Sentiment Analysis of Text Incorporating Emojis: A Machine Learning Approach

Mona M. Abd Elsalam
Student at Department of Computer Science, Faculty of Graduate Studies for Statistical Research, Egypt

Ahmed M. Gadallah
Assoc. Prof. at Department of Computer Science, Faculty of Graduate Studies for Statistical Research, Egypt

Hesham A. Hefny
Prof.at Department of Computer Science, Faculty of Graduate Studies for Statistical Research, Egypt

ABSTRACT
Nowadays, people use emojis in their text to communicate their sentiments or summarize their words. Prior artificial intelligence (AI) strategies only included the order of text, emoticons, pictures, or emoticons with text have always been disregarded, resulting in a slew of feelings being overlooked. This study proposed a calculation and technique for opinion investigation using both text and emoticon. Awell-known sentiment analysis algorithms were inspected, including rule-based and classification algorithms, to assess the effect of enhancing emoticons as extra elements to further develop the algorithm performance. Emoticons were likewise converted into relating sentiment words when constructing features for correlation with those straightforwardly created from emoticon name words. Furthermore, considering various elements of emoticons in texts, all posts were classified in the dataset by their emoticon use and inspected the progressions in algorithm performance. Here, two methods of information were examined in joined mode with AI calculation for observing feelings from the Twitter-based Covid-19 dataset using a few highlights. Several classification techniques are used which are Random Forest, Supported Vector Machine, Gradient Boosting, K-Neighbors and Latent Dirichlet allocation (LDA). This study demonstrates that whenever emoticons are used, their related feeling rules the opinion passed on by text-based information investigation.

Keywords
Sentiment Analysis (SA), Machine Learning (ML), Data Analysis, Emojis, LDA

1. INTRODUCTION
Sentiment analysis targets investigating and separating information from the abstract data distributed on the Internet. Because of its tremendous scope of educational and various applications as well as remarkable development of Web, feeling examination has been a very important exploration field in information mining and natural language processing (NLP) as of late [1].

Sentiment analysis is a type of data mining that uses natural language processing (NLP), computational phonetics, and message examination to extract and dissect abstract data from the web, primarily from web-based media and comparative sources. The broken-down information evaluates the overall population's opinions or responses to specific items, individuals, or thoughts and reveals the logical extreme of the data. Feeling analysis uses information mining cycles and strategies to concentrate and capture information for investigations to comprehend the emotional assessment of an archive or assortment of reports, such as blog entries, surveys, news stories, and online media channels such as tweets and announcements[2]. Sentiment analysis, also known as “opinion mining,” can tell you whether there has been an adjustment in general assessments of any aspect of a business. Peaks or valleys in opinion scores give a spot to start making item upgrades, training deals or customer service representatives, or launching new advertising campaigns. Feeling investigation is not a one-time task. By auditing clients' criticism regularly, it becomes more proactive in terms of changing elements in the commercial center [3].

One of the most common uses of sentiment analysis is to assign a class to a message’s text. Sentiment classification can be a binary (positive or negative) or multi-class (positive or negative or neutral) issue, depending on the dataset and explanation[4][13].Data Analysis is the process which contains collection, cleaning, transformation,and modelling of data to catch significant datafor different cycles for making a decision [5][20]. The primary usage is to gather data from the crude data. Data passes on sequences of steps like gathering, collection, cleaning, analysis, interpretation,and visualization. At first,the requirement for data analysisought to be found out.At this point the chosen information ought to be gathered from any social data source. The vital point is to clean data, the data ought to be blunder free, for that every one of the undesirable subtleties like duplicated records, blank areas, and missteps will be eliminated fromthe gathered dataset. In the Analysis step, the basic investigation will be finished on the cleaned and handled data. In the wake of analysis, the data will be deciphered either in the type of basic words or graphs or tables, and so on. The last step is perception, where the outcomes will be picturedas graphs, charts, and so forth.

2. RELATED WORKS
Data can be classified by dividing it into two main parts, namely, train and test. It could use 70% of the data for preparing the model, and 30% for testing. The final results are those obtained during testing [4]. Opinions can be comprehensively classified into three categories, such as positive, negative, and neutral. At this point in them ethodology of sentiment analysis, each abstract sentence identified is delegated as positive, negative, great, excellent, awful, so bad, like, dislike, extremely like or dislike, and do not care [5].

The authors of [6] gathered and sent Twitter tweets for text preprocessing and isolating the words into tokens utilizing tokenization strategy; it then separated those features by developing the scores and polarity of the tweets. The performance and effectiveness of the processed classifier were assessed in terms of accuracy, precision, recall, and F-measure. Therefore, the authors proposed a new hybrid technique for sentiment analysis classification. Compared to other existing calculations, the combination of PSO, GA, and DT demonstrated better execution.

In [7], the natural language toolkit (NLTK) includes everything required to get started with sentiment analysis using Bag of Words Feature Extraction, such as a movie reviews corpus with reviews ordered into positive and negative classifications and several trainable classifiers. It was begun with a simple Naive Bayes Classifier as a standard, utilizing Boolean word feature extraction.

On Twitter, a sentiment analysis ontology known as Tweet Onto Sense, which interfaces communicated feelings or sentiments, the Twitter messages, the analyzed concepts, and their properties were characterized. A few feeling ontologies currently exist. The emotional classification ontology [8] was chosen among them. The authors of [9] introduced an overall spatial investigation of emoji use by analyzing an enormous dataset of geo-encoded tweets containing emoji. It portrayed the popularity of emojis on Twitter worldwide, showing that they are most famous in South-Eastern Asia and South America. In examining the specificity of the nations in regards to the use of various emoticons, the chosen country bunching results separate the “first world”group, the most significant features of which are somewhat without emotions, the “second world”group, which is explicit only for the highest positive emotions, the “third world”group, which detect both positive and negative emotions, and the “fourth world”group, which is primarily negative with extra, rather fundamental ideas such as home, music instruments, fire, snow, dance, and hand motions. Assessing the probabilityof emoticon event in a such tweet given the country by the following:

- Grouping nations are addressed as emoticon probability distributions.
- Correlation over all the world development is Calculated,indicators and distributions of explicit tweets across nations.

