

**KLASIFIKASI KUALITAS UDARA DI KAWASAN INDUSTRI
MEDAN (KIM) MENGGUNAKAN METODE NAIVE BAYES**

SKRIPSI

DISUSUN OLEH

ZHAHRAH ZHAFIRAH

2009020101



UMSU

Unggul | Cerdas | Terpercaya

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2024

LEMBAR PENGESAHAN

Judul Skripsi : Klasifikasi Kualitas Udara Di Kawasan Industri Medan
(Kim) Menggunakan Metode Naive Bayes
Nama Mahasiswa : Zhahrah Zhafirah
NPM : 2009020101
Program Studi : Teknologi Informasi

**Menyetujui
Komisi Pembimbing**


(Dr. Al-Khowarizmi, S.Kom, M.Kom)
NIDN. 0127099201

Ketua Program Studi


(Fatma Sari Hufagalung, S.Kom, M.Kom)
NIDN. 0117019301

Dekan


(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

**KLASIFIKASI KUALITAS UDARA DI KAWASAN INDUSTRI
MEDAN (KIM) MENGGUNAKAN METODE NAIVE BAYES**

SKRIPSI

Saya menyatakan bahwa karya tulisan ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 1 November 2024

Yang membuat pernyataan



NPM. 200902010

**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Zhahrah Zhafirah
NPM : 2009020101
Program Studi : Teknologi Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**KLASIFIKASI KUALITAS UDARA DI KAWASAN INDUSTRI
MEDAN (KIM) MENGGUNAKAN METODE NAIVE BAYES**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 1 November 2024

Yang membuat pernyataan



Zhahrah Zhafirah

NPM. 2009020129

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Zhahrah Zhafirah
Tempat dan Tanggal Lahir : Sei Kecil, 9 Juli 2002
Alamat Rumah : Medan
Telepon/Faks/HP : 082370917574
E-mail : zhahrazhafira02@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SD HARVARD Medan TAMAT: 2014
SMP : SMPS Galih Agung TAMAT: 2017
SMA : SMK Budi Agung TAMAT: 2020

KATA PENGANTAR



Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Fatma Sari Hutagalung, S.Kom, M.Kom, Ketua Program Studi Teknologi Informasi
4. Bapak Mhd. Basri, S.Si, M.Kom Sekretaris Program Studi Teknologi Informasi
5. Pembimbing saya Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom.
6. Kepada ayah saya Roni Hartono saya ucapkan terimakasih banyak atas seluruh upaya dalam membiayai pendidikan penulis hingga sarjana.
7. Kepada ibu saya Winarsih saya ucapkan beribu terimakasih atas doa dan kasih sayangnya yang tulus, nasihat, motivasi, semangat dan doa yang tiada henti sehingga penulis bisa menyelesaikan studinya hingga sarjana, *I love you so much mom.*
8. Kepada kakak, dan adik-adik saya yang sangat saya sayangi terimakasih telah memberikan motivasi dan perhatiannya.
9. Kepada Sarah Pratiwi dan Nasyiwa Ramadhini selaku sahabat penulis sejak kecil yang selalu selalu memberi semangat,dukungan,dan selalu mendengarkan keluh kesah si penulis agar bisa menyelesaikan skripsi ini, termakasih banyak saya ucapkan.
10. Kepada teman saya Tri Martati Hrp dan Adinda Baghiz Hidayat Lubis juga saya ucapkan terima kasih karna selalu menemani dalam pengerjaan skripsi ini, diucapkan terima kasih.
11. Semua Pihak yang terlibat baik langsung maupun tidak langsung dalam pengerjaan Skripsi ini yang tidak penulis sebutkan satu persatu diucapkan terima kasih.

ABSTRAK

Kemajuan teknologi informasi telah mempengaruhi berbagai aspek kehidupan, termasuk dalam upaya memantau kualitas udara. Udara yang bersih adalah kebutuhan dasar manusia, namun perkembangan teknologi dan peningkatan industri serta jumlah kendaraan bermotor telah menyebabkan penurunan kualitas udara. Pencemaran udara menimbulkan berbagai dampak negatif, termasuk masalah kesehatan dan pemanasan global. Untuk membantu masyarakat dan pemerintah dalam memantau kualitas udara, Penelitian ini menerapkan metode data mining dengan teknik klasifikasi menggunakan Algoritma Naïve Bayes. Metode ini dipilih karena kemampuannya yang efektif dalam memprediksi kualitas udara berdasarkan data historis. Penelitian ini memanfaatkan data dari parameter Indeks Standar Pencemaran Udara (ISPU) untuk merancang model klasifikasi yang mampu membedakan kategori tingkat kebersihan udara, seperti Baik, Moderat, Kurang Sehat, Sangat Tidak Sehat, dan Berisiko. Hasil penelitian ini diharapkan dapat memberikan informasi yang tepat kepada masyarakat tentang kondisi tingkat kebersihan udara di KIM, serta membantu pemerintah dalam upaya pengendalian pencemaran udara.

Kata kunci : Klasifikasi, Naïve Bayes, penambangan data, Indeks Standar Kualitas Udara, Kualitas Udara.

ABSTRACT

Advances in information technology have affected various aspects of life, including efforts to monitor air quality. Clean air is a basic human need, but technological developments and increased industry and the number of motorized vehicles have caused a decline in air quality. Air pollution has various negative impacts, including health problems and global warming. To help the community and government in monitoring air quality, this study applies a data mining method with a classification technique using the Naïve Bayes Algorithm. This method was chosen because of its effective ability to predict air quality based on historical data. This study utilizes data from the Air Pollution Standard Index (ISPU) parameters to design a classification model that is able to distinguish categories of air cleanliness levels, such as Good, Moderate, Less Healthy, Very Unhealthy, and Risky. The results of this study are expected to provide accurate information to the public about the condition of the level of air cleanliness in KIM, as well as assist the government in efforts to control air pollution.

Keywords: Classification, Naïve Bayes, Data Mining, Air Quality Standard Index, Air Quality.

DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITAS.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI.....	iii
RIWAYAT HIDUP	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
BAB I. PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	5
BAB II. LANDASAN TEORI	6
2.1. Data Mining	6
2.2. Naïve Bayes	6
2.3. Klasifikasi.....	7
2.4. Confusion Matrix	7
2.5. Udara	7
2.6. Polusi Udara.....	8

2.7. Python.....	8
2.8. Flowchart	9
BAB III. METODOLOGI PENELITIAN	11
3.1. Metode Penelitian	11
3.2. Pengumpulan Data	11
3.3. Flowchart Sistem Klasifikasi.....	17
BAB IV. HASIL & PEMBAHASAN.....	18
4.1. Pembahasan dan Persiapan Data	18
4.2. Pembagian Data Training dan Testing	19
4.3. Implementasi Metode Naïve Bayes	20
4.3.1. Kategori Kualitas Udara Baik.....	20
4.3.2. Kategori Kualitas Udara Berbahaya	24
4.3.3. Kategori Kualitas Udara Sangat Tidak Sehat.....	27
4.3.4. Kategori Kualitas Udara Sedang	31
4.3.5. Kategori Kualitas Udara Tidak Sehat.....	34
4.4. Confusion Matrix.....	38
BAB V. KESIMPULAN & SARAN	39
5.1. Kesimpulan	39
5.2. Saran.....	39
DAFTAR PUSTAKA	41

DAFTAR TABEL

Tabel 2.8.	Simbol Flowchart	9
Tabel 3.1.	Tahap Penelitian	12
Tabel 3.2.	Data ISPU Bulan Maret 2023	12
Tabel 3.3.	Data ISPU B ulan April 2023	12
Tabel 3.4.	Data ISPU Bulan Mei 2023	13
Tabel 3.5.	Data ISPU Bulan Juni 2023	14
Tabel 4.1.	Rata-rata dan standart Deviasi setiap polutan untuk Kategori Baik	20
Tabel 4.2.	Rata-rata dan standart Deviasi setiap polutan untuk Kategori Berbahaya	24
Tabel 4.3.	Rata-rata dan standart Deviasi setiap polutan untuk Kategori Sangat Tidak Sehat	27
Tabel 4.4.	Rata-rata dan standart Deviasi setiap polutan untuk Kategori Sedang	31
Tabel 4.5.	Rata-rata dan standart Deviasi setiap polutan untuk Kategori Tidak Sehat	34

DAFTAR GAMBAR

Gambar 3.1.	Tahap Penelitian	11
Gambar 3.2.	Flowchart Sistem Klasifikasi	17
Gambar 4.1.	Data Selection	18
Gambar 4.2.	Transformasi data kategori	19
Gambar 4.3.	Variabel X dan Y	19
Gambar 4.4.	Pembagian Data training dan Testing	19
Gambar 4.5.	Confusion Matrix	38

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Kemajuan dalam teknologi informasi telah memberikan pengaruh besar dalam berbagai aspek kehidupan, termasuk di sektor ekonomi, politik, seni, budaya, dan bahkan pendidikan. Lingkungan yang sehat menjadi kebutuhan dasar yang sangat penting bagi semua manusia untuk mempertahankan kelangsungan hidup. Oleh karena itu, udara bersih dan berkualitas sangat diperlukan oleh manusia. (Sang Inez, A., et al., 2021).

