

Review

A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research

Kian Long Tan, Chin Poo Lee *  and Kian Ming Lim 

Faculty of Information Science and Technology, Multimedia University, Melaka 75450, Malaysia

* Correspondence: cplee@mmu.edu.my

Abstract: Sentiment analysis is a critical subfield of natural language processing that focuses on categorizing text into three primary sentiments: positive, negative, and neutral. With the proliferation of online platforms where individuals can openly express their opinions and perspectives, it has become increasingly crucial for organizations to comprehend the underlying sentiments behind these opinions to make informed decisions. By comprehending the sentiments behind customers' opinions and attitudes towards products and services, companies can improve customer satisfaction, increase brand reputation, and ultimately increase revenue. Additionally, sentiment analysis can be applied to political analysis to understand public opinion toward political parties, candidates, and policies. Sentiment analysis can also be used in the financial industry to analyze news articles and social media posts to predict stock prices and identify potential investment opportunities. This paper offers an overview of the latest advancements in sentiment analysis, including preprocessing techniques, feature extraction methods, classification techniques, widely used datasets, and experimental results. Furthermore, this paper delves into the challenges posed by sentiment analysis datasets and discusses some limitations and future research prospects of sentiment analysis. Given the importance of sentiment analysis, this paper provides valuable insights into the current state of the field and serves as a valuable resource for both researchers and practitioners. The information presented in this paper can inform stakeholders about the latest advancements in sentiment analysis and guide future research in the field.

Keywords: sentiment analysis; review; survey; advances; machine learning; deep learning; ensemble learning



Citation: Tan, K.L.; Lee, C.P.; Lim, K.M. A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Appl. Sci.* **2023**, *13*, 4550. <https://doi.org/10.3390/app13074550>

Academic Editor: Francisco García-Sánchez

Received: 20 February 2023

Revised: 1 March 2023

Accepted: 6 March 2023

Published: 3 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sentiment analysis, a subfield of natural language processing, focuses on the automatic identification and categorization of emotions and sentiments expressed through written text. With the exponential growth of social media, the availability of public opinions and sentiments has increased, making sentiment analysis a crucial tool for comprehending public sentiment across various domains such as business, politics, and others.

The sentiment analysis process encompasses several essential steps, including preprocessing, feature extraction, and classification. In the preprocessing stage, the raw text data undergo cleaning to remove irrelevant information such as stop words, special characters, and numbers. This stage also involves transforming the text data into features using techniques such as Term Frequency–Inverse Document Frequency (TF-IDF), GloVe, fastText, and word2vec. In the feature extraction stage, the processed text is then classified into sentiments using machine learning methods, such as logistic regression, naive Bayes, and support vector machines, or deep learning models such as long short-term memory (LSTM) and recurrent neural networks.

This paper presents a comprehensive review of the recent advancements in the field of sentiment analysis. Unlike other review papers [1–3], the works reviewed in this paper are divided into three categories: conventional machine learning, deep learning, and ensemble

learning. This paper delves into the preprocessing techniques, feature extraction methods, classification techniques, and datasets used, as well as the experimental results of each of the works. Additionally, this paper discusses the widely used sentiment analysis datasets and their associated challenges, as well as the limitations of the current works and the potential future research directions in this field.

The main contributions of this paper are:

- A comprehensive overview of the state-of-the-art studies on sentiment analysis, which are categorized as conventional machine learning, deep learning, and ensemble learning, with a focus on the preprocessing techniques, feature extraction methods, classification methods, and datasets used, as well as the experimental results.
- An in-depth discussion of the commonly used sentiment analysis datasets and their challenges, as well as a discussion about the limitations of the current works and the potential for future research in this field.

2. Sentiment Analysis Algorithms

This section provides an overview of the current landscape of sentiment analysis algorithms. Sentiment analysis is a process that seeks to identify and categorize the emotions or sentiments expressed in written texts. To perform this task, the raw text data must undergo several steps, including data preprocessing, feature extraction, and classification.

Data preprocessing is an important step in the sentiment analysis process, as it helps standardize the text data and remove any irrelevant or noisy elements. This step can include techniques such as stemming, lemmatization, and the removal of stop words and special characters. The cleaned text data are then transformed into features or embeddings, which are fed into the classifier for sentiment prediction.

The classifiers used in sentiment analysis can be broadly classified into three categories: machine learning, deep learning, and ensemble learning (as illustrated in Figure 1). Machine learning classifiers, such as logistic regression, naive Bayes, and support vector machine, use mathematical models to predict sentiments. Deep learning classifiers, such as recurrent neural networks and long short-term memory (LSTM) models, leverage artificial neural networks to make sentiment predictions. Ensemble learning methods combine multiple classifiers to achieve better sentiment analysis performance. Figure 2 provides a visual representation of the sentiment analysis process.

The choice of classifier depends on the specific requirements and use case of the sentiment analysis task. This section reviews the existing works in sentiment analysis and highlights the preprocessing techniques, feature extraction methods, and classification algorithms used in each work.

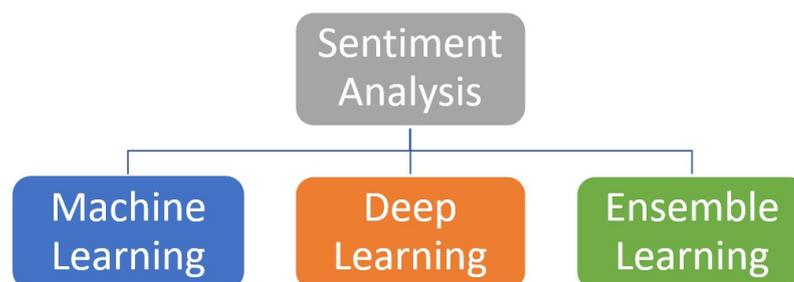


Figure 1. Sentiment analysis approaches can be categorized as machine learning, deep learning, or ensemble learning.

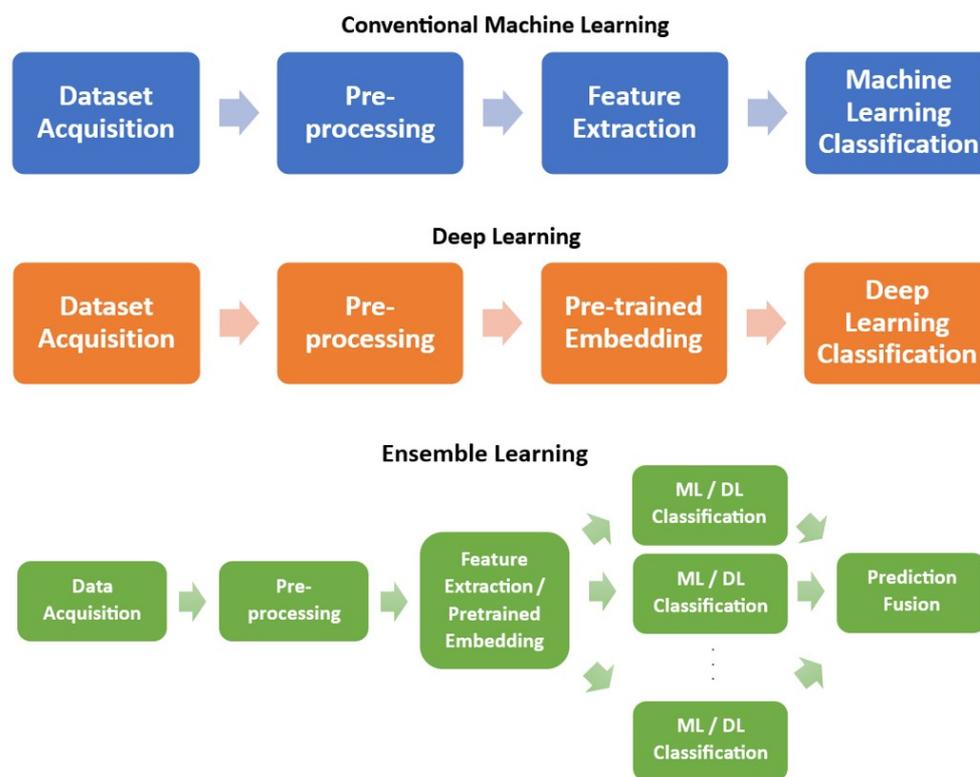


Figure 2. Flow charts of the sentiment analysis process using conventional machine learning, deep learning, and ensemble learning approaches.

2.1. Machine Learning Approach

Machine learning approaches in sentiment analysis start with standardizing the text data through preprocessing and removing any irrelevant information. Then, feature extraction techniques, such as Term Frequency–Inverse Document Frequency (TF-IDF) and N-grams, are applied to represent the text as numerical features that can be fed into machine learning classifiers. Some of the commonly used classifiers in sentiment analysis include support vector machine (SVM), naive Bayes, logistic regression, random forest, and decision trees.

For example, Jung et al. (2016) [4] applied multinomial naive Bayes to sentiment analysis after extracting features from Tweets. Their approach was evaluated on the sentiment140 dataset, which contains 1.6 million Tweets labeled as positive, negative, or neutral. The authors reported that the multinomial naive Bayes model achieved an accuracy of 85% when the dataset was split into a 9:1 training–testing ratio.

Similarly, Athindran et al. (2018) [5] utilized naive Bayes to analyze customer reviews on Twitter. The preprocessing steps included tokenization and stemming, and the naive Bayes model achieved an accuracy of 77%. These works demonstrate the effectiveness of machine learning methods in sentiment analysis, particularly with the naive Bayes classifier.

The study by Vanaja et al. (2018) [6] compared the performance of two popular machine learning algorithms, naive Bayes and support vector machine, in the sentiment analysis of customer reviews on Amazon. The text data were preprocessed to remove stop words and were represented using the a priori algorithm. The results showed that naive Bayes outperformed support vector machine with an accuracy of 90.42% compared to 83.42%. Iqbal et al. (2018) [7] conducted sentiment analysis experiments using three machine learning methods, naive Bayes, support vector machine, and maximum entropy. The experiments were conducted on two datasets, IMDb and Sentiment140, and included preprocessing steps such as tokenization, lemmatization, and text cleaning. The results

showed that the maximum entropy method, when combined with unigram and bigram features, achieved the highest accuracy of 88% on the IMDb dataset and 90% on the Sentiment140 dataset.

