

Prediksi Jumlah Pasien Covid-19 Dengan Menggunakan Klasifikasi Algoritma Machine Learning

Vinna Rahmayanti Setyaning Nastiti¹, Putri Juli Amelia², Aidia Khoiriyah Firdausy³

^{1,2,3}Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang,
Jl. Raya Tlogomas No.246, Babatan, Tegalondo, Kec. Lowokwaru, Kota Malang, Jawa Timur 65144

e-mail: vinastiti@umm.ac.id¹, putrijuliaml@webmail.umm.ac.id², aidiakhoiriyah@webmail.umm.ac.id³

Received : September, 2022

Accepted : October, 2022

Published : October, 2022

Abstract

Corona virus or severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a disease that results in the occurrence of mild to moderate respiratory tract infections. Positive cases of Covid-19 in Indonesia were first detected on March 2, 2020 and continue until 2022. The additional number of deaths caused by COVID-19 has also increased. Therefore, the author is interested in making a predictive model of the cumulative number of COVID-19 patients who died in Indonesia. Therefore, in this study is how to predict the number of patients who die from COVID-19 in Indonesia by creating an appropriate accuracy model to help estimate the number of deaths associated with COVID-19 in Indonesia and assist the government in dealing with cases of new variants of COVID-19. In this study, the authors used the Decision Tree model using entropy criteria as well as Information Gain and Random Forest which resulted in accuracy rates of 91.83% (Decision Tree) and 73.80% (Random Forest). The results, explain that the model used is good. The more the R-squared error value is close to 1, the better the model used will be.

Keywords: Covid-19, Machine Learning, Classification

Abstrak

Virus Corona atau severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) adalah penyakit yang mengakibatkan terjadinya infeksi saluran pernafasan ringan hingga sedang. Kasus positif Covid-19 di Indonesia pertama kali dideteksi pada tanggal 2 Maret 2020 dan masih berlanjut hingga tahun 2022. Penambahan kasus jumlah kematian yang disebabkan oleh COVID-19 ini pun meningkat. Oleh karena itu, penulis tertarik untuk membuat model prediksi jumlah kumulatif pasien COVID-19 yang meninggal di Indonesia. Oleh karena itu, pada penelitian ini adalah bagaimana memprediksi jumlah pasien yang meninggal akibat COVID-19 di Indonesia dengan membuat model akurasi yang tepat untuk membantu memperkirakan jumlah kematian yang dikaitkan dengan COVID-19 di Indonesia dan membantu pemerintah dalam menangani kasus varian baru COVID-19. Pada penelitian ini, penulis menggunakan model Decision Tree dengan menggunakan kriteria Entropy maupun Information Gain dan Random Forest yang menghasilkan tingkat akurasi sebesar 91.83% (Decision Tree) dan 73.80% (Random Forest). Hasil tersebut, menjelaskan bahwa model yang digunakan baik. Semakin nilai error R-squared mendekati 1 maka model yang digunakan akan semakin baik.

Kata Kunci: Covid-19, Pembelajaran Mesin, Klasifikasi

1. PENDAHULUAN

Jenis virus corona baru yang disebut COVID-19 dapat menyebar dari satu manusia ke manusia lainnya melalui orang tersebut. Coronavirus adalah kelas virus yang mempengaruhi sistem pernapasan. Sindrom Pernafasan Timur Tengah (MERS), Sindrom Pernafasan Akut Parah (SARS), infeksi paru-paru (pneumonia), dan bahkan kematian dapat disebabkan oleh virus ini. Akhir Desember 2019 menjadi saksi kemunculan pertama infeksi COVID-19 di kota Wuhan di China, yang segera menyebar ke bagian lain China dan sebagian besar negara lain, termasuk Indonesia [1].

Dua WNI tertular COVID-19 dari warga negara Jepang pada 2 Maret 2020. Hingga 9 April 2022, DKI Jakarta, Jawa Barat, dan Jawa Tengah menjadi kasus terkonfirmasi terbanyak [2].

Menurut statistik WHO, Indonesia adalah salah satu negara Asia dengan jumlah kematian terkait COVID-19 terbesar di dunia. Menyusul 496 kematian yang terjadi pada 18 April 2020, angka kematian terkait COVID-19 di negara itu naik menjadi 8,9%. COVID-19, yang saat ini memiliki 5,54% di Cina. Perbandingan kasus kematian COVID-19 di Indonesia dan China mengungkapkan bahwa hanya 14,8% pasien dengan gejala suspek positif COVID-19 yang tercatat di Indonesia. Hingga 17 April 2020, 5.923 orang di Indonesia dinyatakan positif COVID-19 [3].

