

## Article

# Improving Sentiment Prediction of Textual Tweets Using Feature Fusion and Deep Machine Ensemble Model

Hamza Ahmad Madni <sup>1,\*</sup>, Muhammad Umer <sup>2</sup>, Nihal Abuzinadah <sup>3</sup>, Yu-Chen Hu <sup>4</sup>, Oumaima Saidani <sup>5</sup>, Shtwai Alsubai <sup>6</sup>, Monia Hamdi <sup>7</sup> and Imran Ashraf <sup>8,\*</sup>

- <sup>1</sup> College of Electronic and Information Engineering, Beibu Gulf University, Qinzhou 535011, China
  - <sup>2</sup> Department of Computer Science & Information Technology, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
  - <sup>3</sup> Faculty of Computer Science and Information Technology, King Abdulaziz University, P.O. Box 80200, Jeddah 21589, Saudi Arabia
  - <sup>4</sup> Department of Computer Science & Information Management, Providence University, Sector 7, Taiwan Boulevard, Shalu District, Taichung City 43301, Taiwan
  - <sup>5</sup> Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
  - <sup>6</sup> Department of Computer Science, College of Computer Engineering and Sciences in Al-Kharj, Prince Sattam Bin Abdulaziz University, P.O. Box 151, Al-Kharj 11942, Saudi Arabia
  - <sup>7</sup> Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
  - <sup>8</sup> Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea
- \* Correspondence: hamza@bbgu.edu.cn (H.A.M.); imranashraf@ynu.ac.kr (I.A.)

**Abstract:** Widespread fear and panic has emerged about COVID-19 on social media platforms which are often supported by falsified and altered content. This mass hysteria creates public anxiety due to misinformation, misunderstandings, and ignorance of the impact of COVID-19. To assist health professionals in addressing this epidemic more appropriately at the onset, sentiment analysis can potentially help the authorities for devising appropriate strategies. This study analyzes tweets related to COVID-19 using a machine learning approach and offers a high-accuracy solution. Experiments are performed involving different machine and deep learning models along with various features such as Word2vec, term-frequency, term-frequency document frequency, and feature fusion of both feature-generating approaches. The proposed approach combines the extra tree classifier and convolutional neural network and uses feature fusion to achieve the highest accuracy score of 99%. The proposed approach obtains far better results than existing sentiment analysis approaches.

**Keywords:** sentiment analysis; tweet classification; machine learning; COVID-19



**Citation:** Madni, H.A.; Umer, M.; Abuzinadah, N.; Hu, Y.-C.; Saidani, O.; Alsubai, S.; Hamdi, M.; Ashraf, I. Improving Sentiment Prediction of Textual Tweets Using Feature Fusion and Deep Machine Ensemble Model. *Electronics* **2023**, *12*, 1302. <https://doi.org/10.3390/electronics12061302>

Academic Editor: Alberto Fernandez Hilario

Received: 11 December 2022

Revised: 28 December 2022

Accepted: 11 January 2023

Published: 9 March 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

The spread of novel infectious COVID-19 necessitates an appropriate definition of cases, which are important for clinical diagnosis and health care surveillance. Monitoring the number of cases over time is essential for the development of effective therapies and the rate of dissemination. The World Health Organization (WHO) proclaimed the pandemic of coronavirus on 11 March 2020 [1]. The COVID-19 epidemic has wreaked havoc on the social and economic conditions of nations all over the world. It is one of the worst pandemics the entire planet has ever experienced. This COVID-19 pandemic has had a significant impact on every human's life [2].

The COVID-19 pandemic had a significant impact on each nation's medical and financial position. According to WHO, this pandemic will have an impact on the healthcare systems of about 75% of the world's nations by 2020. Beginning with a few Asian and European nations, COVID-19 will eventually extend to 220 nations worldwide by the

second part of 2020. The most tragic aspect of this virus is that its new COVID-19 form multiplies rapidly and eventually overtakes all other causes of mortality in many nations. By the end of June 2021, this terrible virus will have claimed 3,901,071 lives [3]. Every nation enforces a rigorous lockdown to protect people's lives from harm. As of right now, there have been a growing number of COVID-19-positive instances, including examples of people who have just been infected, people who have died, and those who have recovered from the virus.

The way we interact with and communicate with people has undergone constant change in the modern world, which has undergone a total transformation [4]. Regarding control, reaction, preparation, government upgrades, and media broadcasting, it is difficult to share data advancement concerning the pandemic with the general public [4,5]. There are certain unusual situations for the health-related groups as a result of the emergency scenario brought on by COVID-19 [6]. The technology connected to communication and information has substantially improved in the advanced world in which we live [7]. Similar to this, COVID-19-related misleading material and critical remarks are widely disseminated online. Public stress and concern are increased, and the health authorities' information is tainted [8].

Lockdown encourages users to read and post about COVID-19 experiences on various social media websites. People spent the majority of their time on social media sites because there was so much information available, both true and fraudulent [4]. On social media, people discuss the COVID-19 epidemic, its treatment, and the responses of various nations to this worldwide pandemic. The majority of COVID-19 data are produced on a minute-by-minute basis. There is currently no promising method for analyzing and categorizing the thoughts expressed in text tweets concerning COVID-19. Positive to negative attitudes were found in the sentiment analysis of COVID-19 tweets [9]. An urgent global problem is the increase of false information spread via social media platforms. The public is really upset as a result of these statements.

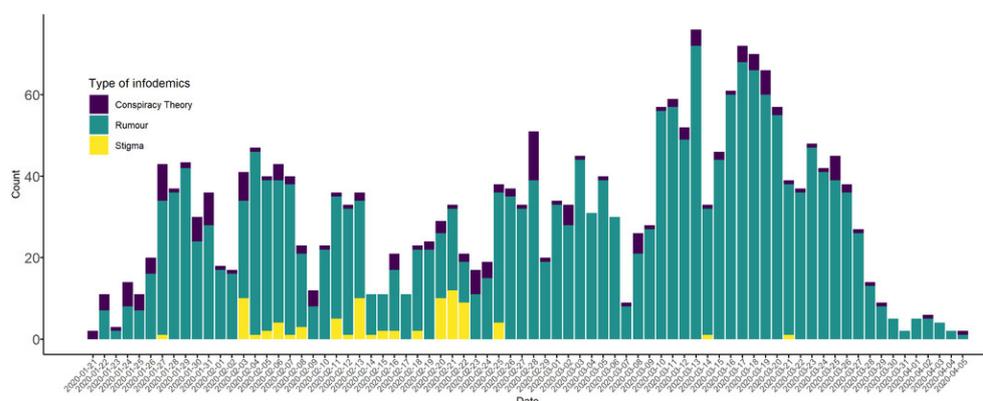
Online data about the COVID-19 pandemic have significantly increased, yet erroneous and unfavorable information is changing people's perspectives. Such false information has the potential to skew the messages that the government and authority deliver [10]. Authorities need to validate original and authentic information since false information might scare COVID patients and other individuals and cause panic [11]. Researchers have studied the public's perception of social media news stories about a coronavirus-caused pandemic in the literature [12–14]. Twitter is a social networking and microblogging site where users may post messages called "tweets".

Twitter has received over 500 million tweets and 200 billion tweets annually, making it a substantial data communication medium used by the general population worldwide [15]. Sadly, because rumors and unfavorable facts are being circulated, it is also the major cause of creating a panic scenario. Although a significant portion of COVID-19 tweets express good feelings, Chakraborty et al. highlighted that users are mostly focused on amplifying the negative tweets that contain offensive terms in the word frequency of tweets [16].

Researchers said that the COVID-19 pandemic condition caused stories and conspiracies to spread quickly, just like the virus did, and that this caused a profound change in the world [17]. Figure 1 depicts the spread of COVID-19-related rumors, stigma, and conspiracy theories that were discovered in 2020, according to the American Journal of Tropical Medicine and Hygiene [18]. The researchers provided computer models for recognizing this bogus messaging. Additionally, they looked at how misinformation regarding COVID-19's political implications influenced public health [19].

Huynh 2020 stated that there are several false reports and speculations concerning COVID-19 circulating on social media sites, making it exceedingly challenging to separate fact from fiction [20]. The public, the government, and healthcare practitioners can avoid needless anxiety if such bad material on social networking sites is authenticated and accurate. These techniques offer accurate information in every scenario involving routine events and are simple to include in conspiracy theories. For these reasons, we used machine

learning models to assess textual tweets, filling the research vacuum. We also used the comparison to validate the performance of the models.



