PLOMS Journal of
Artificial Intelligence
(PLOMS AI)

Article

Deep Neural Network for Arabic Tweets Sentiment Analysis Related to COVD-19

Jawad Hassan Alkhateeb[1,*], Najla Hamandi Alharbi[1], and Abdel-Hamid M. Emara[1,2]
[1]Department of Computer Science Taibah University, Medina, Saudi Arabia; najla.hamandi@gmail.com
[2]Computers and Systems Engineering Department, Faculty of Engineering, Al-Azhar University, Egypt; jkhateeb@taibahu.edu.sa
Correspondence: najla.hamandi@gmail.com
Received: 13-12-2021; Accepted: 10-01-2022; Published: 28-01-2022

Abstract: Aside from the Coronavirus pandemic, several other major crises erupted worldwide. Various industries have been irreparably harmed, and many organizations have succumbed to the calamity. There is an unavoidable need to examine various patterns on social media platforms in order to reduce public anxiety and misconceptions. The study examines the emotional orientations of Arabic persons who use social media, specifically the Twitter platform. From November 2020 to January 2021, we gathered data from Twitter. Tweets have been sent from a number of Arab cities. Natural Language Processing (NLP) and Machine Learning (ML) methods are used to identify whether an opinion's sentiment is favorable, negative, or neutral. This study gathers Arabic tweets and then manually annotates them to classify them as negative, positive, neutral, etc. In this work, word embedding and TFIDF are employed as vector features, with Long Short-Term Memory (LSTM) and Naive Bayes used for classification. This paper offers a learned LSTM model and a Naive Bayes model based on the collected tweets, leveraging two powerful ML methods. The LSTM model is more effective than the Nave Bayes model since its superior performance. This is noted by achieving an accuracy of 99% for the LSTM model. The analysis of this study aids various governments and corporate entities in better understanding public attitude and behavior in the face of the pandemic and making strategic decisions in response. Furthermore, this study focuses on data visualization by exhibiting an emotion plot and a word cloud. This study is an extended version of our work "the Sentiment Analysis of Arabic Tweets Related to COVID-19 Using Deep Neural Network".

Keywords: Arabic Text; Deep Learning; Machine Learning; Opinion mining; NLP; Covid 19; Coronavirus; Naïve Bayes; Sentiment Analysis; LSTM; CNN.

I. Introduction

The COVID-19 emergence and subsequent flare-up, which began in January 2020, has hugely impacted and transformed the planet. According to the World Health Organization (WHO) statistics, the number of confirmed COVID-19 infections would have reached 5 million by mid-May 2020, with over 3 million deaths globally. Despite their positive impact on slowing the spread of the epidemic, they had significant negative consequences for the general population, the economy, and people's daily lives. The public authority has been familiarized with a few required rules to prevent the spreading of the Covid-19, such as prohibitions on parties, social distancing, retail closures, and school closures. There have been anti-lockdown and anti-social-removal protests in several locations throughout the world. Given and having these difficult circumstances, the strategy makers must comprehend the people's reactions to the pandemic so that they can balance the concerns of stopping and ceasing the pandemic on one hand while ensuring people and individuals feel good and imagine people's reactions to specific events and strategy; so that the policymakers can plan earlier.

As a consequence of the notion of social distance, which is influencing public events, business, education, and pretty much every other human-related movement, people are being forced to stay at home. Individuals are also losing their places and becoming contaminated by the crown, increasing personal and community stress. According to the research of social, financial components, various sentiments, including anger, fear, excitement, contempt, concern, and others, can significantly influence an individual's behavior and dynamic. Negative, neutral, and positive are the three major classifications.

The main purpose of this work is to look at the Arabic tweets on Covid-19 and see what people think about it. This analysis is intended to aid the government in comprehending the public's perceptions and making required decisions based on them. All tweets are categorized into three main categories: neutral, negative, and positive. The process will be a lot easier if you organize your tweets. In this research, we looked at what the problems are for Arabic language processing and how to overcome them, as well as how different preprocessing processes can affect the model's classification performance. The "Arabic Language" tweets are the topic of this paper study. We exclusively deal with tweets in Arabic. Our dataset was compiled using random sampling and the query-oriented strategy. Due to the detection of the positive, negative, and neutral tweets being a critical factor, our approach focuses on classification accuracy rather than classification speed. A program has been constructed to label the data after scraping tweets linked to the Covid-19 using the API Twitter and then distributed the data among different people for classification to make the process easier. TFIDF and word embedding are two separate feature extraction algorithm types developed in this study. Then, for the classification of tweets into distinct attitudes, we constructed an ML classifier called Nave Bayes and a deep learning (DL) sequence model called LSTM. Finally, we've included a descriptive comparison and outcomes analysis. This study raised awareness about the epidemic and attempted to detect people's feelings on the Covid-19 on Twitter. The government and organizations must detect people's impressions. As a result, the proposed models for tweet detection may be beneficial for tweet classification systems. For government-related organizations, our recommended models may be relevant.

This paper is structured as follows: Section 2 describes the background and all the related essential theories. Section 3 discusses the proper literature review. Section 4 introduces all the proposed methods which are used in this research. Section 5 presents both results and discussion. Section 6 provides the conclusion and a completer summary.

II. Background

Materials Natural Language Processing (NLP), which analyses textual information for computers to understand human language, focuses on this study [1]. Text Classification (TC), which was used in this work, has many applications. Text Categorization and Text Tagging are names used to characterize TC's ability to categorize texts into one or more predefined categories [2]. Sentiment Analysis (SA) is a form of text analysis that employs computer science ideas to extract sentiment from text. We will apply the SA in this study by employing the method of ML [3]. On the other hand, ML is a technique for teaching a computer about a certain topic; it is divided into two types: supervised and unsupervised. In supervised machine learning, data is learned using labels that specify the class of each sample of data; moreover, fresh data is classified using the training set of data. Also, unsupervised machine learning occurs when there are no known class labels in the training data. The LSTM algorithm is also a deep learning method. LSTMs are a special kind of recurrent neural organization (RNN) that may be used to solve classification issues with SA and TC [4][5].