The author of [10] described how sarcasm detection from text has now spread to various informational data structures forms and methods. This collaboration has brought about novelties for automatic sarcasm detection. Identifying sarcasm is critical for predicting the correct sentiment of any text. The advantages and disadvantages of sarcasm detection to sentiment analysis have sparked interest in automatic sarcasm detection as research work. Automatic sarcasm detection is a method for predicting whether a given text is sarcastic. Recognizing sarcasm is an obscure task in NLP.

However, the author of [11] explained a retrieval strategy for predicting what emoji combination will follow given a short text as context. The single emoticon expectation mission was extended to an arrangement that was a lot nearer to true utilization in all over the world[12].

The authors of [13] analyzed some global positive and negative events to determine whether there was a discrepancy in Emoji usage between the two types of events. Emoji was found to help improve overall sentiment scores when used in sentiment analysis. While Emoji characters can be used for expressing negative and positive opinions, using them in sentiment analysis improves the expressivity and by and large opinion sentiment scores of positive feelings more than negative feelings in the analysis.

The author of [14] found that a calculation, a strategy, and the emoji vocabulary are the most important contributions for analyzing opinions of social media information (both text-based and emojis), for datasets gathered from Twitter. It also shows the effect of analyzing sentiments while considering emojis alongside the message. Machine language (ML) and deep learning (DL) algorithms were used to direct the review. The proposed framework used elements and models on the gathered audits in view of messages and emojis to decide the opinions. The general outcome shows that considering emoji alongside messages has a significant influence and effects on sentiment analysis. It is moreover observed that deep learning(DL)calculations perform better compared to Machine language (ML) calculations with big data classes issue.

Python software was developed to process text collections into a corpus. The product determines emoji sentiments as adjusted qualities based on the opinions of the texts where the emoticons appear. The vocabularies created using the technique can be used in dictionary-based opinion research. The valences of the emoticons are gotten from the valences from the texts that contain the specific emoticons, which are investigated using the VADER strategy, which determines the valences of short casual texts [15].

The authors of [16] direct SA of a tweet bi-sense issue using three arrangement calculations, namely, support vector machine (SVM), Naïve Bayes, flipped learning module (FLM) using three types of information description translation studies (DTs), such as texts, emoticons, and two emoticons and texts. Generally, SVM accomplished the best results as it performs well for bi-order (business intelligent classification) problems.

However, in [17], the authors attempt to manage the mining of feelings from the social media site Twitter on specific #hashtage. By utilizing tweets, enormous data are produced, which must be handled and afterward delegated positive, negative, or neutral. The primary point of the composition is to utilize various elements, such as emojis and shoptalk other than standard Bag of Words, and to perform powerful sentiment classification utilizing these elements. The authors used a feature selection technique to observe ideal highlights and these ideal elements were grouped. The work shows that coordinating emojis and slang with the traditional Bag of Words model upgrades classification precision.

3. CHALLENGES IN SENTIMENT ANALYSIS

Sentiment analysis is an always developing subfield of natural language processing. Many explorationworks have handled the issuesof sentiment analysis from text, emojis, pictures, and sound or recordings independently. There have been very few studies on emojis for finding sentiments. Furthermore, the related works segment indicatesthat there is a degree for expansion in the field of SA for text and emoji. Hence, the following are the objectives and targets of this review:

- Sentiment analysis or opinion investigationutilizing bi-mode (text and emojis) via web-based entertainment information or data.
- Developing and creating emojisvocabulary as a lexicon.
- Improving the classification precision of sentiment analysis by utilizing Machine Learning algorithm.

4. METHODOLOGY

This section describes the proposed strategy for the study issue, such as the dataset, the preprocessing stage, and the processing stage utilizing ML algorithms.

4.1The Dataset

The dataset contains 10000 tweets about COVID-19 [18]. The dataset was stored in CSV format that is the input for the program.This data has three classes of sentiment (‘1’ is Positive, ‘2’ is Negative, ‘3’ is Neutral). To import data from CSV file to data frame, The Pandas library was used.

4.2 Pre-Processing

The most vital phase in any classification study is the Pre-Processing stage, which is basic for getting a successful result. An overview of the used preprocessing strategies is displayedin the following subsections:

4.2.1 Data Cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct.The texts extracted from tweets have to becleaned.The following steps are followed:

- Tokenization: convert text to list of tokens.
- Cleaning data by removing special characters like hashtags, usernames, and symbols.
- Remove all stop words.
- Remove numbers and punctuations.
- Tokens converted to lower case.
- Token normalization by Stemming.
- Token normalization by Lemmatization.

4.2.2 Sentiment Detection

At this level, each paragraph of the dataset is analyzed for subjective and objective words. Subjective expression words are wanted but objective expression words are disposed of. Sentiment analysis is made at various levels utilizing common computational procedures like Bigrams, Trigrams, sentiment polarity and word clouds and so on.

4.2.3 Vectorization

To constructa single classifier of texts and emojis, emojis were changed over into their separate Unicode, and utilized bag of words for tokenization.

4.3 Machine Learning

In this study, some classification algorithms are used, which are Supported Victor Machine(SVM), Random Forest, Gradient Boosting, K-Neighbors, Gaussian Naive Bayes, andLogistic Regression.

