PENERAPAN ALGORITMA NAIVE BAYES TERHADAP KUALITAS UDARA DI JAKARTA DAN REKOMENDASI AKTIVITAS MASYARAKAT

Belia Putri Salsabila, Prudencia Belva Cynara Trana Putri , Neo Ramadhani, Anggraini Puspita Sari
Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur
Jl. Rungkut Madya No.1 Gn. Anyar, Kec. Rungkut, Surabaya, Jawa Timur 60294
anggarini.puspita.if@upnjatim.ac.id

ABSTRAK

Kualitas udara merupakan faktor penting bagi kesehatan manusia, namun wilayah perkotaan seperti Jakarta mengalami penurunan kualitas udara disebabkan oleh aktivitas industri serta emisi karbon dari kendaraan. Penelitian ini bertujuan untuk mengklasifikasikan dan memprediksi tingkat kualitas udara serta memberikan rekomendasi aktivitas bagi masyarakat menggunakan teknik data mining Naive Bayes. Data diperoleh dari platform kualitas udara Satudata.jakarta.go.id dengan total 259 sampel data. Parameter yang digunakan meliputi PM10, SO2, CO, O3, dan NO2. Data dibagi menjadi 65% untuk pelatihan dan 35% untuk pengujian. Model Naive Bayes ini mencapai tingkat akurasi sebesar 92%, menunjukkan bahwa efektivitas kualitas udara dapat diklasifikasikan menjadi beberapa kategori, seperti "Baik", "Sedang", "Tidak Sehat", dan "Sangat Tidak Sehat", serta memberikan rekomendasi yang akurat, yang menunjukkan kemampuan tinggi untuk melakukan klasifikasi kualitas udara. Oleh karena itu, dengan menggunakan data kualitas udara secara real-time, penelitian ini dapat memberikan rekomendasi kepada masyarakat bagaimana cara yang aman untuk menghadapi polusi udara dan meningkatkan kesadaran mereka. Hasil ini menunjukkan bahwa model ini dapat diandalkan untuk membantu masyarakat dan kebijakan mengurangi dampak negatif penurunan kualitas udara dan memitigasi dampak kesehatan yang disebabkan oleh paparan terus-menerus pada polusi udara.

Kata kunci : *Kualitas udara, Naive Bayes, Klasifikasi, Rekomendasi aktivitas*

1. PENDAHULUAN

Udara merupakan hal penting bagi kehidupan. Udara bersih tidak bercampur dengan zat atau gas merugikan seperti debu, karbon dioksida (CO2), nitrogen dioksida (NO2), dan gas lainnya. Namun, udara di alam tidak selalu bersih, yang menyebabkan penurunan kualitas udara. Penurunan ini terjadi karena aktivitas manusia seperti asap rokok, kegiatan industri, transportasi, dan pembakaran lahan atau hutan [1]. Kualitas udara termasuk salah satu faktor krusial bagi kesehatan manusia. Pada tahun 2016, Organisasi Kesehatan Dunia (WHO) mengestimasikan bahwa lebih dari 90% populasi dunia telah terpapar kualitas udara yang tidak memenuhi standar pedoman WHO.

Indonesia menghadapi polusi udara sebagai salah satu dari banyak masalah lingkungan serius yang dialami oleh kota-kota besar akibat pertumbuhan populasi, peningkatan aktivitas ekonomi, serta aktivitas transportasi dan industri terkait [2]. Dikutip dari laporan Badan Pusat Statistik Indonesia 2024, Jakarta sebagai ibu kota Indonesia dan juga kota terbesar di Indonesia, memiliki populasi sekitar 10,6 juta jiwa per 2024 dengan laju pertumbuhan penduduk 0,31% empat tahun terakhir, memiliki tingkat ekonomi aktif, yang turut serta meningkatkan pencemaran udara di kota tersebut. Berdasarkan penelitian yang dilakukan oleh Santoso dan rekan-rekannya (2020) yang berjudul "Assessment of Urban Air Quality in Indonesia."