Sentiment Analysis

The analysis of public opinion SA is known as a statistical analysis of people's feelings, attitudes, and perceptions about a certain individual. Typically, opinion mining helps collect information regarding a subject's advantages and disadvantages. Finally, a product is assigned to the customer based on good feedback and a high rating. Multinational firms and businesspeople utilize opinion mining to make advertising easier [11]. Sentiment analysis is a text mining analytic tool for identifying emotions and researching national moods [15]. Figure 1 illustrates the opinion mining technique, with the following duties for each component:

1. The data gathering/collection: Data collecting is a component. The first step in opinion mining is creating a robust and dependable database. Microblogs such as Twitter, weblogs, social networks such as Facebook, and rating sites might all be used to gather the information needed. To obtain relevant information, software intended for gathering and gathering/collecting data from the web and other techniques such as web crawling may be useful [15].
2. The opinion identification: All statements from the provided texts should be separated and identified throughout this phase. The collected reviews may then be scrutinized to see which are fake and unacceptably negative. An opinion is a word that reflects a person's feelings regarding a product, service, or other classification [15].

3. The aspect extractions during this step, all current aspects are specified and extracted using processes. Selecting possible elements to enhance categorization might be extremely useful [15].
4. The opinion classification: Following the preprocessing procedure for identifying and collecting aspects, the opinions are categorized using various techniques. [15].
5. Summary of production: a summary of the material based on the findings of prior tests, the results of the opinion outcomes are compiled into a report at the development summary stage, which can take many forms such as text, charts, and so on. [15].
6. Four assessment factors may be used to evaluate the opinion classification performance regarding the accuracy, precision, recall, and the f-score. [15].

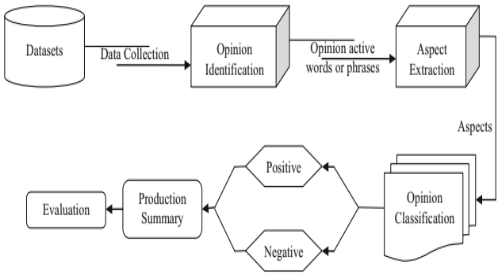

Figure 1. The Opinion Identification [13].

Due to aesthetic and theological grounds, the Arabic Language has its own significance. The fact is that 422 million people speak it as their first Language, and 250 million speak it as a second language [6]. The Holy Quran is written in Classic Arabic, one of the three Arabic language types. The second is Modern Standard Arabic (MSA), which is used in literature, the media, and education, and Colloquial Arabic, which is used in Syria, Lebanon, Iraq, and Algeria, among other locations. There are around five million Arabic users on Twitter. Saudi Arabia (KSA) has 2.4 million Twitter users, accounting for over 40% of all active Arabic users [8]. The Arabic Language has a number of difficult issues, one of which is the Language's intricate morphology, which makes normalization, tokenization, and stemming particularly difficult. Arabic words have a complicated structure that includes conjunctions, prepositions, and affixes (inflectional indicators for tens such as a number and/or gender). Another difficulty is that the Arabic words are derived from words root, which makes extracting root terms from a conventional word is a challenging task. The Short vowels are used as a pronunciation extractor and convey the correct Arabic meaning. There should be no diacritics in Arabic writing on Twitter; moreover, there are numerous dialects with no writing standards. Arabic has a large number of synonyms, making it difficult to categorize a term using the precise keyword. As a result, in addition to the previously described problems, the "linguistic code-switching" between the MSA and other dialects lowers the system's performance. In addition, there aren't adequate tools for morphology and content analysis in Arabic. By considering these problems, a research gap is already there with the sentimental analysis of the Arabic Language.

Twitter is a real-time microblogging social media network where users may express themselves in real-time. It is possible to follow sports teams, fashion companies, influencers, and other entities. In sentiment analysis, supervised, unsupervised, and mixed machine learning algorithms are used to extract characteristics from large text datasets. DL techniques for training with large data have recently received a lot of interest [14].

Machine Learning

The Machine learning (ML) techniques have been applied and used in a variety of disciplines. These technologies include pattern analysis, natural language processing, and computational learning. ML techniques enable machines to function without being explicitly programmed. Using data to create ML models forecast or make judgments based on evidence. With examples like successful internet search, computer vision, self-driving systems, and image categorization, ML has had a tremendous impact on our daily lives in recent years. Furthermore, ML approaches have substantially improved human-level artificial intelligence (AI) [18]. ML algorithms may be conceived of as a set of approaches for assessing potential patterns in a set of data. It uses previously undiscovered patterns to forecast future data (or) make judgments. Both the unsupervised and the supervised ML are the two types of ML. The target value (the label) is considered in supervised learning, but the target value is ignored in unsupervised learning. Both unsupervised learning techniques such as clustering (K-means) and supervised learning algorithms such as selection (Naive Bayes, Decision tree, and so on) are accessible [11].

The supervised text classification method looks through and analyzes material that has been classified as positive, negative, or neutral previously. It also extracts characteristics that model the distinctions between various classes and suggests a technique for finding novel examples that have never been seen earlier. [19]. According to supervised learning, sentiment analysis is a typical statistical classification with many labeled examples. The following classification technique is used as a Probabilistic Classifier in sentiment analysis [20]. The Naive Bayes classification method is the simplest and most often used of the probabilistic classifier algorithms. The Naive Bayes classification analysis computes the posterior likelihood of a class based on the distribution of words in the document. The model uses a bag of words BOWs attributes extraction, which ignores the position of the word in the text. The Bayes Theorem is used to calculate the probability that a given feature collection correlates to a given mark [13]. In the case of DL, however, the characteristics are learned automatically and represented hierarchically in several layers. In this sense, DL offers a clear advantage over traditional ML techniques [23].

