Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Sentiment analysis of coronavirus data with ensemble and machine learning methods

Muhammet Sinan Başarslan *1, Fatih Kayaalp 2

[1] Istanbul Medeniyet University, Department of Computer Engineering, İstanbul, Türkiye, muhammet.basarslan@medeniyet.edu.tr
[2] Düzce University, Department of Computer Engineering, Düzce, Türkiye, fatihkayaalp@duzce.edu.tr

Cite this study: Başarslan, M. S., & Kayaalp, F. (2024). Sentiment analysis of coronavirus data with ensemble and machine learning methods. Turkish Journal of Engineering, 8 (2), 175-185

https://doi.org/10.31127/tuje.1352481

Keywords
Ensemble learning
Machine learning
Sentiment analysis
Text representation
Word embedding

Research Article

Received: 30.08.2023
Revised: 05.01.2024
Accepted: 22.01.2024
Published: 07.04.2024

Abstract

The coronavirus pandemic has distanced people from social life and increased the use of social media. People's emotions can be determined with text data collected from social media applications. This is used in many fields, especially in commerce. This study aims to predict people's sentiments about the pandemic by applying sentiment analysis to Twitter tweets about the pandemic using single machine learning classifiers (Decision Tree-DT, K-Nearest Neighbor-KNN, Logistic Regression-LR, Naïve Bayes-NB, Random Forest-RF) and ensemble learning methods (Majority Voting (MV), Probabilistic Voting (PV), and Stacking (STCK)). After vectorizing the tweets using two predictive methods, Word2Vec (W2V) and Doc2Vec, and two traditional word representation methods, Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BOW), classification models built using single machine learning classifiers were compared to models built using ensemble learning methods (MV, PV and STCK) by heterogeneously combining single machine classifier algorithms. Accuracy (ACC), F-measure (F), precision (P), and recall (R) were used as performance measures, with training/test separation rates of 70%-30% and 80%-20%, respectively. Among these models, the ACC of ensemble learning models ranged from 89% to 73%, while the ACC of single classifier models ranged from 60% to 80%. Among the ensemble learning methods, STCK with Doc2Vec text representation/embedding method gave the best ACC result of 89%. According to the experimental results, ensemble models built with heterogeneous machine learning classifier algorithms gave better results than single machine learning classifier algorithms.

1. Introduction

The development of information technologies has had both negative and positive effects on people. The channels through which they express this situation are social media. With the development of smart devices that allow access to social media at any time, applications and websites on these devices have become the first resort in many areas, especially in the field of health.

During the pandemic, people could not go out of their environment. Working life turned into remote work. Most levels of education, especially universities, turned to distance education. Those infected with the disease were subjected to mandatory isolation. For these reasons that can be increased, people's first choice for socializing was social media such as Twitter [1].

Sentiment analysis was performed on the data collected and tagged on Twitter, a medium where people share their instant ideas and emotions. The fact that social media is an indispensable tool for people and that people constantly express opinions about concepts such as social, economic, health, product, brand, etc. has led to the emergence of the field of sentiment analysis. Natural language processing techniques are used in these studies. Emotions such as positive and negative are predicted from the texts people share. Sentiment analysis studies are carried out on movie and restaurant experiences as well as texts that give opinions about companies. These studies provide positive feedback to companies. For this purpose, in this study, DT, KNN, LR, NB, RF, Support Vector SVM (SVM) algorithms are used to provide diversity in a heterogeneous ensemble system. MV and PV ensemble learners and STCK are used as methods to fuse the ensemble decision.

In the first step of ensemble learning, which consists of two parts: ensemble formation and integration, different base classifiers are used. This step is called ensemble formation, and different sets of patterns are created. In the ensemble integration phase, the final decision of the system is determined by using various integration methods to combine the decisions of the base classifiers. In this process, two other factors are as important as classifier selection and integration methods: the single performance of the base learners and the independence of the ensemble learners' results. High diversity of base learners is usually achieved with traditional ensemble algorithms such as bagging, random subspaces, random forests and rotation forests. In heterogeneous systems, diversity is achieved using different learning algorithms and the results are combined with various decision-making methods such as voting, STCK and boosting. By combining these factors, the ensemble system can achieve a higher classification performance.

Models based on machine learning (SVM, DT, RF, LR, NB, KNN) and ensemble learning (MV, PV, STCK) tweet data were split into two separate training and test sets (70%-30% and 80%-20%). F, ACC, and performance metrics were used to evaluate these models. The contributions of this paper can be listed as:

- The performance of six single machine learning classifiers was compared to sentiment analysis in the same study,
- The performance of various ensemble learning methods was compared with sentiment analysis in the same study,
- The performance of single machine learning classifiers was compared to ensemble learning methods based on the same single learning classifiers in the same study,
- The effect of different text representation/embedding methods on sentiment analysis performance was compared in the same study.

This study used BOW, TF-IDF both of which are frequency-based text representation methods. In addition, the classification process performed using W2V, one of the word embedding methods, and the Doc2Vec method, which allows direct vectorization of documents. Two methods of W2V word embedding (CBOW and Skip-gram) and two methods of Doc2vec embedding (PV-DM, PV-DBOW) were used. The flowchart is shown in Figure 1.

In the second section, information about the literature is given. In the third section, the methodology section, the data set and its types, word embedding and text representation methods are available. In the fourth section, which is the experimental settings section, information is given about the environment in which the experiments are carried out, the language, the performance evaluation metric of the models created, and the separation method of the data set used in the experiments. In the fifth section, experimental results are given. In the last section, the conclusion and discussion section, comparison with the literature and general evaluation are included.