**Figure 1.** Distribution of rumor, stigma, and conspiracy theories related to COVID-19 identified during 2020.

How did social media tweets impact people’s mental health? How do machine learning models help in analyzing public moods and emotions? A dataset comprising the opinions of the general population is necessary to respond to these queries. The significant features that could aid in the precise classification of COVID tweets into positive, negative, and neutral classes are extracted in this study using a tweet-based dataset, various feature engineering techniques individually, feature fusion with multiple textual pre-processing techniques, and diverse feature engineering techniques.

The impact of various classifiers, including the random forest Classifier (RF), extra tree classifier (ETC), gradient boosting machine (GBM), logistic regression (LR), Naive Bayes (NB), stochastic gradient descent classifier (SGD), multi layer perceptron (MLP), Recurrent neural network (RNN), Long short term memory (LSTM), convolutional neural network (CNN), and two voting classifiers, will also be examined in this study. Among voting classifiers, one combines SGD and LR and is known as VC(LR+SGD), while the other combines ETC and CNN and is known as VC(ETC+CNN). Analysis has also been conducted on other feature engineering approaches including term frequency (TF), term frequency-inverse document frequency (TF-IDF), feature fusion of both feature-generating methods (TF + TF-IDF), and Word2vec. The accuracy, precision, recall, and F1-score of the classifiers have all been assessed.

The remaining part of the paper is divided into the following sections: The relevant research is outlined in Section 2 along with a description of their approach. The description of the dataset, the preparation procedures, the specifics of the suggested methodologies, and some background information on the state-of-the-art models that we have utilized are included in Section 3. Results and discussion are given in Section 4. The study concluded in Section 5 with a review of the findings from our research and recommendations for the future.

## 2. Related Work

Natural Language Processing (NLP) is essential to the advancement of people. Almost every aspect of life makes considerable use of AI-based sentiment analysis algorithms. Many sentiment analyses have been conducted during the COVID-19 pandemic to learn more about this infectious illness. This part will provide emotive and textual analysis of tweets relating to COVID-19, as well as machine learning models, natural language processing, and Twitter. To manage Twitter data effectively, there are a few major problems. Textual analytics deals with character analysis and evocation, text visualization, semantics, and grammatical issues, as well as the endogenous and exogenous elements of these tools. Sarcasm and irony recognition [21], sentiment and opinion mining [22], false news detection [23], medical-related text mining and many more applications have made substantial use of text analysis.

Emotion analysis using Twitter data is common practice. For gathering, health-related feedback Twitter maintains the top position. Sentiment analysis of 24,000 tweets on COVID-19 was conducted in India [24]. The psychological impact of COVID-19 on human behavior was analyzed in another study [25]. Due to the COVID-19 news, it appears that the individuals are in a particularly difficult condition and have a significant degree of despair. Information with short text and lengthy text has been classified using a variety of methods. NB and LR perform poorly on the long text, yielding average scores on the short text of 91% and 74%, respectively [26]. The 4 million COVID-19 tweets from the time frame of 1 March 2020, to 21 April 2020 were examined by Xue et al. [27] using 25 distinct hashtags. Five courses were created utilizing the 13 subjects that were determined. The identification of uni-grams, bi-grams, themes, silent subjects, and feelings in tweets is conducted using LDA (Latent Dirichlet Allocation). When addressing health-related concerns, accuracy and outcomes are crucial. Another research study used 2500 short text adverts and 2500 lengthy text messages to identify mood [28]. Depression is the more prevalent feeling. Long-term residence at home, COVID-19 testing results, and unemployment are the key contributors to depression [29]. A model known as Bidirectional Encoder Representations for Transformers (BERT) was developed to investigate emotions. BERT is capable of assigning both single and multiple labels [30]. The model's key selling point is that it can take into account emoticons, which are useful tools for expressing emotions. Pattern-based sentiment analysis using the FP-growth method was developed by Drias et al. [31].

For a certain amount of time, Chakraborty et al. [16] focused on the sentiment analysis of COVID-19 tweets. The fact that they included re-tweets in their analysis is a plus for it. In their research, there are two intervals. The majority of tweets in the first period are negative or neutral, whereas those in the second interval are positive or neutral. The classifier built using deep learning obtained an accuracy of 81%. Support Vector Machine (SVM) is one of the supervised machine learning techniques that Balahur et al. [32] used to analyze the Twitter dataset (SVM). Their methods' performance on Twitter data conclusively demonstrates that the uni-gram and bi-gram method produces superior outcomes than the SVM. Modifiers, tags, and emotional words are included in the findings, which can improve how well the movements are doing. Leskovec et al. [33] presented a network for social media analysis, modeling, and optimization. Their study included a brief explanation of the methods for gathering data from social media sites, analyzing that data, and drawing conclusions from that analysis. They also keep an eye on the way emotions move across the network, and they analyze how polarization is developing in that process to evaluate the tweets' emotional content.

To examine the sentiment, authors [34] used the Malayalam Twitter dataset and several machine learning methods. The classification of the tweets into positive and negative classes using a variety of machine learning algorithms, including NB, SVM, and RF. Working on the sentiment analysis for Twitter were Imamah and Rachman [35]. They view the tweets as being directly about COVID-19. The COVID-19 tweets dataset, which was compiled on 30 April 2020, was the source of the data for this study. The word weighting Term Frequency-Inverse Document Frequency (TF-IDF) and Logistic Regression (LR) techniques were used to categorize the 355,384 tweets that were composed and connected to COVID-19. Their maximum level of accuracy was 94.71%. Those responsible for the COVID-19-related tweets were Shahsavari et al. [17]. The primary goal of their study is to identify false propaganda news in tweets about COVID-19. Therefore, their study is the finest resource to deal with false information on COVID-19.

As false information on COVID-19 might quickly harm public health, Chintalapudi et al. [36] worked on the sentiment analysis of COVID-19 tweets. They used a variety of deep learning and machine learning models in their investigation. They used the BERT (bi-directional encoder representation from the transformer) model to analyze the data in their investigation. BERT is a deep learning model for text analysis, and its performance is compared with other models. They have employed supervised machine learning algorithms

such as SVM, LR, and LSTM for comparison. The accuracy of the BERT is 89%, compared to the accuracy of the SVM, LR, and LSTM, which is 74.75%, 75%, and 65%, respectively.

To gather, store, organize, and analyze Twitter and Twitter user data, Carvalho et al. [37] presented an effective system called MISNIS (intelligent mining of public social networks influence on society). With the help of this technology, a non-technical user may rapidly mine the data and readily record tweets in Portuguese that are in the flow. The COVID-19 public sentiment insights for the categorization of tweets were developed by Samuel et al. [26]. They use tweets to gradually reveal information about equity among the community members in the US throughout the COVID-19 epidemic.

On these lengthy tweets, neither model does well. Government policies about the COVID-19 conversation on Twitter were developed by Lopez et al. [38]. In this research, multilingual Twitter data from many groups and nations are studied to determine the widely-accepted course of action in the epidemic. Kaila et al. [39] worked on the subject of modeling for COVID-19. Out of the 18,000 tweets about COVID-19, they choose a random selection. They also compute the emotions using the NRC sentiment lexicon. Han et al. [40]'s subsequent study focused on Chinese citizens' opinions on COVID-19. They categorize the tweets about COVID-19 into 7 themes. Thirteen subtopics are further broken down into these seven main topics. Depoux et al. [10]'s study on the COVID-19 postings that generate fear among individuals was successful since their panic-instigating post propagated more quickly than other comparable posts. This type of message, which can make people frightened, has a long-lasting impact on the neighborhood. To quickly address the panic, they designed a system that could recognize these rumors, attitudes, and public behavior.

Naseem et al. [41] worked on the benchmark dataset of tweets connected to COVID-19. They made use of information from 90,000 tweets about COVID-19 during February and March 2020. These tweets are divided into three categories: neutral, negative, and positive. This work uses a variety of machine learning models for classification, including SVM, RF, NB, and DT. Word2vec is employed to create the baseline for machine learning classifier feature extraction techniques such as TF-IDF. In their efforts, several deep learning models are also utilized. The BERT and its variation outperform more established techniques such as TF-IDF and word embedding, according to the results. Table 1 summarized the existing studies.

