Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2024, 2024:95

https://doi.org/10.28919/cmbn/8774

ISSN: 2052-2541

COMPARATIVE ANALYSIS OF K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE IN CLASSIFICATION OF COVID 19 DISEASE IN MAKASSAR CITY

MUAMMAR ASHARI ABUSPIN, ERNA TRI HERDIANI*, GEORGINA MARIA TINUNGKI

Department of Statistics, Hasanuddin University, Makassar 90245, Indonesia

Abstract: Coronavirus disease 2019 or better known as COVID-19 is an outbreak that was initially detected in the city of Wuhan, China in December 2019. Before it was called COVID-19, WHO or the World Health Organization gave this new virus a temporary name as 2019 Novel Coronavirus (2019). nCoV). And on April 21 2020 WHO officially called the 2019-nCoV virus COVID-19. There are 4 factors that influence COVID 19 patients and these factors will be considered. To analyze the impact of the factors, the K-Nearest Neighbor (kNN) and Support Vector Machine (SVM) algorithms use JASP. The aim of this research is determine the comparison of classification accuracy levels K-Nearest Neighbor and Support Vector Machine towards Covid 19 patients. The results show that SVM achieved a higher level of accuracy, namely 98.43% compared to the kNN method which produced an accuracy of 98.40%, when applied to COVID 19 patient data in Makassar city.

Keywords: Covid 19; K nearest neighbor; support vector machine.

2020 AMS Subject Classification: 92D30.

*Corresponding author

E-mail address: herdiani.erna@unhas.ac.id

Received July 23, 2024

1. INTRODUCTION

Based on data obtained from WHO, there are 179 countries that have been exposed to the COVID-19 virus. The first COVID-19 instances in Indonesia were found in Depok, West Java, on March 2, 2020. Then, on June 26, 2020, this outbreak quickly extended to every province in Indonesia, resulting in 51.427 positive cases, 21.333 recovered cases, and 2.683 deaths[1].This indicates that this virus has a very high and fast exposure rate. The way it is distributed is also very simple [2]. Spread can be through sneezing, coughing, or interacting with someone who is infected.

And this virus is more susceptible to older people and those who already have a history of serious illness [3]. There are several factors that influence the rapid spread of this virus, namely old age, the number of people traveling to countries that have been infected, having contact with infected people, the presence of comorbidities and so on [4]. These factors can become data and can be processed using data mining.

There are three methods of data mining namely, prediction, Association, and Segmentation [5]. Prediction types are divided into three, namely Classification, Regression, and Time Series [6]. Algorithms used in classification include kNN, Naïve Bayes, Decision Trees, Support Vector Machines, and Neural Networks [7][8]. In carrying out the Classification method, there is an estimation process called simple/single split, namely separating data for training (70%) and testing (30%)[9]. This is used to see the predictions of the classification method. Classification works through recognizing patterns or models of a class [10]. Classification aims to ensure that these patterns can be used to make predictions or classifications of someone based on analysis of training data [11].

Classification is one of the deep techniques text mining and data mining which is used in the process of searching for a model or function that explains or differentiates classes in data and concepts with the aim of using the model in making predictions about data testing [12]. Classification algorithms have their respective advantages and disadvantages in classifying data in text form, including classification using the Support Vector Machine algorithm and kNN which has the highest level of accuracy compared to other algorithms [13]. The advantages of the Support Vector Machine (SVM) algorithm are that it has high accuracy, is efficient in using memory and can handle data that is not normally distributed [14]. Meanwhile, the advantage of the K-Nearest Neighbor (kNN) algorithm is that it is robust against training data that has a lot of noise and large amounts of data [15].

Previous research using the Data Science for COVID-19 (DS4C) dataset taken from Kaggle which was also used in this research was carried out by Al-Najjar and Al-Rousan discussing the prediction of recovery and death of Covid-19 patients in South Korea with an algorithm that used is Artificial Neural Network (ANN) [16]. Using the Support Vector Machine and K-Nearest Neighbor classification models, the research that will be conducted this time will classify using the attributes used, namely Gender, Age, and Comorbidities, with the goal of comparing the accuracy of the two algorithms. Preliminary research will be reviewed. So the researchers decided to use the K-Nearest Neighbor and SVM methods. The performance of the two methods will be compared so that the most effective method for classifying can be identified. What differentiates this research from other research is that it uses three Cross Validations in the algorithm experiment, so that it gets more accurate results. In order to achieve the highest level of accuracy possible using the K-Nearest Neighbor algorithm, this study additionally employed ten trials for the K value, which ranges from K=1 to K=10. The study utilized JASP (Jeffreys's Amazing Statistics Program) software to analyze COVID-19 patient data.

