

Penerapan Stacking untuk Optimasi Model Diagnosa Coronavirus Disease 19 (COVID-19)

Yulianti¹, Sri Mulyati², Teti Desyani³

Teknik Informatika, Universitas Pamulang, Jl. Raya Puspitek No. 46 Buaran, Serpong, Tangerang Selatan, Banten, Indonesia, 15417

e-mail: ¹yulianti@unpam.ac.id, ²dosen00391@unpam.ac.id, ³dosen00839@unpam.ac.id

Submitted Date: January 19th, 2024

Reviewed Date: April 03rd, 2024

Revised Date: April 26th, 2024

Accepted Date: April 30th, 2024

Abstract

Laboratory test results in COVID-19 patients are not specific, but lymphopenia, increased lactate dehydrogenase and increased aminotransferases are often found. Meanwhile, chest imaging examination can show a picture of pneumonia. Until now, there has been no specific therapy in the treatment of COVID-19. There are two of the largest studies on COVID-19 therapy which are currently still running globally. Studies show that the antiviral favipiravir, remdesivir, and tocilizumab may have some benefits for treating COVID-19, and their use has been approved in Indonesia. There have been many diagnostic methods using machine learning that are used to detect whether someone has COVID-19 or not. However, the accuracy of the test may vary depending on when your sample was taken during the course of your disease. If you get tested too soon after exposure to COVID-19, there may not be enough virus in your body to get an accurate result. If this was the case at the time of the test, your test may come back negative, even if you do have the virus. This will be considered a 'false negative' test. It is important to understand that healthcare professionals consider a number of factors in making a diagnosis of COVID-19. In this study using the experimental method by making applications to implement the proposed algorithm. Then test the model using a secondary dataset downloaded from Kaggle and measure the performance of the model.

Keywords: Stacking; Optimization of the Diagnostic Model; COVID-19

Abstrak

Hasil pemeriksaan laboratorium pada pasien COVID-19 tidak spesifik, tetapi sering ditemukan limfopenia, peningkatan laktat dehidrogenase, dan peningkatan aminotransferase. Sedangkan pemeriksaan pencitraan toraks dapat menunjukkan gambaran pneumonia. Sampai saat ini, belum terdapat terapi spesifik dalam penanganan COVID-19. Terdapat dua studi terbesar tentang terapi COVID-19 yang hingga saat ini masih berjalan secara global. Studi menunjukkan bahwa antivirus favipiravir, remdesivir, dan tocilizumab mungkin memiliki beberapa manfaat untuk penanganan COVID-19, dan sudah diperbolehkan penggunaannya di Indonesia. Telah banyak metode diagnosa menggunakan machine learning yang digunakan untuk mendeteksi seseorang terkena COVID-19 atau tidak. Namun, keakuratan tes dapat bervariasi tergantung pada saat sampel Anda diambil selama perjalanan penyakit Anda. Jika Anda dites terlalu cepat setelah terpapar COVID-19, mungkin tidak ada cukup virus di tubuh Anda untuk mendapatkan hasil yang akurat. Jika ini masalahnya pada saat tes, tes Anda mungkin kembali negatif, bahkan jika Anda benar-benar memiliki virus. Ini akan dianggap sebagai tes 'negatif palsu'. Penting untuk dipahami bahwa profesional perawatan kesehatan mempertimbangkan sejumlah faktor dalam membuat diagnosis COVID-19. Pada penelitian ini menggunakan metode eksperimen dengan membuat aplikasi untuk menerapkan algoritma yang diusulkan. Kemudian menguji model menggunakan dataset sekunder yang diunduh dari Kaggle dan mengukur kinerja model.

Kata Kunci: Stacking; Optimasi Model Diagnosa; COVID-19

1 Pendahuluan

Coronavirus disease 2019 (COVID-19) adalah penyakit menular yang disebabkan oleh virus SARS-CoV-2. Kebanyakan orang yang jatuh sakit COVID-19 akan mengalami gejala ringan hingga sedang dan sembuh tanpa pengobatan khusus. Namun, beberapa akan menjadi sakit parah dan memerlukan perhatian medis. Pada tahun 2019, virus corona baru diidentifikasi sebagai penyebab wabah penyakit yang berasal dari China. Virus ini merupakan patogen zoonotik yang memiliki tingkat mutasi tinggi, dan dapat menetap pada manusia dan binatang dengan presentasi klinis beragam, mulai dari asimtomatik, gejala ringan sampai berat, sampai kematian (Casella et al., 2022) (Sahin et al., 2020) (Guo et al., 2020).

Coronavirus disease 2019 (COVID-19) merupakan penyakit infeksi saluran pernapasan yang disebabkan oleh severe acute respiratory syndrome virus corona 2 (SARS-CoV-2), atau sering disebut virus Corona. Coronavirus adalah keluarga virus yang dapat menyebabkan penyakit seperti flu biasa, sindrom pernapasan akut parah (SARS) dan sindrom pernapasan Timur Tengah (MERS).

