



## Penerapan *Stacking Ensemble Learning* untuk Klasifikasi Efek Kesehatan Akibat Pencemaran Udara

Budi Sunarko<sup>1)</sup>, Uswatun Hasanah<sup>2)</sup>, Syahroni Hidayat<sup>3)✉</sup>, Naufal Muhammad<sup>1)</sup>, Muhammad Irfan Ardiansyah<sup>1)</sup>, Muhammad Khikam Hakiki<sup>1)</sup>, Briska Putra Ananda<sup>2)</sup>, dan Luluk Taufiqul Baroroh<sup>3)</sup>

<sup>1)</sup>Program Studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Semarang, Indonesia

<sup>2)</sup>Program Studi Teknik Komputer, Fakultas Teknik, Universitas Negeri Semarang, Indonesia

<sup>3)</sup>Program Studi Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang, Indonesia

### Info Artikel

#### Sejarah Artikel:

Diterima: 27 Juli 2023

Direvisi: 6 September 2023

Disetujui: 29 September 2023

#### Keywords:

Dampak kesehatan,  
Ensemble learning,  
Klasifikasi efek kesehatan,  
Pencemaran udara, Stacking

### Abstrak

Pencemaran udara merupakan masalah serius yang berdampak negatif pada kesehatan manusia. Berbagai jenis polutan udara seperti partikel halus, sulfur dioksida, nitrogen oksida, dan ozon dapat menyebabkan gangguan pernapasan, penyakit jantung, kanker paru-paru, dan masalah kesehatan lainnya. Untuk memahami dampak kesehatan pencemaran udara, klasifikasi efek kesehatan akibat pencemaran udara menjadi penting. Metode klasifikasi ini membagi efek kesehatan berdasarkan jenis polutan, dosis, dan waktu paparan. Penelitian ini mengusulkan penerapan metode klasifikasi dengan *ensemble learning* untuk mengidentifikasi polutan berdampak dan tingkat risiko kesehatannya. *Ensemble learning* adalah teknik pembelajaran mesin yang menggabungkan beberapa model untuk meningkatkan akurasi prediksi. *Stacking ensemble learning* merupakan salah satu metode yang digunakan dalam klasifikasi efek kesehatan pencemaran udara dengan mengintegrasikan beberapa model dasar seperti *Logistic Regression*, *Decision Tree*, *K-Nearest Neighbor*, *Support Vector Machine*, dan *Multi-Layer Perceptron*. Hasil penelitian menunjukkan bahwa model *Stacking* memberikan performa tertinggi dengan akurasi sekitar 99,9% pada dataset baik yang seimbang maupun tidak seimbang. Namun, model *Decision Tree* dan *K-Nearest Neighbor* juga berhasil memberikan performa yang sangat baik. Waktu pelatihan model menjadi pertimbangan penting, di mana *K-Nearest Neighbor* dan *Decision Tree* memiliki waktu yang jauh lebih singkat dibandingkan dengan model *Stacking*.

### Abstract

Air pollution is a serious problem that has negative impacts on human health. Various types of air pollutants such as fine particles, sulfur dioxide, nitrogen oxides, and ozone can cause respiratory problems, heart diseases, lung cancer, and other health issues. To understand the health effects of air pollution, the classification of health effects due to air pollution becomes important. This classification method divides health effects based on the type of pollutant, dosage, and exposure time. This research proposes the application of classification methods using ensemble learning to identify pollutants and their health risk levels. Ensemble learning is a machine learning technique that combines multiple models to improve prediction accuracy. Stacking ensemble learning is one of the methods used in classifying the health effects of air pollution by integrating several basic models such as Logistic Regression, Decision Tree, K-Nearest Neighbor, Support Vector Machine, and Multi-Layer Perceptron. The research results show that the Stacking model provides the highest performance with an accuracy of approximately 99.9% on both balanced and unbalanced datasets. However, the Decision Tree and K-Nearest Neighbor models also perform very well. Training time of the models becomes an important consideration, where K-Nearest Neighbor and Decision Tree models have much shorter training times compared to the Stacking model.

© 2023 Universitas Negeri Semarang

✉ Alamat korespondensi:  
Gedung E11 Lantai 2, Teknik Elektro UNNES  
Kampus Sekaran, Gunungpati, Semarang, 50229  
E-mail: syahronihidayat@mail.unnes.ac.id

ISSN 2252-6811  
E-ISSN 2599-297X

## PENDAHULUAN

Pencemaran udara merupakan masalah lingkungan yang serius dan dapat mempengaruhi kesehatan manusia. Berbagai jenis polutan udara seperti partikel halus, sulfur dioksida, nitrogen oksida, dan ozon dapat menyebabkan berbagai masalah kesehatan seperti gangguan pernapasan, penyakit jantung, kanker paru-paru, dan masalah kesehatan lainnya (Hassan Bhat et al., 2021; Samek, 2016). Data dari *European Environment Agency* menyatakan bahwa polusi udara akibat CO<sub>2</sub> (ppm), CH<sub>4</sub> (ppb) dan N<sub>2</sub>O (ppb) antara tahun 1800 dan 2017 terus meningkat (*Trends in Atmospheric Concentrations of CO<sub>2</sub> (Ppm), CH<sub>4</sub> (Ppb) and N<sub>2</sub>O (Ppb), between 1800 and 2017 — European Environment Agency*, n.d.). Dan menurut *World Health Organization* (WHO) polusi udara menyebabkan kematian prematur mencapai 6.7 juta per tahunnya (*Types of Pollutants*, n.d.). Selain itu, sekitar 2,4 miliar orang terpapar polusi udara rumah tangga dengan tingkat berbahaya saat menggunakan api atau kompor sederhana berbahan bakar minyak tanah, biomassa (kayu, kotoran hewan dan limbah tanaman) dan batu bara (*Types of Pollutants*, n.d.).

