



INNOVATIVE: Journal Of Social Science Research

Volume 5 Nomor 4 Tahun 2025 Page 8805-8817

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Studi Normalisasi pada Pemodelan Klasifikasi
Kondisi Pasien COVID-19 Berdasarkan Data Kualitas Udara
(Studi Kasus: Gradient Boosting, Extra Trees, dan Logistic Regression)

Alfa Krisnanugrah Sakalaty^{1✉}, Alz Danny Wowor²

Universitas Kristen Satya Wacana

Email: 672021258@student.uksw.edu^{1✉}

Abstrak

Penelitian ini meneliti pengaruh penerapan normalisasi data terhadap kinerja model klasifikasi kondisi pasien COVID-19 berbasis data kualitas udara. Data diperoleh dari wilayah Lubang Buaya dan diolah menggunakan metode Min-Max Normalization untuk meningkatkan ketepatan prediksi. Tiga algoritma machine learning Logistic Regression, Gradient Boosting, dan Extra Trees diuji dalam mengklasifikasikan lima kategori kondisi pasien. Hasil penelitian mengungkap bahwa normalisasi mampu meningkatkan akurasi seluruh model secara signifikan, dengan Extra Trees mencatat kinerja tertinggi hingga 92,86% pada kategori Suspek. Meski demikian, kategori Rawat Inap tetap menjadi tantangan akibat ketidakseimbangan distribusi data dan keterbatasan fitur yang tersedia. Secara keseluruhan, Extra Trees tampil sebagai model dengan performa terbaik, disusul oleh Logistic Regression dan Gradient Boosting. Temuan ini menegaskan peran penting normalisasi dan strategi penanganan data tidak seimbang dalam meningkatkan akurasi, sekaligus menunjukkan potensi kualitas udara sebagai indikator pendukung prediksi kondisi pasien COVID-19.

Kata Kunci: *Covid-19, Kualitas udara, Machine Learning, Normalisasi Data*

Abstract

This study examines the impact of data normalization on the performance of classification models for predicting COVID-19 patient conditions based on air quality data. The data were obtained from the Lubang Buaya area and processed using the Min-Max Normalization method to improve prediction accuracy. Three machine learning algorithms Logistic Regression, Gradient Boosting, and Extra Trees were tested in classifying five categories of patient conditions. The findings reveal that normalization significantly improved the accuracy of all models, with Extra Trees achieving the highest performance of 92.86% in the Suspect category. However, the Inpatient category remained challenging due to data imbalance and limited available features. Overall, Extra Trees emerged as the best-performing model, followed by Logistic Regression and Gradient Boosting. These results highlight the critical role of normalization and data imbalance handling strategies in improving model accuracy, as well as the potential of air quality as a supporting indicator for predicting COVID-19 patient conditions.

Keywords: *Air Quality, COVID-19, Data Normalization, Machine Learning*

PENDAHULUAN

Virus corona jenis baru yang memicu penyakit COVID-19 pertama kali muncul pada akhir tahun 2019 di Wuhan, Tiongkok, dan sejak itu berkembang menjadi pandemi global[1]. Virus ini termasuk dalam keluarga *coronaviridae*, yang dapat menginfeksi manusia maupun hewan[2]. Beberapa jenis virus corona telah lama dikenal di dunia kesehatan karena kemampuannya menyebabkan penyakit pernapasan seperti radang paru-paru[3]. COVID-19 sendiri dapat menimbulkan komplikasi serius, terutama pada sistem pernapasan, seperti gagal napas akut, pneumonia, dan *acute respiratory distress syndrome* (ARDS). Infeksi ini juga dapat memengaruhi organ lain sehingga berpotensi mengakibatkan kematian[4]. Penularan COVID-19 terjadi melalui percikan (*droplet*) dari mulut atau hidung individu yang terinfeksi ketika bernapas, berbicara, atau batuk. Droplet tersebut dapat masuk ke tubuh orang lain secara langsung melalui inhalasi atau secara tidak langsung ketika menempel pada permukaan benda, lalu tersentuh oleh tangan yang kemudian menyentuh area wajah seperti mulut, hidung, atau mata[5].

Di Indonesia, permasalahan pencemaran udara telah menjadi isu lingkungan yang menonjol, terutama di wilayah perkotaan padat penduduk seperti DKI Jakarta [6]. Aktivitas industri, transportasi, dan kegiatan masyarakat berkontribusi besar terhadap menurunnya kualitas udara[7]. Becchetti et al. (2020) mengidentifikasi dua alasan utama mengapa polusi udara dapat berhubungan dengan penyebaran COVID-19. Pertama, paparan polusi udara yang berkepanjangan dapat menurunkan fungsi paru-paru, sehingga meningkatkan kerentanan terhadap infeksi pernapasan. Kedua, partikel polutan berpotensi menjadi media

pembawa virus yang dapat bertahan di udara untuk jangka waktu tertentu[8]. Sebagai gambaran wilayah penelitian, Kelurahan Lubang Buaya memiliki luas area 372,2 hektare, terdiri atas 12 Rukun Warga (RW) dan 117 Rukun Tetangga (RT), dengan jumlah penduduk 64.200 jiwa yang terbagi dalam 19.165 kepala keluarga.

