ILKOM JurnalIlmiah Vol. 15, No. 2, August 2023, pp. 240-249
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020;E-ISSN 2548-7779 | P-ISSN 2087-1716

The Effect of The Prediction of The K-Nearest Neighbor Algorithm on Surviving COVID-19 patients in Indonesia

Aris Martono [1,a]; Henderi [2,a,*]; Giandari Maulani [3,a]
[a]*University of Raharja, Modern, Jl. Jenderal Sudirman No.40, Cikokol, Kec. Tangerang, Banten, 15117, Indonesia*
[1] *aris.martono@raharja.info;* [2] *henderi@raharja.info;* [3] *giandari@raharja.info*
** Corresponding author*

Article history: Received June 08, 2022; Revised July 02, 2022; Accepted June 10, 2023; Available online August 16, 2023.

Abstract
This study aims to measure the prediction of survival of covid-19 patients with the best algorithm based on RMSE (Root Mean Square Error). The Covid-19 pandemic has lasted from December 2019 until now and is full of uncertainty about when this pandemic will end, so this research was carried out. In this study, the knowledge discovery database method was used by extracting data sets from Covid-19 patients from March 2020 to March 2021 for each province in Indonesia (Dataset from Kawal Covid-19 SintaRistekbrin) to predict survival during this pandemic as measured by the best algorithms include K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), and Deep Learning. The measurement results using cross-validation and the optimal number of folds is 3 in the form of RSME, showing that the K-NN algorithm is an algorithm with RSME 0.101 +/-0.23 where the error rate is the lowest compared to the two algorithms above. Therefore, the K-NN algorithm was chosen as the algorithm for the predictive measurement of surviving Covid-19 patients.

Keywords: Deep Learning; k-Nearest Neighbor; RMSE; Support Vector Machine.

Introduction

The spread of the Covid-19 pandemic has crossed between countries, so the community's economy has decreased. Every government hopes that this covid-19 pandemic will soon pass/end. With the end of this pandemic, economic and business activities, social and political as well as cultural, can run well and smoothly. Every country competes to maintain and implement health protocols strictly or loosely in addition to vaccinations. Implementing the health protocol avoids the transmission of Covid-19 among humans, while the performance of immunization increases the body's immunity against Covid-19 attacks. This Covid-19 pandemic has lasted from the end of December 2019 until now.

Some previous research aims to observe and analyze the development of the Covid-19 trend and its potential impact on changes in the population composition based on the age structure in Indonesia. In the research, the data used are Covid-19 cases obtained from the Ministry of Health of the Republic of Indonesia and forecasting methods by comparing several models. Study findings show that the trend shows an increase in cases and will continue to increase as long as there is no intervention. Experimentally, death cases dominated by the male population and elderly population are possible causes of changes in population composition based on age structure. Therefore, it is necessary to intervene immediately in the form of policies in the health sector that are more appropriate to maintain Indonesia's human resources.

Research about the prediction of the number of Covid-19 cases can be done using Forecasting Techniques. This study aims to predict and compare Single Exponential Smoothing and Double Exponential Smoothing to the number of Covid-19 cases in Indonesia. The distribution prediction is based on data released by the National Disaster Management Agency in the first 100 days of the spread of Covid-19.

The results of this study can be used as consideration in policy making in dealing with the spread of Covid-19. The Covid-19 pandemic has been going on since the end of February 2020 until now in Indonesia. Researchers have made several models and predictions for this pandemic case, but the results could be more accurate. Each region has different patterns, so it must accommodate these patterns. All aspects of human life are affected by the Covid-19 pandemic due to its spread worldwide.

http://dx.doi.org/10.33096/ilkom.v15i2.1234.240-249

With the spread of Covid-19 in Indonesia, it is necessary to predict its spread. The prediction of active cases of the spread of this virus is used by the Deep Learning method or Machine Learning which is a model like an artificial neural network, k-Nearest Neighbor, and Support Vector Machine. Therefore, this study was conducted to predict surviving covid-19 patients with the dataset "Covid-19 patients from March 2020 to March 2021 in every province in Indonesia" (Dataset from Kawal Covid-19 SintaRistekbrin) and measured by one of the algorithms that the best among the three algorithms: K-Nearest Neighbor, SVM (Support Vector Machine), and Deep Learning.

A. Literature Review

There are several previous studies to analyze this research with previous research, whether there are new findings or the development of previous research or fill in the gaps of several previous studies. The previous research including conducted by Long Short-Term Memory (LSTM) network method[1], Susceptible Infective Recovered Death[2], Particle Swarm Optimization-Based K-NN Algorithm[3], Support Vector Regression (SVR), and Susceptible Infectious Recovered (SIR)[4], Support Vector Machine[5], Deep Learning[6], and different machine learning approached[7].

