

Komparasi Algoritma *Machine Learning* untuk Klasifikasi Gejala *Coronavirus Disease 19 (Covid-19)*

Musriatun Napiah^{1*}, Rachmawati Darma Astuti², Eka Kusuma Pratama³

^{1,2}Program Studi Ilmu Komputer, Fakultas Teknik dan Informatika Universitas Bina Sarana Informatika
Jl. Kramat Raya No.98, Jakarta Pusat, DKI Jakarta 10450, Indonesia

²Program Studi Teknologi Komputer, Fakultas Teknik dan Informatika Universitas Bina Sarana Informatika
Jl. Kramat Raya No.98, Jakarta Pusat, DKI Jakarta 10450, Indonesia

e-mail: 1musriatun.mph@bsi.ac.id, 2rachmawati.rcd@bsi.ac.id, 3eka.eem@bsi.ac.id

Artikel Info : Diterima : 25-05-2023 | Direvisi : 26-06-2023 | Disetujui : 21-07-2023

Abstrak - COVID-19 atau *Corona Virus Disease 19* adalah anggota keluarga besar virus corona yang menyebabkan spektrum penyakit dari ringan hingga parah, termasuk MERS dan SARS. Meskipun penyebab penularan COVID-19 belum dipastikan, diyakini bahwa virus ini ditularkan dari hewan ke manusia, menyebabkan berbagai gejala seperti batuk, pilek, demam, sakit tenggorokan, dan kehilangan penciuman. Penelitian dilakukan untuk mengklasifikasikan gejala COVID-19 menjadi kategori rendah, sedang, dan tinggi pada pasien. Penelitian ini bertujuan mengelompokkan data pasien dan menentukan risiko infeksi COVID-19 berdasarkan tingkat keparahan gejala, yaitu ringan, sedang, dan tinggi. Metode pembelajaran mesin, termasuk algoritma Decision Tree dan SVM, diperkenalkan dan dibandingkan dengan K-Nearest Neighbor (K-NN), Neural Network (NN), Random Forest (RF), dan Naive Bayes. Dataset yang digunakan berisi 127 catatan pasien dari kaggle.com. Hasil pengujian menunjukkan bahwa SVM mencapai akurasi 54%, sedangkan Decision Tree mencapai 98%. Penelitian ini memberikan wawasan penting tentang penilaian risiko infeksi COVID-19 berdasarkan tingkat keparahan gejala, dan penggunaan teknik pembelajaran mesin diharapkan meningkatkan kemampuan analisis dan prediksi dalam menghadapi pandemi COVID-19.

Kata Kunci : *Corona Virus Desiase19*, Klasifikasi, *Decision Tree*, *Machine Learning*, dan SVM

Abstracts - COVID-19 or *Corona Virus Disease 19* is a member of the extended family of coronaviruses that cause a spectrum of illnesses from mild to severe, including MERS and SARS. While the cause of COVID-19 transmission has not been confirmed, it is believed that the virus is transmitted from animals to humans, causing various symptoms such as cough, runny nose, fever, sore throat and loss of smell. Research was conducted to classify COVID-19 symptoms into low, medium, and high categories in patients. This study aims to classify patient data and determine the risk of COVID-19 infection based on the severity of symptoms, namely mild, moderate, and high. Machine learning methods, including Decision Tree and SVM algorithms, are introduced and compared with K-Nearest Neighbor (K-NN), Neural Network (NN), Random Forest (RF), and Naive Bayes. The dataset used contains 127 patient records from kaggle.com. The test results showed that SVM achieved 54% accuracy, while Decision Tree achieved 98%. This research provides important insights into the risk assessment of COVID-19 infection based on symptom severity, and the use of machine learning techniques is expected to improve analysis and prediction capabilities in the face of the COVID-19 pandemic.

