-level learning, but the general-purpose lexicon approaches build lexicons that are too generalize for domain-based tasks. The aim is to combine and have the best of both approaches while confining it to a limited number of domains rather than generalizing. Twitter comments have restrictions on the count of words, which reflects in the way it is communicated; blogs usually use formal language and are informative, etc. The objective of this research is to leverage on the relationships between domains within the similar genre and effort up to construct a framework equipped with a scalable mechanism for learning more related domain-specific sentiment lexicons.

1.4 Objectives

The following are our work objective:

1. Provide a comprehensive review of existing literature, providing scholars with a thorough overview of the tools and resources available to them for multimodal SA.

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2. A summary of the main fusion techniques used to combine audio, video, and text data should be provided. Include a summary of pertinent previous research and provide insights into the classification strategies used for sentiment classification.

3. To propose a lexicon based approach using acronyms, emoticons and sentiments to classify the tweets and to improve the accuracy.

4. To propose a lexicon based approach using acronyms with negations to improve the accuracy.

5. To propose a best sentiment score measurement model to improve the accuracy.

2. Related Works

The concept of SA was introduced in works, and Opinion mining was coined freshly. Early studies on SA concentrated on goods and films, especially when it came to predicting the adjectives' Sentiment Orientation (SO) [16]. A significant method employed in this context is the use of Point Mutual Information (PMI) to determine how sentimentally oriented a phrase is. The utilization of supervised learning approaches using a variety of n-gram feature sets has been crucial in attaining an 83% accuracy rate in binary sentiment classification of documents, with a particular emphasis on unigram features [17]. As SA research progressed, it expanded into different domains, encompassing blogs and news articles. Notably, early research efforts were primarily concentrated on analysing longer documents, user blogs and also in movie reviews too [18].

Researchers explored the idea that it's usually simpler to categorize sentiment in shorter documents than in lengthier ones, yet improving performance in sentiment classification for micro text presents a significant challenge. The study specifically investigated the use of positive and negative emotions, and hashtags in tweets, for sentiment labels, serving as sample components. With blogs becoming an increasingly popular form of communication, has proven to be a valuable resource for SA [19]. Bloggers often express their feelings and emotions towards various entities, and the content of blogs commonly includes comments and reviews of goods, occasions, or offerings. Services for microblogging notably exemplified by Twitter and others, serve as abundant sources for researchers in the field of opinion mining. These platforms provide a wealth of data to understand public sentiment. Furthermore, microblogging systems facilitate social media interactions and marketing tactics [20].

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The research explores into the use of emoticon analysis to classify tweets as either good or bad. SA tasks are categorized according to their levels of use, specifically, feature, document, phrase, and word levels word-level SA is very useful for tweets. Despite Twitter posts include a variety of symbols, abbreviations, and partial phrases that do not follow rigorous grammatical norms, SA on a word level proves effective. These complexity lead to a unique lexicon and provide difficulties for tweet SA experts [21]. A model "bag of words" emerges as the simplest solution for SA, streamlining information retrieval and text mining processes. However, it is important to note that this simpler model treats text as an unordered set of words, fails to maintain the words of sentences in their proper sequence. Research has explored convolutional networks are used for sentence-level SA [22].

This highlights the importance of fine-tuning these parameters based on the unique features and requirements of the SA task and dataset at hand. For the automatic extraction of sentiment from text, lexicon-based and text classification approaches are the two main approaches that researchers typically use [23]. In the former, the SO of documents can be determined in calculating the sentiment values associated with words or phrases. Often called a statistical or machine learning approach, the latter builds classifiers using annotated examples of sentences or text. Text polarity can be detected with remarkable accuracy by classifiers, especially those that use on supervised machine learning techniques [24]. However, the specific domain dependent on it affects how well machine learning classifiers work. The flip side, the lexicon-based approach has enhanced performance across various domains. It involves calculating sentiment values from words or phrases and can be easily enhanced by incorporating additional knowledge sources for sentiment classification. This feature makes lexicon-based methods suitable for cross-domain SA [25].

The proficiency in managing contextual valence shifters but also in effectively analysing both blog posts and video game reviews is carried out in the lexicon-based method. In the lexicon-based method, adjectives are commonly employed as SO indicators. The process typically includes compiling SO values into a dictionary [26]. Subsequently, the whole list of adjectives found in the target text is recognized, and each is given the appropriate SO value based on the specified dictionary. Although tweet scoring based on lexicons and the R method have been explored, some studies rely on a straightforward average sentiment score in their work [27]. To categorize tweets into positive, neutral sentiments, negative sentiment a hybrid approach is used. Although the system's effectiveness is demonstrated through a case study, the study falls

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

short of addressing the role of slangs in sentiment classification. Notably, research on the study of online slang for SA is scarce. This study uses a lexicon-based strategy to classify tweets' sentiment by combining a variety of lexicons and dictionaries. The study also includes identifying and rating slang terms and acronyms that are used in tweets. The outcomes are then presented into various categories like negative, neutral sentiments or positive sentiments [28].

Utilizing one of the basic methods for SA is a lexicon. This approach requires a vocabulary with terms that are both positive and negative each of it assigned to a corresponding sentiment value. A range of methods for developing dictionaries has been proposed, involving both manual and automated approaches [29].

After representing the message in this way, each word or phrase in the message is assigned a sentiment value based on dictionary definitions, both good and negative. Machine learning models are hindered by their dependence on labeled data, which can be challenging to acquire in sufficient quantities and with accurate annotations. Additionally, the advantage of a lexicon-based approach lies in its ease of comprehension and modification by a human, a significant factor in our decision [30][40]. We discovered it was more straightforward to develop a suitable lexicon than to gather and label a pertinent corpus [39]. Considering that the general social media data is sourced from users globally, the algorithm's effectiveness is limited if it can only handle the English language. Therefore, it is crucial for SA algorithms to be readily adaptable to various languages. In subsequent sections in our paper, we will explore how a SA technique based on lexicons can be effectively translated for use in several languages by utilizing string similarity functions [31].

In the context of machine learning, a supervised method is been applied when dealing with a finite set of classes such as positive, neutral and negative. This approach relies on labelled data for training the classifiers effectively. On the other hand, unsupervised techniques are employed in situations where obtaining labelled training documents proves challenging. Unsupervised learning, in contrast, does not necessitate prior training for text mining. Specifically, in document-level SA, unsupervised processes primarily focus on determining the SO of particular phrases that lies in the document [32].

The process of recognizing emotions from the data in a text form is broadly employed method for categorizing people's feelings, representing a rapidly advancing field within Natural Language Processing (NLP). This analysis of text not only aids companies in understanding

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

their customers' needs but also assists political parties in conducting opinion polls. Researchers have identified a variety of emotions in text automatically and analysed their significance in predicting sentiments from documents [33][38].

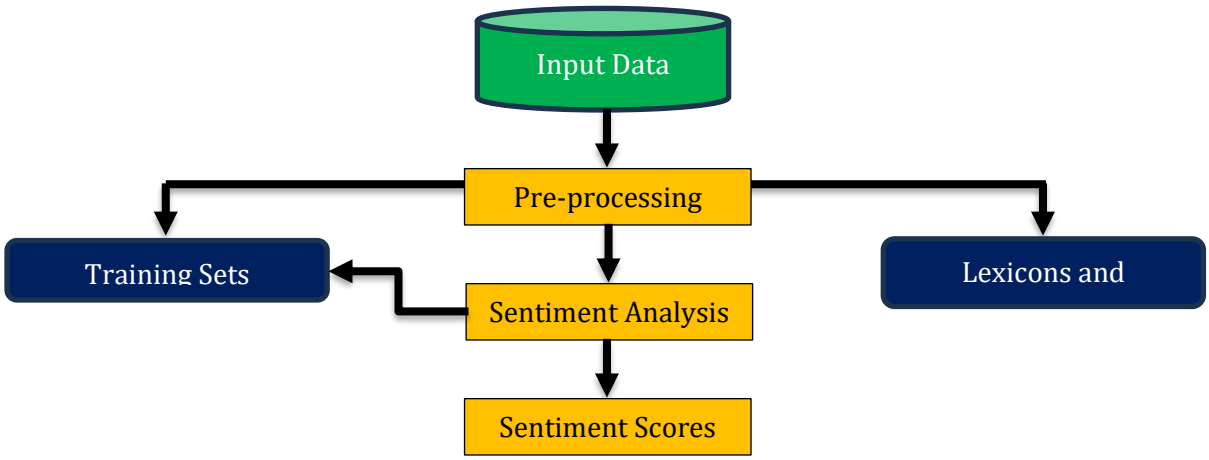


Figure 2: Basic building blocks of SA

Figure 2 illustrates the various categories of emotions expressed by individuals. Numerous research studies have been conducted to identify the sentiments conveyed by various words and sentences within documents. According to the authors, SA in text is primarily grounded in sentiment lexicons, with some approaches incorporating specific rules. SA based on text uses both supervised and unstructured methods. Many predictions are based on the analysis of sentiment in text for applications like forecasting box-office revenue, anticipating election results predictions also includes predicting stock market performance.