**Table 1.** Comparative analysis of the existing approaches.

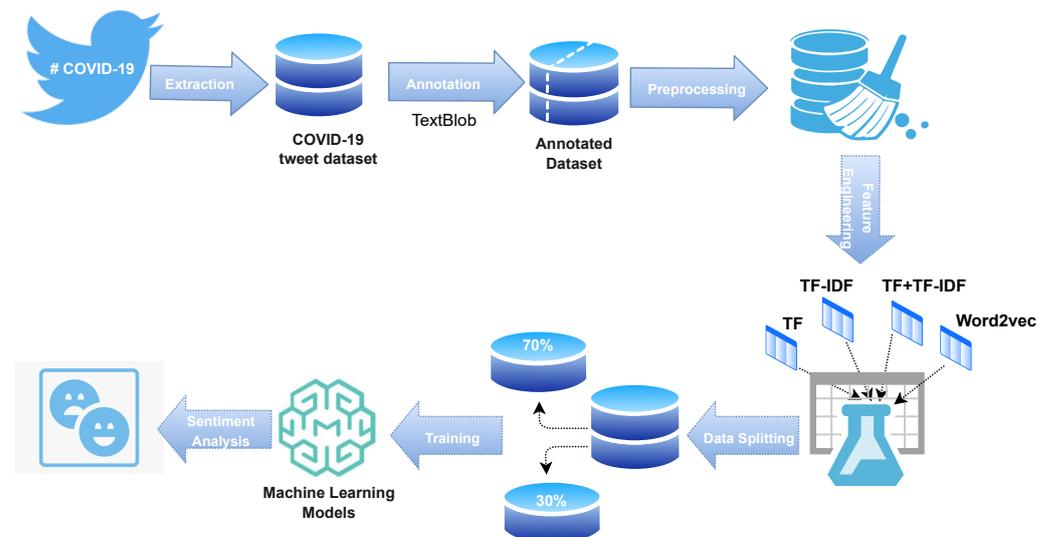
Ref.	Methods	Dataset	Findings
[24]	Wordcloud	24,000 tweets extracted using hashtags	Positive sentiments were prominent.
[27]	LDA	4 million tweets extracted using hashtags	Identified topics, themes and sentiments
[28]	Linear regression models using TF-IDF	Real World Worry Dataset	Introduced ground truth dataset
[16]	Fuzzy logic based model	226,668 tweets (December 2019 to May 2020)	The proposed model achieved 79% accuracy to perform sentiment analysis
[35]	LR and TF-IDF	COVID-19 tweets collected at 30 April 2020	The proposed model achieved 84.71% accuracy in performing sentiment analysis
[36]	BERT, LR, LSTM and SVM	3090 tweets extracted on 12 January 2021	BERT outperformed with 89% accuracy in classifying Indian tweets.
[26]	NB, KNN and LR	COVID-19 tweets February–March 2020	NB achieved 91% accuracy in finding fear-sentiment progress during COVID-19
[38]	NLP tool	Tweets collected from 22 January to 13 March 2020	Authors introduced the multilingual dataset
[39]	LDA Analysis	Random sample of 18,000 tweets	This study concluded that Twitter contains mostly accurate information related to coronavirus.
[40]	LDA and RF	Weibo texts (Sina-Weibo)	This study performed spatial and temporal analysis of COVID-19 related social media text.
[41]	SVM, NB, DT, RF, CNN, and BiLSTM	COVIDSENTI dataset	Authors analyzed public behavior during COVID-19 and introduced a new dataset.

### 3. Methods and Techniques

The experiment's methodologies and procedures are discussed in this section. In addition, the dataset description, preparation procedures, classifiers, and performance assessment matrices utilized in the trials are covered in depth.

#### 3.1. Overview of the Proposed Methodology

This study aims to provide an answer to the subject of how the COVID-19 epidemic has affected people's feelings and attitudes. A dataset including opinions about COVID-19 from the general public is needed to respond to this topic. As a result, the machine learning models used in this study's sentiment analysis of tweets connected to COVID-19. The COVID-19 tweets dataset from the IEEE data repository was used in the current study to carry out the sentiment analysis job. Then, to reduce the dataset's sparsity and identify important characteristics, several methods based on thorough preparation are applied. This research project will examine the impact of several feature engineering strategies (TF, TF-IDF, TF + TF-IDF, and Word2vec) both separately and in feature fusion during the training phase of learning algorithms. Using supervised machine learning models, the outcomes of each feature creation strategy are produced. In this work, classifiers such as RF, GBM, ETC, NB, LR, SGD, MLP, RNN, LSTM, CNN, VC(LR+SGD), and VC(ETC+CNN) were utilized. The dataset's training-to-testing ratio is 70% training to 30% testing. Accuracy, precision, recall, and F1-score are the experiment assessment criteria that will provide you with all the information you need about the experiment. Figure 2 displays the proposed model's whole design.



**Figure 2.** Workflow diagram of the proposed VC(ETC+CNN) architecture model.

#### 3.2. Dataset Description

The COVID-19-related Twitter dataset [42] was downloaded from the IEEE DataPort website. Tweet id and sentiment score of tweets make up this information. To ensure that the content in a tweet is discussing a COVID-19-related scenario, a variety of hashtags and keywords are employed in the tweet derivation process. To access complete tweet-related data, the IEEE platform's Tweet ID is hydrated. The collection has 11,858 records in total. Keywords that are used to acquire data are *sarscov2*, *COVID-19*, *quarantine*, *pandemic*, *lockdown*, and *n95*, etc. The dataset description is shown in Table 2.

**Table 2.** Number of tweets for each class using TextBlob.

Technique	Positive	Negative	Total
TextBlob	7876	3982	11,858

### 3.3. Annotating Tweets

TextBlob [43] is used to identify each instance of a tweet in terms of whether it is favorable or bad. Calculating scores between  $-1$  and  $1$  may determine phrase polarity. When the score is less than  $0$ , sentiment is classified as negative. Positive emotion is indicated if the sentiment score is greater than  $0$ :

$$Label_{Target_i} = \begin{cases} \text{Negative,} & Prob_i < 0 \\ \text{Positive,} & Prob_i > 0 \end{cases} \quad (1)$$

where  $Target_i$  is the  $i$ th tweet and  $Prob_i$  is the polarity score of  $Target_i$ .

TextBlob is a Python library and is used in text processing. To perform natural language processing tasks such as sentiment analysis, noun-phrase extraction, translation, and classification, it provides an API.

### 3.4. Data Preprocessing

Data items such as individuals, samples, observations, events, instances, vectors, points, patterns, or records are all gathered together in a dataset. These data objects are the number of features or attributes that match the fundamental properties of the data items. These traits or properties are sometimes referred to as data dimensions or dataset characteristics [44]. The majority of datasets come from numerous sources with data in various forms, as well as some raw information that is useless for machine learning models. For the machine learning models to extract usable information from the data, the quality of the raw data must be improved through data preparation. Data preprocessing in machine learning is the structuring and purging of the dataset from the stuff that is unimportant to the study. To make the raw data acceptable for the training of machine learning models, it must be transformed into something intelligible and useful. The dataset used in the current study was taken from the IEEE data port and is semi-structured/unstructured and contains a lot of extraneous data. The prediction technique does not significantly rely on unnecessary data.

Text preparation is needed to get over this restriction since big datasets increase training time and “stop words” decrease prediction accuracy. Stemming, changing all capital letters to lowercase, adding punctuation, and removing words that do not add more significance to the text are just a few of the pre-processing duties.

### 3.5. Graphical Representation of Data

We exhibit it visually so that you may see tweets about the COVID-19 dataset in greater depth. We begin by listing the most frequently used phrases from the tweet sentiment collection. The most often used phrases in the conversation are “coronavirus” and “COVID-19”. These words provide information on the keywords and subjects that were discussed the most on internet forums during the COVID-19 epidemic period. The paragraph that received the greatest attention in tweets about COVID-19 is seen in Figure 3. The word clouds of the most common positive and negative phrases in each tweet are shown in Figure 4.