Keywords : *Corona Virus Desiase19*, Classification, *Decision Tree*, *Machine Learning*, and SVM

PENDAHULUAN

COVID-19 adalah bagian dari keluarga besar virus yang menjadi penyebab berbagai penyakit, mulai dari yang ringan seperti common cold atau radang selaput lendir hingga penyakit serius seperti MERS dan SARS. Virus ini dapat ditularkan dari hewan ke manusia, dan penularan antarmanusia terjadi dalam tingkat yang sangat terbatas. Namun, hingga saat ini, mekanisme penularan COVID-19 masih belum sepenuhnya dipahami dengan pasti. Banyak yang beropini virus ini berasal dari Wuhan, China dan bahwa penularannya dari hewan ke manusia



karena banyak kasus yang terjadi disana (PH, Suwoso, Febrianto, Kushindarto, & Aziz, 2020). Dengan munculnya gejala-gejala seperti tenggorokan kering, suhu panas tinggi, penciuman terganggu, Organisasi Kesehatan Dunia (WHO) menghadapi tantangan tahun Covid-19 ini dengan menerapkan langkah-langkah proteksi dasar yang dapat diambil secara mandiri. Langkah-langkah tersebut meliputi membersihkan tangan secara rutin menggunakan alkohol atau sabun dan air, menjaga jarak aman dari orang yang mengalami gejala batuk atau bersin, menerapkan etika batuk atau bersin, dan mencari perawatan medis jika mengalami gejala yang masuk dalam kategori suspek. Selain itu, WHO juga menyarankan untuk sementara menghindari kerumunan dan tetap tinggal di rumah untuk menjaga kesehatan dan keselamatan keluarga, karena keluarga memiliki peran yang sangat penting dalam situasi ini. (Sari, Ayu Riana et al., 2020). Pasien yang mengalami penyakit jantung dan gangguan metabolik kronis, serta mengalami peradangan akut dan penurunan fungsi organ seperti jantung, ginjal, hati, dan hematologi sejak awal perawatan, berisiko lebih tinggi mengalami kematian akibat infeksi Covid-19. (Satria, Tutupoho, & Chalidyanto, 2020). Perlu adanya penanganan khusus yang dilakukan oleh pemerintah ataupun masyarakat untuk menghadapi wabah ini agar tidak terus berlarut semakin parah.

Orang Dalam Pemantauan, Pasien Dalam Pengawasan, Orang Tanpa Gejala, dan Positif merupakan penamaan kategori sebagai stigma orang positif covid-19 berdasarkan panduan WHO dan penelitian (Abudi, Mokodompis, & Magulili, 2020). Banyak dampak yang terasa sebagai pencetus masalah pikososial yaitu makin tingginya prevelensi terhadap kasus terkonfirmasi positif di berbagai wilayah secara cepat, maraknya informasi yang tidak *valid* dan keberhasilan program perawatan dan pengobatan pasien positif Covid-19 yang tidak pasti maupun ketidakjelasan kapan pandemi ini selesai, secara langsung memiliki andil terhadap sumber ancaman menjadikan masyarakat menjadi mudah gelisah, cemas, panik, stres, merasa tertekan, takut dan beraksi maladaptif menghadapi situasi pandemi seperti ini. Dalam penelitian yang dilakukan oleh Tiodora, ditemukan bahwa golongan berisiko tinggi terinfeksi Covid-19 adalah mereka yang memiliki penyakit bawaan, perokok, dan lansia (Siagian, 2020). Penggunaan masker masih menjadi perdebatan yang luas sejak awal pandemi Covid-19, dan dianjurkan untuk menerapkan physical distancing dengan minimal jarak 1,8 meter atau 6 kaki dengan orang lain atau pasien serta mengurangi dalam interaksi aktif kurang dari 30 menit. (Atmojo et al., 2020).

