

## Prediksi penyebaran Sub Varian omicron di Indonesia menggunakan Machine Learning

Muhammad Furqan Rasyid<sup>\*1</sup>, Imran Djafar<sup>2</sup>, Andi Asvin Mahersatillah<sup>3</sup>

<sup>1,3</sup>Manajemen Informatika, Universitas Dipa Makassar;

<sup>2</sup>Teknik Informatika, Universitas Dipa Makassar

e-mail: <sup>\*1</sup>muhammad.furqan@undipa.ac.id, <sup>2</sup>imrandjafar@undipa.ac.id,

<sup>3</sup>andiasvin@undipa.ac.id

### Abstrak

Belakangan ini, kembali terjadi penambahan kasus baru terkonfirmasi positif covid-19 yang disebabkan oleh sub varian omicron BA.4 dan BA.5. Meskipun masih baru, namun tingkat penyebaran sub varian ini sangat cepat, menyebabkan analisis perkembangannya sulit diketahui. Penelitian ini bertujuan membuat sistem prediksi kasus baru sub varian omicron. Terdapat tantangan dipenelitian ini, yaitu keterbatasan dataset. Data mengenai penambahan kasus baru setiap harinya masih sedikit. Dataset hanya berjumlah 17, data dari tanggal 14 sampai 30 juni 2022. Semua data yang digunakan diperoleh dari media sosial BNPB (Badan Nasional Penanggulangan Bencana). Sebuah adaptasi dari machine learning yaitu algoritma SVR (Super Vector Regression) digunakan sebagai metode perhitungan prediksi. Data yang telah ada dikelompokkan menjadi dua, 14 sampai 27 juni sebagai data latih dan 28 sampai 30 juni sebagai data uji. Berdasarkan Pengujian yang telah dilakukan menggunakan analisis regresi, di peroleh akurasi sebesar 67.4 %. Akurasi yang diperoleh menunjukkan bahwa masih perlu di lakukan perbaikan dan pengembangan untuk memperoleh akurasi yang lebih baik.

**Kata kunci**—BA.4 dan BA.5, Machine Learning, algoritma SVR, analisis regresi.

### Abstract

Recently, there has been another increase in the number of new confirmed cases of COVID-19 caused by the Omicron BA.4 and BA.5 sub-variants. Even though it is still new, the rate of spread of this sub-variant is very fastly, making it difficult to analyze its development. This study aims to making a prediction system for new cases of sub-variant omicron. There are challenges in this research, the limitations of the dataset. Data regarding increase in new cases every day is still small. There are only 17 datasets, data from June 14 to 30, 2022. All data used were obtained from the social media of BNPB (National Disaster Management Agency). An adaptation of machine learning, namely the SVR (Super Vector Regression) algorithm is used as a predictive calculation method. The existing data were grouped into two, June 14 to 27 as training data and June 28 to 30 as test data. Based on the tests that have been carried out using regression analysis, an accuracy of 67.4% was obtained. The accuracy shows that improvements and developments still need to be made to better accuracy.

**Keywords**—BA.4 and BA.5, Machine Learning, SVR algorithm, regression analysis.

### 1. Pendahuluan

Virus covid-19 telah melanda negara kita tercinta sudah lebih dari dua tahun. Pertama kali menjangkit masyarakat Indonesia pada 2 maret 2020 [1]. Virus ini meluas cukup besar dalam kurung waktu yang singkat. Membuat pemerintah menjadikan situasi ini sebagai pandemi. Pandemi sangat berdampak signifikan terhadap kehidupan masyarakat. Hampir semua kegiatan yang dilakukan sehari-hari berubah semenjak keadaan ini berlangsung. Jumlah terkonfirmasi positif kasus covid-19 sempat mengalami penurunan, namun baru-baru ini terjadi peningkatan yang cukup signifikan. Hal ini disebabkan oleh adanya sub-varian omicron yang diberi kode BA.4 dan BA.5. Virus corona ini terus menerus mengalami perubahan seperti mutasi. Terkadang mutasi ini menciptakan subvarian baru. Varian ini terdeteksi pertama kali pada tahun 2022 dari specimen yang dikumpulkan dinegara Afrika Selatan. Begitu cepat tingkat penyebarannya, hingga saat itu pula langsung menyebar keseluruh provinsi di afrika selatan. Hal ini memberikan bukti bahwa versi yang terbaru mungkin lebih menular dan lebih mungkin menyebabkan infeksi terobosan daripada varian Omicron yang ada sebelumnya [2].