Untuk mengkaji keterkaitan antara kualitas udara dan penyebaran COVID-19, teknologi *machine learning*, khususnya model klasifikasi, digunakan karena kemampuannya mengolah data multivariat dan hubungan non-linear secara efektif. Tidak seperti metode statistik tradisional[9]. algoritma *machine learning* mampu mengidentifikasi pola tersembunyi antara variabel lingkungan dan kondisi pasien[10]. Namun, keberhasilan model sangat dipengaruhi oleh tahap pra-pemrosesan data, salah satunya melalui teknik normalisasi. Dalam penelitian ini digunakan metode Min-Max Normalization untuk menyamakan skala antar fitur, sehingga model seperti Logistic Regression dan Gradient Boosting dapat bekerja secara optimal tanpa bias akibat perbedaan skala nilai atribut[11]. Pemilihan algoritma dalam penelitian ini mempertimbangkan karakteristik dan keunggulannya. Gradient Boosting Classifier adalah metode *ensemble* yang menggabungkan model lemah secara bertahap untuk membentuk prediktor yang kuat, efektif untuk data kompleks dengan hubungan non-linear[12]. Logistic Regression dipilih karena kesederhanaannya serta kemampuannya memodelkan hubungan linier secara efisien, sehingga cocok dijadikan *baseline model*[13]. Sementara itu, Extra Trees Classifier menggunakan pendekatan berbasis pohon keputusan dengan tingkat keacakan tinggi dalam pemilihan fitur dan titik pembelahan, yang membuatnya cepat, akurat, dan efektif dalam berbagai kasus klasifikasi, termasuk analisis data medis[14].

METODE PENELITIAN

Identifikasi Masalah

Langkah awal yang dilakukan adalah mengidentifikasi permasalahan utama dalam penelitian. Fokus permasalahan yang diangkat adalah bagaimana pengaruh kualitas udara terhadap kondisi pasien COVID-19, serta seberapa akurat algoritma klasifikasi dalam memodelkan hubungan tersebut.

Pengumpulan Data

Tahap ini mencakup proses perolehan data yang dilakukan melalui pengambilan data kualitas udara (PM_{10} , SO_2 , CO , NO_2 , O_3) dan data covid-19 (Suspek, Positif, Rawat Inap, Sembuh, Meninggal) di Kelurahan Lubang Buaya, Jakarta Timur, karena memiliki kepadatan

penduduk tinggi, riwayat kasus COVID-19 signifikan, dan ketersediaan data kualitas udara yang akurat dari stasiun pemantauan setempat. dengan melakukan pendekatan kuantitatif, untuk memanfaatkan data numerik yang menggambarkan fenomena secara objektif. Data yang dikumpulkan terdiri dari kualitas udara dan status kondisi pasien covid-19. Seluruh informasi disajikan dalam bentuk tabel untuk mempermudah proses analisis dan klasifikasi menggunakan algoritma machine learning.

Processing Data

Sebelum dilakukan pemodelan, data terlebih dahulu melalui tahapan pembersihan (cleaning) untuk menghapus entri kosong (null) atau tidak valid yang dapat mengganggu akurasi model. Setelah itu, diterapkan proses normalisasi Min-Max untuk menyetarakan rentang nilai antar fitur, agar model tidak berat sebelah terhadap fitur tertentu[15]. Tahapan ini sangat krusial karena kualitas data sangat memengaruhi performa akhir model klasifikasi.

Penerapan Model

Pada tahap ini, dilakukan implementasi beberapa algoritma machine learning guna melakukan klasifikasi terhadap kondisi pasien. Algoritma yang digunakan dalam penelitian ini meliputi Logistic Regression, Gradient Boosting, dan Extra Trees Classifier. Pemilihan algoritma ini didasarkan pada keunggulannya dalam mengolah data klasifikasi serta kemampuannya menangani data dengan distribusi dan skala yang bervariasi.

Evaluasi Model

Tahapan akhir dalam metode ini adalah mengevaluasi performa masing-masing model dengan menggunakan akurasi. Evaluasi dilakukan untuk membandingkan efektivitas algoritma dalam memprediksi status kondisi pasien, serta untuk mengidentifikasi model dengan kinerja terbaik dalam klasifikasi berdasarkan data kualitas udara[16].

HASIL DAN PEMBAHASAN

Processing Data

Pada proses ini dilakukan pemeriksaan untuk memastikan bahwa data yang diperoleh dari stasiun data Lubang Buaya mempunyai nilai (*non-null*), Langkah ini dilakukan untuk hasil prediksi yang lebih akurat dalam proses pelatihan data pada setiap algoritma, setelah didapat data yang kosong maka akan dilakukan tahap *Cleaning* pada data dengan menghapus data-data yang kosong. Proses *Cleaning* data ini dilakukan untuk menghasilkan

dataset yang bersih yang dapat digunakan pada tahap berikutnya, yaitu tahap pemodelan, di mana informasi bermanfaat akan dikumpulkan[17].