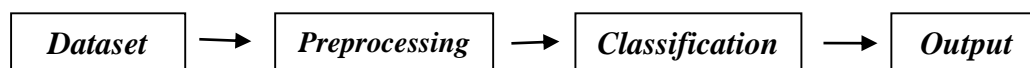
Pengamatan manual untuk melakukan diagnosis terhadap penderita apakah mengalami gejala ringan, sedang, atau tinggi dengan masing-masing gejala yang dihadapi kurang efektif dalam hal ini untuk dilakukan, maka dari itu dengan adanya perkembangan teknologi seperti *machine learning* maka bisa dilakukan diagnosis untuk mengklasifikasi gejala yang dihadapi penderita dengan metode – metode pengujian yang ada. *Machine Learning* bisa digunakan sebagai salah satu sarana yang cocok untuk pemantauan, pengelolaan berkelanjutan, dalam pembuatan kebijakan (Pritalia, 2022)

Dari uraian diatas, penelitian ini mencoba melakukan pembaruan untuk menerapkannya pada eksperimen dengan menggunakan dua metode algoritma baru yaitu *Decision Tree* dan SVM, alasan penggunaan metode *Decision Tree* dan metode SVM adalah karena memiliki kelebihan dalam generalisasi data yang tinggi, dan dapat menghasilkan model klasifikasi yang baik meskipun dilatih dengan data yang relatif sedikit. (Singgalen et al., 2023), sehingga diharapkan dapat memberikan akurasi yang lebih baik dari penelitian yang pernah dilakukan sebelumnya (Angraini, Akbar, Wijaya, Syaputra, & Sobri, 2021) untuk mengklasifikasi Gejala *Coronavirus Disease 19*.

Dalam pengujian metode yang diterapkan menggunakan software Jupyter Notebook untuk mempermudah menganalisis sehingga memberikan hasil yang terbaik, sebelum masuk ke tahapan klasifikasi data dibagi menjadi 3 kategori gejala yaitu gejala rendah, gejala sedang, dan gejala tinggi. Hasil pengujian dengan metode *Support Vector Machine* (SVM) memberikan nilai akurasi sebesar 54%., sedangkan dengan menggunakan *Decision Tree* yang merupakan algoritma terbaik memberikan nilai akurasi sebesar 98%.

METODE PENELITIAN

Penelitian ini menggunakan metode algoritma *machine learning* untuk mengklasifikasikan data pasien yang terinfeksi Covid19 dengan mencari pola data pada *dataset* yang diambil dari *kaggle.com*.



Sumber: Penelitian (2023)
Gambar 1. Metode Penelitian

1. Pengumpulan Data

Data yang di diperoleh dalam penelitian ini bersumber dari situs web yang bisa diakses secara umum, alasan menggunakan dataset ini karena data tersebut baru beberapa kali dikembangkan sehingga peneliti ingin mencoba untuk mengembangkannya dengan memaparkan tentang Covid-19.

Dataset Covid-19 Patient Symptoms diperoleh dari negeri India dan berisi 127 data pasien yang terinfeksi Covid-19. Dataset ini dapat diunduh dari situs web <https://www.kaggle.com/bitsofishan/covid19-patient-symptoms> dan telah dikelompokkan ke dalam tiga kategori klasifikasi berdasarkan penelitian sebelumnya (Anggraini et al., 2021), yaitu tingkat rendah (low risk), tingkat sedang (medium risk), dan tingkat tinggi (high risk).

a. Kategori Gejala Rendah:

Gejala ini paling umum terlihat pada pasien yang tergolong positif Covid-19 dalam klasifikasi ini. Terdapat 5 gejala, termasuk suhu tubuh ≥ 39 C, batuk kering, sakit tenggorokan, perkembangan gejala, dan perubahan pada nafsu makan. Pada dataset, sekitar 0,27% pasien melaporkan tiga gejala yang sama.

b. Kategori Gejala Sedang:

Gejala ini paling sering muncul pada data pasien yang tergolong Covid-19 dalam klasifikasi ini. Terdapat beberapa gejala, termasuk suhu tubuh pasien ≥ 39 C, batuk kering, sakit tenggorokan, rasa kantuk, dan memiliki riwayat perjalanan. Pada dataset, sekitar 0,14% pasien melaporkan empat gejala secara bersamaan.

c. Kategori Gejala Tinggi:

Dalam dataset, sekitar 0,16% pasien melaporkan gejala tinggi, termasuk suhu tubuh ≥ 39 C, batuk kering, sakit tenggorokan, kesulitan dalam pernapasan, dan memiliki riwayat perjalanan.