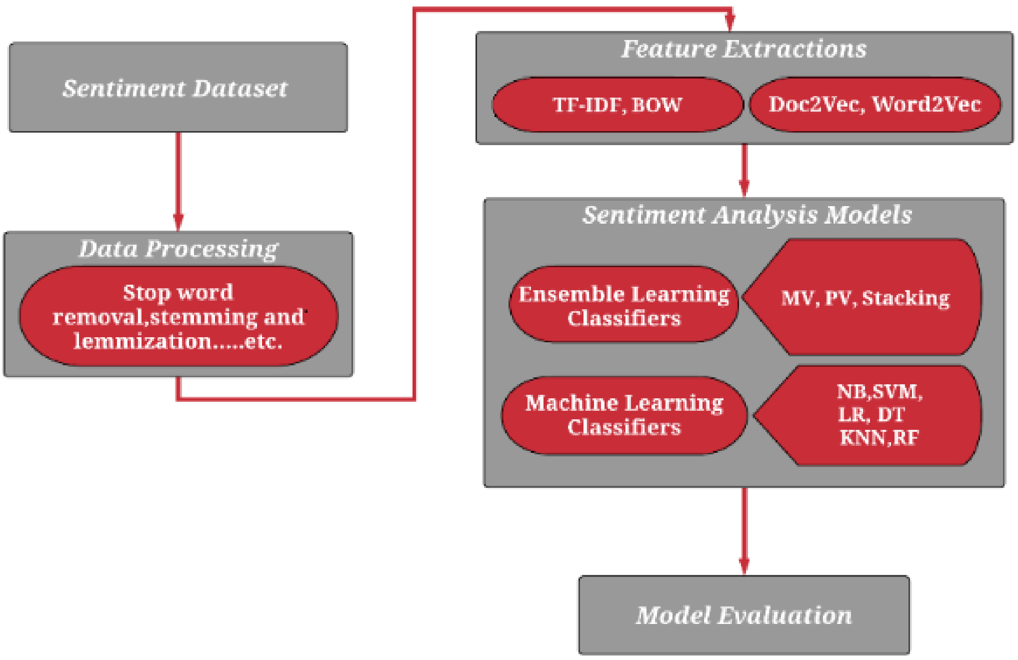

Figure 1. Study flowchart diagram.

After the text representation methods on Coronavirus made with Twitter data during the pandemic period, models and details on sentiment analysis with various Machine Learning methods will be explained in this section.

After TF-IDF and BOW, 65% ACC result was obtained in the model created with Long short-term memory (LSTM) [2]. In another study with TF-IDF, 65% ACC value was obtained in the model created with stochastic gradient descent (SGD) [3]. After word representation with TF-IDF method, 78.5% ACC was obtained as a result of emotion classification with LR machine learning [4]. The first study [5] obtained 74.29% ACC and the second [6] obtained 84% ACC in the model created with TF-IDF and ngrams before NB machine learning. In their experimental study on sentiment analysis using SVM and NB methods together with MV ensemble learning on Twitter data, they achieved the best result with 87.7%

accuracy [7]. In another study, 84.38 ACC results were obtained in the model created with SVM after TF-IDF and W2V [8]. Among the models they created using different ensemble learning methods for sentiment analysis on tweet data during the pandemic period, they obtained 83.3% with voting and 83.2% with STCK [9]. TF IDF on 40,000 coronavirus tweets with Bi-gram Among the models created with NB, RF, and SVM, the best results were obtained with SVM with 87.80% ACC [10].

Classification Performance achieved using TF-IDF feature extraction and SGD is 85.141% ACC [11]. In another study, they obtained an ACC value of 70.6% with TDF-IDF and artificial neural networks (ANN) [12]. Models were created with KNN, SVM, DT, and NB after TF-IDF and BOW with Covid data, and they got better results than other models with 63% ACC with SVM [13]. In sentiment analysis models made with DT, LR, SVM, NB models and BOW, LR gave better results with 81.8% ACC [14]. In the models built with Arabic tweet data and RF, LR, NB, Voting, and BOW word representations, Voting performed better with a 74% ACC [15].

In this section, we describe sentiment analysis studies based on different machine learning classifiers using different text representation and word embedding methods on Twitter tweets about coronavirus. Popular studies have used ensemble learning to improve the performance of machine learning algorithms. In this study, we compare the performance of models generated by single machine learning methods and different ensemble methods according to TF-IDF and W2V in different vector dimensions.

2. Method

This section describes the ensemble and machine learning algorithms, the text representations, and the data set. Before classification, 44955 tweets were preprocessed. Symbols, punctuation, numbers and stop words were removed. All characters were converted to lower case. Lemmization was performed. Removed usernames and hashtags. NLTK was used in the preprocessing steps.

2.1. Dataset

The study used open-source data. [16]. The data set contains 44955 data and 6 attributes as shown in Table 2. In the study, only the label attribute with tweet and sentiment class was used. Table 1 gives information about the Twitter data.

As shown in Table 2, the emotion class distribution of the tweet data consists of 3 classes: positive, negative, and neutral.

The percentage distribution of the coronavirus tweet data is shown in Figure 2.

Figure 3 shows word clouds according to the sentiment classes of the posted tweets. As can be seen in all three sentiment classes, words for basic needs such as "supermarket", "pandemic", "grocery", "store", "shopping" are used. In addition to protecting health, this situation has revealed feelings about the places where he and his family meet their needs such as hygiene, especially food. This situation is another research topic that needs to be studied.

Table 1. Data set information.

Attribute	Definition of attribute
Username	Twitter Users integer type name
Screen Name	The name of the query that everyone on Twitter sees in integer type
Location	Tweet location
Tweet At	When the tweet was posted
Tweet	Content of the tweet

Table 2. Sentiment class distribution.