2. PRELIMINARIES

A. Data Source and variable

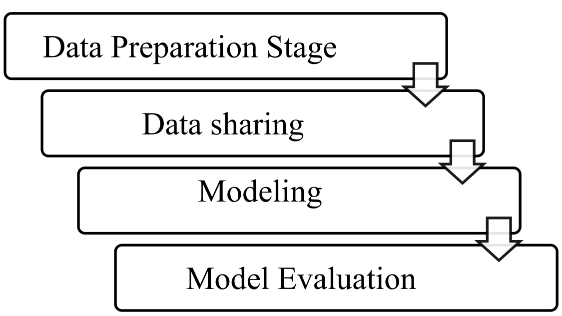
This study makes use of secondary data from the 2020–2021 COVID 19 patient population that was collected from the Makassar City Health Service. Table 1 provides an explanation of the variables' identification.

Table 1. Identification of Data

Variables	Overview	Measures	
y	Status	1	Dead
		0	Life
x_1	Gender	1	Women
		2	Men
x_2	Age	Year	
x_3	Comorbid	1	Positive with comorbidities
		2	Positive with non comorbidities
x_4	Length of Treatment	Day	

B. Research Stages

This research focuses on model development data mining by comparing classifications Support Vector Machine and K-Nearest Neigbor by [17]. The analysis method is structured as follows:



C. kNN

Steps in the kNN algorithm [18]:

1. Determine the number of neighbors (K) that will be used to consider class determination.
2. Calculate the distance from the new data to each data point in the dataset.
3. Take a number of K data with the closest distance, then determine the class of the new data.

To find how close or far the distance between points in class k is usually calculated using

Euclidean distance. Euclidean distance is a formula for finding the distance between 2 points in two-dimensional space. The formula for determining Euclidean distance is as follows [19]:

$$\hat{Y}(P) = \sqrt{\sum_{i=1}^{n} (Xtrainingi - Ytestingi)^2}$$

Information:

$\hat{Y}$ (P) = Status

$X_{trainingi}$ = Data Training

$Y_{testingi}$ = Data Testing

i = Data Variables

n = Data Dimensions

D. SVM

The steps of the SVM algorithm include [20]:

(1) selecting an appropriate kernel function,

(2) determining parameters and constraints,

(3) solving the optimization problem to find the optimal hyperplane, and

(4) making predictions based on the learned model.

E. Classification Performance

To determine whether or not the model utilized is good, the accuracy of the classification model is evaluated using the confusion matrix approach. The example confusion matrix that follows is based on [21].

Table 2. Illustration of the confusion matrix

Class	**Current Positive (1)**	**Current Negative (0)**
Positive Predictions (1)	TP (True)	FP (False)
Negative Prediction (0)	FN (False)	TN (True)

Where:

TP: The model predicted correctly and it was correct

TN: The model predicted negative and it was correct

FP: The model predicted positive and that was wrong

FN: The model predicted negative and that was wrong

Based on the values (TP), (TN), (FP), and (FN), a value is obtained *Accurasi, Precision*, *Recall,* and *F-Measure* [22].

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

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

$$Recall = \frac{(TP)}{(TP + FN)}$$

$$F - Measure = 2\,\frac{(Recall \times Precision)}{(Recall + Precision)}$$

3. MAIN RESULTS

A. kNN Model Testing

K-Nearest Neighbor classification models To predict whether COVID-19 patients would be classified as Alive or Dead, provide training data. The steps involved in creating a model are as follows: the data is separated into two sections, data training and data testing, following a series of procedures including pre-processing and cleaning. To produce a model, data is needed training, while for data testing used in evaluating the model that has been created, then selecting the value of K (nearest neighbor) 10 times, namely K=1, K=2, K=3, K=4, K=5, K=6, K=7, K=8, K=9, and K=10 and the best K value is taken based on value Accuracy the best.