Pada tahap ini juga penulis melakukan normalisasi dengan tujuan untuk menyeimbangkan rentang nilai pada setiap atribut. Ini dilakukan agar hasil yang dihasilkan lebih akurat dan mengurangi waktu komputasi perhitungan pada model. Sehingga nilai atribut pada dataset yang diperoleh memiliki rentang yang seimbang tanpa mengubah informasi dalam dataset tersebut[18].

Tabel 1. Hasil accuracy modelling sebelum normalisasi

Modelling	70%:30%	75%:25%	80%:20%	85%:15%	90%:10%
Logistic Regression	9.52%	8.57%	10.71%	14.29%	7.14%
Gradient Boosting	4.76%	2.86%	7.14%	0.00%	7.14%
Extra Trees	7.14%	11.43%	14.29%	4.76%	7.14%

Tabel 1 memperlihatkan bahwa pada model Logistic Regression, nilai akurasi tertinggi tercapai sebesar 14,29% pada rasio data 85%:15%, sedangkan nilai terendah tercatat sebesar 7,14% pada rasio 90%:10%. Pada model Gradient Boosting, akurasi tertinggi sebesar 7,14% diperoleh pada rasio 80%:20%, sedangkan akurasi terendah yaitu 0,00% terjadi pada rasio 85%:15%. Sementara itu, model Extra Trees mencapai akurasi tertinggi 14,29% pada rasio 80%:20%, dan akurasi terendah 4,76% pada rasio 85%:15%.

Tabel 2. Hasil accuracy modelling sesudah normalisasi

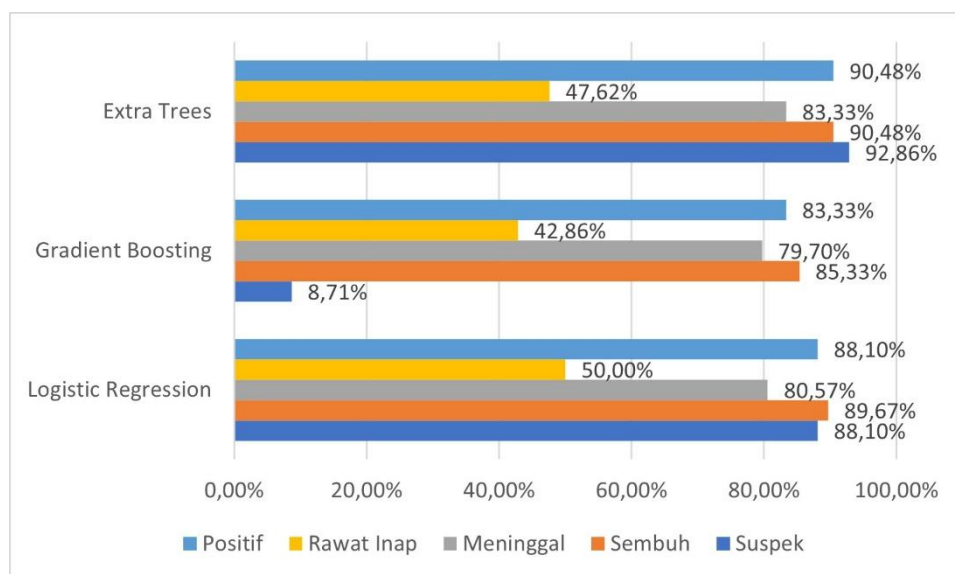
Modelling	70%:30%	75%:25%	80%:20%	85%:15%	90%:10%
Logistic Regression	50.00%	40.00%	39.29%	23.81%	28.57%
Gradient Boosting	42.86%	48.57%	42.86%	38.10%	42.86%
Extra Trees	47.62%	54.29%	46.43%	40.67%	50.00%

Tabel 2 menunjukkan hasil accuracy Setelah di normalisasi, ketiga model mengalami peningkatan akurasi secara signifikan. Model Extra Trees mencatat peningkatan paling signifikan dengan data tertinggi, mencapai 54.29% pada rasio data 75%:25%. Sedangkan Gradient Boosting mencapai akurasi 48.57% dengan rasio 75%:25%, dan Logistic Regression 50.00% dengan rasio data 70%:30%. Hal ini membuktikan bahwa normalisasi data sangat krusial dalam meningkatkan performa model machine learning, terutama dalam data yang memiliki rentang nilai fitur yang sangat bervariasi.