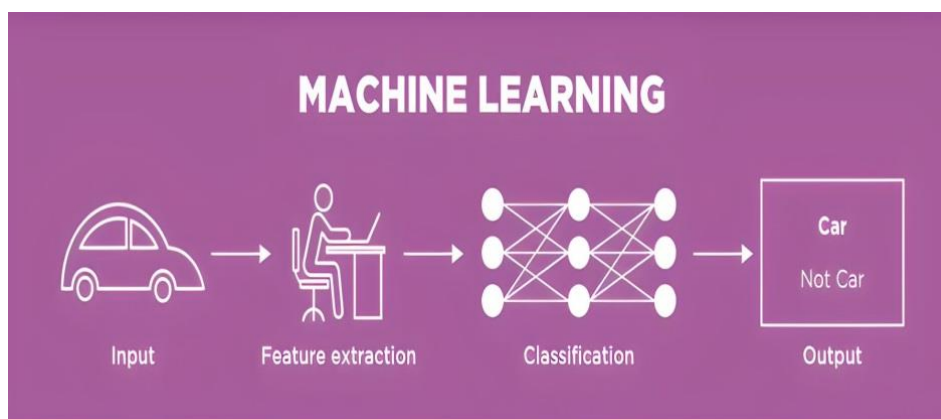
2. Pre-Processing

Preprocessing merupakan salah satu tahapan pengolahan data. jika data tersebut memiliki berbagai permasalahan yang mengganggu hasil pengolahan data yang sering tidak lengkap dan bentuknya tidak beraturan. Pre-processing sebagai proses data akan diolah sebelum dilakukannya analisis (Diwandanu & Wisudawati, 2023). Pada tahap ini, data mentah akan diubah menjadi format yang lebih user-friendly. Proses ini menjadi langkah awal sebelum memasuki tahap klasifikasi.

3. Arsitektur Algoritma

a. Machine Learning

Machine learning menjadi aspek yang semakin krusial dalam bidang ilmu data yang terus berkembang. Proyek-proyek penambangan data menggunakan teknik statistik untuk mengajarkan mesin cara mengklasifikasikan data, memberikan prediksi, dan mengungkap pola yang tersembunyi. (Qalam, Keagamaan, Bioimaging, Kunci, & Learning, 2023). Machine learning adalah bentuk kecerdasan buatan yang bertujuan untuk menggantikan atau meniru cara manusia dalam memecahkan masalah dan otomatisasi. Konsep utama dari machine learning adalah meniru cara belajar dan generalisasi yang dimiliki oleh manusia atau makhluk cerdas lainnya. (Napiah, Purnama, Raharjo, & Bismi, 2022).



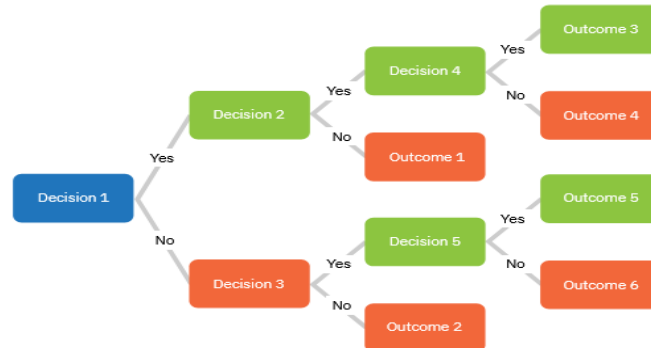
Sumber: (Napiah et al., 2022)

