A Survey on Text Processing Using Deep Learning Techniques

Akshita Tyagi[1], Terrance Frederick Fernandez[2][0000-0002-7317-3362], K.Shantha Kumari[3], Amit Kumar Tyagi[4][0000-0003-2657-8700]

[1]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, 600127, Tamil Nādu, India
[2]Institute of computer science and engineering, Saveetha school of Engineering, Saveetha Institute of Medical and Technical sciences, Thandalam,Chennai-602105

[3]Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur-603203
[4]Department of Fashion Technology, National Institute of Fashion Technology, New Delhi, India

akshitatyagi222@gmail.com, frederick@pec.edu, shanthak@srmist.edu.in, amitkrtyagi025@gmail.com

Abstract. We report an experiment in which we attempted to determine emotion class at the phrase level. The method is based on a mixture of machine learning and key word analysis. A substantial annotated data set exists in which a statement was manually classified beyond the six fundamental emotions: joy, love, rage, fear, surprise and sadness. Using the annotated data set, create an emotion vector for the key word in the input sentence. Calculate the emotion vector using an algorithm. of a phrase from the emotion vector of a word. The sentence was then classified into relevant emotion classes based on the emotion vector. In comparison to an individual method, the results are demonstrated and determined to be satisfactory. The goal of this article is to showcase many of the most significant text document categorization approaches and methodologies, as well as to raise awareness of some of the intriguing difficulties that remain unanswered, particularly in the fields of machine learning techniques and text representation.

Keywords: Sentence Level, Emotion Detection, Emotion Vector, Machine Learning, Natural Language Processing

1 Introduction

The Internet and the World Wide Web are generating a massive quantity of data from users who contribute text relating to product reviews, thoughts, attitudes, and other services. This data is being processed and analyzed by a variety of techniques. To evaluate the material and comprehend these processes, NLP and information retrieval technologies are applied. Sentiment analysis' fundamental challenge is divided into two categories: positive opinion and negative opinion. This study compares sentiment classification approaches based on lexicon and sentiment classification methods based on machine learning. Several methods and methodologies are used.

This document classifies and categorizes characteristics. To tackle sentiment analysis difficulties, many approaches are used. Emotions are a part of human nature and play a vital role in behavior science. Emotions are a set of thoughts, feelings, experiences, behaviors, cognitions, and conceptualizations that define a state of mind. The three basic methodologies used when creating text-based ED systems, as well as their

merits and shortcomings, have been described. The present state-of-the-art is also discussed, with an emphasis on the applicable methodologies, datasets used, key contributions, and limits. Twitter sentiment analysis is a very new and difficult study subject. Because social media sites such as Twitter include a large volume of text sentiment data in the form of tweets, it is possible to determine people's feelings or opinions on a certain event. Opinion mining is another term for sentiment analysis, is beneficial for film reviews, product reviews, customer service reviews, and opinions about any event. This allows us to determine if a certain item or service is excellent, terrible, or preferable. It may also be used to find out what people think about any event or person, as well as the polarity of text, whether positive, negative, or neutral. Sentiment analysis is a sort of text classification that may categorize text into various emotions. Sentiment analysis is a way of transforming, extracting and understanding views from a text using NLP and classifying them as positive, negative, or natural feelings. The two primary approaches of sentiment analysis have been described as a machine learning and a lexicon-based approach methodology. The lexicon-based approach counts the negative and positive words connected with the data, whereas the machine learning technique employs algorithms to classify and extract sentiment from data. Scholars have now been developing a new sentiment analysis algorithm that is both accurate and useful. One of the NLP approaches is the feature extraction algorithm. To extract sentiment, extract subject-specific features, from each lexicon that contains sentiment, and relate the sentiment to a certain matter the feature can be used. It outperformed machine learning algorithms, achieving accuracy of up to 87 percent for online review articles and 919 percent for general web page and news item reviews. This method concentrated on generic text and eliminated several challenging circumstances in order to get better results, such as confusing sentences or sentences with no feeling.

2 Sentiment analysis is divided into numerous categories

Different methods of sentiment analysis are utilized in the market to analyze people's feelings. Other sorts of sentiment analysis, in addition to regular opinions – positive, negative, or neutral – aid in understanding people's underlying sentiments, genuine purpose, and feelings.

2.1 2.1 Sentiment with a finer granularity

One of the most fundamental and often utilized techniques of measuring client attitude is to ask them. This analysis offers us a better grasp of the client comments we've received. The feelings are categorized using publicly available categories such as positive, neutral, and negative. Another method to scale consumer input is to provide a rating choice ranging from 1 to 5. This method is used by the majority of e-commerce companies to determine their clients' feelings.

2.2 2.2 Sentiment analysis for emotion identification

This is a more sophisticated approach of identifying emotion in a text. This type of analysis assists in detecting and comprehending people's emotions. Anger, sorrow, happiness, frustration, fear, panic, and concern are all possible emotions to include. The benefit of employing this is that a business can better understand why a consumer feels a certain way. However, analyzing people's emotions via emotion detection is challenging since individuals use a variety of phrases with varied meanings, such as sarcasm.