On the others hand, there were several previous studies that made performance comparisons of some methods in machine learning and the prediction model. It was including comparison between Support Vector Machine (SVM), Artificial Neural Network (ANN), and Multiple Liner Regression (ML)[8], Linear Regression, Gradient Boosting, Random Forest, Bootstrap Aggregation (Bagging), Huber Regression, Bayesian Regression, and SVM[9], Machine Learning Based Regression Model and Deep Learning Model Based Long Short Term Memory (LSTM) Deep Learning[10], Fuzzy Logic model, Artificial Neural Networks, Multiple Regression Analysis, Case-Based Reasoning, Hybrid Models, Genetic Fuzzy Model, Scalable Boosting Trees (XGBoost) and Random Forest [11], Linear Regression, k-Nearest Neighbor Regression and Decision Tree Regression Algorithm[12], Random Forest, SVM, Naïve Bayes, Logistic Regression, KNN, XGBoost, Decision Tree and AdaBoost[13], supervised machine learning algorithm such k-Nearest Neighbor, Random Forest, Decision Tree, Naïve Bayes, Support Vector Machine and Logistic Regression[14], Logistic Regression, k-Nearest Neighbor, Support Vector Machine, Gaussian Naïve Bayes, Decision Tree, Random Forest and Artificial Neural Network[15], and the Principal Component Analysis and Sequential Minimal Optimization methods[16]. But in the group research, the comparison algorithm performance between K-Nearest Neighbor, Support Vector Machine, and Machine Learning for prediction model using the Covid-19 Data set was not conducted.

There is a lot of criteria performance that was conducted in comparing performance between methods in Machine Learning or Deep Learning. The comparison performance method has conducted based on Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percent Error (MAPE)[8], Quality Metrics and Execution Time[9], Coefficient of Determination values and Mean Absolute Error values[12], Accuracy Value[14], [16], precision, recall, ROC Curve and PRC[15], [17], Mean Absolute Error (MAE) and Mean Square Error (MSE)[18], Cross Validation[19], Root Mean Squared Error [20], Mean Accuracy[21], Internal and External Verification[22], Accuracy[23], Accuracy, Sensitivity, dan Specificity[24], and Precision, Recall, f1 Score, RMSE, Kappa Coefficient, Matthew Correlation Coefficient, Receiver Operating Characteristics (ROC) Curve and Accuracy[25]. Based on a lot of the previous research, this study is a predictive measurement of ongoing events for patients who survive and is national by using the best algorithm from several algorithms.

Method

Knowledge Discovery in Technology is a potentially valuable identification process with great significance and inevitability in the computer-aided medical diagnosis system[26]. The data is a collection of facts or egg cases in a database, and a pattern is an expression in some language that describes a part of the data or a model that can be applied to another. However, these uses include creating a pattern, properly designing a data model, determining a data structure, or some high-level schema explanation of a data set. The computer-aided and online data analysis system can help with knowledge discovery in database tasks, like selection, transformation, data mining, interpretation, and evaluation[27].

Therefore, this study uses the methods contained in several stages of the knowledge discovery in the database process, where all operations are extracted or identified potential patterns, knowledge, and information from large datasets. Knowledge discovery suggests advanced and efficient tools, methods, and technologies for accessing and processing data in a database[28]. The compilation of knowledge and information from knowledge discovery in the database is valid, new, easy to understand, and valuable. The stages of compiling knowledge include selection data, pre-processing data, transformation, and interpretation and evaluation. Data, data warehouse, data mining, and knowledge.

The first selection of the data. It was done because not all existing data can be used. Therefore, it is necessary to select the data where this data selection activity is the creation of data sets that are stored in storage media. Second, pre-processing or cleaning. In this stage, the selected data is cleaned, including deleting duplicate data, correcting inconsistent data, and correcting incorrect data. In this data-cleaning process, data enrichment can be done by adding relevant information.

The third is transformation. In this stage, we use data mining with various neutral network algorithms and K-NN to analyse the data, and the process of extracting and searching knowledge and information uses certain algorithms. For interpretation an evaluation, the knowledge, and information obtained by the data mining process are displayed in a form that is easily understood by interested parties, such as a form graphic display or a decision tree by the provisions. Knowledge and information obtained from the data mining process are trained whether or not it is by reality or hypotheses.

