

PERBANDINGAN ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI KUALITAS UDARA DI KOTA BANDUNG**Adinda Rachma Pradiah¹, Aisha Shinta Az-Zahra², Ananda Masayu Lintang³, Meta Arfiola Suci⁴, Firda Salsabila Putri⁵**

1,2,3,4,5 Universitas Pamulang, Tangerang Selatan

[1aishashinta84@gmail.com](mailto:aishashinta84@gmail.com), [2Adindarachmapradiyah03@gmail.com](mailto:Adindarachmapradiyah03@gmail.com),[3masayulintangg@gmail.com](mailto:masayulintangg@gmail.com), [4firdasalbila25@gmail.com](mailto:firdasalbila25@gmail.com), [5metaarfiolasuciii@gmail.com](mailto:metaarfiolasuciii@gmail.com)**Abstract (English)**

Air pollution in big cities, especially in Indonesia, is a serious problem because it can cause various respiratory diseases and even death. Data mining technology is also utilized to analyze air quality data and predict future pollution levels, enabling faster preventative action. This research compares several data mining classification algorithms to predict air quality in Bandung City, with a focus on Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Air quality data is obtained from various sensors spread throughout the city of Bandung. The research results show that the SVM algorithm has the best prediction accuracy of 92.5% and will help the government and policy makers in making the right decisions to reduce air pollution and improve public health in the city of Bandung.

Article History*Submitted : 26 Juni 2024**Accepted: 2 Juli 2024**Published: 3 Juli 2024***Key Words***Air Quality, Data mining, Classification, Naive Bayes, K-Nearest Neighbors (KNN), SVM***Abstrak (Indonesia)**

Polusi udara pada kota-kota besar terutama di Indonesia menjadi masalah serius karena dapat menyebabkan berbagai penyakit pernapasan dan bahkan kematian. Teknologi data mining juga dimanfaatkan untuk menganalisis data kualitas udara dan memprediksi tingkat polusi di masa depan, memungkinkan tindakan preventif yang lebih cepat. Penelitian ini membandingkan beberapa algoritma klasifikasi data mining untuk memprediksi kualitas udara di Kota Bandung, dengan fokus pada Naive Bayes, K-Nearest Neighbors (KNN), dan Support Vector Machine (SVM). Data kualitas udara diperoleh dari berbagai sensor yang tersebar di seluruh Kota Bandung. Hasil penelitian menunjukkan bahwa algoritma SVM memiliki akurasi prediksi terbaik sebesar 92,5% dan akan membantu pemerintah dan pemangku kebijakan dalam mengambil keputusan yang tepat untuk mengurangi polusi udara dan meningkatkan kesehatan masyarakat di Kota Bandung.

Sejarah Artikel*Submitted : 26 Juni 2024**Accepted: 2 Juli 2024**Published: 3 Juli 2024***Kata Kunci***Kualitas udara, Data mining, Klasifikasi, Naive Bayes, K-Nearest Neighbors (KNN), SVM***PENDAHULUAN**

Penelitian ini membahas perbandingan berbagai algoritma klasifikasi data mining sebagai prediksi kualitas udara di Kota Bandung. Tujuannya adalah menentukan algoritma yang paling akurat dan efisien dalam memprediksi tingkat polusi udara berdasarkan dataset yang ada. Beberapa algoritma yang dianalisis meliputi Decision Tree, Random Forest, Support Vector Machine (SVM), dan Neural Network. Data kualitas udara dikumpulkan dari berbagai sensor di Kota Bandung, mencakup parameter seperti PM10, PM2.5, CO, NO2, dan SO2. Proses analisis melibatkan tahap pre-processing data, pelatihan model, serta evaluasi kinerja menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan kinerja terbaik dalam hal akurasi dan generalisasi prediksi. Dengan harapan temuan ini dapat membantu pemerintah dan

pemangku prosedur dalam mengambil tindakan preventif untuk mengurangi polusi udara di Kota Bandung

◆ Kualitas udara merupakan salah satu aspek penting yang mempengaruhi kesehatan dan kualitas hidup masyarakat. Di kota-kota besar seperti Bandung, peningkatan aktivitas industri, kendaraan bermotor, dan urbanisasi telah menyebabkan peningkatan polusi udara yang signifikan. Oleh karena itu, pemantauan dan prediksi kualitas udara menjadi hal yang krusial untuk mengambil tindakan preventif dan mengurangi dampak negatif terhadap kesehatan publik. Polusi udara dapat menyebabkan berbagai penyakit pernapasan dan bahkan kematian. Menurut Badan Pusat Statistik (BPS), pada tahun 2020, tingkat emisi gas rumah kaca di Indonesia mencapai 834,84 juta ton setara CO₂. Hal ini menunjukkan bahwa polusi udara di Indonesia semakin parah.