2.3 2.3 Analyses based on aspects

This sort of sentiment analysis is primarily concentrated on the features of a certain product or service. One of the most fundamental and often utilized techniques of measuring client attitude is to ask them., as well

as mechanized procedures such as customer care chores, allowing us to acquire valuable insights on the go. Businesses may use aspect-based sentiment analysis to discover which components of their products or services are causing them dissatisfaction, and it can help them progressively resolve those issues. Problems with new software programs, such as malfunctions or serious problems, can also be handled.

2.4 2.4 Sentiment analysis based on intent

The automated classification of textual material based on the customer's intent is known as intent classification. An intent classifier can examine writings and reports organically and classify them into intentions like Purchase, Downgrade, Unsubscribe, and so on. This is useful for understanding the goals behind a huge number of the client's inquiries, automating measures, and gaining valuable experience. When it comes to areas like customer assistance and sales, intent categorization allows firms to be more customer friendly. It enables them to respond to leads more quickly and handle big numbers of queries.

3 Approaches to sentiment analysis

The main strategy is rules-based and utilizes a dictionary of words named by sentiment to decide the sentiment of a sentence. Sentiment scores regularly should be joined with extra principles to relieve sentences containing negations, sarcasm, or dependent clauses [18-21].

3.1 Rule-based approach

The rules incorporate the following NLP methods.:

- Tokenization, stemming, and part-of-speech tagging.
- Lexicons.

Because the sequential merger of words is not considered in rule-based systems, they are extremely basic. Superior processing methods can be employed, as well as the most up-to-date regulations, to permit newer vocabularies and modes of expression. The inclusion of new rules, on the other hand, might have an impact on previously achieved outcomes and make the entire system highly convoluted. Rule-based systems need continuous fine-tuning and maintenance, which will necessitate funding at regular periods.

3.2 Machine Learning approach

Machine learning strategies rely on machine learning algorithms rather than human designed rules. A sentiment analysis problem is typically described as a classification problem, in which the classifier receives text input and assigns it to one of three classes: positive, negative, or neutral [7].

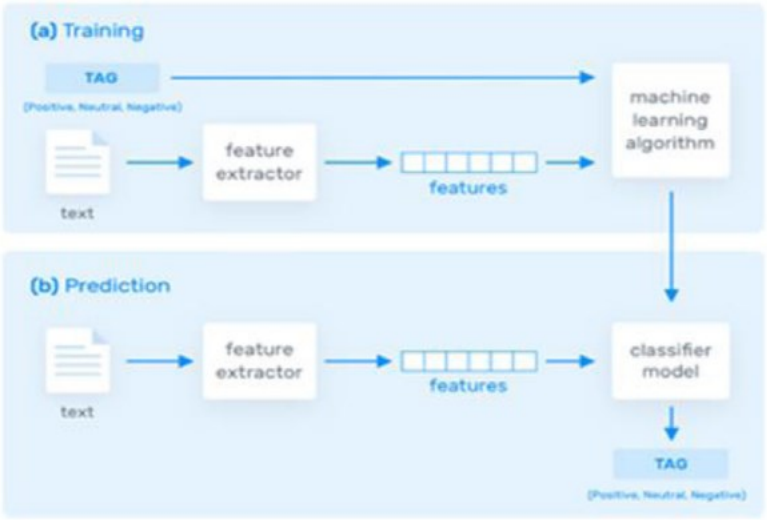

Fig. 1. Machine Learning Method

- During the training phase, our model learns to match a certain input data set to the associated output data set, as shown in Fig. 1 (a). The textual input is transformed into a features vector by the feature extractor. The feature tag and vector pair pairs are then sent into the algorithm, which generates a model.
- The feature extractor translates concealed textual inputs into feature vectors in the prediction process represented in Fig. 1 (b). The model is then fed these vectors, which generate prediction tags for the relevant vectors.

3.3 Lexicon based approach

The following was the approach used to complete the sentiment classification challenge. To begin, all text data training weights and classified text are calculated. After then, the full textual data is stored in a one-dimensional emotion field. The mean weights of the training text data were then determined for each sentiment category. The classified text belonged to the 1-D emotion field's closest category [17]. The accompanying graphic (2) depicts the machine learning and lexicon-based approaches in action.

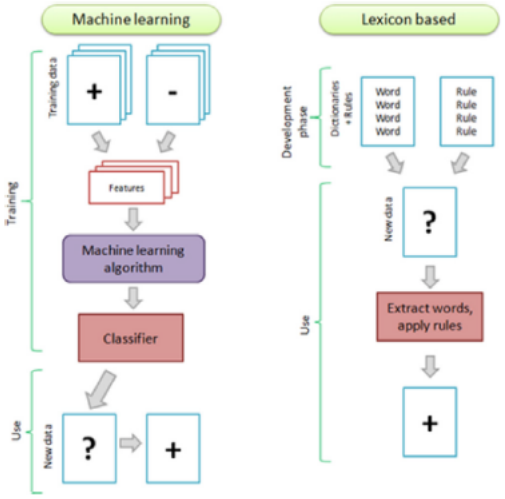

Fig. 2. Machine Learning Approach vs Lexicon Based Approach.