Gambar 2. Decision Tree

b. Decision Tree

Decision Tree adalah sebuah struktur pohon dalam bentuk flowchart yang menggunakan node internal

untuk menggambarkan pengujian atribut dan cabang-cabangnya merepresentasikan hasil dari pengujian tersebut. Node daun pada pohon ini menggambarkan kelas atau distribusi kelas. Root node adalah node paling atas yang tidak memiliki sisi masuk tetapi memiliki beberapa sisi keluar. Internal node memiliki satu sisi masuk dan beberapa sisi keluar, sementara leaf node hanya memiliki satu sisi masuk tanpa sisi keluar. (Qadrini L, Sepperwali A, & Aina A, 2021).



Sumber: Penelitian (2023)
Gambar 3. *Decission Tree*

c. *SVM (Support Vector Machine)*

Support Vector Machine (SVM) adalah sebuah metode yang bertujuan untuk mencari hyperplane optimal yang dapat memisahkan dua kelas dalam ruang input (input space). Metode klasifikasi SVM menggunakan data training untuk membentuk model klasifikasi, dan model tersebut digunakan untuk memprediksi kelas data baru yang belum pernah ada sebelumnya, yang dikenal sebagai *data testing*. (Suryati, Ari Aldino, Penulis Korespondensi, & Suryati Submitted, 2023). SVM dapat diterapkan dalam kasus klasifikasi dan regresi, namun penggunaannya sebagian besar lebih sering untuk tujuan klasifikasi. Algoritma ini bekerja dengan membuat garis terbaik atau batas keputusan yang disebut sebagai hyperplane, yang membagi ruang n-dimensi menjadi kategori-kategori sehingga memudahkan dalam menentukan kategori data baru (Metode et al., 2023).

4. Proses Training

Proses training adalah suatu tahap dimana untuk mendapati hasil akurasi lebih tinggi dari hasil klasifikasi data yang dilakukan pada penelitian.

5. Proses Testing

Proses testing adalah suatu proses klasifikasi menggunakan bias dan bobot hasil pelatihan data. Ini merupakan langkah terakhir dalam penelitian untuk menguji keakuratan klasifikasi dengan menggunakan nilai indeks yang dihasilkan oleh model CNN yang telah dilatih.

HASIL DAN PEMBAHASAN

Hasil pengujian yang dilakukan menggunakan Jupyter Notebook dari dataset yang terdiri dari 127 pasien menunjukkan adanya tiga label klasifikasi, yaitu tingkat rendah (low risk), tingkat sedang (medium risk), dan tingkat tinggi (high risk), dengan menggunakan dua algoritma yakni Decision Tree dan SVM. Sementara itu, dalam penelitian sebelumnya, empat algoritma klasifikasi telah digunakan, yaitu k-nearest neighbor, random forest, neural network, dan naive bayes, pada dataset yang sama berjumlah 127 pasien..

1. *Decission Tree*

Pada algoritma ini ditahap awal testing dengan melakukan import pada *numpy.matplot*, *pandas*, dll, dengan menggunakan data sebanyak 127 pasien hasil akurasi yang di dapat adalah sebesar 98 % . Dapat dilihat pada Gambar 5.

```
print('==== Report ====')
print(classification_report(dataset['RESULT'],y_predict,target_names=target_names))

print('==== Confution Matrix ====')
print(confusion_matrix(dataset['RESULT'],y_predict))

print("==== Cohen Kappa ====")
print([cohen_kappa_score(dataset['RESULT'],y_predict)])
```

	precision	recall	f1-score	support
NORMAL	0.97	1.00	0.99	33
SATDIUM 1	0.99	0.99	0.99	69
STADIUM 2	1.00	0.96	0.98	25
accuracy			0.98	127
macro avg	0.99	0.98	0.98	127
weighted avg	0.98	0.98	0.98	127

Sumber: Penelitian (2023)