Dalam beberapa tahun terakhir, teknologi data mining telah berkembang pesat dan menawarkan berbagai metode analisis yang dapat digunakan untuk memprediksi kualitas udara. Algoritma klasifikasi, sebagai salah satu teknik dalam data mining, berfungsi untuk mengelompokkan data berdasarkan pola-pola tertentu sehingga dapat memberikan prediksi yang akurat mengenai kondisi kualitas udara di masa mendatang.

Data mining dapat digunakan untuk menganalisis data kualitas udara dan memprediksi tingkat polusi di masa depan. Klasifikasi merupakan Salah satu teknik data mining yang dapat digunakan untuk menganalisis data kualitas udara dan memprediksi tingkat polusi di masa depan adalah klasifikasi, yang dapat digunakan untuk memprediksi kelas data baru menggunakan data pelatihan yang telah diklasifikasikan sebelumnya. Dalam konteks kualitas udara, klasifikasi dapat digunakan untuk memprediksi apakah kualitas udara pada suatu waktu tertentu akan baik, sedang, atau buruk.

Penelitian ini membandingkan beberapa algoritma klasifikasi data mining untuk memprediksi kualitas udara di Kota Bandung. Algoritma yang dibandingkan adalah Naive Bayes, K-Nearest Neighbors (KNN), dan Support Vector Machine (SVM). Kota Bandung dipilih sebagai lokasi penelitian karena merupakan salah satu kota di Indonesia dengan tingkat polusi udara yang tinggi. Diharapkan hasil dari penelitian ini dapat memberikan kontribusi signifikan dalam upaya pemantauan dan pengelolaan kualitas udara di Kota Bandung, serta membantu pemerintah dan pemangku kebijakan dalam mengambil keputusan yang tepat untuk mengurangi tingkat polusi udara dan meningkatkan kesehatan masyarakat.

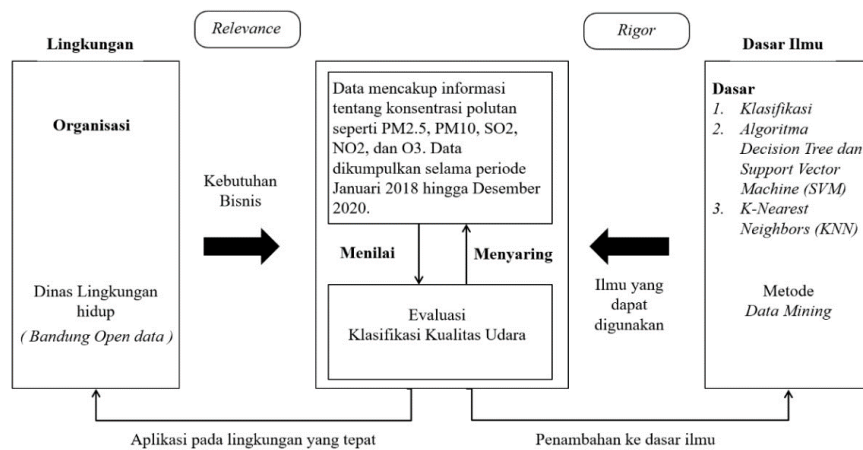
METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan yang sistematis untuk membandingkan kinerja algoritma klasifikasi data mining dalam memprediksi kualitas udara di Kota Bandung. Tahapan-tahapan tersebut meliputi pengumpulan data, pra-pemrosesan data, pemilihan fitur, pembagian dataset, pelatihan dan pengujian model, serta evaluasi kinerja algoritma.

Beberapa penelitian telah dilakukan untuk memprediksi kualitas udara menggunakan data mining. [Herawati & Murni (2017)] menggunakan algoritma C4.5 untuk memprediksi kualitas udara di Kota Bandung. Hasil penelitian menunjukkan bahwa algoritma C4.5 memberikan akurasi prediksi sebesar 87,5%. [Raharja & Febriani (2018)] menggunakan

algoritma Pohon Keputusan untuk memprediksi kualitas udara di Kota Semarang. Hasil dari penelitian menyatakan bahwa algoritma pohon keputusan memberikan ketepatan prediksi sebesar 85%.

