

Menerapkan Metode Klasifikasi pada Data Uji Emisi Kendaraan di Jakarta dengan Menggunakan Jupyter Notebook

Adamara Muzaky¹, Firman Arifianto², Retno Hendrowati³, Muhammad Darwis⁴

Program Studi Teknik Informatika Universitas Paramadina¹²³⁴

adamara.muzaky@students.paramadina.ac.id¹, firman.arifianto@students.paramadina.ac.id²,
retno.hendrowati@paramadina.ac.id³, muhammad.darwis@paramadina.ac.id⁴

Informasi Artikel:

Received: 15 Maret 2024, Accepted: 04 April 2024, Published: 09 April 2024

Abstract

This research aims to apply classification methods using Jupyter Notebook on vehicle emission test data in Jakarta. The rapid growth of vehicles and urbanization in Jakarta has led to increased air pollution, triggering concerns among the government and city residents. This research aims to identify hidden patterns, relationships, and trends in emission test data through a Data Mining approach with classification analysis methods. It is hoped that this research will significantly contribute to understanding the characteristics of vehicle emissions in Jakarta, including identifying dominant pollutants and factors influencing vehicle emissions.

Keywords: Uji Emisi Kendaraan, Klasifikasi, *Data Mining*, *Jupyter Notebook*

1. Pendahuluan

Uji emisi kendaraan saat ini menjadi topik pembicaraan umum di masyarakat, khususnya di Jakarta. Upaya pemerintah untuk mengurangi polusi udara melalui uji emisi kendaraan menjadi sangat penting seiring pertumbuhan kendaraan dan urbanisasi yang pesat di ibu kota. Penelitian ini bertujuan menerapkan metode klasifikasi pada data uji emisi kendaraan di Jakarta menggunakan *Jupyter Notebook*. Pertumbuhan jumlah kendaraan sejalan dengan peningkatan penggunaannya menjadi sumber masalah serius terkait tingginya tingkat emisi kendaraan yang mencemari kualitas udara di Jakarta.

Meskipun emisi kendaraan merupakan kontributor utama, perlu diakui bahwa faktor lain, seperti pembakaran sampah dan emisi dari pabrik, juga turut berkontribusi terhadap pencemaran udara. Uji emisi, sebagai serangkaian tes untuk mengukur jumlah polutan yang dikeluarkan oleh kendaraan saat beroperasi [1], menjadi krusial untuk menjaga agar kendaraan mematuhi standar emisi yang ditetapkan oleh pemerintah atau badan pengawas lingkungan.

Uji emisi bertujuan untuk mengurangi polusi udara yang dihasilkan oleh kendaraan bermotor [2]. Di sisi lain, uji emisi juga memberikan informasi akurat tentang kondisi kendaraan dan

pembakaran bahan bakar dalam mesin kendaraan. Oleh karena itu, uji emisi kendaraan dapat dianggap sebagai kontribusi positif dalam upaya mengendalikan pencemaran udara.

Kendaraan, sebagai kontributor utama emisi polutan udara, memberikan dampak negatif terhadap kesehatan manusia dan kelestarian lingkungan. Polusi udara yang diakibatkan oleh emisi kendaraan dapat menyebabkan masalah kesehatan, mulai dari penyakit pernapasan hingga risiko kanker paru-paru yang dihadapi oleh penduduk kota. Dengan jumlah penduduk yang terus meningkat di Jakarta, risiko pencemaran udara semakin meningkat. Serta meningkatnya jumlah pembangunan juga berdampak langsung terhadap masalah perekonomian masyarakat. Sebab itu, program urbanisasi dilakukan oleh masyarakat desa untuk mencari pekerjaan agar meningkatkan ekonomi mereka di Jakarta [3].