Untuk memahami dampak kesehatan dari pencemaran udara, diperlukan klasifikasi efek kesehatan akibat pencemaran udara. Klasifikasi efek kesehatan akibat pencemaran udara sangat penting untuk memahami dampak buruk pencemaran udara pada kesehatan manusia (Du et al., 2021). Dalam klasifikasi ini, efek kesehatan akibat pencemaran udara dikategorikan berdasarkan jenis polutan, dosis, dan waktu paparan (Khojasteh et al., 2021).

Metode klasifikasi efek kesehatan akibat pencemaran udara dapat membantu para ahli lingkungan dan kesehatan dalam menentukan risiko kesehatan yang terkait dengan tingkat pencemaran udara tertentu. Salah satu teknologi yang dapat digunakan untuk membantu dalam pengklasifikasian tersebut adalah pembelajaran mesin (Gladkova & Saychenko, 2022; Li et al., 2023). Dengan menggunakan pembelajaran mesin, dapat dilakukan analisis data udara yang sangat besar dan kompleks untuk mengidentifikasi polutan udara yang berdampak pada kesehatan dan menentukan tingkat risiko kesehatan yang terkait dengan polutan tersebut (Kumar & Pande, 2022; Lei et al., 2022).

Salah satu metode pembelajaran mesin yang digunakan dalam klasifikasi ini adalah *Ensemble Learning* (Hadj Sassi & Chaari Fourati, 2022). *Ensemble learning* adalah teknik pembelajaran mesin yang memanfaatkan beberapa model pembelajaran mesin untuk menghasilkan prediksi yang lebih akurat. Dalam

klasifikasi efek kesehatan pencemaran udara. *Ensemble learning* dapat digunakan untuk mengkombinasikan beberapa model pembelajaran mesin yang berbeda untuk meningkatkan akurasi prediksi risiko kesehatan yang terkait dengan paparan polutan udara (Hulkkonen et al., 2022; Singh et al., 2013; Zhang et al., 2022).

Penerapan *ensemble learning* dalam klasifikasi efek kesehatan pencemaran udara sangat penting karena dapat meningkatkan akurasi prediksi risiko kesehatan yang terkait dengan paparan polutan udara (Masmoudi et al., 2020). Dengan mengkombinasikan beberapa model pembelajaran mesin yang berbeda, *ensemble learning* dapat mengatasi masalah ketidakpastian dan kelemahan dalam satu model, sehingga menghasilkan prediksi yang lebih akurat (Ke et al., 2022).

*Ensemble Learning* adalah teknik pembelajaran mesin di mana beberapa model pembelajaran mesin digabungkan untuk meningkatkan akurasi dan konsistensi prediksi. Dalam konteks klasifikasi efek kesehatan akibat pencemaran udara, *Ensemble Learning* dapat digunakan untuk mengintegrasikan beberapa model pembelajaran mesin yang berbeda, seperti *Decision Tree*, *Random Forest*, *Support Vector Machine* (SVM), dan *Neural Networks* (Gokul et al., 2023; Méndez et al., 2023).

Dalam penelitian terbaru, beberapa studi telah dilakukan untuk mengaplikasikan metode *Ensemble Learning* pada klasifikasi efek kesehatan akibat pencemaran udara (Cui & Wang, 2021; Singh et al., 2013). Hasil penelitian menunjukkan bahwa metode ini dapat meningkatkan akurasi dan konsistensi prediksi dan dapat dijadikan sebagai alternatif yang lebih andal dan terpercaya dalam mengevaluasi dampak kesehatan dari pencemaran udara (Zhang et al., 2022).

Klasifikasi efek kesehatan akibat pencemaran udara menggunakan metode *ensemble learning* telah menunjukkan kemampuan untuk meningkatkan akurasi prediksi efek kesehatan akibat pencemaran udara (Du et al., 2021; Lin et al., 2021). Selain itu, kebaruan dalam klasifikasi efek kesehatan akibat pencemaran udara menggunakan metode *ensemble learning* mencakup pengembangan teknik baru untuk mengatasi masalah ketidakseimbangan *dataset* dan peningkatan kinerja model dengan menggunakan teknik augmentasi data dan

pengelolaan fitur (García et al., 2018; Liu et al., 2023; Magnolia et al., 2023).

Metode *ensemble learning* memiliki banyak tipe, diantaranya *Bagging*, *Boosting*, *Stacking*, *AdaBoost*, dan yang terbaru adalah tipe *Dynamic* (Alves Ribeiro et al., 2020; Liu et al., 2023; Wu et al., 2021). Setiap metode memiliki kelebihan dan kekurangannya masing-masing. Oleh karena itu, pada penelitian ini akan diimplementasikan metode *stacking ensemble learning* untuk mengklasifikasi efek kesehatan yang muncul akibat pencemaran udara.