Tetapi dalam perkembangan Di era teknologi saat ini, kualitas udara yang baik semakin sulit dijumpai karena pesatnya pembangunan industri dan peningkatan jumlah kendaraan. Oleh karena itu dampak dari semakin pesatnya teknologi komputer dapat mempermudah pekerjaan manusia seperti adanya perancangan sistem klasifikasi kualitas udara yang akan mempermudah masyarakat dalam memperhatikan kualitas baik dan buruknya udara yang dihirup untuk sehari-hari. Karena Udara adalah substansi yang tidak dapat terlihat secara langsung, tetapi dampak dari pencemaran udara dapat dirasakan secara langsung ketika terjadi kontaminasi. (Lahal A, 2021).

Polusi udara merujuk pada kondisi di mana udara terkontaminasi oleh zat kimia, fisik, atau biologis yang dapat membahayakan makhluk hidup dan lingkungan. Dampak negatif dari polusi ini meliputi masalah kesehatan serta kenaikan suhu udara yang signifikan, yang berkontribusi pada pemanasan global. Fenomena ini umum terjadi di kota-kota besar dan area industri yang sering mengeluarkan emisi

gas hasil produksi. Selain itu, jumlah kendaraan bermotor yang semakin meningkat di daerah perkotaan juga berkontribusi pada masalah ini. (Kurniawan, M. O. P., 2024).

Pertumbuhan jumlah kendaraan bermotor di Indonesia berkontribusi terhadap peningkatan emisi gas karbon monoksida (CO), yang menyebabkan pencemaran udara meningkat dari 70% menjadi 80%. Tingginya tingkat polusi udara berakibat pada menurunnya kualitas udara. Dalam rangka mengatasi masalah ini, pengukuran dan pemantauan kualitas udara perlu dilakukan secara berkelanjutan. Indonesia telah menetapkan standar resmi untuk pengukuran kualitas udara, yaitu Indeks Standar Pencemaran Udara (ISPU), sesuai dengan keputusan yang berlaku. Peraturan MENLHK Republik Indonesia Peraturan P.14 tahun 2020 mengatur tentang Indeks Standar Pencemaran Udara, yang meliputi Partikulat (PM10), Partikulat (PM2.5), Karbon Monoksida (CO), Sulfur Dioksida (SO₂), Nitrogen Dioksida (NO₂), Ozon (O₃), dan Hidrokarbon (HC). Dalam peraturan tersebut, kategori nilai Indeks Standar Pencemaran Udara (ISPU) dibagi menjadi beberapa kategori, yaitu Kualitas Udara Baik, Kualitas Udara Sedang, Kualitas Udara Tidak Sehat, Kualitas Udara Sangat Tidak Sehat, dan Kualitas Udara Berbahaya. Penetapan status kualitas udara ini diharapkan mempermudah dan memberikan informasi yang konsisten kepada masyarakat, serta menjadi salah satu langkah pemerintah dalam memantau pencemaran udara. Untuk membantu pemerintah dan masyarakat dalam memahami apakah kualitas udara aman untuk kesehatan atau tidak, diperlukan klasifikasi yang memanfaatkan Teknologi Informasi. Qosim, A. L. (2021).

Berdasarkan isu yang ada, oleh karena itu penelitian ini membutuhkan pengelolaan Data yang dapat memberikan wawasan, yang umumnya dikenal sebagai data mining, berfungsi untuk mengeksplorasi informasi secara mendalam dari kumpulan data yang besar. Penerapan data mining menjadi solusi untuk memahami kualitas udara pada Kawasan Industri Medan (KIM) dengan menggunakan teknik klasifikasi. Menurut (Satrika E., 2023) klasifikasi merupakan Proses pengelompokan data ke dalam beberapa kategori yang relevan dengan masalah yang dihadapi bertujuan untuk menemukan model atau fungsi yang mampu menjelaskan setiap kategori data tersebut dan memisahkannya, sehingga dapat digunakan untuk memprediksi kategori data di masa mendatang. Dalam melakukan klasifikasi kualitas udara KIM menggunakan Algoritma Naïve Bayes.

Metode Naive Bayes dipilih karena dianggap memiliki potensi yang baik dalam pengklasifikasian menggunakan pendekatan probabilitas dan statistik yang diperkenalkan oleh ilmuwan Inggris, Thomas Bayes, yang berfokus pada prediksi kemungkinan di masa depan berdasarkan pengalaman masa lalu. Tujuan dari Naive Bayes Gaussian adalah untuk menggambarkan atau menghitung secara langsung atribut data dengan karakteristik kontinu. (Widiawati, F., et al., 2023).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dibahas sebelumnya, dirumuskan masalah sebagai berikut: Bagaimana cara membangun model klasifikasi untuk menilai kualitas udara menggunakan metode Naïve Bayes, dan apa tingkat akurasi?

1.3 Batasan Masalah

Dalam penelitian ini, telah ditetapkan batasan masalah yang menjadi perhatian utama, yaitu:

1. Data yang digunakan adalah kumpulan data yang berisi pengukuran polusi udara, termasuk Particulate Matter (PM10), Particulate Matter (PM2.5), Sulfur Dioksida (SO₂), Karbon Monoksida (CO), Ozon (O₃), dan Nitrogen Dioksida (NO₂).Max, Critical, dan Kategori yang berada di Kawasan Industri Medan.
2. Data yang digunakan merupakan data Indeks Standar Pencemaran Udara (ISPU) pada bulan Maret 2023 hingga bulan Juni 2023.
3. Adapun metode analisis ini menggunakan pemodelan algoritma Naïve Bayes.

1.4 Tujuan Penelitian

Berdasarkan latar belakang yang telah diuraikan di atas, tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun sistem pengklasifikasian kualitas udara yang efisien dan efektif di Kawasan Industri Medan (KIM) untuk memberikan data yang akurat berdasarkan data yang ada.
2. Mengetahui tingkat efektifitas metode Naïve Bayes dalam pengklasifikasian kualitas udara di Kawasan Industri Medan (KIM).

1.5 Manfaat Penelitian

1. Membantu untuk mengetahui baik buruknya kualitas udara yang berada di Kawasan Industri Medan (KIM)
2. Dengan adanya sistem klasifikasi kualitas udara ini, diharapkan dapat membantu masyarakat dalam memantau kualitas udara dan potensi dampaknya terhadap kesehatan masyarakat, termasuk resiko terhadap penyakit penapasan dan kondisi kesehatan lainnya.
3. Meningkatkan kesadaran akan pentingnya udara bersih bagi kualitas hidup.

BAB II

LANDASAN TEORI

2.1 Data Mining

Pengumpulan informasi penting dari sekumpulan data dikenal sebagai data mining. Berbagai disiplin seperti statistik, basis data, pembelajaran mesin, pengenalan pola, kecerdasan buatan, dan visualisasi berkontribusi dalam proses data mining. Data yang akan diproses oleh data mining harus mengikuti langkah-langkah dalam Knowledge Discovery in Databases (KDD) dan terbagi menjadi dua jenis, yaitu data pelatihan (training) dan data pengujian (testing). Prinsip Pareto menyatakan bahwa 80% hasil kinerja seseorang berasal dari 20% usaha yang dilakukan. Pembagian data pelatihan terdiri dari 80% dari keseluruhan dataset, sedangkan data pengujian mencakup 20% dari dataset tersebut.