Testing 1 kNN

In Test 1, there were 48,913 training data and 5,434 testing data, with a 90% training and 10% testing data distribution. Regarding the Confusion Matrix results, which are detailed in Table 3 below:

Table 3. Confusion Matrix kNN Test 1

Confusion Matrix

		Predicted	
		Life	**Dead**
Observed	Life	5310	16
	Dead	76	32

Table 3 is a calculation according to testing data, in Table 2 It is recognized that from 5,434 data, 5310 were classified as Alive according to the predictions made by the method K-Nearest Neighbor, then 76 data were predicted to be Alive but turned out to be Dead, 32 data classified as Dead were predicted accordingly, and 16 data were predicted to be Dead but turned out to be Alive.

Test results using the model K-Nearest Neighbor where the best K value K=5 is obtained accuracy = 98.30%, precission = 98.80%, recall = 98.30% as in Figure 1 and Accuracy and AUC values in table 4 below:

Figure 1. Accuracy Precision, Recall on kNN Test 1

Model Performance Metrics ▼

	Hidup	Mati	Average / Total
Support	5326	108	5434
Accuracy	0.983	0.983	0.983
Precision (Positive Predictive Value)	0.986	0.667	0.980
Recall (True Positive Rate)	0.997	0.296	0.983
False Positive Rate	0.704	0.003	0.353
False Discovery Rate	0.014	0.333	0.174
F1 Score	0.991	0.410	0.980
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	0.799	0.799	0.799
Negative Predictive Value	0.667	0.986	0.826
True Negative Rate	0.296	0.997	0.647
False Negative Rate	0.003	0.704	0.353
False Omission Rate	0.333	0.014	0.174
Threat Score	31.607	0.296	15.952
Statistical Parity	0.991	0.009	1.000

Note. All metrics are calculated for every class against all other classes.

Table 4. Accuracy and AUC values of kNN Test 1

	K Nearest Neighbor (kNN)
Accuracy	98.30%
AUC	0.799

·**Testing 2 kNN**

In Test 2, there were 43,378 training data and 10,869 testing data, with an 80% training and 20% testing data distribution. Regarding the Confusion Matrix results, which are detailed in Table 5 below:

Table 5. Confusion Matrix kNN Test 2

Confusion Matrix

		Predicted	
		Life	**Dead**
Observed	Life	10610	40
	Dead	125	94

Table 5 is a calculation according to testing data. In Table 5 It is recognized that from 10,869 data, 10,610 were classified as Alive according to predictions made using the KNearest Neighbor, then 125 data were predicted to be Alive but turned out to be Dead, 94 data classified as Dead were predicted accordingly, and 40 data were predicted to be Dead but turned out to be Alive.

Test results using the model K-Nearest Neighbor where the best K value K=4 is obtained accuracy = 98.50%, precission = 98.30%, recall = 98.50% as in Figure 2 and Accuracy and AUC values in table 6 below:

Figure 2. Accuracy, Precision, Recall on kNN Test 2

Model Performance Metrics ▼

	Hidup	Mati	Average / Total
Support	10650	219	10869
Accuracy	0.985	0.985	0.985
Precision (Positive Predictive Value)	0.988	0.701	0.983
Recall (True Positive Rate)	0.996	0.429	0.985
False Positive Rate	0.571	0.004	0.287
False Discovery Rate	0.012	0.299	0.155
F1 Score	0.992	0.533	0.983
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	0.804	0.804	0.804
Negative Predictive Value	0.701	0.988	0.845
True Negative Rate	0.429	0.996	0.713
False Negative Rate	0.004	0.571	0.287
False Omission Rate	0.299	0.012	0.155
Threat Score	36.586	0.459	18.522
Statistical Parity	0.988	0.012	1.000

Note. All metrics are calculated for every class against all other classes.

Table 6 Accuracy and AUC kNN values for Test 2

	K Nearest Neighbor (SVM)
Accuracy	98.50%
AUC	0.804

· **Testing 3 kNN**

In Test 3, there were 16,304 testing data and 38,043 training data, with a 70% distribution of training data and a 30% distribution of testing data. Regarding the Confusion Matrix results, which are detailed in Table 7 below:

Table 7. Confusion Matrix kNN Test 3

Confusion Matrix

		Predicted Life	Predicted Dead
Observed	Life	15944	38
	Dead	215	107

Table 7 is a calculation according to testing data. In Table 6 It is recognized that from 16,304 data, 15,944 were classified as Alive according to predictions made using the KNearest Neighbor, then 215 data were predicted to be Alive but turned out to be Dead, 107 data classified as Dead were predicted accordingly, and 38 data were predicted to be Dead but turned out to be Alive.