Sentiment class	Text
Positive	19592
Negative	17031
Neutral	8332

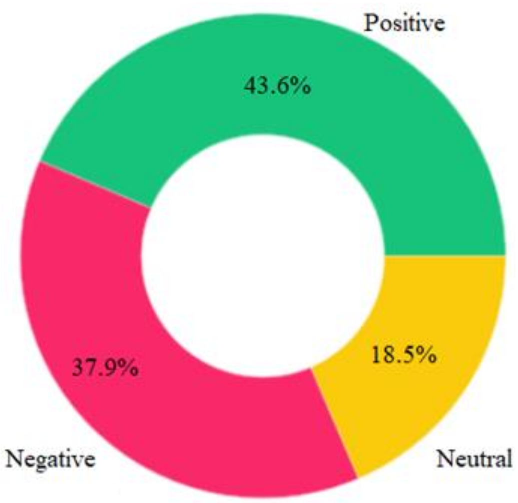

Figure 2. Class distribution.

Figure 3. Class word cloud.

2.2. Text representation

In text classification, text representation and embedding improve classification performance. A distinction can be made between frequency-based and estimation-based methods. In frequency-based methods, vectorization is performed according to the word frequency during the classification process, while in embedding, vectorization is performed based on the neural network. In this study, TF-IDF and W2V were used [17].

2.2.1. Term frequency inverse document frequency

TF-IDF is calculated by taking into account the words present in the document, including all the words associated with that particular document. This method facilitates the identification of the target word in the context of the document. Specifically, TF (Term Frequency) is determined by counting the occurrences of the target word within the document, providing insight into its frequency in that specific context. Conversely, IDF (Inverse Document Frequency) is derived by locating the relevant records within the entire set of word records, highlighting the importance of the term in relation to the broader corpus [17].

2.2.2. Word2Vec

W2V is a prediction-based text representation method. It is an unattended neural network model consisting of an input, an output, and a hidden layer. Mikolov and his team have proposed two new model architectures. W2V is performed by two different methods as Skip-Gram and CBOW. These two approaches are based on different application of input and output variables, but they basically use the same neural network [18]. The CBOW model tries to predict the central word from the words around the central word. The CBOW model makes better predictions on small data sets. On the other hand, it is difficult to understand words that have more than one meaning. The representation of the model is shown in Figure 4 [18-19].

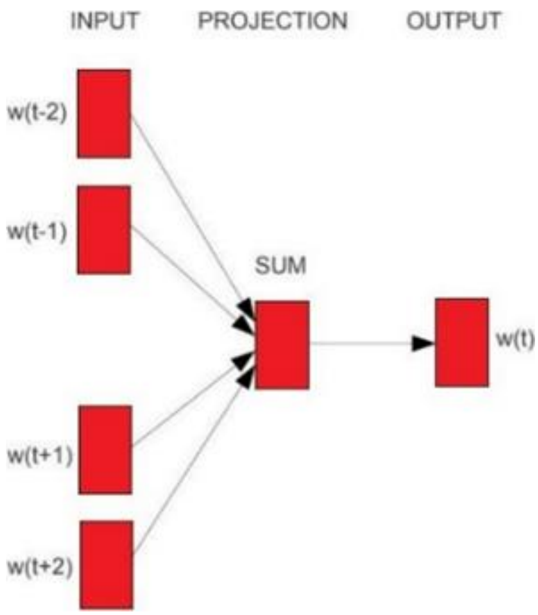

Figure 4. CBOW model [13].

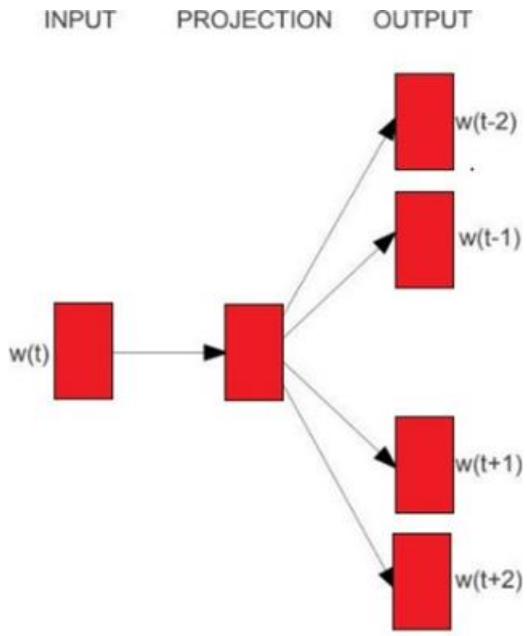

Figure 5. Skip-gram model [13].

The skip-gram model tries to predict the words found around a word. The skip-gram model works better with larger amounts of data. At the same time, the skip-gram model is better at understanding words that have 2 or more meanings than the CBOW. The representation of the model is shown in Figure 5 [19].

In the study, vector sizes of 100 and 200 and window size of 5 were taken.

2.2.3. Doc2Vec

Based on the W2V architecture, each document is created by adding a separate document vector. These document vectors are represented by numerical values in vector space. The Doc2Vec model has two methods, the distributed memory version of the paragraph vector (PV-DM) and the distributed bag-of-words version of the paragraph vector (PV-DBOW), which are shown in Figure 6 and Figure 7, respectively [20]. The PV-DM in the Doc2Vec model corresponds to the CBOW in the W2V model, while the PV-DBOW method is implemented as a skip-gram method [21]. In the study, the hyperparameter settings for Doc2Vec were taken as vector size 100 and 200, window size 5, min-count 5.

2.3. Machine learning

Machine learning is a mathematical approach to modeling decision making in the human brain and neural networks. It compares each neuron in the human brain to a simple digital processor and the brain to a computing machine. In 1950, Alan Turing introduced the idea of machines thinking like humans. The emergence of machine learning as we know it today dates back to the 1980s [21]. In this study, SVM, NB, KNN, LR from machine learning methods are used. These classifiers are explained in this section.