Setiap hari, ada lebih banyak individu yang terinfeksi virus, lebih banyak kasus pemulihan, dan bahkan lebih banyak kematian. Memprediksi jumlah kasus COVID-19 yang diantisipasi akan membantu menyelesaikan masalah ini. Prediksi adalah perkiraan metodis tentang apa yang akan terjadi di masa depan yang dibuat dengan menggunakan informasi dari masa lalu dan masa kini. Prediksi mungkin tidak selalu memberikan respons yang jelas karena kesalahan (perbedaan antara apa yang sebenarnya terjadi dan hasil yang diprediksi) harus dijaga seminimal mungkin. tetapi mencari solusi yang mendekati apa yang akan terjadi [4].

Berdasarkan permasalahan tersebut, solusi dapat ditemukan dengan membandingkan Decision Tree dan Random Forest, dua pendekatan untuk meramalkan jumlah pasien yang meninggal karena COVID-19, menggunakan teknik data mining. Data yang digunakan adalah data dari setiap area yang dapat diakses di situs web Kaggle, namun karena volume data yang

besar, metode Big Data dapat digunakan untuk menganalisis data ini untuk memberikan informasi yang berharga dan dapat dipahami [5].

Kedua teknik tersebut telah banyak digunakan pada penelitian sebelumnya, seperti penelitian "Klasifikasi Diagnosis Penyakit COVID-19 Menggunakan Metode Decision Tree" oleh Rasyid Ridho dan Hendra. -19 tergantung pada keadaan pengguna, jenis kelamin, usia, dan gejala. Pengujian dihitung dengan menggunakan pendekatan matriks konfusi, dan hasilnya menunjukkan tingkat akurasi 90% [6].

Kajian Nanik Wuryani dan Sarifah Agustiani berjudul "Random Forest Classifier untuk Deteksi Pasien COVID-19 Berdasarkan Gambar CT Scan" memuat temuan investigasi yang dimulai dengan memasukkan satu fitur ke dalam eksperimen, kemudian menggabungkannya dengan fitur lain, kemudian membandingkannya menggunakan klasifikasi dengan algoritma lain seperti *K-Nearest Neighbor* (KNN), *Decision Tree*, *Linear Discriminant Analysis* (LDA), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Naive Bayes*, menunjukkan bahwa Random Forest Classifier dengan fitur Haralic dan Color Histogram menghasilkan akurasi 96,9%. KNN selanjutnya memiliki akurasi 96,5%, Decision Tree 95,5%, dan Naive Bayes 42,4%.

Pada saat ini, kasus COVID-19 di Indonesia kian melandai, penambahan kasus dan angka kematian akibat COVID-19 ini juga menurun tiap harinya. Berdasarkan survey internal yang telah dilakukan Kemenko PMK di 18 Rumah Sakit DKI Jakarta pada bulan Februari 2022, saat ini angka kematian akibat COVID-19 di Indonesia telah turun di peringkat ke-14[7]. Namun juga masih terdapat varian baru dari COVID-19, lebih tepatnya subvarian Omicron BA.4 dan BA.5 yang baru – baru ini sudah terdeteksi masuk ke Indonesia[8]. Selain itu pasien COVID-19 varian Omicron yang sudah sembuh atau pulih tidak menutup kemungkinan masih bisa mengalami beberapa gejala jangka panjang atau yang disebut *long covid*. Salah satu kondisi yang banyak dikeluhkan pasien Omicron yang sudah sembuh adalah nyeri tubuh dan kelelahan[9].

Maka dari itu, tujuan dari penelitian ini adalah selain untuk membandingkan dua metode data mining yaitu *Decision Tree* dan *Random Forest* yang diharapkan dapat membantu memprediksi jumlah kasus pasien yang meninggal akibat COVID-19 juga untuk membantu pemerintah dalam membuat kebijakan agar kasus COVID-19 di Indonesia

tidak mengalami kenaikan kembali dan pandemi segera berakhir.