Test results using the model K-Nearest Neighbor where the best K value K=6 is obtained accuracy = 98.40%, precission = 98.20%, recall = 98.40% as in Figure 3 and Accuracy and AUC values in table 8 below:

Figure 3. Accuracy, Precision, Recall on kNN Test 3

Model Performance Metrics ▼

	Hidup	Mati	Average / Total
Support	15982	322	16304
Accuracy	0.984	0.984	0.984
Precision (Positive Predictive Value)	0.987	0.738	0.982
Recall (True Positive Rate)	0.998	0.332	0.984
False Positive Rate	0.668	0.002	0.335
False Discovery Rate	0.013	0.262	0.138
F1 Score	0.992	0.458	0.982
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	0.823	0.823	0.823
Negative Predictive Value	0.738	0.987	0.862
True Negative Rate	0.332	0.998	0.665
False Negative Rate	0.002	0.668	0.335
False Omission Rate	0.262	0.013	0.138
Threat Score	34.068	0.368	17.218
Statistical Parity	0.991	0.009	1.000

Note. All metrics are calculated for every class against all other classes.

Table 8. Accuracy and AUC kNN values for Test 3

	K Nearest Neighbor
	(kNN)
Accuracy	98.40%
AUC	0.823

From the results of tests that have been carried out 3 times, the average values of Accuracy, Precision, Recall are obtained as shown in Table 8 as follows:

Table 9. Average Value of Accuracy, Precision, Recall and AUC on kNN

Test	Training %	Testing %	Accuracy%	Precision%	Recall%	AUC value
1	90	10	98,30%	98,80%	98,30%	0,799
2	80	20	98,50%	98,30%	98,50%	0,804
3	70	30	98,40%	98,20%	98,40%	0,823
Average			98,40%	98,43%	98,40%	0,809

Table 9 illustrates that the K-Nearest Neighbor model is capable of classifying COVID-19 patient data. The average result after three tries to alter the training and testing data was 98.40% accuracy, 98.43% precision, 98.40% recall, and 0.809 AUC.

B. SVM Model Testing

Regarding categorization schemes Vector Machine Support To predict whether COVID-19 patients would be classified as Alive or Dead, provide training data. The stages in building a model include: Following a series of procedures, including pre-processing and cleaning, the data is separated into two components: data training and data testing. To produce a model, data is needed training, while for data testing used in evaluating the model that has been created. This test was carried out to find out how much training and testing data influences the classification carried out by Support Vector Machine.

- **Testing 1 SVM**

In Test 1, there were 48,913 training data and 5,434 testing data, with a 90% training and 10% testing data distribution. Regarding the Confusion Matrix results, which are detailed in Table 10 below:

Table 10. Confusion Matrix SVM Test 1

Confusion Matrix

		Predicted	
		Life	**Dead**
Observed	Life	5314	10
	Dead	65	45

Table 10 is a calculation according to testing data, in Table 9 It is recognized that from 5,434 data, 5314 were classified as Alive according to the predictions made by the method Support Vector Machine, then 65 data were predicted to be Alive but turned out to be Dead, 45 data classified as Dead were predicted accordingly, and 65 data were predicted to be Dead but turned out to be Alive.

Test results using the model Support Vector Machine results were obtained accuracy = 98.60%, precission = 98.40%, recall = 98.60% as in Figure 4 and Accuracy and AUC values in table 11 below:

Figure 4. Accuracy, Precision, Recall on SVM Test 1

Model Performance Metrics ▼

	Hidup	Mati	Average / Total
Support	5324	110	5434
Accuracy	0.986	0.986	0.986
Precision (Positive Predictive Value)	0.988	0.818	0.984
Recall (True Positive Rate)	0.998	0.409	0.986
False Positive Rate	0.591	0.002	0.296
False Discovery Rate	0.012	0.182	0.097
F1 Score	0.993	0.545	0.984
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	0.704	0.704	0.704
Negative Predictive Value	0.818	0.988	0.903
True Negative Rate	0.409	0.998	0.704
False Negative Rate	0.002	0.591	0.296
False Omission Rate	0.182	0.012	0.097
Threat Score	37.957	0.529	19.243
Statistical Parity	0.990	0.010	1.000

Note. All metrics are calculated for every class against all other classes.