Virus dapat menyebar dari mulut atau hidung orang yang terinfeksi dalam partikel cairan kecil ketika mereka batuk, bersin, berbicara, bernyanyi, atau bernapas. Partikel-partikel ini berkisar dari tetesan pernapasan yang lebih besar hingga aerosol yang lebih kecil. Anda dapat terinfeksi dengan menghirup virus jika Anda berada di dekat seseorang yang memiliki COVID-19, atau dengan menyentuh permukaan yang terkontaminasi dan kemudian mata, hidung, atau mulut Anda. Virus menyebar lebih mudah di dalam ruangan dan di tempat ramai.

Virus ini dikenal sebagai sindrom pernafasan akut yang parah coronavirus 2 (SARS-CoV-2). Penyakit yang ditimbulkannya disebut coronavirus disease 2019 (COVID-19). Pada Maret 2020, Organisasi Kesehatan Dunia (WHO) menyatakan wabah COVID-19 sebagai pandemi.

Hasil pemeriksaan laboratorium pada pasien COVID-19 tidak spesifik, tetapi sering ditemukan limfopenia, peningkatan laktat dehidrogenase, dan peningkatan aminotransferase. Sedangkan pemeriksaan pencitraan toraks dapat menunjukkan gambaran pneumonia (Casella et al., 2022) (World Health Organization, 2020).

Karakteristik gambaran COVID-19 pada CT scan toraks nonkontras adalah ground glass opacification (GGO) bilateral, multilobar, dengan distribusi perifer atau posterior. Walaupun kurang spesifik, ultrasonography (USG) dan rontgen toraks juga dapat membantu menegakkan diagnosis COVID-19. Diagnosis baku emas COVID-19 adalah mendeteksi virus RNA dengan pemeriksaan nucleic acid amplification test (NAAT) dengan metode real time polymerase chain reaction (RT-PCR) (Casella et al., 2022) (World Health Organization, 2020).

Sampai saat ini, belum terdapat terapi spesifik dalam penanganan COVID-19. Terdapat dua studi terbesar tentang terapi COVID-19 yang hingga saat ini masih berjalan secara global. Studi menunjukkan bahwa antivirus favipiravir, remdesivir, dan tocilizumab mungkin memiliki beberapa manfaat untuk penanganan COVID-19, dan sudah diperbolehkan penggunaannya di Indonesia (McIntosh et al., 2020) (Mendez, 2020) (Oxford University, 2022)

Pasien COVID-19 tanpa gejala dan derajat ringan umumnya hanya disarankan isolasi di rumah dan menggunakan obat simtomatik. Pasien dengan gejala derajat sedang sampai berat membutuhkan terapi oksigen, sehingga disarankan untuk dirawat inap dan terkadang diperlukan tindakan intubasi dan ventilasi mekanik apabila terjadi gagal napas atau acute respiratory distress syndrome (Cennimo, 2024) (Burhan et al., 2020)

Adanya infeksi COVID-19 yang meluas telah mendorong upaya di seluruh dunia untuk mengendalikan dan mengelola virus, dan diharapkan dapat mengekangnya sepenuhnya. Salah satu penelitian penting adalah penggunaan machine learning (ML) untuk memahami dan melawan COVID-19. Ini saat ini merupakan bidang penelitian aktif. Meskipun sudah ada banyak survei dalam literatur, ada kebutuhan untuk mengikuti jumlah publikasi yang berkembang pesat tentang aplikasi ML terkait COVID-19. Makalah ini menyajikan ulasan laporan terbaru tentang algoritma ML yang digunakan dalam kaitannya dengan COVID-19. Kami fokus pada potensi ML untuk dua aplikasi utama: diagnosis COVID-19 dan prediksi risiko dan tingkat keparahan kematian, menggunakan data klinis dan laboratorium yang tersedia. Aspek yang terkait dengan tipe algoritme, kumpulan data pelatihan,

dan pemilihan fitur dibahas. Saat kami meliput karya yang diterbitkan antara Januari 2020 dan Januari 2021, beberapa poin penting telah terungkap. Sebagian besar algoritme pembelajaran mesin yang digunakan dalam dua aplikasi ini adalah algoritme pembelajaran yang diawasi. Model yang sudah mapan belum digunakan dalam implementasi dunia nyata, dan sebagian besar penelitian terkait bersifat eksperimental. Fitur diagnostik dan prognostik yang ditemukan oleh model ML konsisten dengan hasil yang disajikan dalam literatur medis. Keterbatasan aplikasi yang ada adalah penggunaan set data yang tidak seimbang yang rentan terhadap bias seleksi.