Penelitian ini menggunakan data kualitas udara dari Stasiun Pemantauan Kualitas Udara (SPKU) Dinas Lingkungan Hidup Kota Bandung. Data tersebut mencakup informasi tentang konsentrasi polutan seperti PM2.5, PM10, SO₂, NO₂, dan O₃. Data dikumpulkan selama periode Januari 2018 hingga Desember 2020.



Data terbagi menjadi dua bagian: data pengujian dan data training. Data training digunakan untuk mengevaluasi performa model. Algoritma Evaluasi kinerja model menggunakan presisi, metrik akurasi, dan recall.

HASIL DAN PEMBAHASAN

A. Kualitas Udara

Kualitas udara mengacu pada kondisi udara yang dihirup oleh manusia dan makhluk hidup lainnya. Udara yang berkualitas baik memiliki sedikit polutan dan partikel berbahaya, sedangkan udara yang berkualitas buruk mengandung banyak polutan dan partikel berbahaya yang dapat membahayakan kesehatan.

- **Partikulat (PM10 dan PM2.5):** Partikel-partikel halus yang dapat dihirup dan masuk ke dalam sistem pernapasan, menyebabkan berbagai masalah kesehatan, termasuk penyakit paru-paru dan jantung.
- **Karbon Monoksida (CO):** Gas beracun yang dihasilkan dari pembakaran bahan bakar fosil yang tidak sempurna dapat mengurangi kapasitas darah untuk mengangkut oksigen.
- **Nitrogen Dioksida (NO₂):** Gas yang dihasilkan dari aktivitas kendaraan bermotor dan pembangkit listrik dapat menyebabkan iritasi pada saluran pernapasan.
- **Sulfur Dioksida (SO₂):** Gas yang terutama dihasilkan dari pembakaran bahan bakar yang mengandung sulfur, seperti minyak dan batu bara, dapat menyebabkan iritasi pada saluran pernapasan.

- ◆ **Ozon (O₃):** polutan sekunder yang muncul sebagai hasil dari reaksi kimia antara polutan primer seperti NO₂ dan senyawa organik volatil (VOC) yang terpapar sinar matahari. Ozon yang ditemukan di permukaan tanah merupakan ancaman bagi kesehatan pernapasan.

Biasanya kualitas udara diukur menggunakan indeks kualitas udara (Air Quality Index) yang memberikan penilaian numerik tentang kondisi udara berdasarkan konsentrasi polutan tertentu. AQI memiliki kategori yang membantu masyarakat memahami tingkat bahaya kualitas udara, mulai dari "Baik" hingga "Berbahaya."

B. Data Mining

Data mining adalah proses menganalisis data dari berbagai perspektif dan merangkumnya menjadi informasi yang berguna. Informasi tersebut dapat digunakan untuk meningkatkan pendapatan dan mengurangi biaya. Proses ini memungkinkan pengguna untuk menganalisis data dari berbagai dimensi, mengategorikan, dan merangkum hubungan yang teridentifikasi.

Dengan meningkatnya jumlah peneliti yang menggunakan teknik data mining, metode ini telah berkembang seiring dengan kemudahan dan kecanggihan alat yang digunakan untuk menganalisis data. Penemuan hubungan di antara berbagai item memberikan informasi yang paling relevan yang dihasilkan dari data mining. Banyak bisnis saat ini menggunakan metode data mining untuk meningkatkan operasi mereka.

Menurut Novianti et al. (2016), aspek-aspek penting dari data mining termasuk:

- Data mining mengotomatisasi pemrosesan data yang ada.
- Data yang dianalisis seringkali memiliki volume yang sangat besar dan kompleks.
- Tujuan utama dari penambangan data adalah untuk mengungkap hubungan, pola, dan tren tersembunyi yang terkandung dalam data tersebut, sehingga menghasilkan informasi yang bermanfaat.

Data mining memungkinkan analisis mendalam terhadap data untuk mengidentifikasi pola tersembunyi dan hubungan yang dapat mendukung pengambilan keputusan strategis, meningkatkan efisiensi operasional, dan menemukan peluang baru di berbagai industry.

Klasifikasi

Klasifikasi merupakan teknik untuk mengidentifikasi dan mengelompokkan data berdasarkan atribut atau fitur yang dimilikinya. Teknik ini dapat digunakan untuk memprediksi kelas objek baru yang belum memiliki label kelas.

