KLASIFIKASI DATA TINGKAT KUALITAS UDARA DI TANGERANG SELATAN MENGGUNAKAN ALGORITMA NAIVE BAYES

Fitri Widiawati, Rudi Kurniawan, Tati Suprapti
Teknik Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No.10B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat 45131, Indonesia
fitriwidiawati76@gmail.com

ABSTRAK

Kota Tangerang Selatan yang memiliki padat penduduk yang tinggi yang harus kita perhatikan kondisi kesehatannya. Indeks Standar Pencemaran Udara (ISPU) merupakan indikator yang mengukur tingkat pencemaran udara berdasarkan konsentrasi beberapa polutan udara banyak data yang muncul tentang kualitas udara pada Tangerang Selatan yang selalu menurun karena berbagai faktor dintaranya faktor fisik dan faktor manusia sehingga menyebabkan pencemaran udara bagi masyarakat setempat. Penelitian ini bertujuan untuk mengklasifikasikan kualitas udara di Tangerang Selatan menggunakan algoritma Naive Bayes dan mengevaluasi faktor-faktor yang berpengaruh terhadap kualitas udara. Model Naive Bayes mencapai tingkat akurasi sebesar 79.38%, menunjukkan keberhasilan dalam mengidentifikasi kondisi kualitas udara. Evaluasi dilakukan melalui *presisi, recall,* dan *F1-score* untuk setiap kelas, memberikan gambaran rinci tentang kinerja model. Temuan menunjukkan presisi yang tinggi pada kelas "Sedang" dan "Baik," menandakan kemampuan model dalam mengenali kondisi tersebut secara akurat. Faktor-faktor yang mempengaruhi kualitas udara diidentifikasi melalui evaluasi model, mencakup partikulat matter (PM2.5 dan PM10), sulfur dioksida (SO2), nitrogen dioksida (NO2), karbon monoksida (CO), dan ozon (O3). Penggunaan atribut-atribut ini memungkinkan model memberikan label dengan nilai Class Moderate (0.583), Class Good (0.358), dan Class Unhealthy (0.058). Hasil penelitian menunjukkan bahwa algoritma Naive Bayes dapat memberikan hasil klasifikasi yang akurat untuk data tingkat kualitas udara di Tangerang Selatan dan hasil analisis ini dapat menjadi dasar bagi masyarakat untuk mengambil tindakan *preventif* dan *mitigasi* terhadap potensi dampak negatif dari pencemaran udara.

***Kata Kunci** : Klasifikasi, Naive Bayes, Data Mining Dan Indeks Standar Pencemaran Udara.*

1. PENDAHULUAN

Udara merupakan hal yang sangat penting bagi kehidupan. Udara yang bersih adalah udara yang tidak tercampur dengan substansi atau gas yang dapat membahayakan, seperti partikel debu, karbon dioksida (CO2), nitrogen dioksida (NO2), dan gas-gas lainnya. Namun, pada kenyataannya, udara alam tidak selalu dalam kondisi bersih. Pada kenyataannya, udara yang terdapat di alam tidak selalu bersih. Hal ini dapat menimbulkan penurunan kualitas udara [1].

Udara telah mencapai tingkat polusi yang tinggi dan tidak sehat bagi kelompok sensitif. Banyak yang merasakan gejala seperti kesulitan bernapas atau iritasi tenggorokan. Konsentrasi PM2.5 di Tangerang saat ini 15,9 kali lipat dari nilai pedoman kualitas udara tahunan WHO. Berdasarkan KEPMEN Negara Lingkungan Hidup NOMOR P.14 tahun 2020 Tentang Indeks Standar Pencemaran Udara yang terdiri dari Partikulat (PM10), Karbon monoksida (CO), Sulfur Dioksida (SO2), Nitrogen Dioksida (NO2), dan Ozon (O3). Sedangkan kategori nilai Rentang ISPU kualitas udara berdasarkan ketetapan pemerintah tersebut diantaranya kualitas udara kategori baik, kualitas udara kategori sedang, kualitas udara kategori tidak sehat, kualitas udara kategori sangat tidak sehat, dan kualitas udara kategori berbahaya (Qosim et al., 2021). Klasifikasi melibatkan langkah-langkah untuk menemukan serangkaian model, pola, atau fungsi yang menggambarkan serta membedakan objek data. Tujuannya adalah mengelompokkan objek-objek tersebut ke dalam kelas-kelas tertentu dari sejumlah kelas yang ada.. Akurasi klasifikasi dihitung dengan menentukan persentase *instance* yang diberi label kelas yang benar. Tujuan klasifikasi untuk dapat memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui Naive Bayes menjadi salah satu algoritma yang terdapat dalam metode klasifikasi (Sang et al., 2021). Metode Naive Bayes dipilih karena dinilai berpotensi baik untuk pengklasifikasian dengan metode probabilitas dan statistik yang dikemukan oleh ilmuan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya. Tujuan dari Naive Bayes Gausian adalah untuk menggambarkan atau menghitung secara langsung atribut data yang memiliki karakteristik kontinu. (Kirono et al., 2022).