The following was the approach used to complete the sentiment classification challenge. To begin, all text data training weights and classified text are calculated. After that, the complete textual data is stored in a one-dimensional emotion field. The mean weights of the training text data for each sentiment category were then calculated. The classified text belonged to the closest category in the 1-D emotion field. The process of machine learning and lexicon-based approach is depicted in the above figure (2).

4 All Approaches Advantages and Limitations

Table 1 shows the advantages and disadvantages of the Rule-Based technique, Machine Learning approach, and Lexicon-Based Approach. By referring the below table, we can see the limitations and advantages of different algorithms so that we can choose the better one.

Table 1. Approaches' Benefits and Limitations of different algorithms

S.NO	Approaches	Advantages	Limitations
1.	Rule Based Approach	Data isn't required for training. Exceptional precision It's a great way to collect data since you can set up the system with rules and then let data flow in as people use it.	The recall rate is lower. The task of listing all of the requirements is laborious and time-consuming.
2.	Machine Learning Approach	It is not necessary to use a dictionary. Demonstrate a high level of categorization precision.	Many times, a classifier trained on textual input in a single field does not function with other fields.

3.	Lexicon Based Approach	It is not necessary to provide a name for the knowledge or the technique of learning.	Requires incredible semantic assets that aren't widely available.

5 Text-Based Emotion Detection (TBED)

This section gives a general introduction of emotion models, which describe how emotions are recognized. Some datasets are indicated for academics looking for data for studies Reference [1] identifies several different ways to describe emotions.

According to the search results, 202 of the 1810 results accessible on IEEE Xplore for the search keyword "ED" over the whole year range were focused on "ED from texts." Similarly, out of a total of 5481 results for "ED," the Scopus database returned 593 "ED from texts" results. Figures 3 and 4 exhibit graphs depicting the distribution over a ten-year period (i.e., from 2010 to 2020).

As opposed to text-based ED, the results demonstrate that multimodal types of ED, such as speech, body language, facial expressions, and so on, are commonly worked on. The scarcity is due to the fact that, unlike multimodal approaches, texts may not depict specific emotional cues, making emotion identification from texts significantly more challenging in contrast to other methods. Furthermore, the challenges of extracting emotions from grammatically incorrect texts, brief messages, sarcasm in written documents, contextual information, and other sources might be exhausting. Inadequate understanding of appropriate text extraction methods for the field, which is still in its infancy due to a lack of study, is a key roadblock in accurately recognizing emotions from written texts.

$$r : A*T \rightarrow E$$

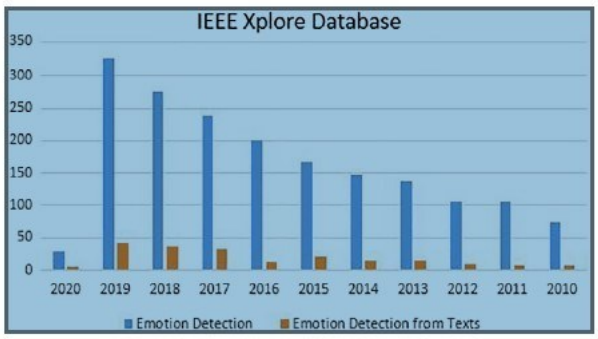

Fig. 3. In the IEEE Xplore Database, a graph depicting the discrepancy of research in emotion detection and emotion detection from texts.

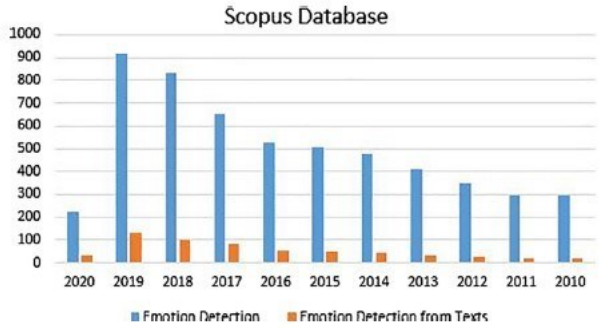

Fig. 4. In the Scopus Database, a graph depicting the discrepancy of research in emotion recognition and emotion detection from texts.

T is for the text from which emotions are to be drawn, and A stands for the author of T. While the problem may appear simple at first, determining the appropriate relationship under which an author can be significantly associated with their written texts in order to determine their emotions can be difficult. The variable r represents the relationship between the author and their written texts, which frequently express emotions, and it states that, while the problem may appear simple at first, determining the appropriate relationship under which an author can be significantly associated with their written texts in order to determine their emotions can be difficult. Text-based education has distinct hurdles as a result of all of these concerns. Despite its challenges, the field has made great progress in improving human-computer interaction. These include detecting and providing timely assistance to individuals who may be suicidal,7,8 detecting insulting sentences in conversations,10 chatbots for psychiatric counseling,29 and so on, all of which are still in the early stages of development.