## METODE PENELITIAN

*Ensemble learning* adalah teknik pembelajaran mesin yang melibatkan penggabungan beberapa model untuk meningkatkan performa tugas prediksi secara keseluruhan (Murad et al., 2021; Rismayati et al., 2022). Pada penelitian ini metodologi penelitian dapat dijelaskan sebagai berikut:

1. Penentuan Masalah: Langkah pertama dalam setiap proyek pembelajaran mesin adalah menentukan masalah yang akan diselesaikan melalui *literature review*. Pada penelitian ini akan diselesaikan masalah polusi udara.
2. Persiapan data: Persiapan data merupakan proses pembersihan data, penanganan nilai yang hilang, pengkodean variabel kategori, dan penskalaan data. Data yang digunakan adalah data polusi udara yang dengan tingkat dan label kelas yang berbeda-beda. Jumlah kelas akan dianalisis untuk memastikan kondisi *dataset* seimbang atau tidak seimbang. Penerapan rekayasa penyeimbangan data akan dilakukan jika *dataset* yang dimiliki tidak seimbang.
3. Pemilihan model dasar: Pembelajaran ensemble melibatkan penggabungan beberapa model dasar, Model dasar yang akan diterapkan antara lain *Decision Tree*, *Support Vector Machine*, *neural network*, dan model regresi.
4. Pelatihan model dasar: Setelah memilih model dasar, dilakukan pelatihan terhadap *dataset* polusi udara yang telah ditentukan. Akan diterapkan subkumpulan kondisi data acak untuk melatih setiap model guna memastikan bahwa model tersebut berbeda dan beragam. Model dasar yang memberikan hasil terbaik akan digunakan pada proses penggabungan model.
5. Penggabungan model: Beberapa cara untuk menggabungkan prediksi dari model dasar,

yaitu rata-rata (*average*), rata-rata terbobot (*weighted average*), susun (*stacking*) dan metode dinamis. Pada penelitian ini diterapkan metode *stacking*.

6. Evaluasi kinerja: Setelah menggabungkan prediksi model dasar, dilakukan evaluasi kinerja model ensemble berupa nilai akurasi, presisi, daya ingat (*recall*), dan skor F1. Selain itu, waktu eksekusi metode *ensemble learning* akan dievaluasi.
7. Menyesuaikan hyperparameter: Pembelajaran ensemble melibatkan beberapa *hyperparameter* yang perlu disetel, seperti jumlah model dasar, skema pembobotan, dan model meta. Akan diterapkan teknik seperti *cross-validation* untuk menyetel *hyperparameter*. Jumlah *fold* masing-masing adalah 10 *fold*, 5 *fold*, dan 3 *fold*.
8. Pengujian model: kinerja model menggunakan *dataset* pengujian terpisah. Untuk mendapatkan perkiraan kinerja model yang tidak bias pada data baru. Presentase data pengujian adalah 10%, 20%, dan 30%.

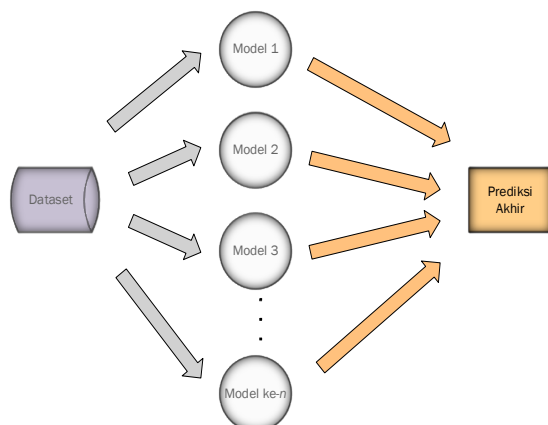
### A. Stacking Ensemble Learning

*Stacking (Stacked Generalization)* melibatkan pelatihan model meta untuk membuat prediksi berdasarkan *output* dari beberapa model dasar. Model dasar biasanya merupakan jenis model yang berbeda, atau model yang sama dilatih dengan parameter atau sub-kumpulan data yang berbeda. Proses kerja *stacking* (penumpukan) terdiri dari membagi data pelatihan menjadi dua subset atau lebih, melatih beberapa model dasar pada satu subset data, menggunakan model terlatih untuk membuat prediksi pada subkumpulan data lainnya, menggabungkan prediksi dari model dasar ke dalam kumpulan data baru, melatih model meta pada kumpulan data baru ini, menggunakan nilai target aktual sebagai label, dan menggunakan meta-model untuk membuat prediksi pada data uji. Secara umum metode *ensemble learning*, termasuk *stacking*, bekerja seperti ditunjukkan pada Gambar 1.

Gagasan di balik *stacking* adalah bahwa setiap model dasar mungkin memiliki kekuatan dan kelemahan, dan dengan menggabungkan prediksi mereka, meta-model dapat belajar menimbang kekuatan dan kelemahan ini dengan tepat (Zhang et al., 2022). Ini dapat menghasilkan kinerja keseluruhan yang lebih baik daripada model individual mana pun.

Model dasar yang umum digunakan adalah *Regression Model (RM)*, *Decision Tree (DT)*,

*Support Vector Machine (SVM), Multilayer Perceptron (MLP), K-Nearest Neighbor (k-NN).*



Gambar 1. Skema umum metode *ensemble learning*

#### B. *Logistic Regression*

*Logistic Regression* adalah salah satu algoritma pembelajaran mesin yang umum digunakan untuk masalah klasifikasi. Meskipun namanya mengandung kata "regresi," *Logistic Regression* sebenarnya digunakan untuk tugas klasifikasi, bukan regresi. Algoritma ini cocok untuk masalah di mana variabel target bersifat biner (dua kelas) atau ordinal (Marini et al., 2022).

#### C. *Decision Tree*

*Decision Tree* (pohon keputusan) adalah salah satu algoritma pembelajaran mesin yang digunakan untuk masalah klasifikasi dan regresi. Algoritma ini menghasilkan model dalam bentuk struktur pohon, di mana setiap simpul internal mewakili keputusan berdasarkan fitur-fitur data, dan setiap daun mewakili label kelas (untuk klasifikasi) atau prediksi nilai (untuk regresi).