Rathi et al. (2018) [8] evaluated the sentiment analysis performance of three machine learning algorithms, decision tree, AdaBoost, and support vector machine, on three datasets: Sentiment140, the Polarity dataset, and the University of Michigan. The text was preprocessed and represented using Term Frequency–Inverse Document Frequency (TF-IDF) features. The results showed that support vector machine achieved the highest accuracy of 82%, followed by decision tree with 84% accuracy and AdaBoost with 67% accuracy.

In another work, Tariyal et al. (2018) [9] conducted experiments to compare the performance of several machine learning models for sentiment analysis. The models included linear discriminant analysis, classification and regression trees, k-nearest neighbors, support vector machine, random forest, and C5.0. The authors collected a dataset of 1150 product review Tweets, which underwent preprocessing steps such as stop-word and punctuation removal, case folding, and stemming. The text was then represented as a term-document matrix and fed into different models. The results showed that classification and regression trees, which is a type of decision tree algorithm, achieved the highest accuracy of 88.99%.

In a similar study, Hemakala and Santhoshkumar (2018) [10] conducted a comprehensive evaluation of several machine learning models for sentiment analysis. The models included decision tree, random forest, support vector machine, k-nearest neighbors, logistic regression, Gaussian naive Bayes, and AdaBoost. The authors collected a dataset of 14,640 Tweets related to Indian Airlines, which were labeled as positive, negative, or neutral. The text was preprocessed using techniques such as lemmatization and stop-word removal. The results showed that the AdaBoost model achieved the highest accuracy of 84.5%, demonstrating that boosting models perform better in terms of sentiment classification accuracy.

Rahat et al. (2019) [11] compared the performance of two machine learning models, multinomial naive Bayes and support vector classifier (SVC) with linear kernels, for sentiment analysis. The dataset consisted of approximately 10 K positive, negative, and neutral Tweets, which were collected from Twitter. The text was preprocessed using techniques such as stemming, URL removal, and stop-word removal. The results showed that SVC achieved an accuracy of 82.48% and outperformed the multinomial naive Bayes model, which obtained an accuracy of 76.56%.

In the study conducted by Makhmudah et al. (2019) [12], the authors proposed the use of a support vector machine (SVM) method for the sentiment analysis of Tweets related to homosexuality in Indonesia. The dataset was preprocessed to remove stop words and undergo lemmatization and stemming and was represented by Term Frequency–Inverse Document Frequency (TF-IDF) features. The results showed that the SVM method achieved an impressive accuracy of 99.5% on the dataset.

Wongkar and Angdresey (2019) [13] conducted a comparison of three machine learning methods, support vector machine (SVM), naive Bayes, and k-nearest neighbors (KNN), for sentiment analysis. The dataset was a collection of Tweets related to the presidential candidates in Indonesia in 2019 and was divided 80:20 for training and testing. The preprocessing steps included text parsing, tokenization, and text mining. The results showed that naive Bayes achieved the highest accuracy of 75.58%, followed by KNN with 73.34% and SVM with 63.99%.

In the study by Madhuri (2019) [14], the author compared the performance of four machine learning methods, C4.5, naive Bayes, support vector machine (SVM), and random forest, on a dataset collected from Twitter related to Indian Railways. The dataset was divided into positive, negative, and neutral sentiments. The results showed that the highest accuracy was achieved by SVM with 91.5%, followed by C4.5 with 89.5%, random forest with 90.5%, and naive Bayes with 89%.

In Gupta et al. (2019) [15], a comprehensive study was conducted to evaluate the performance of four machine learning algorithms for sentiment analysis, including decision

tree, logistic regression, support vector machine, and neural network algorithms. The sentiment140 dataset was utilized as the sample data for the study, which was preprocessed and transformed into TF-IDF features. The results indicated that the neural network model outperformed the other algorithms, achieving the highest accuracy of 80.

Prabhakar et al. (2019) [16] introduced a novel approach to sentiment analysis by combining boosting and bagging methods in an AdaBoost model. The study focused on sentiment analysis of US airline Twitter data and the dataset was preprocessed to remove irrelevant information and then analyzed using data-mining techniques. The authors used a 75-25 split for the data, with 75% of the data utilized for training and 25% for testing. The AdaBoost model showed promising performance, with an F-score of 68%.

Hourrane et al. (2019) [17] applied the ridge classifier method for sentiment analysis on two datasets: IMDb and Sentiment140. The preprocessing stage involved various techniques such as tokenization, case folding, and the removal of URLs, HTML tags, and irrelevant words. The resulting data were then represented using the TF-IDF method. The experiments showed that the ridge classifier achieved an accuracy of 90.54% on the IMDb dataset and 76.84% on the Sentiment140 dataset, demonstrating its effectiveness for sentiment analysis.

The study by Alsalman (2020) [18] aimed at enhancing the accuracy of sentiment analysis for Arabic Tweets. To achieve this goal, the multinomial naive Bayes method was used on a dataset consisting of 2000 Tweets labeled as positive or negative. The raw data were preprocessed through tokenization using 4-grams and stemming using the Khoja stemmer, before being represented as TF-IDF features and classified using a fivefold cross-validation process. The proposed method showed promising results with an accuracy of 87.5% on the dataset.

Saad (2020) [19] conducted a comprehensive study on the sentiment analysis of US airline Twitter data using six different machine learning models, including support vector machine (SVM), logistic regression, random forest, XgBoost, naive Bayes, and decision tree. The authors applied various preprocessing techniques, such as stop-word removal, punctuation removal, case folding, and stemming, to prepare the data for analysis. Feature extraction was carried out using the bag-of-words method on a dataset collected from CrowdFlower and Kaggle, consisting of 14,640 samples with three sentiment classes: positive, negative, and neutral. The results showed that SVM achieved the highest accuracy of 83.31%, followed by logistic regression with an accuracy of 81.81% when the dataset was split into a 70% training and 30% testing set.

Alzyout et al. (2021) [20] performed sentiment analysis related to violence against women using several machine learning models, including support vector machine (SVM), k-nearest neighbor, naive Bayes, and decision tree. The dataset was collected from Arabic Tweets and preprocessed using tokenization, stemming, and stop-word removal. Feature extraction was performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method and the results showed that SVM achieved the highest accuracy of 78.25% on the self-collected dataset.

In the study conducted by Jemai et al. (2021) [21], the authors aimed to develop a sentiment analyzer that could classify the polarity of text with high accuracy. To achieve this, they employed five different machine learning methods, including naive Bayes, multinomial naive Bayes, Bernoulli naive Bayes, logistic regression, and linear support vector classification. The experiments were conducted using the ‘twitter samples’ corpus in the Natural Language Toolkit, which consisted of 10,000 Tweets with equal numbers of positive and negative Tweets. Before feeding the data into the models, a thorough preprocessing phase was conducted, which included tokenization, stop-word removal, URL removal, symbol removal, case folding, and lemmatization. The results showed that the naive Bayes method achieved the highest accuracy of 99.73% on the corpus, making it the best-performing algorithm in the study.

In conclusion, the authors of the reviewed studies mainly focused on data preprocessing, feature extraction, and a comparison of the accuracies of the different machine learning

algorithms in sentiment analysis. The results showed that different algorithms achieved different levels of accuracy, ranging from 67% to 99.5%. Some of the highest-performing algorithms included maximum entropy, naive Bayes, and support vector machine. A summary of these conventional machine learning approaches can be found in Table 1.

Table 1. Summary of machine learning approaches.

Literature	Features	Classifier	Dataset	Accuracy (%)
Jung et al. (2016) [4]		MNB	Sentiment140	85
Athindran et al. (2018) [5]		NB	Self-collected dataset (from Tweets)	77
Vanaja et al. (2018) [6]	A priori algorithm	NB, SVM	Self-collected dataset (from Amazon)	83.42
Iqbal et al. (2018) [7]	Unigram, Bigram	NB, SVM, ME	IMDb Sentiment140	88 90
Rathi et al. (2018) [8]	TF-IDF	DT AdaBoost SVM	Sentiment140, Polarity Dataset, and University of Michigan dataset	84 67 82
Hemakala and Santhoshkumar (2018) [10]		AdaBoost	Indian Airlines	84.5
Tariyal et al. (2018) [9]		Regression Tree	Own dataset	88.99
Rahat et al. (2019) [11]		SVC MNB	Airline review	82.48 76.56
Makhmudah et al. (2019) [12]	TF-IDF	SVM	Tweets related to homosexuals	99.5
Wongkar and Angdresey (2019) [13]		NB	Twitter (2019 presidential candidates of the Republic of Indonesia)	75.58
Madhuri (2019) [14]		SVM	Twitter (Indian Railways)	91.5
Gupta et al. (2019) [15]	TF-IDF	Neural Network	Sentiment140	80
Prabhakar et al. (2019) [16]		AdaBoost (Bagging and Boosting)	Skytrax and Twitter (Airlines)	68 F-score
Hourrane et al. (2019) [17]	TF-IDF	Ridge Classifier	IMDb Sentiment 140	90.54 76.84
Alsalmán (2020) [18]	TF-IDF	MNB	Arabic Tweets	87.5
Saad et al. (2020) [19]	Bag of Words	SVM	Twitter US Airline Sentiment	83.31
Alzyout et al. (2021) [20]	TF-IDF	SVM	Self-collected dataset	78.25
Jemai et al. (2021) [21]		NB	NLTK corpus	99.73

2.2. Deep Learning Approach

Deep learning has emerged as a popular method for sentiment analysis due to its ability to learn representations of textual data. In these methods, textual data is preprocessed and then encoded using pretrained embeddings such as GloVe and word2vec. These embeddings are then fed into deep learning models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM), and gated recurrent units (GRUs) for representation learning and classification.