Negara-negara di seluruh dunia telah terkena dampak virus, mengakibatkan berbagai tindakan diberlakukan, termasuk penguncian negara, jam malam, dan pembatasan perjalanan. Meskipun gejala umum infeksi COVID-19 biasanya ringan, bagi beberapa pasien infeksi dapat menyebabkan komplikasi serius, dan terkadang mematikan. Mengelola jumlah kasus COVID-19 yang melonjak adalah tantangan besar yang membuat fasilitas perawatan kesehatan di seluruh dunia kewalahan; namun, masih ada informasi yang cukup tentang virus tersebut. Sejak munculnya infeksi COVID-19, para peneliti dari berbagai disiplin ilmu telah mengeksplorasi virus baru ini. Machine learning (ML) adalah cabang dari kecerdasan buatan (AI) yang berfokus pada produksi sistem yang mampu belajar dari contoh dan meningkatkan tanpa diprogram secara eksplisit (Burhan et al., 2020). Machine learning (ML) telah berhasil diterapkan di banyak bidang, termasuk perawatan kesehatan (Mendez, 2020) dan informatika medis (Oxford University, 2022). Satu arah penelitian penting memanfaatkan Machine learning (ML) untuk memahami dan melawan COVID-19. Banyak lini penelitian telah dimulai untuk penerapan dan pengembangan algoritme Machine learning (ML) terkait COVID-19.

Telah banyak metode diagnosa yang digunakan untuk mendeteksi seseorang terkena COVID-19 atau tidak. Namun, keakuratan tes dapat bervariasi tergantung pada saat sampel Anda diambil selama perjalanan penyakit Anda. Jika Anda dites terlalu cepat setelah terpapar COVID-19, mungkin tidak ada cukup virus di tubuh Anda untuk mendapatkan hasil yang akurat. Jika ini masalahnya pada saat tes, tes Anda mungkin kembali negatif, bahkan jika Anda benar-benar

memiliki virus. Ini akan dianggap sebagai tes 'negatif palsu'. Penting untuk dipahami bahwa profesional perawatan kesehatan mempertimbangkan sejumlah faktor dalam membuat diagnosis COVID-19.

Untuk meningkatkan kinerja model diagnosa COVID-19 dapat dilakukan dengan menggabungkan algoritma Machine Learning. Salah satu teknik yang digunakan untuk menggabungkan algoritma machine learning adalah Stacking. Stacking adalah salah satu teknik pembelajaran mesin ensemble paling populer yang digunakan untuk memprediksi beberapa node untuk membangun model baru dan meningkatkan kinerja model. Stacking memungkinkan kami melatih beberapa model untuk memecahkan masalah serupa, dan berdasarkan keluaran gabungannya, ini membangun model baru dengan kinerja yang lebih baik. Dalam stacking, sebuah algoritma mengambil luaran dari sub-model sebagai masukan dan mencoba mempelajari cara terbaik untuk menggabungkan prediksi masukan untuk membuat prediksi luaran yang lebih baik.

Model yang diusulkan akan diuji menggunakan dataset sekunder yang diunduh dari Kaggle dan mengukur kinerja model. Dataset gejala (symptoms) COVID-19 diunduh dari Kaggle dengan link <https://www.kaggle.com/datasets/hemanthhari/symptoms-and-covid-presence> terdiri dari 21 kolom (atribut) dan 5434 baris (record).

Stacking juga dikenal sebagai generalisasi bertumpuk dan merupakan bentuk lanjutan dari teknik Model Averaging Ensemble di mana semua sub-model berpartisipasi secara setara sesuai bobot kinerjanya dan membangun model baru dengan prediksi yang lebih baik. Model baru ini ditumpuk di atas yang lain, inilah alasan mengapa dinamai stacking. Arsitektur model stacking dirancang sedemikian rupa sehingga terdiri dari dua atau lebih model dasar pembelajar (base learner) dan model meta yang menggabungkan prediksi model dasar.

2 Metodologi

Pada penelitian ini digunakan pendekatan kuantitatif. Pada pendekatan penelitian kuantitatif, dilakukan analisa kuantitatif secara teliti terhadap beberapa generasi informasi yang berbentuk kuantitatif. Umumnya pendekatan kuantitatif memiliki tiga bentuk yang berbeda, yaitu pendekatan inferensial, pendekatan eksperimental,

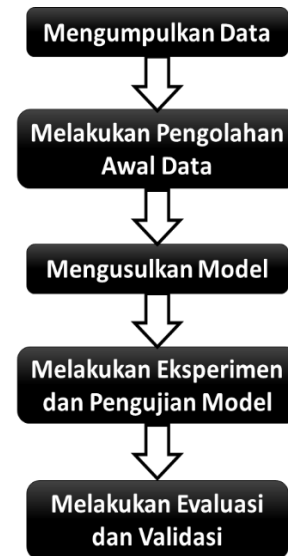
dan pendekatan simulasi. Pada pendekatan inferensial, sampel yang diperoleh digunakan untuk membuat dugaan karakteristik populasi, relasinya, dan lain-lain. Pada pendekatan ini, peneliti tidak memiliki kontrol atas karakteristik, variabel, dan responden yang diteliti. Pendekatan eksperimental ditandai dengan adanya kontrol atas lingkungan penelitian oleh peneliti. Eksperimen adalah suatu proses yang sistematis di mana peneliti memiliki kontrol atas variabel berdasarkan pertimbangan agar sesuai dengan tujuan penelitian. Simulasi berarti operasi model numerik yang mewakili struktur proses dinamis. Pada pendekatan simulasi, lingkungan buatan dibuat di mana informasi yang diperlukan dapat dihasilkan.