Kasus sub varian omicron pertama kali ditemukan diIndonesia pada 13 juni 2022. Varian virus tersebut menjangkit 4 orang, Tiga orang tidak menunjukkan gejala dan gejala ringan seperti badan pegal

serta sakit tenggorokan di satu orang. Penyebarannya sampai sekarang terus bertambah seiring berjalannya waktu. Meskipun belum ditemukan indikasi dari kedua sub varian omicron ini yang menyebabkan kesakitan lebih parah dibanding varian omicron yang lain, kita di himbau agar tetap waspada. Penambahan kasus baru yang tidak mampu diprediksi jumlahnya menyebabkan tenaga medis, penyediaan layanan kesehatan, dan fasilitas tidak mampu diperkirakan. Hal tersebut berakibat penanganan sedikit berjalan lambat

Beberapa penelitian mengenai sistem prediksi. Penelitian yang dilakukan oleh Debanjan Parbat dan Monisha Chakraborty [3], penelitian ini menerapkan metode SVR (Super Vector Regression) dengan Fungsi dasar radial sebagai kernel dan confidence interval 10% fitting kurva untuk memprediksi jumlah total kematian, orang yang sembuh, jumlah kasus terkonfirmasi harian. Data yang dikumpulkan mulai dari tanggal 1 Maret 2020 hingga 30 April 2020 (61 hari). Data telah dibagi menjadi dataset dan data uji dengan ukuran data latih 60% dan data uji 40%. Parameter kinerja model dihitung sebagai mean square error, root mean square error, regression score dan persentase akurasi. Model ini memiliki akurasi diatas 97% untuk kasus sembuh, kematian, dan jumlah kumulatif kasus terkonfirmasi. Sedangkan untuk pertambahan kasus harian memperoleh akurasi sebesar 87%. SVR ini sangat efisien dan memiliki akurasi yang lebih tinggi daripada regresi linier atau polinomial. Seperti yang diketahui, penelitian ini hanya dilakukan di India.

Penelitian lainnya yang dilakukan oleh Smita Rath, Alakananda Tripathy, Alok Ranjan Tripathy [4], analisis statistik harian kasus baru, kasus aktif dan kasus meninggal akibat virus covid-19 di negara yang sama yaitu India namun dengan menggunakan salah satu algoritma machine learning yaitu model MLR (multiple linear regression). Sebelumnya menggunakan metode MLR, mereka membandingkan antar regresi linier dan model regresi linier berganda (MLR) di mana skor model  $R^2$  (Regression Score  $^2$ ) cenderung memperoleh 0.99 dan 1.0 hal ini menunjukkan model prediksi yang kuat untuk memprediksi kasus aktif berikutnya. Penelitian ini menyimpulkan bahwa metode MLR memperoleh akurasi yang sangat baik dalam memprediksi kasus COVID-19 di hari berikutnya. Menentukan hubungan antara dependen (aktif) dengan variabel independen (terkonfirmasi positif, pasien covid-19 yang meninggal dunia dan pasien pulih dari virus ini) merupakan faktor kolerasi yang kuat. Penelitian ini mengumpulkan data dari 22 maret 2020 hingga 4 juli 2020 (105 hari) untuk memprediksi penyebaran kasus virus covid-19.

Memprediksi jumlah kasus baru terkonfirmasi sub varian omicron ini memiliki tantangan tersendiri. Penyebaran Sub varian omicron masih baru di Indonesia, jumlah data yang tersedia pun masih kurang dan terbatas. Meskipun kedua penelitian sebelumnya memperoleh akurasi yang baik, namun metode SVR menggunakan dataset yang paling sedikit. Dengan menggunakan metode tersebut, penelitian ini bertujuan untuk mengetahui akurasi yang diperoleh untuk memprediksi penyebaran sub-varian omicron, BA.4 dan BA.5. Sebagai perhitungan akurasi, kami menggunakan analisis regresi. Di akhir bagian, kami membahas kesimpulan mengenai penggunaan metode untuk perhitungan prediksi sub varian omicron di Indonesia.