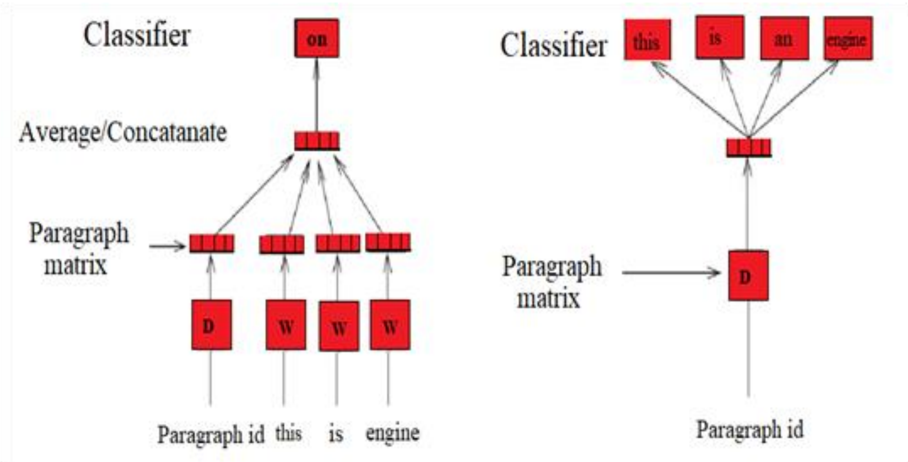

Figure 6. PV-DM model and PV-DBOW model [20].

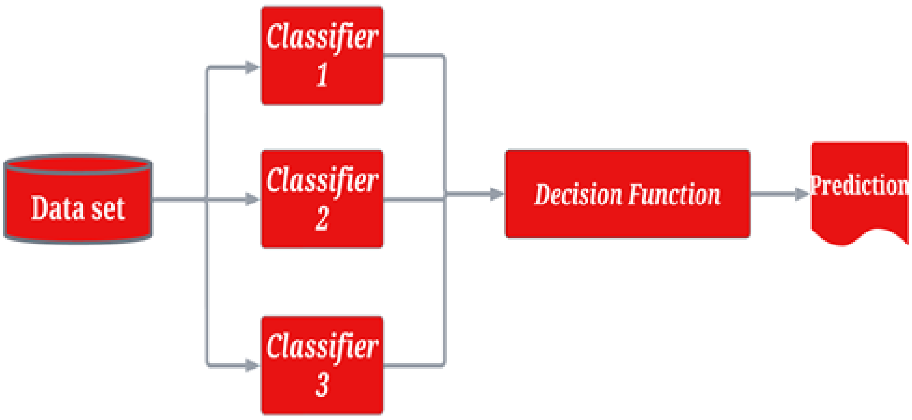

Figure 7. Ensemble learning.

2.3.1. Naïve bayes

Named in honor of the English mathematician Edmund Bayes, the algorithm belongs to the class of statistical classification methods. The Bayesian classifier is based on Bayes' theorem [21].

The above formula is used to decide whether a given x $(x = [x(1), x(2), \ldots, x(L)]^T \in r^L)$ belongs to class S_i When the independence proposition is used statistically in the Bayes decision theorem, this type of classification is called NB classification. In a mathematical expression (Equation 1) [22].

$$P(x|S_i)P(S_i) > P(x|S_j)P(S_j), \forall j \neq i \tag{1}$$

The term $P(x|S_i)$ in Equation (1) is rewritten as in Equation (2).

$$P(x|S_i) \approx \prod_{k=1}^{L} P(x_k|S_i) \tag{2}$$

In this way, Bayes' theorem takes the form of Equation (3).

$$P(S_i)\prod_{k=1}^{L}P(x_k|S_i)\,P(S_j)\prod_{k=1}^{L}P(x_k|S_j) \tag{3}$$

$P(S_i)$ and $P(S_j)$ is the prior probability of classes I and j. Their values can be easily calculated from the studied data set [22]. The NB classifier is used in many fields in artificial intelligence studies, including disease diagnosis [23-24].

2.3.2. Support vector machine

In classification studies based on the principle of least intrinsic risk, it creates a hyperplane for the separation between classes. In the hyperplane determines in which class the new sample will be placed [25]. SVM excel at both linear and nonlinear classification tasks [22].

2.3.3. K-nearest neighbor

KNN, one of the supervised learning methods, is used for classification and regression. It determines the k nearest neighbors by looking at the distances between the data set and the problem-specific data and includes that data in that class. Methods such as Euclidean Manhattan and Minkowski distance are used to calculate the distance [26].

2.3.4. Logistic regression

It was first introduced in 1958 by statistician David Cox. Logistic regression uses the "maximum likelihood" method. Logistic regression uses the sigmoid function to classify, the sigmoid function is an "S" shaped curve [22].

2.3.5. Decision tree

Tree-based learning algorithms are widely employed in supervised machine learning. At the core of these algorithms is the decision tree, a hierarchical structure designed to partition a data set with numerous records into smaller subsets through a series of decision rules. Essentially, a decision tree is a tool that, through sequential decision steps, segments substantial data sets into more manageable groups of records [27].

2.4. Ensemble learning

It is a type of learning designed to use different classifiers together to solve classification and regression problems, or to improve the performance of a classifier. Ensemble learning optimizes efficiency by strategically combining multiple expert or machine learning models to improve the performance of a single inferior model [28]. In addition to the distinction between homogeneous and heterogeneous, ensemble learning differs according to the decision function, such as voting, averaging, etc., and is classified as such. However, it is usually classified as bagging, boosting, STCK. Figure 7 shows an illustration of this classification of the ensemble process.

2.4.1. Voting

Voting is an ensemble learning method for classification problems. It makes predictions based on the creation of two or more sub-models and voting the result with the mean or mode of the predictions of these sub-models [29].