Diharapkan penelitian ini mampu memberikan kontribusi dalam mengidentifikasi polutan dominan, perubahan pola seiring waktu, dan faktor-faktor yang memengaruhi emisi kendaraan di Jakarta. Melalui pendekatan ini, diharapkan dapat memberikan sumbangan penting dalam menjaga kualitas udara di daerah perkotaan, terutama di Jakarta sebagai pusat aktivitas urban yang padat penduduk. Dengan memanfaatkan teknologi informasi dan metode klasifikasi, penelitian ini juga bertujuan meningkatkan efisiensi analisis data emisi kendaraan, mendukung pengambilan kebijakan yang lebih akurat dan berkelanjutan.

2. Kajian Pustaka

2.1. Data Mining

Data mining merupakan suatu alat yang memungkinkan pengguna untuk mengakses data dalam jumlah besar dengan cepat melalui alat atau aplikasi yang menggunakan analisis statistik pada data [4]. Sebagai pendekatan utama, metode klasifikasi didukung oleh aplikasi *Jupyter Notebook*. Penelitian ini bertujuan untuk mengidentifikasi pola tersembunyi, hubungan, atau tren dalam data uji emisi, dengan harapan memberikan kontribusi penting dalam pemahaman karakteristik emisi kendaraan di Jakarta.

2.2. Jupyter Notebook

Jupyter Notebook dipilih sebagai alat analisis utama karena berfungsi sebagai *editor* dalam bentuk aplikasi *web* yang berjalan di *localhost* pada komputer dalam penelitian ini. *Jupyter Notebook* memungkinkan pengguna untuk menulis kode *Python*, *equations*, visualisasi, serta berfungsi sebagai *markdown editor* [5]. Keunggulan *Jupyter Notebook* terletak pada kemampuannya dalam menampilkan kode yang ditulis dan hasil yang tepat, serta sangat cocok digunakan oleh pemula karena dapat dijalankan pada sel per sel.

2.3. Python

Pemilihan *Python* sebagai bahasa pemrograman dan penggunaan *library scikit-learn* yang terintegrasi dalam *Jupyter Notebook* memberikan fleksibilitas dan efisiensi dalam mengimplementasikan metode klasifikasi seperti *Decision Tree*, *Random Forest*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbors (KNN)* [6]. *Python* juga bersifat *open-source* atau gratis untuk digunakan secara publik, sehingga memiliki banyak fitur dan kode yang dapat diperluas untuk meningkatkan kapabilitasnya.