Penerapan Model

Pada proses ini dilakukan penerapan modelling ketiga algoritma klasifikasi pada data yang telah dinormalisasi. Implementasi dilakukan menggunakan algoritma machine learning, tujuan utama penerapan algoritma ini adalah membandingkan efektivitas ketiga model dalam mengklasifikasikan kondisi pasien berdasarkan data kualitas udara dan data pasien covid-19 yang telah dinormalisasi. Evaluasi dilakukan berdasarkan akurasi prediksi terhadap lima target kondisi pasien: Suspek, Positif, Rawat Inap, Sembuh, dan Meninggal. Tabel berikut menunjukkan hasil akurasi dari model Logistic Regression terhadap lima target klasifikasi:

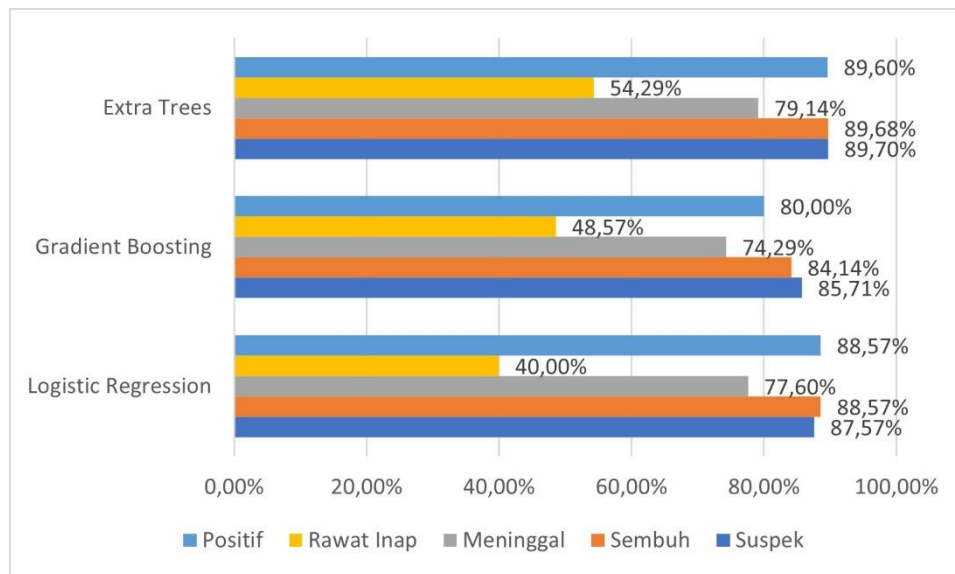
Komparasi Algoritma



Gambar 1. Grafik Algoritma Perbandingan Pembagian Data 70%:30%

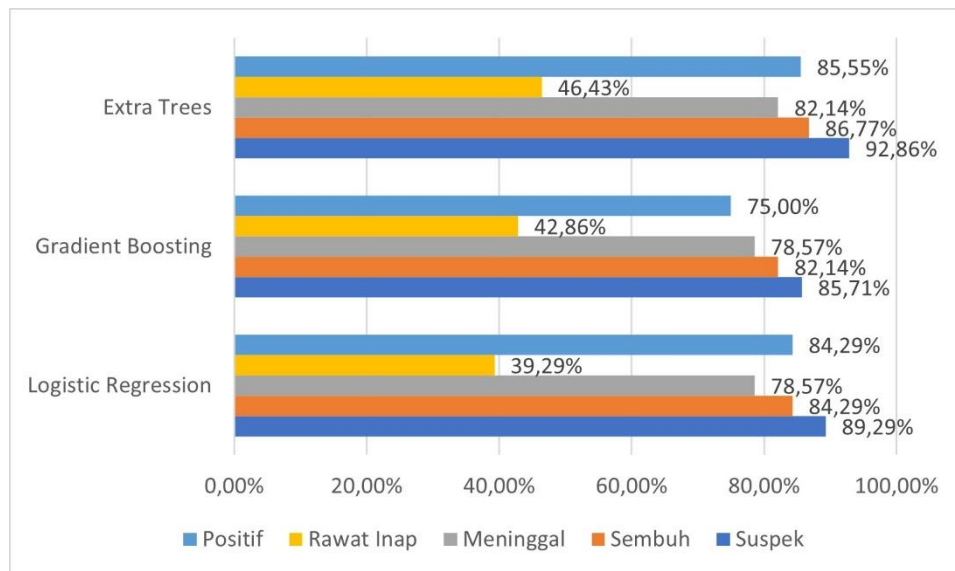
Gambar grafik di atas menampilkan perbandingan persentase keakuratan atau performa tiga algoritma yaitu: Logistic Regression, Gradient Boosting, dan Extra Trees dalam memprediksi beberapa kategori kasus COVID-19 pada pembagian data 70% data latih dan 30% data uji. Kategori yang digunakan adalah: Positif, Rawat Inap, Meninggal, Sembuh dan Suspek. Extra Trees menunjukkan performa paling konsisten dan tertinggi dalam kategori Suspek yaitu 92,86% dan Sembuh yaitu 90,48 % dan performa paling rendah nya terdapat pada kategori rawat inap dengan akurasi 47,62%. Logistic Regression menunjukkan hasil yang relatif stabil di semua kategori, kategori sembuh menunjukkan hasil terbaik 89,67% dan kategori suspek yaitu 88,10%, tetapi rawat inap menunjukkan hasil yang lebih rendah 50%. Gradient Boosting sendiri menunjukkan performa yang tidak stabil,

terutama pada kategori suspek, dengan akurasi yang sangat rendah yaitu 8,71% dan performa terbaik terdapat pada kategori sembuh dengan akurasi 85,33%.