2.4.1.1. Majority voting

The same or different Base classifiers vote for a class and the majority with the MV wins. Statistically, the target label of the classes together can be said to be distributed by voting to guess. Equation (4), we estimate the $\hat{y}$ Class label through the majority (plurality) vote of each classification Cl.

$$\hat{y} = mode\{Cl_1(x), Cl_2(x), \ldots\ldots Cl_m(x)\} \tag{4}$$

If we combine the three-class results of an example of education as in Equation (5) [30].

$$\hat{y} = mode\{0,0,1\} = 0 \tag{5}$$

By MV, we would classify the sample into "Class 0". The MV process is shown in Figure 8 [31].

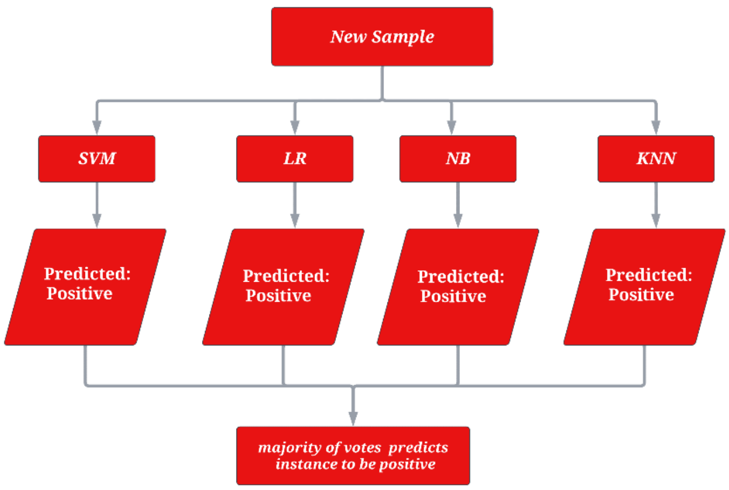

Figure 8. MV process.

2.4.1.2. Probabilistic voting

In the PV scheme, each base classifier provides a probability estimate indicating the likelihood that a given data point belongs to a particular target class. These probability predictions are assigned weights proportional to the size of the classifier and aggregated. The final prediction is determined by selecting the target label associated with the highest sum of weighted probabilities.

Soft voting, on the other hand, involves predicting class labels by considering the expected probabilities (p) provided by each classifier. It's important to note that this approach is only advisable if the classifiers are appropriately calibrated [30].

$$\hat{y} = arg\,\max_{i} \sum_{j=1}^{m} w_j p_{ij} \tag{6}$$

w_j is the weight that can be assigned to the j th classifier in Equation (6) [31]. The PV procedure is shown in Figure 9.

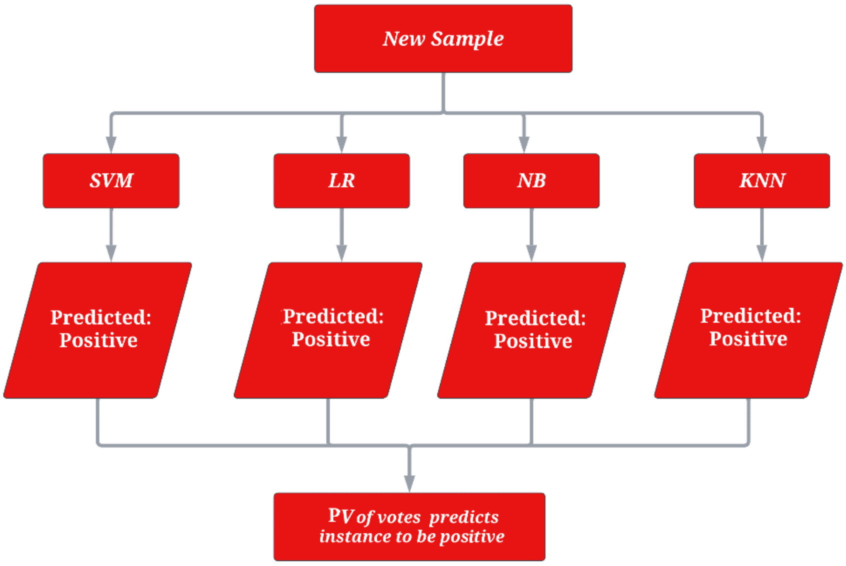

Figure 9. PV process.

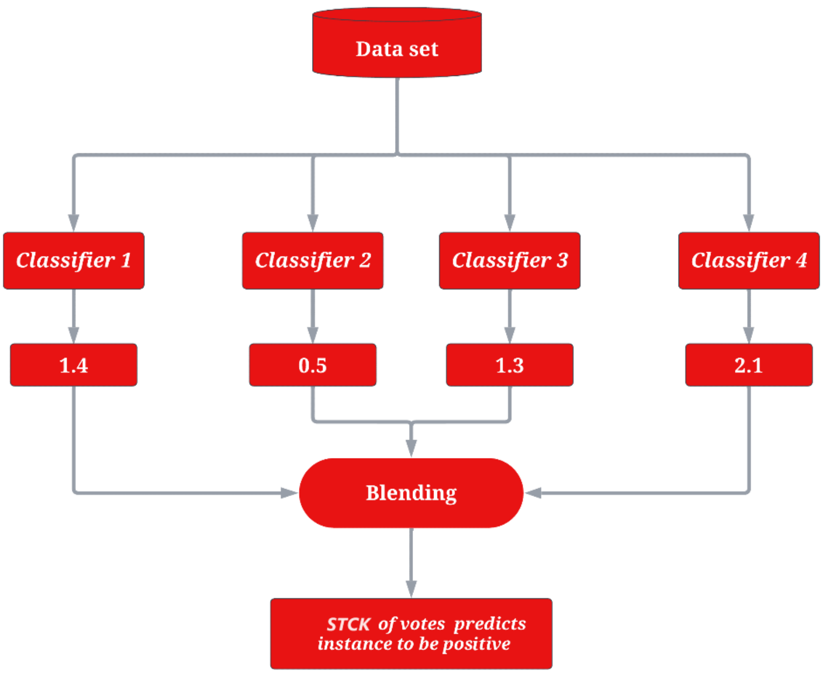

Figure 10. STCK process.