5.1 Datasets for text-based ED (Emotion Detection) research

The collecting of data relevant to the course is the next critical stage in recognizing emotions from text after settling on the model to represent emotions. For research purposes, there are a few structured annotated datasets for Emotion Detection that are freely available. This section lists the most important publicly accessible datasets and their characteristics. Table 2 contains a table with the datasets, their attributes and the emotion models they reflect.

Table 2. Datasets for identifying emotions in texts that are publicly available

S.NO	Dataset	Feature	Emotion Model
1	ISEAR21	7665 phrases annotated for fear, joy, rage, sorrow, guilt, disgust, and shame reactions were acquired from 37 nations through cross-cultural study.	Distinct
2	SemEval-2017 Task 4	1250 texts were selected from Twitter, Google News, news headlines, and other notable publications. The six primary emotions identified by Ekman have been labeled.	Distinct
3	EmoBank	articles, blogs, news headlines, travel guides, newspapers, letters and fiction are just a few examples of what you may find on the internet.	Dimensional

4	WASSA-2017 Emotion Intensities (EmoInt)	Tweets were used to create this map, which was labeled for emotions including happiness, sadness, fear, and rage.	Distinct
5	Affect data from Cecilia Ovesdotter Alm	The emotions are divided into five categories: angry, afraid, pleased, sad, disgusted, and startled.	Distinct
6	DailyDialog	There are 13118 dialogues in this collection, all of which have been annotated for happiness, sorrow, rage, contempt, fear, surprise, and other emotions.	Distinct
7	Crowd-Flower	It's made up of 39,740 tweets that have been annotated for thirteen13 emotions.	Distinct
8	Grounded emotions	2557 total tweets were gathered and examined to determine if they were in a joyful or sad mood.	Distinct
9	Emotional Stimulation	Data for the emotion lexical unit was created using Frame Nets' annotated data. There are 1594 emotion-labeled sentences in this collection.	Distinct
10	The Valence and Arousal dataset	2895 Facebook Posts were used to create this page.	Dimensional
12	MELD data	Friends talks and utterances were used to compile this list.	Distinct
13	Emotion Lines	Conversations from the Friends TV show and Facebook messenger chats were used to compile this list.	Distinct
14	SMILE dataset	Tweets concerning the British Museum were used to compile this list.	Distinct
15	Dens Dataset	The data consists of 9710 paragraphs categorized as pleasure, anger, sadness, anticipation, fear, surprise, disgust, love, and neutral from Wattpad stories and Project Gutenberg books.	Distinct
16	Aman Emotion Dataset	Blogposts were used to create this piece.	Distinct

6 Feature Set

Python was used to create the feature extraction. SNoW takes only operational features as input, resulting in a feature set with an average size of 30 features. A list of features is provided below. These had been incorporated as Boolean values having sustainable value ranges. In order to gain higher generalization coverage, the ranges often overlapped.

- The story's first phrase
- Combinations of specific characteristics
- In a sentence, direct speech (i.e., the entire quote).
- Type of narrative with a theme (There are three main categories and fifteen sub-types.)
- (! and?) are notable punctuation marks.
- Uppercase word in its entirety

- Number of words in a sentence (0-1, 2-3, 4-8, 9-15, 16-25, 26-35, >35)
- Story range progress.
- Count the number of Vs in a phrase (excluding participles) (0-1, 0-3, 0-5, 0-7, 0-9, > 9)
- Count the number of balances of contrary forces words (1, 2, 3, 4, 5, 6)
- Emotion or feelings terms from WordNet
- Affective words and interjections.

A variety of storey advances and witticisms in feature conjunctions were coupled with counts of positive and negative words. Feature groups 1, 3, 5, 6, 7, 8, 9, 10, and 14 are obtained directly from phrases in the narrative, having features 9, 10, and 14 being tagged with the SNoW POS-tagger. The number of active verbs in a sentence is represented by Group 10. Verb domination, together with quote and punctuation, aims to represent the idea that emotion is frequently accompanied by greater activity and involvement. There are three main story categories in the current collection (JOKES AND ANECDOTES, ORDINARY FOLK-TALES, AND ANIMAL TALES), in addition to 15 subclasses (e.g., subclass of the ORDINARY FOLK-TALE is a supernatural helpers).

Words are plainly significant in semantic tasks. We looked at specific word lists in addition to examining 'content words.' Synonyms and hyponyms were manually extracted for nouns and any verbal homonyms that were identical.