Deep Learning

The methods for deep learning (DL) were originally suggested recently by the end of the twentieth century. In 2006, Hinton proposed a revolutionary deep structured learning architecture called deep belief network DBN. The DBN has been applied and used to make DL breakthroughs. Research on neuromorphic systems also helps the development of deep network models [18]. DL is considered as a subset of ML that focuses on learning data representations rather than particular methods for specific tasks. Unsupervised, partly supervised, and supervised learning are the three forms of learning. DL is also essential for analyzing vast volumes of data. Recently, a number of DL models, such as recurrent neural network (RNN) [20], have been utilized for sentiment analysis. DL algorithms have lately made great progress with popular implementations compared to traditional ML, and artificial intelligence (AI) approaches. NLP, speech recognition (SR), computation vision (CV), and picture interpretation are just a few of the applications that have been developed [18]. The term embedding, which is a distributional vector representation that stores syntactic and semantic properties of words into low-dimensional and dense vectors, has aided DL in NLP applications [21]. Neural networks (artificial intelligence), the notion of DL, were developed using artificial neural networks (ANNs). ANNs have been a major study subject in recent decades. ANNs are widely employed for classification tasks [20]. They are modeled after biological neural networks found in the human central nervous system. To establish a regular neural network NN, neurons must be used to generate real-valued activations, and the weights must be changed to ensure that the NNs behave as expected. The neuron is the aim of a neural network, which is made up of several neurons. Each neuron is composed of a collection of weights used to compute/calculate the characteristics of its inputs. The output is created based on the given data and distributions. The most recent breakthroughs in neural networks and deep learning have resulted in outstanding performance across many NLP jobs. Traditional NLP approaches have become more accurate due to RNNs [22]. RNNs are unique in that they may operate on a sequence of vectors across time. The output unit's findings are utilized as inputs in the network with the hidden layer's inputs. RNN methods have recently been applied in NLP [23].

Handcrafted engineering features are created by employing a variety of feature extraction methods, followed by applying learning algorithms in typical ML approaches. The features which are retrieved by common ML and DL algorithms vary greatly. However, in the case of DL, the qualities are learned automatically and represented hierarchically in several layers. In this regard, DL outperforms classic ML algorithms [23].

Twitter has shown to be useful for a variety of tasks, including crisis correspondence organization, public emotion monitoring, detecting anomalies and providing early warning, and so on. Twitter has also been used as a source of information in monitoring public response and well-being during disasters. Storms [10], floods [11], seismic earthquakes [12], terrorist attacks [13], and disease outbreaks are only a few examples. Scientists are attempting to tackle the problem from a fresh perspective, which involves uniquely utilizing Twitter data. During COVID-19, Catherine [14] used Twitter data to examine and define five unique techniques to analyze the subjects, important words and characteristics, data dispersion, propagation, and network activity. These researchers, Lisa [15] and Ramez [16] spoke on the spread of disinformation and evaluation of Coronavirus using

Twitter. Emotion Theory is a theory that defines the many feelings that individuals exhibit in various situations. There are six fundamental emotions. Emotion Theory addresses the many emotions that individuals display in various situations. Mainly, there are six emotions. Paul Ekman and his team are some of the main researchers [17]. He suggested that the six main feelings of people are happiness, sadness, fear, anger, disgust, and surprise. Ekman discusses that different sentiments have unique qualities transmitted to varying degrees. NLP technologies have shown to be beneficial in the field of healthcare, particularly in the area of hands-free communication. Chatbots are now being developed using a mix of NLP and AI. As a result, detecting emotions and feelings in content is a crucial topic in NLP, and there has been a large body of research on commenting on writings based on estimations and developing machine devices to identify feelings and estimations naturally. [18] [19] created a dataset for four types of emotions, namely, anger, fear, joy, and sorrow. Each piece of information is defined by the feeling class's name and the emotion's strength expressed using the Best-Worst Scaling (BWS) technique [20].

III. Methodology

Data Collection/gathering and Annotation

The Tweets from the popular Arabic hashtags such as Covid-19 and Corona were gathered. The Twitter API will be used to gather a total of 30,000 tweets. Mainly, 20,000 tweets were split into 40 files and distributed to various individuals for Annotation. The Python application was used to collect labeled tweets, which will then be transformed into data frames. The data histogram plotted in Figure 2 reveals that the majority of Arab peoples had unfavorable opinions towards the covid-19 epidemic. According to the histogram, positive tweets account for fewer than 1500, negative tweets for more than 6000, and neutral tweets for less than 6000. Figure 2 summarizes this.

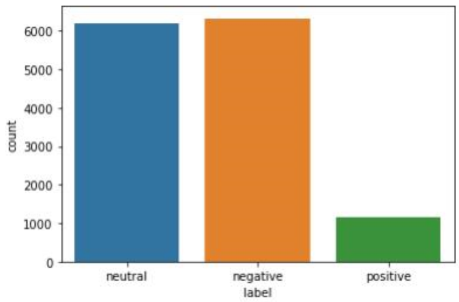

Figure 2. The Tweets counting

The recommended technique that we employed is summarized in Figure 3 as a flowchart. Using Google Colab, this study employs a sentiment analysis approach. It's a Python framework for combining ML with cloud computing. Sentiment analysis begins with the collection of a Twitter dataset relating to the specific issue that is being studied19. Second, cleaning the dataset by removing punctuations, numerical values, and emojis from the text, among other things. In addition, tokenization and the removal of Arabic stop words from the dataset were performed in conjunction with the stemming and lemmatization processes. Third, use the features extraction procedure before feeding the dataset to the ML models. Finally, the ML models are trained, and the results are analyzed.