Kualitas udara yang buruk memiliki efek yang signifikan terhadap kesehatan manusia. Pada tahun 2016, studi yang dilaksanakan oleh World Bank dan Institute for Health Metrics and Evaluation menunjukan bahwa polusi udara adalah penyebab kematian dini terbesar kelima secara global, mengakibatkan sekitar 5,5 juta kematian setiap tahun. Pada tahun 2023, Kementerian Kesehatan mencatat dari sepuluh gangguan kesehatan dengan jumlah tertinggi per 100.000 penduduk, diantaranya adalah penyakit respirasi. Penyakit respirasi ini berupa penyakit paru obstruktif kronik (PPOK) yang terjadi sebanyak 145 kasus dengan 78,3 ribu kematian, kanker paru-paru memiliki 18 kasus dengan 28,6 ribu kematian, selanjutnya pneumonia mencatat 5.900 kasus dengan 52,5 ribu kematian, sementara asma memiliki 504 kasus dengan 27,6 ribu kematian sedangkan Coarse particulate matter (PM10) menyerang ke dalam saluran pernapasan, mengakibatkan penyakit respirasi yang parah [3].

Penggunaan algoritma Naive Bayes pada penelitian ini merupakan langkah strategis untuk mengoptimalkan prediksi kualitas udara di Jakarta. Naive Bayes sebagai salah satu algoritma yang sangat efisien untuk klasifikasi dan memiliki keunggulan dalam kecepatan pelatihan, bahkan ketika bekerja dengan dataset besar dan beragam fitur. Kemampuan algoritma ini untuk memberikan hasil prediksi secara realtime menjadikannya sangat berguna dalam membangun model pembelajaran mesin yang cepat dan responsif [4]. Tujuan penelitian ini adalah untuk mengembangkan model prediksi kualitas udara di Jakarta menggunakan algoritma Naive Bayes, yang diharapkan mampu memberikan hasil prediksi secara cepat dan akurat. Algoritma ini memungkinkan analisis yang lebih mendalam terhadap pola dan tren polusi udara sehingga dapat memberikan pemahaman bagi pembuat kebijakan dan masyarakat untuk meningkatkan kualitas udara secara keseluruhan.

2. TINJAUAN PUSTAKA

Penelitian ini menggunakan metode Naive Bayes untuk menganalisis dan mengoptimasi kualitas udara di Jakarta, yang telah menjadi perhatian serius akibat polusi udara tinggi. Algoritma Naive Bayes adalah pengklasifikasi statistik yang memprediksi probabilitas keanggotaan suatu kelas berdasarkan teorema [5]. Algoritma Naive Bayes memiliki efisiensi memori tinggi karena hanya menyimpan beberapa statistik sederhana.

2.1. Data Mining

Data mining merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Selain itu, data mining juga digambarkan sebagai kumpulan teknik yang digunakan untuk mendeteksi kejadian-kejadian yang ganjil dalam data yang telah dikumpulkan [6]. Proses Knowledge Discovery in Databases (KDD). Proses ini meliputi pembersihan data (data cleaning), integrasi data (data integration), dan pemilihan data (data selection) [7].

2.2. Naive Bayes Classifier

Naive Bayes Classifier adalah teknik pengkategorian berdasarkan teorema Bayes yang menggunakan probabilitas dan statistik untuk memprediksi kejadian di masa depan dari data sebelumnya. Ciri khasnya adalah asumsi kuat terhadap independensi antara setiap kondisi yang mendasarinya [8]. Asumsi ini menyatakan bahwa fitur-fitur dalam dataset saling bebas, meskipun jarang terjadi di dunia nyata, Naive Bayes sering memberikan kinerja baik. Teknik ini digunakan untuk mengembangkan sistem klasifikasi dalam kondisi tertentu, berdasarkan teorema Bayes. Metode Naive Bayes dari teorema Bayes, sebagai berikut.

$$P(H|X)=(P(X \mid H).P(H))/(P(X))$$

Keterangan:
X : Data dengan class yang belum diketahui
H : Hipotesis data
P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteori probabilitas)
P(H) : Probabilitas hipotesis H (prior probabilitas)
P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis
H P(X) : Probabilitas X

2.3. Confusion Matrix

Confusion Matrix membandingkan jumlah prediksi benar dan salah dari model klasifikasi. Barisnya mewakili kelas data aktual, kolomnya kelas yang diprediksi. Ini membantu menilai kinerja model dan menentukan langkah perbaikan seperti menambah data atau menyesuaikan parameter [9].

