Jurnal CoreIT, Vol.9, No.1, June 2023
ISSN 2460-738X (Print)
ISSN 2599-3321 (Online)

Python Model Predicts Covid-19 Cases since Omicron in Indonesia

Muhammad Furqan Rasyid[1*]
[1] Department of Informatics Management, Dipa Makassar University, Indonesia
[1*] muhammad.furqan@undipa.ac.id

Abstract. The proposed work uses Support Vector Regression model to predict the new cases, recovered cases, and deaths cases of covid-19 every day during sub-variant omicron spread in Indonesia. We collected data from June 14, 2022, to August 12, 2022 (60 Days). This model was developed in Python 3.6.6 to get the predictive value of the issues mentioned above up to September 21, 2022. The proposed methodology uses a SVR model with the Radial Basis Function as the kernel and a 10% confidence interval for curve fitting. The data collected has been divided into 2 with a size of 40% test data and 60% training data. Mean Squared Error, Root Mean Squared Error, Regression score, and percentage accuracy calculated the model performance parameters. This model has an accuracy above 87% in predicting new cases and recovered patients and 68% in predicting daily death cases. The results show a Gaussian decrease in the number of cases, and it could take another 4 to 6 weeks for it to drop to the minimum level as the origin of the undiscovered omicron sub-variant. RBF (Radial Basis Function) very efficient and has higher accuracy than linear or polynomial regression as kernel of SVR.

Keywords: SVR (Support Vector Regression), Python 3.6.6, SVR Kernel

Received Aug 2022 / **Revised** Jan 2023 / **Accepted** May 2023

INTRODUCTION

Covid-19, or coronavirus disease 2019, is a disease that attacks the human lungs. At the end of 2019, this virus was first discovered in Wuhan city, China [1]. Due to its rapid spread, in early 2020, WHO (World Health Organization) declared the status of this virus to be a pandemic. The COVID-19 pandemic has become a significant global crisis since World War II [2]. This condition significantly affects human life and the world economy. People are isolated and have to reduce their mobility. Research suggests that being isolated due to the COVID-19 pandemic will increase the risk of adverse mental health [3].

The first case of Covid-19 in Indonesia in March 2020. Which initially infected four people and continued to increase rapidly until 2021. The number of cases slowly decreased when the vaccine began to be implemented, from hundreds of thousands decreasing drastically to tens of people per day. In the middle of 2022, there was an increase in the number of new cases of Covid-19. An increase has occurred since the discovery of sub-variant omicron, namely BA.4 and BA.5. The increasing number of cases has made many people curious about the development of COVID-19. Some of the most curious questions: (i) when will the infection rate due to the omicron sub-variant reach it is maximum, (ii) how long will it take for this sub-variant to stop spreading, (iii) how many individuals are infected; (iv) what is the daily increase in deaths [4].

Previous studies regarding predictions using the SVR method, such as [5], used the SVR method to predict landslide disaster. Research by [6], found that the combined WOA-SVR method was better than other machine learning models for house price prediction. Research by [7], applied an improved SVR method to a short-term hybrid prediction model of an ice storage air conditioner. Research by [8], predicts the increase in Covid-19 cases in India using the SVR method. Research by [9], evaluated the LR and SVR models to predict the COVID-19 pandemic. Research by [10] used the SVR kernel to determine the duration of rehabilitation for COVID-19 patients. All of these studies make predictions for different cases. All of these studies make predictions for other issues. However, there has been no System for predicting covid-19 cases in Indonesia since the appearance of Sub-Variant Omicron. The prediction system will help answer community questions. Meanwhile, the prediction system also helps the government make public policy decisions in health [11] and economics [12] to counteract other effects triggered by the indirect spread of

DOI: 10.24014/coreit.v9i1.18908

the virus, such as psychological stress, economic losses and economic losses, and negative impacts on daily activities[13].