4.4Experiments

This section describes the experiments, which consists of classifying tweets as positives, negative, or neutral from combination of text and emoji over different splits with different techniques.

5. CLASSIFICATION TECHNIQUES

The classification is the last step in deciding the opinion of the text. Classification to assign input data with a class label.The followed process of classification in this research is defined by analgorithm. See Algorithm 1.

Algorithm1: SA for text and emoticons.

Input: Custom made dataset with emoticon, ML_classifiers list
Output: Sentiment (Positive/Negative/Neutral)
1. Begin
2. Remove all stop words using NLTK stopwords corpus
3. Tokenize all the samples and for the new dataset for training
5. For test, train split ratios do
6. For classifier CL in ML_classifiers list do
7.Train the classifier CL based on training data Evaluate the trained classifier CL using the test data
8.Record the Accuracy and score of the classifier
9.end for
10 end for
11.Compare the results of ML Classifiers through Criteria Specified
12. Save plots and graphs
13. end

6. EVALUATION

The data collected for the performance evaluation of sentiment analysis consisted of 10000 tweets with three types of sentiments. The tweets were collected for the COVID-19, with “1” for positive, “2”for negative and “3” for neutral. Emojis were added to the data set in preparation stage.Just subjective tweets with their emojis were utilized for evaluation, while the objective tweets were discarded. The dataset was divided into different sizes, and classifiers were applied to each one, and the results are shown in Tables (1), (2), and (3). The performance evaluation with the optimal features of accuracy over different splits using MI is shown in Figure 1, while the accuracy of different classifiers over different splits using MI is shown in figure 2 which also shows the accuracy over the depth in test and train dataset of using Random Forest classifier is also shown in Figure 2, the accuracy over epochs in test and train dataset of using Supported Vector Machin classifier is shown in Figure 3 shows, the accuracy over the depth in test and train dataset of using Gradient Boosting classifier is shown in Figure 4, the accuracy over N-neighbors in test and train dataset of using K-Neighbors classifier is shown in Figure 5.

Table 1.Results for 0.4 Test Size Split

Metric	Random Forest	Supported Vector Machin	Gradient Boosting	K-Neighbors
Accuracy	0.91	0.31	0.85	0.37
F-Score	0.91	0.37	0.85	0.54

Table 2.Results for 0.3 Test Size Split

Metric	Random Forest	Supported Vector Machin	Gradient Boosting	K-Neighbors
Accuracy	0.92	0.31	0.85	0.37
F-Score	0.92	0.37	0.85	0.54

Table 3.Results for 0.2 Test Size Split

Metric	Random Forest	Supported Vector Machin	Gradient Boosting	K-Neighbors
Accuracy	0.93	0.31	0.86	0.37
F-Score	0.93	0.37	0.86	0.54