6.1 Extraction of feature

The goal of pre-processing is to make the border of each language structure explicit and to remove as many language-dependent elements as feasible, such as tokenization, stop words removal, and stemming [10]. The first stage in pre-processing is FE, which converts text materials into a readable word format. Pre-processing procedures include eliminating stop words and stemming words [12]. Text categorization materials are represented by a large number of characteristics, the majority of which may be irrelevant or noisy [9]. The elimination of a significant number of terms, ideally based on statistics, is known as DR.

to generate a low-dimensional vector [13]. Because successful dimension reduction makes the learning work more efficient and saves more storage space, DR approaches have lately received a lot of attention [14]. The steeps most used for feature extractions (Fig.6) are:

- **Tokenization:** It is the process of treating a document as a string and then partitioning it into tokens.
- **Removing stop words:** Stop words like "the," "a," "and" and so on are commonly used, thus the unnecessary words should be omitted.
- **Stemming word:** Using a stemming method to transform diverse word forms into canonical forms that are comparable. The process of conflating tokens to their base form, such as connection to connect, computing to compute, and so on, is called this stage.

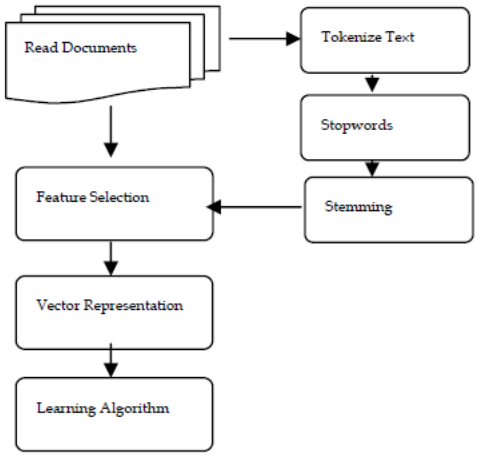

Fig. 5. Text document classification

6.2 Feature Selection

After feature extraction, the next stage in text classification preprocessing is feature selection to generate vector space, which improves a text classifier's scalability, efficiency, and accuracy. A decent feature selection approach should take domain and algorithm properties into account [15]. The primary concept behind FS is to choose out a subset of characteristics from the source papers.

FS is carried out by retaining the words with the highest score based on a preset estimate of the word's value [9].The chosen attributes preserve the physical meaning of the data and improve understanding of the learning process [11]. The large dimensionality of the feature space is a key issue in text categorization. Almost every text domain includes a large number of characteristics, the majority of which are neither relevant or advantageous for text classification tasks, and even small noise features can degrade classification accuracy significantly [16]. As a result, FS is widely utilised in text classification to minimise feature space dimensionality and enhance classifier efficiency and accuracy.

7 Comparison analysis

The lexicon-based technique, machine learning-based approach, and hybrid-based approach are the three types of approaches to sentiment analysis. This research examines the differences between lexicon-based and machine-learning-based techniques.

7.1.1 Lexicon Based Approach

A sentiment dictionary with sentiment terms is the lexicon-based strategy. The emotion words are given a score that signifies positive, negative, or neutral feeling. The collection of sentiment terms, phrases, and even idioms are generated the communication of lexicon sentiment. There are two types of lexicon-based approaches these are dictionary-based classification and corpus-based classification as shown in figure 5.

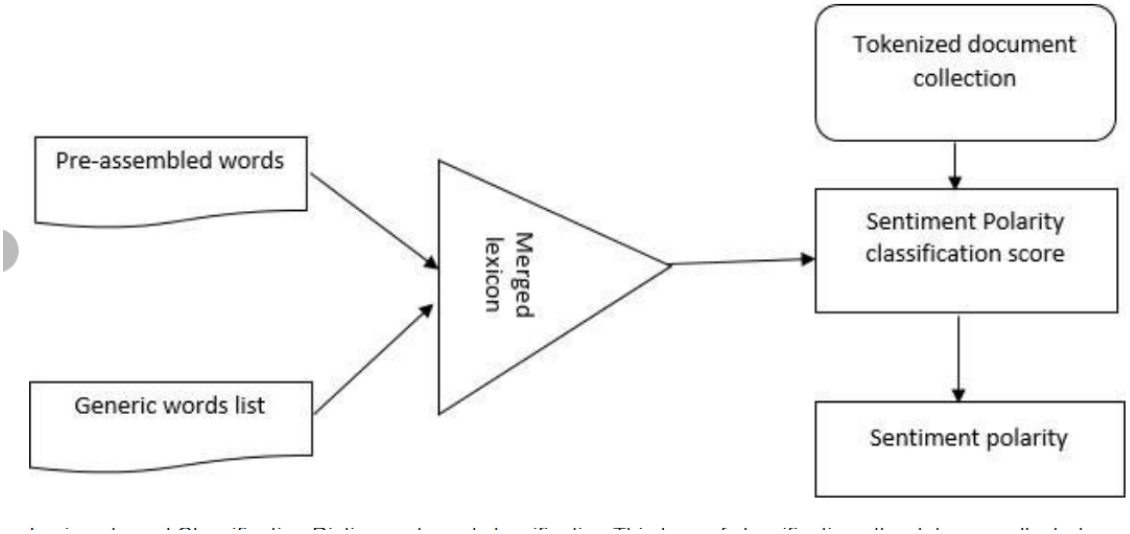

Fig. 6. Lexicon based Classification

7.1.2 Dictionary-based classification

The data is manually gathered in this sort of categorization, and the information is searched for synonyms and antonyms in a sentiment dictionary. WordNet and SentiwordNet are the two dictionaries in question.