Confusion Matrix atau matriks kebingungan, adalah metodologi untuk mengevaluasi performa model kategori biner atau lebih pada suatu dataset. Untuk menilai efisiensi algoritma, penting menghitung berbagai metrik seperti accuracy, precision, recall, dan F1 score [10] Akurasi menunjukkan ketepatan model dalam mengklasifikasikan hasil dengan benar, atau dapat didefinisikan sebagai proporsi prediksi benar dari keseluruhan prediksi. Berikut rumusnya:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Keterangan:
TP (True Positives) : Jumlah prediksi positif yang benar.
TN (True Negatives) : Jumlah prediksi negatif yang benar.
FP (False Positives) : Jumlah prediksi positif yang salah.
FN (False Negatives) : Jumlah prediksi negatif yang salah.

Precision menunjukkan rasio prediksi positif yang benar dibandingkan dengan jumlah keseluruhan hasil yang diprediksi positif. Dengan kata lain, precision menggambarkan persentase prediksi positif yang benar sesuai dengan kriteria yang diminta. Berikut rumusnya:

$$Precision = \frac{TP}{TP+FP}$$

Selanjutnya, recall menunjukkan keberhasilan model untuk menemukan kembali suatu informasi atau dapat disebut sebagai rasio prediksi positif yang benar dibandingkan dengan keseluruhan data yang benar–benar positif berikut rumusnya:

$$\text{F1 Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Kemudian, F1 Score menunjukkan perbandingan rata–rata dari precision dan recall. F1-Score digunakan sebagai acuan kinerja atau performa algoritma bila jumlah data *false negative* dan *false positive* tidak mendekati..

3. METODE PENELITIAN

Bagan alir atau flowchart adalah representasi grafis dari langkah-langkah dalam suatu sistem atau prosedur, menggunakan simbol untuk menunjukkan operasi dan aliran kontrol. Flowchart berguna untuk memeriksa, mengembangkan, merekam, atau mengawasi proses di berbagai sektor.

Gambar 1. Bagan alir

Berikut ini adalah keterangan dari bagan alir diatas

3.1. Pengumpulan Data

Dataset dapat diperoleh dari satu data Jakarta, dataset tersebut meliputi berbagai parameter seperti PM10, SO2, CO, O3, NO2. Data kategori untuk kualitas udara biasanya didapatkan dari standarisasi kualitas yang sudah ditentukan oleh badan lingkungan hidup atau organisasi kesehatan. Di Indonesia, data ini sering kali diambil dari:

a. Indeks Standar Pencemar Udara (ISPU): ISPU standar yang digunakan oleh Kementerian Lingkungan Hidup dan Kehutanan (KLHK) Indonesia untuk mengukur dan melaporkan kualitas udara. ISPU mengelompokkan kualitas udara ke dalam beberapa kategori berdasarkan konsentrasi polutan udara seperti PM10, SO2, CO, O3, dan NO2.

b. Air Quality Index (AQI): AQI adalah indeks yang digunakan secara internasional untuk mengukur dan melaporkan kualitas udara. AQI mengklasifikasikan kualitas udara ke dalam beberapa kategori berdasarkan konsentrasi polutan udara utama.

Tabel 1. Dataset

Tanggal	Pm10	So2	CO	O3	No2	Max	Critical	Categori
08/01/2018	62	32	16	48	6	62	PM10	Sedang
08/02/2018	45	32	11	37	7	45	PM10	Baik
08/03/2018	36	34	12	55	9	55	O3	Sedang
08/04/2018	37	38	9	43	7	43	O3	Baik
08/05/2018	37	33	13	47	9	47	PM10	Baik
08/06/2018	51	13	8	51	72	72	PM10	Sedang
08/07/2018	26	72	7	29	5	72	SO2	Sedang
08/08/2018	22	72	11	42	80	72	SO2	Sedang
08/09/2018	18	70	7	38	5	70	SO2	Sedang
dst…								

c. Setiap polutan memiliki ambang batas yang berbeda untuk setiap kategori berdasarkan pedoman yang ditetapkan oleh ISPU atau AQI. Klasifikasi Kategori Berdasarkan Konsentrasi Polutan Setiap kategori dalam ISPU atau AQI didasarkan pada ambang batas konsentrasi polutan tertentu. Berikut adalah salah satu contohnya pada PM10, sebagai berikut :
   - Baik: 0-50 µg/m³
   - Sedang: 51-150 µg/m³
   - Tidak Sehat: 151-350 µg/m³
   - Sangat Tidak Sehat: 351-420 µg/m³
   - Berbahaya: >420 µg/m³

3.2. Persiapan Data

Data dimuat ke dalam struktur Data Frame menggunakan pustaka pandas dan kolom kategori (target) dihapus dari dataset sebagai langkah awal pre-processing.