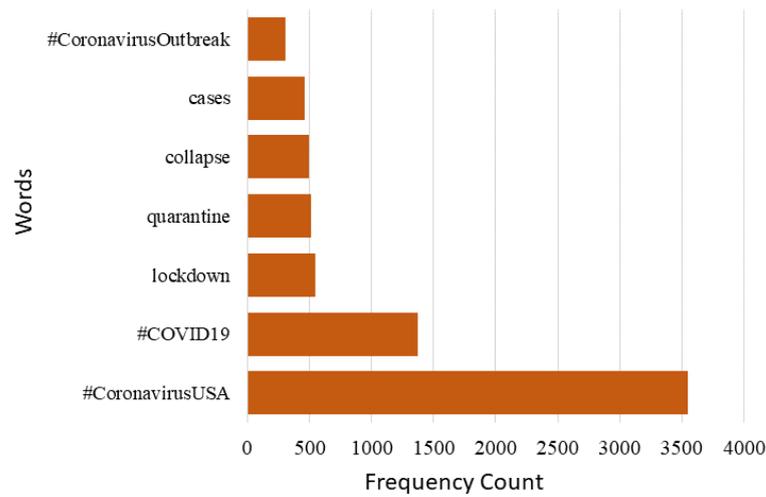


Figure 3. Frequency count of most often used terms in the dataset.

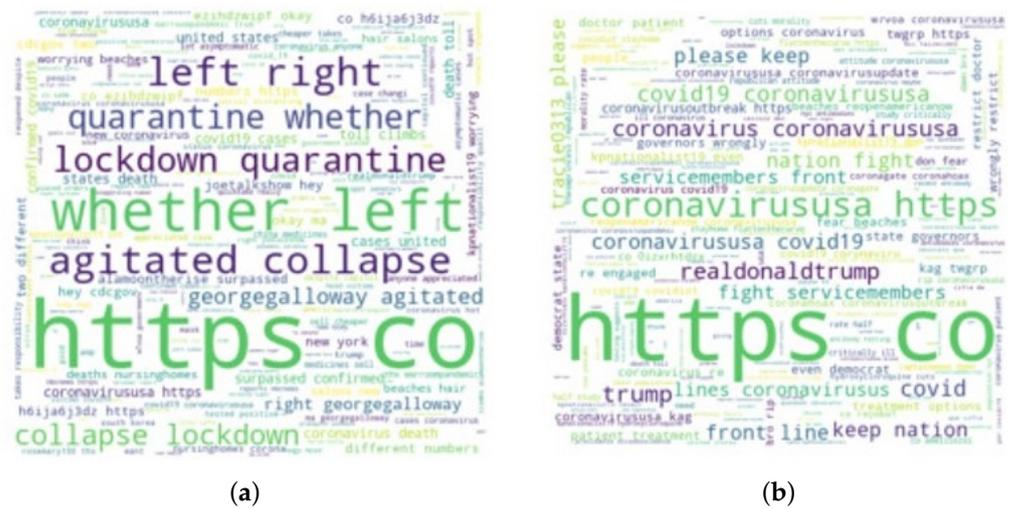


Figure 4. Word cloud of (a) positive tweets and (b) negative tweets.

### 3.6. Feature Extraction Techniques

For training, supervised machine learning techniques require vector representations of the textual input. The textual information must be converted into numbers for this purpose without losing any of the original information. Several methods, such as Bag-Of-Words (BOW), which assigns a vector value to each word, can be used to convert the data. The BOW approach does, however, have certain limits because tweets may only contain a few characters, which decreases their effectiveness. The accuracy of the BOW-based technique is limited by the lack of adequate word occurrences in textual comments or tweets [45]. Therefore, we utilize TF to handle the transition [46]. It turns a group of text documents into a matrix of integers, where each number represents the sum of all occurrences of each word in the document, and so counts the frequency of words in the document. In our suggested strategy, we also employed TF, TF/IDF, and their fusion for the feature representation [47]. TF-IDF reduces the weight of words that are often used in almost all texts and increases the weight of terms that are used in a smaller sample of documents. It penalizes some frequent terms by giving them lower weights while elevating some uncommon words in a given manuscript.

### 3.7. Data Splitting

Data splitting is the process of dividing the input dataset into two portions, mostly for the proposed models’ assessment needs. The training and test sets are these two

components. A predictive machine learning model is trained and developed using the training set, whilst the model is evaluated using the test set. The test set's size is kept small, while the training set's size is kept huge. We divided the dataset for the current investigation 70:30 into training and testing, accordingly.

### 3.8. Classifiers

To categorize tweets addressing COVID-19, this study makes use of a variety of machine learning models, including ensemble learning classifiers, regression-based models, and probability-based models. We assess the effectiveness of several deep learning and machine learning classifiers, which we briefly describe. These models are implemented in Python using *Scikit learn* [48,49]. Several machine learning approaches are combined into one predictive model using ensemble learning methods, which reduce variance (bagging), and bias (boosting), and enhance predictions (stacking) [50]. In this work, the usage of ensemble learning-based models for tweet sentiment analysis about COVID-19 is taken into consideration. The following machine-learning models were used in this study:

#### 3.8.1. Random Forest

RF [51] is one of the meta estimators that combines the data from many decision trees to increase framework effectiveness and over-fitting. It operates by fitting decision tree classifiers on a variety of input data samples. It then takes the average of each decision tree classifier's output, acting as an ensemble learner. It works by creating a huge number of decision trees, each of which provides the output class for classification or predicts the mean for regression at its nodes. RF is a classification technique that works with data by building several decision trees. Because of its simplicity and variety, RF is one of the most popular machine learning algorithms; it delivers suitable results even without modifying hyperparameters.

#### 3.8.2. Gradient Boosting Machine

GBM [52] is an ensemble model that integrates a loss function to optimally create an additive model. It operates iteratively, employing the loss function to optimize the error rate at each iteration. To reduce the prediction error, the gradient boosting algorithm's goal is to describe the results of the target variable for the following model. In that situation, the result of the target variable is dependent on a significant change in prediction, which also affects the total error. It provides a high score when the error rate for the following target prediction decreases significantly. Consequently, when the predictions of the following prediction model are more closely related to the target variable, the prediction error is reduced.

#### 3.8.3. Extra Tree Classifier

In terms of working and tree, ETC [53] is comparable to the random forest model and a tree-based model. In contrast to the random forest, it constructs trees from the actual data sample without utilizing bootstrap data, earning the additional name of a very randomized tree. From randomly chosen data, the root node of the trees is chosen using the Gini index. It was suggested to create trees while taking into account the fact that numerical input was used and choosing the best cut-point to prevent variation at each node and lessen the computational load. In issues with many dimensions and complexity, this model has produced reliable solutions. It generates multi-linear piecewise approximations as opposed to random forest's constant ones.

#### 3.8.4. Logistic Regression

LR [54] is used for classification problems and operates on a probability-based model. The logistic function is utilized to represent binary data. The link between dependent and independent variables is shown using a sigmoid function. The correlation coefficient, which measures the link between the target variable and the independent variable, is used

in LR. The correlation coefficient, which ranges from  $-1$  to  $1$ , calculates the link between two variables and shows how well the expected and actual values match up. It is a linear model, where  $Y$  is the target variable and  $X$  is the independent variable, and it may be written as  $Y = a + bX$ .

#### 3.8.5. Naive Bayes

Based on the “Bayes” theorem, NB [55] is a potent algorithm. It functions by determining the conditional probability and likelihood that a data object belongs to a specific class. The highest probability class is regarded as the ultimate. This particular feature is assumed to be unrelated to any other character in the data. If the data support the assumption, the model performs well even on a short training dataset.

#### 3.8.6. Stochastic Gradient Descent

SGD classifier [56] is a machine learning technique that identifies the ideal parameter for tying the predicted and actual outputs together. It possesses smoothness qualities and optimizes the objective function. When learning from huge datasets, it does so more quickly than gradient descent. Additionally, it converges more quickly and creates a batch from the dataset required to determine the gradient at each iteration level.

#### 3.8.7. Multilayer Perceptron

The neural network called MLP uses basic models to examine and resolve challenging problems, including prediction, classification, etc. It is a fully connected feed-forward network that can learn nonlinear relationships. MLP can be used for binary as well as challenging multi-class tasks [57].

#### 3.8.8. Recurrent Neural Network

An RNN is a neural network that has been designed to be run repeatedly, with elements from each run feeding into the following one. In particular, a portion of the input to the same hidden layer in the subsequent run comes from hidden layers from the prior run. RNNs are especially helpful for assessing sequences because the hidden layers may learn from the neural network’s past iterations on earlier portions of the sequence [58].

#### 3.8.9. Long Short Term Memory

Three gate control mechanisms are added to long short-term memory (LSTM): the forget gate, input gate, and output gate. It also presents the choice of dependent information on cell state control, which successfully solves the issue of gradient explosion and gradient disappearance [59].