Tujuan dari metode kuantitatif adalah untuk dapat memahami bagaimana sesuatu dikonstruksi, bagaimana dibangun, dan bagaimana cara kerjanya. Penelitian kuantitatif umumnya didorong oleh hipotesis, kemudian dibuat rumusan dan pengujian secara ketat untuk menunjukkan bahwa hipotesisnya salah. Sehingga usaha yang dilakukan adalah membuktikan bahwa hipotesis yang dibuat adalah salah, jika hipotesisnya tahan uji, maka hipotesis tersebut dianggap benar. Tetapi jika tidak tahan uji, maka hipotesisnya dianggap salah.

Sudut pandang kuantitatif menekankan bahwa pengukuran merupakan dasar yang dapat digunakan untuk menunjukkan hubungan antara observasi dan formalisasi model, teori, dan hipotesis. Penelitian dan metode kuantitatif akan menghasilkan pengembangan model, teori, dan hipotesis yang berkaitan dengan fenomena alam.

Pendekatan yang akan digunakan pada penelitian ini adalah pendekatan eksperimen. Pada penelitian eksperimen dilakukan dengan cara menginvestigasi hubungan sebab-akibat menggunakan pengujian yang dikontrol oleh peneliti. Pada penelitian semiekperimental sering mendapatkan kendala pada tidak cukupnya akses terhadap sampel, masalah etika, dan sebagainya. Untuk pengembangan, evaluasi, dan pemecahan masalah proyek biasanya dilakukan dengan eksperimen.

Tujuan dari penelitian ini adalah mengembangkan aplikasi Deteksi Dini Anak Autis. Suatu penelitian harus mengikuti aturan yang diakui agar dapat diakui/diterima, maka penelitian ini dilakukan sesuai tahapan yang ditunjukkan pada Gambar 3.



Gambar 1 Tahapan Penelitian

Tahapan pada Gambar 1 dijelaskan sebagai berikut:

a. Pengumpulan data

Pengumpulan data bukan hanya sekedar mengambil data yang ada, tetapi harus mampu mendeskripsikan data yang ada, dan memiliki kontribusi terhadap pengetahuan. Data tersebut harus dapat memberikan penjelasan, hubungan, perbandingan, prediksi, generalisasi, dan teori (Dawson, 2009, p. 18). Berdasarkan sumbernya, data dibedakan menjadi dua, yaitu:

- Data primer, yaitu data yang dikumpulkan langsung dari sumber data. Pengumpulan data ini memerlukan waktu, dan biaya yang lebih banyak dari data sekunder. Contoh sumber data primer adalah kuisioner, observasi, wawancara, dan eksperimen yang dilakukan langsung oleh peneliti.
- Data sekunder, yaitu data yang diperoleh dari peneliti/pihak lain, walaupun data tersebut sebelumnya digunakan dengan tujuan yang berbeda. Data sekunder dapat diperoleh relatif lebih cepat, dan dengan biaya rendah. Contoh sumber data sekunder adalah kantor statistik baik pemerintah maupun swasta, perpustakaan, toko buku, maupun internet.

Beberapa alasan penggunaan data sekunder adalah:

- Karena pengumpulan data primer relatif lebih sulit, memakan waktu, dan mahal.
- Karena ada kemungkinan memiliki data yang tidak cukup.
- Karena lebih masuk akal untuk menggunakan data yang ada, walaupun datanya dalam bentuk yang berbeda.
- Karena lebih dapat menjelaskan, atau pelengkap dari data primer dikumpulkan.
- Karena dapat mengkonfirmasi, memodifikasi atau bertentangan dengan yang ditemukan.
- Karena memungkinkan untuk memusatkan perhatian pada analisis dan interpretasi.
- Karena tidak dapat melakukan penelitian pada sesuatu yang telah dilakukan/terjadi.
- Karena lebih banyak data yang dikumpulkan daripada yang pernah digunakan.

b. Pengolahan awal data

Data yang sudah dikumpulkan diolah menggunakan algoritma soft-computing untuk mengurangi data yang tidak relevan, atau data dengan atribut yang hilang. Pengolahan juga dapat berupa konversi nilai-nilai redundan (berlebihan), atau nilai yang terlalu beragam ke dalam kelompok yang lebih kecil untuk mempermudah pembentukan model.

c. Model/metode yang diusulkan

Untuk menggambarkan alur model/metode yang diusulkan dan menjelaskan cara kerja model/metode yang diusulkan. Model/metode ini digambarkan secara skematik dan disertai dengan formula penghitungan. Model/metode yang diusulkan akan dibentuk dari data yang sudah diolah, dan hasil pengolahan model akan diukur dengan model yang ada saat ini.