## 2. Metode Penelitian

### 2.1 *Pandemic Covid-19 di Indonesia*

Covid-19 adalah penyakit yang disebabkan oleh virus SARS-COV-2 atau virus corona. Virus ini dapat menginfeksi hewan seperti anjing dan kucing. Selain itu, virus ini dapat menular ke sesama manusia [5]. Kasus pertama di dunia untuk virus covid-19 ditemukan di kota Wuhan, China sekitar akhir tahun 2019. Di awal tahun 2020 WHO (World Health Organization) mengategorikan virus ini sebagai pandemi. Pandemi covid-19 merupakan sebuah situasi dimana virus ini menyebar dengan cepat. WHO menegaskan bahwa pandemi ini bukanlah sebuah istilah sembarang yang bisa digunakan begitu saja. Status pandemi dikeluarkan melihat kondisi virus covid-19 yang mewabah ke seluruh negara.

Di Indonesia, Presiden mengeluarkan kepres 11/2020 tentang status pandemi. Di waktu itu, status pandemi ditetapkan sebagai kedaruratan kesehatan masyarakat. Berbagai istilah berbeda dalam penanganan covid-19 di Indonesia, mulai dari PSBB (Pembatasan Sosial Berskala Besar) hingga PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat).

### 2.2 *Sub Varian Omicron*

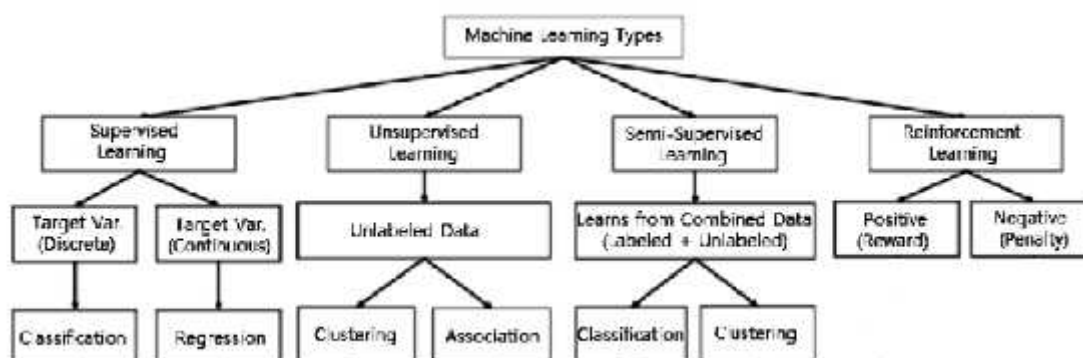
Sejak pertama kali muncul, virus covid-19 telah bermutasi dan berubah bentuk. Sudah ada beberapa varian utama dari virus ini seperti Alpha, Delta dan omicron. Varian-varian ini telah menciptakan gelombang besar infeksi. Varian-varian tadi bermutasi, menghasilkan sub-sub varian baru. BA.4 dan BA.5 merupakan sub varian dari omicron. Itulah mengapa sub varian ini diberi julukan sebagai

“son of omicron” atau “Stealth Omicron [6]. Subvarian ini pertama kali ditemukan di benua Asia dan Afrika. Kedua subspecies virus corona ini ditambahkan ke daftar pengawasan organisasi kesehatan dunia, WHO pada Maret 2022. Di Eropa, Sub spesies ini menjadi perhatian khusus untuk penyebaran dan pencegahan. Gejala dan penanganan sub varian Omicron BA.4 dan BA.5 sebenarnya sama dengan varian virus COVID-19 lainnya.

Di Indonesia, sub varian Omicron menyebar begitu cepat, menyebabkan lonjakan baru covid-19[7]. Meskipun tergolong rendah bila dibandingkan dengan negara lain, pemerintah tetap mengeluarkan himbauan agar masrakat tetap waspada terkait sub-varian baru ini.