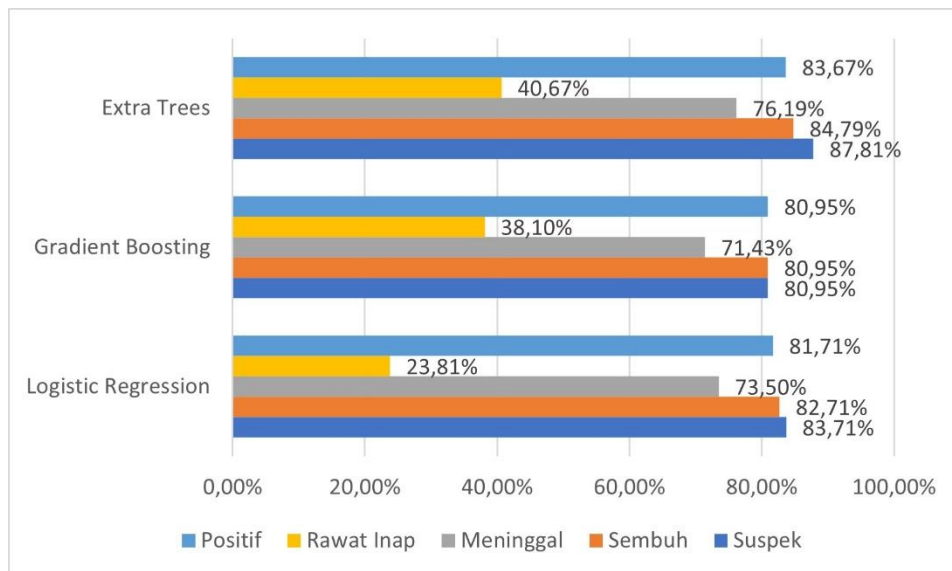
Gambar 2. Grafik Algoritma Perbandingan Pembagian Data 75%:25%

Gambar grafik di atas menampilkan presentase dari ketiga algoritma *machine learning* dengan lima kategori diantaranya Positif, Rawat Inap, Meninggal, sembuh, dan Suspek pada pembagian data dengan presentase 75% data latih dan 25% data uji. Pada algoritma Extra Trees Suspek memiliki akurasi tertinggi yaitu 89,70% dan akurasi terendahnya terdapat pada kategori Rawat inap dengan presentase 54,29%. Pada algoritma klasifikasi Gradient Boosting sendiri suspek tetap menjadi kategori tertinggi jika di bandingkan dengan kategori lainnya yaitu dengan presentase 85,71%, namun jika di bandingkan dengan presentase akurasi suspek menggunakan algoritma Extra Trees suspek pada algoritma Gradient Boosting lebih rendah nilai akurasi nya. Dan untuk nilai presentase akurasi terendah pada algoritma Gradient Boosting sendiri terdapat pada kategori Rawat Inap yaitu 48,57%. Berbeda dengan Gradient Boosting dan Extra Trees di mana kategori suspek merupakan presentase akurasi tertinggi jika di bandingkan dengan kategori lainnya sedangkan pada algoritma Logistic Regression kategori positif merupakan kategori yang memiliki presentase akurasi tertinggi dengan nilai 88,57%. Sama halnya dengan algoritma Extra Trees dan Gradient Boosting kategori rawat inap pada algoritma Logistic Regression juga menjadi kategori yang memiliki nilai akurasi terendah jika di bandingkan dengan kategori lainnya yaitu dengan nilai presentase 40%.



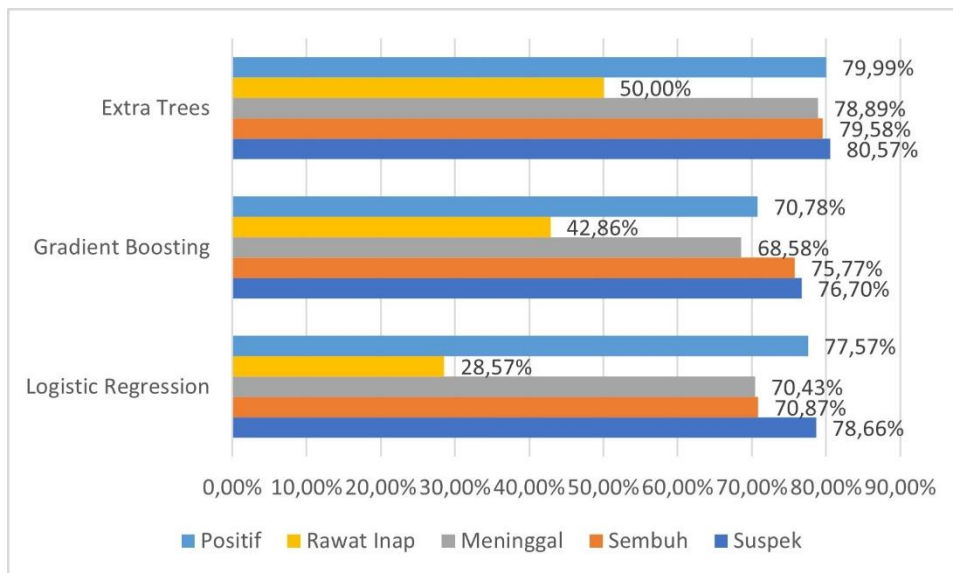
Gambar 3. Grafik Algoritma Perbandingan Pembagian Data 80%:20%