3.3. Pembagian Data

Penyebaran data dalam penelitian ini memenuhi beberapa tujuan signifikan, yaitu memfasilitasi penilaian kemanjuran model dalam meramalkan data yang tidak dikenal, menawarkan pemahaman tentang kapasitas model untuk merampingkan. Selain itu, prosedur ini membantu mencegah overfitting, keadaan di mana model terlalu berkonsentrasi pada data pelatihan dan tidak dapat memprediksi data baru secara tepat [11]. Data dibagi menjadi dua bagian, yaitu : data pelatihan dan data pengujian. Dari 259 data total, 91 data digunakan untuk data pelatihan. Hal ini menunjukkan bahwa 65 persen dari proporsinya untuk data pelatihan dan 35 persen untuk data pengujian.

3.4. Preprocessing Data

Fitur numerik disesuaikan agar berada pada skala yang seragam, sementara fitur kategorikal dikonversi ke format yang lebih mudah dipahami oleh model. Langkah-langkah ini memastikan bahwa semua jenis data dapat diproses dengan baik, meningkatkan efektivitas serta akurasi prediksi model machine learning.

3.5. Pemodelan Naive Bayes

Model klasifikasi Naive Bayes diimplementasikan dalam penelitian ini untuk memprediksi kualitas udara dan memperoleh hasil akurasi pada data uji guna mendukung pengambilan keputusan yang tepat dalam manajemen lingkungan. Proses pelatihan model dilakukan menggunakan data pelatihan, diikuti dengan evaluasi kinerja model menggunakan data pengujian.

3.6. Evaluasi Model

Kinerja model dievaluasi menggunakan berbagai metrik, termasuk akurasi, confusion matrix, dan classification report, untuk memberikan gambaran menyeluruh tentang seberapa baik model dalam memprediksi kualitas udara. Selain itu, visualisasi performa model dilakukan menggunakan plot confusion matrix untuk mempermudah interpretasi hasil dan identifikasi area.

3.7. Prediksi dan Rekomendasi

Setelah dilatih dengan data pelatihan, model kemudian diterapkan pada data uji untuk membuat prediksi terkait kualitas udara. Hasil prediksi tersebut tidak hanya berfungsi sebagai prediksi kualitas udara, tetapi juga sebagai dasar untuk memberikan rekomendasi aktivitas kepada masyarakat Jakarta.

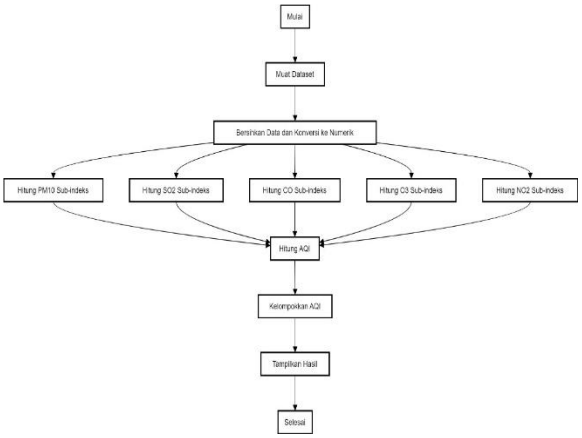
4. HASIL DAN PEMBAHASAN

Penelitian ini berfokus pada penggunaan algoritma Naive Bayes untuk memprediksi kualitas udara di Jakarta. Hasil dan pembahasan meliputi tahap pre-processing data, pelatihan model dengan algoritma Naive Bayes, evaluasi kinerja model, dan analisis hasil prediksi atau kategori rekomendasi.

4.1. Tahap Pre-processing Data

Tahap ini digunakan untuk menetapkan bahwa dataset yang diterapkan dalam analisis dan model prediksi memiliki kualitas yang baik, bersih dari nilai hilang, dan siap digunakan dalam tahap selanjutnya penelitian ini. Tahap pre-processing data ini dimulai dengan memuat dataset yang berisi informasi tentang kualitas udara di Jakarta dari file CSV menggunakan library pandas.