In Ramadhani and Goo's study (2017) [22], sentiment analysis was performed on a dataset of 4000 Tweets in both Korean and English. The preprocessing phase included tokenization, case folding, stemming, and the removal of stop words, numbers, and punctuation marks. A multilayer perceptron (MLP) model with three hidden layers was used in the study, which was optimized using Stochastic Gradient Descent (SGD). The results showed that the MLP model achieved an accuracy of 75.03% on the dataset.

Similarly, in Demirci et al.'s study (2019) [23], an MLP model was used for sentiment analysis on Turkish Tweets. The dataset consisted of 3000 positive and negative Tweets with the hashtag "15Temmuz". The preprocessing phase included Turkish deasciification, tokenization, stop-word and punctuation removal, and stemming. The word2vec pretrained model was used to convert the text into embeddings, which were then used as inputs to the MLP consisting of six dense layers and three dropout layers. The results showed that the MLP achieved an accuracy of 81.86% on the dataset.

In Raza et al. (2021) [24], an MLP was utilized to perform sentiment analysis on COVID-19-related Tweets. The dataset consisted of 101,435 Tweets and was preprocessed to remove HTML tags and non-letters and to tokenize and stem the text. The text data were then represented using two feature extractors, the count vectorizer and TF-IDF vectorizer, which were classified separately using an MLP model with five hidden layers and ReLU activation. The results showed that the count vectorizer representation with MLP achieved the highest accuracy of 93.73%.

Dholpuria et al. (2018) [25] conducted a sentiment analysis study comparing several machine learning and deep learning methods, including naive Bayes, support vector machine, logistic regression, k-nearest neighbors, an ensemble model, and a CNN. The dataset used was the IMDb dataset containing 3000 reviews with positive and negative labels. The text was preprocessed to remove irrelevant characters, symbols, repeating words, and stop words, and then represented using the count vectorizer. The results showed that the CNN model achieved the highest accuracy of 99.33%.

Harjule et al. (2020) [26] compared machine learning and deep learning methods for sentiment analysis using two datasets: Sentiment140 and Twitter US Airline Sentiment. The dataset was preprocessed using tokenization and case folding and stop-word, URL, hashtag, and punctuation removal. The methods compared were multinomial naive Bayes, logistic regression, support vector machine, long short-term memory (LSTM), and an ensemble of multinomial naive Bayes, logistic regression, and support vector machine with majority voting. On the Sentiment140 dataset, LSTM recorded the highest accuracy of 82%, whereas support vector machine achieved the highest accuracy of 68.9% on the Twitter US Airline Sentiment dataset.

Uddin et al. (2019) conducted a study that focused on sentiment analysis of 5000 Tweets in the Bangla language using LSTM [27]. The authors preprocessed the data by removing spaces and punctuation marks and split the dataset into 80% for training, 10% for validation, and 10% for testing. After tuning the hyperparameters, the authors found that an LSTM architecture with five layers, each with a size of 128, a batch size of 25, and a learning rate of 0.0001, achieved the highest accuracy of 86.3%.

In another study, Alahmary and Al-Dossari (2019) analyzed sentiment in the Saudi dialect using LSTM and bidirectional long short-term memory (BiLSTM) models [28]. The authors used 60,000 Tweets with positive and negative sentiments in the Saudi dialect as their dataset and preprocessed the data by removing numbers, punctuation, special symbols, and non-Arabic letters. The data were then text-normalized to replace word forms with their lemmas and encoded using a word2vec pretrained model for word embedding. The dataset was randomly split into 70% for training and 30% for testing and then trained and tested on both the LSTM and BiLSTM models. The results showed that the BiLSTM model performed better, achieving an accuracy of 94%.

Yang (2018) proposed a recurrent neural filter-based CNN and LSTM model for sentiment analysis [29]. The study utilized a recurrent neural network (RNN) as a convolutional filter in the model. The training and testing were performed using the Stanford Sentiment Treebank dataset, which was encoded using GloVe word embedding. The proposed model comprised an embedding layer using a GloVe pretrained model, a pooling layer, and an LSTM layer. To prevent overfitting, the authors adopted the Adam optimizer and early stopping. The model achieved an accuracy of 53.4% on the Stanford Sentiment Treebank dataset.

In the paper by Goularas and Kamis (2019) [30], a hybrid model was proposed for sentiment analysis that combined the strengths of CNNs and LSTM. The data used in the study comprised 32,000 Tweets obtained from the International Workshop on Semantic Evaluation competition. To prepare the data, the authors performed preprocessing steps such as removing URLs, emoticons, and special characters, as well as converting all the text to lowercase characters. To represent the text, the hybrid model leveraged two different pretrained word embeddings, namely word2vec and GloVe. The results indicated that the hybrid CNN and LSTM model with GloVe embedding achieved an accuracy of 59%.

Similarly, Hossain and Bhuiyan (2020) [31] proposed a hybrid model that combined CNN and LSTM for sentiment analysis. The dataset used in the study comprised 100

restaurant reviews collected from the Foodpanda and Shohoz Food apps. The preprocessing steps involved removing unimportant words and symbols, as well as converting the text into embeddings using the word2vec algorithm. The proposed hybrid model consisted of multiple layers, including an embedding layer with the word2vec pretrained model, a convolutional layer, max-pooling layer, LSTM layer, dropout layer, and classification layer. The model achieved an impressive accuracy of 75.01% on the self-collected dataset.

Tyagi et al. (2020) [32] presented a hybrid model that combined a CNN with bidirectional long short-term memory (BiLSTM) for sentiment analysis on the Sentiment140 dataset. The dataset comprised 1.6 million positive and negative Tweets and underwent preprocessing steps such as case folding and stemming and the removal of stop words, numbers, URLs, Twitter usernames, and punctuation marks. The proposed hybrid model consisted of multiple layers, including an embedding layer with the GloVe pretrained model, a one-dimensional convolutional layer, a BiLSTM layer, fully connected layers, dropout layers, and a classification layer. The results indicated that the model achieved an accuracy of 81.20% on the Sentiment140 dataset.

The study by Rhanoui et al. (2019) [33] presented a hybrid model that combined the strengths of convolutional neural networks (CNNs) and bidirectional long short-term memory (BiLSTM) networks for sentiment analysis. The experiments were conducted on a dataset that contained 2003 articles and news articles with positive, negative, and neutral sentiments. The text data were preprocessed by converting them into word embeddings using a pretrained doc2vec model. The proposed hybrid model comprised several layers, including a convolutional layer, max-pooling layer, BiLSTM layer, dropout layer, and classification layer. The experiments showed that the hybrid model achieved an impressive accuracy of 90.66% on the dataset.

Jang et al. (2020) [34] improved on the hybrid CNN and BiLSTM model by incorporating an attention mechanism. The experiments were conducted on the IMDb dataset, which contained 50,000 positive and negative reviews. The text data were preprocessed by encoding them using a word2vec pretrained embedding model. The proposed hybrid model was further optimized using techniques such as the Adam optimizer, L2 regularization, and dropout. The results showed that the proposed model achieved an accuracy of 90.26% on the IMDb dataset, demonstrating the effectiveness of the attention mechanism in sentiment analysis.

Chundi et al. (2020) [35] developed a hybrid model that combined CNNs and BiLSTMs for sentiment analysis. The model was constructed using several key components, including a convolutional layer that consisted of 3 filters with a filter size of 128 and a ReLU activation function, a max-pooling layer with a pool size of 3, and a dropout layer with a rate of 0.2 to prevent overfitting. The experiments were conducted on a dataset that contained 10,401 comments in English, Kannada, and a mixture of both languages. The text data underwent preprocessing steps such as case folding and the removal of digits, special characters, symbols, and duplicated comments. The proposed convolutional BiLSTM model achieved an accuracy of 77.6% on the dataset, demonstrating its effectiveness in sentiment analysis on multilingual data.

Thinh et al. (2019) [36] presented a hybrid model that combined a one-dimensional convolutional neural network (1D-CNN) and an RNN for sentiment analysis. The model was evaluated on the IMDb dataset, which consisted of 50,000 movie reviews with positive and negative sentiments. The 1D-CNN part of the model included 128 and 256 filters for its convolutional layers. The RNN part consisted of LSTM, Bi-LSTM, and GRU, with 128 units in each layer. The results showed that the 1D-CNN with the GRU model outperformed the other models, achieving an accuracy of 90.02% on the IMDb dataset.

Janardhana et al. (2020) [37] proposed a hybrid model consisting of a CNN and an RNN for sentiment analysis. The model was trained and evaluated on a dataset of 12,500 positive and negative movie reviews. The text was preprocessed by removing stop words and punctuation and then represented using the GloVe pretrained model. The proposed convolutional RNN model achieved an accuracy of 84% on the movie review dataset.

Chowdhury et al. (2020) [38] introduced a Bi-LSTM model for sentiment analysis on Twitter data. The model was trained and evaluated on the Twitter US Airline Sentiment dataset, which is publicly available on Kaggle. The text data were preprocessed by applying techniques such as tokenization and stop-word and punctuation removal. The text was then represented using three different word embeddings—word2vec, GloVe, and sentiment-specific word embedding. The results showed that the Bi-LSTM model achieved an accuracy of 81.20% on the Twitter US Airline Sentiment dataset, demonstrating its effectiveness for sentiment analysis on Twitter data.

Vimali and Murugan (2021) [39] proposed a Bi-LSTM model for analyzing the sentiment of Amazon e-commerce reviews. The authors collected a dataset of 23,485 reviews, which were divided into three classes based on the rating points: negative, neutral, and positive. Data cleaning was carried out by tokenizing the data and removing special characters, followed by the generation of word embeddings using the word2vec pretrained model. The BiLSTM model achieved an impressive accuracy of 90.26% on the self-collected dataset.

Similarly, Anbukarasi and Varadhaganapathy (2020) [40] utilized a character-based deep bidirectional long short-term memory (DBLSTM) method for analyzing the sentiment of Tamil Tweets. The authors collected a dataset of 1500 Tweets, which were divided into positive, negative, and neutral classes. The data were preprocessed to remove unnecessary symbols and characters and the cleaned data were represented using the word2vec pretrained model as the word embedding for the DBLSTM method. The model achieved an accuracy of 86.2% on the Tamil Tweets dataset.