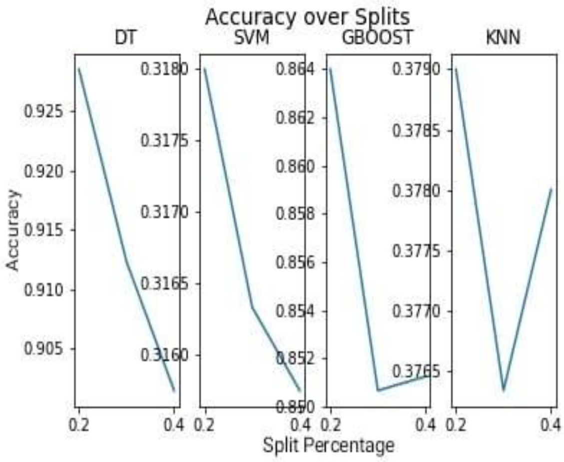

Figure 1:Accuracy over Depth

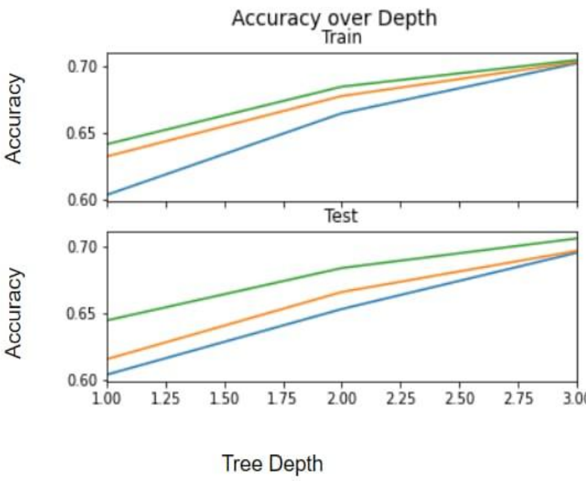

Figure 2: Accuracy over Epochs

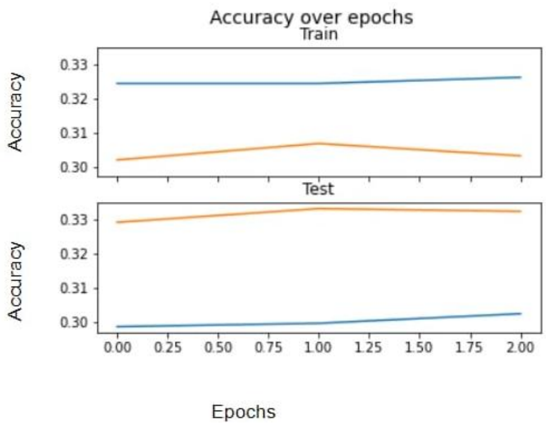

Figure 3: Accuracy over Depth

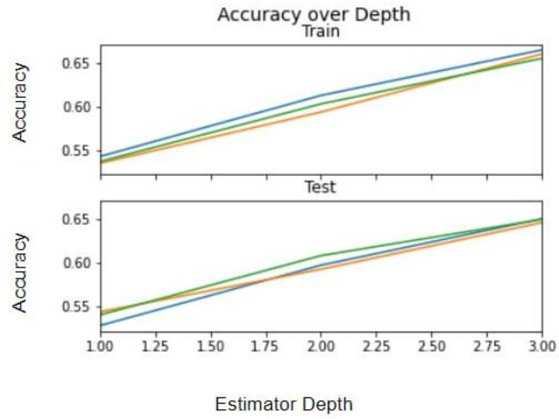

Figure 4 Accuracy over N Neighbors using Gradient Boosting Classifier

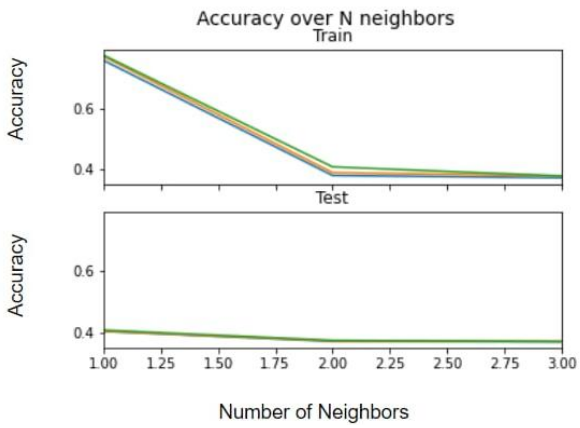

Figure 5: Accuracy over N Neighbors using K-Neighbors classifier

The data is divided to 200 topics and several classifiers are applied on it and theirs score results are recorded as shown in figure 6.

```
compare_models()
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lightgbm	Light Gradient Boosting Machine	0.6888	0.8542	0.6672	0.6876	0.6848	0.5213	0.5238	4.157
et	Extra Trees Classifier	0.6702	0.8214	0.6465	0.6664	0.6635	0.4913	0.4944	3.303
rf	Random Forest Classifier	0.6662	0.8261	0.6429	0.6660	0.6609	0.4850	0.4886	4.708
gbc	Gradient Boosting Classifier	0.6614	0.8252	0.6311	0.6696	0.6517	0.4739	0.4833	21.723
lr	Logistic Regression	0.6584	0.8295	0.6357	0.6539	0.6527	0.4741	0.4763	1.207
ridge	Ridge Classifier	0.6568	0.0000	0.6296	0.6526	0.6476	0.4691	0.4735	0.056
lda	Linear Discriminant Analysis	0.6549	0.8247	0.6329	0.6526	0.6505	0.4691	0.4714	0.416
ada	Ada Boost Classifier	0.6216	0.7826	0.5933	0.6191	0.6126	0.4140	0.4190	1.547
svm	SVM - Linear Kernel	0.6176	0.0000	0.5969	0.6116	0.6131	0.4137	0.4147	1.223
nb	Naive Bayes	0.6156	0.7849	0.6075	0.6160	0.6116	0.4166	0.4196	0.055
dt	Decision Tree Classifier	0.6061	0.7042	0.5909	0.6045	0.6050	0.3986	0.3989	0.955
qda	Quadratic Discriminant Analysis	0.5845	0.7859	0.5829	0.6009	0.5677	0.3722	0.3876	0.266
knn	K Neighbors Classifier	0.5205	0.6852	0.5033	0.5368	0.5157	0.2625	0.2700	3.226
dummy	Dummy Classifier	0.3771	0.5000	0.3333	0.1422	0.2065	0.0000	0.0000	0.034

Figure 6: Different Classifiers with theirs score

After that sentiment polarity distribution is observed as shown in figure 7.

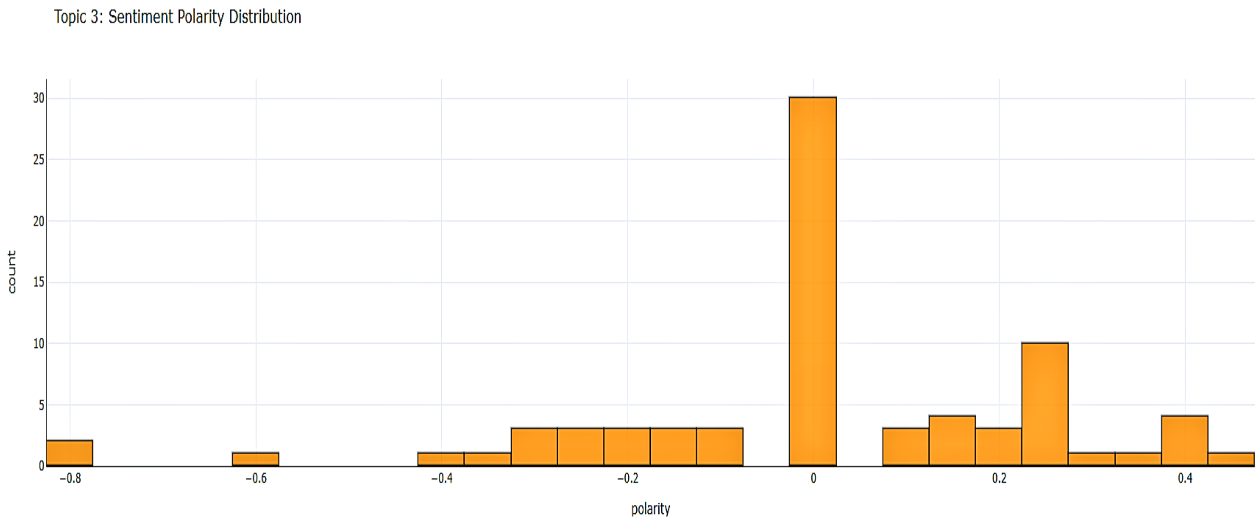

Figure 7: Sentiment Polarity Distribution