Gambar di atas menunjukkan perbandingan algoritma yaitu, Extra Trees, Gradient Boosting dan Logistic Regression. Dalam memperdiksi tingkat akurasi untuk masing-masing algoritma pada setiap kategori. Algoritma Extra Trees memiliki nilai yang relatif tinggi hampir disetiap kategori dengan nilai akurasi tertinggi terdapat pada kategori suspek 92,86% dan nilai akurasi terendah yang berada dikategori rawat inap yaitu 46,43%. Disusul dengan Gradient Boosting yang mempunyai nilai akurasi tertinggi yang sama dengan Extra Trees yang berada pada kategori suspek 85,71% dan kategori sembuh dengan nilai akurasi 82,14% tetapi masih relatif dibawah Extra Trees. Dan algoritma Logistic Regression menunjukan performa nilai akurasi yang relatif konsisten dibeberapa kategori dengan nilai akurasi tertinggi berada pada kategori suspek 89,29% dan akurasi terendah terdapat pada kategori Rawat inap 39,29%. Secara umum, hasil ini mengindikasikan bahwa Extra Trees unggul dalam memprediksi sebagian besar kategori, Logistic Regression memiliki kestabilan performa, sedangkan Gradient Boosting tetap kompetitif meski cenderung berada sedikit di bawah kinerja Extra Trees. Tantangan terbesar ketiga algoritma terletak pada prediksi kategori *Rawat Inap*, yang kemungkinan disebabkan oleh distribusi data yang tidak seimbang atau keterbatasan fitur prediktif yang tersedia.



Gambar 4. Grafik Algoritma Perbandingan Pembagian Data 85%:15%

Tiga algoritma Logistic Regression, Gradient Boosting, dan Extra Trees berhasil mengkategorikan pasien ke dalam kategori berikut: Positif, Rawat Inap, Meninggal, Sembuh, dan Suspek, seperti yang ditunjukkan pada grafik. Nilai-nilai ini menunjukkan persentase perkiraan yang tepat untuk masing-masing kategori. Semua algoritma memiliki tingkat akurasi yang lebih rendah dalam memprediksi kasus "Rawat Inap", dengan tingkat akurasi berkisar antara 23,81% hingga 40,67%. Ini menunjukkan bahwa model menghadapi kesulitan untuk menemukan pola dalam kategori ini. Extra Trees muncul sebagai pemenang secara keseluruhan. Extra Trees di atas memiliki akurasi tertinggi di kategori Positif, Sembuh, dan Suspek, dan juga memimpin di kategori Meninggal. Hasil akurasi tertinggi terdapat pada kategori suspek dengan nilai 87,81% dan hasil akurasi terendahnya terdapat pada kategori rawat inap yaitu 40,67%. Performa dari Logistic Regression cukup seimbang, meskipun kasus Rawat Inap sangat jarang terjadi. Pada logistic Regression sendiri performa terbaik terdapat pada kategori suspek dengan akurasi 83,71% dan akurasi terendah terdapat pada kategori rawat inap yaitu 23,81%. Gradient Boosting meningkatkan kategori Rawat Inap sedikit dibandingkan Logistic Regression, tetapi masih jauh di bawah Extra Trees, dan secara keseluruhan, performanya lebih sering di bawah kedua algoritma lainnya. Pada algoritma ini sendiri kategori yang memiliki akurasi tertinggi terdapat pada kategori positif, suspek, dan sembuh. Sementara akurasi performa terendah terdapat pada kategori rawat inap yaitu 38,10%.