7.1.3 Corpus-based classification

This comes close to the goals of dictionaries in a single topic. The terms refer to Latent Semantic Analysis (LSA) and a method based on semantics, both of which are statistical and semantic methodologies.

7.1 Machine Learning based Classification

Machine learning algorithms are the most effective approach for classifying sentiments into positive, negative, and neutral categories in text. Machine learning necessitates a dataset for training and testing. The learning dataset is referred to as the document-based learning dataset, and the validation performance is referred to as the validation performance. The reviews are categorized using machine learning techniques. A supervised learning algorithm and an unsupervised learning algorithm are the two types of machine learning algorithms. SVM, Maximum Entropy, Nave Bayes, and KNN are examples of supervised learning algorithms. HMM, Neural Networks, PCA, SVD, ICA, and other unsupervised machine learning methods are examples [2].

7.2 Support Vector Machine

It is one of the most effective classification approaches for machine learning algorithms. In classical learning approaches, SVM is based on structural risk minimization, which determines the hypothesis with the lowest chance of mistakes [3]. It is based on minimizing the empirical risk, which is the learning set's

performance. Quadratic optimization issues result. Complexity categorization issues require a greater number of patterns and a larger scale. On SVM, the feature space dimensionality has no bearing.

7.3 Comparative table for different classification Algorithm

Below the Table 3 is showing the accuracy percentage of different algorithms applied. Through this accuracy we can determine which algorithm gives the best result and we can also see their attribute totals also for each algorithm like totals of positive, negative, and neutral words and because of this total we can calculate the accuracy percentage of different algorithm.

Table 3. Comparison of different algorithms

S.NO.	Algorithm	Total words	Positive Words	Negative	Neutral Words	Accuracy Percentage (%)
1	NB [10]	5576	2115	1199	103	96
2	SMO [10]	5576	1987	1254	127	96
3	Random forest [13]	5576	1419	1210	186	96
4	Random Tree [13]	5576	1204	1213	34	100
5	Keyword [13]	5576	2250	720	806	97
6	Emotion [13]	5576	1456	530	1700	93
7	Senti-word[13]	5576	2110	513	203	87
8	SVM [5]	45000	23514	21486	-	76
9	Maximum Entropy [5]	45000	22606	22226	-	75
10	CNN-KNN	3500	600	600	-	91

7.2.2 Naïve Bayes

It's a straightforward and efficient categorization algorithm. It is typically used to classify documents at the document level. It estimates the probability of words and categories in a text document. It is heavily reliant on approaches based on features. It requires quick and accurate categorization. Large datasets are not required.

7.2.3 K-Nearest Neighbor

In a comparable test document, it identifies the labels category that is related to the training document. In KNN [4], this approach classifies objects into object-based classes. It is a sort of lazy learning in which just

the function is estimated locally, and all computations differ until classification. It's used to calculate the Euclidean or Manhattan distance [5].

7.2.4 Maximum Entropy

This categorization method demonstrates the usefulness of NLP applications. There is no connection between the characteristics. When the conditional independence assumption is fulfilled, performance may be improved. With training data, the feature functions model equals the predicted values [5].

7.2.5 Decision Tree Learning

It's a tree-based method that consists of a collection of child and root nodes. It concentrates on the desired outcomes. The text property is represented by every internal node in the decision tree model, which is a flow chart structure model. Each branch represents the text's conclusion, whereas the left node represents the child node or class distributions. ID3, CART, and C4.5 are three well-known DT algorithms. The ID3 method is a relatively simple technique that is used to partition data into categories. The Gini coefficient is utilized as the test attribute for selection criteria using the CART algorithm. The ID3 is given to C4.5. The gain ratio is used as a splitting criterion.

7.4 Semantic Orientation Approach

It's a categorization based on unsupervised learning. The training dataset is not required. It identifies the positive and negative measures that are incorporated in the verb to defend.

7.5 Keyword-based Classification

The categories are known as "Bag of Words." Because the terms are domain independent, they are classed as either positive or negative. It gives precise spelling categorization and assigns equal weight to each word [6].

7.6 Emotions based Classifications

This is referred to as fundamental emotions. The set is used to categories the feelings as good or negative. Positive and negative emotions are carefully categorized. In the text files, good and negative emotions are represented by a set of symbols [6].

This section provides a basic overview of emotion models, which specify how emotions are detected. Some datasets in table 4 are highlighted for academics that want data for their field study.