Keuntungan dari *Decision Tree* adalah interpretabilitasnya yang tinggi. Model pohon keputusan dapat diuraikan secara visual sehingga mudah dipahami oleh manusia. Selain itu, *Decision Tree* juga dapat menangani data kategorikal dan numerik tanpa memerlukan pra-pengolahan data yang rumit (Mercol et al., 2008).

#### D. *Support Vector Machine*

*Support Vector Machine (SVM)* adalah salah satu algoritma pembelajaran mesin yang kuat dan sering digunakan untuk tugas klasifikasi dan regresi. SVM merupakan metode pembelajaran berbasis pengawasan (*supervised learning*) yang digunakan untuk memisahkan dua kelas dengan mencari *hyperplane* (bidang pemisah) optimal yang dapat memaksimalkan *margin* (jarak) antara dua kelas. SVM adalah algoritma yang cukup populer

dalam dunia pembelajaran mesin karena keahliannya dalam melakukan klasifikasi dan regresi, terutama ketika data memiliki jumlah fitur yang besar (Amira et al., 2020; Worasawate et al., 2022).

#### E. *Multilayer Perceptron*

*Multilayer Perceptron (MLP)* adalah salah satu jenis arsitektur dari jaringan saraf tiruan (*neural networks*) dalam pembelajaran mesin. MLP adalah model yang terdiri dari satu lapisan input, satu atau lebih lapisan tersembunyi (*hidden layers*), dan satu lapisan *output*. MLP merupakan salah satu algoritma pembelajaran mendalam (*deep learning*) karena memiliki lebih dari satu lapisan tersembunyi.

MLP memiliki kemampuan untuk menyelesaikan masalah yang sangat kompleks dan mampu memodelkan hubungan yang sangat non-linear dalam data. Meskipun demikian, pelatihan dan tuning MLP memerlukan pemilihan parameter yang tepat serta bisa menjadi proses yang intensif secara komputasional, terutama pada *dataset* yang besar dan dengan arsitektur yang kompleks (Maleki et al., 2019).

#### F. *K-Nearest Neighbor*

*K-Nearest Neighbor (KNN)* adalah salah satu algoritma pembelajaran mesin yang digunakan untuk masalah klasifikasi dan regresi. KNN adalah algoritma yang berbasis *instance-based learning*, yang berarti modelnya tidak mempelajari distribusi data, tetapi menyimpan seluruh *dataset* pelatihan sebagai pengetahuan untuk melakukan prediksi pada data baru.

Algoritma KNN relatif mudah diimplementasikan dan dapat memberikan hasil yang baik pada data yang cukup besar. Namun, kelemahan utamanya adalah sensitif terhadap skala data dan bisa menjadi lambat pada *dataset* dengan jumlah sampel yang sangat besar. Selain itu, pemilihan parameter K yang tepat juga bisa mempengaruhi performa algoritma KNN. Oleh karena itu, dalam prakteknya, penting untuk melakukan eksplorasi dan evaluasi parameter-parameter KNN untuk mencapai hasil yang optimal (Shah et al., 2020).

#### G. *Cross Validation*

*Cross-validation* adalah Teknik yang digunakan luas pada pembelajaran mesin dan statistika untuk mengevaluasi performa sebuah model prediksi. Bekerja dengan membagi *dataset* menjadi beberapa bagian *dataset* lebih kecil yang digunakan baik untuk pelatihan maupun pengujian model. Beberapa tipe metode *cross-validation*, yaitu *K-fold cross validation*, dilakukan

dengan membagi data menjadi k himpunan bagian yang sama dan digunakan untuk kasus yang lebih umum. *Stratified cross-validation*, dilakukan dengan membagi data menjadi k himpunan bagian dengan mempertahankan proporsi sampel untuk setiap kelas. Metode ini digunakan untuk data yang tidak seimbang.

#### H. Evaluasi

Metrik evaluasi yang digunakan pada penelitian ini dirangkum pada Tabel 1.

Tabel 1. Metode Evaluasi

Metrik	Rumus	Keterangan
Akurasi	$Acc = \frac{TP + TN}{TP + TN + FP + FN}$	Mengukur seberapa sering model benar dalam memprediksi kelas yang benar.
Presisi	$Pr = \frac{TP}{TP + FP}$	Mengukur seberapa akurat model dalam memprediksi positif
Recall	$Rc = \frac{TP}{TP + FN}$	Mengukur seberapa baik model dalam mengidentifikasi semua sampel positif.
F1-Score	$F1\ score = 2 * \left( \frac{Pr * Rc}{Pr + Rc} \right)$	Memberikan gambaran tentang keseluruhan performa model

#### I. Dataset

*Dataset* yang digunakan adalah *Air Pollution Dataset* yang diperoleh dari Kaggle. Lebih detail *dataset* yang digunakan ditampilkan pada Tabel 2. *Dataset* memiliki fitur CH<sub>4</sub>, CO<sub>2</sub>, dan CO sebanyak 9065 baris dengan jumlah kelas 12. Tabel 3 menampilkan penjelasan label kelas dan jumlah masing-masing kelas. Dari label kelas dapat diketahui bahwa efek polusi udara bisa dari tidak memberikan efek sekalipun “*No Health Effect*” hingga yang paling parah dapat menyebabkan kematian dalam waktu singkat “*Death within 1-3 minutes of exposure*”. Selain itu, distribusi antara setiap kelas tidak seimbang (*imbalance*) sehingga pada prosesnya perlu menerapkan rekayasa *imbalance dataset*.