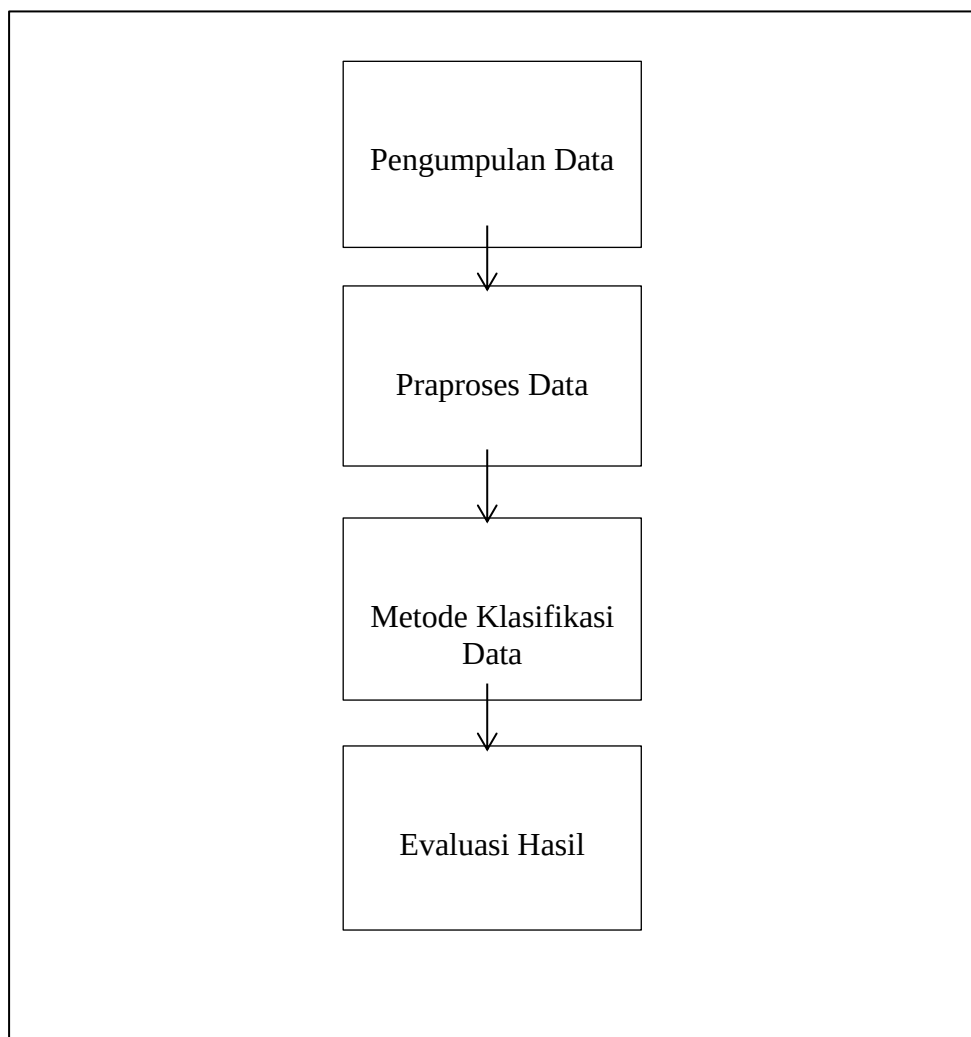
2.4. Penelitian Terdahulu

Penelitian serupa telah dilakukan oleh Agung Prabowo, Sumita Wardani, Rico Wijaya Dewantoro, Wilfredo Wesly, dan Leonardo dengan tujuan untuk membandingkan model

Random Forest dengan *Decision Tree C4.5* guna menentukan metode yang lebih baik dan unggul dalam hal ketepatan akurasi untuk memprediksi keberhasilan metode klasifikasi pada penyakit infertilitas [7]. Penelitian ini juga melibatkan model SVM dan KNN untuk memperluas cakupan analisis, yang sejalan dengan pendekatan yang diambil dalam penelitian yang dilakukan oleh penulis.

3. Metode Penelitian

Penulis mengadopsi metode penelitian ini untuk memastikan kelancaran dan struktur penelitian sesuai dengan tahapan-tahapan yang telah direncanakan sebelumnya. Tujuan utama penerapan metode ini adalah memastikan keberlanjutan dan keteraturan setiap langkah penelitian, sehingga hasil yang akurat dapat dicapai. Panduan visual, seperti yang terlihat pada Gambar 1, membantu penulis dan pihak terkait memahami urutan langkah-langkah penelitian secara jelas.



3.1. Pengumpulan Data

Dataset sangat diperlukan untuk melakukan analisis kebutuhan secara mendalam agar dapat memahami persyaratan dan tujuan dari penelitian yang dilakukan [8]. Pada penelitian ini, penulis menggunakan *dataset uji* emisi kendaraan yang memiliki ekstensi csv, diperoleh dari situs <https://satudata.jakarta.go.id/>.

Dataset ini mencakup hasil uji emisi dari 2300 sampel kendaraan, di mana setiap baris mewakili satu pengujian emisi. Penjelasan mengenai atribut-atribut dalam *dataset* ini dapat ditemukan pada Tabel 1.

Tabel 1. Tabel Uji Emisi Kendaraan

Nama Unit	Tipe Data	Keterangan
tanggal_uji	tanggal	Tanggal dilakukannya uji emisi kendaraan.
nomor_kendaraan	string	Plat nomor kendaraan.
merek	kategorikal	Merek kendaraan.
bahan_bakar	kategorikal	Jenis bahan bakar yang digunakan oleh kendaraan (contoh: Solar).
tipe	kategorikal	Tipe kendaraan (contoh: Sedan, Truk, <i>Pick Up</i>).
tahun_pembuatan	numerik	Tahun pembuatan kendaraan.
alat_uji	string	Jenis alat uji emisi yang digunakan pada saat pengujian.
ket	kategorikal	Keterangan hasil uji emisi (Lulus/Tidak Lulus).