A. k-Nearest Neighbor

The supervised learning algorithm whose results are obtained from the majority classification and the k-Nearest Neighbour (k-NN) category to assess the predictions of the new instance is called the k-NN algorithm. The completion of this algorithm using the Euclidean Distance formula with dimensions greater than two is as below.

$$\text{Euclidean Distance} = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \quad (1)$$

B. Deep Learning

Deep learning as a supervised learning algorithm, if every activity function in a deep feed-forward neural network is non-linear, as well as functions in the output layer, a network that serves a simple non-linear model for some inputs. It is described as x input neurons/layers, y output neurons/layers, and hidden layers with z for each neuron see in **Figure 1**. Let f be a set of non-linear activation functions where f and some input $x_1, x_2, \ldots, x_n$ for some function, namely $f(\alpha)$. There is $x = [x_1, x_2, \ldots, x_n]$, the input value of the neuron multiplied by the weight w = [w_1, w_2, …w_n], so that is the non-linear activation function $f(\alpha)$.

$$\propto = \sum_{i=1}^{n} [W_i, X_i] + b \quad (2)$$

where:

b: bias represents the neuron's activation threshold

Bias units are included in each non-output layer of the network. The weights linking neurons and biases with other neurons fully determine the output of the entire network. Then, the weights of the network and deep learning of the feed-forward neural network are show in **Figure 1** and **Figure 2**.

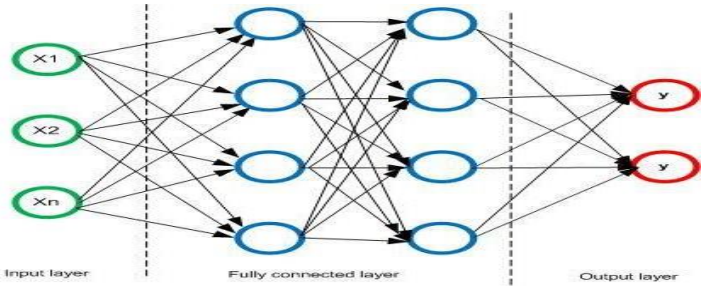

Figure 1. Deep Learning-Multi-layer of Feedforward Neural Network

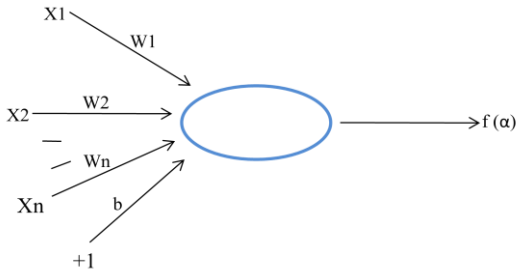

Figure 2. The Model Neuron

C. Support Vector Machine

The supervised machine learning model used for classification and regression is called a support vector machine. But more often used for classification. The goal is to determine the hyperplane (n-dimensional space), which is distinguished by positive and negative classes (binary value classification) called Margin Maximizing hyperplane. Some 2-dimensional, 3-dimensional, and N-dimensional planes of equations are as below.

The equation of 2-dimensional plane:

$y = mx + c$

where m: slope; c: y-intercept

The general form of this Equation 3:

$$W_0 + w_1.x_1 + w_2.x_2 = 0 \quad (3)$$

$x_2 = -w_0/w_2 - (w_1/w_2).x_1$

$c = -w_0/w_2$

$m = -w_1/w_2$

See **Figure 3**, The Plane of 2-dimensions, as below.

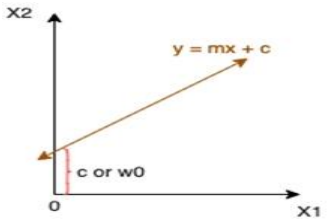

Figure 3. Plane of 2-dimensions

The Equation 4 of a 3-dimensional plane:

$$W_0 + w_1.x_1 + w_2.x_2 + w_3.x_3 = 0 \quad (4)$$

$w_0 + w_T.x = 0 \; or \; b + w_T.x = 0$

See **Figure 4**, The Plane of 3-dimensions, as below.

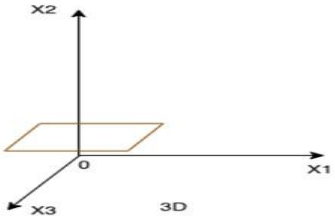

Figure 4. The Plane of 3-dimensions

The Equation 5 of hyperplane:

$$W_0 + w_1.x_1 + w_2.x_2 + w_3.x_3 + w_n.x_n = 0 \quad (5)$$

$w_0 + w_T.x = 0 \; or \; b + w_T.x = 0$

$w_T.x + b = 1$: *hyperplane of positive direction or* $\pi +$

$w_T.x + b = -1$: *hyperplane of negative direction or* $\pi -$

See **Figure 5**, Two parallel hyperplanes below.