Table 11. Accuracy and AUC SVM Values for Test 1

	Support Vector Machine (SVM)
Accuracy	98.60%
AUC	0.704

·Testing 2 SVM

In Test 2, there were 43,378 training data and 10,869 testing data, with an 80% training and 20% testing data distribution. Regarding the Confusion Matrix results, which are detailed in Table 12 below:

Table 12. Confusion Matrix SVM Test 2

Confusion Matrix

		Predicted	
		Life	**Dead**
Observed	Life	10598	36
	Dead	148	87

Table 12 is a calculation according to testing data. In Table 12 It is recognized that from 10,869 data, 10,598 were classified as Alive according to predictions made using the Support Vector Machine, then 148 data were predicted to be Alive but turned out to be Dead, 87 data classified as Dead were predicted accordingly, and 36 data were predicted to be Dead but turned out to be Alive.

Test results using the model Support Vector Machine results were obtained accuracy = 98.30%, precission = 98.00%, recall = 98.30% as in Figure 5 and Accuracy and AUC values in table 13 below:

Figure 5. Accuracy, Precission, Recall on SVM Test 2

Model Performance Metrics

	Hidup	Mati	Average / Total
Support	10634	235	10869
Accuracy	0.983	0.983	0.983
Precision (Positive Predictive Value)	0.986	0.707	0.980
Recall (True Positive Rate)	0.997	0.370	0.983
False Positive Rate	0.630	0.003	0.317
False Discovery Rate	0.014	0.293	0.153
F1 Score	0.991	0.486	0.980
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	0.683	0.683	0.683
Negative Predictive Value	0.707	0.986	0.847
True Negative Rate	0.370	0.997	0.683
False Negative Rate	0.003	0.630	0.317
False Omission Rate	0.293	0.014	0.153
Threat Score	31.922	0.395	16.159
Statistical Parity	0.989	0.011	1.000

Note. All metrics are calculated for every class against all other classes.

Table 13. Accuracy and AUC values for SVM Test 2

	Support Vector Machine (SVM)
Accuracy	98.30%
AUC	0.683

.

Testing 3 SVM

In Test 3, there were 16,304 testing data and 38,043 training data, with a 70% distribution of training data and a 30% distribution of testing data. Regarding the Confusion Matrix results, which are detailed in Table 14 below:

Table 14. Confusion Matrix SVM Test 3

Confusion Matrix

		Predicted	
		Life	**Dead**
Observed	Life	15901	62
	Dead	195	146

Table 14 is a calculation according to testing data. In Table 13 It is recognized that from 16,304 data, 15,901 were classified as Alive according to predictions made using the Support Vector Machine, then 195 data were predicted to be Alive but turned out to be Dead, 146 data classified as Dead were predicted accordingly, and 62 data were predicted to be Dead but turned out to be Alive.

Test results using the model Support Vector Machine results were obtained accuracy = 98.40%, precission = 98.20%, recall = 98.40% as in Figure 6 and Accuracy and AUC values in table 15 below:

Figure 6. Accuracy, Precission, Recall on SVM Test 3

Model Performance Metrics

	Hidup	Mati	Average / Total
Support	15963	341	16304
Accuracy	0.984	0.984	0.984
Precision (Positive Predictive Value)	0.988	0.702	0.982
Recall (True Positive Rate)	0.996	0.428	0.984
False Positive Rate	0.572	0.004	0.288
False Discovery Rate	0.012	0.298	0.155
F1 Score	0.992	0.532	0.982
Matthews Correlation Coefficient			NaN
Area Under Curve (AUC)	0.712	0.712	0.712
Negative Predictive Value	0.702	0.988	0.845
True Negative Rate	0.428	0.996	0.712
False Negative Rate	0.004	0.572	0.288
False Omission Rate	0.298	0.012	0.155
Threat Score	35.179	0.458	17.818
Statistical Parity	0.987	0.013	1.000

Note. All metrics are calculated for every class against all other classes.