#### 3.8.10. Convolutional Neural Network

CNN comprises convolutional layers, pooling layers, activation layers, dropout layers, and flattened layers. The primary layer, known as convolution, is used to extract features, while the pooling layer shrinks the size of those features, the dropout layer lessens overfitting, and the flattened layer turns data into an array. ReLU is employed as an activation function and a 0.2 dropout rate is used for the dropout layer in this study.

#### 3.8.11. Voting Classifiers

A voting classifier is a group of different models that are put together to provide aggregate predictions [60]. It operates by averaging the results from each classifier in the ensemble and then forecasts the result based on the results of the majority of votes. The advantages of LR and SGD are combined in the voting classifier VC(LR+SGD) used in this work. Another voting classifier VC(ETC+CNN) makes use of ETC and CNN for final prediction. It operates by averaging the two models’ outcomes. The model output is chosen to be the goal value with the highest probability. It combines the advantages of the integrated models and yields effective outcomes.

### 3.9. Proposed Framework

The proposed method uses a voting classifier termed VC(ETC+CNN) that combines the ET and CNN models while utilizing soft voting, as seen in Figure 5. The end result will be the class with the highest likelihood. The working of the suggested VC(ETC+CNN) is explained by Algorithm 1, which is represented as

$$\hat{p} = \operatorname{argmax}\left\{\sum_i^n \text{ETC}_i, \sum_i^n \text{CNN}_i\right\}. \quad (2)$$

where  $\sum_i^n \text{ET}_i$  and  $\sum_i^n \text{CNN}_i$  both give prediction probabilities against each test sample. After that, the probabilities for each test example by both ETC and CNN pass through the soft voting criteria.

---

**Algorithm 1** Ensembling of ETC and CNN as VC(ETC+CNN).

---

**Input:** input data  $(x, y)_{i=1}^N$

$M_{ET} = \text{Trained\_ETC}$

$M_{CNN} = \text{Trained\_CNN}$

**for**  $i = 1$  to  $M$  **do**

**if**  $M_{ET} \neq 0$  &  $M_{CNN} \neq 0$  &  $\text{training\_set} \neq 0$  **then**

$\text{ProbCNN} - \text{Pos} = M_{CNN}.\text{probability}(\text{Pos} - \text{class})$

$\text{ProbCNN} - \text{Neg} = M_{CNN}.\text{probability}(\text{Neg} - \text{class})$

$\text{ProbET} - \text{Pos} = M_{ETC}.\text{probability}(\text{Pos} - \text{class})$

$\text{ProbET} - \text{Neg} = M_{ETC}.\text{probability}(\text{Neg} - \text{class})$

    Decision function =  $\max\left(\frac{1}{N_{\text{classifier}}}, \sum_{\text{classifier}} (\text{Avg}(\text{ProbCNN} - \text{Pos}, \text{ProbETC} - \text{Pos}), \text{Avg}(\text{ProbCNN} - \text{Neg}, \text{ProbETC} - \text{Neg}))\right)$

**end if**

  Return final label  $\hat{p}$

**end for**

---

The functionality of the VC(ETC+CNN) can be discussed with an example. When a given sample passes through the ETC and CNN, a probability score is assigned to each class (that can be positive or negative). Let ETC's probability score be 0.6, 0.4 for *ProbLR - Pos* and *ProbLR - Neg* classes and CNN's probability score be 0.5 and 0.5 for *ProbSGD - Pos*, and *ProbSGD - Neg*, respectively. Then, the average probability for the two classes can be calculated as

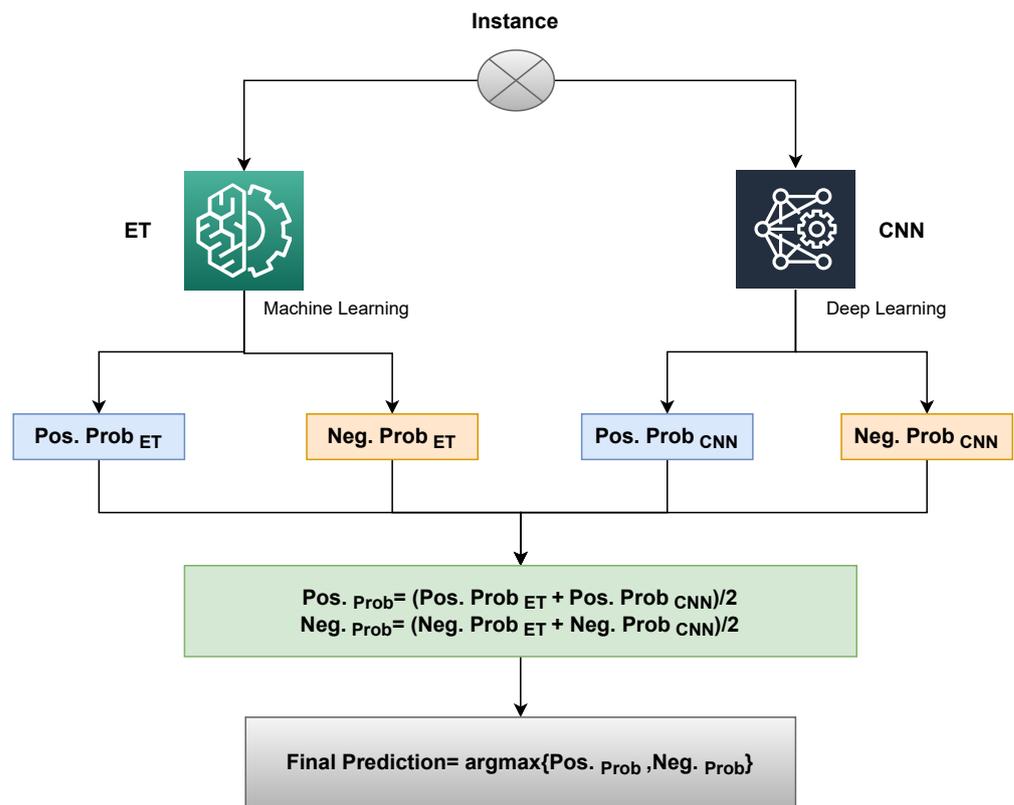
$$\text{Avg-Pos} = (0.6 + 0.5)/2 = 0.55$$

$$\text{Avg-Neg} = (0.4 + 0.5)/2 = 0.45$$

The final prediction will be positive class as shown below:

$$\text{VC(ETC+CNN)} = \operatorname{argmax}\{0.55, 0.45\} \quad (3)$$

The proposed VC(ETC+CNN) makes the final decision by combining the predicted probabilities of both classifiers and decides the final class based on the maximum average probability for a class.



**Figure 5.** Architecture of the proposed VC(ETC+CNN).

#### 4. Results and Discussion

The main objective of this study is to use machine learning models to do sentiment analysis on tweets linked to COVID-19. The COVID-19 tweets dataset from the IEEE data repository was used in the current study to carry out the sentiment analysis job. Tweets are first preprocessed through a lengthy process. After that, Textblob is used to tag tweets. In this work, classifiers such as RF, GBM, ETC, NB, LR, SGD, MLP, RNN, LSTM, CNN, VC(LR+SGD), and VC(ETC+CNN) were utilized. Each model receives individual and features fusion training on a variety of feature engineering strategies (TF, TFIDF, TF + TF-IDF, and Word2vec). Accuracy, precision, recall, and F1-score are used to evaluate the supervised machine learning and deep learning models. To analyze the sentiment of COVID-19 tweets, experimental findings utilizing several feature engineering strategies that include TF, TF-IDF, and TF +TF-IDF are examined.

##### 4.1. Performance Evaluation Matrices

This research used accuracy, precision, recall, and F1 score as the performance evaluation matrices. Four terms are the basis for these matrices such as True Positive (TP), True Negativity (TN), False Positive (FP), and False Negative (FN).

**True Positive (TP):** it refers to correctly classified positive instances;

**True Negativity (TN):** it refers to correctly classified negative instances;

**False Positive (FP):** it refers to misclassified positive instances;

**False Negative (FN):** it refers to misclassified negative instances.

Based on these terms, we can evaluate the Accuracy, Precision, Recall, and F-score.