2.3 Machine Learning

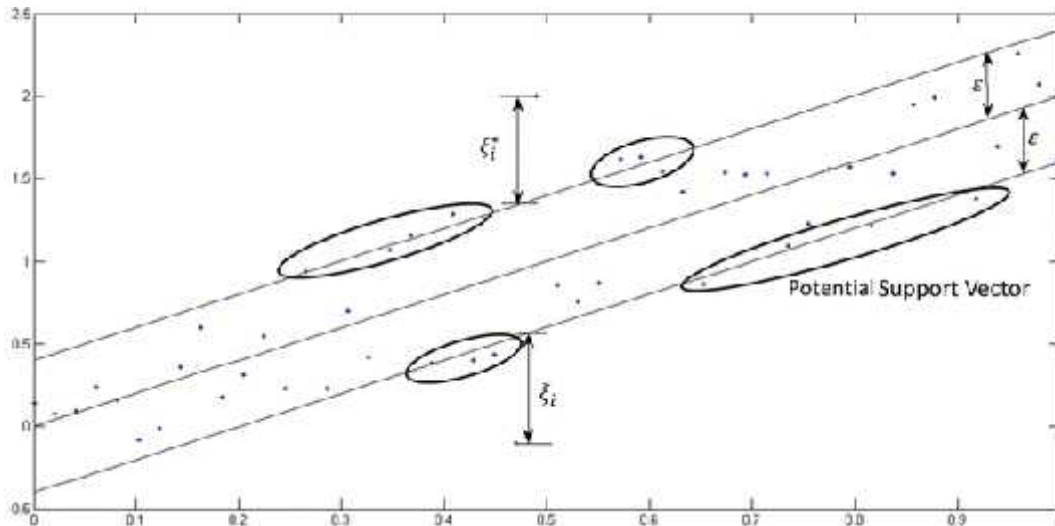
Pembelajaran mesin atau yang lebih dikenal dengan Machine Learning (ML) adalah sebuah teknologi dimana mesin dilatih dengan sendirinya tanpa arahan manusia. Teknologi ini sudah banyak dikembangkan untuk membantu pekerjaan manusia seperti perhitungan prediksi. Kesuksesan machine learning dalam prediksi bergantung dari data dan algoritma pembelajaran yang digunakan [8]. Machine learning merupakan bagian dari kecerdasan buatan (AI) [9] yang mengacu kepada software dan hardware yang dapat menirukan manusia. Pembelajaran mesin, telah berkembang pesat dalam beberapa tahun terakhir dalam konteks analisis data dan pemrosesan data, memungkinkan aplikasi berfungsi dengan cara yang cerdas [10]. Machine learning dan kecerdasan buatan merupakan istilah yang sangat populer khususnya didunia teknologi. Algoritma machine learning terbagi menjadi empat kategori: Supervised learning, unsupervised learning, Semi-supervised learning, and Reinforcement learning. Gambaran umum mengenai kategori machine learning di disajikan di gambar 1 [11].



Gambar 1. Kategori ML

2.4 SVR

SVR atau Super Vector Regression merupakan jenis algoritma yang banyak digunakan untuk pengambilan keputusan. Algoritma ini merupakan sebuah teknik analitik untuk menyelidiki hubungan antara satu atau lebih variabel prediktor dan variabel dependen bernilai nyata (kontinu) [12]. SVR merupakan suatu model adaptasi dari Machine Learning. Penggunaan model machine Learning SVR untuk memprediksi telah banyak digunakan [13]. Ilustrasi mengenai SVR di sajikan digambar 2 [14].



Gambar 2. Ilustrasi SVR

Grafik diatas menunjukkan sebuah garis diagonal di tengah (hyperplane) yang diapit oleh dua garis batas + dan garis batas -. Simbol  $\epsilon$  sebagai jarak antara hyperplane dengan 2 garis batas tadi. Ada beberapa datapoin yang dilingkari yang menjadi potential support vectors. Berarti titik-titik (data points) ini merupakan data poin yang bisa menjadi calon pembatas, sehingga semua data poin bisa masuk ke dalam satu kluster, dengan tetap sebisa mungkin meminimasi nilai  $\xi$  nya. Sehingga jika divisualisasikan, garis hyperplanenya sebisa mungkin melewati semua titik-titik data (data points). Berikut adalah langkah-langkah algoritma SVR:

1. Menyiapkan training set

$$\tau = \{ \vec{X} + \vec{Y} \}$$

2. Memilih kernel dan parameternya serta regularisasinya (regularization)
3. Membuat korelasi matriks

$$\vec{K}$$

4. Melatih modelnya untuk mendapatkan koefisien

$$\vec{a} = \{ \alpha_i \}$$

5. Gunakan koefisien di atas, kemudian buat estimatornya

$$f(\vec{X}, \vec{a}, x^*) = y^*$$

## 2.5. Analisis regresi

Analisis regresi adalah suatu metode atau teknik analisis hipotesis penelitian untuk menguji ada tidaknya pengaruh antara variabel satu dengan variabel lain, dinyatakan dalam bentuk persamaan regresi (matematik). Dipenelitian ini menggunakan dua analisis regresi, MSE (Mean Squared Error), RMSE (Root Mean Squared error). Keduanya merupakan merupakan contoh parameter yang biasa digunakan sebagai indikator untuk akurasi model prediksi. Dalam model regresi, untuk mengetahui keakuratan kinerja model suatu algoritma dapat mengambilnya dari nilai yang memiliki kesalahan terkecil. Semakin kecil nilai error yang diperoleh maka semakin dekat nilai atau jarak antara nilai aktual dengan nilai prediksi.

## 3. Hasil dan Pembahasan

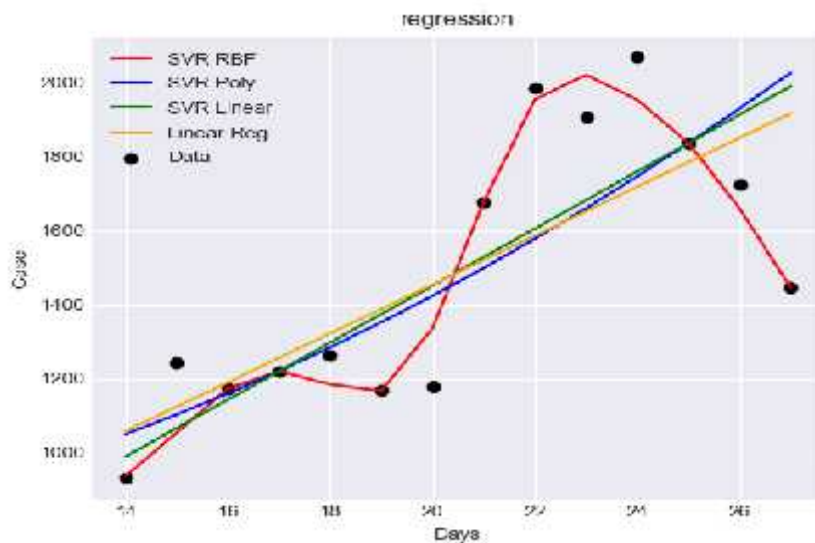
Data yang digunakan diperoleh dari BNPB (Badan Nasional Penanggulangan Bencana). BNPB merupakan sebuah lembaga non departemen yang mempunyai tugas membantu Presiden Republik Indonesia untuk mengkoordinasi perencanaan dan pelaksanaan kegiatan penanganan bencana dan kedaruratan secara terpadu di Indonesia. Kami mengumpulkan data berdasarkan tanggal pertama kali

kasus sub varian omicron BA.4 dan BA.5 terdeteksi di Indonesia. Total yang di gunakan sebanyak 17 data (dari tanggal 14 juni hingga 30 juni 2022). Data tersebut di kelompokkan menjadi dua, data training dan data uji. Untuk data training, 14 data digunakan (dari tanggal 14-27 juni 2022). Data training yang digunakan di tampilkan di tabel 1. Tabel tersebut menunjukkan tanggal dan jumlah kasus baru terkonfirmasi positif harian.