2. METODE PENELITIAN

Metode data mining digunakan dalam pendekatan pengolahan data penelitian ini, yang dimulai dengan studi pustaka, pengumpulan data, pengembangan model data mining, dan uji coba dan validasi hasil. Berikut adalah penjelasan dari masing-masing tahapan dalam metodologi penelitian ini:

2.1 Kaji Literatur

A. Pembelajaran Mesin

Pembelajaran mesin adalah studi tentang algoritma dan model statistik yang digunakan oleh komputer untuk melakukan tugas tanpa instruksi, pola, dan inferensi (ML) yang eksplisit. Algoritma pembelajaran mesin membangun model matematika dari "data pelatihan" untuk membuat prediksi tanpa diprogram secara eksplisit [10].

B. Prediksi

Memprediksi apa yang akan terjadi adalah proses prediksi. Mengurangi ketidakpastian dan menciptakan perkiraan yang lebih tepat tentang masa depan adalah tujuan dari prediksi data. Menggunakan metode Decision Tree di bidang pendidikan, model Algoritma Neural Network di bidang kesehatan, algoritma MKNN-X di bidang meteorologi, klimatologi, dan geofisika, metode Regresi Linier Berganda di bidang perdagangan, dan penelitian model analitik untuk pendugaan produksi garam di Indonesia. Kabupaten Kupang yang menggunakan regresi linier hanyalah beberapa contoh dari sekian banyak bidang yang dapat digunakan prediksi [11].

C. Decision Tree C4.5

C4.5 adalah evolusi dari ID3 dan standar untuk algoritme pembelajaran terawasi yang baru. C4.5 adalah klasifikasi prediksi, segmentasi, atau algoritma pengelompokan. Dasar algoritma C4.5 masih diperdebatkan. Pertanyaan yang menjadi inti perhatian pohon adalah salah satu klasifikasi, dan jawabannya adalah kelas atau segmen. Setiap atribut dalam Decision Tree C4.5 menggunakan *Impurity Gain Ratio*. *Information gain* adalah metrik

pemilihan fitur yang digunakan untuk memilih fitur uji untuk setiap simpul di pohon. Fitur dengan informasi gain maksimum dipilih sebagai node uji fitur. Gain (S, A) adalah ekstraksi informasi dari atribut A sehubungan dengan data keluaran S. Informasi diperoleh dari data keluaran atau variabel terikat S, yang diklasifikasikan dengan atribut A, dilambangkan dengan gain (S,A)[12].

Dengan menggunakan rumus di bagian bawah ini, kita dapat mengurangi keuntungan [13]:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

D. Random Forest

Pohon keputusan yang membentuk Hutan Acak (RF) masing-masing secara independen bergantung pada hasil vektor acak yang dihasilkan dari distribusi yang sama. Kemampuan Random Forest untuk memberikan tingkat kesalahan yang cukup rendah adalah karena fungsi yang dipilih secara acak untuk mengurutkan setiap node atau kumpulan node [9]. Untuk meningkatkan akurasi pada metode random forest, hyperparameter yang digunakan meliputi *n_estimator* dan *max_depth*. Parameter ini adalah *n_estimator* yang menyimpulkan semua pohon keputusan berikutnya, sedangkan *max_depth* menunjukkan kedalaman setiap level regresi linier[14].

Table 1 Hyperparameter regresi random forest

HyperParameter Random Forest	
HyperParameter	Nilai
Max_depth	10
N_estimators	100

2.2 Pengumpulan Dataset

Anda bisa mendapatkan dataset pasien COVID-19 untuk Indonesia dari situs web Kaggle (www.kaggle.com). Dengan menggunakan teknik machine learning, dataset ini selanjutnya akan dilatih untuk membuat model prediksi jumlah pasien yang meninggal akibat COVID-19. Hanya karakteristik yang relevan yang dikumpulkan

dari kumpulan data yang diarsipkan dan digunakan dalam proses kategorisasi. Tabel berikut menunjukkan dataset pasien COVID-19 di Indonesia:

Table 2 Dataset Kaggle

No	Column	Jumlah Data	Dtype
1	Date	31822	Object
2	Location_ISO_Code	31822	Object
3	New_Cases	31822	Int64
4	New_Deaths	31822	Int64
5	New_Active_Cases	31822	Int64
6	Total_Cases	31822	Int64
7	Total_Deaths	31822	Int64
8	Total_Recovered	31822	Int64
9	Total_Active_Cases	31822	Int64
10	Province	30893	Object
11	Country	31822	Object
12	Continent	31822	Object
13	Island	30893	Object
14	Time_Zone	30893	Object
15	New_Cases_per_Million	31822	Float64
16	Total_Cases_per_Million	31822	Float64
17	New_Deaths_per_Million	31822	Float64
18	Total_Deaths_per_Million	31822	Float64
19	Case_Fatality_Rate	31822	Float64
20	Case_Recovered_Rate	31822	Float64
21	Longitude	31822	Float64
22	Latitude	31822	Float64

2.3 Pembangunan Model Data Mining

Confusion Matrix, yang merupakan matriks klasifikasi benar-salah yang dihasilkan oleh metode ini, digunakan dalam pembuatan model Data Mining. *Confusion Matrix* adalah teknik yang digunakan untuk menghitung tingkat akurasi, presisi, dan nilai recall untuk ide data mining.

Table 3 Confusion Matrix

		Observed	
		True	False
Predicted Class	True	Positive (TP)	Positive (TV)
	False	Negative (FN)	Negative (TN)

Keterangan:

1. True Positive (TP) mengacu pada proporsi data positif yang diklasifikasikan oleh sistem dengan benar.

2. TN adalah singkatan dari True Negative, yaitu jumlah data negatif yang dikategorikan secara akurat oleh sistem.
3. FN adalah singkatan dari False Negative, yang merupakan jumlah data negatif yang salah dikategorikan oleh sistem.
4. FP adalah singkatan dari False Positive, yang mengacu pada jumlah data positif yang salah dikategorikan oleh sistem.

Dengan kata lain, dengan membandingkan data yang dikategorikan dengan benar dengan seluruh kumpulan data, nilai akurasi dapat ditentukan. Persamaan [15] dapat digunakan untuk menghitung nilai akurasi, presisi, dan recall.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} * 100\%$$

$$\text{Recall} = \frac{TP}{FN+TP} * 100\%$$

3. HASIL DAN PEMBAHASAN

Langkah-langkah yang akan dilakukan dalam mengantisipasi kondisi pasien COVID-19 dan mengetahui karakteristik pasien atau pemeriksaan untuk memudahkan penelitian berjalan dengan lancar, metodis, dan mencapai tujuan yang telah ditetapkan. Pada penelitian ini, penulis menggunakan pustaka python untuk mengimport file dataset, menganalisa dataset, memvisualisasikan dataset, melakukan persiapan dataset dan sampai pembuatan model.

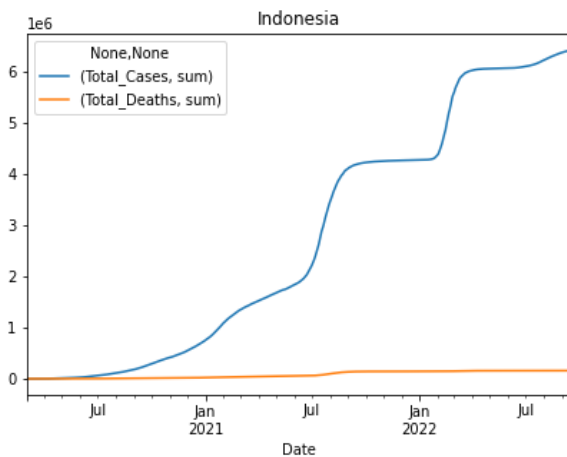
Setelah dataset didapatkan, dilakukannya *preprocessing* dan pemodelan *decision tree*. Pada tahap *preprocessing*, data dianalisis mulai dari type data yang digunakan, missing value yang terdapat pada dataset, setelah itu dilakukannya *converting* pada dataset dari bentuk *string* menjadi *numeric* dengan tujuan mempermudah dataset pada proses *splitting* dan visualisasi data, selain itu juga digunakan pada proses *decision tree classification* yang berguna untuk membantu prediksi terkait peningkatan data pasien yang meninggal akibat COVID-19 di Indonesia.