Furthermore, Kumar and Chinnalagu (2020) [41] proposed a sentiment analysis bidirectional long short-term memory (SAB-LSTM) model. The model consisted of 196 Bi-LSTM units, 128 embedding layers, 4 dense layers, and a classification layer with a softmax activation function. The authors used a dataset of 80,689 samples from 5 sentiment classes collected from various sources such as Twitter, YouTube, Facebook, news articles, etc. The dataset was divided into a 90% training set and a 10% testing set and the experiments revealed that the SAB-LSTM model outperformed traditional LSTM models in sentiment analysis.

Hossen et al. (2021) [42] explored the use of RNNs for sentiment analysis of customer reviews collected from hotel booking websites. In their study, the authors employed various preprocessing techniques, such as lemmatization and stemming, as well as punctuation and stop-word removal, to clean the data before feeding it into two deep learning models: long short-term memory (LSTM) and a gated recurrent unit (GRU). The LSTM model had 30 hidden layers, whereas the GRU model had 25 hidden layers. The results of the experiments showed that the LSTM and GRU models achieved accuracies of 86% and 84%, respectively, on the self-collected dataset.

In another study, Younas et al. (2020) [43] investigated the performance of two deep learning models, multilingual BERT (mBERT) and XLM-RoBERTa (XLM-R), for sentiment analysis of multilingual social media text. The dataset used in this study was scraped from Twitter during the 2018 general election in Pakistan and consisted of 20,375 Tweets in both English and Roman Urdu. The Tweets were classified into positive, negative, and neutral classes and were split into 80% for training and 20% for testing. After fine-tuning the learning rate of the mBERT model to 2×10^{-5} and the XLM-R model to 2×10^{-6} , the results showed that mBERT achieved 69% accuracy, whereas XLM-R achieved 71% accuracy.

Dhola and Saradva (2021) [44] conducted a comparative study to evaluate the performance of four machine learning models for sentiment analysis: support vector machine, multinomial naive Bayes, long-short term memory (LSTM), and BERT. The Sentiment140 dataset, which consists of 1.6 million Tweets classified as positive or negative, was used in the study. After preprocessing the data using tokenization, stemming, lemmatization, and stop-word and punctuation removal, the results showed that the BERT model performed the best, achieving 85.4% accuracy and outperforming the other models.

The study by Tan et al. (2022) [45] introduced a cutting-edge approach to sentiment analysis by blending the strengths of two popular models: the robustly optimized BERT approach (RoBERTa) and long short-term memory (LSTM). The hybrid model was designed

to tackle the challenge of analyzing the sentiment of text sequences by combining the self-attention and dynamic masking capabilities of RoBERTa with the ability of LSTM to capture long-range dependencies in the encoded text. The proposed hybrid model was tested on three widely used sentiment analysis datasets: IMDb, Twitter US Airline, and Sentiment140. The results showed that the hybrid model achieved impressive accuracy scores of 92.96%, 91.37%, and 89.70%, respectively, on these datasets. These results serve as strong evidence of the efficacy of the hybrid model in accurately capturing the sentiment of text sequences.

Kokab et al. (2022) [46] proposed a BERT-based convolutional bi-directional recurrent neural network (CBRNN) model that combined a pretrained BERT model and a neural network consisting of dilated convolutional and Bi-LSTM layers for sentiment analysis. The model extracted sentence-level semantics and contextual features from the data and generated embeddings, and then used dilated convolution to extract local and global contextual semantic features. Bi-LSTM was used for the entire sequencing of the sentences. The model achieved an accuracy of 97% on US airline reviews, 90% on self-driving car reviews, 96% on US presidential election reviews, and 93% on IMDb.

A sentiment transformer graph convolutional network (ST-GCN) was presented by AlBadani et al. (2022) [47] to model sentiment analysis as a heterogeneous graph and learn document and word embeddings using the sentiment graph transformer neural network. The model learned useful connections between nodes and determined the soft selection of edge types and complex relations to learn node representation for sentiment classification. Laplacian eigenvectors were used to fuse node positional information. The message-passing technique was used to learn the node representation on a heterogeneous graph and a transformer was used to aggregate local substructures with appropriate position encoding. The model achieved state-of-the-art results on real-world datasets achieving 95.43% on SST-B, 94.94% on IMDB, and 72.7% on Yelp 2014.

Tiwari and Nagpal (2022) [48] proposed the knowledge-enriched attention-based hybrid transformer (KEAHT) model for sentiment analysis. The model used a pretrained BERT model to train on a minimum training corpus. The model incorporated explicit knowledge from Latent Dirichlet Allocation (LDA) topic modeling and lexicalized domain ontology to address the issues of inaccurate polarity scoring and utility-based topic modeling in sentiment analysis. To further improve performance, the external knowledge sources, including sentiment network graphs, text-length distribution, word count, and higher-polarity Tweets, were consolidated. The proposed model was evaluated on two benchmark datasets related to the COVID-19 vaccine and Indian farmer protests, achieving testing accuracies of 91% and 81.49%, respectively.

A two-stage emotion detection methodology for sentiment analysis was introduced by Tesfagerghis et al. (2022) [49]. The first stage was an unsupervised zero-shot learning model based on a sentence transformer that returned probabilities for 34 emotions. The second stage trained a machine learning classifier on the sentiment labels using ensemble learning. The proposed hybrid semi-supervised method achieved the highest accuracy of 87.3% on the English SemEval 2017 dataset. The method was evaluated on three benchmark datasets using multiple classifiers, including machine learning, neural network, and ensemble learning. The best accuracy achieved was 0.87 for the binary sentiment classification problem and 0.63 for the three-class sentiment classification problem, with the sets of 10 and 6 emotions, respectively.

In the study by Maghsoudi et al. (2022) [50], the researchers used public content from Twitter to analyze insomnia-related Tweets based on time intervals. They performed sentiment analysis using pretrained transformers and the Dempster–Shafer theory (DST) to classify emotions as positive, negative, or neutral. They validated their pipeline on 300 annotated Tweets and achieved 84% accuracy. Using logistic regression, they found that the odds of posting negative Tweets about insomnia were higher in the peripandemic interval than in the prepandemic interval.

Jing and Yang (2022) [51] proposed a light-transformer model for sentiment analysis, which combined word-vector representation and positional embedding to extract sentence features. The model was based on the transformer architecture but only used the encoder

module and simplified the number of layers and parameters. The proposed method showed a 0.3–1.0% improvement in classification accuracy compared to traditional methods such as LSTM and CNN while greatly reducing the number of model parameters. The method yielded an accuracy of 76.40% on the NLPCC2014 Task2 dataset.

In recent years, deep learning has become a popular method for sentiment analysis. Various models, such as CNNs, RNNs, LSTM, and GRUs have been used for representation learning and classification. Pretrained embeddings such as GloVe and word2vec have been used to encode the preprocessed textual data. In several studies, MLP was used for sentiment analysis on different datasets and languages, including Korean, English, Turkish, COVID-19-related, and Bangla Tweets. The accuracies of the MLP model ranged from 75.03% to 93.73%. Other methods, such as naive Bayes, support vector machine, logistic regression, k-nearest neighbors, and an ensemble model were also tested in some studies, with the best results achieved with a CNN with 99.33% accuracy. In addition, LSTM and Bi-LSTM were used for sentiment analysis on different datasets, including the Saudi dialect and English. The Bi-LSTM model achieved the highest accuracy of 94% in one study. Table 2 provides a summary of the deep learning approaches.

Table 2. Summary of deep learning approaches.

Literature	Embedding	Classifier	Dataset	Accuracy (%)
Ramadhani et al. (2017) [22]		MLP	Korean and English Tweets	75.03
Demirci et al. (2019) [23]	word2vec	MLP	Turkish Tweets	81.86
Raza et al. (2021) [24]	Count Vectorizer and TF-IDF Vectorizer	MLP	COVID-19 reviews	93.73
Dholpuria et al. (2018) [25]		CNN	IMDb (3000 reviews)	99.33
Harjule et al. (2020) [26]		LSTM	Twitter US Airline Sentiment Sentiment140	82 66
Uddin et al. (2019) [27]		LSTM	Bangla Tweets	86.3
Alahmary and Al-Dossari (2018) [28]	word2vec	BiLSTM	Saudi dialect Tweets	94
Yang (2018) [29]	GloVe	Recurrent neural filter-based CNN and LSTM	Stanford Sentiment Treebank	53.4
Goularas and Kamis (2019) [30]	word2vec and GloVe	CNN and LSTM	Tweets from semantic evaluation	59
Hossain and Bhuiyan (2019) [31]	word2vec	CNN and LSTM	Foodpanda and Shohoz Food	75.01
Tyagi et al. (2020) [32]	GloVe	CNN and BiLSTM	Sentiment140	81.20
Rhanoui et al. (2019) [33]	doc2vec	CNN and BiLSTM	French articles and international news	90.66
Jang et al. (2020) [34]	word2vec	hybrid CNN and BiLSTM	IMDb	90.26
Chundi et al. (2020) [35]		Convolutional BiLSTM	English, Kannada, and a mixture of both languages	77.6
Thinh et al. (2019) [36]		1D-CNN with GRU	IMDb	90.02
Janardhana et al. (2020) [37]	GloVe	Convolutional RNN	Movie reviews	84
Chowdhury et al. (2020) [38]	word2vec, GloVe, and sentiment-specific word embedding	BiLSTM	Twitter US Airline Sentiment	81.20
Vimali and Murugan (2021) [39]		BiLSTM	Self-collected	90.26
Anbukkarasi and Varadhaganapathy (2020) [40]		DBLSTM	Self-collected (Tamil Tweets)	86.2
Kumar and Chinnalagu (2020) [41]		SAB-LSTM	Self-collected	29 (POS) 50 (NEG) 21 (NEU)
Hossen et al. (2021) [42]		LSTM GRU	Self-collected	86 84

Table 2. Cont.