Latent Dirichlet allocation (LDA) is a generative measurable model. It permits sets of perceptions to be made sense of by unnoticed gatherings that make sense of why a few pieces of the data are comparable. If perceptions are words gathered into records, it sets that each report is a combination of few subjects and that each word's presence is owing to one of the report's topics [19]. So, the stop words are removed and using LDA, because of it is an example of a topic model, and dividing the text to 200 topics. Then three samples of the 200 topics are observed with the repeated words in each topic with different majority as shown in figures 8, 9 and 10.

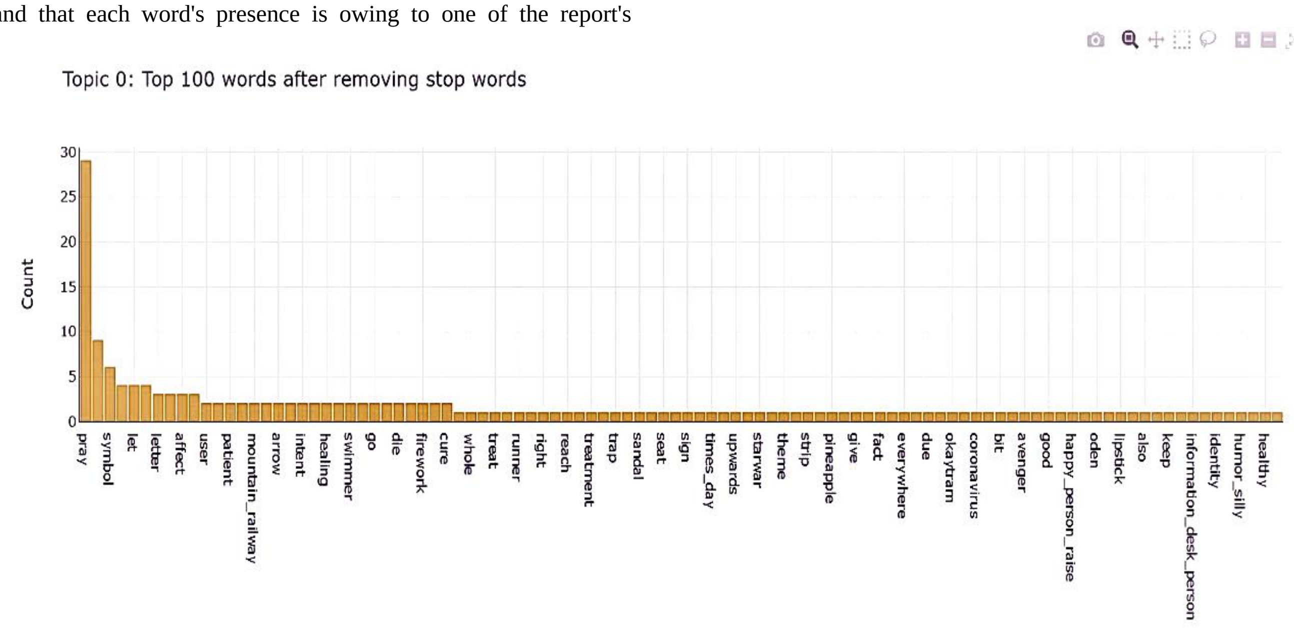

Figure 8 Topic 0: Top 100 words after removing stop words

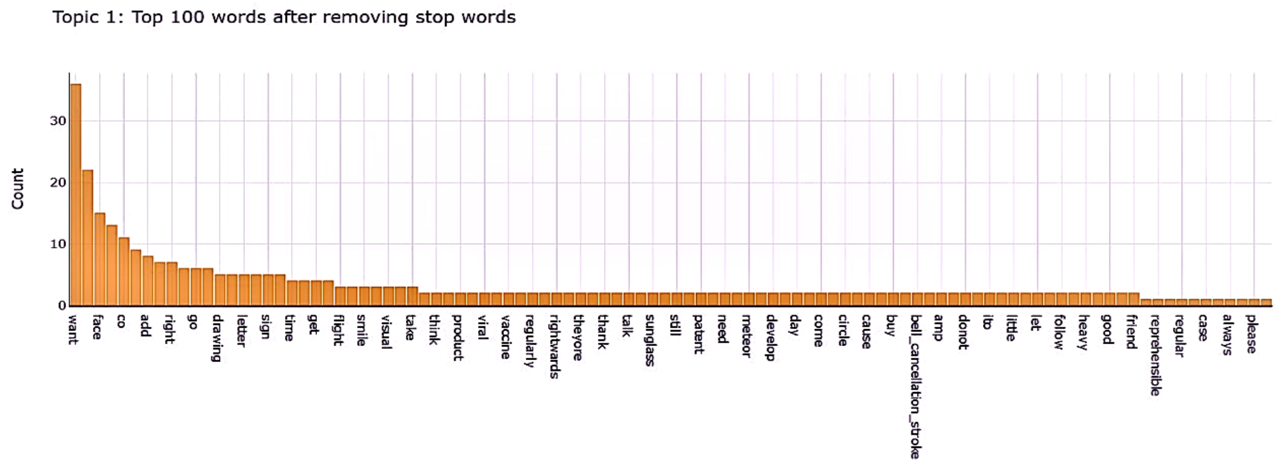

Figure 9 Topic 1: Top 100 words after removing stop words

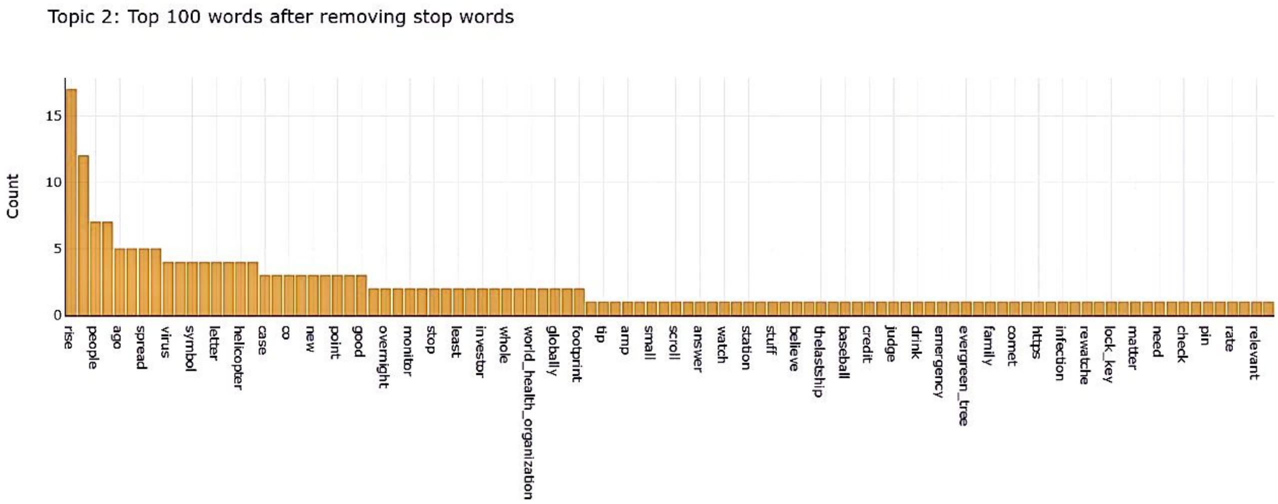

Figure 10 Topic 2: Top 100 words after removing stop words

The frequency distribution of every bigram is gotten in a string is commonly used for simple statistical analysis of text or the two composite words which repeated simultaneously. Then, the three samples of the 200 topics are observed as shown in figures 11, 12 and 13.

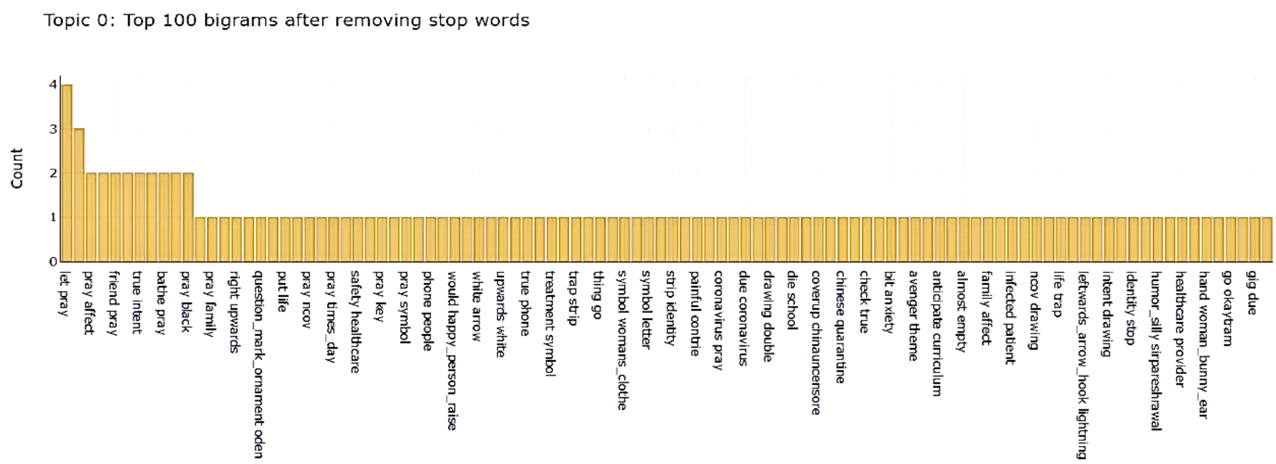

Figure 11 Topic 0: Top 100 bigrams after removing stop words

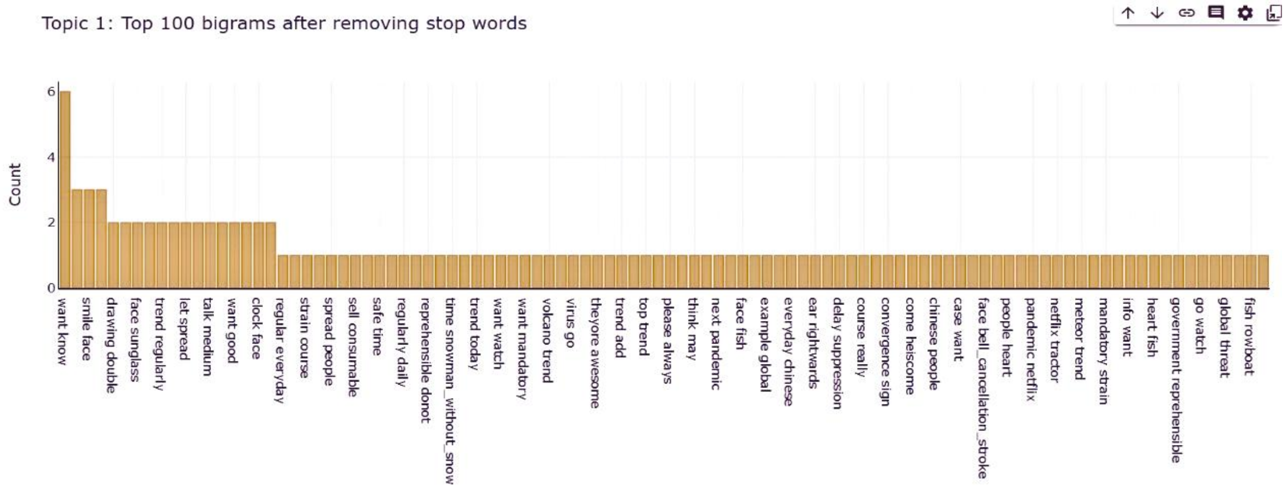

Figure 12 Topic 1: Top 100 bigrams after removing stop words

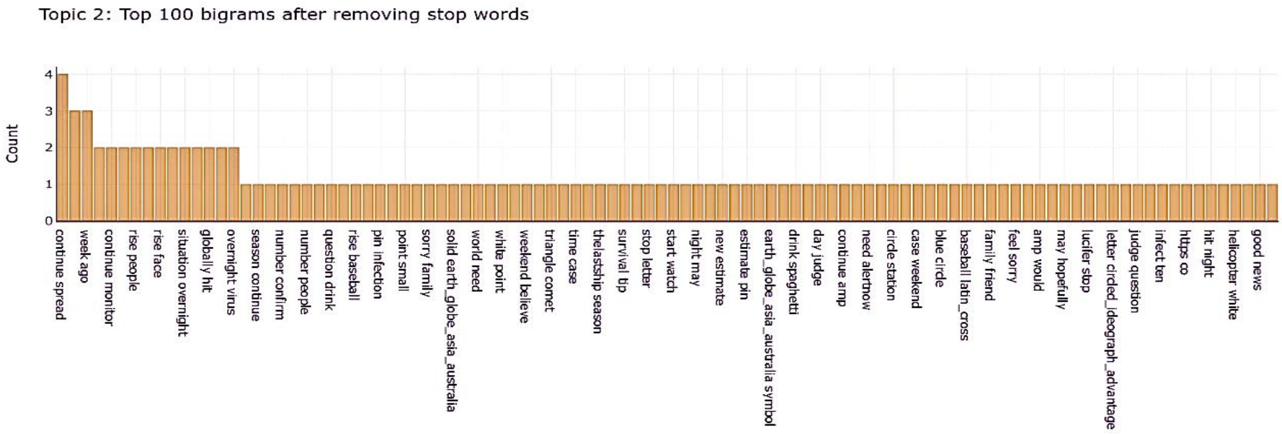

Figure 13 Topic 2: Top 100 bigrams after removing stop words

By obtaining the word cloud which is a visual representation of the words used in the particular text. the size of each word indicates to its relative frequency in the dataset.Figure 14 shows words aredisplayed in various sizes. The greater and bolder the word shows up, the more frequently it's referenced inside a given text and the more significant it is.

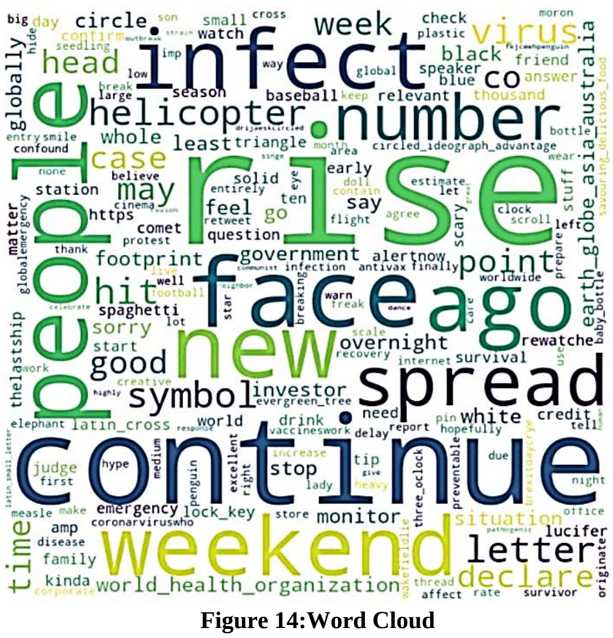

Figure 14:Word Cloud

7. COMPARISONWITHOTHERWORK

By applying and developing several classifiers on the dataset which are Supported Victor Machine (SVM), Random Forest, Gradient Boosting, K-Neighbors, Gaussian Naive Bayes, and Logistic Regression. The results is extracted in a tabular form as shown in Table (4)and is compared with the results of another workas shown in Table (5).