Tabel 2. *Dataset* Efek Kesehatan yang Ditimbulkan Akibat Pencemaran Udara

Nama Dataset	Fitur Data-set	Jumlah baris	Jumlah Label/Class	Sumber
CH4_CO_health_eff ect_NEW.csv	CH4, CO2, CO	9065 baris	12	<a href="https://www.kaggle.com/airpollutionhealth/ch4-co-co2-health-effects">https://www.kaggle.com/airpollutionhealth/ch4-co-co2-health-effects</a> [21]

Tabel 3. Penjelasan Label Setiap Kelas dan Rincian Jumlahnya

No	Label Kelas	Jumlah
1	<i>Headache, nausea, and dizziness after 5-10 minutes; collapse and unconsciousness after 30 minutes of exposure. Death within 1 hour.</i>	1742
2	<i>Death within 30 minutes.</i>	1562
3	<i>Possible health effects with long-term exposure.</i>	1477
4	<i>Headache, nausea, and dizziness after 20 minutes of exposure. Death within 1-2 hours.</i>	1427
5	<i>No Health Effect</i>	979
6	<i>Loss of consciousness after 1 hour of exposure.</i>	515
7	<i>Headache and nausea after 1-2 hours of exposure. Life threatening in 3 hours.</i>	458
8	<i>Death within 1-3 minutes of exposure.</i>	266
9	<i>Headache, nausea, and dizziness after 45 minutes; collapse and unconsciousness after 1 hour of exposure. Death within 2-3 hours.</i>	207
10	<i>Immediate physiological effects, unconsciousness. Death within 1-3 minutes of exposure.</i>	206
11	<i>Dizziness, naseau, fagitue, headache after 2-3 hours of exposure.</i>	159
12	<i>Slight headache after 1-2 hours.</i>	66
Total		9065

## HASIL DAN PEMBAHASAN

*Dataset* tidak seimbang adalah *dataset* di mana distribusi kelas tidak proporsional, sementara *dataset* seimbang adalah *dataset* dengan distribusi kelas yang hampir proporsional. Hasil pengujian menggunakan metrik evaluasi klasifikasi seperti akurasi, presisi, recall, dan F1-score yang sudah diintegrasikan di dalam kode program, serta mencatat waktu yang diperlukan oleh setiap model untuk melatih data pada berbagai pengaturan *fold* (lipatan), yaitu *fold* 10, *fold* 5, dan *fold* 3, ditunjukkan pada Tabel 4 dan Tabel 5.

Tabel 4. Hasil Penerapan *Ensemble Learning Stacking* dan Beberapa Model Dasar dengan Dataset Tidak Seimbang

Model	Dataset tidak Seimbang					
	Fold 10	Waktu (s)	Fold 5	Waktu (s)	Fold 3	Waktu (s)
Stacking	0,999	563,44	0,999	245,46	0,999	120,39
LR	0,329	11,24	0,331	5,08	0,333	3,83
KNN	0,966	0,97	0,964	0,964	0,961	1,36
DT	0,999	0,38	0,999	0,19	0,999	0,24
SVM	0,537	60,67	0,536	31,44	0,536	19,97
MLP	0,406	34,88	0,39	15,52	0,376	6,7

Tabel 5. Hasil Penerapan *Ensemble Learning Stacking* dan Beberapa Model Dasar dengan Dataset Seimbang

Model	Dataset Seimbang					
	Fold 10	Waktu (s)	Fold 5	Waktu (s)	Fold 3	Waktu (s)
Stacking	0,999	2624,93	0,999	1108,68	0,999	513,25
LR	0,203	31,55	0,203	11,13	0,204	6,97
KNN	0,985	2,01	0,984	1,85	0,981	1,75
DT	0,999	1,05	0,999	0,47	0,999	0,25
SVM	0,559	395,43	0,516	193,33	0,435	116,06
MLP	0,423	107,45	0,412	45,58	0,391	22,99

#### A. Performa Model pada Dataset Tidak Seimbang

Hasil pengujian pada *dataset* tidak seimbang menunjukkan bahwa model *Stacking* memiliki performa tertinggi dengan tingkat akurasi sekitar 99,9% untuk semua *fold* data (*Fold 10*, *Fold 5*, dan *Fold 3*). Teknik ini terbukti efektif dalam mengatasi ketidakseimbangan kelas pada *dataset* dan memberikan hasil yang konsisten pada berbagai *fold*. Selain itu, model *Decision Tree* (DT) juga menunjukkan performa yang sangat baik dengan akurasi 99,9% untuk semua *fold* data. Begitupula *K-Nearest Neighbors* (KNN) juga menunjukkan performa yang tinggi dengan akurasi sekitar 96,6% untuk *Fold 10*, 96,4% untuk *Fold 5*, dan 96,1% untuk *Fold 3*.

Metode *stacking*, DT, dan KNN memberikan performa yang seimbang. Akan tetapi, dalam hal ini *stacking* dapat disimpulkan memberikan performa yang paling baik dikarenakan dapat menggabungkan kelebihan dan mereduksi kekurangan yang dimiliki oleh

metode basis. Sehingga dengan kelebihan ini metode *stacking* dapat memberikan performa serupa untuk kondisi data yang beragam karena meningkatkan performa prediktif keseluruhan dengan mengurangi *overfitting* dan bias.