3.2. *Praproses Data*

Praproses Data merupakan tahap awal pada *data mining*, dengan cara mengatasi permasalahan-permasalahan yang ada dan dapat mengganggu hasil dari proses data yang dilakukan. Data yang digunakan juga belum bersih dari *missing value*, duplikasi, dan inkonsistensi data. Dengan dilakukannya praproses data, semua data yang akan digunakan akan dibersihkan, sehingga data yang digunakan menjadi data yang benar-benar memiliki kualitas dan hasil *mining* atau pengetahuan yang lebih [9]. Langkah-langkah ini bertujuan meningkatkan kualitas data, mengurangi ketidaksempurnaan, dan memastikan model *machine learning* dapat memproses data dengan baik.

Dengan melakukan praproses data secara efektif, penulis dapat memastikan bahwa data yang digunakan untuk melatih model adalah data yang bersih, baik, dan dapat dipahami, sehingga meningkatkan peluang keberhasilan model *machine learning*. Pada langkah ini, penulis melakukan praproses data pada *dataset*, termasuk penanganan nilai yang hilang. Proses pengkodean data menggunakan metode sangat penting untuk mengidentifikasi jumlah nilai yang hilang dalam setiap kolom *dataset*. Selain itu, *encoding* data variabel, seperti penerapan *one-hot encoding*, diperlukan untuk mengatasi variabel kategorikal pada kolom 'merek', 'bahan_bakar', dan 'tipe'.

3.3. *Klasifikasi*

Klasifikasi merupakan salah satu metode yang signifikan dalam *data mining*, di mana tujuannya adalah memprediksi kelas yang tidak diketahui [10]. Dalam konteks *dataset* uji emisi kendaraan, klasifikasi digunakan untuk memasukkan kendaraan ke dalam kategori "Lulus" atau "Tidak Lulus" berdasarkan parameter-parameter yang diukur.

Pemilihan metrik evaluasi menjadi krusial dalam analisis klasifikasi, dengan akurasi dipilih karena memberikan gambaran umum sejauh mana model dapat mengklasifikasikan data dengan benar. Selain akurasi, terdapat beberapa metrik penting lainnya:

a) Precision

Mengukur sejauh mana hasil positif yang diprediksi oleh model benar-benar positif, memberikan informasi tentang akurasi model dalam mengidentifikasi kelas positif.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

b) Recall

Mengukur sejauh mana model dapat menemukan semua instance positif yang sebenarnya, memberikan gambaran tentang seberapa baik model menangkap semua *instance* positif.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

c) F1-score

Merupakan rata-rata harmonis dari metrik *precision* dan *recall*, digunakan untuk mengukur keseimbangan antara *precision* dan *recall*, memberikan gambaran performa model secara keseluruhan.

$$F1 - Score = \frac{2x(Precision \times Recall)}{Precision + Recall}$$

d) Support

Jumlah *instance* dalam kelas sebenarnya pada *dataset*, memberikan informasi tentang umumnya suatu kelas dalam *dataset* dan digunakan untuk mengatasi ketidakseimbangan pada kelas.

Dalam evaluasi klasifikasi, kombinasi *precision*, *recall*, dan *F1-score* memberikan gambaran yang holistik tentang kinerja model pada setiap kelas. *Support* membantu memahami distribusi kelas dalam *dataset*.