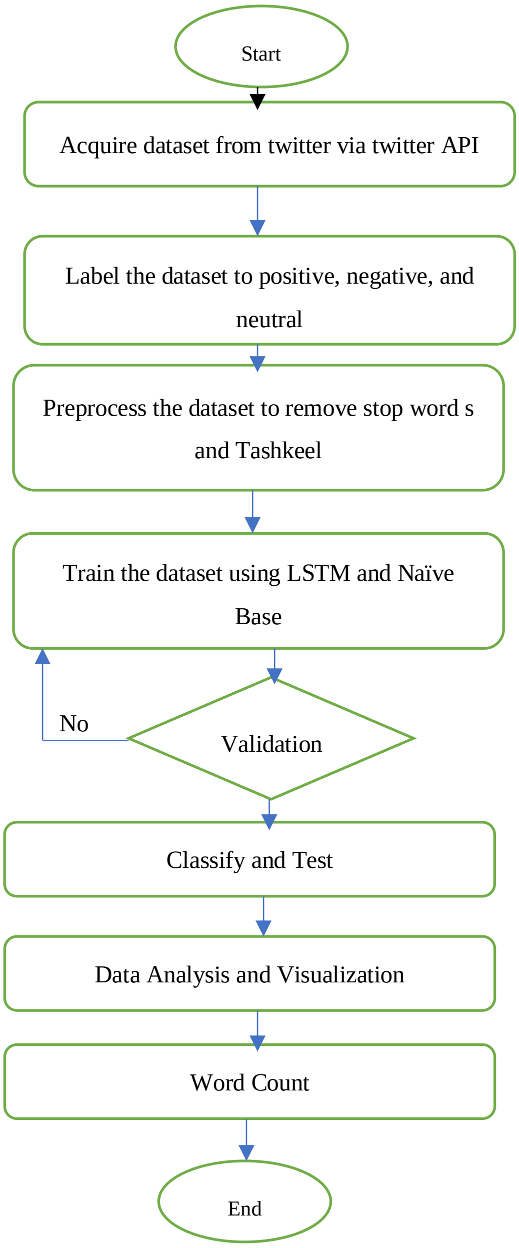

Figure 3. The Proposed Model

Preprocessing the Data

Ordinary users' text is in its raw form, and most of them abridge various words and phrases to save time. Furthermore, a large amount of text is disrupted during the text parsing process utilizing certain scrapper API. The raw text must go through a series of processes before being fed into feature extraction algorithms. Actually, the raw text is transformed into a format recognized by various feature extraction algorithms after a number of processes. The initial step in text preprocessing is to transform the raw text into appropriate tokens. Text in the form of phrase tokens is occasionally required, and text in word tokens is sometimes required. Whether we require a phrase, word, character, or any other tokenizer depends on the situation. After that, we convert the text to its simplest form, detect and eliminate noise, delete characters and words that don't convey any valuable or relevant information, normalize the words, and do a variety of additional tasks. The next sections go through some of the preparation procedures employed in this study.

Stop Words Removing

In text data, stop-words are typically regarded as noise. Actually, the noise is anything that hinders gaining insights from the text and, if not dealt with properly, can reduce the usefulness of the overall algorithm. There are a variety of criteria that may be used to determine if a word is a stop-word. For instance, if a term appears too frequently in a text, it will be deemed a stop-word. A word can be deemed a stop-word if its frequency in a text is too low. There is no uniform rule for classifying any word as a stop-word or choosing whether or not to delete one. Stop-words are defined differently by different NLP frameworks. In some cases, such as spam email filtering, emotion classification, review classification, or other text classification issues, eliminating stops words is advantageous. The following are some of the advantages of eliminating stop-words:

- It can save processing time since many words that have no bearing on the general meaning of the phrase will be eliminated as stop-words, leaving just a few tokens to be processed.
- It can enhance the algorithm's accuracy by training it on only the most crucial and meaningful words, allowing the model to learn more quickly [21].

It isn't always the case that stop words are deleted. Stop-words are beneficial, if not necessary, in some situations. Stop-words must be kept in mind while dealing with language translation, text summary, and question answering issues, among other things. Taking away stop-words in situations where they are required would have devastating consequences. Consider the following sentence: "Cats are not like dogs." When the stop-words ‘are' and ‘not' are removed, the whole meaning of the phrase is inverted. As a result, we should proceed with caution while deleting or obtaining stop-words. Because Arabic has a large number of lexical tokens, it has a large number of stopwords in its lexicon. Stop-words in Arabic generally contain the following characteristics [22].

- They don't have any significance as a single word.
- The frequency with which they appear in a sentence is quite high.
- Contribute to the completion of a sentence's or phrase's structure.
- They are rarely used as keywords.
- Even when coupled with other stop-words, it cannot create a sentence.

Extracting Features

The term "features" refers to the qualities that differentiate one item from another. Certain feature extraction approaches are now available for digital image processing, speech/audio processing, and text processing. In this study, documents, phrases, words, and characters in the text were used. Therefore, these things may be used as features. The challenge now is how to employ these things in ML. For feature extraction, we utilized two algorithms: Word embedding and TFIDF ML, and DL algorithms seldom interact with strings directly. Text data must be translated to a numeric format for these methods to work. Text is transformed to numbers, which are then processed further before being utilized by ML or DL algorithms. The TF-IDF and word embedding are two of these approaches that are used in this study.

TF-IDF

For text data, TF-IDF is a commonly applied feature extraction approach. This method has previously been used in a variety of applications. [23] To identify hate speech on Twitter, researchers used n-grams and then converted them to TF-IDF. They compared the outcomes of different classifiers. The acronym TF-IDF stands for Term frequency-inverse document frequency. It is divided into two sections: the term frequency (TF) and the inverse document frequency (IDF).

Term Frequency (TF)

Term Frequency Inverse Document Frequency (TF-IDF) is a mixture of two words: Term Frequency and Inverse Document Frequency. The word "term frequency" will be introduced first. The term frequency (TF) is used to determine how many times a phrase appears in a document [22]. The total frequency of a word appearing in a document is measured by TF, as the name implies. Suppose there is a document labeled ‘A1' with a total word count of 5000. If the word "Dog" appears ten times in a text, the TF for the word "Dog" is the ratio of the frequency of the word "Dog" to the total number of words in the document. As a result, the TF for the word 'Dog' will be 50/5000= 0.01. In this manner, the weight of a single word will be determined concerning the entire text length.