This paper aims to build a prediction model for COVID-19 cases in Indonesia since the assault of Sub-Variant Omicron using the Support Vector Regression algorithm implemented in python 3.6.6. Method Section will discuss the modeling and analysis steps. Result And Discussion Section will discuss the results of the research. The author concludes the overall objective of the work in the Conclusion Section.

METHODS

In this work, we use the python programming language to run all the methods used. Below we will discuss the explanation of the method used.

A. Dataset of Sub-Variant Omicron

A dataset is a collection of data based on past information and managed into new knowledge. Computational techniques, such as machine learning datasets, have previously affected predictive ability [14]. Sometimes, the number of datasets will affect the results of future predictions [15].

Since it spread, COVID-19 has mutated and changed shape. Some of the main variants of this virus such as Alpha, Delta, and Omicron. One example of these variants mutated again is the Sub-variant of Omicron, namely BA.4 and BA.5. That is why this sub-variant is nicknamed “son of omicron” or “Stealth Omicron [16].

The dataset is collect from the social media of BNPB (National Disaster Management Agency). This department provides information on the development of Covid-19 cases every day. We collected data in the form of the number of new cases, patients recovered, and patients are dying every day. This data was collected from 14 June to 12 August 2022 (60 days) because the Omicron sub-variant first appeared in Indonesia on 13 June 2022.

B. Data Processing

Data processing is a method for converting or translating raw data into usable information. When the data is natural, it is useless, so it must be processed to become useful information. In recent years, several promising approaches to data processing have been developed [17].

In the data processing stage, we create two columns: column X and column y. Column X shows the date of data collection. The total is 60 data, from June 14 to August 12, 2022. Column y shows the number of daily cases. Next, Column X and y are combined back into vector 60. Grouping datasets into training (60%) and test (40%) using the train_test_split(). We use this allocation to make the test data size as manageable as with datasets. Function imported from the class model selection from Python's sklearn package.

C. SVR

SVR (Super Vector Regression) [18] is a popular algorithm for predicting linear and non-Linear regression types. The SVR algorithm is an element of the Support Vector Machine (SVM) algorithm, where the support vector is the point closer to the hyperplane. The SVR concept maximizes the hyperplane to get support for vector data [19]. SVR is different from Principal Component Analysis (PCA) which only extracts classical features and many data representation techniques for use in pattern recognition [20]. Find more discussion on SVR and SVM algorithm on [21].

D. Visualization

The representation of data or information through designs or visuals is known as data visualization [22]. This paper uses python's matplotlib library to do just that. The most widely used Python data visualization library is matplotlib. This library was created and developed by John Hunter together with several contributors [23]. Using the scatter function to plot the data regression that matches the predicted data value. Figure 1 shows the actual and predicted points of this job, respectively.