Pemilihan metrik disesuaikan dengan kebutuhan analisis dan dirangkum melalui dua metode:

e) Macro Average

Rata-rata sederhana dari suatu metrik evaluasi, dihitung secara terpisah untuk setiap kelas dan memberikan bobot yang sama pada setiap kelas tanpa memperdulikan jumlah *instance* pada setiap kelas dalam *dataset*.

$$Macro\ Average = \frac{Precision\ ClassN}{N\ (Jumlah\ Kelas)}$$

f) Weighted Average

Rata-rata yang diberi bobot berdasarkan jumlah *instance* pada setiap kelas, memberikan bobot lebih besar pada kelas dengan lebih banyak *instance*. Metode ini memberikan representasi lebih akurat tentang kinerja model pada keseluruhan *dataset*, terutama ketika terjadi ketidakseimbangan distribusi kelas.

$$\text{Weighted Average} = \frac{\text{Precision ClassN} \times \text{Support ClassN}}{\text{Total Support}}$$

Pada kasus *multiclass classification*, *weighted average* seringkali lebih informatif daripada *macro average* karena memberikan perhatian lebih pada kelas yang lebih besar.

3.4. Cross Validation

Cross-validation digunakan untuk mengukur seberapa baik model dapat berperforma pada data yang tidak terlihat sebelumnya. Ini memberikan gambaran kuat tentang kemampuan umum model dalam membuat prediksi yang baik dan membantu mengevaluasi performa model pada beberapa *subset* data uji yang berbeda.

Berikut adalah beberapa jenis model klasifikasi yang digunakan dalam analisis data:

a) Decision Tree (C4.5)

Decision Tree adalah salah satu model klasifikasi dan prediksi yang dapat mengubah data menjadi pohon keputusan dengan aturan-aturan keputusan. Terdiri dari beberapa aturan yang membagi secara besar, populasi heterogen menjadi populasi yang lebih kecil dan menyebabkan jumlah yang lebih besar dari suatu variabel target [11]. Model ini menggunakan parameter *default* dengan *random_state=42* untuk menjaga reproduktibilitas hasil.

b) Random Forest

Random Forest adalah salah satu model dari metode klasifikasi yang merupakan kumpulan dari model *Decision Tree* yang beroperasi menjadi suatu gabungan secara fungsional. *Random Forest* memiliki hasil yang lebih baik daripada model lainnya karena menggunakan *Decision Tree* yang tidak memiliki korelasi [12]. Model ini menggunakan *parameter default* dengan *random_state=42*.

c) Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu model dari metode klasifikasi yang berakar dari teori pembelajaran statistik, yang diharapkan memberikan hasil yang lebih baik dibandingkan dengan model lain [13]. Model SVM menggunakan *parameter default* dengan *random_state=42*.

d) K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) adalah salah satu model klasifikasi yang dapat mengelompokkan *data testing* pada *data training* berdasarkan pada jarak *data testing* ke beberapa data tetangga terdekatnya. Model ini sangat sering digunakan untuk memprediksi klasifikasi dan menggunakan *data training* yang sudah ada dengan *output* yang sudah diketahui [14]. KNN menggunakan *parameter default* tanpa modifikasi.

4. Hasil dan Pembahasan

4.1. Praproses Data

Tabel 2. Jumlah *Missing Values*

Jumlah <i>Missing Values</i> per Kolom:	
tanggal_uji	0
nomor_kendaraan	0
merek	0
bahan_bakar	0
tipe	0
tahun_pembuatan	0
alat_uji	0
ket	0
dtype: int64	

Tabel 2 menunjukkan bahwa tidak ada nilai yang hilang pada setiap kolom *dataset*, dengan nilai 0 pada setiap kolom, menunjukkan bahwa *dataset* kendaraan tidak memiliki nilai yang hilang.