Table 15. Accuracy and AUC values for SVM Test 3

	Support Vector Machine (SVM)
Accuracy	98.30%
AUC	0.712

According to the outcomes of three separate tests, the average values of Accuracy, Precision, Recall are obtained as shown in Table 16 as follows:

Table 16. Average Value of Accuracy, Precision, Recall and AUC on SVM

Test	**Training %**	**Testing %**	**Accuracy%**	**Precision%**	**Recall%**	**Nilai AUC**
1	90	10	98,60%	98,40%	98,60%	0,704
2	80	20	98,30%	98%	98,30%	0,683
3	70	30	98,40%	98,20%	98,40%	0,712
Rata-Rata			98,43%	98,20%	98,43%	0,700

Table 16 illustrates the model Support Vector Machine's capability to categorize COVID-19 patient data. The average result after three tries to alter the training and testing data was 98.43% accuracy, 98.20% precision, 98.43% recall, and 0.700 AUC.

C. Comparison Results of SVM and kNN

From the test results for each method, tests were carried out 8 times on each method, then a comparison was carried out between them Support Vector Machine with method K-Nearest Neighbor as in Table 17 below:

Table 17. Comparison Results of SVM and K-Nearest Neighbor Testing

Test	Training %	Testing %	Support Vector Machine				K-Nearest Neighbor			
			Accuracy	Precision	Recall	AUC	accuracy	Precision	recall	AUC
1	90	10	98,60%	98,40%	98,60%	0,704	98,30%	98,80%	98,30%	0,799
2	80	20	98,30%	98%	98,30%	0,683	98,50%	98,30%	98,50%	0,804
3	70	30	98,40%	98,20%	98,40%	0,712	98,40%	98,20%	98,40%	0,823
Average			98,43%	98,20%	98,43%	0,700	98,40%	98,43%	98,40%	0,809

Based on Table 4.40, it shows that the comparison of training data greatly influences the accuracy, precision and recall values of the two methods. From the results of trials carried out, the level of accuracy in classification of COVID 19 patient data in the city of Makassar between algorithms Support Vector Machine and K-Nearest Neighbor that the average value Accuracy and Recall Support Vector Machine outperform of K-Nearest Neighbor namely
98.43% and 98.43%, while the average value Accuracy and Recall K-Nearest Neighbor of 98.40% and 98.40%, For average value Precision and AUC Value K-Nearest Neighbor outperform of Support Vector Machine namely 98.43% and 0.809, while the average value Precision and AUC Value Support Vector Machine namely 98.20% and 0.700.

4. CONCLUSION

In this research, it was created using two algorithm models, namely Support Vector Machines and k-Nearest Neighbor using data from COVID 19 patients in the city of Makassar. The resulting models were compared to find out the best algorithm in determining factors for Covid 19 patients. To measure the performance of the two algorithms, the Cross Validation and Confusion Matrix testing methods were used. Support Vector Machines has value Accuracy highest and lowest methods k-Nearest Neighbor. From the results of this analysis it can be concluded that the method Support Vector Machine is a good method for classifying Covid 19 patient data compared to methods k-Nearest Neighbor.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interest.

REFERENCES

[1] T. Toharudin, R.S. Pontoh, S. Zahroh, et al. Impact of large scale social restriction on the COVID-19 cases in East Java, Commun. Math. Biol. Neurosci. 2020 (2020), 56. https://doi.org/10.28919/cmbn/4837.

[2] M. Coccia, An index to quantify environmental risk of exposure to future epidemics of the COVID-19 and similar viral agents: Theory and practice, Environ. Res. 191 (2020), 110155. https://doi.org/10.1016/j.envres.2020.110155.

[3] M.G. Ali, M.O. Ahmad, S.N. Husain, Spread of corona virus disease (COVID – 19) from an outbreak to pandemic in the Year 2020, Asian J. Res. Infect. Dis. 3 (2020), 37–51. https://doi.org/10.9734/ajrid/2020/v3i430135.

[4] S.A. Rasmussen, J.C. Smulian, J.A. Lednicky, et al. Coronavirus Disease 2019 (COVID-19) and pregnancy: what obstetricians need to know, Amer. J. Obstetr. Gynecol. 222 (2020), 415–426. https://doi.org/10.1016/j.ajog.2020.02.017.

[5] S.I. Murpratiwi, I.G. Agung Indrawan, A. Aranta, Analisis pemilihan cluster optimal dalam segmentasi pelanggan toko retail, J. Pendidik. Teknologi. Kejuruan. 18 (2021), 152-163. https://doi.org/10.23887/jptk-undiksha.v18i2.37426.