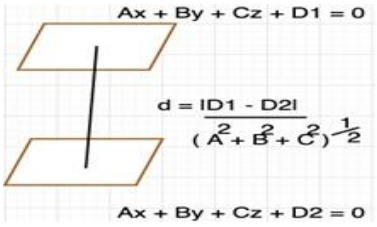

Figure 5. Two Parallel hyperplanes

Where $D1 = -1, D2 = 1, A = w_0, B = w_1, C = w_2$

So that:

$$d = \frac{|D1 - D2|}{\sqrt{w0^2 + w1^2 + \cdots + wn^2}} \quad (6)$$

$$d = \frac{2}{||\, w \,||} \quad (7)$$

Results and Discussion

This study used a regression learning model to determine RMSE by measuring the three algorithms, namely k-NN, Deep Learning, and SVM. The RMSE results of the three algorithms with the lowest value are the best. The

research stage using the knowledge discovery in the database process, as discussed in the previous sub-chapter, includes: selection activity by selecting the source of the Covid-19 data set from March 2020 to March 2021 for each province in Indonesia.

The dataset consists of five common attributes and one unique attribute and a total of 34 data with attribute descriptions as treated, get well, died, isolation, percentage of Covid-19 patients who died (CFR), and the percentage of Covid-19 patients who live as unique attributes/labels (RI). The data analysis is carried out, and there is all the data.

A. Dataset

The application used to "load the dataset" is Rapid miner. The step taken is to execute this application. After the application is active, the operator "read csv" is opened, then the file "Covid_19PropRegresi.csv" is opened using the Import Configuration Wizard. The opening of this dataset file results in a six-attribute dataset that shows in **Table 1**.

Table 1. Attribute Dataset Covid_19 PropRegresi

Row Now	RI	Hospitalized (*Dirawat*)	Get Well (*Sembuh*)	Died (*Meninggal*)	Isolation (*Isolasi*)	CFR
2	0.850	3012	20121	631	0	0.030
3	0.830	3561	21051	743	0	0.030
4	0.830	3317	3317	6\78	0	0.020
5	0.910	595	4052	136	0	0.030
6	0.660	258	12730	430	0	0.020
7	0. 890	6259	221567	4024	19554	0.020
8	0.740	4670	3139	67	0	0.020
9	0.820	1057	118090	3445	3331	0.020
10	0.830	18437	128350	8909	0	0.060
11	0.860	17950	91442	73381	3307	0.070
12	0.890	383	3344	38	0	0.010

B. Data Pre-processing

The first stage in the regression learning model is data pre-processing. This can be seen from the records in the dataset. These steps taken are the first, looking at the class distribution, showing that the average data distribution is uneven. The second notice if there are any missing data, indicating that there is no missing data, and then make a check for duplication of data. If there is a duplication of data, removing duplicates is used to remove duplicate data. There is no duplication. The statistical dataset of Covid-19 after pre-processing is in **Figure 6**. The covid-19 dataset is in the form of a CSV file shared with the multiply operator to the three algorithms.

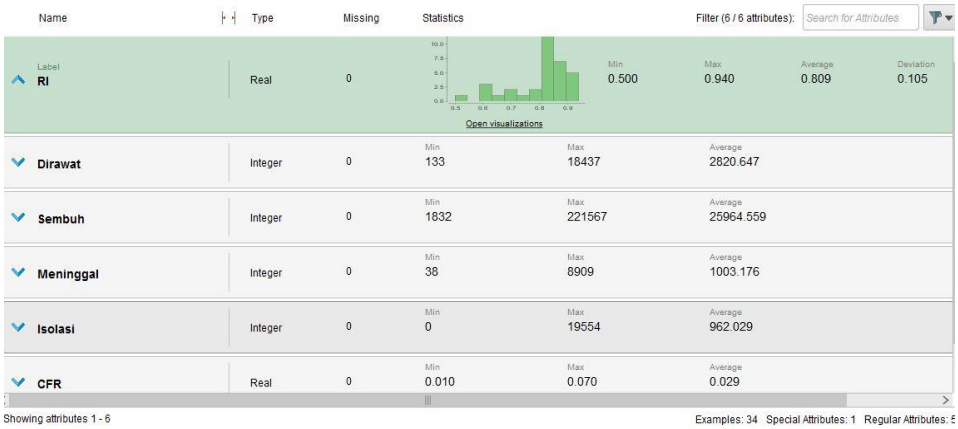

Figure 6. The Statical Dataset of Covid-19 After Pre-processing

C. Learning Algorithm

This study uses the k-NN algorithm, Deep Learning, and Support Vector Machine with the application of Cross-Validation where the optimal number of folds is obtained, namely three subprocesses. The dataset is partitioned into five subsets of equal size. First subset, a single subset is retained as the test data. The remaining 4 subsets are used as training data sets for the KNN.