Di sisi lain, model *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Multi-Layer Perceptron* (MLP) menunjukkan performa yang lebih rendah dibandingkan dengan model *Stacking*, DT, dan KNN. LR dan SVM, keduanya adalah model linier, cenderung kurang efektif dalam menghadapi *dataset* tidak seimbang karena sensitivitas terhadap distribusi kelas. MLP, yang merupakan model jaringan saraf tiruan, mungkin memerlukan lebih banyak data atau penyetelan *hyperparameter* yang lebih hati-hati untuk mencapai performa yang lebih baik pada *dataset* tidak seimbang. Oleh karena itu, model basis ini dapat dipertimbangkan lebih jauh untuk diterapkan sebagai model basis dalam sebuah model *ensemble stacking*.

#### B. Performa Model pada Dataset Seimbang

Hasil pengujian pada *dataset* seimbang menunjukkan bahwa model *Stacking* tetap memiliki performa tertinggi dengan tingkat akurasi sekitar 99,9% untuk semua *fold* data (*Fold 10*, *Fold 5*, dan *Fold 3*). Performa yang konsisten ini menegaskan keefektifan metode *Stacking* dalam meningkatkan performa prediksi, bahkan pada *dataset* yang seimbang. Model *Decision Tree* (DT) juga menunjukkan hasil yang sangat baik dengan akurasi 99,9% untuk semua *fold* data. KNN juga mempertahankan performa yang tinggi pada *dataset* seimbang, dengan akurasi sekitar 98,5% untuk *Fold 10*, 98,4% untuk *Fold 5*, dan 98,1% untuk *Fold 3*.

Model *Logistic Regression* (LR) menunjukkan peningkatan performa pada *dataset* seimbang, dengan akurasi sekitar 20,3% untuk semua *fold* data. Meskipun peningkatan ini signifikan, performa LR masih jauh di bawah model lainnya. SVM menunjukkan hasil yang bervariasi pada *dataset* seimbang, dengan akurasi sekitar 55,9% untuk *Fold 10*, 51,6% untuk *Fold 5*, dan 43,5% untuk *Fold 3*. Performa yang lebih rendah ini menunjukkan bahwa SVM mungkin kurang cocok untuk *dataset* seimbang atau memerlukan penyetelan parameter yang lebih hati-hati.

MLP juga menunjukkan peningkatan performa pada *dataset* seimbang, dengan akurasi sekitar 42,3% untuk *Fold 10*, 41,2% untuk *Fold 5*, dan 39,1% untuk *Fold 3*. Meskipun terjadi peningkatan, performa MLP masih lebih rendah dibandingkan dengan model *Stacking*, DT, dan KNN.



Berdasarkan hasil pengujian dengan *dataset* seimbang, bahwa *dataset* seimbang dapat meningkatkan performa model basis. Akan tetapi, hal tersebut tidak berlaku di beberapa model basis lainnya terutama *Stacking*. Hasil ini menegaskan bahwa metode *ensemble learning* khususnya *stacking* selalu dapat memberikan hasil akurasi yang tinggi meskipun kondisi *dataset* seimbang dan tidak seimbang.

### C. Perbandingan Waktu Pelatihan Model

Dalam hal waktu pelatihan, terdapat perbedaan yang cukup signifikan antara model-model yang diuji. Pada *dataset* tidak seimbang, model *Stacking* dan LR memiliki waktu pelatihan yang relatif lama dibandingkan dengan model lainnya. *Stacking* memerlukan waktu pelatihan terlama, dengan sekitar 563,44 detik untuk *Fold* 10, 245,46 detik untuk *Fold* 5, dan 120,39 detik untuk *Fold* 3. LR juga memerlukan waktu yang cukup lama dengan sekitar 11,24 detik untuk *Fold* 10, 5,08 detik untuk *Fold* 5, dan 3,83 detik untuk *Fold* 3.

Di sisi lain, model KNN dan DT memerlukan waktu yang jauh lebih singkat dalam melatih data pada semua *fold*. KNN memiliki waktu pelatihan yang sangat cepat, yaitu sekitar 0,97 detik untuk *Fold* 10, 0,964 detik untuk *Fold* 5, dan 1,36 detik untuk *Fold* 3. DT memiliki waktu pelatihan yang paling singkat, dengan sekitar 0,38 detik untuk *Fold* 10, 0,19 detik untuk *Fold* 5, dan 0,24 detik untuk *Fold* 3.

Pada *dataset* seimbang, pola waktu pelatihan tidak berbeda secara signifikan dengan *dataset* tidak seimbang. Model *Stacking* dan LR tetap memerlukan waktu pelatihan terlama, sedangkan KNN dan DT tetap menjadi yang tercepat. SVM dan MLP juga memiliki waktu pelatihan yang relatif serupa dengan *dataset* tidak seimbang. Peningkatan waktu pemrosesan model pada *dataset* seimbang diakibatkan oleh

penambahan jumlah data akibat rekayasa tak seimbang menjadi seimbang.

## KESIMPULAN

Dari hasil penelitian ini, dapat disimpulkan bahwa model *Stacking*, DT, dan KNN menunjukkan performa yang sangat baik dalam menghadapi *dataset* baik yang seimbang maupun tidak seimbang. Model *Stacking* konsisten memberikan akurasi tertinggi pada berbagai kondisi *dataset*, sementara DT dan KNN juga berhasil memberikan performa yang tinggi. Waktu pelatihan model juga menjadi pertimbangan penting dalam memilih model yang tepat. Model KNN dan DT menonjol karena memiliki waktu pelatihan yang sangat cepat, sedangkan model *Stacking* dan LR memerlukan waktu yang lebih lama. Meskipun demikian metode *stacking* tetap mengungguli model basis DT dan KNN dikarenakan kelebihan yang dimiliki oleh metode *stacking*.