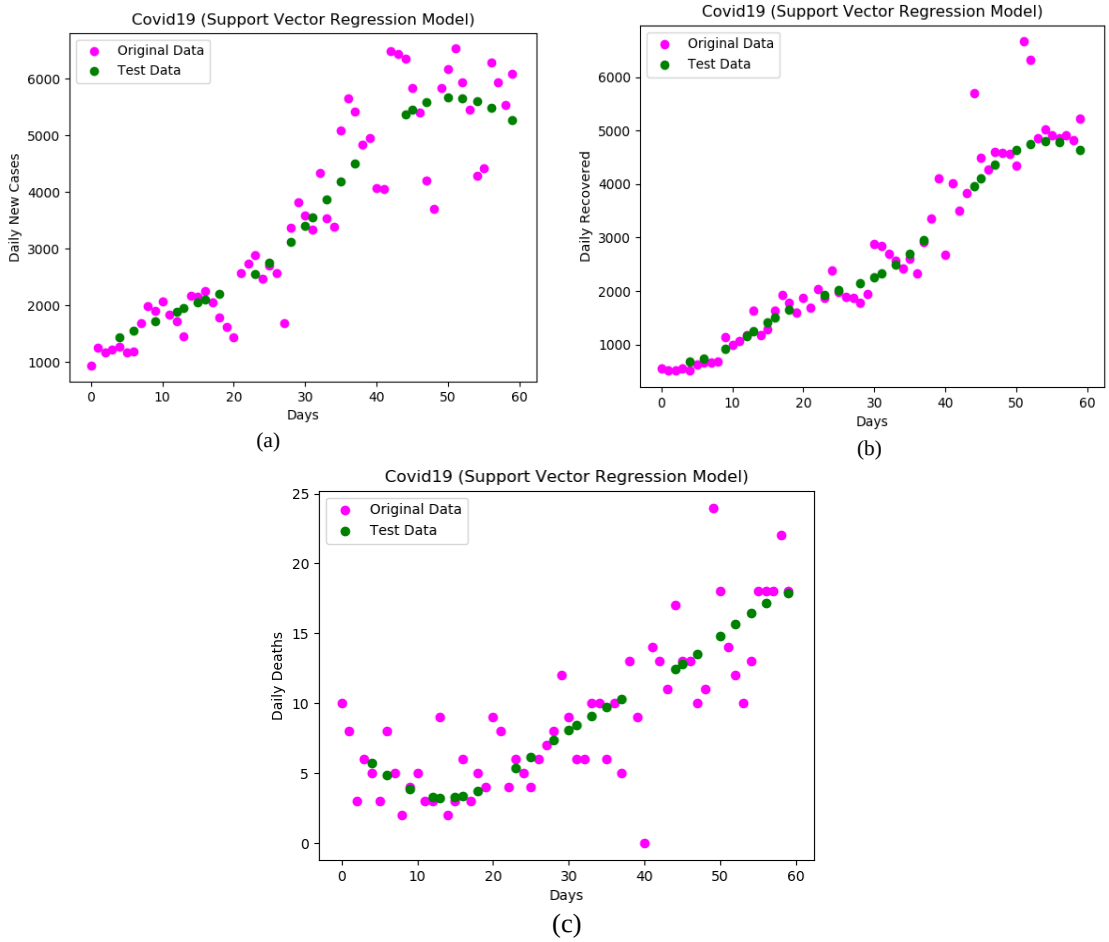

Figure 1. The regression plot that corresponds to the data on the increase (A) of daily new cases; (B) patients who recovered; (C) and daily deaths.

E. Evaluation of model performance

The performance parameters of the models assess whether they are reliable in estimating the results. This paper uses MSE, RMSE, R2 value, and accuracy percentage. Table 1 shows all the calculations.

Table 1. Parameters of support vector regression model with RBF kernel and 10% EPSILON

Data	MSE	RMSE	Reg. score	%Accuracy
Daily New Cases	0.115577	0.339966	0.880081	88%
Daily Recovered	0.114426	0.3382702	0.893172	89%
Daily Deaths	0.235084	0.48485	0.68297	68%

F. Prediction

Predicting the future value of a time series involves several steps of manipulating the data to obtain a cumulative trend that matches the past trend of the original data set. The previous data set was in cumulative form, but since we applied the RBF multiplier in our model, it is clear that the predicted time series follows a Gaussian trend. The downtrend can be maintained with a transition, as discussed below. We have implemented several steps in the algorithm that can help us achieve our goal.

Here, we obtain time series projections for each case for 60 days from June 14 or 61 from the start date. Therefore, we want to combine the last 60 days' forecast with the last 61 days. The prediction column consists of descending values. So we calculate the time series difference and then use the absolute value of the time series difference. The time series divergence will reverse and give us an uptrend, which saturates after a certain value. We then perform the cumulative sum of the time series elements and add the maximum elapsed time series values to them. It helps us stay on top of trends and visualize them in a cumulative form. Figure 2 shows the graph of the past value and the predicted value.