Accuracy is a widely used parameter that is used to evaluate classifier performance. It is calculated by:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4)$$

Precision and recall are extensively used parameters for the classifier performance evaluation. Precision calculates the predicted positive case. Precision calculated as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

Out of all the above-mentioned matrices, the F1-score is calculated as well. It is a statistical measure used in the classification. It takes precision and recall of the model in its calculation and calculates the value between 0 and 1. It is calculated as:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

#### 4.2. Comparison of Classifiers Using TF

Table 3 compares machine learning models that use TF in terms of Accuracy, Precision, Recall, and F1-score. The results of the experiments show that the ETC model performed better than other models that used TF with a 94.06% Accuracy. However, SGD also achieved the second-highest results using TF, with a classification accuracy of 93.79% for tweets about COVID-19. SGD and ETC have Precision, Recall, and F1-score values of 94%. However, when it came to the sentiment analysis of the COVID-19 tweets, NB and GBM performed poorly. GBM has an 85% F1-score, 86.03% Accuracy, 88% Precision, and 86% Recall. 87.88% Accuracy, 88% Precision, 0.88% Recall, and 0.89% F1-score were displayed by NB.

**Table 3.** Comparison of classifiers using TF.

Model	Accuracy	Precision	Recall	F-Score
RF	92.52%	93%	93%	92%
ETC	94.06%	94%	94%	94%
GBM	86.03%	88%	86%	85%
LR	92.49%	93%	92%	92%
NB	87.88%	88%	88%	87%
SGD	93.79%	94%	94%	94%
MLP	83.62%	81%	85%	83%
RNN	87.72%	84%	89%	86%
LSTM	90.21%	85%	91%	88%
CNN	93.99%	94%	94%	94%
VC (LR+SGD)	92.41%	93%	92%	92%
VC(ETC+CNN)	96.62%	96%	96%	96%

#### 4.3. Comparison of Classifiers Using TF-IDF

Results are shown in Table 4 for the classification of COVID-19-related tweets by supervised machine learning models employing TF-IDF characteristics. After analyzing the outcomes, it can be concluded that TFIDF enhanced SGD and ETC model outcomes. SGD ranked second in tweet classification accuracy with a TF-IDF score of 94.01%. While NB and GBM did not outperform other machine learning models utilizing TF-IDF, they did perform better than the outcomes achieved with TF. ETC has the best precision (95% of value), followed by SGD (94%), and RF (92%). GBM achieves the lowest level of accuracy (88%). The ETC model achieves the greatest value of recall, 95%, followed by SGD (94%), which came in second, and RF (92%), which came in third. The ETC model likewise attains the greatest F1-score, which is 95%.

**Table 4.** Comparison of classifiers using TF-IDF.

Model	Accuracy	Precision	Recall	F-Score
RF	91.99%	92%	92%	92%
ETC	94.74%	95%	95%	95%
GBM	86.53%	88%	87%	86%
LR	89.91%	91%	90%	89%
NB	88.92%	89%	89%	88%
SGD	94.01%	94%	94%	94%
MLP	85.25%	79%	84%	83%
RNN	89.77%	88%	91%	90%
LSTM	90.11%	89%	92%	91%
CNN	95.65%	96%	96%	96%
VC (LR+SGD)	90.02%	91%	90%	90%
VC(ETC+CNN)	96.89%	96%	94%	95%

#### 4.4. Comparison of Classifiers Using Word2vec

Word2vec is also used to assess the effectiveness of supervised machine-learning models for sentiment analysis of COVID-19 tweets. Word2vec has established itself as a successful text classification method [61]. Table 5 experimental findings reveal that the supervised machine-learning models did not produce reliable results. The accuracy of the ETC model using word2vec, which was 88.64%, was lower than the accuracy of the ETC models using TF and TFIDF, which were 94.06% and 94.74%, respectively. The experimental findings make it abundantly evident that using the Word2vec feature representation approach does not increase the efficiency of any classifier. The ETC classifier still achieves the greatest F1-Score with word2vec (88%), which is 7% lower than the F1-score obtained using TF-IDF.

**Table 5.** Comparison of classifiers using Word2vec.

Model	Accuracy	Precision	Recall	F-Score
RF	87.44%	88%	87%	87%
ETC	88.64%	90%	89%	88%
GBM	84.45%	85%	84%	84%
LR	82.21%	82%	82%	82%
NB	68.94%	72%	69%	70%
SGD	83.36%	83%	83%	83%
MLP	80.44%	79%	85%	82%
RNN	85.56%	82%	84%	83%
LSTM	89.37%	83%	87%	85%
CNN	90.23%	92%	94%	93%
VC(LR+SG)	82.41%	82%	82%	82%
VC(ETC+CNN)	89.22%	91%	91%	91%

#### 4.5. Performance Comparison of Classifiers Using Feature Fusion

We also conducted experiments employing feature fusion (TF + TF-IDF) to compare classifiers to demonstrate the usefulness, efficiency, and robustness of machine learning models. It is evident that LR, SGD, and VC(LR+SGD) beat other models with a 92% value in terms of Accuracy, Precision, Recall, and F1-score utilizing feature fusion (TF + TF-IDF) while assessing sentiments from tweets connected to COVID-19. As demonstrated in Table 6, ETC beat other models in terms of Precision, Recall, and F1-score, with a 92% value.

With feature fusion, NB and GBM did not perform well, and their outcomes are consistent with those of TFIDF. Using feature fusion, NB outperformed Word2vec with a performance of 88.39% accuracy compared to 68.94% accuracy. For the sentiment analysis of COVID-19 tweets, GBM has also demonstrated marginally superior performance utilizing feature fusion as compared to Word2vec.

**Table 6.** Comparison of classifiers using feature fusion.

Model	Accuracy	Precision	Recall	F-Score
RF	90.83%	91.34%	91.65%	90.49%
ETC	91.93%	92.10%	92.23%	92.15%
GBM	85.49%	86.47%	85.34%	84.40%
LR	92.07%	92.22%	92.34%	92.28%
NB	88.39%	88.61%	88.56%	88.58%
SGD	92.18%	92.21%	92.45%	92.33%
MLP	93.65%	91.41%	95.77%	93.56%
RNN	91.22%	90.34%	90.48%	90.41%
LSTM	92.59%	91.87%	95.97%	93.92%
CNN	97.77%	96.14%	98.37%	97.29%
VC (LR+SGD)	92.10%	92.55%	92.62%	92.58%
VC(ETC+CNN)	99.99%	99.99%	99.96%	99.98%

#### 4.6. Performance Comparison

A performance comparison is also carried out for the proposed approach. Existing studies utilize a large range of machine learning models for sentiment analysis. For example, Ref. [62] utilizes RF, XGboost, support vector classifier (SVC), ETC, and decision tree (DT). The study reports the best results using ETC with a 93% accuracy score. Similarly, Ref. [16] made use of a support vector machine (SVM) for sentiment analysis and obtained 79% accuracy. Comparison results given in Table 7 indicate that the proposed voting classifier in this study shows better results than existing studies and archives the highest accuracy of 99%.

**Table 7.** Comparison of proposed approach with state-of-the-art approaches.

Reference	Model	Accuracy
[62]	RF	92%
[62]	XGboost	92%
[62]	SVC	89%
[62]	ETC	93%
[62]	DT	91%
[16]	SVM	79%
Proposed	VC(ETC+CNN)	99%

#### 4.7. Results of Cross-Validation

A 10-fold cross-validation is carried out to validate the performance of the proposed approach, and results are presented in Table 8. It can be observed that the proposed model provides an average accuracy of 99.5% while the average values for precision, recall, and F score are 99.5%, 99.2%, and 99.5%, respectively.