Tabel 1. Dataset

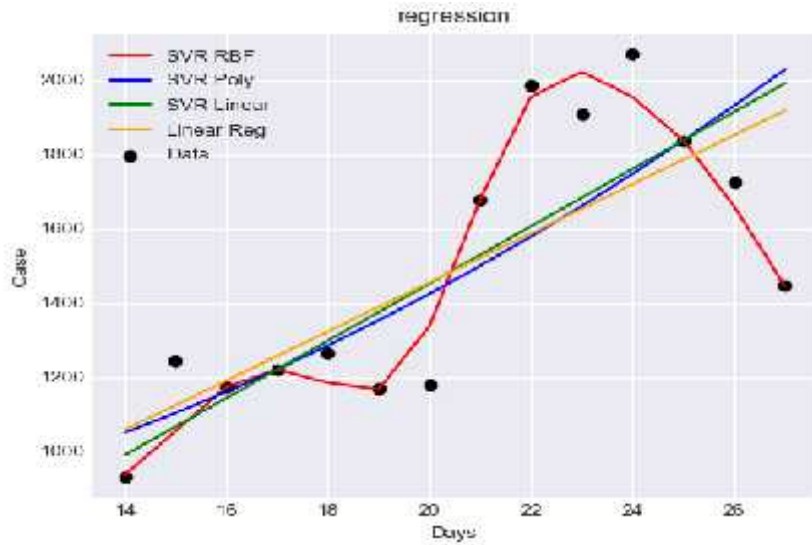
Dates	Cases
6/14/2022	930
6/15/2022	1242
6/16/2022	1173
6/17/2022	1220
6/18/2022	1264
6/19/2022	1167
6/20/2022	1180
6/21/2022	1678
6/22/2022	1985
6/23/2022	1907
6/24/2022	2069
6/25/2022	1831
6/26/2022	1726
6/27/2022	1445

Sebagai pengujian, kami menggunakan data yang berjarak tidak jauh dengan data latih. Tanggal pengujian di ambil dari tanggal 28 sampai 30 juni 2022. Grafik pengujian ditampilkan di gambar 3 sampai gambar 5. Grafik pengujian untuk tanggal 28 juni 2022 ditampilkan digambar 3.



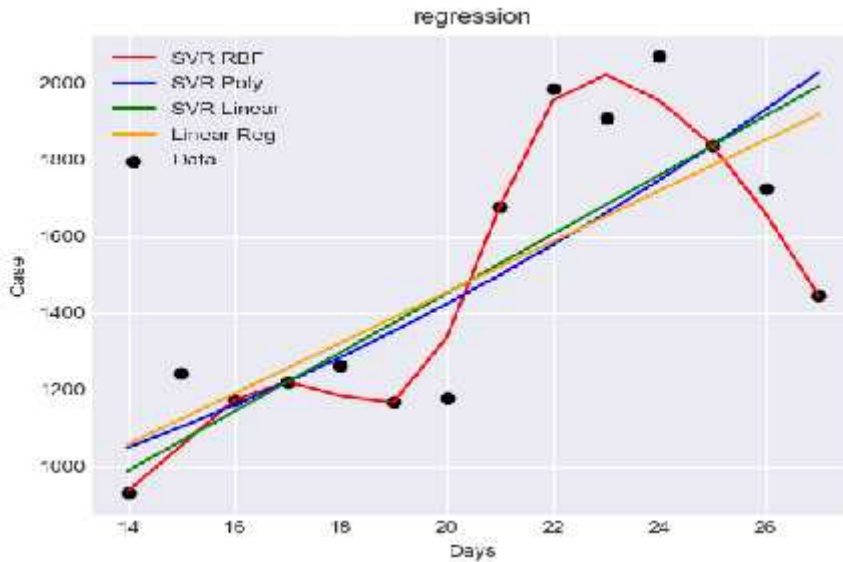
Gambar 3. Grafik Tanggal 28

Grafik pengujian untuk tanggal 29 juni 2022 ditampilkan di gambar 4.



Gambar 4. Grafik Tanggal 29

Terakhir, grafik pengujian untuk tanggal 30 juni 2022 ditampilkan di gambar 5.



Gambar 5. Grafik tanggal 30

Table pengujian terdiri dari tiga kolom. Kolom pertama menampilkan tanggal pengujian. Kolom kedua menampilkan hasil prediksi menggunakan metode SVR. Kolom ketiga menampilkan banyak kasus baru terkonfirmasi positif. Data lengkap mengenai pengujian yang dilakukan ditampilkan di tabel 2.

Tabel 2. Tabel Pengujian

Tanggal	Hasil Prediksi	Data real
6/28/2022	1289	2167
6/29/2022	1252	2149
6/30/2022	1291	2248

Setelah pengujian dilakukan, analisis regresi digunakan untuk mengetahui akurasi model prediksi jumlah kasus harian BA.4 BA.5. Pertama, pengujian untuk mengetahui MRE dan RMSE. Selanjutnya, perhitungan regression score dilakukan untuk mengetahui akurasi model algoritma SVR untuk prediksi. Hasil perhitungan di tampilan di tabel 3.