3. Proposed System

The process begins with the gathering of text, video/image data and audio which is then organized into distinct datasets. Pre-processing is carried out after data collection, and then the feature extraction stage is carried out, in which relevant features are extracted from the incoming data. A single feature vector for sentiment classification is formed by consolidating the features pull out from various audio, video/image data and textual data for sentiment classification. The determination of sentiment polarity, encompassing positive, negative, or neutral, is done in the final step, relying on which of these two methodologies can be used either with the help of machine learning or by techniques of DL. Furthermore, there is a summarization of available databases for the execution of analysis of sentiments across multiple modes is as depicted in Table 1.

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 1: Databases related to the analysis of sentiments across multiple modes.

Sl.No	Name	Weblink
1	MOUD	web.eecs.umich.edu
2	CMU-MOSI	www.amir-zadeh.com
3	ICT-MMMO	multicomp.cs.cmu.edu
4	CMU-MOSEI	multicomp.cs.cmu.edu
5	IEMOCAP	sail.usc.edu
6	EmoReact-Children	multicomp.cs.cmu.edu
7	Emotion Dataset MVSA-Multiview Social Dataset	mcrlab.net/multi-view-social-data/
8	RECOLA VAM	diuf.unifr.ch sail.usc.edu
9	RAVDEES	zenodo.org
10		

Conversion of Audio, video and image to text [41]

Optical Character Recognition (OCR) is a technology used to convert different types of documents, such as scanned paper documents, PDF files, or images captured by a digital camera, into editable and searchable data. Here's how OCR generally works:

Image Preprocessing: The input image is preprocessed to enhance its quality and improve OCR accuracy. This preprocessing may involve tasks such as noise removal, skew correction, binarization (converting the image to black and white), and deskewing (straightening tilted text lines).

Text Detection: The OCR system identifies regions of the image that contain text using techniques such as edge detection, connected component analysis, or machine learning-based object detection algorithms.

Character Segmentation: Within the text regions, individual characters are segmented from each other to prepare them for recognition. This step is particularly important when dealing with handwritten text or documents with overlapping characters.

Character Recognition: Each segmented character is analyzed and recognized by comparing its features with a predefined set of character templates. OCR systems may use various techniques

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

for character recognition, including template matching, feature extraction, neural networks, or statistical models like Hidden Markov Models (HMMs).

Text Reconstruction: Recognized characters are combined to form words, sentences, and paragraphs, reconstructing the text content of the input image.

OCR technology has numerous applications across various industries, including:

Document digitization: Converting printed documents into digital formats for archiving, indexing, and searchability.

Text extraction: Extracting text from images or PDF files for data mining, information retrieval, or content analysis.

Handwritten text recognition: Recognizing handwritten text in forms, documents, or historical manuscripts.

Automatic number plate recognition (ANPR): Extracting text from vehicle license plates for applications like traffic management, toll collection, and law enforcement.

Popular OCR software and libraries include Tesseract OCR (an open-source OCR engine developed by Google), Adobe Acrobat OCR, Amazon Textract, Microsoft Azure Cognitive Services, and Google Cloud Vision API. These tools provide APIs and SDKs for integrating OCR capabilities into applications and workflows.

Post processing: Finally, the recognized text undergoes postprocessing to correct errors, improve formatting, and enhance overall accuracy. Postprocessing techniques may include spell-checking, language modelling, context-based correction, and dictionary lookup. Table 2 provides characteristics of three benchmark datasets

- Stanford Sentiment Treebank (SST-1): An extension of MR includes labels as very positive, positive, neutral, negative, very negative.
- MR: Movie Reviews (MR) classified into either positive or negative categories.
- SST-2: Similar to SST-1, except neutral reviews are omitted and binary labels are used.

This includes converting cases labelled as "very positive" to "positive" and cases labelled as "very negative" to "negative."

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 2: Characteristics of three benchmark datasets process

Dataset	T	I	N	V	V_{pre}	Test
MR	2	21	10653	18766	18778	CV
SST-1	5	19	11876	17841	17842	2212
SST-2	2	20	9623	16199	16188	1835

Note: T – Target class; I – Length of sentences (Average); N- Size of dataset; V – Size of Vocabulary; V_{pre} – Word sector pretrained (count); Test- Size of test set

The combination of the three modalities mentioned above is used for sentiment classification. Different stages involved in SA using multimodal data shown in Figure 3.

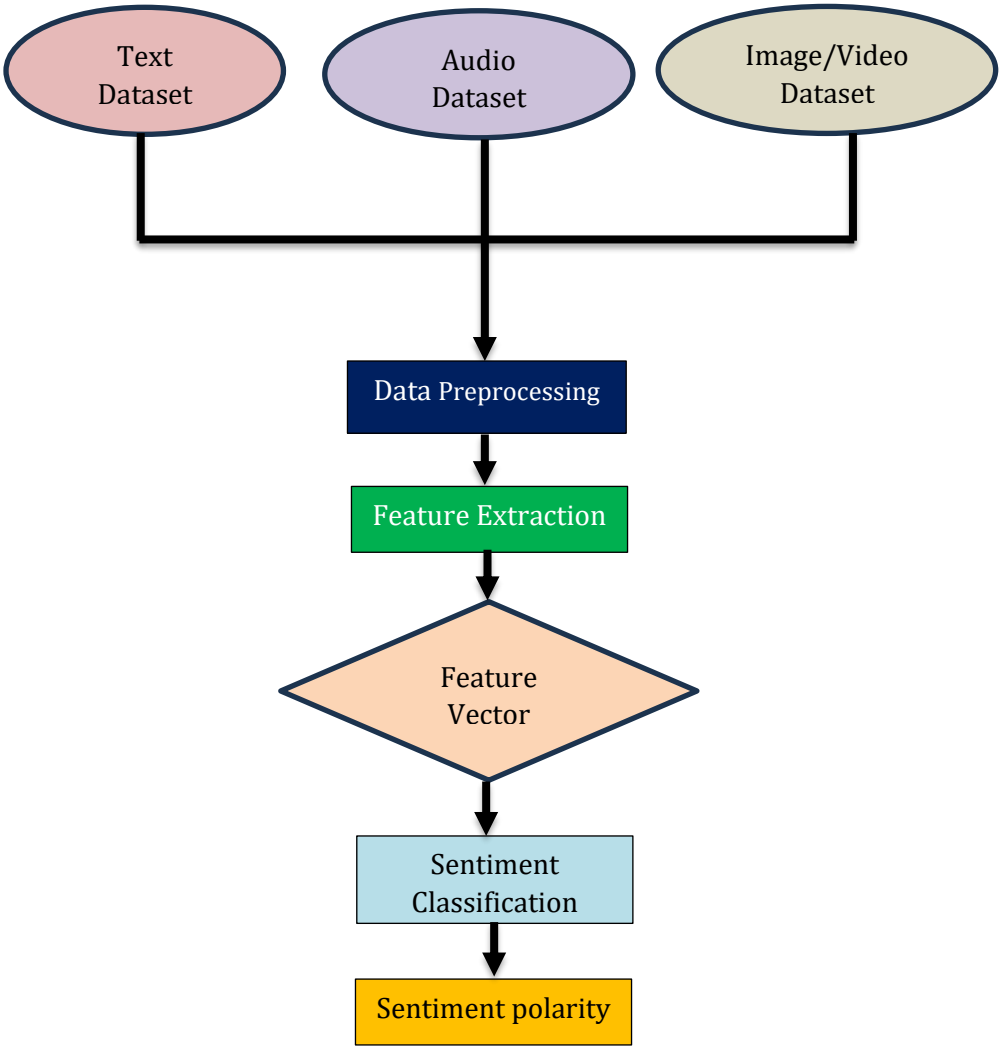


Figure 3. Representation of SA at various levels

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.1 Pre-processing

It is a crucial step in data mining to ensure accurate and meaningful results. To avoid inaccurate and deceptive results, it is essential to pre-process the data before analysis. Pre-processing steps employed as given below:

- Exclude non-English tweets that are irrelevant from the dataset.
- Deletion of repeated or same tweets in the dataset.
- Elimination of punctuation marks, numeric expressions, and stop words, during pre-processing.
- Using a single space character in place of many spaces.
- Correction of spelling by replacing characters repeated more than twice any word with one or two times, where possible. Repeated characters are typically used for emphasis.
- Identification of words that are capitalised, often used for expressing strong feelings.
- Removal of RT (retweet) and hash tags (#) symbols.
- Replacement of all tokens (begins with "http://", "https://", "http:", "http", or www)
- Replacing negations

3.2 Multimodal Fusion Techniques

The integration of features from text, audio, and video modalities is a crucial step before going forward on the classification task. The overall precision of the result is improved through the multimodal fusion process, which incorporates additional information.

3.2.1 Early fusion (Feature level)

The early fusion method, also known as feature level, combines features from different varied modalities to create a single, unified feature vector. The primary advantage of this strategy lies in its ability to ensure precise results by identifying the early-stage correlations among features in multimodal data. When compared to existing approaches, feature-level fusion produced the best results for unimodal fusion and showed an improvement in processing speed. Utilization of the early fusion to integrate facial colour, head movement and physiology for affect classification, leading to statistically higher accuracy when taken its compared with unimodal approaches.

Received: August 04, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.2.2 Decision level or late fusion

The decision level approach focuses on the features from each modality where they undergo a independent classification before being combined. A distinct classification is then carried out for every modality. This strategy has the advantage of allowing each modality to learn its own characteristics by using the best classifier for it. However, this method is known for its additional time consumption due to the utilization of different classifiers. When it comes to sentiment prediction, the researchers have shown that late fusion yields better results than its other fusion methodology, such as early fusion.

3.2.3 Hybrid fusion

Combining the strengths of both feature and decision-level fusion techniques, the hybrid fusion approach is used for enhancing sentiment prediction tasks. Through the fusion of modalities, the researchers achieved an F1-measure of 65.7%, surpassing the unimodal score. They accomplished this by combining multimodal information, applying sentiment ontology to videos and spectrogram characteristics to voice. The performance outcomes were improved by the integration of input modalities since each modality provided complementing information to the others. In the process to identify human emotions from our speech and the varied facial expressions, the researchers developed a model that incorporates a area of hybrid feature. This fusion has outperformed unimodal-based approaches, showcasing superior performance.

3.3.3 Model level fusion

The recognition was used to evaluate the efficacy of this proposed method, and the results were impressive even when there was an audio noise. The study produced improved results for emotion recognition using audio-visual data obtained from a variety of people, each of whom displayed distinct affect states. The outcomes display its dominance over other prevailing unimodal methods.

3.3.4 Rule-based fusion

Multimodal fusion is accomplished by Rule-based Fusion using techniques like majority voting and weighted fusion. When combining features from different modalities, operations like multiplication or summation are used (weighted fusion). By using this method the process is cost-effective, assigning normalized weights to individual modalities. However, a challenge is ensuring the proper normalization of weights for improved execution. In majority voting-based

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

fusion, the decision made by the majority of classifiers is crucial. The authors combined speech and 2D gestures attained while interactions with a system for gaming purposes. They made a computer system that can understand and respond to the gestures people make. To identify monologues in videos, the authors used linear weighting of face scores and normalized speech models. They came to the conclusion that the product weighting strategy might not perform and the weighted sum method.

3.4 Multi-Aspect Sentiment Analysis of social media text using Lexicon based FasText

The sentiment lexicon, that was developed manually with around 6300 words was created using SentiWordNet as a reference point. Each word in the lexicon is assigned a sentiment value ranging from −100 (most negative) to 100 (most positive). From experience some words that seem positive or negative might actually mean something neutral in a sentence. Determining sentiments in such cases can be challenging. To handle this, we figured out a conditional probability (marked as P) for every word in the lexicon, together with its sentiment value, as described in Equation (1).

$$P(w) of\ positive\ w \qquad (1)$$

$$P(w) of\ negative\ w \qquad (2)$$

To determine the likelihood of each positive word, we employed a set of labelled data. The likelihood that a message chosen at random and containing that word is positive is indicated by this probability. Similarly, for negative words also the method is applied. The aim of the investigation was to determine if the integration of this data with the sentiment classification procedure may enhance the processing of messages that contain a combination of positive and negative attitudes. Based on whether or not emoticons were present in a communication, the training set was automatically created. The Twitter API was queried with a list of emoticons, and the messages that were gathered were automatically classified as positive or negative depending on the kind of emoticon they included. Then, as shown in Equation (3), the conditional probability was calculated based on whether the term was positive or negative.

$$P(w) = \frac{P(positive \cap w)}{P(w)} = \frac{\#w_P}{\#w} \qquad (3)$$

$$P(w) = \frac{P(Negative \cap w)}{P(w)} = \frac{\#w_N}{\#w} \qquad (4)$$

Received: August 04, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Here is the representation of these variables, $\#w_P$ represents the quantity of positive messages with the word "w" in the sample and $\#w_N$ represents the number of messages containing the word w that are negative. Our method involves creating a corpus-based lexicon with sentiment terms from language and content domains. It is imperative that this corpus has a high degree of relevance to the language and content domains. The language domain is concerned with the lexical and syntactical decisions made in language, whereas the content domain is concerned with the discourse environment. Our approach is concentrated on maintaining this kind of data in the freshly created sentiment lexicon.

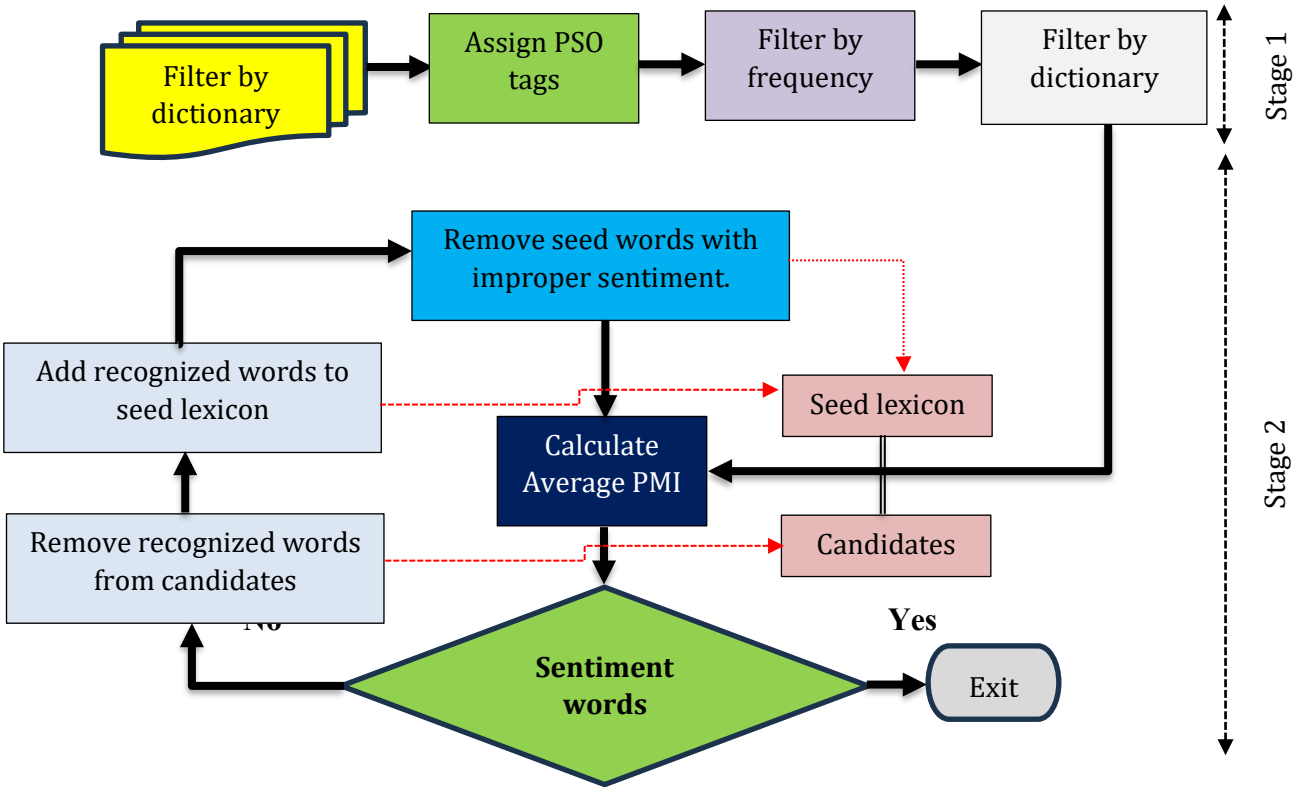


Figure 4: The proposed lexicon generation method

Figure 4 depicts our approach; including two phases namely the 1.candidate extraction and 2. Sentiment recognition. The details of all the steps involved in our method:

1) The initial step is extracting the candidates from the corpus, the developing corpus.

2) Identify how the candidate words are connected or related to the sentiment words in the seed lexicon.

3) Evaluate the leaning of each candidate towards sentiment, incorporate words that mirror that sentiment into the seed lexicon, and eliminate terms that deviate from that sentiment.

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

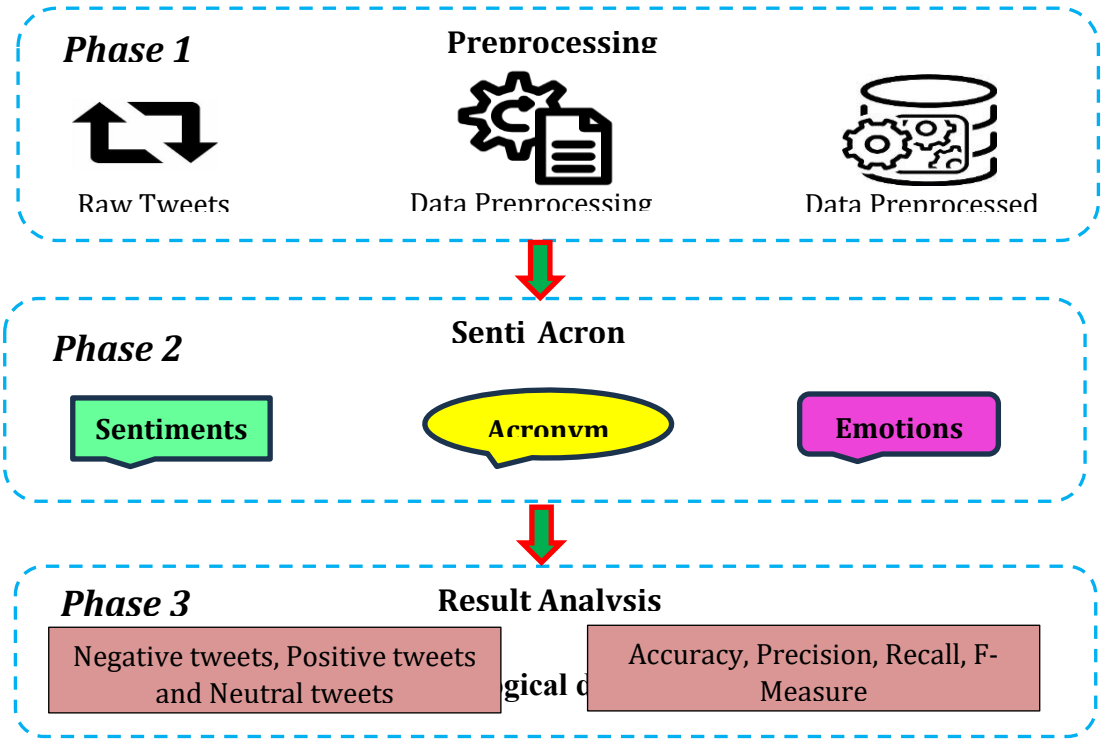
4) Then it starts by identifying the connection as in Step 2 then followed by step 3, this is reiterated till no additional new words are included

In general lexicon-based approaches, in a sentence, negation is often handled by flipping the polarity of the lexical item next to the negator (e.g., good: 100 and not good: −100). We proposed a fresh strategy in our research.

$$F_N(S) = \{max\left\{\frac{s+100}{2}, 10\right\} if\ S < 0\ min\left\{\frac{s-100}{2}, -10\right\} if\ S > 0 \qquad (5)$$

Here F_N - final negation , S - sentiment value from the lexicon.

In recent times, Facebook Research released an open-source word embeddings project called FastText. FastText is designed for efficient and rapid acquisition of word representations, enhancing text classification. Unlike merely instructing users on how to interpret word representations, FastText embedding's consider a word's intrinsic structure. This proves particularly beneficial for languages with intricate morphology, ensuring autonomous learning for the representations of various word structural forms. This attribute develops learning on languages with extensive inflections.



Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Meanwhile the entire procedure of the pre-processing and proposed approach has broadly been categorized into three phases. Figure 5 exhibits the methodological diagram of the three phases. The phases are explained below:

Phase 1: This phase provided an explanation of data extraction. The gathered tweets undergo pre-processing and are saved in both.txt and.csv formats. The pre-processing stage consists of stemming, negation, user name, stop word removal, and URL removal. The ensuing subsections provide a brief explanation of each step.

Phase 2: A brief explanation of the Senti_Acron Approach is provided in the section dedicated to managing sentiment, acronyms, and emoticons. The tweets have been categorized with the help of a mathematical formula.

Phase 3: This session results section classifies and discusses the tweets that are good, negative, and neutral. To evaluate the accurateness of the Senti_Acron method, polarity metrics that are frequently employed are calculated, including accuracy, precision, F-measure and recall.

Algorithm

Input: Reviews or tweets taken from social media

Output: Sentiment Score

Step 1: Initialize Intensifiers List (IniL); Negations List (NegL)

{

Step 2: Score_Sent(tweet/reviews)

Step 3: Pre-processed the tweet/reviews as ptext

Step 4: Assign tokenize (ptext) into token

Step 5: if (Tasks)

{

Step 5.1: Case 1: if word is in NegL then polarity of word+1 reverse

Step 5.2: Case 2: if word is in IniL then polarity of word+1 modify

Step 5.3: Case 3: if words (all letters) are in upper case then fraction added to word score

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Step 5.4: Case 4: Word score enhanced if repeated letters present

Step 5.5: Case 5: Exclamation count exclam(ptext) as Xc

}

Step 6: f**or** finding word in tokens

{

Step 6.1: i**f** emoticons words then assign emoticon score to score

}

E**lse**

{

Step 6.2: Searching an opinion lexicons in dictionaries

Step 6.3: If word in lexicon based Fastext assign score

{

Step 6.3.1: Assign lexicon based Fastext score to score

Step 6.3.2: .do from task 5.1 to 5.4

}

Step 6.4: If word not found in lexicon based Fastext assign score then check antonyms, synonyms and assign score.

{

Step 6.4.1: Assign lexicon based Fastext score to score

Step 6.4.2: .do from task 5.1 to 5.4

}

Step 6.5: If word not found in lexicon based Fastext then check in SentiWordNet and compute its score.

{

Step 6.5.1: Assign SentiWordNet score to score

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Step 6.5.2: .do from task 5.1 to 5.4

}

Step 6.6: If word not found in SentiWordNet then search in either Slang's dictionary or Web and compute its score.

{

Step 6.6.1: Assign Slang's score to score

}

Step 6.7: If word not found then assign score as zero

}

Step 7: increment Tweetscore as Tweetscore + score

Step 8: Compute score (Xc+1)/2* tweetscore

}

3.5 Feature Selection

Variable selection or attribute selection are other names for Feature Selection (FS). The process of choosing and building the features of Item ID, Sentiment Source, and Sentiment Text is known as FS. The subset features are relevant to the FS model. Thus, the unigram feature is chosen in this work to find the sentiments from the textual data on the data set shown in Table 3.