Tabel 3. *Encode Dataset*

Jumlah tiap tipe:		Jumlah tiap merek:	
tipe_MPV/ MINI BUS	1186	merek_TOYOTA	789
tipe_TRUCK	341	merek_DAIHATSU	377
tipe_PICK UP	239	merek_MITSUBISHI	352
tipe_LAINNYA	157	merek_ISUZU	212
tipe_BOX	153	merek_HINO	147
tipe_SUV	78	merek_HONDA	108
tipe_SEDAN	74	merek_SUZUKI	89
tipe_JEEP	20	merek_NISSAN	61
tipe_BUS	16	merek_HYUNDAI	55
tipe_HATCHBACK	13	merek_KOMATSU	12
tipe_WAGON	6	merek_WULING	10
tipe_CONVERTIBLE	5	merek_FORD	9
tipe_KATO SR 300L	4	merek_SUMITOMO	8
tipe_KOBELCO SK130L	3	merek_MERCEDEZ BENZ	8
tipe_KOBELCO BM 5000	2	merek_KIA	8
tipe_SUMITOMO LS-108RH	2	merek_MAZDA	6
tipe_IHI CC 500	2	merek_VOLKSWAGEN	5
		merek_BMW	4
		merek_PEUGEOT	4
		merek_KATO	4
		merek_VOLVO	4
		merek_KOBELCO	4
		merek_CHEVROLET	3
		merek_KALMAR	3
		merek_CHANGGHAI	2
		(BELCO)	
		merek_TCM	2
		merek_RENAULT	2
		merek_GROVE	2

merek_HELI (FORKLIFT)	2
merek_LIEBHEER	2
merek_SUBARU	1
merek_GOLDEN DRAGON	1
merek_HUAKE (FORKLIFT)	1
merek_MANITOU	1
merek_IHI	1
merek_BOBCAT	1
merek_JCB	1

dtype: int64

Tabel 3 menunjukkan hasil jumlah tiap tipe dan merek kendaraan pada dataset setelah proses encoding variabel kategorikal dilakukan. Dengan mengevaluasi output ini, penulis juga dapat memperoleh wawasan tentang distribusi jenis dan merek kendaraan dalam dataset. Informasi ini dapat menjadi dasar untuk melakukan analisis lebih lanjut terkait pengembangan model machine learning.

4.2. Klasifikasi

Tabel 4. Hasil Accuracy

Model	Precision		Recall		F1-Score		Support		Accuracy
	Lulus	Tidak Lulus	Lulus	Tidak Lulus	Lulus	Tidak Lulus	Lulus	Tidak Lulus	
Decision Tree	97%	67%	99%	32%	98%	43%	442	19	97%
Random Forest	97%	67%	99%	32%	98%	43%	442	19	97%
SVM	96%	0%	1%	0%	98%	0%	442	19	96%
KNN	97%	62%	99%	26%	98%	37%	442	19	96%

Tabel 5. Hasil Macro Average & Weighted Average

Metrik	Random Forest		Decision Tree		SVM		KNN	
	Macro Avg	Weighted Avg	Macro Avg	Weighted Avg	Macro Avg	Weighted Avg	Macro Avg	Weighted Avg
Precision	82%	96%	82%	96%	48%	92%	80%	95%
Recall	65%	97%	65%	97%	50%	96%	63%	96%
F1-Score	71%	96%	71%	96%	49%	94%	68%	96%
Support	461	461	461	461	461	461	461	461

Tabel 4 dan 5 menunjukkan bahwa *output* di atas memberikan evaluasi performa dari empat model klasifikasi yang digunakan terhadap *dataset* uji.

a) Decision Tree

Decision Tree memiliki akurasi sekitar 97%, dengan presisi (Lulus) sekitar 97% dan (Tidak Lulus) 67%, *recall* (Lulus) sekitar 99% dan (Tidak Lulus) 32%, serta *F1-Score* (Lulus) sekitar 98% dan (Tidak Lulus) 43%. *Macro average* dengan presisi 82%, *recall* 65%, dan *F1-Score* 71%. *Weighted average* dengan presisi 96%, *recall* 97%, dan *F1-Score* 96%.

b) *Random Forest*

Random Forest memiliki nilai dan makna serupa dengan *Decision Tree* karena keduanya memberikan hasil yang mirip.

c) *Support Vector Machine (SVM)*

SVM memiliki akurasi sekitar 96%, dengan presisi (Lulus) sekitar 96% dan (Tidak Lulus) 0%, *recall* (Lulus) sekitar 100% dan (Tidak Lulus) 0%, serta *F1-Score* (Lulus) sekitar 98% dan (Tidak Lulus) 0%. *Macro average* dengan presisi 48%, *recall* 50%, dan *F1-Score* 49%. *Weighted average* dengan presisi 92%, *recall* 96%, dan *F1-Score* 94%.