Literature	Embedding	Classifier	Dataset	Accuracy (%)
Younas et al. (2020) [43]		mBERT	Pakistan elections in 2018 (Tweets)	69
		XLM-R		71
Dhola and Saradva (2021) [44]		BERT	Sentiment140	85.4
Tan et a. (2022) [45]		RoBERTa-LSTM	IMDb	92.96
			Twitter US Airline Sentiment	91.37
			Sentiment140	89.70
Kokab et al. (2022) [46]	BERT	CBRNN	US airline reviews	97
			Self-driving car reviews	90
			US presidential election reviews	96
			IMDb	93
AlBadani et al. (2022) [47]	ST-GCN	ST-GCN	SST-B	95.43
			IMDB	94.94
			Yelp 2014	72.7
Tiwari and Nagpal (2022) [48]	BERT	KEAHT	COVID-19 vaccine Indian Farmer Protests	91 81.49
Tesfagergish et al. (2022) [49]	Zero-shot transformer	Ensemble learning	SemEval 2017	87.3
Maghsoudi et al. (2022) [50]	Transformer	DST	Self-collected	84
Jing and Yang (2022) [51]	Light-Transformer	Light-Transformer	NLPCC2014 Task2	76.40

3. Ensemble Learning Approach

The ensemble learning method in sentiment analysis has been proven to be effective in improving the performance of predictions by combining the results of multiple models. The basic idea behind this approach is to leverage the strengths of different models for a more accurate prediction. An example of how ensemble learning can be applied in sentiment analysis is by using a voting-based approach. In this approach, multiple models, such as random forest, naive Bayes, and support vector machines (SVM), are trained on the same dataset. During the prediction phase, each model predicts the sentiment of the input text and the final prediction is made by taking the majority vote of all the models.

In the study by Alrehili and Albalawi (2019) [52], the authors employed an ensemble of five machine learning models including naive Bayes, random forest, support vector machine, bagging, and boosting. The authors used a dataset of 34,661 customer reviews collected from Amazon, which was preprocessed using techniques such as stemming, case folding, stop-word removal, and N-grams. The results showed that the majority voting ensemble method using unigram achieved an accuracy of 89.4%.

Similarly, Bian et al. (2019) [53] evaluated the performance of logistic regression, support vector machine, k-nearest neighbors, and an ensemble of these models with majority voting for sentiment analysis. The dataset used consisted of 6328 positive and negative samples and the TF-IDF vectorizer was applied for feature extraction. The results of the 10-fold cross-validation experiments showed that the ensemble model achieved the highest accuracy of 98.99%.

In 2021, Gifari and Lhaksmana [54] evaluated four machine learning models, multinomial naive Bayes, k-nearest neighbors, logistic regression, and an ensemble of these models, for sentiment analysis. The dataset consisted of positive and negative movie reviews taken from IMDb, which underwent preprocessing procedures, including tokenization, stop-word removal, and word stemming. The TF-IDF vectorizer was utilized to extract features. The ensemble model achieved the highest accuracy of 89.40% on the IMDb dataset.

Parveen et al. (2020) [55] conducted a study to assess the performance of five individual machine learning models and an ensemble of these models in sentiment analysis. The models included Bernoulli naive Bayes, multinomial naive Bayes, linear support vector classification, logistic regression, and Nu-support vector classification. The research used a dataset of movie reviews, consisting of 1000 positive and 1000 negative samples, sourced from the corpora community. The analysis utilized the top 3000 most frequent words as

features. The ensemble model employing the majority voting method achieved the highest accuracy of 91% on the dataset.

In another study, Aziz and Dimililer (2020) [56] proposed an ensemble learning algorithm with stochastic gradient descent, logistic regression, naive Bayes, decision tree, random forest, and support vector machine. Three datasets, SemEval-2017 4A, SemEval-2017 4B, and SemEval-2017 4C, were used, which were preprocessed using tokenization, word substitution, and stemming, as well as stop-word, punctuation, number, and repeated-word removal. Five different features were extracted from the text, including N-grams, part-of-speech tagging, TF-IDF, and bag-of-words and lexicon-based features. The authors used two ensemble learning methods, simple-majority voting ensemble and weighted-majority voting ensemble, in the experiments. The results showed that the ensemble model with weighted-majority voting achieved the highest accuracy of 72.95%, 90.8%, and 68.89% on the SemEval-2017 4A, SemEval-2017 4B, and SemEval-2017 4C datasets, respectively.

Varshney et al. (2020) [57] conducted a study on the effectiveness of machine learning techniques for sentiment analysis, including naive Bayes, logistic regression, stochastic gradient descent, and an ensemble model using majority voting. The Sentiment140 dataset was preprocessed by removing unwanted columns, usernames, hyperlinks, and null values, followed by converting the textual data to lowercase, which were vectorized using TF-IDF feature extraction to identify significant features. The results of the experiments indicated that the ensemble model outperformed the other models, with a positive class recall of 80% on the Sentiment140 dataset.

In Athar et al. (2021) [58], the authors proposed an ensemble model composed of logistic regression, naive Bayes, random forest, XGBoost, and multilayer perceptron to analyze the sentiment of movie reviews. The experiments were performed on the IMDb dataset, which consisted of 25,000 positive and 25,000 negative reviews. The dataset was preprocessed using tokenization and stemming, and by removing URLs, stop words, and punctuation. The feature extraction phase was conducted using the TF-IDF vectorizer. The ensemble model achieved an impressive accuracy of 89.9% on the IMDb dataset.

Nguyen and Nguyen (2018) [59] tackled the problem of Vietnamese sentiment analysis by implementing an ensemble of machine learning and deep learning models. The feature extraction phase utilized the TF-IDF vectorizer, whereas logistic regression and support vector machine were used for classification in the machine learning component. In the deep learning component, the Vietnamese text was represented by word2vec embeddings, which were then passed into the CNN and LSTM models for classification. Three ensemble techniques were evaluated, including the average/mean rule, the max rule, and the voting rule. The results showed that the ensemble model with the mean rule achieved an accuracy of 69.71% on the Vietnamese Sentiment Dataset (DS1), whereas the ensemble model with the voting rule achieved accuracies of 89.19% and 92.80% on the Vietnamese Sentiment Food Reviews dataset (DS2) and Vietnamese Sentiment dataset (DS3), respectively.

Kamruzzaman et al. (2021) [60] conducted a comparative study on the performance of six different models for sentiment analysis. Three of the models were conventional ensemble models, a voting ensemble, a bagging ensemble, and a boosting ensemble of logistic regression, support vector machine, and random forest. The other three models were neural network ensemble models, including 7-Layer CNN + LSTM + attention layer, 7-Layer CNN + GRU, and 7-Layer CNN + GRU + GloVe embedding. The datasets used in the study were the Grammar and Online Product Reviews dataset and the Restaurant Reviews in Dhaka, Bangladesh, dataset, both of which underwent several preprocessing steps such as tokenization, case folding, lemmatization, and stop-word and special character removal. The results indicated that the 7-Layer CNN + GRU + GloVe model achieved the highest accuracy of 94.19% on the Grammar and Online Product Reviews dataset, and the 7-Layer CNN + LSTM + attention layer model achieved the highest accuracy of 96.37% on the Restaurant Reviews in Dhaka, Bangladesh, dataset.

In another study, Al Wazrah and Alhumoud (2021) [61] evaluated the performance of four different models for sentiment analysis of Arabic Tweets, including a stacked gated

recurrent unit (SGRU), a stacked bidirectional gated recurrent unit (SBI-GRU), Arabic bidirectional encoder representations from transformers (AraBERT), and an ensemble of these models. The Arabic Sentiment Analysis dataset, which contained 56,674 samples with positive, negative, and neutral classes, was used for the evaluation. To preprocess the data, the automatic sentiment refinement (ASR) technique was applied to remove repeated Tweets, stop words, hashtags, and unrelated content. The SGRU and SBI-GRU models used the pretrained word embedding for Arabic text, AraVec, to represent the text. The results showed that the ensemble model (SGRU + SBI-GRU + AraBERT) performed best with 90.21% accuracy.

In Tan et al. (2022) [62], the authors proposed an ensemble model of RoBERTa-LSTM, RoBERTa-BiLSTM, and RoBERTa-GRU for sentiment analysis. The text sequences were first transformed into contextual embeddings by the RoBERTa model and then encoded by the LSTM, BiLSTM, and GRU models for classification. The ensemble model achieved high accuracy scores of 94.9%, 91.77%, and 89.81% on the IMDb, Twitter US Airline Sentiment, and Sentiment140 datasets, respectively.

Ensemble learning is a promising approach to sentiment analysis, leveraging the strengths of multiple models to achieve better performance. Several studies on sentiment analysis were conducted using ensemble models with different machine learning algorithms, including naive Bayes, support vector machine, random forest, logistic regression, and XGBoost. The majority voting scheme was the most commonly used method for combining the results of multiple models in an ensemble. The accuracies of the ensemble models used in sentiment analysis ranged from 80% to 98.99%. Some studies preprocessed the datasets using techniques such as case folding, stop-word removal, stemming, N-grams, and TF-IDF vectorization. Ensemble models were applied to various datasets, including customer reviews from Amazon, movie reviews from the corpora community, and Twitter data. A summary of the ensemble learning approaches discussed is presented in Table 3.

Table 3. Summary of ensemble learning approaches.