Table 4. Comparison of results in different models and classifiers

Model	Precision	Recall	F-Score	Accuracy
SVM	0.47	0.318	0.37	0.318
Random Forest	0.93	0.93	0.93	0.93
Gaussian Naive Bayes	0.62	0.629	0.61	0.624
Logistic Regression	0.95	0.95	0.95	0.95

Table 5. Comparison of results in different models and classifiers of another work [14]

Model	Precision	Recall	F-Score	Accuracy
SVM	0.73	0.71	0.74	0.78
Random Forest	0.73	0.66	0.68	0.76
Naive Bayes	0.59	0.57	0.60	0.52
Logistic Regression	0.74	0.66	0.68	0.78

The Tables Shows that our data formed with different methods combining emoticons and text gets better or close performance to the prior work results regarding in Random Forest, Logistic Regressionand Naive Bayes. But shows less performance on Supported Victor Machine (SVM).

8. CONCLUSION

Sentiment Analysis or Opinion Miningcomprehends thesentiments, feelings, replications as well as decisionselicited from texts or different information used indata analysisor mining, web mining, social media analyticssince opinions are to pass judgment on humanbehavioral comportment.They can be sorted into three classes which are positive, negative, or neutral. This study mainly contributes algorithms and methods for classifying and analyzing sentiments of social media data in textual and emojis formats, such as the COVID-19 dataset from Twitter with three classes of sentiment (‘1’ is Positive, ‘2’ is Negative, ‘3’ is Neutral). Pre-Processing is the first step in classification which includes data cleaning. The study was conducted using Machine Learning algorithms. The proposed system used several features and classifiers such as Supported Victor Machine (SVM), Random Forest, Gradient Boosting, K-Neighbors, Gaussian Naive Bayes, and Logistic Regression on the gathered audits in light of text and emojis to decide the sentiments and opinions.The dataset was divided into different sizes, and classifiers were applied to each one. Random Forest Classifier records higher results with Accuracy and F-score over different dataset splits. On other hand after removingstop words from the datasetand using LDA to devide text to 200 topics to get most common words and declare the