d. Eksperimen dan pengujian model

Menjabarkan bagaimana eksperimen yang dilakukan hingga terbentuknya model, serta menjelaskan cara menguji model yang terbentuk.

e. Evaluasi dan validasi hasil

Evaluasi dilakukan dengan mengamati hasil prediksi menggunakan algoritma soft-computing. Validasi dilakukan dengan mengukur hasil prediksi dibandingkan dengan data asli/awal. Pengukuran kinerja dilakukan dengan membandingkan nilai kesalahan (error) hasil prediksi masing-masing algoritma sehingga dapat diketahui algoritma yang lebih akurat.

Sejak munculnya COVID-19, para peneliti dalam pembelajaran mesin dan radiologi telah bergegas mengembangkan algoritme yang dapat membantu diagnosis, triase, dan pengelolaan penyakit (1). Akibatnya, ribuan model diagnostik dan prognostik menggunakan radiografi dada dan CT telah dikembangkan. Namun, tanpa pendekatan standar untuk pengembangan atau evaluasi, sulit, bahkan bagi para ahli, untuk menentukan model mana yang paling bermanfaat secara klinis. Di sini, kami berbagi keprihatinan utama kami dan menyajikan beberapa solusi yang mungkin.

Pandemi ini terus menantang sistem medis di seluruh dunia dalam banyak aspek, termasuk peningkatan tajam dalam permintaan tempat tidur rumah sakit dan kekurangan peralatan medis, sementara banyak petugas kesehatan sendiri telah terinfeksi. Dengan demikian, kapasitas untuk keputusan klinis segera dan penggunaan sumber daya kesehatan yang efektif sangat penting. Tes diagnosis COVID-19 yang paling tervalidasi, menggunakan reverse transcriptase polymerase chainreaction (RT-PCR), telah lama kekurangan di negara berkembang. Hal ini berkontribusi pada peningkatan tingkat infeksi dan penundaan tindakan pencegahan kritis. Skrining yang efektif memungkinkan diagnosis COVID-19 yang cepat dan efisien dan dapat mengurangi beban pada sistem perawatan kesehatan. Model prediksi yang menggabungkan beberapa fitur untuk memperkirakan risiko infeksi telah dikembangkan, dengan harapan dapat membantu staf medis di seluruh dunia dalam melakukan triase pasien, terutama dalam konteks sumber daya perawatan kesehatan yang terbatas. Model-model ini menggunakan fitur-fitur seperti pemindaian tomografi komputer (CT), gejala klinis, tes laboratorium, dan integrasi fitur-fitur ini. Namun, sebagian besar model sebelumnya didasarkan pada data dari pasien yang dirawat di rumah sakit,

sehingga tidak efektif dalam skrining SARS-CoV-2 dalam populasi umum.

Random Forest adalah model terawasi yang mengimplementasikan pohon keputusan dan metode bagging. Idanya adalah bahwa dataset pelatihan disampel ulang sesuai dengan prosedur yang disebut "bootstrap". Setiap sampel berisi subset acak dari kolom asli dan digunakan agar sesuai dengan pohon keputusan. Jumlah model dan jumlah kolom adalah hyperparameter yang akan dioptimalkan.

Akhirnya, prediksi pohon dicampur bersama-sama menghitung nilai rata-rata (untuk regresi) atau menggunakan pemungutan suara lunak (untuk klasifikasi). Ide bagging adalah bahwa, dengan merata-ratakan output dari pohon keputusan tunggal, kesalahan standar berkurang dan begitu juga varians model menurut tradeoff bias-variens. Itu sebabnya Random Forest menjadi sangat terkenal dalam beberapa tahun terakhir.

Setiap pohon di hutan acak dapat menghitung pentingnya fitur sesuai dengan kemampuannya untuk meningkatkan kemurnian daun. Ini adalah topik yang berkaitan dengan cara kerja Classification And Regression Trees (CART). Semakin tinggi peningkatan kemurnian daun, semakin tinggi pentingnya fitur tersebut. Ini dilakukan untuk setiap pohon, kemudian dirata-ratakan di antara semua pohon dan, akhirnya, dinormalisasi menjadi 1. Jadi, jumlah skor kepentingan yang dihitung dengan Hutan Acak adalah 1.

Setelah kami mengetahui pentingnya setiap fitur, kami melakukan pemilihan fitur menggunakan prosedur yang disebut Penghapusan Fitur Rekursif. Dalam artikel ini, saya akan berbicara tentang versi yang menggunakan validasi silang k-fold.