Gambar 4. Model Architecture Decission Tree

2. SVM

Setelah data diuji dengan menggunakan metode *decision tree*, maka untuk metode selanjutnya dengan menggunakan metode *svm*, data sebanyak 127 pasien hasil akurasi yang di dapat adalah sebesar 54 %.

```
print('==== Report ====')
print(classification_report(dataset['RESULT'],y_predict,target_names=target_names))

print('==== Confution Matrix ====')
print(confusion_matrix(dataset['RESULT'],y_predict))

print("==== Cohen Kappa ====")
print([cohen_kappa_score(dataset['RESULT'],y_predict)])
```

	precision	recall	f1-score	support
NORMAL	0.00	0.00	0.00	33
SATDIUM 1	0.54	1.00	0.70	69
STADIUM 2	0.00	0.00	0.00	25
accuracy			0.54	127
macro avg	0.18	0.33	0.23	127
weighted avg	0.30	0.54	0.38	127

Sumber: Penelitian (2023)

Gambar 5. Model Architecture SVM

Dari pengujian yang sudah dilakukan dengan kontribusi penelitian penambahan 2 metode baru yaitu *Decision Tree* dan *SVM*, dimana lebih unggul metode *Decission Tree* dengan akurasi sebesar 98% .

Tabel 1. Hasil Penelitian

Metode Penelitian	Hasil Akurasi
<i>Decission Tree</i>	98%
SVM	54%

Sumber: Penelitian (2023)

KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan dengan menggunakan *Dataset Covid-19 Patient Symptons* dari India yang dikumpulkan berjumlah 127 data pasien menggunakan 3 kelas *dataset*, tingkat rendah (*low risk*), tingkat sedang (*medium risk*), dan tingkat tinggi (*high risk*) yang didapat dari *kaggle*, dengan menambahkan 2

metode baru yaitu algoritma *Decision Tree* dan SVM dimana dari penelitian sebelumnya menggunakan 4 metode yaitu K-Nearest Neighbor (K-NN), Neural Network (NN), Random Forest (RF), dan Naive Bayes.

Pada penelitian ini juga dilakukan proses *pre-processing* dengan membagi data menjadi 2 bagian yaitu data *training* dan *testing* yang kemudian di klasifikasi dengan menggunakan metode *Decision Tree* dan *SVM*. Hasil akurasi yang diperoleh untuk metode *Decision Tree* sebesar 98 % lebih tinggi dibandingkan dengan menggunakan metode SVM hanya memperoleh nilai akurasi sebesar 54 %.

Hasil penelitian ini melengkapi hasil penelitian yang sudah dilakukan sebelumnya (Angraini et al., 2021) nilai akurasi tertinggi 73,68% dengan menggunakan metode *neural network*. Maka dapat disimpulkan bahwa hasil dari penelitian ini lebih baik yaitu dengan menghasilkan akurasi tertinggi menggunakan metode *Decision Tree* sebesar 98%.

Untuk penelitian selanjutnya dapat melakukan proses *pre-processing* yang lain untuk menghilangkan *noise* pada data, aplikasi yang lebih ringan, dan bisa juga melakukan eksperimen dengan metode yang lain untuk meningkatkan hasil akurasi.