Hasil dari perhitungan yang telah dilakukan terhadap data ISPU DKI Jakarta dengan menggunakan teknik klasifikasi *data mining* menggunakan algoritma Naive Bayes, dapat disimpulkan dari permodelan uji skenario menghasilkan hasil yang baik. Hasil dari pengklasifikasian pada data Indeks Standar Pencemaran Udara pada kota DKI Jakarta menghasilkan yaitu, dengan rata-rata akurasi 88%, precision 85%, recall 96%, f1-score 90% [2]. Metode K-Nearest Neighbor dengan Manhattan Distance dan nilai K sebesar 7 memiliki tingkat akurasi yang lebih baik dibandingkan metode Naive Bayes yakni sebesar 97.3396 % berbanding dengan 91.862 % [3]. Penerapan teknik data mining dalam klasifikasi

kualitas udara di DKI Jakarta menunjukkan bahwa algoritma Decision Tree memberikan hasil yang lebih unggul dibandingkan dengan algoritma SVM. Hal ini terlihat dari peningkatan kinerja pada parameter Precision, Recall, F1-Measure, dan akurasi dalam melakukan klasifikasi kualitas udara di wilayah tersebut.Pada penelitian ini, hasil yang didapatkan bahwa rasio terbaik untuk melakukan klasifikasi kualitas udara di DKI Jakarta dari masing-masing algoritma, pada algoritma Decision Tree menggunakan rasio 90:10 dan pada algoritma SVM menggunakan rasio 60:40 karena mencapai tingkat akurasi paling tinggi dari berbagai rasio, seperti 60:40, 70:30, 80:20, dan 90:10. Pada algoritma Decision Tree mendapatkan nilai Precision sebesar 99,02%, Recall 99,73%, F1-Measure 99,37%, Akurasi 99,40% dan pada algoritma SVM mendapatkan nilai Precision sebesar 95,82%, Recall 88,89%, F1- Measure 92,22% dan Akurasi 94,93%. Berdasarkan hasil evaluasi performansi dan klasifikasi performa, algoritma Decision Tree dan algoritma SVM memiliki classifier dengan klasifikasi performa sangat baik berdasarkan rentang nilai dari *confusion matrix* [4]

Penelitian ini bertujuan untuk meningkatkan akurasi dan relevansi model klasifikasi Naïve Bayes dalam memprediksi tingkat kualitas udara di Tanggerang Selatan. Peningkatan tingkat ketepatan dalam prediksi memiliki dampak praktis yang signifikan, termasuk memberikan informasi yang lebih tepat waktu kepada masyarakat dan pihak berwenang, sehingga mereka dapat mengambil tindakan pencegahan atau langkah-langkah mitigasi yang efektif terhadap potensi dampak negatif kualitas udara terhadap kesehatan dan lingkungan.

Menerapkan pendekatan klasifikasi dengan memanfaatkan algoritma Naive Bayes, akan menghimpun data mengenai tingkat kualitas udara dari beragam sumber, data cuaca, dan parameter-parameter lingkungan lainnya di wilayah Tangerang Selatan. Pendekatan eksperimental akan mencakup proses pemberian label kelas pada data, yang didasarkan pada standar kualitas udara yang berlaku. langkah-langkah pelatihan model dengan menggunakan data yang telah diberi label untuk mengoptimalkan parameter dari model Naive Bayes. Evaluasi kinerja model selanjutnya akan dilakukan dengan menggunakan teknik analisis data, termasuk metrik klasifikasi seperti akurasi, presisi, recall, dan F1-score. Dengan mengintegrasikan berbagai faktor yang mempengaruhi melalui pendekatan klasifikasi yang lebih canggih menggunakan algoritma Naive Bayes.