Literature	Feature Extractor	Classifier	Dataset	Accuracy (%)
Alrehili et al. (2019) [52]		NB + SVM + RF + Bagging + Boosting	Self-collected	89.4
Bian et al. (2019) [53]	TF-IDF	LR + SVM + KNN	COVID-19 reviews	98.99
Gifari and Lhaksmana (2021) [54]	TF-IDF	MNB + KNN + LR	IMDb	89.40
Parveen et al. (2020) [55]		MNB + BNB + LR + LSVM + NSVM	Movie reviews	91
Aziz and Dimililer (2020) [56]	TF-IDF	NB + LR + SGD + RF + DT + SVM	SemEval-2017 4A SemEval-2017 4B SemEval-2017 4C	72.95 90.8 68.89
Varshney et al. (2020) [57]	TF-IDF	LR + NB + SGD	Sentiment140	80
Athar et al. (2021) [58]	TF-IDF	LR + NB + XGBoost + RF + MLP	IMDb	89.9
Nguyen and Nguyen (2018) [59]	TF-IDF, word2vec	LR + SVM + CNN + LSTM (Mean)	Vietnamese Sentiment	69.71
		LR + SVM + CNN + LSTM (Vote)	Vietnamese Sentiment Food Reviews	89.19
		LR + SVM + CNN + LSTM (Vote)	Vietnamese Sentiment	92.80
Kamruzzaman et al. (2021) [60]	GloVe	7-Layer CNN + GRU + GloVe	Grammar and Online Product Reviews	94.19
	Attention embedding	7-Layer CNN + LSTM + Attention Layer	Restaurant Reviews	96.37
Al Wazrah and Alhumoud (2021) [61]	AraVec	SGRU + SBI-GRU + AraBERT	Arabic Sentiment Analysis	90.21
Tan et al. (2022) [62]		RoBERTa-LSTM + RoBERTa-BiLSTM + RoBERTa-GRU	IMDb	94.9
			Twitter US Airline Sentiment Sentiment140	91.77 89.81

4. Sentiment Analysis Datasets

Sentiment analysis datasets provide a foundation for training and evaluating sentiment analysis models. Despite the use of self-collected datasets in some sentiment analysis works, there are several publicly available and widely used sentiment analysis datasets that have proven to be valuable resources. These datasets include the Internet Movie Database (IMDb), Twitter US Airline Sentiment, Sentiment140, and SemEval-2017 Task 4 datasets.

4.1. Internet Movie Database (IMDb)

The Internet Movie Database (IMDb) dataset [63] is a comprehensive and well-balanced collection of 50,000 movie reviews. This dataset has been split into two equal parts, each containing 25,000 reviews, half of which are classified as positive and the other half as negative.

This dataset poses a significant challenge for sentiment analysis models due to the complex nature of the reviews. The movie reviews in the IMDb dataset are a combination of both storylines and personal opinions, which can make it challenging for models to accurately determine the overall sentiment of the review. The dataset contains two columns, "review" and "sentiment", and the complexity of the language used in the reviews adds a layer of difficulty to sentiment analysis models.

4.2. Twitter US Airline Sentiment

The Twitter US Airline Sentiment dataset, collected in 2017 by CrowdFlower, offers a rich collection of customer reviews of six major American airlines. This dataset provides a valuable resource for evaluating the performance of sentiment analysis models in a real-world scenario, making it a highly sought-after resource for researchers in the field.

The sentiment classes in this dataset include positive, negative, and neutral, with sample sizes of 2363, 9178, and 3099, respectively. The main challenge of this dataset is class imbalance, with the majority of the samples belonging to the negative sentiment class, which can affect the accuracy of sentiment analysis models.

In addition to class imbalance, the brevity and informality of the Tweets in the dataset present further challenges for sentiment analysis models. Tweets are often written in an informal style and are short in length, which can result in contexts and important sentiment-carrying words being missed. These challenges can result in sentiment analysis models misclassifying Tweets, making it more difficult to accurately determine the sentiment expressed in these texts.

4.3. Sentiment140

The Sentiment140 dataset [64] collected by Stanford University is a large and comprehensive collection of customer sentiment data gathered from Twitter. It consists of 1.6 million samples, evenly split between the positive and negative sentiment classes.

The short and informal nature of the Tweets in the Sentiment140 dataset can present challenges for sentiment analysis models. Tweets are typically written in a more casual and concise manner compared to other forms of text, which can result in ambiguity in the polarity of the sentiment. This can make it difficult for sentiment analysis models to accurately identify the sentiment of a Tweet, as important sentiment-carrying words or contextual information may be missed.

In addition to the brevity and informality of the Tweets, the Sentiment140 dataset also presents a unique challenge to sentiment analysis models due to the real-world scenarios it represents. Customer sentiment data gathered from social media platforms such as Twitter provide valuable insights into how customers perceive and interact with companies and their products. The Sentiment140 dataset provides an opportunity for researchers and practitioners to evaluate the performance of sentiment analysis models in a real-world scenario, where the data are more representative of the types of data they are likely to encounter in practice.

4.4. SemEval-2017 Task 4

The International Workshop on Semantic Evaluation (SemEval) has been a leading forum for evaluating the performance of sentiment analysis models on social media text data. One of the primary tasks hosted by SemEval is sentiment analysis on Twitter, which has been a shared task since 2013. The latest edition of this shared task is the SemEval-2017 Task 4 dataset, as referenced in [65].

Unlike previous SemEval datasets, which only included English data, the SemEval-2017 Task 4 dataset is a multilingual dataset, containing both English and Arabic data. The authors of the paper focused on the English portion of the dataset.

The SemEval-2017 Task 4 dataset is a comprehensive benchmark dataset for sentiment analysis models, with a total of five subtasks covering different aspects of sentiment analysis. These subtasks include message polarity classification, topic-specific message polarity classification, and Tweet quantification. Each of these subtasks has a different number of sentiment polarity classes, providing a comprehensive evaluation of sentiment analysis models in different scenarios.

A summary of the sentiment analysis datasets is provided in Table 4. The table provides a concise overview of the characteristics of the datasets, making it easy for researchers and practitioners to compare the different datasets and choose the most appropriate one for their needs.

Table 4. Summary of sentiment analysis datasets.

Dataset	Classes	Strongly Positive	Positive	Neutral	Negative	Strongly Negative	Total
IMDb	2	-	25,000	-	25,000	-	50,000
Twitter US Airline Sentiment	3	-	2363	3099	9178	-	14,160
Sentiment140	2	-	800,000	-	800,000	-	1,600,000
SemEval-2017 4A	3	-	22,277	28,528	11,812	-	62,617
SemEval-2017 4B	2	-	17,414	-	7735	-	25,149
SemEval-2017 4C	5	1151	15,254	19,187	6943	476	43,011
SemEval-2017 4D	2	-	17,414	-	7735	-	25,149
SemEval-2017 4E	5	1151	15,254	19,187	6943	476	43,011

5. Limitations and Future Research Prospects

Sentiment analysis research has made significant progress in the field but there are still challenges that limit its accuracy and applicability. These limitations include the difficulties in handling poorly structured and sarcastic texts, the lack of fine-grained sentiment analysis, the dependence on annotated data, the limitations of word embeddings, and the potential for bias in the training data.

- **Poorly Structured and Sarcastic Texts:** Many sentiment analysis methods rely on structured and grammatically correct text, which can lead to inaccuracies in analyzing informal and poorly structured texts, such as social media posts, slang, and sarcastic comments. This is because the sentiments expressed in these types of texts can be subtle and require contextual understanding beyond surface-level analysis.
- **Coarse-Grained Sentiment Analysis:** Although positive, negative, and neutral classes are commonly used in sentiment analysis, they may not capture the full range of emotions and intensities that a person can express. Fine-grained sentiment analysis, which categorizes emotions into more specific categories such as happy, sad, angry, or surprised, can provide more nuanced insights into the sentiment expressed in a text.
- **Lack of Cultural Awareness:** Sentiment analysis models trained on data from a specific language or culture may not accurately capture the sentiments expressed in texts from other languages or cultures. This is because the use of language, idioms, and expressions can vary widely across cultures, and a sentiment analysis model trained on one culture may not be effective in analyzing sentiment in another culture.
- **Dependence on Annotated Data:** Sentiment analysis algorithms often rely on annotated data, where humans manually label the sentiment of a text. However, collecting

and labeling a large dataset can be time-consuming and resource-intensive, which can limit the scope of analysis to a specific domain or language.

- **Shortcomings of Word Embeddings:** Word embeddings, which are a popular technique used in deep learning-based sentiment analysis, can be limited in capturing the complex relationships between words and their meanings in a text. This can result in a model that does not accurately represent the sentiment expressed in a text, leading to inaccuracies in analysis.
- **Bias in Training Data:** The training data used to train a sentiment analysis model can be biased, which can impact the model's accuracy and generalization to new data. For example, a dataset that is predominantly composed of texts from one gender or race can lead to a model that is biased toward that group, resulting in inaccurate predictions for texts from other groups.

To address these challenges, future research can focus on improving the accuracy and scope of sentiment analysis. This can include the development of fine-grained sentiment analysis models that capture the emotional intensity of a text, quantifying the sentiment distribution across different topics, handling ambiguous and sarcastic texts, expanding the scope of cross-lingual sentiment analysis, and improving the applicability of sentiment analysis to social media platforms.

- **Fine-Grained Sentiment Analysis:** The current sentiment analysis models mainly classify the sentiment into three coarse classes: positive, negative, and neutral. However, there is a need to extend this to a fine-grained sentiment analysis, which consists of different emotional intensities, such as strongly positive, positive, neutral, negative, and strongly negative. Researchers can explore various deep learning architectures and techniques to perform fine-grained sentiment analysis. One such approach is to use hierarchical attention networks that can capture the sentiment expressed in different parts of a text at different levels of granularity.
- **Sentiment Quantification:** Sentiment quantification is an important application of sentiment analysis. It involves computing the polarity distributions based on the topics to aid in strategic decision making. Researchers can develop more advanced models that can accurately capture the sentiment distribution across different topics. One way to achieve this is to use topic modeling techniques to identify the underlying topics in a corpus of text and then use sentiment analysis to compute the sentiment distribution for each topic.
- **Handling Ambiguous and Sarcastic Texts:** Sentiment analysis models face challenges in accurately detecting sentiment in ambiguous and sarcastic texts. Researchers can explore the use of reinforcement learning techniques to train models that can handle ambiguous and sarcastic texts. This involves developing models that can learn from feedback and adapt their predictions accordingly.
- **Cross-lingual Sentiment Analysis:** Currently, sentiment analysis models are primarily trained on English text. However, there is a growing need for sentiment analysis models that can work across multiple languages. Cross-lingual sentiment analysis would help to better understand the sentiment expressed in different languages, making sentiment analysis accessible to a larger audience. Researchers can explore the use of transfer learning techniques to develop sentiment analysis models that can work across multiple languages. One approach is to pretrain models on large multilingual corpora and then fine-tune them for sentiment analysis tasks in specific languages.
- **Sentiment Analysis in Social Media:** Social media platforms generate huge amounts of data every day, making it difficult to manually process the data. Researchers can explore the use of domain-specific embeddings that are trained on social media text to improve the accuracy of sentiment analysis models. They can also develop models that can handle noisy or short social media text by incorporating contextual information and leveraging user interactions.