2.4.1.3. Stacking

This voting is similar to the voting used in classification problems. In addition to the selection of multiple sub-models, STCK allows you to specify an additional model to learn how to best combine the predictions of the sub-models. Because a metamodel is used to best combine the predictions of the sub-models, this method is sometimes referred to as mixture, as is mixture of predictions [32]. The STCK process is illustrated in Figure 10.

3. Experimental settings

In the experimental results, the preprocessing and feature extraction is completed and the performance of the data set is evaluated ACC, F by classifying it with ensemble and machine learning classifiers after hold-out test training separation. The experiments were coded using the Python sci-kit learn library. The experiments were performed on an AMD Ryzen CPU computer.

3.1. Performance metrics

ACC is the ratio of the true negative (T_N) and true positive (T_P) fields correctly predicted by the model to the sum of the false negative (F_N), and false positive (F_P) values contained in these fields. The ACC value is given by Equation (7) [27].

$$Acc = \frac{T_N + T_P}{T_P + T_N + F_N + F_P} \tag{7}$$

P is the ratio of (T_P) to (T_P) and (F_P) as given in Equation (8) [32].

$$P = \frac{T_P}{T_P + F_P} \tag{8}$$

R refers to the ratio of (T_P) to (F_P) and (T_P) given by Equation (9) [33].

$$R = \frac{T_P}{T_P + F_N} \tag{9}$$

F has values between 0 and 1. It is the harmonic mean of P and R defined by Equation (10) [33].

$$F = \frac{2 * Precision * Recall}{Precision + Recall} \tag{10}$$

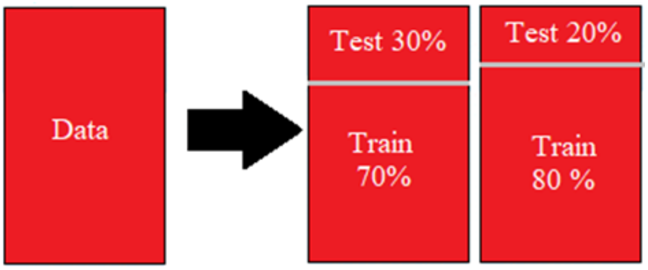

Figure 11. Train and test set separations.

Data sets were divided into training and test sets to develop models using classifier algorithms, followed by evaluation of model performance. Holdout was preferred in this research. Figure 11 shows the separation of training and test sets used in the study.

4. Results

In the Coronavirus data set, text representation methods and word embedding methods, then STCK and Voting (MV, PV) ensemble learning (background is blue) hold created using DT, KNN, NB, LR, RF, SVM machine learning methods (background is green) 70%-30% and 80%-20% ACC (Table 3, Table 4), F (Table 5, Table 6), shows the results.

Table 3 shows the ACC results of the sentiment analysis models generated by 80%-20% train-test separation after frequency-based text representation (TF-IDF, BOW) and word embedding in 100 and 200 vector sizes (Word2Vec and Doc2Vec). MV showed better classification performance after BOW and TF-IDF methods. STCK performed better on both Word2vec and Doc2vec methods and all vector sizes.

Table 4 shows the ACC results of the generated sentiment analysis models with 70-30% train-test separation after frequency-based text representation (TF-IDF, Bow) and vector sizes of 100 and 200 words (Word2Vec, Doc2Vec). STCK ensemble learning performed better in all models created after word embedding and text representation methods. But MV and PV also gave results close to STCK.

As shown in Table 5, the STCK ensemble learning model, which was developed after all text representation methods, performed better than the other models. Among the machine learning models, although SVM has given the best results, DT, KNN, which came after LR and SVM, gave the worst results. Ensemble learning methods outperform machine learning methods.

Comparison of the 200-vector size CBOW Word2Vec model with STCK 80%-20% train-test separation, which gave the best results in the study, and similar studies on coronavirus are shown in Table 7.

Among the models created for sentiment analysis between Table 3 and Table 7, the ensemble models produced ACC ranging from 73% to 89%, while the ACC of the single classifier models ranged from 60% to 80%. In terms of F performance, the results ranged from 68% to 78% for the ensemble models and from 60% to 76% for the single classifiers. After all text representation methods, ensemble learning models built with STCK gave better results in all performance evaluation criteria. According to the experimental results, ensemble models built with heterogeneous machine learning algorithms gave better results than machine learning algorithms used as a single classifier.

5. Conclusion and Discussion

Sentiment analysis of various social media topic data is so popular today. In this study, an evaluation of six single machine learning classifiers (SVM, DT, RF, LR, NB, KNN) versus ensemble learning methods (MV, PV, STCK) based on listed single classifiers on Twitter data about coronavirus with frequency-based text representation and word embedding methods have been investigated with different experiments.

The experimental results obtained in the study are presented between Table 3 and Table 6. It has been observed that using machine learning algorithms in heterogeneous ensemble learning algorithms gives a significant advantage in terms of classification performance compared to single classifiers. And STCK with Skip-Gram gave the best performance among all the models.

In the study, we obtained the best model with STCK ensemble learning after Word2Vec CBOW at 100 and 200 vector size. Table 7 shows a comparison of the presented model with other coronavirus studies in the literature.

Table 3. ACC results of models (80:20 Train: Test).