Hasilnya dapat memberikan kontribusi yang signifikan pada pemahaman tentang kualitas udara di Tangerang Selatan melalui pendekatan klasifikasi menggunakan algoritma Naive Bayes. Penelitian ini dapat memberikan pandangan yang lebih akurat dan terperinci tentang faktor-faktor yang memengaruhi kualitas udara. Referensi tentang ANALISIS DAN KOMPARASI ALGORITMA KLASIFIKASI DALAM INDEKS PENCEMARAN UDARA DI DKI JAKARTA, bahwa Neural Network Backpropagation, K-Nearest Neighbors, Support Vector Machine, dan Naive Bayes juga masih dapat digunakan sebagai model klasifikasi yang baik karena mendapatkan nilai akurasi yang tinggi di atas 90% dan nilai kappa di atas 0.8 dan nilai RMSE di bawah 0.3, yaitu dengan mengidentifikasi perbedaan kondisi lingkungan antara DKI Jakarta dan Kota Tangerang yang mungkin memengaruhi kualitas udara [5]. Penerapan klasifikasi tersebut menggunakan algoritma Naive Bayes dimana data dari Indeks Standar Pencemaran Udara (ISPU) yang bersifat kontinu cocok digunakan algoritma Naive Bayes. Klasifikasi pada dataset Indeks Standar Pencemaran Udara (ISPU) perlu dievaluasi untuk menilai sejauh mana keefektifan penerapan algoritma Naive Bayes dalam mengklasifikasikan dataset ISPU [6]

2. TINJAUAN PUSTAKA

Penelitian yang dilakukan oleh [7] membahas tentang Analisis dan Perbandingan Algoritma Klasifikasi dalam Indeks Pencemaran Udara di Daerah Khusus Ibukota Jakarta (DKI Jakarta). Tujuan dari penelitian ini adalah untuk menilai algoritma klasifikasi yang paling efektif dalam memprediksi kualitas udara yang terbagi menjadi kategori baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya dengan menggunakan metode data mining. Hasil penelitian menunjukkan bahwa performa terbaik terdapat pada algoritma klasifikasi Random Forest dalam memprediksi tingkat indeks pencemaran udara di DKI Jakarta, diikuti oleh algoritma klasifikasi Naive Bayes dan Decision Tree.

Penelitian yang dilakukan oleh [8] membahas tentang Sistem Prediksi Tingkat Pencemaran Polusi Udara dengan Algoritma Naive Bayes di Kota Makassar. Tujuan dari penelitian ini adalah untuk mengembangkan sistem prediksi tingkat polusi udara yang dapat membantu masyarakat dan pemerintah dalam mengambil keputusan yang tepat terkait dengan kesehatan dan lingkungan. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes dapat digunakan untuk memprediksi tingkat polusi udara di Kota Makassar dengan akurasi yang cukup tinggi. Selain itu, penelitian ini juga menyarankan pengembangan lebih lanjut dengan memperoleh data training atau data set dari beberapa titik di Kota Makassar.

2.1. Data Mining

Data mining merupakan suatu proses ekstraksi informasi penting dari kumpulan data yang besar. Proses ini melibatkan berbagai konsep seperti statistika, basis data, machine learning, pengenalan pola, kecerdasan buatan, dan visualisasi data. Data yang akan diolah melalui data mining harus mengikuti langkah-langkah Knowledge Discovery in Databases (KDD) dan telah dibagi menjadi dua bagian, yaitu data pelatihan (training) dan data pengujian (testing). Prinsip Pareto juga diterapkan dalam konteks data

mining, yang menyatakan bahwa 80% hasil kinerja dapat berasal dari 20% usaha yang dilakukan. Proses pengolahan data ini memiliki potensi untuk menghasilkan informasi berharga, seperti dalam kasus pengklasifikasian kualitas udara di DKI Jakarta [9].