Table 4 Hasil converting string menjadi numeric

No	Column	Jumlah Data	Dtype
1	Date	31822	Int64
2	Location_ISO_Code	31822	Int64
3	New_Cases	31822	Int64
4	New_Deaths	31822	Int64
5	New_Active_Cases	31822	Int64
6	Total_Cases	31822	Int64
7	Total_Deaths	31822	Int64
8	Total_Recovered	31822	Int64
9	Total_Active_Cases	31822	Int64
10	Province	30893	Int64
11	Country	31822	Int64
12	Continent	31822	Int64
13	Island	30893	Int64
14	Time_Zone	30893	Int64
15	New_Cases_per_Million	31822	Float64
16	Total_Cases_per_Million	31822	Float64
17	New_Deaths_per_Million	31822	Float64
18	Total_Deaths_per_Million	31822	Float64
19	Case_Fatality_Rate	31822	Float64
20	Case_Recovered_Rate	31822	Float64
21	Longitude	31822	Float64
22	Latitude	31822	Float64

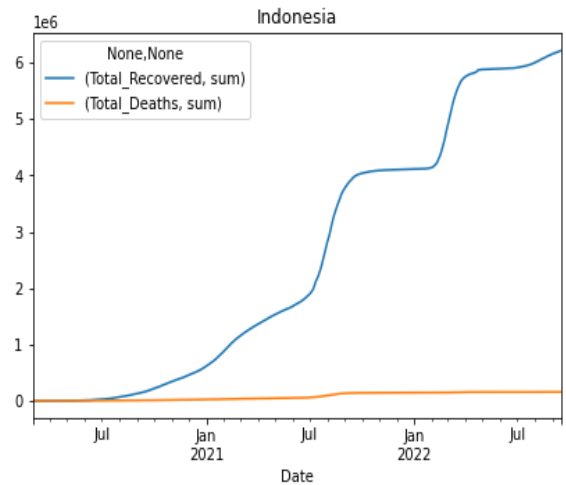
Langkah selanjutnya yaitu memvisualisasikan dataset COVID-19 di Indonesia sebagai berikut:

- Total kasus aktif dan total kematian akibat COVID-19 di Indonesia



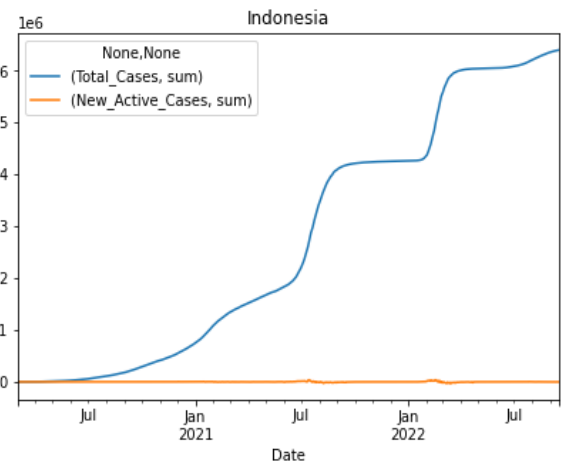
Gambar 1 Kasus COVID-19 di Indonesia

- Total kasus sembuh dan total kematian akibat COVID-19 di Indonesia



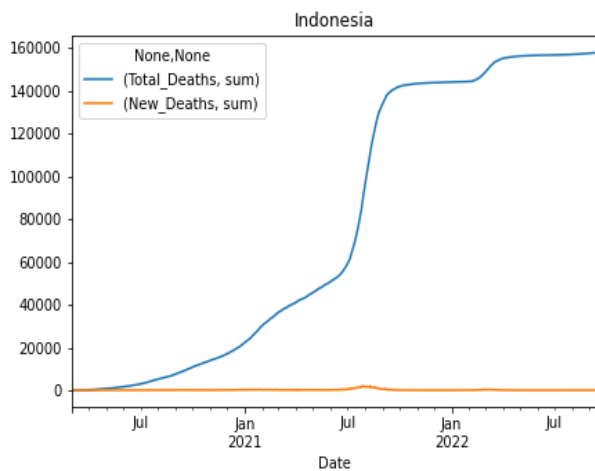
Gambar 2 Kasus COVID-19 di Indonesia

- Total kasus aktif dan kasus aktif baru akibat COVID-19 di Indonesia



Gambar 3 Total kasus COVID-19 di Indonesia

- Total kasus kematian dan kasus kematian baru akibat COVID-19 di Indonesia



Gambar 4 Total kasus COVID-19 di Indonesia

Setelah proses visualisasi, penulis melakukan ujicoba dataset dengan menggunakan model algoritma *Decision Tree* dan *Random Forest* dengan membagi data menjadi 80% sebesar 25.458 data untuk *training* dan 20% sebesar 6.364 data untuk *testing*. Algoritma *Random Forest* menghasilkan tingkat akurasi sebesar 71.60% sedangkan untuk uji coba menggunakan algoritma *Decision Tree* menghasilkan tingkat akurasi sebesar 90.07%.