repeated words over the dataset. The Bi-gram shows the most two words repeated together or the most two composite words over 3 topics which is sample from 200 topics. Word Cloud shows the distribution polarity frequency of word over the dataset. In the future, the propose is investigating the effectivenessof the proposed sentiment analysisof text incorporating emojis with different approach and different tasks like rating prediction and helpfulness prediction, aswell as different levels including sentence and aspect-level sentimentanalysis.

9. Author Contributions

The paper conceptualization, methodology, software, Vectorization, Experiments, resources, data preparation, data cleaning, classifications, visualization, have been done by 1st author. The supervision, and project administration, have been done by 2nd and 3rdauthors.

10. Acknowledgments

This research is supported by Faculty of Graduate Studies for Statistical Research, Department of Computer Science, Cairo University.

11. REFERENCES

[1] Mohammed Basiri, Shahla Nemati, Moloud Adbar, Erik Cambria, “ABCDM: An Attention- based Bidirectional CNN-RNN Deep Model for Sentiment analysis”, Future Generation Computer Systems 115, 279–294, 2021.

[2] Pratik Thakora, Dr. Sreela Sasib,“Ontology-based Sentiment Analysis Process for Social Media Content”, Procedia Computer Science, Volume 53, Pages 199–207, INNS Conference on Big Data, 2015

[3] Anatoliy Surikov, Evgeniia Egorova, “Alternative method sentiment analysis using emojis and emoticons”, Procedia Computer Science 178, 182-193, 9th International Young Scientist Conference on ComputationalScience (YSC), 2020.

[4] “Social, Culturaland Behavioral Modeling”, 12th International Conference, 2019.

[5] Rizwan Khan, “Social Media Analysis with AI: Sentiment Analysis Technique for The Analysis of Twitter COVID-19 Data”, Journal of Critical Reviews, VOL7, ISSUE 09, ISSN- 2394-5125,2020.

[6] Senthil Nagarajan, UshaGandhi, “Classifying streaming of Twitter data based on sentiment analysis using hybridization”, Neural Computing and Applications 31:1425–1433, 2019.

[7] https://streamhacker.com/2010/05/10/text-classification-sentiment-analysis-naive-bayes-classifier/