This table shows the comparison of different approaches with several kinds of datasets, findings and as well as limitations which is clearly stated below and because of this information one can decide to which algorithm should be implemented for the better results. In this table we have taken different datasets for the algorithms to check for the better findings and limitations with different dataset we have taken for the algorithm we selected. Table 4 provides a summary of the state-of-the-art literature in the domain of text-based ED as discussed in the article. The work has been arranged according to the year of publication (in descending order, from 2019) in order to aid in the comprehension of the field taking into account the progress of research.

Table 4. An overview of recent developments in text-based emotion identification

S.NO.	Approach	Dataset(s)	Findings	Limitations
1	Machine Learning	Emo-Dis-HI data	Cross-lingual embeddings and transfer learning were used to demonstrate that information gleaned from valuable resource languages may be applied to other fields of language. An F1 score of 0.53 was obtained.	Words are used without consideration for their context.
2	Machine Learning	Tweets	With an accuracy of 72.06 percent versus 55.50 percent, the NB machine learning technique outperformed the KNN machine learning technique.	Contextual information in sentences is extracted in a limited way.
3	Rule Based	ISEAR data	In the ISEAR dataset, emotions were detected with a strong focus on phrasal verbs	Words are used without consideration for their context.
4	Machine Learning	Tweets	For text-based emotion recognition, there was a BERT and HRLCE model given. For the joyful, furious, and sad emotion classes, they received an F1 score of 0.779.	There are a lot of misclassifications.
5	Machine Learning	Task 3 dataset for SemEval-2019	An attention-based paradigm for categorizing emotions was presented. They scored 0.7582 on the F1 scale.	Doesn't perform well when it comes to identifying happiness.
6	Machine Learning	Texts in many languages on Facebook	A bi-directional transformer BERT architecture was proposed. Hindi texts had a 0.4521 F1 score, whereas English texts received a score of 0.5520.	A BERT design for bi-directional transformers was proposed. Hindi texts had a 0.4521 F1 score, whereas English texts received a score of 0.5520.

7	Hybrid	Tweets	Both online and offline, SVM, NB, and Text emotions were detected using a Decision Tree. A 90 percent accuracy rate was found.	loose semantic feature extraction
8	Hybrid	News Headlines	I classified emotions into six groups by using SVM classifier.	Improve your performance with a robust classification technique.
9	Machine Learning	YouTube Comments	Accuracy of Emotion Classification: 59.2 percent 65.97 percent and 54.24 percent, respectively, for multiclass emotion labels.	Accuracy outcomes that are satisfactory
10	Machine Learning	Interviews, forums, and article comments were used to create a dataset.	The SVM had an accuracy of more than 75% and a recall of more than 80%, while the Tree Bagger and the Multilayer Neural Network both have recall and accuracy of above 75%.	In the model or in the design, there is no semantic representation.
11	Hybrid	Tweets	To extract actionable emotion patterns from Tweets, we used the NRC emotion lexicon and SVM.	Generalization is challenging due to the small number of emotion classifications.
12	Machine Learning	Emoji Prediction is a shared task during SemEval-2018.	For recognizing emotions in emojis, proposed a label wise attention LSTM method.	With regularly used emojis, the model did not operate properly.

8 Issues in Text Sentiment Analysis

This section discusses some of the outstanding concerns that have been found and suggests some potential study directions for emotion detection researchers that work with text. The most up-to-date technology talks in this article revealed that field research is primarily divided into two parts. Language representation and categorization are the two parts of the process. The extraction of contextual information is critical during language representation since it provides the foundation for increasing categorization accuracy. The need to offer a comprehensive method for extracting this contextual information from text has been regarded as a critical concern. The use of transformer-based embeddings improved the quality of contextual data extraction significantly. 10,87,89 However, various constraints, such as out of vocabulary (OOV) restrictions, increased level of complexity, and, most crucially, technique in tiny networks, overfitting is a

problem, impact the usage of transformers. 99,100. An ensemble of attention and neuro-fuzzy networks101 might assist lessen the limiting effects of transformers because of the mentioned limitations.

As a result, categorization performance improves. Prior to classification, the attention networks should focus on extracting important the neuro-fuzzy networks, on the other hand, have unique characteristics should provide clearer intelligibility and categorization of the recovered characteristic.

9 Conclusion

The results of the performed systematic literature review include research on sentiment analysis in social media. The following three contributions are made by the paper. First, we'll go through the approach for assessing social media sentiment. Although several methods have been proposed by researchers, the most frequent methods used in Lexicon-based methods are SentiWordnet and TF-IDF, while Nave Bayes and SVM are used in machine learning. The data itself determines whether type of sentiment analysis is acceptable. Both strategies yielded comparable results in terms of accuracy. The structure of the text, as well as the time and amount of data, must all be considered. Combining lexical and machine learning methods to increase the quality and accuracy of the outcome is recommended. If the data structure is jumbled, there is a little amount of data, and you only have a short amount of time to analyze, the lexicon-based strategy is advised. Machine learning-based methods are better suited to larger data since they take more time and data to train. Combining lexical and machine learning methods to increase the quality and accuracy of the outcome is recommended.