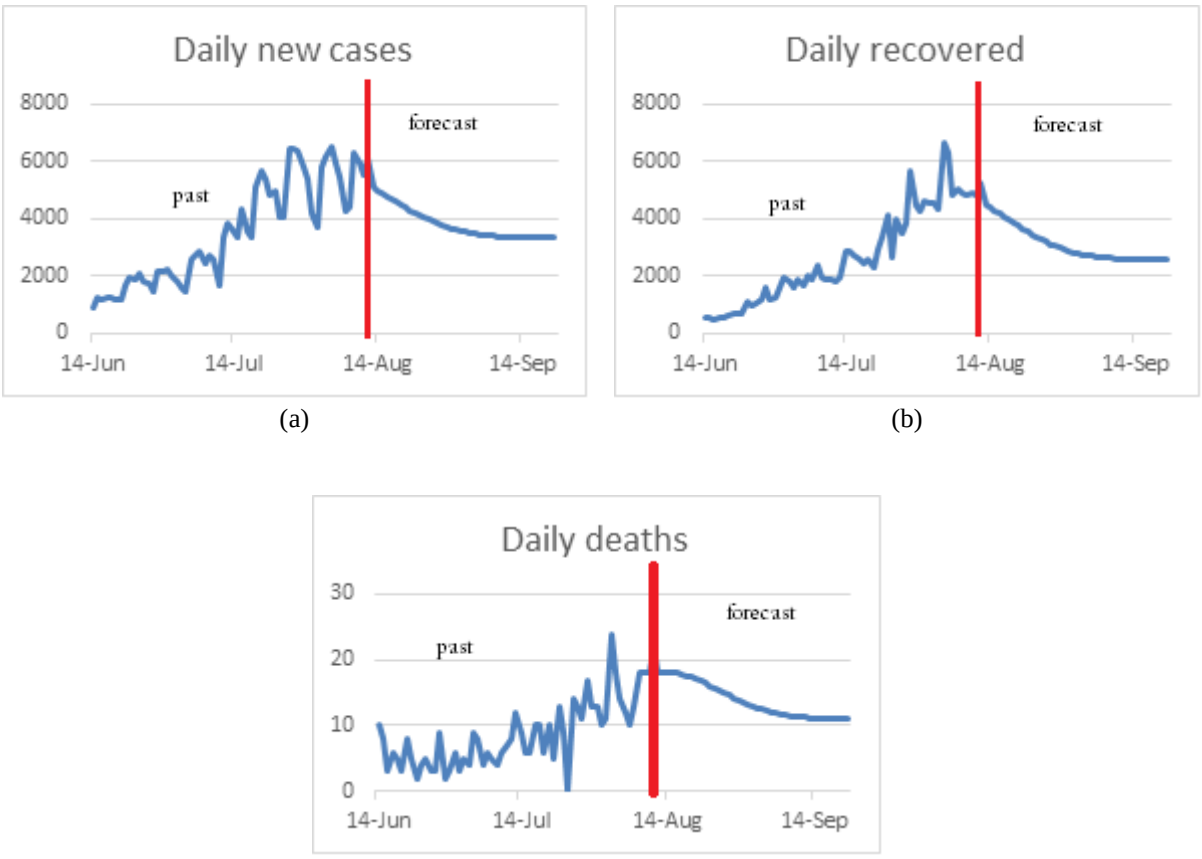

Figure 2. (A) Daily new cases; (B) Daily recovered patient; (C) Daily case of death

RESULT AND DISCUSSION

The results show that the model performs well in predicting recovered cases. Meanwhile, poor fitting occurs in the case of daily deaths. It happens because the case of daily mortality shows extreme changes every day (sometimes increasing or decreasing excessively), which reduces the accuracy of the model's predictability. However, the change interval for daily mortality data is only between1-24. The model predicts that the overall calculation will decrease for the next 40 days. If we look at the graph shown in section 2 (prediction), it shows that the whole will experience a decline over the next 40 days (21 September 2022). The predicted results for his 3 are 3337, 2579, and 11 (respectively).