d) *K-Nearest Neighbors (KNN)*

KNN memiliki akurasi sekitar 96%, dengan presisi (Lulus) sekitar 97% dan (Tidak Lulus) 62%, *recall* (Lulus) sekitar 99% dan (Tidak Lulus) 26%, serta *F1-Score* (Lulus) sekitar 98% dan (Tidak Lulus) 37%. *Macro average* dengan presisi 80%, *recall* 63%, dan *F1-Score* 68%. *Weighted average* dengan presisi 95%, *recall* 96%, dan *F1-Score* 96%.

4.3. *Cross Validation*

Tabel 6. Hasil *Cross Validation*

Model	Cross-Validation Accuracy	Deviasi Standar
<i>Decision Tree</i>	95%	2%
<i>Random Forest</i>	96%	2%
SVM	96%	0%
KNN	96%	1%

Pada Tabel 6, *output* dari *cross-validation* menunjukkan hasil akurasi pada setiap model klasifikasi tanpa melibatkan data uji. Angka yang ada dalam kurung menunjukkan deviasi standar, digunakan untuk mengukur sejauh mana hasil pengukuran dapat bervariasi.

a) *Decision Tree*

Decision Tree Classifier memiliki *Cross-validation accuracy* sekitar 95% dengan deviasi standar sekitar 2%.

b) *Random Forest*

Random Forest Classifier memiliki *Cross-validation accuracy* sekitar 96% dengan deviasi standar sekitar 2%.

c) *Support Vector Machine (SVM)*

SVM memiliki *Cross-validation accuracy* sekitar 96% tanpa deviasi standar yang diindikasikan.

d) *K-Nearest Neighbors (KNN)*

KNN memiliki *Cross-validation accuracy* sekitar 96% dengan deviasi standar sekitar 1%.

5. **Kesimpulan**

Penelitian ini berhasil memberikan pemahaman mendalam mengenai karakteristik emisi kendaraan di Jakarta melalui penerapan metode klasifikasi. Implikasi hasil penelitian ini memberikan dasar yang kuat untuk pengambilan kebijakan yang lebih efektif dalam menanggulangi dampak pencemaran udara di Jakarta.

Published by:

Dari hasil klasifikasi, *Decision Tree* dan *Random Forest* menunjukkan tingkat akurasi yang tinggi, sementara *SVM* memiliki performa yang lebih rendah terutama dalam mengklasifikasikan kategori "Tidak Lulus." *KNN* juga memberikan akurasi yang baik, meskipun memiliki *recall* yang rendah pada kategori yang sama. Secara keseluruhan, keempat model menunjukkan akurasi yang konsisten sekitar 95-96%.

Hasil *cross-validation* menegaskan bahwa keempat model klasifikasi yang digunakan menunjukkan performa yang baik dan konsisten dalam melakukan analisis klasifikasi data. *Random Forest* tetap unggul dalam akurasi dibandingkan dengan *SVM*, *KNN*, dan *Decision Tree*. Stabilitas hasil, yang diukur melalui deviasi standar, pada keempat model cukup baik, menunjukkan keandalan model pada berbagai dataset yang digunakan dalam penelitian.

Penelitian ini memberikan dasar pemahaman dan solusi terhadap permasalahan emisi kendaraan di Jakarta. Rekomendasi untuk penelitian mendatang mencakup eksplorasi metode klasifikasi tambahan guna meningkatkan performa prediksi. Selain itu, pengembangan sistem prediktif berbasis model klasifikasi yang mendukung pengambilan kebijakan secara *real-time* juga dapat menjadi fokus penelitian yang menarik. Keseluruhan, penelitian ini menunjukkan bahwa penerapan metode klasifikasi dengan pendekatan *Data Mining* dan *Jupyter Notebook* mampu memberikan kontribusi signifikan dalam penanganan masalah pencemaran udara di Jakarta.