Table 3: Example of emoticons

Positive Emoticons	Negative Emoticons	Neutral Emoticons
<3	:(	:/
:)	:-(	:P
:-)	=(	;-)
=]	=/	:0
:D	:@	:|
;)	>:(	:S

Received: August 04, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Algorithm demonstrates the pseudo-code for proposed method and its related processes and notations. Every tweet is regarded as "t" and the pre-processed tweets are seen as "T". The sentence level computation is the foundation of the approach. In the tweets, every word is regarded as 'w'. Initially apply unigram feature to select the sentiment individually, then it has matched with dictionary and replace the lexicon instead of acronyms and emoticons. The proposed formula, which is covered in the next section, is used to determine the polarity of sentiments. Based on the following characteristics, the proposed technique determines the sentiment's polarity.

The tweets are categorized by the scoring module as positive, negative, or neutral. The system is divided into smaller modules that include vocabulary scoring, slang, managing synonyms and antonyms, handling negations, intensifiers and, plurals and singulars addressing forms. During the first stage, each word in the lexicon, whether positive or negative, is given a number between +1 and -1. Emoticons were categorized as either good or negative, with a score of +1 or -1 assigned to each, respectively. SentiWordNet (SWN) is used to determine the sentiment score in cases where the word in question is not present in either the lexicon or the list of emonticons. Each Wordnet synset is assigned three numerical values by SWN: pos(w), neg(w), and obj(w).

$$obj(w) = 1 - [pos(w) + neg(w)] \quad (6)$$

Every item in SWN is multisensorial. A word is regarded as positive or negative if its obj(w) score is less than the threshold value of 0.5; otherwise, it is deemed objective and receives a score of zero.

$$pos\_score(w) = \sum_{x=1}^{n} \quad pos(p_x)/n \quad (7)$$

$$neg\_score(w) = \sum_{x=1}^{n} \quad neg(p_x)/n \quad (8)$$

$$obj\_\_score(w) = \sum_{x=1}^{n} \quad obj(p_x)/n \quad (9)$$

$$Score(S_x) = W_{sx} + E_{sx} + S_{sx} + L_{rx} \quad (10)$$

$$S_{sx} = \left[(pos\_score(S_{sx}) + pf) - (neg_{score(S_{sx})} + nf\right] \quad (11)$$

Where W_{sx} – xth word score, E_{sx} – emoticon and , S_{sx} - slang respectively. L_{rx} – xth word repeated letters. The fraction of positive and negative documents that contain the specified

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

slang is represented by the terms pf and nf, respectively. Equation (12) displays the whole formula for determining the emotion of the full document.

$$Score(tweet) = \frac{(1+I_c)}{2} * \sum_{x=1}^{|t|} \quad Score(S_x) \qquad (12)$$

Where $I_c \; - \; count\ of\ exclamations\ in\ the\ tweet$ of sentiment scoring and |t| - the length of number of tokens in tweet.

4. Results and Discussions

4.1 Performance Evaluation

A unique table designed to show the model's performance. The binary classification confusion matrix is displayed in Table 4. Precision, F-score, recall, and Matthew correlation coefficient (MCC) are only a few of the performance metrics used by researchers to assess the effectiveness of classifiers.

Table 4. Classification (binary) confusion matrix

Human says		**Machine Says**	
		Pos	**Neg**
	Pos	True Positive (TP)	False Negative (FN)
	Neg	False Positive (FP)	True Negative (TN)

$$Precision, p = \frac{TP}{TP+FP} \qquad (13)$$

$$Recall, r = \frac{TP}{TP+FN} \qquad (14)$$

$$F - Score = \frac{2rp}{r+p} + \frac{2*TP}{2*TP+FP+FN} \qquad (15)$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP+TN} \qquad (16)$$

Matthews Correlation Coefficient (MCC)

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \qquad (17)$$

Two sentiment classification situations were used in our experiments: pos against neg and positive against negative versus neutrality. To demonstrate that accuracy is closely correlated with lexicon size and coverage, a number of lexicons were utilized in the initial phase of SA

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

of tweets using a basic scoring system. The effects for pos against neg sentiment classification are presented in Table 5.

Table 5: Comparison of scoring accuracy of proposed and existing systems

Type	Accuracy
Lexicon-1	0.6435
Emoticons	0.3254
Lexicon-2	0.3456
Hybrid-2	0.7458
Hybrid-1	0.5164
Proposed system	0.7179

Lexicon-1 covers more sentiment words than Lexicon-2, the findings above demonstrate that Lexicon-1 is more accurate than Lexicon-2. In tweets and online evaluations, emoticons are frequently put to use. Using an integrated strategy based on lexicons, the experiment's second phase involves binary (pos versus neg) and multi-class ((pos versus neg versus neutrality) classification shown in Tables 6 and 7.

Table 6: Binary classification performance

Measure	Results		
	Overall	Pos	Neg
Accuracy	0.9254	0.9084	0.9631
Precision	0.9864	0.9861	0.8269
Recall	0.9052	0.9013	0.9614
F-score	0.9478	0.9458	0.8875
FPR	0.0563	0.0265	0.0248
MCC	0.8379	0.8565	0.8325

Table 7: Proposed system performance WO/S: Without Slangs, W/S: With Slangs

Measure	Results							
	Overall		Pos		Neg		Neutral	
	WO/S	W/S	WO/S	W/S	WO/S	W/S	WO/S	W/S
Accuracy	0.8864	0.8574	0.8964	0.8647	0.9625	0.9387	0.66	0.66
Precision	0.9851	0.9857	0.9851	0.9842	0.7508	0.7325	0.8247	0.6947

Received: August 04, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Recall	0.9245	0.9035	0.8968	0.8624	0.9627	0.9384	0.66	0.66
F-score	0.9254	0.9325	0.9357	0.9192	0.8461	0.8264	0.7325	0.6767

The binary sentiment classification yields an impressive total accuracy of 92%, characterized by a notably low false positive rate of only 5%. The Matthews Correlation Coefficient (MCC) for binary classification stands at 0.8204, indicating an enhanced predictive performance. There should be fewer false positives given the high precision for positive mood. As the datasets lack imbalanced classes, accuracy is utilized as the primary metric for assessment. A basic model utilizing static vectors, specifically CNN-static fastText, demonstrates remarkable performance, delivering competitive results compared to more intricate deep learning models that rely on sophisticated pooling schemes or necessitate pre-computed parse trees. The confusion matrices for the three tested datasets using dynamic fastText are illustrated in Figures 6-8. These matrices show the models' performance on a given dataset. The classifier performs better if the main diagonal is darker.