2.2 Naïve Bayes

Algoritma Naive Bayes adalah metode klasifikasi yang menggunakan probabilitas dan statistik untuk menganalisis perbandingan antara data pelatihan dan data pengujian. Proses ini terdiri dari beberapa langkah untuk mendapatkan probabilitas tertinggi, yang kemudian diinterpretasikan sebagai informasi yang dihasilkan. Algoritma Naive Bayes dapat digunakan untuk mengklasifikasikan data baik yang kontinu, seperti data numerik, maupun yang kategorikal. Di samping itu, algoritma ini juga dapat memodelkan atau menghitung atribut data kontinu secara langsung. Dalam hal pengklasifikasian kualitas udara di Kawasan Industri Medan (KIM). Naive Bayes diterapkan untuk menguji data Indeks Standar Pencemaran

Udara (ISPU) dengan maksud menilai kinerja klasifikasi. (Widiawati F., et al., 2023)

2.3 Klasifikasi

Klasifikasi merupakan suatu proses yang digunakan untuk mencari peran yang menerangkan suatu kelompok data, dengan maksud untuk dapat memprediksi kategori dari objek tertentu atau data yang label atau nilainya belum teridentifikasi. Untuk meraih tujuan proses klasifikasi tersebut menghasilkan sebuah model atau algoritma ditujukan mampu membedakan data kedalam kelas- kelas yang berbeda berdasarkan fungsi tertentu. Berbagai model dapat digunakan, seperti aturan "jika-maka," pohon keputusan, atau rumus tertentu. matematis (Aziiz, Abdul, et al.2022).

2.4 Confusion matrix

Confusion matrix merupakan tabel yang digunakan untuk menghitung kinerja suatu model data atau algoritma. Setiap baris dalam matriks tersebut menggambarkan kelas sebenarnya dari data, sementara setiap kolom menggambarkan kelas yang diprediksi dari data (atau sebaliknya). (Maulana, Andika S, et al.2023)

2.5 Udara

Kehadiran udara memiliki peran yang sangat berpengaruh dalam menjaga kelangsungan hidup semua makhluk hidup, terutama dalam penyediaan oksigen. Oksigen dihasilkan melalui proses pengubahan energi cahaya menjadi makanan oleh tumbuhan dan alga, yang mengambil atau mengikat karbon dioksida (CO₂). Oksigen memiliki peran penting bagi makhluk hidup karena dibutuhkan dalam proses respirasi. Setiap hari, kita perlu menghirup udara sebagai suatu kebutuhan.

Secara umum, udara dapat dikelompokkan menjadi dua jenis, yaitu udara bersih dan udara kotor. (Maulana, Andika S, et al.2023)

2.6 Polusi Udara

Polusi udara, yang juga dikenal sebagai pencemaran udara, merupakan salah satu bentuk pencemaran lingkungan yang saat ini semakin meningkat. Pencemaran terjadi ketika polutan, baik dalam bentuk zat maupun gas, masuk ke dalam suatu lingkungan dan menyebabkan penurunan kualitas lingkungan tersebut. Polutan itu sendiri adalah zat atau bahan yang terdapat dalam konsentrasi melebihi ambang batas pada waktu dan lokasi yang tidak sesuai, sehingga berfungsi sebagai bahan pencemar, seperti bahan kimia, debu, panas, dan suara. Pencemaran udara mengacu pada fenomena masuknya atau tercampurnya unsur-unsur berbahaya ke dalam atmosfer, yang dapat menurunkan kualitas udara. Pencemaran ini dapat terjadi di mana saja, baik di dalam ruangan seperti rumah dan kantor maupun di luar ruangan seperti di daerah perkotaan dan jalan raya. Umumnya, polutan yang mencemari udara terdiri dari gas dan asap, yang berasal dari proses pembakaran bahan bakar yang tidak sempurna, asap dari cerobong pabrik, pembangkit listrik, hingga asap kendaraan. Polutan gas dan asap ini merupakan hasil dari oksidasi berbagai komponen bahan bakar, termasuk CO₂ (karbon dioksida), CO (karbon monoksida), SO_x (oksida sulfur), dan NO_x (oksida nitrogen). (Rames & J. P. 2020)

2.7 Python

Python adalah bahasa pemrograman tingkat tinggi yang dapat menjalankan berbagai instruksi secara langsung (dikenal sebagai interpretatif) menggunakan pendekatan Object Oriented Programming dan semantik dinamis, yang meningkatkan keterbacaan sintaksnya. Sebagai bahasa pemrograman tingkat tinggi,

Python dirancang agar mudah dipelajari karena sudah dilengkapi dengan sistem manajemen memori. otomatis (Simanjuntak., et al., 2022).

2.8 Flowchart

Diagram alir adalah serangkaian langkah prosedur dari suatu program yang diilustrasikan dalam bentuk grafis. Umumnya, diagram alir digunakan untuk menjelaskan proses pemrosesan. Diagram alir mencakup diagram yang menggambarkan langkah-langkah kerja yang menunjukkan alur suatu proses, menggunakan simbol-simbol yang disusun secara teratur dari keseluruhan sistem. (Siahaan, J.K.,2023).

Beberapa simbol yang terdapat dalam flowchart ialah sebagai berikut:

Tabel 2.8 Simbol Flowchart

No.	Simbol	Keterangan
1.		Flow: Simbol yang digunakan untuk menggunakan antara simbol yang satu dengan simbol yang lain. Simbol ini disebut juga dengan connecting line.
2.		On-Page Reference: Simbol untuk menandakan keluar-masuk atau penyambungan proses dalam lembar kerja yang sama.
3.		Off-Page Reference: Simbol keluar-masuk atau penyambungan proses dalam lembar kerja yang berbeda.
4.		Terminator: Simbol yang menyatakan awal atau akhir suatu program

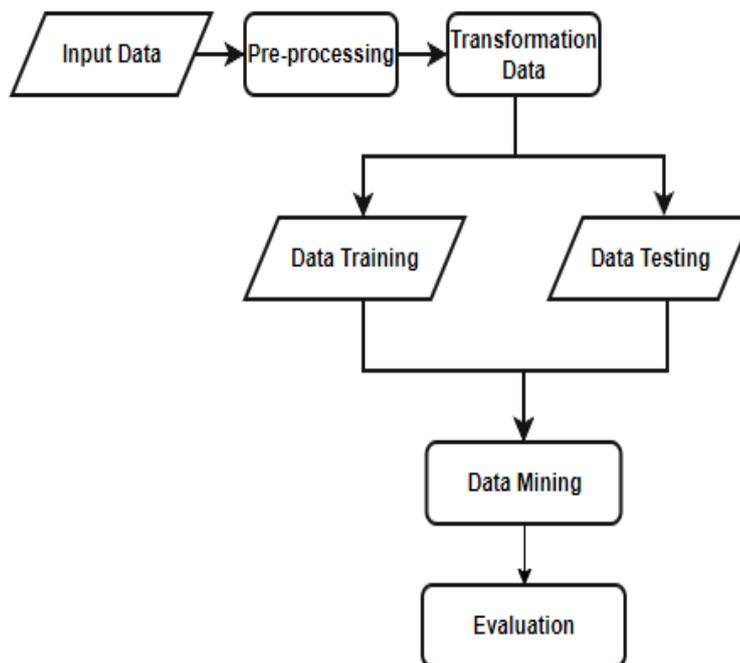
No.	Simbol	Keterangan
5.		Process: Simbol yang menyatakan suatu proses yang dilakukan komputer.
6.		Decision: Simbol yang menunjukkan kondisi tertentu yang akan menghasilkan dua kemungkinan jawaban yaitu iya atau tidak.
7.		Input/Output: Simbol yang menyatakan proses input atau output tanpa tergantung peralatan.
8.		Manual Operation: Simbol yang menyatakan suatu proses yang tidak dilakukan oleh komputer.
9.		Document: Simbol yang menyatakan bahwa input berasal dari dokumen dalam bentuk fisik, atau output yang perlu dicetak.
10.		Predefine Proses: Simbol untuk pelaksanaan suatu bagian (sub-program) atau prosedur.
11.		Display: Simbol yang menyatakan peralatan output yang digunakan.

BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Pendekatan penelitian ini diterapkan untuk mengukur tingkat kualitas udara yang ada di KIM melalui penerapan metoden Klasifikasi yang dilakukan dengan memanfaatkan algoritma *naïve bayes*. Berikut adalah langkah-langkah penelitian yang ditunjukkan pada Gambar 1:



Gambar 3.1 Tahap Penelitian

3.2 Pengumpulan Data

Data Indeks Standar Pencemar Udara (ISPU) dalam penelitian ini dikumpulkan dari Dinas Lingkungan Hidup Medan. Secara umum, dataset ini berisi 100 data yang dicatat terdiri dari 10 atribut dan 1 kelas. Data tersebut diambil dari bulan Maret hingga Juni. Atribut yang terdapat dalam dataset ini meliputi tanggal,

stasiun, PM10, PM2.5, SO2, CO, O3, NO2, maksimum, kritis, dan kategori, seperti yang ditunjukkan pada tabel 3.2 berikut:

Tabel 3.2 Data ISPU Bulan Maret 2023

Tanggal	PM ₁₀	PM _{2.5}	SO ₂	CO	O ₃	NO ₂	Max	Critical	Categori	Stasiun
1-03-2023	23	21	113	1947	73	140	1947	CO	Berbahaya	KIM
2-03-2023	34	32	106	1231	49	82	1231	CO	Berbahaya	KIM
3-03-2023	19	17	79	1162	44	141	1162	CO	Berbahaya	KIM
4-03-2023	16	15	74	2335	44	141	2335	CO	Berbahaya	KIM
5-03-2023	25	24	136	3362	101	79	3362	CO	Berbahaya	KIM
6-03-2023	31	28	88	5275	116	133	5275	CO	Berbahaya	KIM
7-03-2023	53	49	41		1	36	53	PM ₁₀	Sedang	KIM
8-03-2023	53	49	44			39	53	PM ₁₀	Sedang	KIM
9-03-2023	67	64	54	5203	227	121	5203	CO	Berbahaya	KIM
10-03-2023	40	38	148	3603	139	115	3603	CO	Berbahaya	KIM
11-03-2023	21	18	19			20	21	PM ₁₀	Baik	KIM
12-03-2023	29	26	131	3381	169	100	3381	CO	Berbahaya	KIM
13-03-2023	31	28	170	5585	225	144	5585	CO	Berbahaya	KIM
14-03-2023	40	37	141	3723	133	132	3723	CO	Berbahaya	KIM
15-03-2023	36	33	175	6180		136	6180	CO	Berbahaya	KIM
16-03-2023	36	33	175	6180		136	6180	CO	Berbahaya	KIM
17-03-2023	41	38		7156		164	7156	CO	Berbahaya	KIM
18-03-2023	37	34		7613		161	7613	CO	Berbahaya	KIM
19-03-2023	41	38	177	6453		140	6453	CO	Berbahaya	KIM
20-03-2023	45	42	170	6090	217	135	6090	CO	Berbahaya	KIM
21-03-2023	59	56	124	2039	91	95	2039	CO	Berbahaya	KIM
22-03-2023	34	33	107	260	30	83	260	CO	Berbahaya	KIM
23-03-2023	17	15	112	689	45	164	689	CO	Berbahaya	KIM
24-03-2023	19	18	176	6711		158	6711	CO	Berbahaya	KIM
25-03-2023	23	22				192	192	NO ₂	Tidak Sehat	KIM
26-03-2023	22	21	88	3511	183	71	3511	CO	Berbahaya	KIM
27-03-2023	26	24	78	220	24	62	220	CO	Berbahaya	KIM
28-03-2023	46	40	114	604	44	88	604	CO	Berbahaya	KIM
29-03-2023	38	43	68	3069	111	55	3069	CO	Berbahaya	KIM
30-03-2023	28	35	68	3706	139	54	3706	CO	Berbahaya	KIM
31-03-2023	34	49	113	5875	228	91	5875	CO	Berbahaya	KIM

Tabel 3.3 Data ISPU Bulan April 2023

Tanggal	PM ₁₀	PM _{2.5}	SO ₂	CO	O ₃	NO ₂	Max	Critical	Categori	Stasiun
01-04-2023	52	53					53	PM ₁₀	Sedang	KIM
02-04-2023	43	48	120	5571	187	91	5571	CO	Berbahaya	KIM
03-04-2023	33	34	94	5954	222	73	5954	CO	Berbahaya	KIM
04-04-2023	45	43	123			96	123	SO ₂	Sedang	KIM
05-04-2023	26	25	140			114	140	SO ₂	Sedang	KIM
06-04-2023	39	36	31	1363	47	19	1363	CO	Berbahaya	KIM

07-04-2023	35	32	33	701	19	59	701	CO	Berbahaya	KIM
08-04-2023	26	24				2	26	PM10	Baik	KIM
09-04-2023	48	43	98			63	98	SO2	Sedang	KIM
10-04-2023	38	38	63	3724	112	58	3724	CO	Berbahaya	KIM
11-04-2023	58	57	74	3894	136	60	3894	CO	Berbahaya	KIM
12-04-2023	47	41	83	2347	83	55	2347	CO	Berbahaya	KIM
13-04-2023	48	39					48	PM10	Baik	KIM
14-04-2023	40	32	113	4848	189	83	4848	CO	Berbahaya	KIM
15-04-2023	46	26					46	PM10	Baik	KIM
16-04-2023	47	34	143	5757	235	112	5757	CO	Berbahaya	KIM
17-04-2023	47	32					47	PM10	Baik	KIM
18-04-2023	42	22	60	294	8	102	294	CO	Sangat Tidak Sehat	KIM
19-04-2023	40	28	40	2850	30	26	2850	CO	Berbahaya	KIM
20-04-2023	34	16	11	1118	28	27	1118	CO	Berbahaya	KIM
21-04-2023	35	16	47	1217	40	129	1217	CO	Berbahaya	KIM
22-04-2023	42	14	54	2830	85	34	2830	CO	Berbahaya	KIM
23-04-2023	22	7	60	4151	87	48	4151	CO	Berbahaya	KIM
24-04-2023	20	5	5	140	10	3	140	CO	Tidak Sehat	KIM
25-04-2023	28	6					28	PM10	Baik	KIM
26-04-2023	44	6					44	PM10	Baik	KIM
27-04-2023	38	5					38	PM10	Baik	KIM
28-04-2023	28	4					28	PM10	Baik	KIM
29-04-2023	30	3	74	4645	170	57	4645	CO	Berbahaya	KIM
30-04-2023	34	3					34	PM10	Baik	KIM

Tabel 3.4 Data ISPU Bulan Mei 2023

Tanggal	PM ₁₀	PM _{2.5}	SO ₂	CO	O ₃	NO ₂	Max	Critical	Categori	Stasiun
01-05-2023	34	3					34	PM10	Baik	KIM
02-05-2023	31	3					31	PM10	Baik	KIM
03-05-2023	26	2					26	PM10	Baik	KIM
04-05-2023	36	3					36	PM10	Baik	KIM
05-05-2023	16	2					16	PM10	Baik	KIM
06-05-2023	13	1					13	PM10	Baik	KIM
07-05-2023	14	1					14	PM10	Baik	KIM
08-05-2023	14	1					14	PM10	Baik	KIM
09-05-2023	24	2					24	PM10	Baik	KIM
10-05-2023	20	1					20	PM10	Baik	KIM
11-05-2023	24	2					24	PM10	Baik	KIM
12-05-2023	38	3	40	2807	104	38	2807	CO	Berbahaya	KIM
13-05-2023	71	5	39	3453	67	34	3453	CO	Berbahaya	KIM
14-05-2023	20	1					20	PM10	Baik	KIM
15-05-2023	18	2	6	325	8	5	325	CO	Berbahaya	KIM
16-05-2023	22	1					22	PM10	Baik	KIM
17-05-2023	19	1	5	284	9	3	284	CO	Sangat Tidak Sehat	KIM
18-05-2023	14	1					14	PM10	Baik	KIM
19-05-2023	15	1	15	6585		118	6585	CO	Berbahaya	KIM