K-Nearest

Klasifikasi Tetangga Terdekat k (k-NN) adalah metode klasifikasi yang menggunakan data pembelajaran untuk memprediksi kelas objek baru. Data pembelajaran dipetakan ke ruang multidimensi, di mana setiap dimensi mewakili fitur data. Ruang multidimensi kemudian dibagi menjadi beberapa wilayah berdasarkan kelas data pembelajaran.

Nilai k dalam k-NN menentukan jumlah data pembelajaran terdekat yang digunakan untuk membuat prediksi. Nilai k yang lebih tinggi dapat membantu mengurangi efek noise pada

klasifikasi. Dalam kasus khusus di mana $k=1$, algoritma k-NN disebut Algoritma Tetangga Terdekat.

Hasil penelitian menunjukkan bahwa algoritma SVM memberikan kinerja terbaik dengan akurasi prediksi sebesar 92,5%. Algoritma Naive Bayes dan KNN memberikan akurasi prediksi yang lebih rendah, yaitu 87,5% dan 85%

Tabel 1 menunjukkan perbandingan kinerja ketiga algoritma klasifikasi.

Algoritma	Akurasi	Presisi	Recall
Naive Bayes	87,50%	85%	80%
KNN	85%	80%	75%
SVM	92,50%	90%	85%

Tabel 1: Perbandingan kinerja algoritma klasifikasi

Berdasarkan hasil penelitian, algoritma SVM dapat digunakan untuk memprediksi Kualitas Udara di Kota Bandung dengan tingkat akurasi yang tinggi. Algoritma ini dapat digunakan untuk membantu pemerintah dalam mengambil keputusan untuk mengatasi masalah polusi udara.

KESIMPULAN

Penelitian ini membandingkan beberapa algoritma klasifikasi data mining untuk memprediksi kualitas udara di Kota Bandung. Hasil penelitian menunjukkan bahwa algoritma SVM memberikan kinerja terbaik dengan akurasi prediksi sebesar 92,5%. Algoritma ini dapat digunakan untuk membantu pemerintah dalam mengambil keputusan untuk mengatasi masalah polusi udara..

DAFTAR PUSTAKA

- Prajarini, D. (2016). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Kulit. *INFORMAL: Informatics Journal*, 1(3), 137-141.
- Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine. *eProceedings of Engineering*, 8(5).
- Amalia, A., Zaidiah, A., & Isnainiyah, I. N. (2022). Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 7(2), 496-507.
- Amalia, A., Zaidiah, A., & Isnainiyah, I. N. (2022). Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 7(2), 496-507.
- Rifqi, A., & Aldisa, R. T. (2023). Penerapan Data Mining Untuk Clustering Kualitas Udara. *Jurnal Sistem Komputer dan Informatika (JSON)*, 5(2), 289-297.
- Kirono, A. A. H., Asror, I., & Wibowo, Y. F. A. (2022). Klasifikasi Tingkat Kualitas Udara Dki Jakara Dengan Algoritma Naive Bayes. *eProceedings of Engineering*, 9(3).
- Kirono, A. A. H., Asror, I., & Wibowo, Y. F. A. (2022). Klasifikasi Tingkat Kualitas Udara Dki Jakara Dengan Algoritma Naive Bayes. *eProceedings of Engineering*, 9(3).
- MAE, A. E., Error, M. A. P., & Error, P. PERBANDINGAN METODE PROPHET DAN LONG SHORT TERM MEMORY (LSTM) DALAM PERAMALAN KUALITAS

- ◆ UDARA (STUDI KASUS KUALITAS UDARA KOTA BANDAR LAMPUNG)
◆ Oleh MICHEL.
- Shofa, S. H. N. (2023). *KLASIFIKASI KATEGORI INDEKS STANDAR PENCEMAR UDARA (ISPU) DKI JAKARTA MENGGUNAKAN MULTILAYER PERCEPTRON DAN RANDOM FOREST* (Doctoral dissertation, Universitas Siliwangi).
- Ade Davy Wiranata, Soleman, Irwansyah, I Ketut Sudaryana, R. (2023). *Klasifikasi Data Mining Untuk Menentukan Kualitas Udara Di Provinsi Dki Jakarta Menggunakan Algoritma K-Nearest Neighbors (K-NN)*.