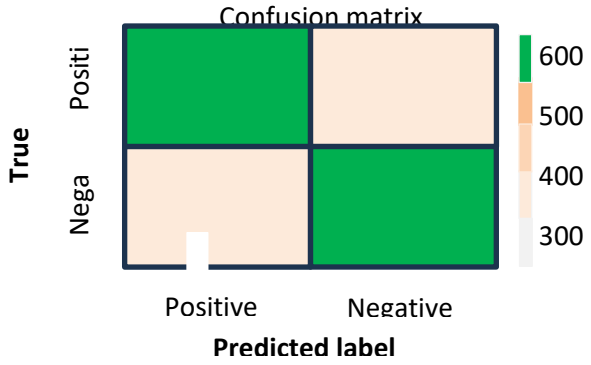


Figure 6: Dynamic FastText on MR dataset using proposed method

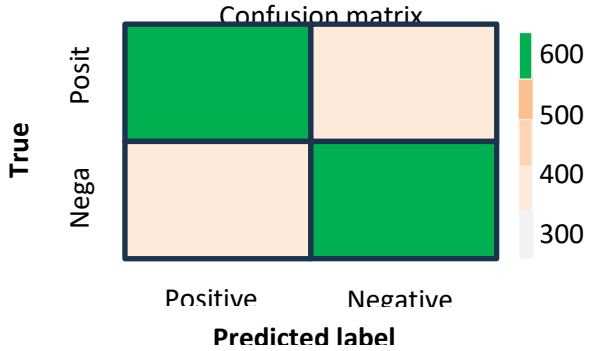


Figure 7: Dynamic FastText on SST-2 dataset using proposed method

Received: August 04, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

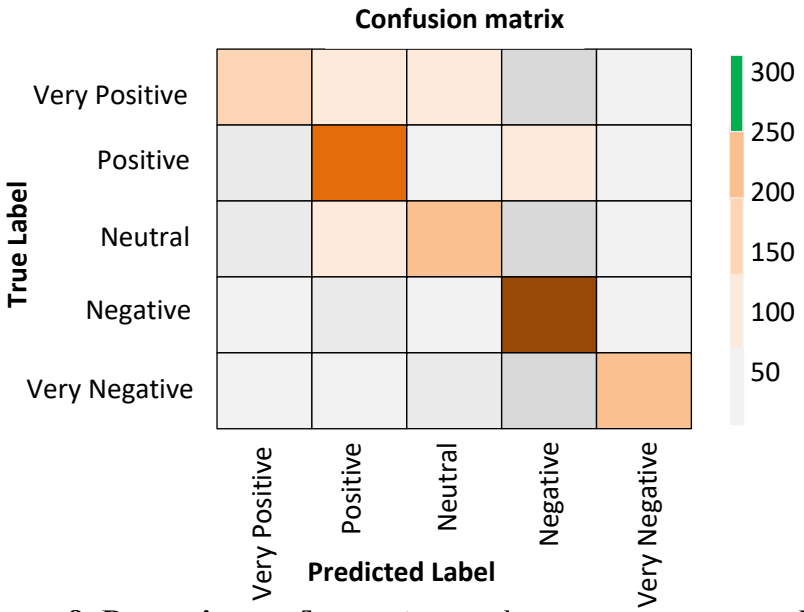


Figure 8: Dynamic Fast Text on SST-1 dataset using proposed method

Examined using the MR and SST-2 datasets, the model exhibits excellent performance. However, the great granularity of its classes makes the SST-1's confusion matrix less than ideal. The outcomes show that the pre-trained vectors may be used independently across datasets and are strong feature extractors. They can, however, also be adjusted in accordance with a particular task. Table 8 displays the tweets correctly classified using the raw data, while Figure 9 visually represents the distribution of correctly classified tweets. On the X-axis, we have the classified tweets, and the Y-axis represents the count of classified tweets. Table 9 and Figure 10 shows the values of MCC for different numbers of non-neutral words and values of p. Comparison of Number of words and Average sentiment shown in Figure 11.

Table 8: Correctly labelled tweets

Tweets	Total
Sentiment words	501457
Acronyms	206479
Emoticons	86578

Received: August 04, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

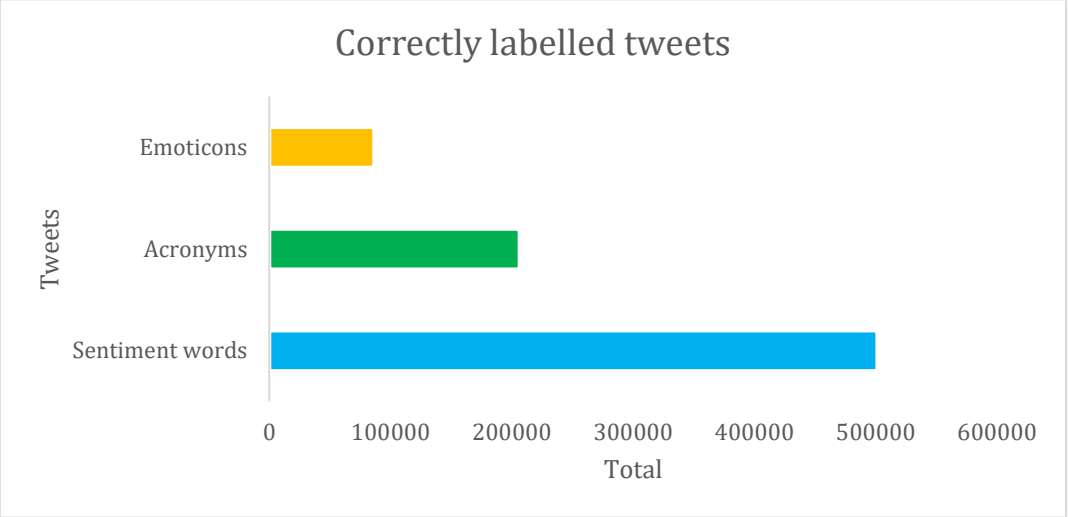


Figure 9: Correctly labelled tweets

Table 9: MCC for different numbers of non-neutral words and p value

		Values of p					
		2.00	2.5	3	3.5	4	4.5
No. of words	1	0.60	0.63	0.67	0.70	0.73	0.75
	2	0.73	0.78	0.83	0.88	0.92	0.97
	3	0.83	0.90	0.97	1.03	1.10	1.16
	4	0.92	1.00	1.01	1.18	1.27	1.35
	5	1.00	1.12	1.22	1.33	1.44	1.55

Table 10: Comparison of Number of words and Average sentiment

		Average sentiment				
		20	40	60	80	100
No. of words	1	15	28	42	56	70
	2	17	36	53	70	88
	3	20	42	62	83	100
	4	24	48	71	95	100
	5	27	54	80	100	100

Received: August 04, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

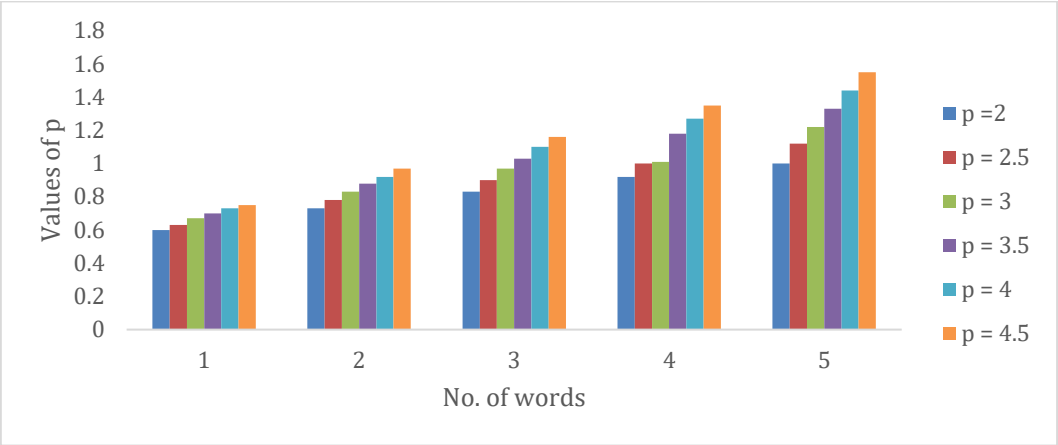


Figure 10: MCC for different numbers of non-neutral words and p value

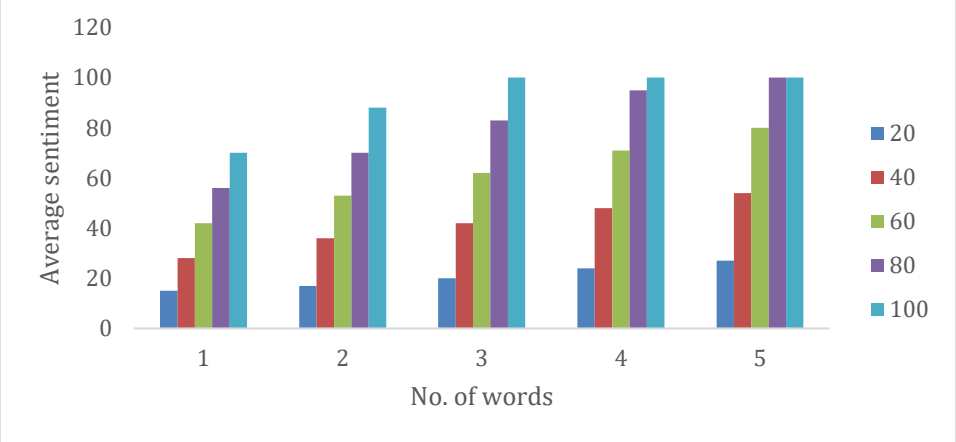


Figure 11: Comparison of Number of words and Average sentiment

$$F_P = min\left\{\frac{A_p}{2-loglog\,(3.5\times W_P+X_P)}, 100\right\} \quad (18)$$

$$F_N = max\left\{\frac{A_N}{2-loglog\,(3.5\times W_N+X_N)}, -100\right\} \quad (19)$$