Langkah pertama adalah menampilkan beberapa baris pertama data untuk memastikan data telah dimuat dengan benar. Selanjutnya, dilakukan pengecekan untuk mengidentifikasi apakah terdapat nilai yang hilang dalam dataset. Nilai yang hilang dalam data diimputasi dengan nilai rata-rata untuk memastikan konsistensi data yang akan digunakan



Gambar 2. Flowchart Pre-processing Data

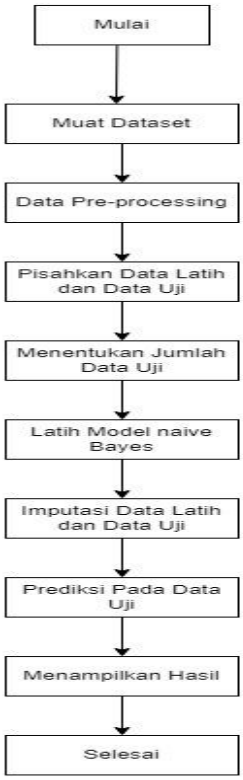
dalam pelatihan dan pengujian model. Informasi umum tentang dataset, seperti jumlah baris, kolom, tipe data, dan jumlah nilai yang hilang di setiap kolom juga ditampilkan. Kolom ‘tanggal’ diatur sebagai indeks untuk memudahkan analisis berbasis waktu. Data kemudian dibersihkan dan dipersiapkan untuk analisis lebih lanjut dengan mengkonversi nilai dalam kolom-kolom yang relevan ke tipe data numerik.

Dataset ini terdiri dari 259 data kualitas udara yang diambil dalam rentang waktu dari 1 Agustus 2018 hingga 16 April 2019. Setelah itu, sub-indeks untuk setiap polutan dihitung berdasarkan standar yang telah ditetapkan. Air Quality Index (AQI) diterapkan sebagai nilai maksimum dari sub-indeks PM10, SO2, CO, O3, dan NO2. Tahap selanjutnya, nilai AQI dikelompokkan ke dalam klasifikasi kualitas udara seperti “Baik”, “Sedang”, “Tidak Sehat”, “Sangat Tidak Sehat”, dan “Berbahaya”, berdasarkan kriteria standar yang ditetapkan.

Data kemudian dipisah menjadi fitur X dan label Y, dimana fitur adalah parameter kualitas udara (PM10, SO2, CO, O3, NO2) dan label adalah klasifikasi kualitas udara. Setelah pemisahan, data dibagi menjadi set pelatihan dan set pengujian dengan perbandingan 65:35, yang berarti sekitar 91 data digunakan untuk tujuan pelatihan berbagai model dan sekitar 35% data digunakan untuk melatih model. Pembagian ini memastikan bahwa model dapat dievaluasi menggunakan data yang belum diamati sebelumnya. Oleh karena itu, kinerja model bisa dievaluasi dengan lebih akurat.

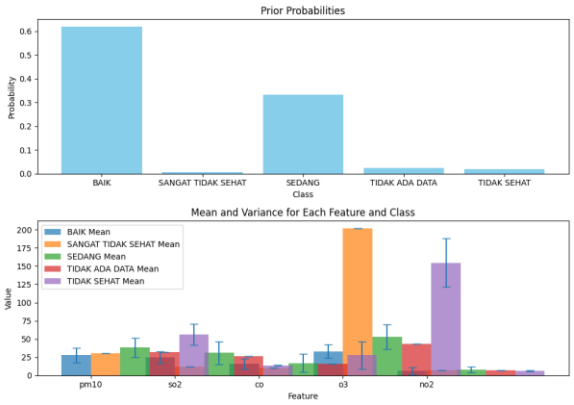
4.2. Pelatihan Model dengan Algoritma Naive Bayes

Pelatihan model dengan algoritma Naive Bayes merupakan pendekatan yang efektif dalam mengklasifikasikan data berdasarkan probabilitas. Dalam konteks penggunaan model ini untuk menganalisis kualitas udara di Jakarta, langkah pertama adalah memisahkan data menjadi dua bagian, yaitu data pelatihan yang digunakan untuk melatih model dan data pengujian yang digunakan untuk menilai kinerjanya. Data latih akan memahami pola dan hubungan antara parameter kualitas udara seperti PM10, SO2, CO, O3, dan NO2 dengan variabel target kategori kualitas udara (misalnya, "Baik", "Sedang", "Tidak Sehat", dan “Sangat Tidak Sehat”).