Second, we figure out which social media sites are most widely utilized to collect data for sentiment analysis. Twitter is the most widely used social media channel for information gathering. The bulk of the articles in this evaluation make use of Twitter as their primary social media platform. Because of Twitter's enormous availability, accessibility, and diversity of content, this is the case. Millions of tweets are sent out every day on virtually any topic. As a result, social media is quickly becoming into a vital source of information. Blogs, WordPress, YouTube, and other social media sites, on the other hand, attract less attention. Because the content of each social networking site may differ, it's worthwhile to investigate different possibilities and discoveries.

10 References

1. Borod JC. *The Neuropsychology of Emotion*. Oxford, UK: Oxford University Press; 2000.
2. Ajayi Adebowale, Idowu S.A, Anyaehie Amarachi A., “Comparative Study of Selected Data Mining Algorithms Used for Intrusion Detection”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-3, Year: 2013.
3. Ahmad Ashari, Iman Paryudi, A Min Tjoa, “Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 11, Year: 2013.
4. V. Garcia, and C. Debreuve, “Fast k Nearest Neighbor Search using GPU”, IEEE, Year: 2008.
5. C.EMELDA “A Comparative Study on Sentiment Classification and Ranking on Product Reviews” International Journal of Innovative Research in Advanced Engineering (IJIRAE)ISSN: 2349-2163 Volume 1 Issue 10, Year: 2014
6. Vishal A. Kharde, and S.S. Sonawane “Sentiment Analysis of Twitter Data: A Survey of Techniques” International Journal of Computer Applications (0975 – 8887) Volume 139 –No.11, Year: 2016
7. Aless.ia D’Andrea, Fernan.do Ferri, Pat.rizia Grifoni, Tiz.iana Guzzo, Approaches, Tools and Applications for Sentiment Analysis Implementation, International Journal of Computer Applications 125 (2015) 26–33.
8. Gw.anghoon Yoo, Jee.sun Nam, in: A Hybrid Approach to Sentiment Analysis Enhanced by Sentiment Lexicons and Polarity Shifting Devices. The 13th Workshop on Asian Language Resources, Miyazaki, Japan, May 2018, pp. 21–28. Kiyoaki Shirai.

9. Montanes, E., Ferandez, J., Diaz, I., Combarro, E.F and Ranilla, J., “Measures of Rule Quality for Feature Selection in Text Categorization”, 5th international Symposium on Intelligent data analysis, Germeny-2003, Springer-Verlag 2003, Vol2810, pp.589-598, 2003.
10. Wang, Y., and Wang X.J., “A New Approach to feature selection in Text Classification”, Proceedings of 4th International Conference on Machine Learning and Cybernetics, IEEE- 2005, Vol.6, pp. 3814-3819, 2005.
11. Liu, H. and Motoda, “Feature Extraction, constraction and selection: A Data Mining Perpective.”, Boston, Massachusetts (MA): Kluwer Academic Publishers.
12. Lee, L.W., and Chen, S.M., “New Methods for Text- Categorization Based on a New Feature Selection Method and New Similarity Measure Between Documents”, IEA/AEI, France 2006.
13. Manomaisupat, P., and Abmad k., “Feature Selection for text Categorization Using Self Orgnizing Map”, 2nd International Conference on Neural Network and Brain, 2005, IEEE press Vol 3, pp.1875-1880, 2005.
14. Yan, J., Liu, N., Zhang, B., Yan, S., Chen, Z., Cheng, Q., Fan, W., and Ma, W., “OCFS: Optimal Orthogonal centroid Feature selection for Text Categorization.” 28 Annual International conference on Reserch and Informational reterival, ACM SIGIR, Barizal, pp.122-129, 2005.
15. Zi-Qiang Wang, Xia Sun, De-Xian Zhang, Xin Li “An Optimal Svm-Based Text Classification Algorithm” Fifth International Conference on Machine Learning and Cybernetics, Dalian, pp. 13-16, 2006.
16. Jingnian Chen a, b, Houkuan Huang a, Shengfeng Tian a, Youli Qua Feature selection for text classification with Naïve Bayes” Expert Systems with Applications 36, pp.5432–5435, 2009.
17. R.K. Dash, T.N. Nguyen, K. Cengiz, A. Sharma, Fine-tuned support vector regression model for stock predictions, in: Neural Computing and Applications, 2021, pp. 1–15.
18. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, & I. Polosukhin. (2017). Attention Is All You Need.
19. G. Ranjan, T.N. Nguyen, H. Mekky, Z.L. Zhang, On virtual id assignment in networks for high resilience routing: a theoretical framework, in: GLOBECOM 2020-2020 IEEE Global Communications Conference, IEEE, 2020, pp. 1–6.
20. B.D. Parameshachari, H.T. Panduranga, S. liberata Ullo, Analysis and computation of encryption technique to enhance security of medical images, IOP Conference Series: Materials Science and Engineering, 925, IOP Publishing, 2020.