Inverse Document frequency

There are a number of words in the paper that are used a lot yet don't have any useful meaning. Stop words such as 'the,' 'a,' 'of,' 'not,' and others are overused, but their role is to provide structure to a statement without carrying any major weight in the context of the content. The issue with TF is that all such words will have more weight because of their high frequency compared to other powerful words. Inverse term frequency is a method that gives less weight to terms that frequently appear in a document and more weight to words that appear less frequently. We may compute a word's IDF as follows: Using this method, words that often appear in all publications will be given less weight than words that appear less frequently.

Term Frequency - Inverse Document Frequency (TF-IDF)

As the name suggests, the word frequency and inverse document frequency are combined in TF-IDF. It is used to determine the significance of a word in a given text among a group of papers. Now, when the term "laptop" appears many times in a document or sentence but only once or twice in other documents, it implies that this word is more essential to this document. Simply, the calculation of the TF with IDF, as mentioned [24], is a simple technique of determining the TF-IDF of a word.

Word Embedding

TF-IDF works well with tiny texts, but it is computationally costly when dealing with longer documents because it applies a new weight to each word. Furthermore, it provides no semantic information about the word's context. If we compute the distance between the words "cats" and "carrot," for example, the distance will be independent of the meaning of the terms. As a result, the relationship between 'cat' and 'carrot' will be the same as the relationship between 'cat' and 'dog,' while the distance between 'cat' and 'dog' should be considerably less than the distance between 'cat' and 'carrot.' Words must be represented in a format other than basic numbers. One method is to represent each word as a distinct vector, with the direction of the vectors indicating whether the words are similar or different. To examine classification using feature vectors, a supervised technique is used. Feature must first be retrieved for this procedure, which may be done with a unigram, bigram, or trigram. The phrase "unigram method" refers to the practice of considering each word as a separate unit without regard for context. The supervised method delivers more accurate results when tested on a dataset with the same domain as the training dataset. Its accuracy will be improved by training with a big dataset and extracting numerous characteristics [27][28].

Data Classification

Recurrent Neural Networks

Recurrent Neural Networks (RNN) is an ANN that connects nodes in a directed graph sequentially. It's referred to as a linked chain of neural network blocks. Each node communicates with the following node. Because the text is naturally sequential, this design allows RNN to operate chronologically and obtain sequential data, giving it a 'natural' approach when dealing with textual data. Simple RNNs can model lengthy data sequences; however, they cannot represent large sequences in real-time applications [21]. RNNs come in a variety of forms, but we're primarily interested in the Long Short-Term Memory network, which we utilize in our research. As you read this research, you grasp each word depending on your understanding of the preceding term. Human beings do not rethink their thoughts every minute. You don't throw anything away and start all over again. Your viewpoints are adamant. This is something that standard neural networks are incapable of accomplishing, and it appears to be a significant weakness. Recurrent neural networks are used to solve this problem. They're networks with loops that keep data from being lost. A simple recurrent network [27] is made up of one input unit, one output unit, and one recurrent hidden unit, as illustrated in Figure 4.

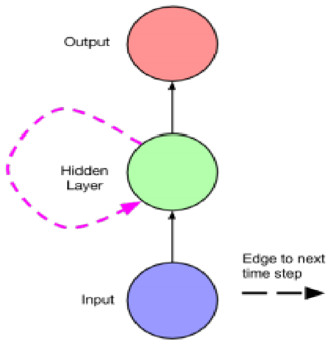

Figure 4. Simple Recurrent Neural Network (RNN) [27]

Long short-term memory (LSTM)

Word embedding is used to train neural networks based on LSTMs. We require one LSTM classifier to train an ambiguous word if no deep-network techniques are used. The LSTM is an RNN that does not have the problem of vanishing gradients and has been utilized in NLP applications (Fei Long, 2019). For time t, the input gate it, output gate of, a forget gate f t, and a cell ct are used to prevent the vanishing gradient issue. Backpropagation is used to train the three gates and memory cells that weigh on the data utilized for training. The input of LSTM is the vector xt, and the hidden output is ht. The ability of LSTM can cope successfully with lengthy dependencies, such as syntactic dependencies, is recognized to be beneficial for text analytics applications like ambiguity removal. The different LSTM value components are computed using Equations 1-5. Subscripts in a weights matrix indicate the components that are associated. For example, Whi is the weight matrix between the input gate and the concealed output.

$$i_{t=}\ \sigma(w_{xixt}\ + w_{hi}h_{t-1}\ + w_{cict-1} + b_i) \quad (1)$$

$$f_{t=}\ \sigma\big(w_{xfxt}\ + w_{hf}h_{t-1}\ + w_{cfct-1} + b_f\big) \quad (2)$$

$$c_t{=}f_t\ \ c_{t-1}{+}i_t \tanh\ (w_{xcxt}\ + w_{hc}h_t\ + b_c) \quad (3)$$

$$o_{t=}\ \sigma(w_{xoxt}\ + w_{ho}h_{t-1}\ + w_{coct} + b_o) \quad (4)$$

$$h_t\ = O_t \tanh\,(c_t) \quad (4)$$

Figure 5 summarizes the network's architecture, including RNN with two LSTM memory cells in the hidden layer.