Models	Text representation/Embedding methods									
	BOW	TF-IDF	W2V				Doc2Vec			
			CBOW		Skip-Gram		PV-DM		PV-DBOW	
			100	200	100	200	100	200	100	200
LR	78	75	67	69	67	68	62	64	61	67
SVM	77	77	67	69	66	68	62	64	61	66
RF	79	78	72	72	72	72	67	67	64	71
NB	77	77	72	74	65	75	64	65	63	66
KNN	63	69	67	68	69	69	62	62	59	67
DT	71	69	62	61	60	61	65	65	64	69
STCK	**83**	**82**	84	**89**	**84**	**85**	**79**	**81**	**83**	**83**
MV	82	80	83	84	82	84	75	76	73	79
PV	81	81	83	84	82	84	75	76	83	79

Table 4. ACC results of models (70:30 Train: Test).

Models	Text representation/Embedding methods									
	BOW	TF-IDF	Word2Vec				Doc2Vec			
			CBOW		Skip-Gram		PV-DM		PV-DBOW	
			100	200	100	200	100	200	100	200
LR	79	76	69	70	67	69	62	64	61	67
SVM	78	78	68	70	67	69	62	63	61	66
RF	80	80	72	72	73	73	67	67	72	71
NB	77	77	73	74	67	74	65	66	63	66
KNN	62	69	67	67	68	67	60	69	69	67
DT	71	70	61	61	61	61	64	66	64	69
STCK	83	82	84	85	84	85	81	80	83	83
MV	82	81	83	84	83	84	76	77	73	79
PV	82	80	82	83	80	83	75	77	73	79

Table 5. F results of models (80:20 Train: Test).

Models	Text representation/Embedding methods									
	BOW	TF-IDF	W2V				Doc2Vec			
			CBOW		Skip-Gram		PV-DM		PV-DBOW	
			100	200	100	200	100	200	100	200
LR	72	71	73	76	73	71	71	70	68	67
SVM	72	72	76	76	73	73	69	71	70	68
RF	72	72	74	75	74	72	70	65	67	65
NB	67	66	76	74	77	72	64	65	63	66
KNN	63	63	67	68	68	69	62	62	61	64
DT	61	64	62	61	60	61	60	60	60	65
STCK	72	72	76	77	76	78	74	75	73	75
MV	71	71	74	76	74	75	73	74	72	73
PV	70	71	73	73	72	73	72	72	71	72

Table 6. F results of models (70:30 Train: Test).

Models	Text representation/Embedding methods									
	BOW	TF-IDF	W2V				Doc2Vec			
			CBOW		Skip-Gram		PV-DM		PV-DBOW	
			100	200	100	200	100	200	100	200
LR	72	71	71	72	70	71	72	73	68	72
SVM	72	72	72	74	71	72	72	73	69	73
RF	71	71	70	73	70	71	71	71	64	71
NB	67	67	69	68	69	70	65	66	62	66
KNN	71	69	67	67	68	67	60	61	60	66
DT	70	70	61	61	61	60	60	60	60	60
STCK	73	72	77	77	76	76	76	75	72	74
MV	72	71	73	74	73	74	75	74	70	68
PV	71	71	73	74	73	74	75	74	71	72

In Table 7, it is seen that the proposed method can compete with the literature when compared with other coronavirus studies in the literature.

As future work, we plan to improve the performance of sentiment analysis by using various deep learning algorithms in hybrid forms after feature selection on social media data along with text representation methods. Also, performance analysis of models based on transfer learning can be investigated on different data sets.

Table 7. Presented model versus literature.

References	Model	ACC (%)
[2]	LSTM	65
[3]	SGD	65
[4]	LR	78.5
[5]	NB	74.29
[6]	Trigram NB	84
[7]	MV (SVM, NB)	87.7
[8]	SVM	84.38
[9]	MV	83.3
	STCK	83.5
[10]	SVM	87.8
[11]	SGD	85.141
[12]	ANN	70.6
[13]	SVM	63
[14]	LR	81.8
[15]	Voting	74
The Presented model	STCK with CBOW	89