2.2. Algoritma Naive Bayes

Algoritma Naive Bayes merupakan suatu metode klasifikasi yang memanfaatkan probabilitas dan statistik untuk membandingkan data training dengan data testing. Proses ini melibatkan beberapa tahap persamaan untuk memperoleh probabilitas tertinggi, yang kemudian dianggap sebagai informasi yang dihasilkan. Algoritma Naive Bayes dapat diterapkan pada klasifikasi data baik yang bersifat kontinu, seperti data numerik, maupun kategorikal. Selain itu, algoritma ini mampu memodelkan atau menghitung langsung atribut data yang bersifat kontinu. Dalam konteks pengklasifikasian kualitas udara di DKI Jakarta, Naive Bayes digunakan untuk melakukan uji terhadap data Indeks Standar Pencemaran Udara (ISPU) dengan tujuan mengevaluasi performansi klasifikasi [10].

2.3. Klasifikasi

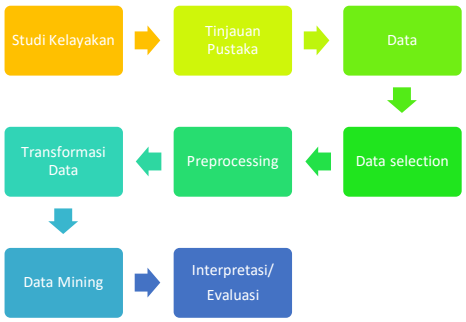
Klasifikasi merupakan suatu proses yang bertujuan untuk menemukan model atau fungsi yang dapat menjelaskan serta menggambarkan konsep atau kelas data untuk tujuan tertentu. Algoritma dalam data mining yang digunakan untuk klasifikasi termasuk K-Nearest Neighbors (K-NN) dan Naive Bayes. K-NN merupakan algoritma pembelajaran mesin yang mengadopsi pendekatan berbasis instansi atau non-parametrik, sementara Naive Bayes merupakan algoritma klasifikasi berdasarkan probabilitas yang membandingkan data training dan data testing. Klasifikasi dapat diaplikasikan untuk melakukan prediksi dan analisis terkait kualitas udara, seperti yang telah diuji dalam penelitian sebelumnya terkait kualitas udara di DKI Jakarta [11].

2.4. RapidMiner

RapidMiner adalah salah satu perangkat yang digunakan dalam tahap desain dan analisis data. Dalam ranah data mining, RapidMiner digunakan untuk membaca file Excel, mempartisi data menjadi set pelatihan dan pengujian, menerapkan model yang telah di-train pada data baru, dan mengevaluasi kinerja algoritma klasifikasi seperti K-Nearest Neighbors (K-NN). Pengguna dapat menentukan lembar kerja dan rentang sel yang akan diproses menggunakan RapidMiner, dan alat ini memberikan hasil evaluasi akurasi berdasarkan metode K-fold. Oleh karena itu, RapidMiner menjadi sebuah alat yang bermanfaat dalam melakukan analisis data dan mengevaluasi performa algoritma klasifikasi [12].

3. METODE PENELITIAN

Metode penelitian untuk klasifikasi data tingkat kualias udara dengan menggunakan algoritma Naive Bayes melibatkan serangkaian langkah yang sistematis untuk mengumpulkan data, memproses data, melatih model dan mengevaluasi hasil. Uraian tahapan metode yang akan digunakan pada penelitian ini dapat dilihat pada Gambar 1 sebagai berikut:



Gambar 1. metode penelitian

4. HASIL DAN PEMBAHASAN

4.1. Hasil

Hasil penelitian dituliskan berdasarkan tahapan penelitian dari langkah awal penelitian hingga akhir.