**Table 8.** Significance of VC(ETC+CNN) with k-fold validation.

K-Folds	Accuracy	Precision	Recall	F-Score
1st-Fold	0.994	0.994	0.994	0.999
2nd-Fold	0.998	0.992	0.995	0.993
3rd-Fold	0.992	0.993	0.996	0.994
4th-Fold	0.991	0.994	0.994	0.999
5th-Fold	0.998	0.999	0.992	0.995
6th-Fold	0.999	0.993	0.991	0.997
7th-Fold	0.999	0.998	0.984	0.991
8th-Fold	0.999	0.995	0.992	0.993
9th-Fold	0.994	0.999	0.991	0.995
10th-Fold	0.992	0.993	0.991	0.997
<b>Average</b>	<b>0.995</b>	<b>0.995</b>	<b>0.992</b>	<b>0.995</b>

#### 4.8. Discussion

It is evident from the findings presented above that machine learning models were successful in extracting public sentiment from a tweet-based dataset. A wide range of feature engineering approaches, including TF, TF-IDF, TF + TF-IDF, and Word2vec, have been used in several trials. Using TF, ETC, RF, LR, SGD, and VC (LR+ SGD) were able to obtain accuracy, precision, recall, and F1-score values greater than 92%. Except for GBM, all other supervised learning models employed in the experiment performed better when using TF-IDF than TF. However, employing Word2vec for sentiment analysis of tweets has not shown satisfactory results for supervised machine learning models. Results indicate that Word2vec is not enhancing classifier effectiveness when categorizing tweet-based data. We also assessed models employing feature fusion of TF and TF-IDF to demonstrate the efficacy of supervised machine learning models. The feature fusion improved the performance of LR, SGD, and VC(LR+SGD). Table 6 demonstrates that LR and SGD have obtained strong results, with 92% values for Accuracy, Recall, Precision, and F1-score. However, the values of their results are comparable to those of their voting ensemble, VC(LR+SGD). The ETC classifier outperforms all other models using TF, TFIDF, TF + TF-IDF, and Word2vec, according to experimental data. Using TF-IDF, it can be seen that ETC outperformed all other models and can analyze tweet sentiment with 94.74% accuracy, 95% precision, recall, and F1-score.

#### 5. Conclusions

The infodemic during the COVID-19 outbreak caused substantial havoc on the masses by creating confusion among the masses and lengthening the duration of the pandemic and negatively influencing public health. There is a need for an automatic method to decrease the consequences of a health emergency in preventing the spread of incorrect information. The instability of the coronavirus pandemic has the potential to spark a significant global emergency. The development of a system that can gauge the public's feelings and attitudes during such pandemics is required in light of COVID-19's potential to jeopardize the global economic, social, and healthcare systems. The goal of this study is to design an approach to evaluate tweets about COVID-19 using the sentiment information for COVID-19. Machine learning models are used to perform sentiment analysis for this purpose including RF, GBM, ETC, NB, LR, SGD, etc. In addition, several feature engineering approaches including TF, TF-IDF, TF + TF-IDF, and Word2vec are applied. Several experiments are performed to evaluate the suitability and efficiency of various models and feature engineering approaches. Experimental results suggest that the ensemble of ETC and CNN shows the best results when used with the feature fusion technique. The proposed approach is suitable to analyze tweets and their accompanied sentiments which can be very helpful to devise strategies to deal with public sentiments accordingly. This study lacks the capabilities to handle multilingual tweets, which might be seen as a promising area for future research. Similarly, the influence of other factors such as dataset size, use of emojis, and the impact of preprocessing is not performed in this research.

**Author Contributions:** Conceptualization, H.A.M. and M.U.; Data curation, O.S. and S.A.; Formal analysis, H.A.M. and Y.-C.H.; Investigation, O.S. and M.H.; Methodology, Y.-C.H.; Project administration, Y.-C.H.; Software, M.U. and S.A.; Supervision, I.A.; Validation, S.A. and I.A.; Visualization, O.S. and M.H.; Writing—original draft, H.A.M. and M.U.; Writing—review and editing, N.A. and I.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the College of Electronic and Information Engineering, Beibu Gulf University, Qinzhou 535011, China, and by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R125), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Institutional Review Board Statement:** Not applicable

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets can be found by the authors at request.

**Acknowledgments:** Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R125), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This study is also supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R333/1444).

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

## References

1. Bai, Y.; Yao, L.; Wei, T.; Tian, F.; Jin, D.Y.; Chen, L.; Wang, M. Presumed asymptomatic carrier transmission of COVID-19. *JAMA* **2020**, *323*, 1406–1407. [CrossRef] [PubMed]
2. Lades, L.K.; Laffan, K.; Daly, M.; Delaney, L. Daily emotional well-being during the COVID-19 pandemic. *Br. J. Health Psychol.* **2020**, *25*, 902–911. [CrossRef]
3. Depoux, A.; Martin, S.; Karafillakis, E.; Preet, R.; Wilder-Smith, A.; Larson, H. COVID-19 Coronavirus/Death Toll. 2021. Available online: <https://www.worldometers.info/coronavirus/coronavirus-death-toll/> (accessed on 5 November 2022).
4. Donthu, N.; Gustafsson, A. Effects of COVID-19 on business and research. *J. Bus. Res.* **2020**, *117*, 284–289. [CrossRef] [PubMed]
5. Staszkiwicz, P.; Chomiak-Orsa, I.; Staszkiwicz, I. Dynamics of the COVID-19 Contagion and Mortality: Country Factors, Social Media, and Market Response Evidence From a Global Panel Analysis. *IEEE Access* **2020**, *8*, 106009–106022. [CrossRef]
6. Guo, Y.R.; Cao, Q.D.; Hong, Z.S.; Tan, Y.Y.; Chen, S.D.; Jin, H.J.; Tan, K.S.; Wang, D.Y.; Yan, Y. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak—An update on the status. *Mil. Med. Res.* **2020**, *7*, 11. [CrossRef] [PubMed]
7. Mittal, M.; Battineni, G.; Goyal, L.M.; Chhetri, B.; Oberoi, S.V.; Chintalapudi, N.; Amenta, F. Cloud-based framework to mitigate the impact of COVID-19 on seafarers' mental health. *Int. Marit. Health* **2020**, *71*, 213–214. [CrossRef]
8. Garcia, L.P.; Duarte, E. Infodemic: Excess Quantity to the Detriment of Quality of Information about COVID-19. *Epidemiol. Serv. Saude* **2020**, *29*, e2020186. [CrossRef]
9. Hung, M.; Lauren, E.; Hon, E.S.; Birmingham, W.C.; Xu, J.; Su, S.; Hon, S.D.; Park, J.; Dang, P.; Lipsky, M.S. Social network analysis of COVID-19 Sentiments: Application of artificial intelligence. *J. Med. Internet Res.* **2020**, *22*, e22590. [CrossRef]
10. Apuke, O.D.; Omar, B. Fake news and COVID-19: Modeling the predictors of fake news sharing among social media users. *Telemat. Inform.* **2021**, *56*, 101475. [CrossRef]
11. Al-Zaman, M. COVID-19-Related social media fake news in India. *J. Media* **2021**, *2*, 100–114. [CrossRef]
12. Depoux, A.; Martin, S.; Karafillakis, E.; Preet, R.; Wilder-Smith, A.; Larson, H. The Pandemic of Social Media Panic Travels Faster than the COVID-19 Outbreak. *J. Travel Med.* **2020**, *27*, taaa031. [CrossRef] [PubMed]
13. Gao, J.; Zheng, P.; Jia, Y.; Chen, H.; Mao, Y.; Chen, S.; Wang, Y.; Fu, H.; Dai, J. Mental health problems and social media exposure during COVID-19 outbreak. *PLoS ONE* **2020**, *15*, e0231924.
14. Ahmad, A.R.; Murad, H.R. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: Online questionnaire study. *J. Med. Internet Res.* **2020**, *22*, e19556. [CrossRef] [PubMed]
15. Stats, I.L. Twitter Usage Statistics. Available online: [https://www.internetlivestats.com/twitter-statistics/?\\_ga=2.265985167.1893892026.1661193312-937589960.1661193312](https://www.internetlivestats.com/twitter-statistics/?_ga=2.265985167.1893892026.1661193312-937589960.1661193312) (accessed on 24 July 2022).
16. Chakraborty, K.; Bhatia, S.; Bhattacharyya, S.; Platos, J.; Bag, R.; Hassani, A.E. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. *Appl. Soft Comput.* **2020**, *97*, 106754. [CrossRef] [PubMed]
17. Shahsavari, S.; Holur, P.; Tangherlini, T.R.; Roychowdhury, V. Conspiracy in the time of corona: Automatic detection of COVID-19 conspiracy theories in social media and the news. *arXiv* **2020**, arXiv:2004.13783.
18. Islam, M.S.; Sarkar, T.; Khan, S.H.; Kamal, A.H.M.; Hasan, S.M.; Kabir, A.; Yeasmin, D.; Islam, M.A.; Chowdhury, K.I.A.; Anwar, K.S.; et al. COVID-19-related infodemic and its impact on public health: A global social media analysis. *Am. J. Trop. Med. Hyg.* **2020**, *103*, 1621. [CrossRef]