We also compare the accuracy of three SVR kernels, namely RBF, Poly, and Linear. Table 2 shows the results of the comparison.

Table 2. SVR kernels accuracy

Kernel	New cases Accuracy	Recovery Accuracy	Deaths Accuracy
RBF	88%	89%	68%
Linear	83%	87%	64%
Poly	54%	64%	56%

CONCLUSION

The proposed methodology predicts the total daily cases of COVID-19 in Indonesia since the omicron subvariant was detected. Predictions are made to determine the number of new cases, deaths, and patients recovering daily. Based on the latest trends, sophisticated machine learning models that support vector regression are used to predict the future. The proposed methodology manages the variation in the dataset. Although using a small number of datasets, the proposed model obtains a fairly high accuracy. The proposed model has an accuracy of 88% for predicting new cases, 89% for predicting recovered patients, and 68% for daily death cases. The spread of this virus is very high, and if we take proper precautions with

physical and sanitary conditions, we can reduce the spike in cases and, thus, the rate of its development. The proposed method can help reduce the spread of omicron subvariant.

REFERENCES

[1] N. Zhu *et al.*, "A Novel Coronavirus from Patients with Pneumonia in China, 2019," *N Engl J Med*, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.

[2] S. Boccaletti, W. Ditto, G. Mindlin, and A. Atangana, "Modeling and forecasting of epidemic spreading: The case of Covid-19 and beyond," *Chaos Solitons Fractals*, vol. 135, p. 109794, Jun. 2020, doi: 10.1016/j.chaos.2020.109794.

[3] C. McArthur *et al.*, "Evaluating the Effect of COVID-19 Pandemic Lockdown on Long-Term Care Residents' Mental Health: A Data-Driven Approach in New Brunswick," *Journal of the American Medical Directors Association*, vol. 22, no. 1, pp. 187–192, Jan. 2021, doi: 10.1016/j.jamda.2020.10.028.

[4] L. Li *et al.*, "Propagation analysis and prediction of the COVID-19," *Infectious Disease Modelling*, vol. 5, pp. 282–292, Jan. 2020, doi: 10.1016/j.idm.2020.03.002.

[5] A.-L. Balogun *et al.*, "Spatial prediction of landslide susceptibility in western Serbia using hybrid support vector regression (SVR) with GWO, BAT and COA algorithms," *Geoscience Frontiers*, vol. 12, no. 3, p. 101104, May 2021, doi: 10.1016/j.gsf.2020.10.009.

[6] X. Wang, S. Gao, S. Zhou, Y. Guo, Y. Duan, and D. Wu, "Prediction of House Price Index Based on Bagging Integrated WOA-SVR Model," *Mathematical Problems in Engineering*, vol. 2021, p. e3744320, Oct. 2021, doi: 10.1155/2021/3744320.

[7] R. Cheng, J. Yu, M. Zhang, C. Feng, and W. Zhang, "Short-term hybrid forecasting model of ice storage air-conditioning based on improved SVR," *Journal of Building Engineering*, vol. 50, p. 104194, Jun. 2022, doi: 10.1016/j.jobe.2022.104194.

[8] D. Parbat and M. Chakraborty, "A python based support vector regression model for prediction of COVID19 cases in India," *Chaos Solitons Fractals*, vol. 138, p. 109942, Sep. 2020, doi: 10.1016/j.chaos.2020.109942.

[9] N. P. Dharani, P. Bojja, and P. Raja Kumari, "Evaluation of performance of an LR and SVR models to predict COVID-19 pandemic," *Materials Today: Proceedings*, Feb. 2021, doi: 10.1016/j.matpr.2021.02.166.

[10] E. Setti *et al.*, "Predicting post COVID-19 rehabilitation duration with linear kernel SVR," in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, Jul. 2021, pp. 1–5. doi: 10.1109/BHI50953.2021.9508602.