4.1.1. Data

Data berbentuk file Excel dengan nama Data Kualitas Udara di Tangerang Selatan diperoleh dari Kaggle, dengan jumlah rekord sebanyak 1096 dan terdiri dari 10 atribut. Informasi pada Tabel 4.1 sebagai berikut:

Tabel 1 Dataset Kualitas Udara

No.	Atribut	Tipe Data
1.	Date	Polynominal
2.	PM2.5	Integer
3.	PM10	Integer
4.	SO2	Integer
5.	CO	Integer
6.	O3	Integer
7.	NO2	Integer
8.	Max	Integer
9.	Critical Component	Polynominal
10.	Kualitas Udara	Polynominal

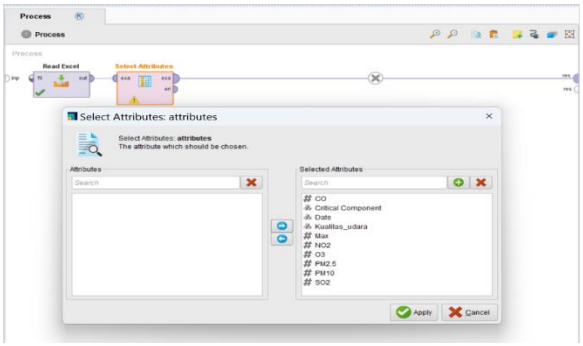
Agar dataset tersebut ada di repository Rapidminer, maka perlu dilakukan import data kualitas udara di tangerang.xlsx menggunakan *operator ReadExcel*, bisa dilihat pada Gambar 2, sebagai berikut:

Gambar 2 Operator *ReadExcel* pada Rapidminer

4.1.2. *Data Selection*

Proses data selection merupakan tahapan yang dilakukan untuk menyeleksi data yang tidak relevan ataupun data yang salah input, yang nantinya data tersebut akan diolah menggunakan rapidminer. Pada

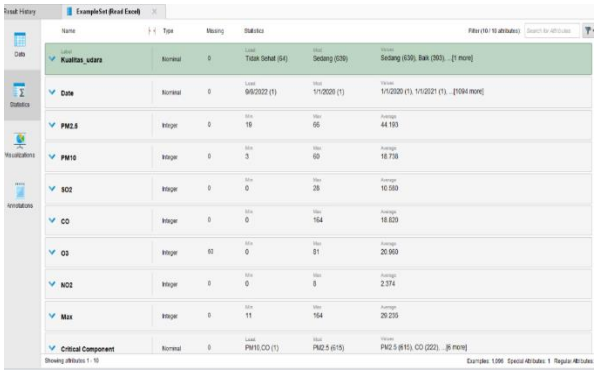
tahap ini, proses pemilihan data yang relevan dan Sesuai dengan tujuan penelitian yang telah ditetapkan, data awal yang digunakan dalam penelitian ini yaitu dataset kualitas udara di wilayah tanggerang sebanyak 1097 data yang terdiri dari 10 atribut yaitu date, PM2.5, PM10, SO2, CO, O3, NO2, MAX, Critical Component. Operator yang digunakan dalam seleksi data yaitu menggunakan select atribut, bisa dilihat pada Gambar 3 sebagai berikut:



Gambar 3 Model Proses Data *Selection*

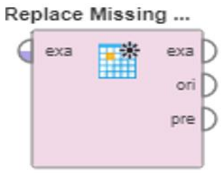
4.1.3. *Pre-Processing*

Proses pembersihan atau cleansing data dari nilai yang hilang atau tidak konsisten dilakukan pada tahap pre-processing. Sebelum melaksanakan langkah ini, dilakukan analisis terlebih dahulu untuk menentukan apakah atribut dalam dataset yang dipilih mengandung nilai yang hilang atau tidak konsisten, dengan mengamati data *statistic* dilihat pada Gambar 4 setelah diketahui ada data yang missing kemudian dilakukan proses pre-processing dengan menggunakan operator *Replace Missing Value*. Bisa dilihat pada Gambar 4 dibawah ini:



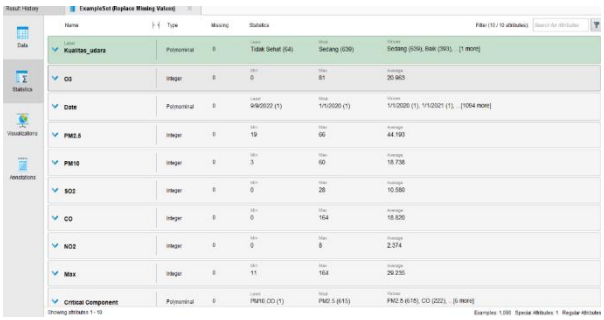
Gambar 4 Data *Statistic*

Berdasarkan hasil dari gambar 4. data statistic terlihat bahwa terdapat data yang missing pada atribut O3 sebanyak 60 data maka langkah selanjutnya dilakukan proses replace missing value.