19. Havey, N.F. Partisan public health: How does political ideology influence support for COVID-19 related misinformation? *J. Comput. Soc. Sci.* **2020**, *3*, 319–342. [CrossRef]
20. Huynh, T.L. The COVID-19 risk perception: A survey on socioeconomics and media attention. *Econ. Bull.* **2020**, *40*, 758–764.
21. Naseem, U.; Razzak, I.; Eklund, P.; Musial, K. Towards improved deep contextual embedding for the identification of irony and sarcasm. In Proceedings of the 2020 IEEE International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–7.
22. Naseem, U.; Khan, S.K.; Razzak, I.; Hameed, I.A. Hybrid words representation for airlines sentiment analysis. In Proceedings of the Australasian Joint Conference on Artificial Intelligence, Adelaide, SA, Australia, 2–5 December 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 381–392.
23. Aggarwal, C.C.; Reddy, C.K. *Data Clustering: Algorithms and Applications*; Chapman and Hall: Boca Raton, FL, USA, 2013.
24. Barkur, G.; Vibha, G.B.K. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian J. Psychiatry* **2020**, *51*, 102089. [CrossRef]
25. Li, S.; Wang, Y.; Xue, J.; Zhao, N.; Zhu, T. The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2032. [CrossRef]
26. Samuel, J.; Ali, G.; Rahman, M.; Esawi, E.; Samuel, Y. COVID-19 public sentiment insights and machine learning for tweets classification. *Information* **2020**, *11*, 314. [CrossRef]
27. Xue, J.; Chen, J.; Hu, R.; Chen, C.; Zheng, C.; Zhu, T. Twitter discussions and concerns about COVID-19 pandemic: Twitter data analysis using a machine learning approach. *arXiv* **2020**, arXiv:2005.12830.
28. Kleinberg, B.; van der Vegt, I.; Mozes, M. Measuring emotions in the COVID-19 real world worry dataset. *arXiv* **2020**, arXiv:2004.04225.
29. Li, I.; Li, Y.; Li, T.; Alvarez-Napagao, S.; Garcia-Gasulla, D.; Suzumura, T. What Are We Depressed About When We Talk about COVID-19: Mental Health Analysis on Tweets Using Natural Language Processing. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 358–370.
30. Feng, Y.; Zhou, W. Is working from home the new norm? An observational study based on a large geo-tagged COVID-19 Twitter dataset. *arXiv* **2020**, arXiv:2006.08581.
31. Drias, H.H.; Drias, Y. Mining Twitter Data on COVID-19 for Sentiment analysis and frequent patterns Discovery. *medRxiv* **2020**. [CrossRef]
32. Balahur, A. Sentiment analysis in social media texts. In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, GA, USA, 14 June 2013; pp. 120–128.
33. Leskovec, J. Social media analytics: Tracking, modeling and predicting the flow of information through networks. In Proceedings of the 20th International Conference Companion on World Wide Web, Hyderabad, India, 28 March–1 April 2011; pp. 277–278.
34. Wirawan, N.C.; Indriati, P.P.A. Analisis Sentimen Dengan Query Expansion Pada Review Aplikasi M-Banking Menggunakan Metode Fuzzy K-Nearest Neighbor (Fuzzy k-NN). *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.* **2018**, *2548*, 964X.
35. Rachman, F.H. Twitter Sentiment Analysis of COVID-19 Using Term Weighting TF-IDF In addition, Logistic Regression. In Proceedings of the 2020 6th IEEE Information Technology International Seminar (ITIS), Surabaya, Indonesia, 14–16 October 2020; pp. 238–242.
36. Chintalapudi, N.; Battineni, G.; Amenta, F. Sentiment Analysis of COVID-19 Tweets Using Deep Learning Models. *Infect. Dis. Rep.* **2021**, *13*, 329–339. [CrossRef]
37. Carvalho, J.P.; Rosa, H.; Brogueira, G.; Batista, F. MISNIS: An intelligent platform for Twitter topic mining. *Expert Syst. Appl.* **2017**, *89*, 374–388. [CrossRef]
38. Lopez, C.E.; Vasu, M.; Gallemore, C. Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset. *arXiv* **2020**, arXiv:2003.10359.
39. Prabhakar Kaila, D.; Prasad, D.A. Informational flow on Twitter–Corona virus outbreak–topic modelling approach. *Int. J. Adv. Res. Eng. Technol. (IJARET)* **2020**, *11*, 128–134.
40. Han, X.; Wang, J.; Zhang, M.; Wang, X. Using social media to mine and analyze public opinion related to COVID-19 in China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2788. [CrossRef] [PubMed]
41. Naseem, U.; Razzak, I.; Khushi, M.; Eklund, P.W.; Kim, J. COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 1003–1015. [CrossRef] [PubMed]
42. Lamsal, R. Design and analysis of a large-scale COVID-19 tweets dataset. *Appl. Intell.* **2021**, *51*, 2790–2804. [CrossRef] [PubMed]
43. Umer, M.; Ashraf, I.; Mehmood, A.; Ullah, S.; Choi, G.S. Predicting numeric ratings for Google apps using text features and ensemble learning. *ETRI J.* **2021**, *43*, 95–108. [CrossRef]
44. Bow, S.T. *Pattern Recognition and Image Preprocessing*; CRC Press: Boca Raton, FL, USA, 2002.
45. Sriram, B.; Fuhry, D.; Demir, E.; Ferhatosmanoglu, H.; Demirbas, M. Short text classification in Twitter to improve information filtering. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, 19–23 July 2010; ACM: New York, NY, USA, 2010; pp. 841–842.
46. Scikit Learn. Scikit-Learn Feature Extraction with countVectorizer. Available online: [https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_extraction.text.Count/](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.Count/) (accessed on 5 April 2019).
47. Scikit Learn. Scikit-Learn Feature Extraction with TF/IDF. Available online: [https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_extraction.text.Tfidf/](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.Tfidf/) (accessed on 5 April 2019).

48. Hackeling, G. *Mastering Machine Learning with Scikit-Learn*; Packt Publishing Ltd.: Birmingham, UK, 2017.
49. Scikit Learn. Scikit-Learn Classification and Regression Models. Available online: [http://scikitlearn.org/stable/supervised\\_learning.html](http://scikitlearn.org/stable/supervised_learning.html) (accessed on 10 April 2019).
50. Araque, O.; Corcuera-Platas, I.; Sanchez-Rada, J.F.; Iglesias, C.A. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Syst. Appl.* **2017**, *77*, 236–246. [[CrossRef](#)]
51. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
52. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
53. Sharaff, A.; Gupta, H. Extra-tree classifier with metaheuristics approach for email classification. In *Advances in Computer Communication and Computational Sciences*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 189–197.
54. Genkin, A.; Lewis, D.D.; Madigan, D. Large-scale Bayesian logistic regression for text categorization. *Technometrics* **2007**, *49*, 291–304. [[CrossRef](#)]
55. Perez, A.; Larranaga, P.; Inza, I. Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes. *Int. J. Approx. Reason.* **2006**, *43*, 1–25. [[CrossRef](#)]
56. Gardner, W.A. Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique. *Signal Process.* **1984**, *6*, 113–133. [[CrossRef](#)]
57. Almaghrabi, M.; Chetty, G. Improving sentiment analysis in Arabic and English languages by using multi-layer perceptron model (MLP). In Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, NSW, Australia, 6–9 October 2020; pp. 745–746.
58. Sharfuddin, A.A.; Tihami, M.N.; Islam, M.S. A deep recurrent neural network with bilstm model for sentiment classification. In Proceedings of the 2018 IEEE International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 21–22 September 2018; pp. 1–4.
59. Jin, N.; Wu, J.; Ma, X.; Yan, K.; Mo, Y. Multi-task learning model based on multi-scale CNN and LSTM for sentiment classification. *IEEE Access* **2020**, *8*, 77060–77072. [[CrossRef](#)]
60. Anderson, J.L. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Clim.* **1996**, *9*, 1518–1530. [[CrossRef](#)]
61. Stein, R.A.; Jaques, P.A.; Valiati, J.F. An analysis of hierarchical text classification using word embeddings. *Inf. Sci.* **2019**, *471*, 216–232. [[CrossRef](#)]
62. Rustam, F.; Khalid, M.; Aslam, W.; Rupapara, V.; Mehmood, A.; Choi, G.S. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PLoS ONE* **2021**, *16*, e0245909. [[CrossRef](#)] [[PubMed](#)]

**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.