20-05-2023	22	2	67	2871	78	51	2871	CO	Berbahaya	KIM
21-05-2023	12	1					12	PM10	Baik	KIM
22-05-2023	20	1				166	166	NO2	Tidak Sehat	KIM
23-05-2023	17	1	16	1500	46	8	1500	CO	Berbahaya	KIM
24-05-2023	24	2	95	5790	206	112	5790	CO	Berbahaya	KIM
25-05-2023	30	2	74	6479	203	80	6479	CO	Berbahaya	KIM
26-05-2023	13	1	60	5187	144	38	5187	CO	Berbahaya	KIM
27-05-2023	14	1	81	4866	123	102	4866	CO	Berbahaya	KIM
28-05-2023	10	1	65	2234	77	46	2234	CO	Berbahaya	KIM
29-05-2023	18	1	14 3	1074	219	98	1074	CO	Berbahaya	KIM
30-05-2023	19	1	16	185	41	7	185	CO	Tidak Sehat	KIM
31-05-2023	20	1	4	200	5	2	200	CO	Tidak Sehat	KIM

Tabel 3.5 Data ISPU Bulan Juni 2023

Tanggal	PM ₁₀	P M 2.5	SO ₂	CO	O ₃	NO ₂	Max	Critical	Categori	Stasiun
01-06-2023	23	1	176			177	177	SO2	Tidak Sehat	KIM
02-06-2023	35	2	21	42	2	34	42	CO	Baik	KIM
03-06-2023	38	2					38	PM10	Baik	KIM
04-06-2023	14	1					14	PM10	Baik	KIM
05-06-2023	18	1					18	PM10	Baik	KIM
06-06-2023	16	1		84			84	CO	Sedang	KIM
07-06-2023	26	1		56			56	CO	Sedang	KIM
08-06-2023	33	2					33	PM10	Baik	KIM
09-06-2023	35	2	57	5078	178	36	5078	CO	Berbahaya	KIM
10-06-2023	23	1		1			23	PM10	Baik	KIM
11-06-2023	22	1	1	34	1	1	34	CO	Baik	KIM
12-06-2023	16	1					16	PM10	Baik	KIM
13-06-2023	29	1					29	PM10	Baik	KIM
14-06-2023	31	1					31	PM10	Baik	KIM
15-06-2023	36	2					36	PM10	Baik	KIM
16-06-2023	41	2		187			187	CO	Tidak Sehat	KIM
17-06-2023	36	2		100			100	CO	Sedang	KIM
18-06-2023	29	1					29	PM10	Baik	KIM
19-06-2023	24	1	49	2308	78	64	2308	CO	Berbahaya	KIM
20-06-2023	24	1	85	98		166	166	NO2	Tidak Sehat	KIM
21-06-2023	15	1	90	56		77	77	NO2	Sedang	KIM
22-06-2023	15	1	146	2971	107	156	2971	CO	Berbahaya	KIM
23-06-2023	29	2	67	4768		27	4768	CO	Berbahaya	KIM
24-06-2023	16	3	243	384	63	129	384	CO	Berbahaya	KIM
25-06-2023	53	4	75	2453	58	37	2453	CO	Berbahaya	KIM
26-06-2023	64	2	83	4814	365		4814	CO	Berbahaya	KIM
27-06-2023	45	1	67	6874	45		6874	CO	Berbahaya	KIM
28-06-2023	78	1	28	2871	736	21	2871	CO	Berbahaya	KIM
29-06-2023	24	4	60	1987	233		1987	CO	Berbahaya	KIM
30-06-2023	54	2	164	89		34	164	SO2	Tidak Sehat	KIM
31-06-2023	34	1	93	854	873	56	873	O3	Berbahaya	KIM

Dataset ini terdiri dari 10 atribut, termasuk atribut kelas, yaitu:

- a. Tanggal: Waktu saat pengukuran kualitas udara dilakukan.
- b. Lokasi: Tempat pengukuran di lokasi tertentu.
- c. PM10: Partikulat halus yang menjadi salah satu parameter yang diukur.
- d. PM2.5: Partikulat halus lainnya yang juga diukur.
- e. SO₂: Sulfur dioksida (dalam bentuk SO₂) yang merupakan salah satu parameter yang diukur.
- f. CO: Karbon monoksida yang termasuk dalam parameter yang diukur.
- g. O₃: Ozon yang merupakan salah satu parameter yang diukur.
- h. NO₂: Nitrogen dioksida yang diukur.
- i. Max: Nilai pengukuran tertinggi dari semua parameter yang diukur pada waktu yang bersamaan.
- j. Critical: Parameter dengan hasil pengukuran indeks tertinggi.
- k. Kategori: Kategori yang ditentukan berdasarkan hasil perhitungan indeks standar pencemaran udara. dan mencakup sebagai nilai dari kelas

Input Data Pada tahap ini, akan diambil data yang relevan untuk analisis dengan cara mengurangi jumlah atribut dan catatan data yang ada, sehingga informasi tetap terjaga. Variabel input yang digunakan mencakup tanggal dan lokasi. PM_{10} , PM_{25} , SO_2 , CO, O_3 , NO_2 , maksimal, kritis. Atribut yang dipilih hanya meliputi tanggal dan lokasi. PM_{10} , PM_{25} , SO_2 , CO, O_3 , NO_2 , hanya yang mendekati dampak nilai dari variabel keluaran yang akan diolah dalam tahap pra-pemrosesan, dan variabel keluaran tersebut adalah kategori.

Pre-Processing pada proses ini digunakan untuk memproses data mentah sebelum data tersebut digunakan dalam proses analisis atau pemodelan. Pra-processing data

memainkan peran penting dalam mempersiapkan data yang berkualitas untuk analisis dan pemodelan yang lebih akurat dan andal.

Data Training Sebelum mengklasifikasikan data yang telah diseleksi, data akan dibagi menjadi dua bagian: data pelatihan (training) dan data pengujian (testing). Pada langkah ini, data pelatihan dibuat untuk melatih algoritma Naïve Bayes. Proporsi data pelatihan yang digunakan adalah 80% dari keseluruhan dataset.

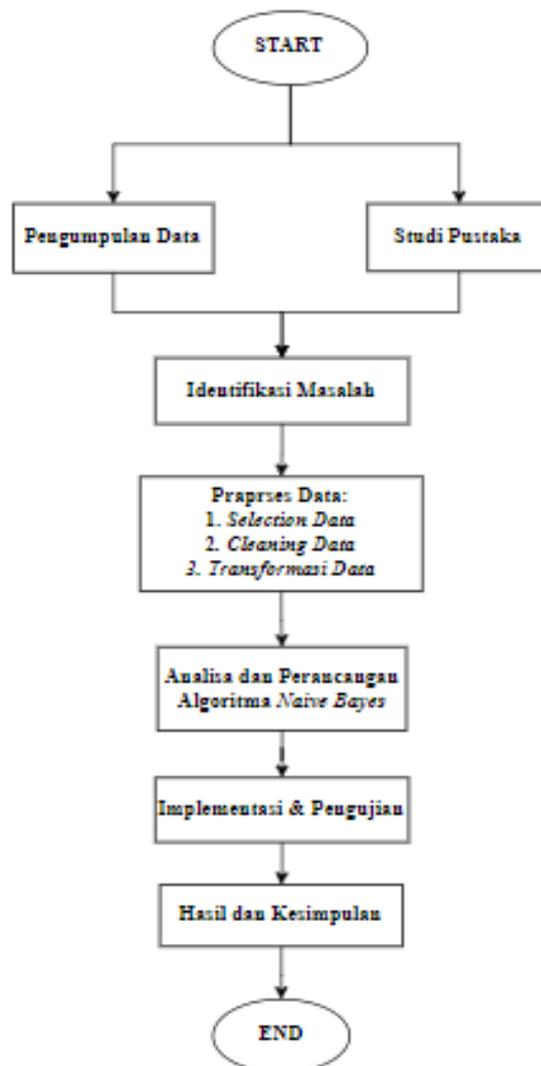
Data Test Pada tahap ini, dilakukan pembuatan data pengujian untuk mengevaluasi kinerja metode Naïve Bayes. Data pengujian dibuat dengan perbandingan sebesar 20% dari total dataset.