Where X_P and X_N are the quantity of intensifiers used in a sentence. There is an overall increase in the value from 70 to 100 for an average feeling of 100. The Equation (20) for determining a sentence's overall good and negative sentiment after the aforementioned review. The Equation (21) is consistently used to combine evidence fragments for both positive and negative words when a message includes mixed sentiments. The maximum and minimum potential values were adjusted to 1 and −1, respectively, during the consolidation of evidence, rather than 100 and −100.

$$e_P = min\left\{\frac{A_p}{2-loglog\,(3.5\times W_P)}, 1\right\} \quad (20)$$

Received: August 04, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$e_N = min\left\{\frac{A_N}{2-loglog\,(3.5\times W_N)}, -1\right\} \qquad (21)$$

Where the terms "e_P" – overall positive and "e_N" - overall negative evidence respectively, in a statement. The concluding evidence indicating whether the context is favourable or unfavourable was determined by combining positive and negative data independently and evaluating the outcomes. The sentiment classification procedure took these two values into account. In Figure 12, the breakdown of labelled tweets is depicted according to sentiments, acronyms, and emoticons. Specifically, sentiment-related tweets comprise 65%, tweets with acronyms constitute 25%, and those featuring emoticons make up 10% of the overall distribution.

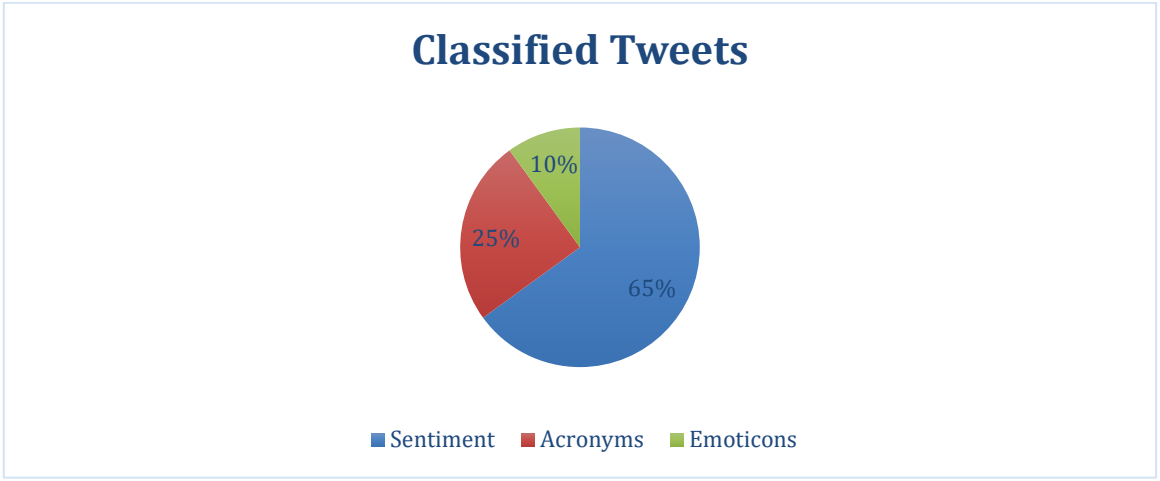


Figure 12: Percentage of classifier Tweets

Figure 13 portrays the performance measures of the gathered tweets. Notably, precision and recall yield identical outcomes, highlighting the usefulness of the proposed tactic.

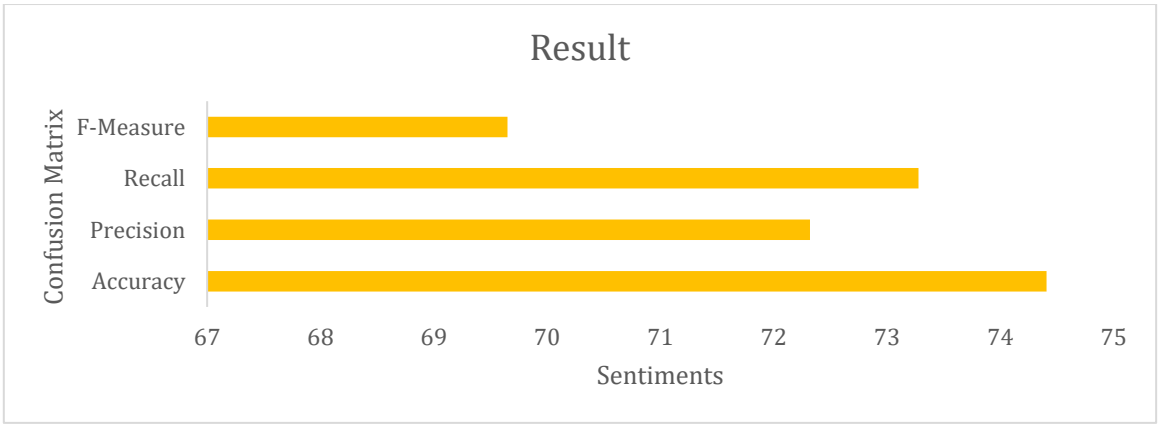


Figure 13: Performance measures of proposed system

Received: August 04, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5. Conclusions

We discovered the difficulties in sentiment categorization brought about by the language and content domains in this study. We proposed a lexicon extension technique to overcome these obstacles and enhance sentiment categorization through the acquisition of domain knowledge. As seeds and baselines, our methodology underwent testing on two substantial unannotated emerging corpora and was assessed against five pre-existing sentiment lexicons. The lexicon-based SA algorithm is enhanced without requiring a significant amount of human effort, which makes it useful for SA practitioners and researchers. Furthermore, it creates opportunities for new studies and uses of SA in areas where it was previously useless. Evaluation results indicate a substantial enhancement in sentiment classification performance when utilizing the expanded lexicons, as compared to the performance with the initial seed lexicons. In this research, we integrated lexicons and dictionaries to establish a framework for sentiment classification. In binary classification, we obtained 92% accuracy, and in multi-class classification, 87% accuracy. The accuracy in negative situations and recall in neutral cases need to be enhanced in the system.

6. Future Directions

The investigations of our study can also be expanded in diverse areas. The performance of the extended lexicon could be further enhanced by using a better seed lexicon. The measurement of word association was conducted through the use of PMI. Additional metrics, such as context similarity, could prove beneficial in this context. In addition to single words, the incorporation of bigrams and trigrams into the candidate set may also offer valuable insights. These candidates must be filtered using specific patterns. Other than dictionaries and growing corpora, domain-specific language resources could also be investigated. In our lexicon expansion procedure, Right now, we can distinguish between good and negative words. In order to improve the process's overall efficacy, we want to investigate the additional neutral word identification in subsequent research. We intend to experiment with additional datasets in the future to broaden this approach. Many advances in research and practice are brought about by our work.

Received: August 04, 2025

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

References

1. Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon-and machine-learning-based approach. *Social Network Analysis and Mining*, *13*(1), 31.
2. Ainapure, B. S., Pise, R. N., Reddy, P., Appasani, B., Srinivasulu, A., Khan, M. S., & Bizon, N. (2023). Sentiment Analysis of COVID-19 Tweets Using Deep Learning and Lexicon-Based Approaches. *Sustainability*, *15*(3), 2573.
3. Chouhan, K., Yadav, M., Rout, R. K., Sahoo, K. S., Jhanjhi, N. Z., Masud, M., & Aljahdali, S. (2023). Sentiment Analysis with Tweets Behaviour in Twitter Streaming API. *Comput. Syst. Sci. Eng.*, *45*(2), 1113-1128.
4. Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P. M. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, 119862.
5. Kukkar, A., Mohana, R., Sharma, A., Nayyar, A., & Shah, M. A. (2023). Improving Sentiment Analysis in Social Media by Handling Lengthened Words. *IEEE Access*, *11*, 9775-9788.
6. Catelli, R., Pelosi, S., Comito, C., Pizzuti, C., & Esposito, M. (2023). Lexicon-based sentiment analysis to detect opinions and attitude towards COVID-19 vaccines on Twitter in Italy. *Computers in Biology and Medicine*, *158*, 106876.
7. Rita, P., António, N., & Afonso, A. P. (2023). Social media discourse and voting decisions influence: sentiment analysis in tweets during an electoral period. *Social Network Analysis and Mining*, *13*(1), 46.
8. Thangavel, P., & Lourdusamy, R. (2023). A lexicon-based approach for sentiment analysis of multimodal content in tweets. *Multimedia Tools and Applications*, 1-24.
9. Jain, R., Kumar, A., Nayyar, A., Dewan, K., Garg, R., Raman, S., & Ganguly, S. (2023). Explaining sentiment analysis results on social media texts through visualization. *Multimedia Tools and Applications*, 1-17.
10. Sapkale, P., Bansal, P., Mukhedkar, M., & Sharma, S. (2023, April). Sentimental Segregation for Social Media Using Lexicon Technology. In *Proceedings of 3rd International Conference*

Received: August 04, 2025

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

on Artificial Intelligence: Advances and Applications: ICAIAA 2022 (pp. 417-426). Singapore: Springer Nature Singapore.