REFERENSI

- Abudi, R., Mokodompis, Y., & Magulili, A. N. (2020). Stigma Terhadap Orang Positif Covid-19. *Jambura Journal of Health Sciences and Research*, 2(2), 77–84. <https://doi.org/10.35971/jjhsr.v2i2.6012>
- Angraini, S., Akbar, M., Wijaya, A., Syaputra, H., & Sobri, M. (2021). Klasifikasi Gejala Penyakit Coronavirus Disease 19 (COVID-19) Menggunakan Machine Learning. *Journal of Software Engineering Ampera*, 2(1), 57–68. <https://doi.org/10.51519/journalsea.v2i1.105>
- Atmojo, J. T., Iswahyuni, S., Rejo, R., Setyorini, C., Puspitasary, K., Ernawati, H., ... Mubarak, A. S. (2020). Penggunaan Masker Dalam Pencegahan Dan Penanganan Covid-19: Rasionalitas, Efektivitas, Dan Isu Terkini. *Avicenna : Journal of Health Research*, 3(2), 84–95. <https://doi.org/10.36419/avicenna.v3i2.420>
- Diwandanu, M. T., & Wisudawati, L. M. (2023). ANALISIS SENTIMEN TERHADAP TWIT MAXIM PADA TWITTER MENGGUNAKAN R PROGRAMMING DAN K NEAREST NEIGHBORS. *Jurnal Ilmiah Informatika Komputer*, 28(1): 1-16., 28, 1–16.
- Metode, I., Vector, S., Svm, M., Citra, P., Pertiwi, A. A., & Utomo, A. N. (2023). *Jurnal Rekayasa Informasi*, Vol . 12 No . 1 April 2023 *FUNDUS RETINA MATA UNTUK DETEKSI PENYAKIT GLAUKOMA IMPLEMENTATION OF SUPPORT VECTOR MACHINE (SVM) METHOD ON FUNDUS IMAGE OF RETINA OF THE EYE FOR THE DETECTION OF GLAUCOMA*. 12(1), 19–27.
- Napiah, M., Purnama, R. A., Raharjo, M., & Bismi, W. (2022). Komparasi Algoritma Untuk Klasifikasi Penyakit Ispa (Infeksi Saluran Pernapasan Akut). *Jurnal Infortech*, 4(2), 105–110.
- PH, L., Suwoso, R. H., Febrianto, T., Kushindarto, D., & Aziz, F. (2020). Dampak Pandemi Covid-19 bagi Perekonomian Masyarakat Desa. *Indonesian Journal of Nursing and Health Sciences*, 1(1), 37–48. <https://doi.org/10.37287/ijnhs.v1i1.225>
- Pritalia, G. L. (2022). *Analisis Komparatif Algoritme Machine Learning pada Klasifikasi Kualitas Air Layak Minum*. 2(1), 43–55.
- Qadrini L, Sepperwali A, & Aina A. (2021). Decision Tree dan Adaboostpada Klasifikasi Penerima Program Bantuan Sosial. *Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial*, 2(7), 1959–1966.
- Qalam, A., Keagamaan, J. I., Bioimaging, V., Kunci, K., & Learning, M. (2023). *VISUALISASI BIOIMAGING DENGAN MENGGUNAKAN MACHINE Fadhriz Qadrul Amien Faculty of Engineering Universitas Indonesia Basari Faculty of Engineering Universitas Indonesia Abstrak*. 17(1), 98–111.
- Sari, Ayu Riana, F. R., Wulandari, A., Pujianti, N., Laily, N., Vina, Anhar, Y., ... Muddin, F. I. I. (2020). Perilaku Pencegahan Covid-19 Ditinjau dari Karakteristik Individu dan Sikap Masyarakat. *Jurnal Penelitian Dan Pengembangan Kesehatan Masyarakat Indonesia*, 1(128), 32–37.
- Satria, R. M. A., Tutupoho, R. V., & Chalidyanto, D. (2020). Analisis Faktor Risiko Kematian dengan Penyakit Komorbid Covid-19. *Jurnal Keperawatan Silampari*, 4(1), 48–55. <https://doi.org/10.31539/jks.v4i1.1587>
- Siagian, T. H. (2020). Corona Dengan Discourse Network Analysis. *Jurnal Kebijakan Kesehatan Indonesia*, 09(02), 98–106.
- Singalen, Y. A., Sudirman, J. J., 51, N., Semanggi, K., Setiabudi, K., Kota, J., ... Jakarta, I. (2023). Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM. *Journal of Information System Research*, 4(2), 614–625. <https://doi.org/10.47065/josh.v4i2.2999>
- Suryati, E., Ari Aldino, A., Penulis Korespondensi, N., & Suryati Submitted, E. (2023). *Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)*. 4(1), 96–106.