Data Mining Data yang telah disiapkan akan diuji menggunakan metode yang telah ditentukan. Metode yang dipilih adalah klasifikasi dengan algoritma Naïve Bayes. Proses pengujian metode ini akan dilakukan dengan menggunakan bahasa pemrograman Python dan library yang tersedia. Mengingat data yang digunakan bersifat kontinu, maka Gaussian Naïve Bayes akan diterapkan dalam permodelan ini.

Evaluasi Pada tahap evaluasi ini melibatkan pembuatan output yang mudah dipahami, yang dihasilkan dari data mining berupa informasi dari pengujian data pelatihan dan data pengujian. Hasil evaluasi model disajikan dalam bentuk tabel matriks. Untuk menghitung nilai hasil evaluasi, digunakan metode pengujian pada hasil model data mining melalui perhitungan confusion matrix. Pada tahap ini, data pelatihan dan data pengujian yang telah diproses menggunakan algoritma naïve bayes akan diujikan performasinya.

3.3 Flowchart Sistem Klasifikasi

Diagram alir berfungsi untuk menggunakan, Mempermudah serangkaian proses atau prosedur agar lebih mudah dipahami dan dilihat berdasarkan urutan langkah-langkah yang ada. Di bawah ini adalah flowchart yang menggambarkan keseluruhan proses dan digunakan sebagai referensi dalam penelitian ini



Gambar 3.2 Flowchart Sistem Klasifikasi

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pembahasan Dan Persiapan Data

Data yang digunakan dalam penelitian ini diperoleh dari Dinas Lingkungan Hidup Medan dengan total 100 data. Di bawah ini adalah dataset yang akan menjadi fokus penelitian ini. Dari total data, akan dialokasikan 80% dari data ini untuk melatih model (data pelatihan) 20% sisanya akan diperuntukkan sebagai data yang digunakan sebagai data yang digunakan untuk menguji model (data pengujian).

	PM10	PM2.5	S02	CO	O3	NO2	Max	Critical	Categori
0	23	21	113	1947	73.0	140	1947	CO	BERBAHAYA
1	34	32	106	1231	49.0	82	1231	CO	BERBAHAYA
2	19	17	79	1162	44.0	141	1162	CO	BERBAHAYA
3	16	15	74	2335	44.0	141	2335	CO	BERBAHAYA
4	25	24	136	3362	101.0	79	3362	CO	BERBAHAYA
...
95	24	1	85	98	0.0	166	166	NO2	TIDAK SEHAT
96	15	1	90	56	0.0	77	77	NO2	SEDANG
97	29	2	67	4768	0.0	27	4768	CO	BERBAHAYA
98	16	3	243	384	63.0	129	384	CO	BERBAHAYA
99	64	2	83	4814	365.0	0	4814	CO	BERBAHAYA

100 rows x 9 columns

Gambar 4.1 Data Selection

Setelah tahap pemrosesan, langkah selanjutnya adalah mengubah data bertipe object menjadi data dalam bentuk angka. Gambar 4.2 menunjukkan proses mengubah nilai 'BAIK' menjadi 0, 'BERBAHAYA' 1, 'SANGAT TIDAK SEHAT' 2, 'SEDANG' 3, 'TIDAK SEHAT' 4, lalu digabungkan ke dalam dataframe dan tidak lupa diubah menjadi tipe data integer.

```

[24] X = datalatih.drop(columns="Categori")
     y = datalatih.Categori

en=LabelEncoder()
datalatih['Categori']=en.fit_transform(datalatih['Categori'])
datalatih

```

	PM10	PM2.5	SO2	CO	O3	NO2	Max	Critical	Categori
0	23	21	113	1947	73.0	140	1947	CO	1
1	34	32	106	1231	49.0	82	1231	CO	1
2	19	17	79	1162	44.0	141	1162	CO	1
3	16	15	74	2335	44.0	141	2335	CO	1
4	25	24	136	3362	101.0	79	3362	CO	1
...
95	24	1	85	98	0.0	166	166	NO2	4
96	15	1	90	56	0.0	77	77	NO2	3
97	29	2	67	4768	0.0	27	4768	CO	1
98	16	3	243	384	63.0	129	384	CO	1
99	64	2	83	4814	365.0	0	4814	CO	1

100 rows x 9 columns

Gambar 4.2 Transformasi data kategori

Selanjutnya melakukan Langkah yang penting adalah mendefinisikan variabel independen (x) dan variabel dependen (y). Gambar 4.3 menggambarkan proses deklarasi variabel x dan y.

```

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
X_train.shape, X_test.shape, y_train.shape, y_test.shape

```

((80, 8), (20, 8), (80,), (20,))

Gambar 4.3 Variabel X dan Y

4.2 Pembagian Data Training dan Testing

Ini adalah langkah dalam machine learning yang bertujuan untuk membagi data menjadi dua bagian, yaitu data pelatihan dan data pengujian. Dalam hal ini, 75% dari data digunakan untuk pelatihan, sedangkan sisanya dialokasikan untuk pengujian.

```

[30] pipeline.score(X_train, y_train)
     0.9625

[31] pipeline.score(X_test, y_test)
     0.8

```

Gambar 4.4 Data Training dan Testing

4.3 Implementasi Metode Naïve Bayes

Penerapan metode Naïve Bayes untuk mengklasifikasikan kualitas udara di kawasan industri medan yang diawali dengan mendefinisikan jumlah data masing-masing kategori, perhitungan metode naïve bayes, menghitung probabilitas prior, menghitung likelihood, menghitung probabilitas posterior. Jumlah data berdasarkan kategori adalah 100 dengan rincian:

- a. BAIK = 22,
- b. BERBAHAYA = 58,
- c. SEDANG = 10,
- d. TIDAK SEHAT = 8,
- e. SANGAT TIDAK SEHAT = 2.

Data sample yang akan diklasifikasi adalah $PM_{10} = 20$, $PM_{2.5} = 18$, $SO_2 = 100$, $CO = 1500$, $O_3 = 50$, $NO_2 = 100$.

4.3.1 Kategori Kualitas Udara Baik

- a. Mendefinisikan Rata-Rata dan Standar Deviasi

Jenis Polutan	Rata-Rata(μ)	Standar Deviasi(σ)
PM_{10}	27,36	10,85
$PM_{2.5}$	7,82	11,71
SO_2	1,86	5,88
CO	3,50	11,24
O_3	0,14	0,47
NO_2	2,59	8,20

Tabel 4.1 Rata-Rata dan standar Deviasi setiap polutan untuk Kategori Baik

b. Menghitung probabilitas prior

$$P(\text{BAIK}) = \frac{22}{100} = 0,22$$

c. Menghitung Likelihood

Perhitungan Likelihood untuk polutan PM₁₀:

$$\mu_{\text{PM}_{10} | \text{BAIK}} = 27.36, \quad \sigma_{\text{PM}_{10} | \text{BAIK}} = 10.85$$

$$P(\text{PM}_{10} = 20 | \text{BAIK}) = \frac{1}{\sqrt{2\pi(10.85)^2}} \times e^{-\frac{(20-27.36)^2}{2(10.85)^2}}$$

$$(20 - 27.36)^2 = (-7.36)^2 = 54.18$$

$$\text{Eksponen} = -\frac{54.18}{2 \times (10.85)^2} = -\frac{54.18}{235.62} \approx -0.23$$

$$e^{-0.23} \approx 0.794$$

$$\frac{1}{\sqrt{2\pi(10.85)^2}} = \frac{1}{\sqrt{2\pi \times 117.77}} \approx \frac{1}{27.24} \approx 0.0367$$

$$P(\text{PM}_{10} = 20 | \text{BAIK}) \approx 0.0367 \times 0.794 \approx 0.0291$$