Gambar 5. Grafik Algoritma Perbandingan Pembagian Data 90:10%

Grafik di atas menunjukkan perbandingan kinerja tiga algoritma pembelajaran mesin untuk mengklasifikasikan lima kategori data COVID-19: Positif, Rawat Nginap, Meninggal, Sembuh, dan Suspek. Masing-masing model diuji untuk mengetahui seberapa akurat mereka dapat memprediksi setiap kategori. Hasil uji ditampilkan dalam bentuk batang horizontal yang menunjukkan persentase keberhasilan. Logistic Regression sangat baik dalam memprediksi kasus Positif 77,57% dan Suspek 78,66%, tetapi sangat buruk dalam memprediksi Rawat inap, dengan akurasi hanya 28,57%, yang merupakan nilai terendah di seluruh model dan kategori ini. Ini menunjukkan bahwa model ini mungkin menghadapi masalah dengan data yang mungkin lebih kompleks atau tidak seimbang. Gradient Boosting menghasilkan hasil yang lebih konsisten. Untuk kategori Sembuh, akurasi relatif tinggi yaitu 76,70%, tetapi untuk kategori Rawat inap 42,86% yang artinya akurasi masih rendah. Sebaliknya, model ini mampu mengalahkan regresi logistik dalam sebagian besar kasus. Sebagai puncak perbandingan ini, Extra Trees tampak paling unggul. Model ini tidak hanya memiliki akurasi tertinggi pada hampir semua kategori, tetapi juga sangat baik dalam kategori yang sulit seperti Rawat inap dengan hasil akurasi 50% dan Suspek 80,57%. Selain itu, dalam kategori lain seperti Sembuh dan Meninggal, secara konsisten melampaui dua model lainnya. Extra Trees jelas menjadi model yang paling konsisten dan tangguh dari ketiga model yang diuji. Model ini sangat akurat di hampir semua kategori, termasuk kategori Rawat Nginap yang sulit diprediksi. Itu juga mengungguli pesaingnya dan menunjukkan kemampuan untuk menangani data kesehatan yang kompleks seperti COVID-19.

Evaluasi Model

Untuk menilai kinerja tiga algoritma pembelajaran mesin Logistic Regression, Gradient Boosting, dan Extra Trees Classifier dalam mengklasifikasikan kondisi pasien COVID-19 berdasarkan data kualitas udara pada lima kategori target: Suspek, Positif, Rawat Inap, Sembuh, dan Meninggal. Evaluasi ini dilakukan dengan berbagai rasio pembagian data latih dan uji. Hasil menunjukkan bahwa penerapan Min-Max Normalization secara signifikan meningkatkan akurasi semua model. Dapat dilihat bahwa nilai akurasi tertinggi hanya 14,29% sebelum normalisasi Extra Trees pada rasio 80%: 20%, sedangkan nilai akurasi setelah normalisasi dapat melampaui 90% pada kategori tertentu. Extra Trees tetap menjadi model terbaik, dengan akurasi tertinggi 92,86% pada kategori Suspek dan performa terbaik pada kategori Positif dan Sembuh, sementara Logistic Regression menunjukkan kestabilan performa dan Gradient Boosting bersifat kompetitif meskipun cenderung lebih fluktuatif. Meskipun akurasi meningkat setelah normalisasi, kategori rawat inap menghadapi masalah terbesar. Nilai tertingginya hanya 54,29% (Extra Trees pada rasio 75%:25%), hal ini terjadi karena ketidakseimbangan data dan keterbatasan fitur prediktif. Secara keseluruhan, penelitian ini menunjukkan bahwa normalisasi data sangat penting untuk meningkatkan kinerja model klasifikasi. Penelitian ini juga menunjukkan bahwa kualitas udara dapat membantu memprediksi kondisi pasien COVID-19 dan bahwa kategori dengan data yang tidak seimbang harus ditangani dengan hati-hati agar kinerja model lebih merata di semua kategori.

SIMPULAN

Penelitian ini membuktikan bahwa Min-Max Normalization signifikan meningkatkan akurasi klasifikasi kondisi pasien COVID-19 berbasis kualitas udara. Extra Trees menjadi model paling konsisten dengan akurasi tertinggi 92,86% pada kategori Suspek, diikuti Logistic Regression yang stabil dan Gradient Boosting yang fluktuatif. Meski akurasi meningkat, kategori Rawat Inap tetap menjadi tantangan akibat ketidakseimbangan data. Kualitas udara terbukti berpengaruh signifikan terhadap prediksi, khususnya pada kategori Suspek dan Sembuh. Normalisasi data menjadi tahap krusial, dan penelitian lanjutan disarankan mengatasi distribusi tidak seimbang serta menambah variabel lingkungan untuk meningkatkan performa model.