Gambar 3. Flowchart Data Latih Naive Bayes

Setelah data dibagi, model Naive Bayes dilatih dengan menggunakan data latih. Model ini memanfaatkan asumsi naif yang menyatakan bahwa semua fitur (parameter kualitas udara) dalam dataset adalah independen satu sama lain terhadap kelas kualitas udara yang ditetapkan. Namun, meskipun asumsi ini sederhana, Naive Bayes seringkali memberikan hasil klasifikasi yang baik dalam praktik, terutama ketika jumlah fitur atau dimensi data tinggi.



Gambar 4. Model naive bayes gaussian

Dalam proses pelatihan, terdapat dua diagram visualisasi dihasilkan, yaitu : Diagram pertama yang menunjukkan prior probabilities dari setiap kelas kualitas udara yang merepresentasikan visual dari seberapa sering masing-masing kategori kualitas udara muncul dalam dataset pelatihan. Prior probabilities ini adalah probabilitas bahwa sebuah sampel baru yang diuji akan termasuk dalam setiap kategori kualitas udara sebelum melihat data baru tersebut. Pada diagram ditampilkan data historis kategori "Baik" mencapai probabilitas 0.6 atau 60% yang berarti kejadi udara dengan kategori "Baik" lebih sering terjadi dibandingkan dengan udara yang memiliki kualitas "Sangat Tidak Sehat", "Sedang", dan "Tidak Sehat". Diagram kedua menampilkan mean (rata-rata) dari setiap fitur (PM10, SO2, CO, O3, NO2) untuk setiap kelas kualitas udara, serta varians dari masing-masing fitur. Rata-rata ini menggambarkan nilai tengah dari distribusi nilai fitur untuk setiap kategori kualitas udara. Varians mengindikasikan sebaran data di sekitar rata-rata untuk setiap kelas. Error bars pada diagram kedua menunjukkan ketidakpastian atau variasi dalam data. Semakin besar error bars, semakin besar variabilitas nilai fitur di dalam kelas tersebut. Hal ini berguna dalam mengevaluasi keakuratan model dalam mengklasifikasikan sampel baru.

4.3. Evaluasi Model

```
Classification Report:
                precision    recall  f1-score   support

          BAIK       0.96      0.94      0.95        50
       SEDANG        0.90      0.92      0.91        38
TIDAK ADA DATA       1.00      1.00      1.00         1
   TIDAK SEHAT       0.50      0.50      0.50         2

      accuracy                           0.92        91
     macro avg       0.84      0.84      0.84        91
  weighted avg       0.92      0.92      0.92        91

Accuracy: 0.9230769230769231
```

Gambar 5. Hasil akurasi evaluasi model

Evaluasi model digunakan untuk memastikan bagaimana performa model tersebut. Pada evaluasi model, digunakan data uji yang tidak digunakan selama proses training. Dengan evaluasi model, dapat diketahui bagaimana model akan berperforma pada data baru. Evaluasi kinerja model akan mencetak confusion matrix dan classificattion report untuk melihat performa model.

Berdasarkan gambar 3 Hasil Akurasi Evaluasi Model yang diperoleh dari program menggunakan perintah python menunjukkan hasil akurasi sebesar 0.9230769230769231 atau 92%. Hasil ini menunjukkan bahwa model program ini mampu mengklasifikasikan data dengan sangat baik. Hal ini ditunjukan oleh nilai dari kategori "Baik" yang menghasilkan precision 0.96, recall 0.94, f1-score 0.95 untuk 50 sampel data, untuk kategori "Sedang" menghasilkan precision 0.90, recall 0.92, f1-score 0.91 untuk 28 sampel data, untuk kategori "Tidak Ada Data" menghasilkan precision 1.00, recall 1.00, f1-score 1.00 untuk 1 sampel data, dan untuk kategori "Tidak Sehat" yang menghasilkan precision 0.50, recall 0.50, f1-score 0.50 untuk 2 sampel data. Dengan macro average menghasilkan nilai precision 0.84, recall 0.84, f1-score 0.84 sedangkan weighted average precision 0.92, recall 0.92, f1-score 0.92. Akurasi sebesar 92% ini menunjukkan bahwa model telah mempelajari pola dalam data dengan baik, meskipun hasil ini juga dapat mengindikasikan kemungkinan overfitting jika ukuran data yang digunakan untuk pengujian sangat kecil atau tidak cukup beragam.