Deep learning with rectifier activation and epoch is ten. The SVM with dot kernel type, C is 0, convergence epsilon is 0.001 with epsilon is zero where Lpos and Lnegative are one. The cross-validation process is then repeated five times, with each of the five subsets used exactly once as the test data. The five results from the five iterations are averaged or combined to produce a single estimation.

The activity of training, testing, and cross-validation in this research was conducted by Rapid Manner. The model of cross-validation, outputs, testing, and regression predictions model is shown in **Figure 8** and **Figure 9**. It

means that training, testing, and cross-validation were conducted for each method, including for k-NN algorithm, Support Vector Machine, and Deep Learning.

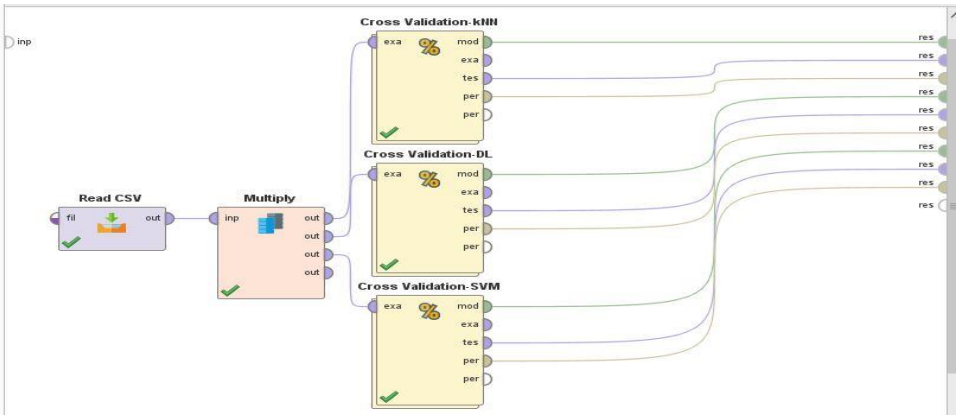

Figure 8. Cross Validation k-NN, Deep Learning, and SVM

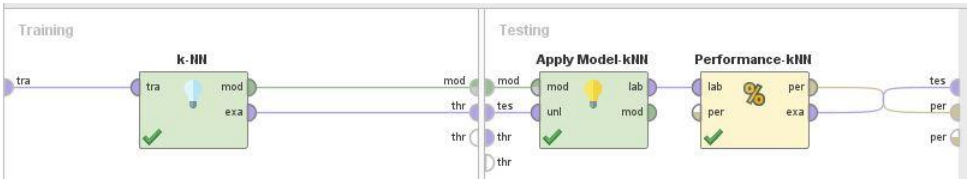

Figure 9. Training and Testing from Cross Validation k-NN

D. Learning Model Algorithm Performance Calculation Results

For the implementation of calculating the performance of the learning model algorithm, it is necessary to execute the models of the three algorithms. The results of calculating the performance of the three algorithms show that the RMSE value of k-NN is 0.101 +/- 0.023. It means the best value for regression prediction measurement is in **Table 2**.

The results in this study supported the other research about the comparison of machine learning algorithms that found out the results identified the performance of SVM and k-NN is different [29], and k-NN has better performance than SVM [30]. Therefore, the k-NN algorithm was chosen to measure the predictions of living covid-19 patients. The results in detail are show in **Table 3**. The other hand, the performance of k-NN in this research still can be improved using Z-score normalization method [31].