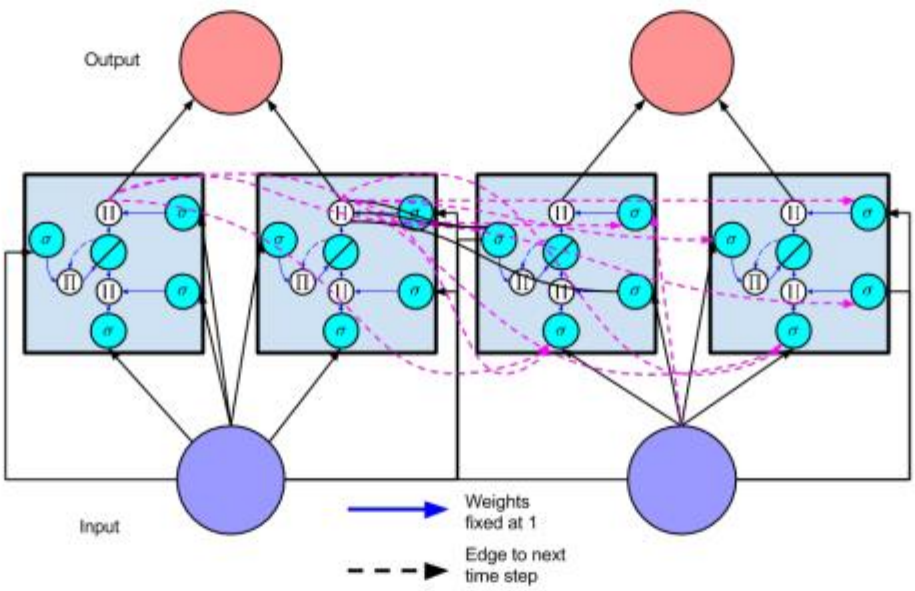

Figure 5. Recurrent Neural Network (RNN) of two memory cells LSTM in the hidden layer [27].

Naïve Bayes

This algorithm is Nave, as its name suggests. It provides all qualities or features equal weighting, and there is no impact of one attribute on another, making it more computationally efficient. The NBC Naive Bayes classifier is a proven, simple, and effective technique for text categorization [29] [30]. It has been in use since the 1950s. This classifier is theoretically based on the Bayes theorem. The greatest a posteriori propriety of the Naive Bayes classifier is used to classify (i.e., features are assigned to a class based on the highest conditional probability). The two NBC models are Multinomial Nave Bayes (which uses a binary representation of characteristics) and Bernoulli Nave Bayes (which uses frequency to describe features). Several studies have employed nBCs to categorize text, recordings, and goods. [31][32]. It uses a probabilistic model to capture unknown features of the text by estimating the likelihood of different outcomes. The statistical link between the likelihood of two occurrences A and B is illustrated in Bayes Theorem. Let P(A) denote the conditional probability of event A as a function of B, and P(B) denote the conditional probability of event B as a function of A. Bayes' theorem is quantitatively expressed in equation (6).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (6)$$

Where P(A|B) is considered the conditional probability, which means the probability of an event A occurs if B is true. Also, it is known as the posterior probability of A given B.

Evaluation of the Model

The Confusion Matrix is a technique to assess the test data model's performance based on a mixture of erroneous and correct predictions. As indicated in Table 1, the classification is divided into two groups in the confusion matrix: Predicted Class and Actual Class. Finally, this confusion matrix serves as our study's assessment model; evaluation measures, such as accuracy, precision, recall, and F score, are utilized to determine the outcome of performance measurements, as shown in Table 1.

Table 1. The Evaluation Measures

Evaluation Measures	Definition
Accuracy	"The number of instances or tweets that are correctly classified."
Precision	"The number of correctly classified positive tweets divided by the number of tweets labeled as positive by the system."
Recall	"The number of correctly classified positive tweets divided by the number of positive tweets in the dataset."
F Score	"It is the harmonic mean of P and R."

Table 2. The Confusion Matrix for Two Classes Classification Problem

"Actual Class / Predicted Class"	C1	-C1
C1	"True Positive (TP)"	"False Negative (FN)"
-C1	"False Positive (FP)"	True Negative (TN)"

IV. Results and Discussion

Word Cloud

Word clouds are a common technique to show the relative importance of words in a collection of texts. The more frequently a word appears, the more space it takes up in the picture. Word clouds may be used to assist us in gaining a sense of what the collection of texts is about, for example. Figure 5 summarizes the word cloud of all tweets, classified as bad and classified as positive—tweets with a favorable classification and tweets with a neutral classification.

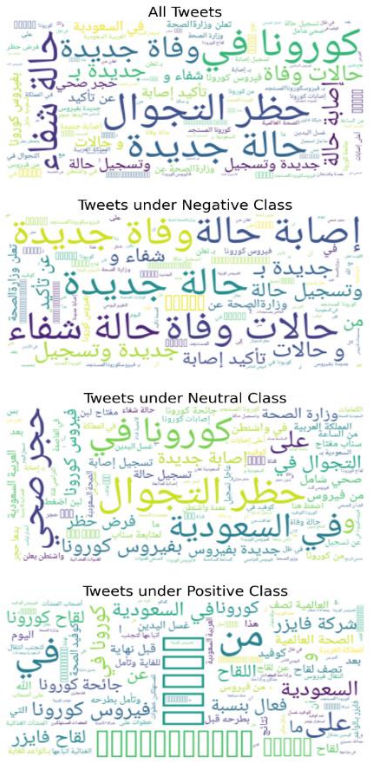

Figure 5. The Word Cloud

Accuracy of the LSTM

We can see that the LSTM model accuracy is 98.90 percent, and the training loss for the LSTM model is reduced from 0.40 to 0.05 when just five epochs are included. At the same time, the validation loss rises from 0.27 to 0.40. Figures 6 and 7 illustrate model loss and accuracy, with a 98 percent accuracy.