Melalui hasil akurasi yang diperoleh, maka algoritma *Decision Tree* ternyata dapat dengan maksimal mampu melakukan prediksi jumlah pasien meninggal akibat COVID-19 di Indonesia dengan baik. Selanjutnya hasil pengujian dinyatakan dalam nilai presisi, recall, dan akurasi yang dinyatakan dalam bentuk confusion matrix seperti dibawah ini:

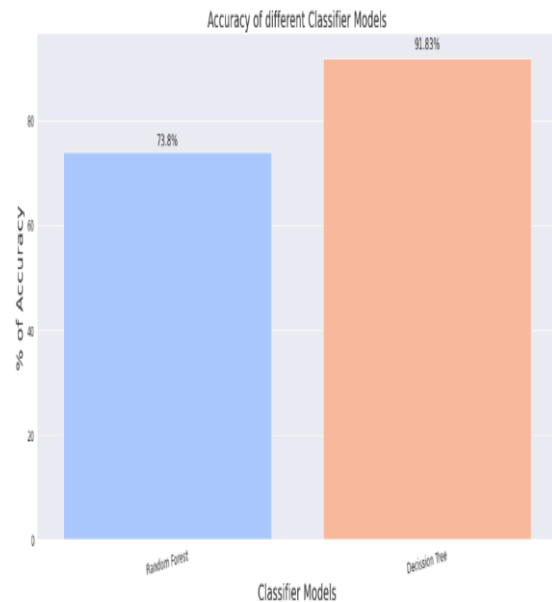
Table 4 Confusion Matrix

	Preci sion	Rec all	F1- score	Akurasi
Decision Tree	71%	74 %	72%	91,83%
Random Forest	92%	92 %	92%	73,80%

Pada Table 4 Confusion Matrix menunjukkan beberapa informasi seperti nilai presisi, recall, dan f1-score dari setiap model algoritma. Confusion matrix memberikan informasi banyaknya nilai yang true positive, true negative, false positive, dan atau false negative dari data pengujian.

Nilai presisi akan menunjukkan seberapa akurat prediksi dalam mengenali data sesuai dengan kelasnya, sedangkan nilai recall akan menunjukkan kemampuan classifier untuk mengambil informasi sesuai dengan kelas aslinya. Akurasi juga menunjukkan performa dari keseluruhan model yang dirancang.

Dari hasil akurasi algoritma data mining dapat diringkas seperti dibawah ini:



Gambar 5 Hasil akurasi model machine learning

4. KESIMPULAN

Pada penelitian ini digunakannya model Data Mining untuk prediksi jumlah pasien yang meninggal akibat COVID-19 di Indonesia dengan menggunakan algoritma *Decision Tree* dan *Random Forest*. Model dibangun menggunakan dataset pasien COVID-19 di Indonesia yang didapatkan dari website www.kaggle.com. Berdasarkan hasil penelitian untuk nilai akurasi algoritma yang mempunyai nilai akurasi, presisi, recall, *f1-score* tertinggi pada dataset yaitu *Decision Tree* dengan tingkat akurasi sebesar 91.83%, sedangkan untuk algoritma *Random*

Forest memiliki hasil akurasi sebesar 73.80%. Hasil dari penelitian ini bermanfaat untuk diterapkan pada masa yang akan datang, untuk membantu pemerintah dalam membuat kebijakan dalam menangani pandemi COVID-19 yang hingga saat ini masih terdapat varian baru yang muncul dan dapat mencegah pertambahan kasus aktif dan kasus meninggal akibat COVID-19. Penelitian ini harus dikaji ulang dengan menggunakan metode *data mining* lain untuk mendapatkan nilai akurasi yang lebih tinggi.

PERNYATAAN PENGHARGAAN

Penulis berharap penelitian selanjutnya akan menggunakan metode pembelajaran mesin yang berbeda untuk meningkatkan akurasi.