Daftar Pustaka

- [1] Suzuki. (2023). "Pentingnya Uji Emisi Mobil!" [Online]. Tersedia: <https://www.suzuki.co.id/tips-trik/pentingnya-uji-emisi-mobil?pages=all>
- [2] Aryasena, A., Iqbal, A. P., & Kusuma, R. R. (2023). "Melacak Kelemahan: Mengapa Pengimplementasian Kebijakan Uji Emisi di DKI Jakarta Tak Efektif?." *Kultura: Jurnal Ilmu Sosial dan Humaniora*, vol. 1, no. 2, pp. 128.
- [3] Safira, I., Adi, S. W., & Winanti, A. (2023). "Efektifitas Peraturan Gubernur Jakarta Tentang Uji Emisi Terhadap Pencemaran Udara di DKI Jakarta." *Triwikrama: Jurnal Ilmu Sosial*, vol. 1, no. 8, pp. 41.
- [4] Zai, C. (2022). "Implementasi Data Mining Sebagai Pengolahan Data." *Jurnal Portal Data*, vol. 2, no. 3, pp. 3.
- [5] Sandika Rofi Pratama, Ahmad Haidar Mirza. (2021). "Penerapan Data Mining Untuk Memprediksi Tingkat Inflasi Menggunakan Metode Regresi Linier Berganda pada BPS Kota Palembang." *Bina Darma Conference on Computer Science*, vol. 3, no. 2, pp. 246.
- [6] Siti Hapsoh Lubis, Muhammad Iqbal. (2022). "Decision Tree Model for Predicting Work Schedules Using Scikit-Learn." *Jurnal Mantik*, vol. 5, no. 4, pp. 2688.
- [7] Agung Prabowo, Sumita Wardani, Rico Wijaya Dewantoro, Wilfredo Wesly, Leonardo. (2023). "Komparasi Tingkat Akurasi Random Forest dan Decision Tree C4.5 Pada Klasifikasi Data Penyakit Infertilitas." *KLIK: Kajian Jurnal Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 218-219.
- [8] Dede Bagus Setiawan. (2023). "Pengembangan Dashboard Interaktif Covid-19 sebagai Alat Informasi Masyarakat dalam Melihat Pandemi di Indonesia." *Inisiatif: Jurnal Dedikasi Pengabdian Masyarakat*, vol. 2, no. 1, pp. 122.

- [9] Simanjuntak, K. P., & Khaira, U. (2021). "Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering." *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 9.
- [10] Alghifari, F., & Juardi, D. (2021). "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes." *Jurnal Ilmiah Informatika*, vol. 9, no. 2, pp. 76.
- [11] Arnomo, S. A., Fajrin, A. A., Siyamto, Y., & Sadikin, S. F. N. (2023). "Evaluasi Model Decision Tree Pada Keputusan Kelayakan Kredit." *Jurnal Desain dan Analisis Teknologi*, vol. 2, no. 2, pp. 201.
- [12] Sanjaya, J., Renata, E., Budiman, V. E., Anderson, F., & Ayub, M. (2020). "Prediksi Kelalaian Pinjaman Bank Menggunakan Random Forest dan Adaptive Boosting." *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 1, pp. 51.
- [13] Sitorus, C. M., Rizal, A., & Jajuli, M. (2020). "Prediksi Risiko Perjalanan Transportasi Online dari Data Telematik Menggunakan Algoritma Support Vector Machine." *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 2, pp. 255.
- [14] Sitepu, R., & Manohar. (2022). "Implementasi Algoritma K-Nearest Neighbors untuk Klasifikasi Pengajuan Kredit." *Jurnal Sistem Informasi, Teknik Informatika dan Teknologi Pendidikan (JUSTIKPEN)*, vol. 1, no. 2, pp. 49.

Copyrights

The author(s) retain copyright for this article, and the journal is granted first publication rights.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>)