[8] Liviu-Adrian Cotfas, Camelia Delcea, Ioan Roxin, Ramona Paun,“Twitter Ontology-Driven Sentiment Analysis”, ResearchGate, 2016.

[9] Nikola Ljubesi, Darja Fiser,“A Global Analysis of Emoji Usage, Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpiriST Shared Task, pages 82–89Association for Computational Linguistics, 2016.

[10] Karthika K., “An efficient approach for sarcasm detection in tweets using polarity flip”, International Journal Of Advance Research, Ideas and Innovations in technology, ISSN: 2454-132X, Impact factor: 4.295, Volume 5 Issue 2, 2019.

[11] Weitsung Lin, Tinghsuan Chao, Jianmin Wu, Tianhuang Su, “Predict Emoji Combination with Retrieval Strategy”, The Association for Natural Language Processing, 2019.

[12] Jayashree Subramanian, Varun Sridharan, Kai Shu,Huan Liu,“Exploiting Emojis for Sarcasm Detection”, 2019.

[13] Serkan Ayvaz, “The Effects of Emoji in Sentiment Analysis”,International Journal of Computer Electrical Engineering,ResearchGate,Volume 9, Number 1,2017.

[14] Mohammad Aman Ullah, Syeda Marium, Shamim Begum, Nibadita Dipa,“An algorithm and method for sentiment analysis using the text and emoticon”, The Korian Institute of Communications and Information Sciences,ICTExpress 6, 357-360, 2020.

[15] Fabian Haak, “Design and Development of an Emoji Sentiment Lexicon”, Institute for Information Management, Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science, 2021.

[16] Mohammed Alzubaidi, Farid Bourennani, "Sentiment Analysis of Tweets Using Emojis and Texts", International Conference on Information Science and Systems (ICISS), 2021.

[17] S.Sayeedunnisa, Dr. Nagratna Hegde, Dr. KhaleelKhan, "Using Slang and Emoticon for Sentiment Analysis of Social Media Data", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, NCAIT Conference Proceedings, 2020.

[18] https://www.kaggle.com/imranzaman5202/starter-code-covid-19-sentiment-analysis/data?select=tweets_tagged.csv

[19] https://en.wikipedia.org/

[20] Nirmal Varghese Babu, E. Grace Mary Kanaga,"Sentiment Analysis in Social Media Data for Depression Detection using Artificial Intelligence: A Review", SN Computer Science 3:74, 2022.