Idanya adalah untuk menyesuaikan model, kemudian menghapus fitur yang kurang relevan dan menghitung nilai rata-rata dari beberapa metrik kinerja di CV. Kemudian kami menghapus fitur penting kedua terakhir, menyesuaikan model lagi dan menghitung kinerja rata-rata. Kami terus melakukan pendekatan ini sampai tidak ada fitur yang tersisa. Kumpulan fitur yang memaksimalkan kinerja di CV adalah kumpulan fitur yang harus kami kerjakan. Harap dicatat bahwa seluruh prosedur harus bekerja dengan nilai yang sama untuk hyperparameter.

2.1 Data Pendukung

Dataset yang digunakan merupakan data sekunder yang diambil dari website Kaggle dengan link

<https://www.kaggle.com/code/midouazerty/symptoms-covid-19-using-7-machine-learning-98/data>.

Tabel 1 Spesifikasi Atribut Dataset

| No. | Atribut | Nilai |
|-----|---|--------|
| 1 | Breathing Problem | Yes/No |
| 2 | Fever | Yes/No |
| 3 | Dry Cough | Yes/No |
| 4 | Sore throat | Yes/No |
| 5 | Running Nose | Yes/No |
| 6 | Asthma | Yes/No |
| 7 | Chronic Lung Disease | Yes/No |
| 8 | Headache | Yes/No |
| 9 | Heart Disease | Yes/No |
| 10 | Diabetes | Yes/No |
| 11 | Hyper Tension | Yes/No |
| 12 | Fatigue | Yes/No |
| 13 | Gastrointestinal | Yes/No |
| 14 | Abroad travel | Yes/No |
| 15 | Contact with COVID Patient | Yes/No |
| 16 | Attended Large Gathering | Yes/No |
| 17 | Visited Public Exposed Places | Yes/No |
| 18 | Family working in Public Exposed Places | Yes/No |
| 19 | Wearing Masks | Yes/No |
| 20 | Sanitization from Market | Yes/No |
| 21 | COVID-19 | Yes/No |