Table 2. The RMSE Measurement of Three Algorithms

Algorithms	RMSE	Description
k-NN	0.101 +/- 0.023	Best
Deep Learning	0.102 +/- 0.037	Good
SVM	0.176 +/- 0.087	Quite good

Table 3. Prediction of Patients Who Survive from March 2020 - March 2021 in each Province

Provinces	Get well	Died	Isolation	CFR	Prediction (RI)		RI	
					(Y)	(%)	(X)	(%)
Aceh	7753	374	0	0.040	0.833	83.3	0.850	85.0
Bali	221567	4024	19554	0.020	0.875	87.5	0.890	89.0
Banten	3139	67	0	0.020	0.879	87.9	0.740	74.0
Bangka Belitung	128350	8909	0	0.060	0.873	87.3	0.830	83.0
Bengkulu	91442	7381	3307	0.070	0.875	87.5	0.860	86.0
DI. Yogyakarta	29613	939	0	0.030	0.883	88.3	0.800	80.0
DKI Jakarta	15385	629	0	0.040	0.800	80.0	0.890	89.0
Jambi	4072	87	0	0.010	0.714	71.4	0.630	63.0
Jawa Barat	8584	400	0	0.030	0.810	81.0	0.690	69.0
Jawa Tengah	17480	726	0	0.040	0.815	81.5	0.860	86.0

Provinces	Get well	Died	Isolation	CFR	Prediction (RI)		RI	
					(Y)	(%)	(X)	(%)
Jawa Timur	9798	200	0	0.020	0.819	81.9	0.830	83.0
Kalimatan Barat	20121	631	0	0.030	0.863	86.3	0.850	85.0
Kalimatan Timur	21051	743	0	0.030	0.863	86.3	0.830	83.0
Kalimantan Tengah	3317	78	0	0.020	0.672	67.2	0.830	83.0
Kalimantan Selatan	4052	136	0	0.030	0.781	78.1	0.910	91.0
Kalimantan Utara	12730	430	0	0.020	0.813	81.3	0.660	66.0
Kepulauan Riau	118090	3445	3331	0.020	0.860	86.0	0.820	82.0
Nusa Tenggara Barat	3344	38	0	0.010	0.671	67.1	0.890	89.0
Sumatera Selatan	24182	583	1391	0.020	0.864	86.4	0.910	91.0
Sumatera Barat	4174	168	0	0.030	0.673	67.3	0.630	63.0
Sulawesi Utara	4981	89	0	0.010	0.791	79.1	0.840	84.0
Sumatera Utara	3917	110	0	0.030	0.781	78.1	0.940	94.0
Sulawesi Tenggara	9680	309	0	0.030	0.775	77.5	0.840	84.0
Sulawesi Selatan	6942	195	0	0.020	0.872	87.2	0.880	88.0
Sulawesi Tengah	5596	321	0	0.050	0.792	79.2	0.790	79.0
Lampung	11162	683	1275	0.050	0.788	78.8	0.820	82.0
Riau	7621	175	0	0.020	0.775	77.5	0.830	83.0
Maluku Utara	39661	711	1886	0.020	0.849	84.9	0.890	89.0
Maluku	6459	478	0	0.050	0.731	73.1	0.710	71.0
Papua Barat	26122	669	933	0.020	0.852	85.2	0.930	93.0
Papua	2634	94	0	0.030	0.859	85.9	0.840	84.0
Suluawesi Barat	5858	107	0	0.020	0.837	83.7	0.900	90.0
Nusa Tenggara Timur	1832	65	1032	0.020	0.861	86.1	0.590	59.0
Gorontalo	2086	114	0	0.030	0.745	74.5	0.500	50.0

Next, specify t_{uji} to compare with t_{tabel}. There is a significant effect or not between prediction (RI) to RI, where is prediction (RI) represented by Y_i and RI represented by X_i. The calculation results in values of t_{uji} are -0.72 and Standard Error Mean (SEM) values is 0.17969 with df = 33, and the value significant (with two tailed) is 0.943. The determination of the table is influenced by the significant level α. If the test is 2-tailed, the significance level is $\alpha = 0.05 / 2 = 0.025$ where $df = N - 1$, namely $34 - 1 = 33$. So $table = t(0.025, 33)$ is 2.035. Thus table> test is $2.035 > -0.72$, so the prediction (RI) for RI has no significant effect.

Conclusion

The research about regression model for prediction system concluded that the covid-19 dataset used for regression prediction to measure RMSE with the better learning model using the k-NN algorithm where the RMSE is k-NN 0.101 +/- 0.023 than SVM that has RMSE 0.176 +/-0.087. In this research, the k-NN algorithm was chosen to process the prediction of Covid-19 patient measurements from March 2020 to March 2021, and the results show the prediction of surviving patients from March 2020 to March 2021 for each province in Indonesian.

Acknowledgment

With the k-NN algorithm, prediction measurements can be made where a national event is taking place to provide government recommendations for further decision-making.

References

[1] Y. Zhang and J. Sun, "A COVID-19 Epidemics Trend Prediction Algorithm Based on LSTM," in *2021 IEEE 4th International Conference on Computer and Communication Engineering Technology (CCET)*, Beijing, China, 2021, doi: 10.1109/CCET52649.2021.9544257.