Gambar 5 Operator *Replace Missing Value* pada Rapidminer

Dari hasil operator Replace Missing Values didapat informasi sebagai berikut

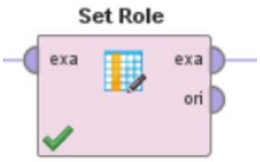


Gambar 6 Data *Statistic*

Dilihat dari hasil gambar 6 Data statistic bahwa atribut O3 setelah dilakukan penerapan operator replace missing value maka atribut tersebut sudah tidak missing hasilnya 0.

4.1.4. *Transformation*

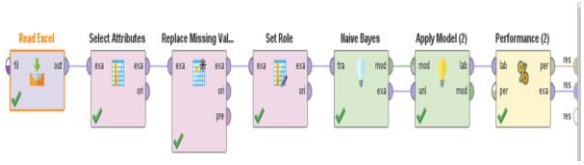
Pada proses tranformasi data dalam penelitian ini menggunakan operator *set role* yang digunakan untuk merubah tipe atribut kualitas udara menjadi label. Adapun operator *set role* terlihat pada Gambar 7 sebagai berikut:



Gambar 7 Operator *set role*

4.1.5. *Data Mining*

Proses pencarian atau pengeksploran Langkah awal dalam penelitian ini adalah mengidentifikasi informasi penting dari kumpulan data yang sangat besar yang telah dipilih dengan menggunakan prosedur atau pendekatan tertentu. Label kualitas udara digunakan untuk mengkategorikan kumpulan data menggunakan algoritma Naïve Bayes. Algoritma Naïve Bayes dan perangkat lunak RapidMiner digunakan untuk pemrosesan data. Model data mining dikembangkan untuk penelitian ini adalah seperti pada Gambar 8 sebagai berikut:



Gambar 8 Model proses data mining Naïve Bayes

Berdasarkan hasil Gambar 9 penerapan dari metode algoritma *Naive Bayes* dalam penelitian ini untuk mengklasifikasikan data kualitas.

Gambar 9 Hasil *Exampleset*

Dilihat dari gambar 9 Hasil Exampleset merupakan hasil dari output exampleset dari operator *performance* bahwa hasil prediksi kualitas udara terklasifikasi menjadi 3 klas yaitu: kualitas udara baik atau sedang dan Kurang Sehat.

Gambar 10 Hasil Visualisasi Kualitas udara

Pada Gambar 10 merupakan gambar hasil visualisasi klasifikasi kualitas udara di daerah tanggerang terklasifikasi menjadi 3 kelas terdiri dari kualitas udara baik, sedang dan tidak sehat.

accuracy: 79.38%

	true Sedang	true Baik	true Tidak Sehat	class preci
pred. Sedang	493	0	0	100.00%
pred. Baik	0	313	0	100.00%
pred. Tidak Sehat	146	80	64	22.07%
class recall	77.15%	79.64%	100.00%	

Gambar 11 Hasil *Accuracy*

Dalam hasil kinerja yang diberikan, terlihat bahwa model atau sistem klasifikasi telah dinilai berdasarkan tiga kategori kualitas udara: "Sedang," "Baik," dan "Tidak Sehat." Berikut adalah interpretasi hasil kinerja tersebut.

a. Presisi (Precision):
   Presisi mengukur seberapa akurat model dalam mengidentifikasi instansi yang benar dari kelas tertentu. Hasil menunjukkan bahwa untuk kelas "Sedang" dan "Baik," model memiliki presisi 100%, artinya semua instansi yang diprediksi sebagai kelas tersebut benar-benar milik kelas tersebut. Namun, untuk kelas "Tidak Sehat," presisi adalah 22.07%, yang menunjukkan bahwa sejumlah besar instansi yang diprediksi sebagai "Tidak Sehat" ternyata sebenarnya bukan.
b. Recall:
   Recall mengukur seberapa baik model dapat menemukan semua instansi yang benar dari suatu kelas. Hasil menunjukkan bahwa untuk kelas "Sedang," recall adalah 77.15%, untuk kelas "Baik" recall adalah 79.64%, dan untuk kelas "Tidak Sehat" recall adalah 100%. Ini menunjukkan bahwa model dapat menemukan sebagian besar instansi dari kelas "Sedang" dan "Baik," tetapi dapat menemukan semua instansi dari kelas "Tidak Sehat."