[6] D.M. Durairaj, B.H.K. Mohan, A convolutional neural network based approach to financial time series prediction, Neural Comput. Appl. 34 (2022), 13319–13337. https://doi.org/10.1007/s00521-022-07143-2.

[7] H. Dharmawan, B. Sartono, A. Kurnia, et al. A study of machine learning algorithms to measure the feature importance in class-imbalance data of food insecurity cases in Indonesia, Commun. Math. Biol. Neurosci. 2022 (2022), 101. https://doi.org/10.28919/cmbn/7636.

[8] M. Bansal, A. Goyal, A. Choudhary, A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning, Decis. Anal. J. 3 (2022), 100071. https://doi.org/10.1016/j.dajour.2022.100071.

[9] J. Richetti, F.I. Diakogianis, A. Bender, et al. A methods guideline for deep learning for tabular data in agriculture with a case study to forecast cereal yield, Computers Electron. Agric. 205 (2023), 107642. https://doi.org/10.1016/j.compag.2023.107642.

[10] J.H. Carter, The immune system as a model for pattern recognition and classification, J. Amer. Med. Inform. Assoc. 7 (2000), 28–41. https://doi.org/10.1136/jamia.2000.0070028.

[11] A. Nurmasani, Y. Pristyanto, Algoritme stacking untuk klasifikasi penyakit jantung pada dataset imbalanced class, Pseudocode. 8 (2021), 21–26. https://doi.org/10.33369/pseudocode.8.1.21-26.

[12] H. Fauzan, A. Adiwijaya, S. Al-Faraby, Pengklasifikasian topik hadits terjemahan bahasa Indonesia menggunakan latent semantic indexing dan support vector machine, J. Media Inform. Budid. 2 (2018), 131-139. https://doi.org/10.30865/mib.v2i4.948.

[13] E.Y. Boateng, J. Otoo, D.A. Abaye, Basic tenets of classification algorithms k-nearest-neighbor, support vector machine, random forest and neural network: a review, J. Data Anal. Inform. Process. 08 (2020), 341–357. https://doi.org/10.4236/jdaip.2020.84020.

[14] B. Gaye, D. Zhang, A. Wulamu, Improvement of support vector machine algorithm in big data background, Math. Probl. Eng. 2021 (2021), 5594899. https://doi.org/10.1155/2021/5594899.

[15] Y. Song, X. Kong, C. Zhang, A large-scale k-nearest neighbor classification algorithm based on neighbor relationship preservation, Wireless Commun. Mobile Comput. 2022 (2022), 7409171. https://doi.org/10.1155/2022/7409171.

[16] H. Al-Najjar, N. Al-Rousan, A classifier prediction model to predict the status of Coronavirus COVID-19 patients in South Korea, Eur. Rev. Med. Pharmacol. Sci. 24 (2020), 3400–3403. https://doi.org/10.26355/eurrev_202003_20709.

[17] S. Jesika, S. Ramadhani, Y.P. Putri, Implementasi model machine learning dalam mengklasifikasi kualitas air, J. Ilm. Karya Mahasiswa. 1 (2023), 382–396.

[18] G. Vinayakumar, A comparison of KNN algorithm and MNL model for mode choice modelling, Eur. Transport. 92 (2023), 1–14. https://doi.org/10.48295/et.2023.92.3.

[19] C. Feng, B. Zhao, X. Zhou, et al. An enhanced quantum K-nearest neighbor classification algorithm based on polar distance, Entropy. 25 (2023), 127. https://doi.org/10.3390/e25010127.

[20] Z. Quan, L. Pu, An improved accurate classification method for online education resources based on support vector machine (SVM): Algorithm and experiment, Educ. Inf. Technol. 28 (2022), 8097–8111. https://doi.org/10.1007/s10639-022-11514-6.

[21] J.H.J. C. Ortega, Analysis of performance of classification algorithms in mushroom poisonous detection using confusion matrix analysis, Int. J. Adv. Trends Comput. Sci. Eng. 9 (2020), 451–456. https://doi.org/10.30534/ijatcse/2020/7191.32020.

[22] G.P. Natakusumah, E. Ernastuti, Implementasi metode CNN multi-scale input dan multi-feature network untuk dugaan kanker payudara, J. Inform. Technol. Computer Sci. 7 (2022), 43-52. https://doi.org/10.31328/jointecs.v7i2.3637.