By addressing these limitations and developing more advanced models, the field of sentiment analysis has the potential to increase its range of real-world applications and provide valuable insights into the sentiment expressed in text.

6. Conclusions

Sentiment analysis is an important area in natural language processing (NLP) due to its broad range of applications. The initial works on sentiment analysis employed traditional machine learning techniques, where the text was preprocessed to remove stop words, normalize the text, and represent it using frequency-based features such as TF-IDF or bag of words. The cleaned text was then fed into machine learning algorithms such as naive Bayes, SVM, etc., for classification. However, with the advancements in NLP, researchers have shifted their focus toward deep learning techniques. In deep learning, text is first encoded into pretrained word embeddings such as GloVe, word2vec, or fastText. These embeddings capture the patterns in the text, which are then fed into deep learning models such as CNN, LSTM, GRU, etc. Some sentiment analysis works also use an ensemble approach by combining predictions from multiple machine learning or deep learning models to improve performance. Some commonly used sentiment analysis datasets include the Internet Movie Database (IMDb), Twitter US Airline Sentiment, Sentiment140, and SemEval-2017 Task 4 datasets.

Despite the progress made in sentiment analysis, existing methods are still vulnerable to poorly structured and sarcastic texts, which highlights the need for more reliable language models. Additionally, most current sentiment analysis works classify sentiments into coarse categories, such as positive, negative, and neutral, but future research should focus on fine-grained sentiment analysis that includes classes with varying emotional intensities, such as strongly positive, positive, neutral, negative, and strongly negative. Sentiment quantification, where the polarity distribution of a topic is calculated, is another area of interest that could be used for strategic decision making.

Author Contributions: Conceptualization, K.L.T. and C.P.L.; methodology, K.L.T., C.P.L. and K.M.L.; software, K.L.T. and C.P.L.; validation, K.L.T., C.P.L. and K.M.L.; formal analysis, K.L.T.; investigation, K.L.T.; resources, K.L.T.; data curation, K.L.T. and C.P.L.; writing—original draft preparation, K.L.T.; writing—review and editing, C.P.L. and K.M.L.; visualization, K.L.T. and C.P.L.; supervision, C.P.L. and K.M.L.; project administration, C.P.L.; funding acquisition, C.P.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research in this work was supported by the Fundamental Research Grant Scheme of the Ministry of Higher Education under award number FRGS/1/2021/ICT02/MMU/02/4; the Deanship of Scientific Research, King Khalid University, Saudi Arabia, under grant number (RGP1/357/43); and the Multimedia University Internal Research Grant under award number MMUI/220021.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ligthart, A.; Catal, C.; Tekinerdogan, B. Systematic reviews in sentiment analysis: A tertiary study. *Artif. Intell. Rev.* **2021**, *54*, 4997–5053. [[CrossRef](#)]
2. Dang, N.C.; Moreno-García, M.N.; De la Prieta, F. Sentiment analysis based on deep learning: A comparative study. *Electronics* **2020**, *9*, 483. [[CrossRef](#)]
3. Chakriswaran, P.; Vincent, D.R.; Srinivasan, K.; Sharma, V.; Chang, C.Y.; Reina, D.G. Emotion AI-driven sentiment analysis: A survey, future research directions, and open issues. *Appl. Sci.* **2019**, *9*, 5462. [[CrossRef](#)]
4. Jung, Y.G.; Kim, K.T.; Lee, B.; Youn, H.Y. Enhanced Naive Bayes classifier for real-time sentiment analysis with SparkR. In Proceedings of the 2016 IEEE International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 19–21 October 2016; pp. 141–146.

5. Athindran, N.S.; Manikandaraj, S.; Kamaleshwar, R. Comparative analysis of customer sentiments on competing brands using hybrid model approach. In Proceedings of the 2018 IEEE 3rd International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 15–16 November 2018; pp. 348–353.
6. Vanaja, S.; Belwal, M. Aspect-level sentiment analysis on e-commerce data. In Proceedings of the 2018 IEEE International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 11–12 July 2018; pp. 1275–1279.
7. Iqbal, N.; Chowdhury, A.M.; Ahsan, T. Enhancing the performance of sentiment analysis by using different feature combinations. In Proceedings of the 2018 IEEE International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 8–9 February 2018; pp. 1–4.
8. Rathi, M.; Malik, A.; Varshney, D.; Sharma, R.; Mendiratta, S. Sentiment analysis of tweets using machine learning approach. In Proceedings of the 2018 IEEE Eleventh International Conference on Contemporary Computing (IC3), Noida, India, 2–4 August 2018; pp. 1–3.
9. Tariyal, A.; Goyal, S.; Tantububay, N. Sentiment Analysis of Tweets Using Various Machine Learning Techniques. In Proceedings of the 2018 IEEE International Conference on Advanced Computation and Telecommunication (ICACAT), Bhopal, India, 28–29 December 2018; pp. 1–5.
10. Hemakala, T.; Santhoshkumar, S. Advanced classification method of twitter data using sentiment analysis for airline service. *Int. J. Comput. Sci. Eng.* **2018**, *6*, 331–335. [[CrossRef](#)]
11. Rahat, A.M.; Kahir, A.; Masum, A.K.M. Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset. In Proceedings of the 2019 IEEE 8th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 22–23 November 2019; pp. 266–270.
12. Makhmudah, U.; Bukhori, S.; Putra, J.A.; Yudha, B.A.B. Sentiment Analysis of Indonesian Homosexual Tweets Using Support Vector Machine Method. In Proceedings of the 2019 IEEE International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), Jember, Indonesia, 16–17 October 2019; pp. 183–186.
13. Wongkar, M.; Angdresey, A. Sentiment analysis using Naive Bayes Algorithm of the data crawler: Twitter. In Proceedings of the 2019 IEEE Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 16–17 October 2019; pp. 1–5.
14. Madhuri, D.K. A machine learning based framework for sentiment classification: Indian railways case study. *Int. J. Innov. Technol. Explor. Eng. (IJITEE)* **2019**, *8*, 441–445.
15. Gupta, A.; Singh, A.; Pandita, I.; Parashar, H. Sentiment analysis of Twitter posts using machine learning algorithms. In Proceedings of the 2019 IEEE 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 13–15 March 2019; pp. 980–983.
16. Prabhakar, E.; Santhosh, M.; Krishnan, A.H.; Kumar, T.; Sudhakar, R. Sentiment analysis of US Airline Twitter data using new AdaBoost approach. *Int. J. Eng. Res. Technol. (IJERT)* **2019**, *7*, 1–6.
17. Hourrane, O.; Idrissi, N. Sentiment Classification on Movie Reviews and Twitter: An Experimental Study of Supervised Learning Models. In Proceedings of the 2019 IEEE 1st International Conference on Smart Systems and Data Science (ICSSD), Rabat, Morocco, 3–4 October 2019; pp. 1–6.
18. AlSalman, H. An improved approach for sentiment analysis of arabic tweets in twitter social media. In Proceedings of the 2020 IEEE 3rd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 19–21 March 2020; pp. 1–4.
19. Saad, A.I. Opinion Mining on US Airline Twitter Data Using Machine Learning Techniques. In Proceedings of the 2020 IEEE 16th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 29–30 December 2020; pp. 59–63.
20. Alzyout, M.; Bashabsheh, E.A.; Najadat, H.; Alaiad, A. Sentiment Analysis of Arabic Tweets about Violence Against Women using Machine Learning. In Proceedings of the 2021 IEEE 12th International Conference on Information and Communication Systems (ICICS), Valencia, Spain, 24–26 May 2021; pp. 171–176.
21. Jemai, F.; Hayouni, M.; Baccar, S. Sentiment Analysis Using Machine Learning Algorithms. In Proceedings of the 2021 IEEE International Wireless Communications and Mobile Computing (IWCMC), Harbin, China, 28 June–2 July 2021; pp. 775–779.
22. Ramadhani, A.M.; Goo, H.S. Twitter sentiment analysis using deep learning methods. In Proceedings of the 2017 IEEE 7th International Annual Engineering Seminar (InAES), Yogyakarta, Indonesia, 1–2 August 2017; pp. 1–4.
23. Demirci, G.M.; Keskin, Ş.R.; Doğan, G. Sentiment analysis in Turkish with deep learning. In Proceedings of the 2019 IEEE International Conference on Big Data, Honolulu, HI, USA, 29–31 May 2019; pp. 2215–2221.
24. Raza, G.M.; Butt, Z.S.; Latif, S.; Wahid, A. Sentiment Analysis on COVID Tweets: An Experimental Analysis on the Impact of Count Vectorizer and TF-IDF on Sentiment Predictions using Deep Learning Models. In Proceedings of the 2021 IEEE International Conference on Digital Futures and Transformative Technologies (ICoDT2), Islamabad, Pakistan, 20–21 May 2021; pp. 1–6.
25. Dholpuria, T.; Rana, Y.; Agrawal, C. A sentiment analysis approach through deep learning for a movie review. In Proceedings of the 2018 IEEE 8th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 24–26 November 2018; pp. 173–181.
26. Harjule, P.; Gurjar, A.; Seth, H.; Thakur, P. Text classification on Twitter data. In Proceedings of the 2020 IEEE 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), Jaipur, India, 7–8 February 2020; pp. 160–164.