Tabel 3. Table RMSE dan MSE

Data	MSE	RMSE	Reg. Score	% accuracy
Kasus Harian	830447.3	911.288831	0.6747686	67.4%

#### 4. Kesimpulan

Berdasarkan pengujian yang dilakukan untuk memprediksi tingkat kasus baru sub omicron di Indonesia memperoleh akurasi yang rendah yaitu 67.4%. Hasil ini cukup rendah bila dibandingkan dengan penelitian sebelumnya. Sebagai pertimbangan, penelitian ini hanya menggunakan 17 data. Jumlah ini lebih kecil dari penelitian sebelumnya yang menggunakan metode yang sama.

#### 5. Saran

Penelitian dapat dikembangkan dengan menambahkan jumlah dataset. Hal selanjutnya yang dapat dilakukan adalah membuat model baru dengan mengkombinasikan algoritma SVR dengan algoritma prediksi lainya atau hanya menggunakan algoritma lainnya.

#### Daftar Pustaka

- [1] "Indonesia.go.id - Kasus Covid-19 Pertama, Masyarakat Jangan Panik." <https://www.indonesia.go.id/narasi/indonesia-dalam-angka/ekonomi/kasus-covid-19-pertama-masyarakat-jangan-panik> (accessed Jul. 11, 2022).
- [2] "Omicron Subvarian BA.4 dan BA.5: Gejala & Diagnosis." <https://aido.id/diseases/omicron-subvarian-ba.4-dan-ba.5/detail> (accessed Jul. 13, 2022).
- [3] 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.
- [4] S. Rath, A. Tripathy, and A. R. Tripathy, "Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model," *Diabetes Metab Syndr*, vol. 14, no. 5, pp. 1467–1474, 2020, doi: 10.1016/j.dsx.2020.07.045.
- [5] C. M. Gibran, S. Setiyawati, and F. Liantoni, "Prediksi Penambahan Kasus Covid-19 di Indonesia Melalui Pendekatan Time Series Menggunakan Metode Exponential Smoothing," *JIUP*, vol. 6, no. 1, p. 112, Mar. 2021, doi: 10.32493/informatika.v6i1.9442.
- [6] S. R. Rahmadania, "Ini 2 Gejala Utama Sub Varian BA.2 alias 'Son of Omicron,'" *detikHealth*. <https://health.detik.com/berita-detikhealth/d-5938598/ini-2-gejala-utama-sub-varian-ba2-alias-son-of-omicron> (accessed Jul. 12, 2022).
- [7] "Kenali Gejala Omicron Subvarian BA.4 dan BA.5." <https://aido.id/health-articles/kenali-gejala-omicron-subvarian-ba.4-dan-ba.5/detail> (accessed Jul. 12, 2022).
- [8] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput Sci*, vol. 2, no. 3, p. 160, 2021, doi: 10.1007/s42979-021-00592-x.
- [9] I. H. Sarker, "Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective," *SN COMPUT. SCI.*, vol. 2, no. 3, p. 154, Mar. 2021, doi: 10.1007/s42979-021-00535-6.
- [10] I. H. Sarker, M. H. Furhad, and R. Nowrozy, "AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions," *SN COMPUT. SCI.*, vol. 2, no. 3, p. 173, Mar. 2021, doi: 10.1007/s42979-021-00557-0.
- [11] "[PDF] Machine Learning: Algorithms and Applications | Semantic Scholar." <https://www.semanticscholar.org/paper/Machine-Learning%3A-Algorithms-and-Applications-Mohammed-Khan/37b7096e8818ee9b12034a0023268d590bf65810> (accessed Jul. 12, 2022).
- [12] F. Zhang and L. J. O'Donnell, "Chapter 7 - Support vector regression," in *Machine Learning*, A. Mechelli and S. Vieira, Eds. Academic Press, 2020, pp. 123–140. doi: 10.1016/B978-0-12-815739-8.00007-9.
- [13] 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.
- [14] M. B. Herlambang, "Machine Learning: Support Vector Regression," *Artificial intelligence and data science*, Oct. 19, 2018. <https://www.megabagus.id/machine-learning-support-vector-regression/> (accessed Jul. 12, 2022).