Testing data digunakan untuk memprediksi kualitas udara berdasarkan parameter-parameter yang telah dikumpulkan sebelumnya. Testing data ini akan mengevaluasi kinerja model Gaussian Naive Bayes (GNB) yang telah dibuat. Hasil prediksi kualitas udara kemudian dibandingkan dengan kualitas udara yang sebenarnya dan dihitung akurasinya menggunakan metrik seperti accuracy score, confusion matrix, dan classificatino report.

4.4. Kategori Rekomendasi

Kategori rekomendasi kualitas udara digunakan untuk memberikan saran kepada masyarakat tentang bagaimana mereka harus berperilaku dalam kondisi kualitas udara yang berbeda-beda. Dalam model prediksi kualitas udara, penggunaan kategori rekomendasi diperuntukkan untuk memberikan saran yang sesuai dengan kualitas udara yang telah diprediksi. Kategori rekomendasi kualitas udara terdiri dari enam level, yaitu: "Baik", "Sedang", "Tidak Sehat", dan "Sangat Tidak Sehat". Setiap kategori memiliki deskripsi dan rekomendasi yang spesifik. Berikut adalah rincian dari deskripsi dan rekomendasi dari masing-masing kategori :

Tabel 2. Rincian Deskripsi dan Rekomendasi dari tiap Kategori

Kategori	Deskripsi	Rekomendasi
Baik	Kualitas udara baik. Risiko polusi udara sangat rendah	Dengan kualitas udara yang baik, disarankan untuk melakukan aktivitas di luar ruangan seperti berolahraga atau berkreasi. Manfaatkan kualitas udara ini untuk kesejahteraan fisik dan mental Anda.
Sedang	Kualitas udara sedang. Polusi udara mungkin berdampak pada individu yang sangat sensitif.	Meskipun kualitas udara masih dapat diterima, individu yang sensitif terhadap polusi udara disarankan untuk melakukan aktivitas di dalam ruangan. Pertimbangkan untuk melakukan aktivitas yang menenangkan seperti yoga atau meditasi.
Tidak Sehat	Kualitas udara dapat berdampak negatif pada kelompok sensitif.	Untuk individu yang sensitif, disarankan untuk membatasi aktivitas luar ruangan. Aktivitas fisik yang intens harus dihindari. Pertimbangkan untuk melakukan aktivitas di dalam ruangan.
Sangat Tidak Sehat	Kualitas udara tidak sehat. Dapat berdampak negatif pada kesehatan umum.	Dengan kualitas udara yang tidak sehat, disarankan untuk semua orang membatasi aktivitas luar ruangan. Jika perlu keluar, pastikan untuk memakai masker dan membatasi waktu di luar.

5. KESIMPULAN DAN SARAN

Penelitian ini menggunakan Algoritma Naive Bayes untuk memantau dan mengoptimalkan kualitas udara di Jakarta, serta memberikan rekomendasi aktivitas bagi masyarakat. Hasilnya menunjukkan bahwa algoritma ini efektif dalam mengklasifikasikan data kualitas udara dan menawarkan panduan praktis bagi masyarakat untuk menjaga kesehatan mereka berdasarkan kategori kualitas udara seperti baik, sedang, dan tidak sehat.

Pemerintah dan badan lingkungan hidup disarankan untuk meningkatkan pemantauan kualitas udara dengan teknologi yang lebih canggih dan algoritma pemrosesan data yang lebih efektif. Selain itu, penting untuk mengedukasi masyarakat tentang dampak polusi udara dan pentingnya mengikuti rekomendasi aktivitas berdasarkan kualitas udara yang dipantau. Upaya untuk mengurangi emisi polutan dari transportasi dan industri melalui kebijakan yang ketat dan penggunaan teknologi ramah lingkungan sangatlah penting. Dengan mengimplementasikan saran-saran ini, diharapkan kualitas udara di Jakarta dapat ditingkatkan, sehingga masyarakat dapat hidup dalam lingkungan yang lebih sehat dan aman.

DAFTAR PUSTAKA

[1] A. Amalia, A. Zaidiah, and I. N. Isnainiyah, "Prediksi kualitas udara menggunakan algoritma K-Nearest Neighbor," *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, vol. 7, no. 2, pp. 496–507, May 2022, doi: 10.29100/jipi.v7i2.2843.