Ringkasnya, meskipun model memiliki presisi yang tinggi untuk kelas "Sedang" dan "Baik," kinerja pada kelas "Tidak Sehat" kurang memuaskan dalam hal presisi. Sebaliknya, *recall* untuk kelas "Tidak Sehat" sangat tinggi, menunjukkan kemampuan model untuk mengidentifikasi sebagian besar instansi dari kelas tersebut. Evaluasi ini memberikan wawasan tentang kekuatan dan kelemahan model dalam mengklasifikasikan setiap kelas kualitas udara.

4.1.6. *Evaluation*

Dengan menggunakan akurasi sebesar 79.38% sebagai ukuran performa keseluruhan model, kita dapat melihat evaluasi lebih rinci dengan mempertimbangkan metrik-metrik lain seperti *presisi*, *recall*, dan *F1-score* untuk setiap kelas kualitas udara. Berdasarkan hasil evaluasi model yang telah diberikan sebelumnya:

1. Presisi (Precision):
   a. Kelas "Sedang": 100.00%
   b. Kelas "Baik": 100.00%
   c. Kelas "Tidak Sehat": 22.07%
2. Recall:
   a. Kelas "Sedang": 77.15%
   b. Kelas "Baik": 79.64%
   c. Kelas "Tidak Sehat": 100.00%
3. F1-Score:
   a. Kelas "Sedang": Dapat dihitung dari presisi dan recall untuk mendapatkan nilai harmonik yang menggabungkan kedua metrik tersebut.
   b. Kelas "Baik": Dapat dihitung dari presisi dan recall untuk mendapatkan nilai harmonik yang menggabungkan kedua metrik tersebut.

c. Kelas "Tidak Sehat": Dapat dihitung dari presisi dan recall untuk mendapatkan nilai harmonik yang menggabungkan kedua metrik tersebut.

Hasil evaluasi ini memberikan pemahaman yang lebih lengkap tentang kinerja model di setiap kelas kualitas udara. Meskipun akurasi umumnya tinggi, fokus pada presisi dan recall dapat membantu mengidentifikasi area di mana model dapat ditingkatkan, terutama jika ada kepentingan khusus pada salah satu kelas tertentu. Juga, menggabungkan nilai *F1-score* memberikan *perspektif holistik* tentang keseimbangan antara *presisi* dan *recall.*

4.2. Pembahasan

Bagaimana kualitas udara di Tangerang Selatan dapat diklasifikasikan menggunakan algoritma *Naive Bayes* berdasarkan parameter-parameter tertentu?

Hasil menunjukkan bahwa model *Naive Bayes* berhasil mengklasifikasikan kualitas udara di Tangerang Selatan dengan tingkat akurasi sebesar 79.38%. Presisi dan recall untuk setiap kelas ("Sedang," "Baik," dan "Tidak Sehat") memberikan gambaran lebih rinci tentang kinerja model terhadap masing-masing kondisi kualitas udara. Terutama, presisi yang tinggi pada kelas "Sedang" dan "Baik" menunjukkan kemampuan model dalam mengidentifikasi kondisi tersebut dengan akurat.

Apa saja faktor-faktor yang berpengaruh terhadap kualitas udara di wilayah Tangerang Selatan yang dapat dijadikan variabel dalam model klasifikasi *Naive Bayes*?