DAFTAR PUSTAKA

- [1] A. Susilo *et al.*, "Coronavirus Disease 2019: Tinjauan Literatur Terkini," *Jurnal Penyakit Dalam Indonesia*, vol. 7, no. 1, pp. 45–67, Apr. 2020, doi: 10.7454/JPDI.V7I1.415.
- [2] "Infeksi Emerging Kementerian Kesehatan RI," 2022. <https://covid19.kemkes.go.id/dashboard/covid-19> (accessed Sep. 30, 2022).
- [3] J. Sindya Widjaya, D. R. Agushinta, and S. Rahayu Puspita Sari, "Sistem Prediksi Jumlah Pasien Covid-19 Menggunakan Metode Trend Least Square Berbasis Web," *Sistemasi: Jurnal Sistem Informasi*, vol. 10, no. 1, pp. 39–51, Jan. 2021, doi: 10.32520/STMSI.V10I1.1036.
- [4] L. Ainiyah and M. Bansori, "Prediksi Jumlah Kasus COVID-19 menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) (Studi Kasus Kabupaten Sidoarjo)," *Jurnal Sains Dasar*, vol. 10, no. 2, pp. 62–68, Nov. 2021, doi: 10.21831/JSD.V10I2.43606.
- [5] R. A. Indraputra and R. Fitriana, "K-Means Clustering Data COVID-19," *Jurnal Teknik Industri*, vol. 10, no. 3, pp. 275–282, Dec. 2020, doi: 10.25105/JTI.V10I3.8428.
- [6] R. Ridho and H. Hendra, "Klasifikasi Diagnosis Penyakit Covid-19 Menggunakan Metode Decision Tree," *JUST IT: Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, vol. 11, no. 3, pp. 69–75, Aug. 2022, doi: 10.24853/JUSTIT.11.3.
- [7] Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan Republik Indonesia, "Jika Covid 19 Sudah Menjadi Endemi, Penanganannya Jadi Seperti Penyakit Biasa," Jun. 23, 2022. kemenkopmk.go.id/jika-covid-19-sudah-menjadi-endemi-penanganannya-jadi-seperti-penyakit-biasa (accessed Oct. 14, 2022).
- [8] Dinas Kesehatan Provinsi Kalimantan Barat, "Varian Baru Omicron BA.4 dan BA.5 Terdeteksi di Indonesia, Berbahayakah?," Jun. 30, 2022. <https://dinkes.kalbarprov.go.id/varian-baru-omicron-ba-4-dan-ba-5-terdeteksi-di-indonesia-berbahayakah/> (accessed Oct. 13, 2022).
- [9] Linda Hasibuan, "Ini Tanda Long Covid yang Bisa Terjadi Usai Sembuh dari Covid," *CNBC Indonesia*, Mar. 04, 2022. <https://www.cnbcindonesia.com/lifestyle/20220304104340-33-319971/ini-tanda-long-covid-yang-bisa-terjadi-usai-sembruh-dari-covid> (accessed Oct. 15, 2022).
- [10] R. Rino Pratama, "Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 302–311, May 2020, doi: 10.30812/MATRIK.V19I2.688.
- [11] Wahyudin and H. Purwanto, "Prediksi Kasus Covid-19 di Indonesia menggunakan Metode Backpropagation dan Regresi Linear," *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)*, vol. 5, no. 2, pp. 331–339, May 2021, doi: 10.52362/JISAMAR.V5I2.420.
- [12] E. P. Cynthia and E. Ismanto, "Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji," *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, vol. 3, p. 1, Jul. 2018, doi: 10.30645/jurasik.v3i0.60.
- [13] Indarto, E. Utami, and S. Raharjo, "Prediksi Risiko Kematian Pasien Stroke Perdarahan dengan menggunakan Teknik Klasifikasi Data Mining,"

- Informasi Interaktif*, vol. 5, no. 2, pp. 86–91, May 2020.
- [14] S. Fachid and A. Triayudi, “Perbandingan Algoritma Regresi Linier dan Regresi Random Forest Dalam Memprediksi Kasus Positif Covid-19,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 68, Jan. 2022, doi: 10.30865/mib.v6i1.3492.
- [15] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, “Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi WSN Menggunakan Confusion Matrix,” *Jurnal Informatika Upgris*, vol. 6, no. 2, Jan. 2020, doi: 10.26877/JIU.V6I2.6552.