Author contributions

Muhammet Sinan Başarslan: Data preprocessing, data analysis, drafting the manuscript. **Fatih Kayaalp:** Defining the methodology, evaluating the results and editing the draft.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Cauberghe, V., Van Wesenbeeck, I., De Jans, S., Hudders, L., & Ponnet, K. (2021). How adolescents use social media to cope with feelings of loneliness and anxiety during COVID-19 lockdown. Cyberpsychology, Behavior, and Social Networking, 24(4), 250-257. https://doi.org/10.1089/cyber.2020.0478
2. Vernikou, S., Lyras, A., & Kanavos, A. (2022). Multiclass sentiment analysis on COVID-19-related tweets using deep learning models. Neural Computing and Applications, 34(22), 19615-19627. https://doi.org/10.1007/s00521-022-07650-2
3. Antonio, V. D., Efendi, S., & Mawengkang, H. (2022). Sentiment analysis for Covid-19 in Indonesia on Twitter with TF-IDF featured extraction and stochastic gradient descent. International Journal of Nonlinear Analysis and Applications, 13(1), 1367-1373. https://doi.org/10.22075/IJNAA.2021.5735
4. Machuca, C. R., Gallardo, C., & Toasa, R. M. (2021). Twitter sentiment analysis on coronavirus: Machine learning approach. In Journal of Physics: Conference Series, 1828(1), 012104. https://doi.org/10.1088/1742-6596/1828/1/012104
5. Barkur, G., & Kamath, G. B. (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. Asian Journal of Psychiatry, 51, 102089. https://doi.org/10.1016/j.ajp.2020.102089
6. Isnain, A. R., Marga, N. S., & Alita, D. (2021). Sentiment analysis of government policy on corona case using Naive Bayes Algorithm. Indonesian Journal of Computing and Cybernetics Systems, 15(1), 55-64. https://doi.org/10.22146/ijccs.60718
7. Siddiqua, U. A., Ahsan, T., & Chy, A. N. (2016). Combining a rule-based classifier with ensemble of feature sets and machine learning techniques for sentiment analysis on microblog. In 2016 19th International Conference on Computer and Information Technology, 304-309. https://doi.org/10.1109/ICCITECHN.2016.7860214
8. Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B. (2019). Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine. Komputek, 3 (2), 52.
9. Rahman, M. M., & Islam, M. N. (2022). Exploring the performance of ensemble machine learning classifiers for sentiment analysis of COVID-19 tweets. In Sentimental Analysis and Deep Learning: Proceedings of ICSADL 2021, 383-396. https://doi.org/10.1007/978-981-16-5157-1_30
10. Bania, R. K. (2020). COVID-19 public tweets sentiment analysis using TF-IDF and inductive learning models. INFOCOMP Journal of Computer Science, 19(2), 23-41.
11. Antonio, V. D. (2021). Performance analysis of TF-IDF feature extraction for stochastic gradient descent classification algorithm on sentiment analysis of Indonesian texts. [Doctoral Dissertation, Universitas Sumatera Utara].
12. Amalia, C., & Sibaroni, Y. (2020). Analisis sentimen data tweet menggunakan model jaringan saraf tiruan dengan pembobotan delta tf-idf. eProceedings of Engineering, 7(2), 7810-7820.
13. Ly, D., & Saad Abdul Malik, T. (2021). How can a module for sentiment analysis be designed to classify tweets about covid19. [Student thesis, University of Borås].
14. Bhardwaj, M., Mishra, P., Badhani, S., & Muttoo, S. K. (2023). Sentiment analysis and topic modeling of COVID-19 tweets of India. International Journal of System Assurance Engineering and Management, 1-21. https://doi.org/10.1007/s13198-023-02082-0
15. AlZoubi, O., Shatnawi, F., Rawashdeh, S., Yassein, M. B., & Hmeidi, I. (2022). Detecting COVID-19 Implication on Education and Economic in Arab World Using Sentiment Analysis Techniques of Twitter Data. In 2022 13th International Conference on Information and Communication Systems, 352-357. https://doi.org/10.1109/ICICS55353.2022.9811166
16. Miglani, A. (2020). Coronavirus tweets nlp-text classification.https://www.kaggle.com/datatattle/covid-19-nlp-textclassification
17. Huanling, T., Hui, Z., Hongmin, W., Han, Z., Xueli, M., Mingyu, L., & Jin, G. (2023). Representation of Semantic Word Embeddings Based on SLDA and Word2vec Model. Chinese Journal of Electronics, 32(3), 647-654. https://doi.org/10.23919/cje.2021.00.113

18. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 26.
19. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
20. Hidayat, T. H. J., Ruldeviyani, Y., Aditama, A. R., Madya, G. R., Nugraha, A. W., & Adisaputra, M. W. (2022). Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier. Procedia Computer Science, 197, 660-667. https://doi.org/10.1016/j.procs.2021.12.187
21. Dündar, A., & Kakışım, A. (2021). Kıyafet Öneri Sistemi için Giyim Metaverilerine dayalı Temsil Öğrenimi. Avrupa Bilim ve Teknoloji Dergisi, (29), 105-110. https://doi.org/10.31590/ejosat.1008736
22. Başarslan, M. S., & Kayaalp, F. (2019). Performance analysis of fuzzy rough set-based and correlation-based attribute selection methods on detection of chronic kidney disease with various classifiers. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science, 1-5. https://doi.org/10.1109/EBBT.2019.8741688
23. Turgut, Z., & Kakisim, A. G. (2024). An explainable hybrid deep learning architecture for WiFi-based indoor localization in Internet of Things environment. Future Generation Computer Systems, 151, 196-213. https://doi.org/10.1016/j.future.2023.10.003
24. Basarslan, M. S., Bakir, H., & Yücedağ, İ. (2019). Fuzzy logic and correlation-based hybrid classification on hepatitis disease data set. The International Conference on Artificial Intelligence and Applied Mathematics in Engineering, 787-800. https://doi.org/10.1007/978-3-030-36178-5_68
25. Rahardi, M., Aminuddin, A., Abdulloh, F. F., & Nugroho, R. A. (2022). Sentiment analysis of Covid-19 vaccination using support vector machine in Indonesia. International Journal of Advanced Computer Science and Applications, 13(6), 534-539.
26. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21-27. https://doi.org/10.1109/TIT.1967.1053964
27. Kakisim, A. G. (2022). Enhancing attributed network embedding via enriched attribute representations. Applied Intelligence, 52(2), 1566-1580. https://doi.org/10.1007/s10489-021-02498-w
28. Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University-Computer and Information Sciences, 35(2), 757-774. https://doi.org/10.1016/j.jksuci.2023.01.014
29. Onan, A. (2020). Mining opinions from instructor evaluation reviews: a deep learning approach. Computer Applications in Engineering Education, 28(1), 117-138. https://doi.org/10.1002/cae.22179
30. Kakisim, A. G., Turgut, Z., & Atmaca, T. (2023). XAI empowered dual band Wi-Fi based indoor localization via ensemble learning. In 2023 14th International Conference on Network of the Future (NoF), 150-158. https://doi.org/10.1109/NoF58724.2023.10302788
31. Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 6(3), 21-45. https://doi.org/10.1109/MCAS.2006.1688199
32. Öztürk, T., Turgut, Z., Akgün, G., & Köse, C. (2022). Machine learning-based intrusion detection for SCADA systems in healthcare. Network Modeling Analysis in Health Informatics and Bioinformatics, 11, 47. https://doi.org/10.1007/s13721-022-00390-2
33. Kayaalp, F., Basarslan, M. S., & Polat, K. (2018). A hybrid classification example in describing chronic kidney disease. In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 1-4. https://doi.org/10.1109/EBBT.2018.8391444