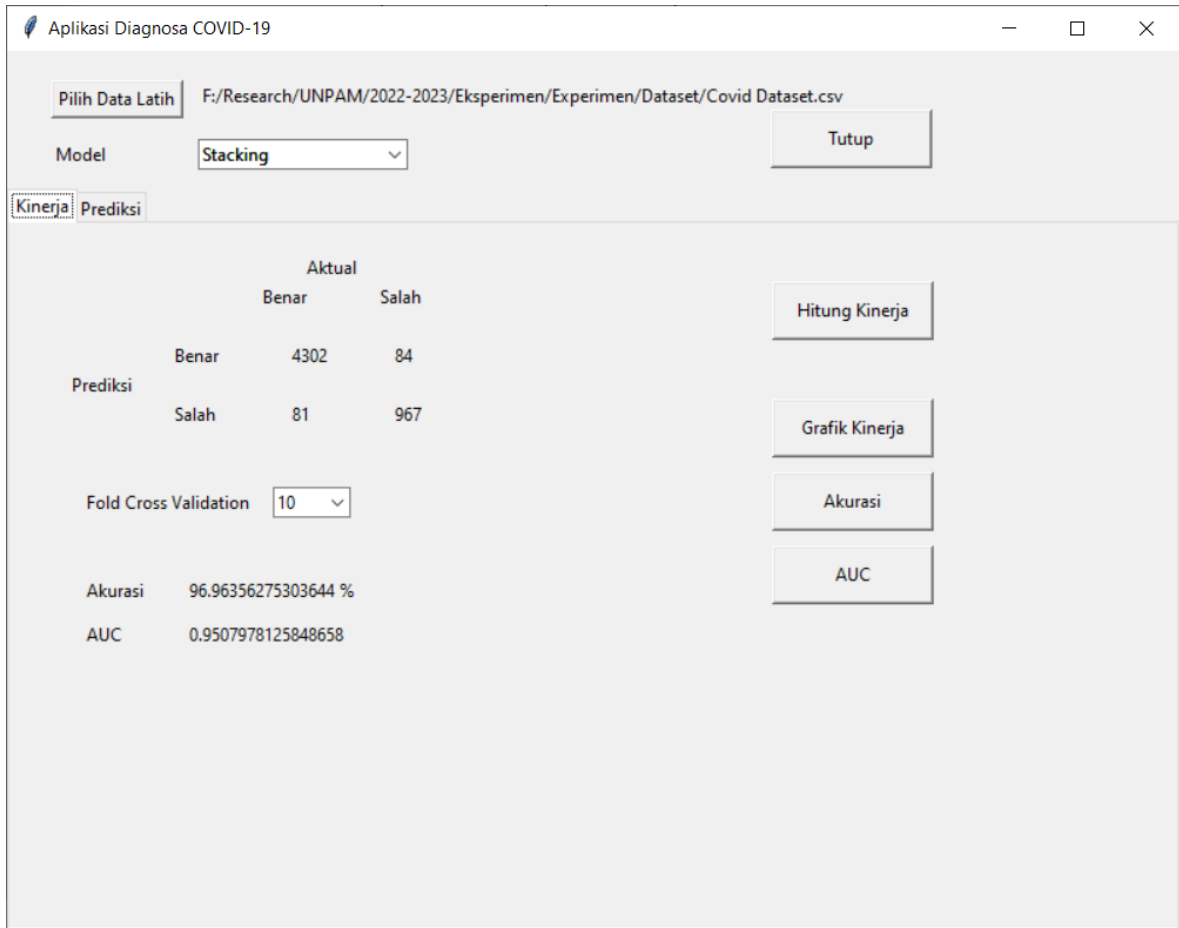
2.2 Model yang Diusulkan

Pada model yang diusulkan digunakan dataset COVID-19 yang telah dikumpulkan. Dataset COVID-19 yang terkumpul dilakukan preprocessing yang mencakup, cleaning, transformasi, dan standardisasi. Dataset akan dibagi menjadi dua bagian menggunakan 10-fold cross validation, yaitu data latih (K-1 fold) dan data uji (1 fold). Data latih digunakan untuk melatih model yang menerapkan KNN, Naïve Bayes, Decision Tree, SVM, Logistic Regression, dan Stacking dari kelima algoritma tersebut. Model yang telah dilatih kemudian divalidasi menggunakan data uji. Hasil validasi digunakan untuk mengukur kinerja model. Kemudian semua

model yang terbentuk dibandingkan untuk mengetahui model terbaiknya.

3 Hasil dan Pembahasan

Pada penelitian ini digunakan bahasa pemrograman Python untuk menerapkan model prediksi/diagnosa COVID-19 yang diusulkan. Tampilan aplikasi ditunjukkan pada Gambar 3.



Gambar 4 Aplikasi Prediksi COVID-19

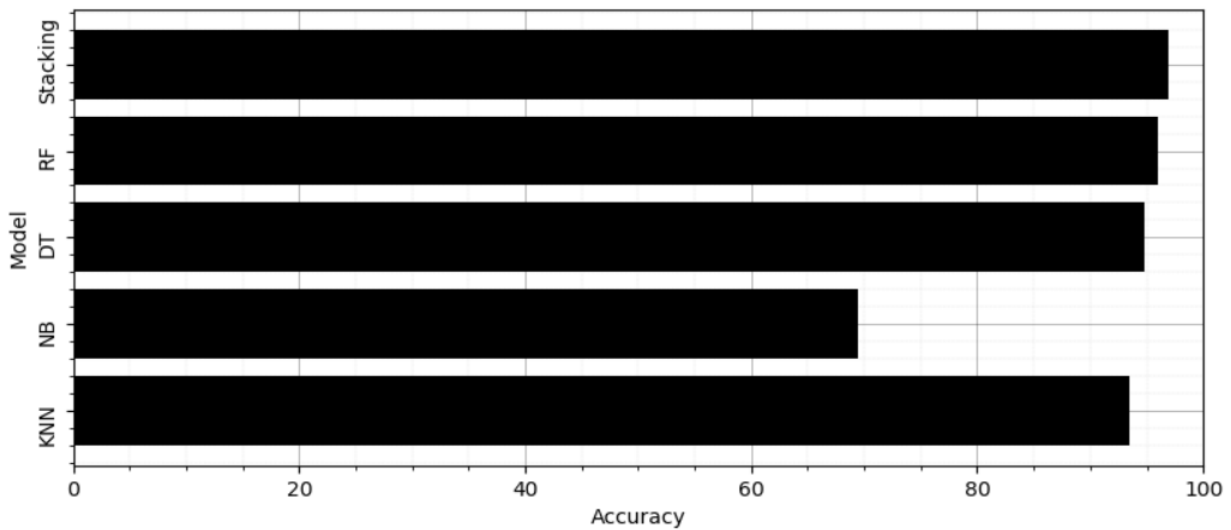
Hasil penerapan model memberikan hasil kinerja seperti pada Tabel 2.