27. Uddin, A.H.; Bapery, D.; Arif, A.S.M. Depression Analysis from Social Media Data in Bangla Language using Long Short Term Memory (LSTM) Recurrent Neural Network Technique. In Proceedings of the 2019 IEEE International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 11–12 July 2019; pp. 1–4.
28. Alahmary, R.M.; Al-Dossari, H.Z.; Emam, A.Z. Sentiment analysis of Saudi dialect using deep learning techniques. In Proceedings of the 2019 IEEE International Conference on Electronics, Information, and Communication (ICEIC), Auckland, New Zealand, 22–25 January 2019; pp. 1–6.
29. Yang, Y. Convolutional neural networks with recurrent neural filters. *arXiv* **2018**, arXiv:1808.09315.
30. Goularas, D.; Kamis, S. Evaluation of deep learning techniques in sentiment analysis from Twitter data. In Proceedings of the 2019 IEEE International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey, 26–28 August 2019; pp. 12–17.
31. Hossain, N.; Bhuiyan, M.R.; Tumpa, Z.N.; Hossain, S.A. Sentiment analysis of restaurant reviews using combined CNN-LSTM. In Proceedings of the 2020 IEEE 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 1–3 July 2020; pp. 1–5.
32. Tyagi, V.; Kumar, A.; Das, S. Sentiment Analysis on Twitter Data Using Deep Learning approach. In Proceedings of the 2020 IEEE 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 18–19 December 2020; pp. 187–190.
33. Rhanoui, M.; Mikram, M.; Yousfi, S.; Barzali, S. A CNN-BiLSTM model for document-level sentiment analysis. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 832–847. [[CrossRef](#)]
34. Jang, B.; Kim, M.; Harerimana, G.; Kang, S.U.; Kim, J.W. Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism. *Appl. Sci.* **2020**, *10*, 5841. [[CrossRef](#)]
35. Chundi, R.; Hulipalled, V.R.; Simha, J. SAEKCS: Sentiment analysis for English–Kannada code switchtext using deep learning techniques. In Proceedings of the 2020 IEEE International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 10–11 July 2020; pp. 327–331.
36. Thinh, N.K.; Nga, C.H.; Lee, Y.S.; Wu, M.L.; Chang, P.C.; Wang, J.C. Sentiment Analysis Using Residual Learning with Simplified CNN Extractor. In Proceedings of the 2019 IEEE International Symposium on Multimedia (ISM), San Diego, CA, USA, 9–11 December 2019; pp. 335–3353.
37. Janardhana, D.; Vijay, C.; Swamy, G.J.; Ganaraj, K. Feature Enhancement Based Text Sentiment Classification using Deep Learning Model. In Proceedings of the 2020 IEEE 5th International Conference on Computing, Communication and Security (ICCCS), Bihar, India, 14–16 October 2020; pp. 1–6.
38. Chowdhury, S.; Rahman, M.L.; Ali, S.N.; Alam, M.J. A RNN Based Parallel Deep Learning Framework for Detecting Sentiment Polarity from Twitter Derived Textual Data. In Proceedings of the 2020 IEEE 11th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 17–19 December 2020; pp. 9–12.
39. Vimali, J.; Murugan, S. A Text Based Sentiment Analysis Model using Bi-directional LSTM Networks. In Proceedings of the 2021 IEEE 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 8–10 July 2021; pp. 1652–1658.
40. Anbukkarasi, S.; Varadhaganapathy, S. Analyzing Sentiment in Tamil Tweets using Deep Neural Network. In Proceedings of the 2020 IEEE Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 11–13 March 2020; pp. 449–453.
41. Kumar, D.A.; Chinnalagu, A. Sentiment and Emotion in Social Media COVID-19 Conversations: SAB-LSTM Approach. In Proceedings of the 2020 IEEE 9th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 4–5 December 2020; pp. 463–467.
42. Hossen, M.S.; Jony, A.H.; Tabassum, T.; Islam, M.T.; Rahman, M.M.; Khatun, T. Hotel review analysis for the prediction of business using deep learning approach. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 1489–1494.
43. Younas, A.; Nasim, R.; Ali, S.; Wang, G.; Qi, F. Sentiment Analysis of Code-Mixed Roman Urdu-English Social Media Text using Deep Learning Approaches. In Proceedings of the 2020 IEEE 23rd International Conference on Computational Science and Engineering (CSE), Dubai, United Arab Emirates, 12–13 December 2020; pp. 66–71.
44. Dhola, K.; Saradva, M. A Comparative Evaluation of Traditional Machine Learning and Deep Learning Classification Techniques for Sentiment Analysis. In Proceedings of the 2021 IEEE 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Uttar Pradesh, India, 28–29 January 2021; pp. 932–936.
45. Tan, K.L.; Lee, C.P.; Anbananthen, K.S.M.; Lim, K.M. RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis with Transformer and Recurrent Neural Network. *IEEE Access* **2022**, *10*, 21517–21525. [[CrossRef](#)]
46. Kokab, S.T.; Asghar, S.; Naz, S. Transformer-based deep learning models for the sentiment analysis of social media data. *Array* **2022**, *14*, 100157. [[CrossRef](#)]
47. AlBadani, B.; Shi, R.; Dong, J.; Al-Sabri, R.; Moctard, O.B. Transformer-based graph convolutional network for sentiment analysis. *Appl. Sci.* **2022**, *12*, 1316. [[CrossRef](#)]
48. Tiwari, D.; Nagpal, B. KEAHT: A knowledge-enriched attention-based hybrid transformer model for social sentiment analysis. *New Gener. Comput.* **2022**, *40*, 1165–1202. [[CrossRef](#)] [[PubMed](#)]

49. Tesfagergish, S.G.; Kapočiūtė-Dzikiėnė, J.; Damaševičius, R. Zero-shot emotion detection for semi-supervised sentiment analysis using sentence transformers and ensemble learning. *Appl. Sci.* **2022**, *12*, 8662. [[CrossRef](#)]
50. Maghsoudi, A.; Nowakowski, S.; Agrawal, R.; Sharafkhaneh, A.; Kunik, M.E.; Naik, A.D.; Xu, H.; Razjouyan, J. Sentiment Analysis of Insomnia-Related Tweets via a Combination of Transformers Using Dempster-Shafer Theory: Pre- and Peri-COVID-19 Pandemic Retrospective Study. *J. Med Internet Res.* **2022**, *24*, e41517. [[CrossRef](#)]
51. Jing, H.; Yang, C. Chinese text sentiment analysis based on transformer model. In Proceedings of the 2022 IEEE 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAL), Sanya, China, 14–16 January 2022; pp. 185–189.
52. Alrehili, A.; Albalawi, K. Sentiment analysis of customer reviews using ensemble method. In Proceedings of the 2019 IEEE International Conference on Computer and Information Sciences (ICCIS), Aljouf, Saudi Arabia, 3–4 April 2019; pp. 1–6.
53. Bian, W.; Wang, C.; Ye, Z.; Yan, L. Emotional Text Analysis Based on Ensemble Learning of Three Different Classification Algorithms. In Proceedings of the 2019 IEEE 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Metz, France, 18–21 September 2019; Volume 2, pp. 938–941.
54. Gifari, M.K.; Lhaksmana, K.M.; Dwifabri, P.M. Sentiment Analysis on Movie Review using Ensemble Stacking Model. In Proceedings of the 2021 IEEE International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS), Bali, Indonesia, 13–14 October 2021; pp. 1–5.
55. Parveen, R.; Shrivastava, N.; Tripathi, P. Sentiment Classification of Movie Reviews by Supervised Machine Learning Approaches Using Ensemble Learning & Voted Algorithm. In Proceedings of the IEEE 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 28–29 February 2020; pp. 1–6.
56. Aziz, R.H.H.; Dimililer, N. Twitter Sentiment Analysis using an Ensemble Weighted Majority Vote Classifier. In Proceedings of the 2020 IEEE International Conference on Advanced Science and Engineering (ICOASE), Duhok, Iraq, 23–24 December 2020; pp. 103–109.
57. Varshney, C.J.; Sharma, A.; Yadav, D.P. Sentiment analysis using ensemble classification technique. In Proceedings of the 2020 IEEE Students Conference on Engineering & Systems (SCES), Prayagraj, India, 10–12 July 2020 pp. 1–6.
58. Athar, A.; Ali, S.; Sheeraz, M.M.; Bhattacharjee, S.; Kim, H.C. Sentimental Analysis of Movie Reviews using Soft Voting Ensemble-based Machine Learning. In Proceedings of the 2021 IEEE Eighth International Conference on Social Network Analysis, Management and Security (SNAMS), Gandia, Spain, 6–9 December 2021; pp. 1–5.
59. Nguyen, H.Q.; Nguyen, Q.U. An ensemble of shallow and deep learning algorithms for Vietnamese Sentiment Analysis. In Proceedings of the 2018 IEEE 5th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City, Vietnam, 23–24 November 2018; pp. 165–170.
60. Kamruzzaman, M.; Hossain, M.; Imran, M.R.I.; Bakchy, S.C. A Comparative Analysis of Sentiment Classification Based on Deep and Traditional Ensemble Machine Learning Models. In Proceedings of the 2021 IEEE International Conference on Science & Contemporary Technologies (ICSCCT), Dhaka, Bangladesh, 5–7 August 2021; pp. 1–5.
61. Al Wazrah, A.; Alhumoud, S. Sentiment Analysis Using Stacked Gated Recurrent Unit for Arabic Tweets. *IEEE Access* **2021**, *9*, 137176–137187. [[CrossRef](#)]
62. Tan, K.L.; Lee, C.P.; Lim, K.M.; Anbananthen, K.S.M. Sentiment Analysis with Ensemble Hybrid Deep Learning Model. *IEEE Access* **2022**, *10*, 103694–103704. [[CrossRef](#)]
63. Maas, A.; Daly, R.E.; Pham, P.T.; Huang, D.; Ng, A.Y.; Potts, C. Learning word vectors for sentiment analysis. In Proceedings of the IEEE 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; pp. 142–150.
64. Go, A.; Bhayani, R.; Huang, L. Twitter sentiment classification using distant supervision. *CS224N Proj. Rep. Stanf.* **2009**, *1*, 2009.
65. Rosenthal, S.; Farra, N.; Nakov, P. SemEval-2017 task 4: Sentiment analysis in Twitter. *arXiv* **2019**, arXiv:1912.00741.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.