11. Azhar, A., Masruroh, S. U., Wardhani, L. K., & Okfalisa, O. (2023). Performance comparison of the Naive Bayes algorithm and the k-NN lexicon approach on Twitter media sentiment analysis. *Science, Technology and Communication Journal*, *3*(2), 33-38.

12. Eljil, K. S., Nait-Abdesselam, F., Hamouda, E., & Hamdi, M. (2023). Enhancing Sentiment Analysis on Social Media with Novel Preprocessing Techniques. *J. Adv. Inf. Technol*, *14*(6), 1206-1213.

13. Rahman, H., Tariq, J., Masood, M. A., Subahi, A. F., Khalaf, O. I., & Alotaibi, Y. (2023). Multi-tier sentiment analysis of social media text using supervised machine learning. *Comput. Mater. Contin*, *74*, 5527-5543.

14. Guo, B., Westland, S., & Lai, P. (2023). Sentiment analysis based on frequency of color names on social media. *Color Research & Application*, *48*(2), 243-252.

15. Sherif, S. M., Alamoodi, A. H., Albahri, O. S., Garfan, S., Albahri, A. S., Deveci, M., ... & Kou, G. (2023). Lexicon annotation in sentiment analysis for dialectal Arabic: Systematic review of current trends and future directions. *Information Processing & Management*, *60*(5), 103449.

16. Astarkie, M. G., Bala, B., Bharat Kumar, G. J., Gangone, S., & Nagesh, Y. (2023, January). A Novel Approach for Sentiment Analysis and Opinion Mining on Social Media Tweets. In *Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 2* (pp. 143-151). Singapore: Springer Nature Singapore.

17. Mohamed, A., Zain, Z. M., Shaiba, H., Alturki, N., Aldehim, G., Sakri, S., ... & Zain, J. M. (2023). Lexdeep: Hybrid lexicon and deep learning sentiment analysis using Twitter for unemployment-related discussions during covid-19. *Computers, Materials & Continua*, *75*(1), 1577-1601.

18. Sanjaya, A. D., Pudjiantoro, T. H., Ningsih, A. K., & Renaldi, F. Sentiment Analysis Of E-Wallets on Twitter social media With Naïve Bayes and Lexicon-Based Methods.

19. Zuhanda, M. K., Syofra, A. H. S., Mathelinea, D., Gio, P. U., Anisa, Y. A., & Novita, N. (2023). Analysis of twitter user sentiment on the monkeypox virus issue using the nrc lexicon. *Jurnal Mantik*, *6*(4), 3854-3860.

Received: August 04, 2025

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

20. Aqlan, W. M. M., Ali, G. A., Rajab, K., Rajab, A., Shaikh, A., Olayah, F., ... & Mangantig, E. (2023). Thalassemia Screening by Sentiment Analysis on Social Media Platform Twitter. *Computers, Materials & Continua*, *76*(1).

21. Omuya, E. O., Okeyo, G., & Kimwele, M. (2023). Sentiment analysis on social media tweets using dimensionality reduction and natural language processing. *Engineering Reports*, *5*(3), e12579.

22. Gunavathie, M. A., Muthulakshmi, S., Janaki, V., & Praveena, M. (2023, June). Exploring Machine Learning Techniques for Twitter Sentiment Analysis to Identify Polarity-A Comprehensive Study. In *2023 8th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1-5). IEEE.

23. Jeyakarthic, M, & Leoraj, A., A Novel Social Media-Based Adaptable Approach for Sentiment Analysis Data, *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Trichirappalli, India, 2023, pp. 1-6, doi: 10.1109/ICEEICT56924.2023.10156976.

24. Bala, D. E., & Stancu, S. (2023, January). Using Twitter Data and Lexicon-Based Sentiment Analysis to Study the Attitude towards Cryptocurrency Market and Blockchain Technology. In *Education, Research and Business Technologies: Proceedings of 21st International Conference on Informatics in Economy (IE 2022)* (pp. 187-198). Singapore: Springer Nature Singapore.

25. PERMANA, A. A., & NOVIYANTO, W. A. (2023). Comparison of the Accuracy of the Lexicon-Based and Naive Bayes Classifier Methods To Public Opinions About Removing Masks on Social Media Twitter. *Journal of Theoretical and Applied Information Technology*, *101*(3), 1174-83.

26. Abdelhady, N., Elsemman, I. E., Farghally, M. F., & Soliman, T. H. A. (2023). Developing Analytical Tools for Arabic Sentiment Analysis of COVID-19 Data. *Algorithms*, *16*(7), 318.

27. Ramin, J. (2023). Optimizing Lexicon-Based Sentiment Analysis for COVID-19 Twitter: Interactions in Health Contexts.

28. Oetama, R. S., Yanfi, Y., UP, M. M. I. A., & Isa, S. M. (2023, February). Sentiment Analysis in Indonesian Trading using Lexicon-based and Support Vector Machine. In *2023*

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

International Conference on Computer Science, Information Technology and Engineering (ICCoSITE) (pp. 744-749). IEEE.

29. Gunasekaran, K. P. (2023). Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review. *arXiv preprint arXiv:2305.14842*.

30. Kastrati, Z., Imran, A. S., Daudpota, S. M., Memon, M. A., & Kastrati, M. (2023). Soaring Energy Prices: Understanding Public Engagement on Twitter Using Sentiment Analysis and Topic Modeling With Transformers. IEEE Access, 11, 26541-26553.

31. Raza, A. A., Habib, A., Ashraf, J., Shah, B., & Moreira, F. (2023). Semantic orientation of crosslingual sentiments: Employment of lexicon and dictionaries. IEEE Access, 11, 7617-7629.

32. Baraibar-Diez, E., Llorente, I., & Odriozola, M. D. (2023). Text emotion analysis in aquaculture communication via Twitter: The case of Spain. Marine Policy, 152, 105605.

33. Albahli, S., & Nawaz, M. (2023). TSM-CV: Twitter Sentiment Analysis for COVID-19 Vaccines Using Deep Learning. Electronics, 12(15), 3372.

34. Esuli, A., & Sebastiani, F. (2006). SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of the Fifth Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy, May 2006.

35. Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta, May 2010.

36. Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Multi-Faceted Polarity Annotation of Lexical Resources. In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta, May 2010.

37. Baccianella, S., Esuli, A., & Sebastiani, F. (2009). Evaluation Measures for Opinion Mining Applications. In Proceedings of the International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco, May 2008.

Received: August 04, 2025

Volume 38 No. 11s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

38. Leoraj, A. & Jeyakarthic, M., Spotted Hyena Optimization with Deep Learning-Based Automatic Text Document Summarization Model. International Journal of Electrical and Electronics Engineering, 10(5), PP 153-164.

39. Jeyakarthic, M., & Leoraj, A. (2024). Knowledge-Infused Corpus Building for Context-Aware Summarization with Bert Model. *Migration Letters*, *21*(S4), PP 1681-1694.

40. Jeyakarthic, M., & Senthilkumar, M., Optimal Bidirectional Long Short Term Memory based Sentiment Analysis with Sarcasm Detection and Classification on Twitter Data, 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-6, doi: 10.1109/MysuruCon55714.2022.9972540.

41. Smith, J. D., & Johnson, R. M. (2023). Transforming Multimedia to Text: Advances in Audio, Video, and Image Conversion. *Journal of Information Technology*, 37(3), 215-230.

Received: August 04, 2025