[11] P. R. Warde *et al.*, "Linking prediction models to government ordinances to support hospital operations during the COVID-19 pandemic," *BMJ Health Care Inform*, vol. 28, no. 1, p. e100248, May 2021, doi: 10.1136/bmjhci-2020-100248.

[12] O. Torrealba-Rodriguez, R. A. Conde-Gutiérrez, and A. L. Hernández-Javier, "Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models," *Chaos, Solitons & Fractals*, vol. 138, p. 109946, Sep. 2020, doi: 10.1016/j.chaos.2020.109946.

[13] M. E. El Zowalaty and J. D. Järhult, "From SARS to COVID-19: A previously unknown SARS-related coronavirus (SARS-CoV-2) of pandemic potential infecting humans – Call for a One Health approach," *One Health*, vol. 9, p. 100124, Jun. 2020, doi: 10.1016/j.onehlt.2020.100124.

[14] C. Gorges, K. Öztürk, and R. Liebich, "Impact detection using a machine learning approach and experimental road roughness classification," *Mechanical Systems and Signal Processing*, vol. 117, pp. 738–756, Feb. 2019, doi: 10.1016/j.ymssp.2018.07.043.

[15] L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed, "Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset," *SN COMPUT. SCI.*, vol. 2, no. 1, p. 11, Nov. 2020, doi: 10.1007/s42979-020-00394-7.

[16] "Ini 2 Gejala Utama Sub Varian BA.2 alias 'Son of Omicron.'" https://health.detik.com/berita-detikhealth/d-5938598/ini-2-gejala-utama-sub-varian-ba2-alias-son-of-omicron (accessed Aug. 24, 2022).

[17] M. Rasyid, Z. Zainuddin, and A. Andani, "Early Detection of Health Kindergarten Student at School Using Image Processing Technology," in *Proceedings of the 1st International Conference on Science and Technology, ICOST 2019, 2-3 May, Makassar, Indonesia*, Makassar, Indonesia, 2019. doi: 10.4108/eai.2-5-2019.2284609.

[18] M. F. Rasyid, D. Imran, and A. A. Mahersatillah, "Prediksi penyebaran Sub Varian omicron di Indonesia menggunakan Machine Learning," *SISITI : Seminar Ilmiah Sistem Informasi dan*

Teknologi Informasi, vol. 11, no. 1, Art. no. 1, Aug. 2022, Accessed: Jan. 11, 2023. [Online]. Available: https://www.ejurnal.dipanegara.ac.id/index.php/sisiti/article/view/936

[19] M. F. Rohmah, I. K. G. D. Putra, R. S. Hartati, and L. Ardiantoro, "Comparison Four Kernels of SVR to Predict Consumer Price Index," *J. Phys.: Conf. Ser.*, vol. 1737, no. 1, p. 012018, Jan. 2021, doi: 10.1088/1742-6596/1737/1/012018.

[20] M. F. Rasyid, "Comparison Of LBPH, Fisherface, and PCA For Facial Expression Recognition of Kindergarten Student," *International Journal Education and Computer Studies (IJECS)*, vol. 2, no. 1, Art. no. 1, May 2022, doi: 10.35870/ijecs.v2i1.625.

[21] "Support Vector Regression (SVR) using linear and non-linear kernels," *scikit-learn*. https://scikit-learn/stable/auto_examples/svm/plot_svm_regression.html (accessed Aug. 24, 2022).

[22] Ali Hassan Sial, Syed Yahya Shah Rashdi, and Dr. Abdul Hafeez Khan, "Comparative Analysis of Data Visualization Libraries Matplotlib and Seaborn in Python," *IJATCSE*, vol. 10, no. 1, pp. 277–281, Feb. 2021, doi: 10.30534/ijatcse/2021/391012021.

[23] R. O. Odegua and F. O. Ikpotokin, "DataSist: A Python-based library for easy data analysis, visualization and modeling," p. 17.