DAFTAR PUSTAKA

- M. K. Sari, "Sosialisasi tentang Pencegahan Covid-19 di Kalangan Siswa Sekolah Dasar di SD Minggiran 2 Kecamatan Papar Kabupaten Kediri," *J. Karya Abdi*, vol. 4, no. 1, pp. 80–83, 2020.
- N. H. Nasution and A. Hidayah, "Gambaran Pengetahuan Masyarakat Tentang Pencegahan Covid-19 Di Kecamatan Padangsidempuan Batunadua, Kota Padangsidempuan," *J. Kesehat. Ilm. Indones. (Indonesian Heal. Sci. Journal)*, vol. 6, no. 1, pp. 107–114, 2021, doi: 10.51933/health.v6i1.419.
- L. Amalia, I. Irwan, and F. Hiola, "Analisis Gejala Klinis Dan Peningkatan Kekebalan Tubuh Untuk Mencegah Penyakit Covid-19," *Jambura J. Heal. Sci. Res.*, vol. 2, no. 2, pp. 71–76, 2020, doi: 10.35971/jjhsr.v2i2.6134.
- W. Zendrato, "Gerakan Mencegah Daripada Mengobati Terhadap Pandemi Covid-19," *J. Educ. Dev.*, vol. 8, no. 2, pp. 242–248, 2020.
- Athena, E. Laelasari, and T. Puspita, "PELAKSANAAN DISINFEKSI DALAM PENCEGAHAN PENULARAN COVID- 19 DAN POTENSI RISIKO TERHADAP KESEHATAN DI INDONESIA Implementation of Disinfection in Prevention of Covid-19 Transmission and Its Potential Health Risk," *J. Ekol. Kesehat.*, vol. 19, no. 1, pp. 1–20, 2020, [Online]. Available: <https://doi.org/10.22435/jek.v19i1.3146> [Diakses pada 28 Maret 2023]
- S. A. Syaharani Nafisa, Hadi Firman, "Eksplorasi Fluktuasi Konsentrasi Pm2.5 Di Provinsi Daerah Khusus Jakarta Selama Oktober 2022," *J. Geod. Undip*, vol. 13, pp. 601–607, 2024.
- P. Agista, N. Gusdini, and M. Maharani, "Analisis Kualitas Udara Dengan Indeks Standar Pencemar Udara (Ispu) Dan Sebaran Kadar Polutannya Di Provinsi Dki Jakarta," *Sustain. Environ. Optim. Ind. J.*, vol. 2, no. 2, pp. 39–57, 2020, doi: 10.36441/seoi.v2i2.491.
- I. P. Ponamon and A. D. Wowor, "Penerapan Model Machine Learning (KNN, Random Forest, dan Naive Bayes) untuk Menganalisis Pengaruh Kualitas Udara terhadap COVID-19: Studi Normalisasi ...," *Kesatria J. Penerapan ...*, vol. 6, no. 2, pp. 537–546, 2025, [Online]. Available: <http://www.pkm.tunasbangsa.ac.id/index.php/kesatria/article/view/596>
<http://www.pkm.tunasbangsa.ac.id/index.php/kesatria/article/download/596/591>
- A. Nata and S. Suparmadi, "Analisis Sistem Pendukung Keputusan Dengan Model

- Klasifikasi Berbasis Machine Learning Dalam Penentuan Penerima Program Indonesia Pintar," *J. Sci. Soc. Res.*, vol. 5, no. 3, p. 697, 2022, doi: 10.54314/jssr.v5i3.1041.
- A. Fahri and Y. Ramdhani, "Visualisasi Data dan Penerapan Machine Learning Menggunakan Decision Tree Untuk Keputusan Layanan Kesehatan COVID-19," *J. Tekno Kompak*, vol. 17, no. 2, p. 50, 2023, doi: 10.33365/jtk.v17i2.2438.
- D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- T. Z. Jasman, M. A. Fadhlullah, A. L. Pratama, and R. Rismayani, "Analisis Algoritma Gradient Boosting, Adaboost dan Catboost dalam Klasifikasi Kualitas Air," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 2, pp. 392–402, 2022, doi: 10.28932/jutisi.v8i2.4906.
- A. Adrian and I. Verawati, "Analisis Performa Logistic Regression dan Random Forest dalam Klasifikasi Kelayakan Penerimaan Kredit," vol. 4, no. 2, pp. 148–158, 2025.
- W. Aprilita, Junadhi, Agustin, and Hadi Asnal, "Analisis Sentimen Layanan Hotel Menggunakan Algoritma Extra Trees: Studi Kasus pada Ulasan Pelanggan," *Indones. J. Comput. Sci.*, vol. 13, no. 3, pp. 4642–4653, 2024, doi: 10.33022/ijcs.v13i3.4014.
- I. Permana and F. N. S. Salisah, "Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation," *Indones. J. Inform. Res. Softw. Eng.*, vol. 2, no. 1, pp. 67–72, 2022, doi: 10.57152/ijirse.v2i1.311.
- S. Rizky Pratama and I. N. Fajri, "Perbandingan Algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) Dalam Memprediksi Kepuasan Pelanggan Comparison of K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) Algorithms in Predicting Customer Satisfaction," *J. Comput. Sci. Informatics Eng.*, vol. 04, no. 3, pp. 135–146, 2025, [Online]. Available: <http://creativecommons.org/licenses/by-sa/4.0/>
- B. Setio and P. Prasetyaningrum, "Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means," *J. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 27–32, 2021, doi: 10.54840/jcstech.v1i1.9.
- F. Adams, R. A. Dwi Anggoro, M. B. Satria, A. W. Oktavia, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma Naïve Bayes, Decision Tree, dan Support Vector Machine," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 2, no. 2, pp. 260–268, 2021, [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/1744>.