Perhitungan Likelihood untuk polutan PM_{2.5}:

$$\mu_{\text{PM}_{2.5} | \text{BAIK}} = 7.82, \quad \sigma_{\text{PM}_{2.5} | \text{BAIK}} = 11.71$$

$$P(\text{PM}_{2.5} = 18 | \text{BAIK}) = \frac{1}{\sqrt{2\pi(11.71)^2}} \times e^{-\frac{(18-7.82)^2}{2(11.71)^2}}$$

$$(18 - 7.82)^2 = 10.18^2 = 103.63$$

$$\text{Eksponen} = -\frac{103.63}{2 \times (11.71)^2} = -\frac{103.63}{274.29} \approx -0.378$$

$$e^{-0.378} \approx 0.685$$

$$\frac{1}{\sqrt{2\pi(11.71)^2}} = \frac{1}{\sqrt{2\pi \times 137.13}} \approx \frac{1}{29.62} \approx 0.0338$$

$$P(\text{PM}_{2.5} = 18 | \text{BAIK}) \approx 0.0338 \times 0.685 \approx 0.0231$$

Perhitungan Likelihood untuk polutan SO₂:

$$\mu_{\text{SO}_2 | \text{BAIK}} = 1.86, \quad \sigma_{\text{SO}_2 | \text{BAIK}} = 5.88$$

$$P(\text{SO}_2 = 100 | \text{BAIK}) = \frac{1}{\sqrt{2\pi(5.88)^2}} \times e^{-\frac{(100-1.86)^2}{2(5.88)^2}}$$

$$(100 - 1.86)^2 = 98.14^2 = 9627.32$$

$$\text{Eksponen} = -\frac{9627.32}{2 \times (5.88)^2} = -\frac{9627.32}{69.15} \approx -139.18$$

$$e^{-139.18} \approx 2.1 \times 10^{-61}$$

$$\frac{1}{\sqrt{2\pi(5.88)^2}} = \frac{1}{\sqrt{2\pi \times 34.57}} \approx \frac{1}{14.73} \approx 0.0679$$

$$P(\text{SO}_2 = 100 | \text{BAIK}) \approx 0.0679 \times 2.1 \times 10^{-61} \approx 1.42 \times 10^{-62}$$

Perhitungan Likelihood untuk polutan CO:

$$\mu_{\text{CO} | \text{BAIK}} = 3.50, \quad \sigma_{\text{CO} | \text{BAIK}} = 11.24$$

$$P(\text{CO} = 1500 | \text{BAIK}) = \frac{1}{\sqrt{2\pi(11.24)^2}} \times e^{-\frac{(1500-3.50)^2}{2(11.24)^2}}$$

$$(1500 - 3.50)^2 = 1496.50^2 = 2239502.25$$

$$\text{Eksponen} = -\frac{2239502.25}{2 \times (11.24)^2} = -\frac{2239502.25}{252.56} \approx -8864.11$$

$$e^{-8864.11} \approx 0 \quad (\text{Teramat sangat kecil hingga mendekati } 0)$$

$$\frac{1}{\sqrt{2\pi(11.24)^2}} = \frac{1}{\sqrt{2\pi \times 126.38}} \approx \frac{1}{27.88} \approx 0.0359$$

$$P(\text{CO} = 1500 | \text{BAIK}) \approx 0 \quad (\text{Nilai yang sangat kecil mendekati } 0)$$

Perhitungan Likelihood untuk polutan O₃:

$$\mu_{\text{O}_3 | \text{BAIK}} = 0.14, \quad \sigma_{\text{O}_3 | \text{BAIK}} = 0.47$$

$$P(\text{O}_3 = 50 | \text{BAIK}) = \frac{1}{\sqrt{2\pi(0.47)^2}} \times e^{-\frac{(50-0.14)^2}{2(0.47)^2}}$$

$$(50 - 0.14)^2 = 49.86^2 = 2486.42$$

$$\text{Eksponen} = -\frac{2486.42}{2 \times (0.47)^2} = -\frac{2486.42}{0.4418} \approx -5629.17$$

$$e^{-5629.17} \approx 0 \quad (\text{Teramat sangat kecil hingga mendekati } 0)$$

$$\frac{1}{\sqrt{2\pi(0.47)^2}} = \frac{1}{\sqrt{2\pi \times 0.22}} \approx \frac{1}{0.83} \approx 1.20$$

$$P(O_3 = 50|\text{BAIK}) \approx 0 \quad (\text{Nilai yang sangat kecil mendekati } 0)$$

Perhitungan Likelihood untuk polutan NO₂:

$$\mu_{\text{NO}_2|\text{BAIK}} = 2.59, \quad \sigma_{\text{NO}_2|\text{BAIK}} = 8.20$$

$$P(\text{NO}_2 = 100|\text{BAIK}) = \frac{1}{\sqrt{2\pi(8.20)^2}} \times e^{-\frac{(100-2.59)^2}{2(8.20)^2}}$$

$$(100 - 2.59)^2 = 97.41^2 = 9489.24$$

$$\text{Eksponen} = -\frac{9489.24}{2 \times (8.20)^2} = -\frac{9489.24}{134.48} \approx -70.56$$

$$e^{-70.56} \approx 2.54 \times 10^{-31}$$

$$\frac{1}{\sqrt{2\pi(8.20)^2}} = \frac{1}{\sqrt{2\pi \times 67.24}} \approx \frac{1}{20.61} \approx 0.0485$$

$$P(\text{NO}_2 = 100|\text{BAIK}) \approx 0.0485 \times 2.54 \times 10^{-31} \approx 1.23 \times 10^{-32}$$

d. Menghitung probabilitas posterior

Rumus Probabilitas Posterior:

$$P(\text{BAIK}|X) = P(\text{BAIK}) \times P(\text{PM}_{10}=20|\text{BAIK}) \times P(\text{PM}_{2.5}=18|\text{BAIK}) \times P(\text{SO}_2=100|\text{BAIK}) \times P(\text{CO}=1500|\text{BAIK}) \times P(\text{O}_3=50|\text{BAIK}) \times P(\text{NO}_2=100|\text{BAIK})$$

Hasilnya adalah:

$$P(\text{BAIK}|X) \approx 0.22 \times 0.0291 \times 0.0231 \times 1.42 \times 10^{-62} \times 0 \times 0 \times 1.23 \times 10^{-32}$$

4.3.2 Kategori Kualitas Udara Berbahaya

a. Mendefinisikan Rata-Rata dan Standar Deviasi

Jenis Polutan	Rata-Rata(μ)	Standar Deviasi(σ)
PM ₁₀	33,47	13,39
PM _{2.5}	23,26	17,34
SO ₂	91,03	51,82
CO	3617,07	2096,57
O ₃	104,18	82,57
NO ₂	85,10	45,49

Tabel 4.2 Rata-Rata dan standar Deviasi setiap polutan untuk Kategori Berbahaya

b. Menghitung probabilitas prior

$$P(\text{BERBAHAYA}) = \frac{58}{100} = 0,58$$

c. Menghitung Likelihood

Perhitungan Likelihood untuk polutan PM₁₀:

$$\mu_{\text{PM}_{10} | \text{BERBAHAYA}} = 33.47, \quad \sigma_{\text{PM}_{10} | \text{BERBAHAYA}} = 13.39$$

$$P(\text{PM}_{10} = 20 | \text{BERBAHAYA}) = \frac{1}{\sqrt{2\pi(13.39)^2}} \times e^{-\frac{(20 - 33.47)^2}{2(13.39)^2}}$$

$$(20 - 33.47)^2 = (-13.47)^2 = 181.37$$

$$\text{Eksponen} = -\frac{181.37}{2 \times (13.39)^2} = -\frac{181.37}{358.60} \approx -0.506$$

$$e^{-0.506} \approx 0.603$$

$$\frac{1}{\sqrt{2\pi(13.39)^2}} = \frac{1}{\sqrt{841.87 \times 2\pi}} = \frac{1}{\sqrt{5290.99}} \approx \frac{1}{72.73} \approx 0.01375$$

$$P(\text{PM}_{10} = 20|\text{BERBAHAYA}) \approx 0.01375 \times 0.603 \approx 0.00829$$