Selain itu, penelitian ini juga menggarisbawahi bahwa masing-masing model memiliki kekuatan dan kelemahan dalam menghadapi *dataset* seimbang dan tidak seimbang. Penggunaan metode *ensemble* seperti *Stacking* dapat memberikan keunggulan dalam menghadapi masalah ketidakseimbangan kelas dan meningkatkan performa model secara keseluruhan.

## UCAPAN TERIMAKASIH

Penelitian ini didanai oleh Universitas Negeri Semarang tahun 2023 melalui hibah nomor DPA 023.17.2.690645/2023.10 REVISI 2 skema Penelitian Dasar dengan nomor kontrak 81.12.4/UN37/PPK.10/2023.

## DAFTAR PUSTAKA

- Alves Ribeiro, V. H., Moritz, S., Rehbach, F., & Reynoso-Meza, G. (2020). A novel dynamic multi-criteria ensemble selection mechanism applied to drinking water quality anomaly detection. *Science of The Total Environment*, 749, 142368. <https://doi.org/10.1016/j.scitotenv.2020.142368>
- Amira, S. A., Utama, S., & Fahmi, M. H. (2020). Penerapan Metode Support Vector Machine untuk Analisis Sentimen pada Review Pelanggan Hotel. *Edu Komputika Journal*, 7(2), 40–48. <https://doi.org/10.15294/edukomputika.v7i2.42608>
- CH4\_CO2\_Health Effects | Kaggle. (n.d.). Retrieved March 16, 2023, from <https://www.kaggle.com/datasets/airpollutionhealth/ch4-co-co2-health-effects>
- Cui, L., & Wang, S. (2021). Mapping the daily nitrous acid (HONO) concentrations across China during 2006–2017 through ensemble machine-learning algorithm. *Science of The Total Environment*, 785, 147325. <https://doi.org/10.1016/j.scitotenv.2021.147325>

- Du, Z., Heng, J., Niu, M., & Sun, S. (2021). An innovative ensemble learning air pollution early-warning system for China based on incremental extreme learning machine. *Atmospheric Pollution Research*, 12(9), 101153. <https://doi.org/10.1016/j.apr.2021.101153>
- García, S., Zhang, Z.-L., Altalhi, A., Alshomrani, S., & Herrera, F. (2018). Dynamic ensemble selection for multi-class imbalanced datasets. *Information Sciences*, 445–446, 22–37. <https://doi.org/10.1016/j.ins.2018.03.002>
- Gladkova, E., & Saychenko, L. (2022). Applying machine learning techniques in air quality prediction. *Transportation Research Procedia*, 63, 1999–2006. <https://doi.org/10.1016/j.trpro.2022.06.222>
- Gokul, P. R., Mathew, A., Bhosale, A., & Nair, A. T. (2023). Spatio-temporal air quality analysis and PM2.5 prediction over Hyderabad City, India using artificial intelligence techniques. *Ecological Informatics*, 76, 102067. <https://doi.org/10.1016/j.ecoinf.2023.102067>
- Hadj Sassi, M. S., & Chaari Fourati, L. (2022). Comprehensive survey on air quality monitoring systems based on emerging computing and communication technologies. *Computer Networks*, 209, 108904. <https://doi.org/10.1016/j.comnet.2022.108904>
- Hassan Bhat, T., Jiawen, G., & Farzaneh, H. (2021). Air Pollution Health Risk Assessment (AP-HRA), Principles and Applications. *International Journal of Environmental Research and Public Health*, 18(4), 1935. <https://doi.org/10.3390/ijerph18041935>
- Hulkkonen, M., Lipponen, A., Mielonen, T., Kokkola, H., & Prisle, N. L. (2022). Changes in urban air pollution after a shift in anthropogenic activity analysed with ensemble learning, competitive learning and unsupervised clustering. *Atmospheric Pollution Research*, 13(5), 101393. <https://doi.org/10.1016/j.apr.2022.101393>
- Ke, H., Gong, S., He, J., Zhang, L., Cui, B., Wang, Y., Mo, J., Zhou, Y., & Zhang, H. (2022). Development and application of an automated air quality forecasting system based on machine learning. *Science of The Total Environment*, 806, 151204. <https://doi.org/10.1016/j.scitotenv.2021.151204>
- Khojasteh, D. N., Goudarzi, G., Taghizadeh-Mehrjardi, R., Asumadu-Sakyi, A. B., & Fehrest-Sani, M. (2021). Long-term effects of outdoor air pollution on mortality and morbidity–prediction using nonlinear autoregressive and artificial neural networks models. *Atmospheric Pollution Research*, 12(2), 46–56. <https://doi.org/10.1016/j.apr.2020.10.007>
- Kumar, K., & Pande, B. P. (2022). Air pollution prediction with machine learning: A case study of Indian cities. *International Journal of Environmental Science and Technology*. <https://doi.org/10.1007/s13762-022-04241-5>
- Lei, T. M. T., Siu, S. W. I., Monjardino, J., Mendes, L., & Ferreira, F. (2022). Using Machine Learning Methods to Forecast Air Quality: A Case Study in Macao. *Atmosphere*, 13(9), 1412. <https://doi.org/10.3390/atmos13091412>
- Li, Y., Sha, Z., Tang, A., Goulding, K., & Liu, X. (2023). The application of machine learning to air pollution research: A bibliometric analysis. *Ecotoxicology and Environmental Safety*, 257, 114911. <https://doi.org/10.1016/j.ecoenv.2023.114911>
- Lin, C.-Y., Chang, Y.-S., & Abimannan, S. (2021). Ensemble multifeatured deep learning models for air quality forecasting. *Atmospheric Pollution Research*, 12(5), 101045. <https://doi.org/10.1016/j.apr.2021.03.008>
- Liu, S. M., Chen, J.-H., & Liu, Z. (2023). An empirical study of dynamic selection and random under-sampling for the class imbalance problem. *Expert Systems with Applications*, 221, 119703. <https://doi.org/10.1016/j.eswa.2023.119703>
- Magnolia, C., Nurhopipah, A., & Kusuma, B. A. (2023). Penanganan Imbalanced Dataset untuk Klasifikasi Komentar Program Kampus Merdeka Pada Aplikasi Twitter. *Edu Komputika Journal*, 9(2), 105–113. <https://doi.org/10.15294/edukomputika.v9i2.61854>