[2] A. Damone, A. Damone, A. Damone, F. Bonino, S. Nuti and G. Ciuti, "Decision-Making Algorithm and Predictive Model to Assess the Impact of Infectious Disease Epidemics on the Healthcare System: The COVID-19 Case Study in Italy," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3661 -

3672, August. 2022, doi: 10.1109/JBHI.2022.3174470.

[3] A. Hernita, M. A. Soeleman and A. Z. Fanani, "Optimization Of Infant Birth Predictions During The Covid-19 Pandemic Using The Particle Swarm Optimization Based K-Nn Algorithm Method," in *2022 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia, 2022, doi: 10.1109/iSemantic55962.2022.9920468.

[4] T. Mantoro, R. T. Handayanto, M. A. Ayu and J. Asian, "Prediction of COVID-19 Spreading Using Support Vector Regression and Susceptible Infectious Recovered Model," in *2020 6th International Conference on Computing Engineering and Design (ICCED)*, Sukabumi, Indonesia, 2020, doi: 10.1109/ICCED51276.2020.9415858.

[5] M. D. Darma, M. R. Faisal, I. Budiman, R. Herteno, J. P. Utami and B. Abapihi, "In Silico Prediction of Indonesian Herbs Compounds as Covid-19 Supportive Therapy using Support Vector Machine," in *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, Depok, Indonesia, 2021, doi: 10.1109/IC2IE53219.2021.9649383.

[6] A. Y. A. Saeed and A. E. B. Alawi, "Covid-19 Diagnosis Model Using Deep Learning with Focal Loss Technique," in *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, Taiz, Yemen, 2021, doi: 10.1109/ICOTEN52080.2021.9493477.

[7] N. Agarwal and R. Dutta, " Comparative Predictive Analysis of Mortality Rate after Covid-19 Vaccination Using Various Machine Learning Approaches," in *2022 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2022, doi: 10.1109/ICCCI54379.2022.9740854.

[8] T. Thongrod, A. Lim, T. Ingviya and B. A. Owusu, "Prediction of PM2.5 and PM10 in Chiang Mai Province: A Comparison of Machine Learning Models," in *2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Phuket, Thailand, 2022, doi: 10.1109/ITC-CSCC55581.2022.9894884.

[9] D. Alekseeva, N. Stepanov, A. Veprev, A. Sharapova, E. S. Lohan and A. Ometov, "Comparison of Machine Learning Techniques Applied to Traffic Prediction of Real Wireless Network," *IEEE Access* , vol. 9, pp. 159495 - 159514, November 2021, doi: 10.1109/ACCESS.2021.3129850.

[10] J. Xie and Q. Wang, "Benchmarking Machine Learning Algorithms on Blood Glucose Prediction for Type I Diabetes in Comparison With Classical Time-Series Models," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3101 - 3124, February 2020, doi: 10.1109/TBME.2020.2975959.

[11] H. H. Elmousalami, "Comparison of Artificial Intelligence Techniques for Project Conceptual Cost Prediction: A Case Study and Comparative Analysis," *IEEE Transactions on Engineering Management* , vol. 68, no. 1, pp. 183 - 196, February 2021, doi: 10.1109/TEM.2020.2972078.

[12] O. Eyecioglu, B. Hangun, K. Kayisli and M. Yesilbudak, "Performance Comparison of Different Machine Learning Algorithms on the Prediction of Wind Turbine Power Generation," in *2019 8th International Conference on Renewable Energy Research and Applications (ICRERA)*, Brasov, Romania, 2019, doi: 10.1109/ICRERA47325.2019.8996541.

[13] I. Ahmed, T. E. Chowdhury, B. B. Routh, N. Tasmiya, S. Sakib and A. A. Chowdhury, "Performance Analysis of Machine Learning Algorithms in Chronic Kidney Disease Prediction," in *2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, BC, Canada, 2022, doi: 10.1109/IEMCON56893.2022.9946591.

[14] S. Kaya and M. Yağanoğlu, "An Example of Performance Comparison of Supervised Machine Learning Algorithms Before and After PCA and LDA Application: Breast Cancer Detection," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Istanbul, Turkey, 2020, doi: 10.1109/ASYU50717.2020.9259883.

[15] P. Cıhan and H. Coşkun, "Performance Comparison of Machine Learning Models for Diabetes Predictio," in *2021 29th Signal Processing and Communications Applications Conference (SIU)*, Istanbul, Turkey, 2021, doi: 10.1109/SIU53274.2021.9477824.