Tabel 2 Kinerja Model

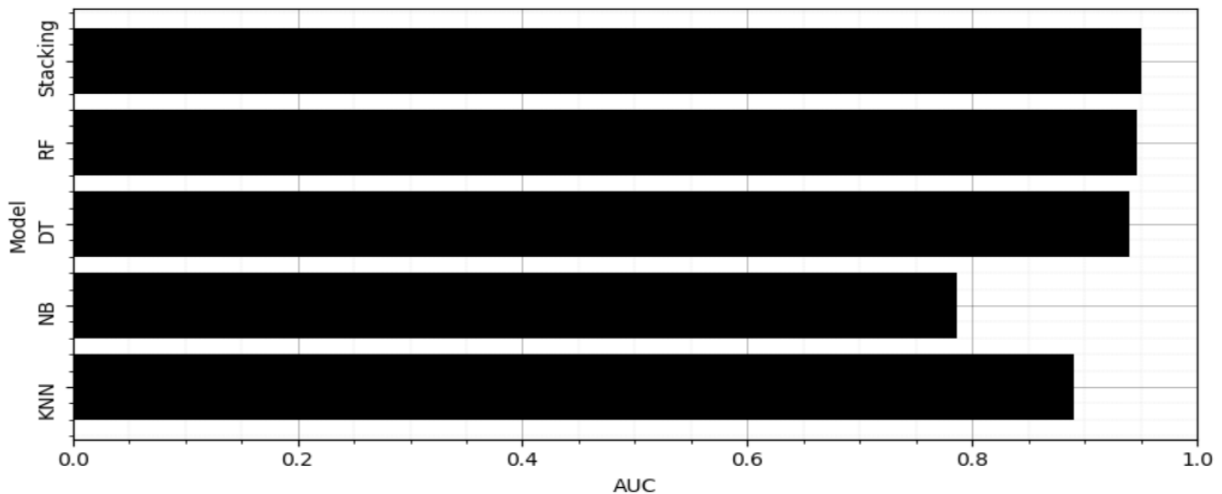
| Model | Akurasi | AUC |
|----------|---------------|---------------|
| KNN | 93,50% | 0,8896 |
| NB | 69,51% | 0,7864 |
| DT | 94,85% | 0,9398 |
| RF | 96,01% | 0,9460 |
| Stacking | 96,96% | 0,9508 |

Kemudian nilai kinerja model divisualisasi menggunakan grafik untuk perbandingannya. Nilai

akurasi ditampilkan pada Gambar 4 dan nilai AUC ditunjukkan pada Gambar 5.



Gambar 4 Akurasi model



Gambar 5 AUC model

Berdasarkan nilai yang didapat, menunjukkan bahwa model yang menerapkan teknik stacking memberikan nilai kinerja yang lebih baik. Maka teknik stacking dianggap dapat meningkatkan kinerja model prediksi/diagnosa COVID-19.

4 Kesimpulan

Berdasarkan hasil eksperimen yang telah dilakukan dapat disimpulkan sebagai berikut:

- Kinerja model prediksi/diagnosa COVID-19 yang menerapkan teknik stacking memberikan nilai akurasi dan AUC lebih baik.

- Model yang menggabungkan algoritma tunggal menggunakan teknik stacking memiliki kinerja yang lebih baik dibandingkan dengan model yang menggunakan algoritma tunggal.

5 Saran

Hasil penelitian menunjukkan bahwa model yang diusulkan dengan menerapkan teknik stacking belum mencapai nilai kinerja yang sempurna, maka disarankan untuk penelitian selanjutnya dapat menerapkan teknik optimasi yang lain. Teknik optimasi yang dapat digunakan misalnya bagging, boosting, atau yang lainnya.

References

- Burhan, E., Susanto, A. D., Nasution, S. A., Ginanjar, E., Pitoyo, W., Susilo, A., & Dkk. (2020). Pedoman Tatalaksana COVID-19. In *Pedoman Tatalaksana COVID-19* (3rd ed.).
- Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C., & Napoli, R. Di. (2022). Features, Evaluation, and Treatment of Coronavirus (COVID-19). *StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–, December 2020*, 1–49.
- Cennimo, D. J. (2024). *Coronavirus Disease 2019 (COVID-19)* (pp. 1–4). Medscape.
- Guo, Y.-R., Cao, Q.-D., Hong, Z.-S., Tan, Y.-Y., Chen, S.-D., Jin, H.-J., Tan, K.-S., Wang, D.-Y., & Yan, Y. (2020). The Origin, Transmission and Clinical Therapies on Coronavirus Disease 2019 (COVID-19) Outbreak – An Update on the Status. *Military Medical Research*, 7(11), 1–10. <https://doi.org/10.1186/s40779-020-00240-0>
- McIntosh, K., Hirsch, M. S., & Bloom, A. (2020). Coronavirus Disease 2019 (COVID-19). *UpToDate*, 1–27. https://www.cmim.org/PDF_covid/Coronavirus_disease2019_COVID-19_UpToDate2.pdf
- Mendez, C. M. (2020). Solidarity” Clinical Trial for COVID -19 Treatments. *Brazilian Journal of Impleantology and Health Sciences*, 1–6.
- Oxford University. (2022). *The RECOVERY Trial - two years on* (pp. 1–2). Oxford University.
- Sahin, A. R., Erdogan, A., Mutlu Agaoglu, P., Dineri, Y., Cakirci, A. Y., Senel, M. E., Okyay, R. A., & Tasdogan, A. M. (2020). 2019 Novel Coronavirus (COVID-19) Outbreak: A Review of the Current Literature. *Eurasian Journal of Medicine and Oncology*, 4(1), 1–7. <https://doi.org/10.14744/ejmo.2020.12220>
- World Health Organization. (2020). Laboratory Testing for 2019 Novel Coronavirus (2019-nCoV) in Suspected Human Cases. In *WHO - Interim guidance* (pp. 1–7). World Health Organization. <https://www.who.int/publications/i/item/WHO-2019-nCoV-lab-testing-2021.1-eng>