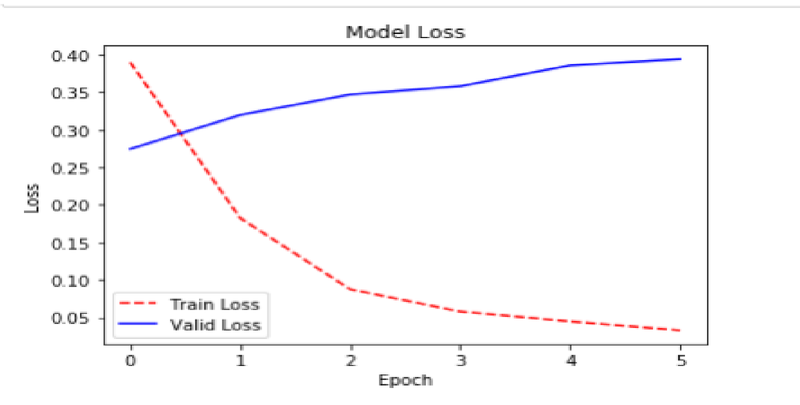

Figure 6. The Model Loss

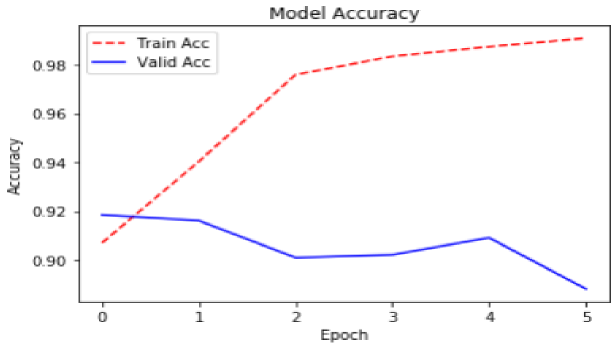

Figure 7. The Model Accuracy

Confusion Matrix of the LSTM

A confusion matrix framework is a table that is frequently used to depict the application of a grouping model (or "classifier") on a large set of test data for which the true data are known and where the values are displayed in a diagonal format that depicts true positive and true negative values that are correctly classified. Positive, neutral, and negative are the three categories in the confusion matrix. The LSTM confusion matrix is shown in Figure 8.

• For the neutral category, 11 predictions were right (neutral), 0 were incorrectly categorized as negative, and 1 was classified as positive.

• For the negative category, 2320 tweets were correctly identified as negative, whereas 59 were incorrectly labeled as positive.

• For the positive category, 158 tweets were correctly identified as positive, while 40 were incorrectly labeled as negative.

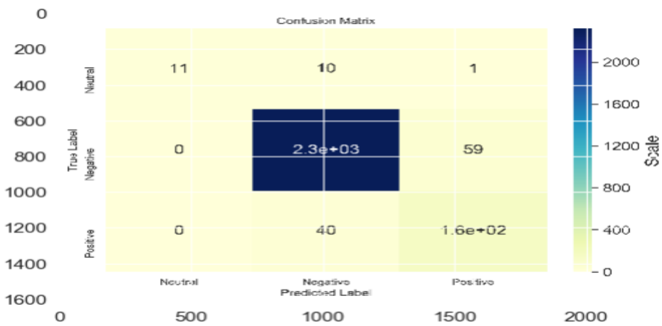

Figure 8. The LSTM Confusion Matrix

Naïve Bayes

The results of the Naive Bayes algorithm are shown, and parameters such as precision, recall, f1-score, and support are assessed; for neutral, the precision is 0.93, recall is 0.79, and f1-score is 0.86, as shown in Figure 9. The accuracy was around 76 percent.

	precision	recall	f1-score	support
neutral	0.93	0.79	0.86	717
negative	0.00	0.00	0.00	6
positive	0.13	0.37	0.19	57
accuracy			0.76	780
macro avg	0.35	0.39	0.35	780
weighted avg	0.87	0.76	0.80	780

Figure 9. Naive Bayes Accuracy

Most Frequent Words

On March 2nd, 2020, Saudi Arabia reported the first case of COVID-19. Prior to finding any case in Saudi Arabia, an early study of tweets from February 2020 revealed that there were 2,879 more negative tweets (57 percent) than good tweets (2174). (43 percent). The most common terms used in good, negative, and neutral tweets are shown in Figures 10-11-12.