Informasi tentang faktor-faktor yang berpengaruh tidak langsung terlihat dari hasil evaluasi model *Naive Bayes* dengan hasil *accuracy* 79.38% bahwa atribut yang mempengaruhi kualitas udara yaitu menggunakan parameter-parameter yang relevan, seperti partikulat matter, partikulat matter (PM10), sulfur dioksida , nitrogen dioksida, karbon monoksida (CO), dan ozon (O3). Dapat menghasilkan label dengan nilai *Class Moderate (0.583), Class Good (0.358),* dan *Class Unhealty (0.058).* Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* dapat memberikan hasil klasifikasi yang akurat untuk data tingkat kualitas udara di Tangerang Selatan.

Sejauh mana keakuratan dan kehandalan model *Naive Bayes* dalam mengklasifikasikan kualitas udara di Tangerang Selatan?

Model *Naive Bayes* mencapai akurasi sebesar 79.38%, yang menunjukkan tingkat keberhasilan yang baik secara keseluruhan. Namun, perlu diperhatikan bahwa akurasi sendiri mungkin tidak cukup untuk mengevaluasi kinerja model secara menyeluruh. Oleh karena itu, evaluasi dilakukan melalui *presisi, recall,* dan *F1-score* untuk setiap kelas, memberikan wawasan tentang kemampuan model dalam menangani setiap kondisi kualitas udara secara *spesifik.*

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan dalam penelitian ini, dapat disimpulkan bahwa penggunaan algoritma Naive Bayes terbukti berhasil dalam mengklasifikasika tingkat kualitas udara di Tangerang Selatan. Berdasarkan Indeks Standar Pencemaran Udara (ISPU), dengan nilai *accuracy* mencapai 79.38%. Uji skenario yang dilakukan menunjukkan kinerja yang positif. Untuk studi berikutnya, disarankan untuk memperkuat temuan penelitian dengan mengintegrasikan analisis statistik lebih komprehensif, seperti uji signifikansi atau analisis regresi. Selain itu, melakukan penelitian tambahan untuk membandingkan keefektifan algoritma Naive Bayes dengan algoritma lainnya.

DAFTAR PUSTAKA

[1] A. A. A. Zaidiah, "Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor," JIPI , 2022.

[2] A. I. Kirono A, "Klasifikasi Tingkat Kualitas Udara Dki Jakara Dengan Algoritma Naive Bayes," eProceedings, vol. vol 9, 2022.

[3] M. S. E. Sodiq, "Perbandingan Metode Naive Bayes Dan K-Nearest Neighbor Pada Klasifikasi Kualitas Udara Di Dki Jakarta," jurnal kesehatan lingkungan, 2020.

[4] S. E. Sang A, "Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine," e-Proceeding of Engineering, 2021.

[5] M. G. Ridho I, "ANALISIS KLASIFIKASI DATASET INDEKS STANDAR PENCEMARAN UDARA (ISPU) DI MASA PANDEMI MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," Jurnal Ilmiah, 2023.

[6] A. A. H. Kirono, "Klasifikasi Tingkat Kualitas Udara Dki Jakara Dengan Algoritma Naive Bayes," e-Proceeding of Engineering, 2022.

[7] S. A. U. S, "Analisis Dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di Dki Jakarta," JIKO (Jurnal Informatika dan Komputer), 2021.

[8] R. R. Aini N, "Sistem Prediksi Tingkat Pencemaran Polusi Udara dengan Algoritma Naïve Bayes pada Kota Makassar," Prosiding Seminar Nasional Komunikasi dan Informatika, 2019.

[9] A. A. H. Kirono, "Klasifikasi Tingkat Kualitas Udara Dki Jakara Dengan Algoritma Naive Bayes," e-Proceeding of Engineering, 2022.

[10] A. A. H. Kirono, "Klasifikasi Tingkat Kualitas Udara Dki Jakara Dengan Algoritma Naive Bayes," e-Proceeding of Engineering, 2022.

[11] W. A. S, "KLASIFIKASI DATA MINING UNTUK MENENTUKAN KUALITAS UDARA DI PROVINSI DKI JAKARTA MENGGUNAKAN ALGORITMA K-

NEAREST NEIGHBORS (K …," Journal of Technology Information , 2023.

[12] A. D. S. S. Wiranata, "KLASIFIKASI DATA MINING UNTUK MENENTUKAN KUALITAS UDARA DI PROVINSI DKI JAKARTA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (K …," Journal of Technology Information, 2023.