- Maleki, H., Sorooshian, A., Goudarzi, G., Baboli, Z., Tahmasebi Birgani, Y., & Rahmati, M. (2019). Air pollution prediction by using an artificial neural network model. *Clean Technologies and Environmental Policy*, 21(6), 1341–1352. <https://doi.org/10.1007/s10098-019-01709-w>
- Marini, R. P., Lavelly, E. K., Baugher, T. A., Crassweller, R., & Schupp, J. R. (2022). Using Logistic Regression to Predict the Probability That Individual ‘Honeycrisp’ Apples Will Develop Bitter Pit. *HortScience*, 57(3), 391–399. <https://doi.org/10.21273/HORTSCI16081-21>
- Masmoudi, S., Elghazel, H., Taieb, D., Yazar, O., & Kallel, A. (2020). A machine-learning framework for predicting multiple air pollutants’ concentrations via multi-target regression and feature selection. *Science of The Total Environment*, 715, 136991. <https://doi.org/10.1016/j.scitotenv.2020.136991>
- Méndez, M., Merayo, M. G., & Núñez, M. (2023). Machine learning algorithms to forecast air quality: A survey. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-023-10424-4>
- Mercol, J. P., Gambini, J., & Santos, J. M. (2008). Automatic classification of oranges using image processing and data mining techniques. *XIV Congreso Argentino de Ciencias de La Computación. XIV Argentine Congress of Computer Sciences (CACIC 2008)*, 1–12.
- Murad, M., Sukmawaty, S., Ansar, A., Sabani, R., & Hidayat, S. (2021). Sistem Pendeteksi Kerusakan Buah Mangga Menggunakan Sensor Gas Dengan Metode DCS - LCA. *JTIM: Jurnal Teknologi Informasi dan Multimedia*, 3(4), 186–194. <https://doi.org/10.35746/jtim.v3i4.169>
- Rismayati, R., Ismarmiaty, I., & Hidayat, S. (2022). Ensemble Implementation for Predicting Student Graduation with Classification Algorithm. *International Journal of Engineering and Computer Science Applications (IJECSA)*, 1(1), 35–42. <https://doi.org/10.30812/ijecsa.v1i1.1805>
- Samek, L. (2016). Overall human mortality and morbidity due to exposure to air pollution. *International Journal of Occupational Medicine and Environmental Health*, 29(3), 417–426. <https://doi.org/10.13075/ijom.1896.00560>
- Shah, K., Patel, H., Sanghvi, D., & Shah, M. (2020). A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification. *Augmented Human Research*, 5(1). <https://doi.org/10.1007/s41133-020-00032-0>
- Singh, K. P., Gupta, S., & Rai, P. (2013). Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmospheric Environment*, 80, 426–437. <https://doi.org/10.1016/j.atmosenv.2013.08.023>
- Trends in atmospheric concentrations of CO2 (ppm), CH4 (ppb) and N2O (ppb), between 1800 and 2017—European Environment Agency.* (n.d.). [Data Visualization]. Retrieved April 18, 2023, from [https://www.eea.europa.eu/data-and-maps/daviz/atmospheric-concentration-of-carbon-dioxide-5#tab-chart\\_5\\_filters=%7B%22rowFilters%22%3A%7B%7D%3B%22columnFilters%22%3A%7B%22pre\\_config\\_polutant%22%3A%5B%22CH4%20\(ppb\)%22%5D%7D](https://www.eea.europa.eu/data-and-maps/daviz/atmospheric-concentration-of-carbon-dioxide-5#tab-chart_5_filters=%7B%22rowFilters%22%3A%7B%7D%3B%22columnFilters%22%3A%7B%22pre_config_polutant%22%3A%5B%22CH4%20(ppb)%22%5D%7D)
- Types of pollutants.* (n.d.). Retrieved April 18, 2023, from <https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants>
- Worasawate, D., Sakunasinha, P., & Chiangga, S. (2022). Automatic Classification of the Ripeness Stage of Mango Fruit Using a Machine Learning Approach. *AgriEngineering*, 4(1), 32–47. <https://doi.org/10.3390/agriengineering4010003>
- Wu, J., Shen, J., Xu, M., & Shao, M. (2021). A novel combined dynamic ensemble selection model for imbalanced data to detect COVID-19 from complete blood count. *Computer Methods and Programs in Biomedicine*, 211, 106444. <https://doi.org/10.1016/j.cmpb.2021.106444>
- Zhang, Y., Liu, J., & Shen, W. (2022). A Review of Ensemble Learning Algorithms Used in Remote Sensing Applications. *Applied Sciences*, 12(17), 8654. <https://doi.org/10.3390/app12178654>