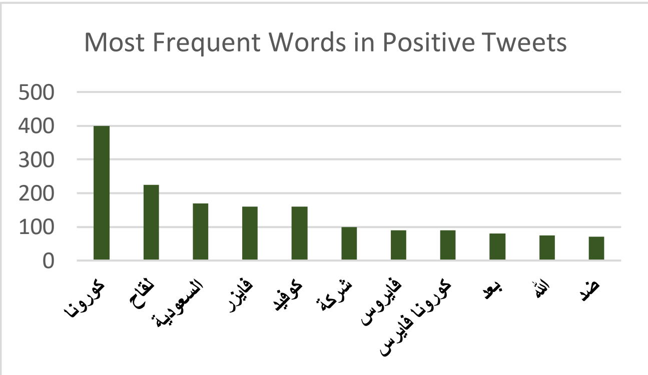

Figure 10. The Most Frequent Words in Positive Tweets

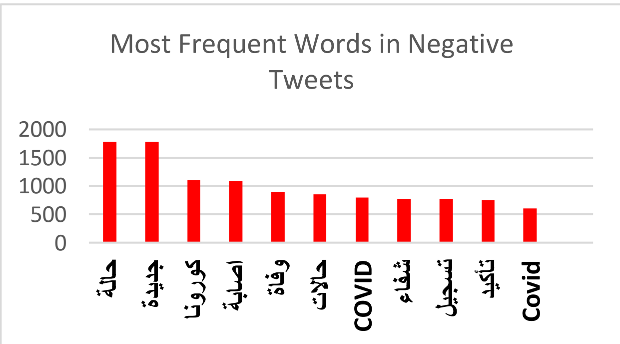

Figure 11. The Most Frequent Words in Negative Tweets

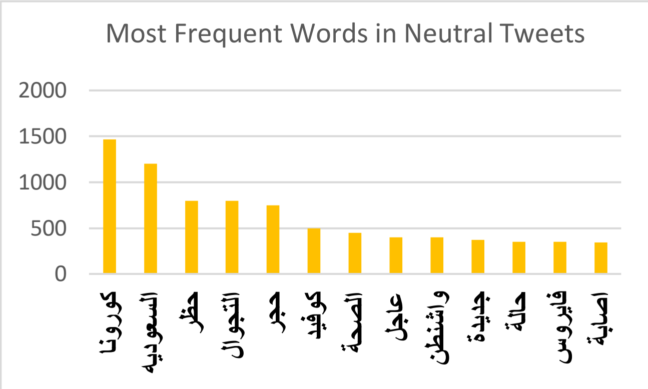

Figure 12. The Most Frequent Words in Neutral Tweets

V. Conclusion

Since the main incidence of COVID-19 was discovered in Wuhan, China, in December 2019, this new illness has sparked a global emergency. The global spread and severity of infections prompted the World Health Organization to declare COVID-19 a pandemic threat. Without mandatory vaccines, countries worldwide rushed to implement a variety of preventative measures to control the spread of the illness and, as a result, avoid a complete failure of their medical care systems. Conclusion analysis, also known as assessment mining, is a strong tool and a unique approach for assessing public perceptions and commitment to important health policies. Pandemics, such as the looming Coronavirus, are a tumultuous and rapidly evolving test requiring close observation of how people perceive the impending danger and react to laws and solutions. Such assessments are necessary to develop appropriate communication content that may address anticipated difficulties. The analysis is carried out using two techniques: LSTM network and naive Bayes, which have 98.905 and 76 percent accuracy, respectively. Furthermore, this study focuses on data visualization by showing an emotion plot and a word cloud. The application's functionality can be expanded in the future by creating a mobile app that supports it. More graphs, such as a map locating good and bad tweets, can be added in the future. Regarding various issues, more categories might be created.

VI. References

[1]. Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.

[2]. CDC, available online at: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html; accessed October 2020.

[3]. World Meter, [available online at: https://www.worldometers.info/coronavirus/; accessed October 2020.

[4]. Abirami, A. M., & Gayathri, V. (2017, January). A survey on sentiment analysis methods and approach. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 72-76). IEEE.

[5]. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.

[6]. Makris, C., Pispirigos, G., & Rizos, I. O. (2020). A distributed bagging ensemble methodology for community prediction in social networks. Information, 11(4), 199.

[7]. Heist, N., Hertling, S., & Paulheim, H. (2018). Language-agnostic relation extraction from abstracts in Wikis. Information, 9(4), 75.

[8]. Kretinin, A., Samuel, J., & Kashyap, R. (2018). When the going gets tough, the tweets get going! An exploratory analysis of tweets sentiments in the stock market. American Journal of Management, 18(5), 23-36.

[9]. De Choudhury, M., Counts, S., & Horvitz, E. (2013, April). Predicting postpartum changes in emotion and behavior via social media. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 3267-3276).

[10]. Wang, Z., Ye, X., & Tsou, M. H. (2016). Spatial, temporal, and content analysis of Twitter for wildfire hazards. Natural Hazards, 83(1), 523-540.
[11]. Skoric, M. M., Liu, J., & Jaidka, K. (2020). Electoral and Public Opinion Forecasts with Social Media Data: A Meta-Analysis. Information, 11(4), 187.
[12]. Pradhan, V. M., Vala, J., & Balani, P. (2016). A survey on sentiment analysis algorithms for opinion mining. International Journal of Computer Applications, 133(9), 7-11.
[13]. Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. Information fusion, 36, 10-25.
[14]. Alharbi, N. H., & Alkhateeb, J. H. (2021, July). Sentiment Analysis of Arabic Tweets Related to COVID-19 Using Deep Neural Network. In 2021 International Congress of Advanced Technology and Engineering (ICOTEN) (pp. 1-11). IEEE.
[15]. Wagh, R., & Punde, P. (2018, March). Survey on sentiment analysis using twitter dataset. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 208-211). IEEE.
[16]. Hemmatian, F., & Sohrabi, M. K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. Artificial Intelligence Review, 52(3), 1495-1545.
[17]. Nassif, A. B., Elnagar, A., Shahin, I., & Henno, S. (2021). Deep learning for Arabic subjective sentiment analysis: Challenges and research opportunities. Applied Soft Computing, 98, 106836.
[18]. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.
[19]. Kolchyna, O., Souza, T. T., Treleaven, P., & Aste, T. (2015). Twitter sentiment analysis: Lexicon method, machine learning method and their combination. arXiv preprint arXiv:1507.00955.
[20]. Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. Knowledge and Information Systems, 60(2), 617-663.
[21]. Badaro, G., Baly, R., Hajj, H., El-Hajj, W., Shaban, K. B., Habash, N., ... & Hamdi, A. (2019). A survey of opinion mining in Arabic: A comprehensive system perspective covering challenges and advances in tools, resources, models, applications, and visualizations. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 18(3), 1-52.
[22]. Bashar, A. (2019). Survey on evolving deep learning neural network architectures. Journal of Artificial Intelligence, 1(02), 73-82.
[23]. Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. Electronics, 8(3), 292.
[24]. Siyam, N., Alqaryouti, O., & Abdallah, S. (2020). Mining government tweets to identify and predict citizens engagement. Technology in Society, 60, 101211.
[25]. Desai, M., & Mehta, M. A. (2016, April). Techniques for sentiment analysis of Twitter data: A comprehensive survey. In 2016 International Conference on Computing, Communication and Automation (ICCCA) (pp. 149-154). IEEE.
[26]. Bing, L. (2012). Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies). University of Illinois: Chicago, IL, USA.
[27]. Alhumoud, S. (2020). Arabic Sentiment Analysis using Deep Learning for COVID-19 Twitter Data. International Journal of Computer Science and Network Security, 132-138.
[28]. Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A., & Al-Ohali, Y. (2017). Arasenti-tweet: A corpus for arabic sentiment analysis of saudi tweets. Procedia Computer Science, 117, 63-72.
[29]. Alhumoud, S. O., Altuwaijri, M. I., Albuhairi, T. M., & Alohaideb, W. M. (2015). Survey on arabic sentiment analysis in twitter. International Science Index, 9(1), 364-368.
[30]. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
[31]. Chakrabarti, S., Roy, S., & Soundalgekar, M. V. (2003). Fast and accurate text classification via multiple linear discriminant projections. The VLDB journal, 12(2), 170-185.
[32]. Alhajji, M., Al Khalifah, A., Aljubran, M., & Alkhalifah, M. (2020). Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19.