[16] A. Sunarya, Henderi and I. Tasyriqan, "The Comparison Between Sequential Minimal Optimization and Multilayer Perceptron Neural Network Methods in Predicting the Commodity Prices," in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, Indonesia, 2019, doi: 10.1109/ICIC47613.2019.8985679.

[17] P. Subarkah, W. R. Damayanti and R. A. Permana, "Comparison of correlated algorithm accuracy Naive Bayes Classifier and Naive Bayes Classifier for heart failure classification," *ILKOM Jurnal Ilmiah*, vol. 14, no. 2, pp. 120-125, August 2022.

[18] P. Marquez, D. Pinos and I.-O. Juan, "Performance comparison in network traffic prediction for polynomial regression to P1P versus ARIMA and MWM," in *2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, Moscow and St. Petersburg, Russia, 2018, doi: 10.1109/EIConRus.2018.8317034.

[19] E. H. Yossy, Y. Heryadi and Lukas, "Comparison of Data Mining Classification Algorithms for Student Performance," in *2019 IEEE International Conference on Engineering, Technology and Education (TALE)*, Yogyakarta, Indonesia, 2019, doi: 10.1109/TALE48000.2019.9225887.

[20] S. A. Septianingrum, M. A. Dzikri, M. A. Soeleman, P. Pujiono and M. Muslih, "Performance Analysis of Multiple Linear Regression and Random Forest for an Estimate of the Price of a House," in *2022 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia, 2022, doi: 10.1109/iSemantic55962.2022.9920454.

[21] G. R. Thummala, R. Baskar and N. Thiyaneswaran, "Prediction of Heart Disease Using Naive Bayes in Comparison with KNN Based on Accuracy," in *2022 International Conference on Cyber Resilience (ICCR)*, Dubai, United Arab Emirates, 2022, doi: 10.1109/ICCR56254.2022.9995841.

[22] Y. Niu, X. Lu, X. Li, W. Su, Z. Meng and S. Zhang, "Modeling and Analysis of Runway Friction Coefficient Prediction Methods Based on Multivariable Coupling," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022, doi: 10.1109/TIM.2022.3148760.

[23] B. C. M and N. Deepa, "Improving Performance Analysis in Classification with Accuracy of Adult Income Salary using Novel Gated Residual Neural Network in Comparison with Logistic Regression Algorithm," in *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, Chennai, India, 2022, doi: 10.1109/ICSES55317.2022.9914103.

[24] P.-Y. Hsu and C. Holtz, "A Comparison of Machine Learning Tools for Early Prediction of Sepsis from ICU Data," in *2019 Computing in Cardiology (CinC)*, Singapore, 2019, doi: 10.22489/CinC.2019.206.

[25] M. Chellamani, S. Murugesan and N. Raju, "Heart disease prediction using Boosting Algorithms: Performance Analysis and Comparison," in *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*, Mysuru, India, 2022, doi: 10.1109/MysuruCon55714.2022.9972536.

[26] Y. Hong and F. Lv, "Data Mining and Knowledge Discovery in Computer Aided Medical Diagnosis System," in *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, India, 2021, doi: 10.1109/I-SMAC52330.2021.9640713.

[27] S. Khlamov, V. Savanevych, O. Briukhovetskyi, I. Tabakova and T. Trunova, "Astronomical Knowledge Discovery in Databases by the CoLiTec Software," in *2022 12th International Conference on Advanced Computer Information Technologies (ACIT)*, Ruzomberok, Slovakia, 2022, doi: 10.1109/ACIT54803.2022.9913188.

[28] V. Gancheva and P. Borovska, "SOA Based System for Big Genomic Data Analytics and Knowledge Discovery," in *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Metz, France, 2019, doi: 10.1109/IDAACS.2019.8924370.

[29] N. Farhat, R. Fatima, S. Kazi and H. Naaz, "A Comparison of Machine Learning Algorithms for Emotional Intelligence Measure," in *2022 Fifth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU)*, Riyadh, Saudi Arabia, 2022, doi: 10.1109/WiDS-PSU54548.2022.00036.

[30] M. F. Suleiman and B. Issac, "Performance Comparison of Intrusion Detection Machine Learning Classifiers on Benchmark and New Datasets," in *2018 28th International Conference on Computer Theory and Applications (ICCTA)*, Alexandria, Egypt, 2018, doi: 10.1109/ICCTA45985.2018.9499140.

[31] Henderi, T. Wahyuningsih and E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *International Journal of Informatics and Information System,* vol. 4, no. 1, pp. 13-20, 2021.