[2] M. Santoso *et al.*, "Assessment of urban air quality in Indonesia," *Aerosol and Air Quality Research*, vol. 20, pp. 2142–2158, Jan. 2020, doi: 10.4209/aaqr.2019.09.0451.

[3] S. Sasmita, D. B. Kumar, and B. Priyadharshini, "Assessment of sources and health impacts of PM10 in an urban environment over eastern coastal plain of India," *Environmental Challenges*, vol. 7, p. 100457, Apr. 2022, doi: 10.1016/j.envc.2022.100457.

[4] P. Sarang, *Thinking data science*. 2023. doi: 10.1007/978-3-031-02363-7.

[5] K. Kartarina, N. K. Sriwinarti, and N. L. P. Juniarti, "Analisis metode K-Nearest Neighbors (K-NN) dan Naive Bayes dalam memprediksi kelulusan mahasiswa," *JTIM Jurnal Teknologi Informasi Dan Multimedia*, vol. 3, no. 2, pp. 107–113, Aug. 2021, doi: 10.35746/jtim.v3i2.159.

[6] M. F. Rifai, H. Jatnika, and B. Valentino, "Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)," *Petir*, vol. 12, no. 2, pp. 131–144, Sep. 2019, doi: 10.33322/petir.v12i2.471.

[7] A. Y. Alfajr, K. Kartini, and A. P. Sari, "CLASSIFICATION OF DISTRACTED DRIVER USING SUPPORT VECTOR MACHINE BASED ON PRINCIPAL COMPONENT ANALYSIS FEATURE REDUCTION AND CONVOLUTIONAL NEURAL NETWORK," *Jurnal Komputer Dan Informatika*, vol. 11, no. 2, pp. 237–245, Oct. 2023, doi: 10.35508/jicon.v11i2.12658.

[8] N. A. F. Watratan, N. A. Puspita. B, and N. D. Moeis, "Implementasi algoritma Naive Bayes untuk memprediksi tingkat penyebaran Covid-19 di Indonesia," *Journal of Applied Computer Science and Technology*, vol. 1, no. 1, pp. 7–14, Jul. 2020, doi: 10.52158/jacost.v1i1.9.

[9] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *Jurnal Informatika*, vol. 7, no. 1, pp. 1–7, Apr. 2020, doi: 10.31311/ji.v7i1.6203.

[10] J. P. Gultom and A. Rikki, "Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan Pola Penerima Beras Raskin," *KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer)*, pp. 11–19, Apr. 2020, doi: 10.54367/kakifikom.v2i1.664.

[11] M. D. Hendriyanto, A. A. Ridha, and U. Enri, "Analisis sentimen ulasan aplikasi Mola pada Google Play Store menggunakan algoritma Support Vector Machine," *INTECOMS Journal of Information Technology and Computer*

Science, vol. 5, no. 1, pp. 1–7, Apr. 2022, doi: 10.31539/intecoms.v5i1.3708.

[12] J. Lelieveld *et al.*, "Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions," *European Heart Journal*, vol. 40, no. 20, pp. 1590–1596, Mar. 2019, doi: 10.1093/eurheartj/ehz135.

[13] N. N. Rasjid, N. N. Arifin, and N. N. Cahya, "KLASIFIKASI NASABAH BANK LAYAK KREDIT MENGGUNAKAN METODE NAIVE BAYES," *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, vol. 1, no. 1, pp. 01–10, Mar. 2021, doi: 10.55606/juisik.v2i2.187.

[14] D. A. Punkastyo, F. Septian, and A. Syaripudin, "Implementasi data mining menggunakan algoritma Naïve Bayes untuk prediksi kelulusan siswa," *Journal of System and Computer Engineering (JSCE)*, vol. 5, no. 1, pp. 24–35, Jan. 2024, doi: 10.61628/jsce.v5i1.1073.

[15] G. Shaddick, M. L. Thomas, P. Mudu, G. Ruggeri, and S. Gumy, "Half the world's population are exposed to increasing air pollution," *Npj Climate and Atmospheric Science*, vol. 3, no. 1, Jun. 2020, doi: 10.1038/s41612-020-0124-2.

[16] World Health Organization: WHO, "WHO releases country estimates on air pollution exposure and health impact," *WHO*, Sep. 27, 2016. [Online]. Available: https://www.who.int/news/item/27-09-2016-who-releases-country-estimates-on-air-pollution-exposure-and-health-impact