Perhitungan Likelihood untuk polutan $\text{PM}_{2.5}$:

$$\mu_{\text{PM}_{2.5}|\text{BERBAHAYA}} = 23.26, \quad \sigma_{\text{PM}_{2.5}|\text{BERBAHAYA}} = 17.34$$

$$P(\text{PM}_{2.5} = 18|\text{BERBAHAYA}) = \frac{1}{\sqrt{2\pi(17.34)^2}} \times e^{-\frac{(18 - 23.26)^2}{2(17.34)^2}}$$

$$(18 - 23.26)^2 = (-5.26)^2 = 27.65$$

$$\text{Eksponen} = -\frac{27.65}{2 \times (17.34)^2} = -\frac{27.65}{601.48} \approx -0.046$$

$$e^{-0.046} \approx 0.955$$

$$\frac{1}{\sqrt{2\pi(17.34)^2}} = \frac{1}{\sqrt{1899.44 \times 2\pi}} = \frac{1}{\sqrt{11933.88}} \approx \frac{1}{109.24} \approx 0.00915$$

$$P(\text{PM}_{2.5} = 18|\text{BERBAHAYA}) \approx 0.00915 \times 0.955 \approx 0.00874$$

Perhitungan Likelihood untuk polutan SO_2 :

$$\mu_{\text{SO}_2|\text{BERBAHAYA}} = 91.03, \quad \sigma_{\text{SO}_2|\text{BERBAHAYA}} = 51.82$$

$$P(\text{SO}_2 = 100|\text{BERBAHAYA}) = \frac{1}{\sqrt{2\pi(51.82)^2}} \times e^{-\frac{(100 - 91.03)^2}{2(51.82)^2}}$$

$$(100 - 91.03)^2 = 8.97^2 = 80.45$$

$$\text{Eksponen} = -\frac{80.45}{2 \times (51.82)^2} = -\frac{80.45}{5378.22} \approx -0.015$$

$$e^{-0.015} \approx 0.985$$

$$\frac{1}{\sqrt{2\pi(51.82)^2}} = \frac{1}{\sqrt{5378.22 \times 2\pi}} = \frac{1}{\sqrt{33796.09}} \approx \frac{1}{183.88} \approx 0.00544$$

$$P(\text{SO}_2 = 100|\text{BERBAHAYA}) \approx 0.00544 \times 0.985 \approx 0.00536$$

Perhitungan Likelihood untuk polutan CO :

$$\mu_{\text{CO}|\text{BERBAHAYA}} = 3617.07, \quad \sigma_{\text{CO}|\text{BERBAHAYA}} = 2096.57$$

$$P(\text{CO} = 1500|\text{BERBAHAYA}) = \frac{1}{\sqrt{2\pi(2096.57)^2}} \times e^{-\frac{(1500 - 3617.07)^2}{2(2096.57)^2}}$$

$$(1500 - 3617.07)^2 = (-2117.07)^2 = 4481938.45$$

$$\text{Eksponen} = -\frac{4481938.45}{2 \times (2096.57)^2} = -\frac{4481938.45}{8796900.81} \approx -0.509$$

$$e^{-0.509} \approx 0.601$$

$$\frac{1}{\sqrt{2\pi(2096.57)^2}} = \frac{1}{\sqrt{8796900.81 \times 2\pi}} = \frac{1}{\sqrt{55269079.07}} \approx \frac{1}{7435.58} \approx 0.000134$$

$$P(\text{CO} = 1500|\text{BERBAHAYA}) \approx 0.0001345 \times 0.601 \approx 0.0000809$$

Perhitungan Likelihood untuk polutan O₃:

$$\mu_{\text{O}_3|\text{BERBAHAYA}} = 104.18, \quad \sigma_{\text{O}_3|\text{BERBAHAYA}} = 82.57$$

$$P(\text{O}_3 = 50|\text{BERBAHAYA}) = \frac{1}{\sqrt{2\pi(82.57)^2}} \times e^{-\frac{(50 - 104.18)^2}{2(82.57)^2}}$$

$$(50 - 104.18)^2 = (-54.18)^2 = 2935.56$$

$$\text{Eksponen} = -\frac{2935.56}{2 \times (82.57)^2} = -\frac{2935.56}{13638.54} \approx -0.215$$

$$e^{-0.215} \approx 0.807$$

$$\frac{1}{\sqrt{2\pi(82.57)^2}} = \frac{1}{\sqrt{13638.54 \times 2\pi}} = \frac{1}{\sqrt{85673.84}} \approx \frac{1}{292.73} \approx 0.00342$$

$$P(\text{O}_3 = 50|\text{BERBAHAYA}) \approx 0.00342 \times 0.807 \approx 0.00276$$

Perhitungan Likelihood untuk polutan NO₂:

$$\mu_{\text{NO}_2|\text{BERBAHAYA}} = 85.10, \quad \sigma_{\text{NO}_2|\text{BERBAHAYA}} = 45.49$$

$$P(\text{NO}_2 = 100|\text{BERBAHAYA}) = \frac{1}{\sqrt{2\pi(45.49)^2}} \times e^{-\frac{(100 - 85.10)^2}{2(45.49)^2}}$$

$$(100 - 85.10)^2 = (14.90)^2 = 222.01$$

$$\text{Eksponen} = -\frac{222.01}{2 \times (45.49)^2} = -\frac{222.01}{4139.24} \approx -0.054$$

$$e^{-0.054} \approx 0.947$$

$$\frac{1}{\sqrt{2\pi(45.49)^2}} = \frac{1}{\sqrt{4139.24 \times 2\pi}} = \frac{1}{\sqrt{26024.47}} \approx \frac{1}{161.27} \approx 0.0062$$

$$P(\text{NO}_2 = 100|\text{BERBAHAYA}) \approx 0.0062 \times 0.947 \approx 0.00587$$

- d. Menghitung probabilitas posterior

Rumus probabilitas posterior:

$$P(\text{BERBAHAYA}|X) = P(\text{BERBAHAYA}) \times P(\text{PM}_{10} = 20|\text{BERBAHAYA}) \times P(\text{PM}_{2.5} = 18|\text{BERBAHAYA}) \times P(\text{SO}_2 = 100|\text{BERBAHAYA}) \times P(\text{CO} = 1500|\text{BERBAHAYA}) \times P(\text{O}_3 = 50|\text{BERBAHAYA}) \times P(\text{NO}_2 = 100|\text{BERBAHAYA})$$

Implementasi rumus probabilitas posterior:

$$P(\text{BERBAHAYA}|X) \approx 0.58 \times 0.00829 \times 0.00874 \times 0.00536 \times 0.0000809 \times 0.00276 \times 0.00587$$

Hasilnya adalah:

$$P(\text{BERBAHAYA}|X) \approx 0.58 \times 7.08 \times 10^{-14} \approx 4.11 \times 10^{-14}$$

4.3.3 Kategori Kualitas Udara Sangat Tidak Sehat

- a. Menjelaskan rata-rata dan deviasi standar.

Tabel 4.3 Menunjukkan Rata-Rata dan deviasi standar dari masing-masing polutan untuk Kategori Sangat Tidak Sehat.

Jenis Polutan	Rata-Rata(μ)	Standar Deviasi(σ)
PM ₁₀	30,50	16,62
PM _{2.5}	11,50	14,85
SO ₂	32,50	38,89
CO	289,00	7,07
O ₃	8,50	0,71
NO ₂	52,50	70,00

b. Menghitung probabilitas prior

$$P(\text{SANGAT TIDAK SEHAT}) = \frac{2}{100} = 0,02$$

c. Menghitung Likelihood

Perhitungan Likelihood untuk polutan PM_{10} :

$$\mu_{\text{PM}_{10} | \text{SANGAT TIDAK SEHAT}} = 30.50, \quad \sigma_{\text{PM}_{10} | \text{SANGAT TIDAK SEHAT}} = 16.26$$

$$P(\text{PM}_{10} = 20 | \text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi}(16.26)^2} \times e^{-\frac{(20-30.50)^2}{2(16.26)^2}}$$

$$(20 - 30.50)^2 = (-10.50)^2 = 110.25$$

$$\text{Eksponen} = -\frac{110.25}{2 \times (16.26)^2} = -\frac{110.25}{528.40} \approx -0.209$$

$$e^{-0.209} \approx 0.811$$

$$\frac{1}{\sqrt{2\pi}(16.26)^2} = \frac{1}{\sqrt{2\pi \times 264.40}} \approx \frac{1}{40.87} \approx 0.0245$$

$$P(\text{PM}_{10} = 20 | \text{SANGAT TIDAK SEHAT}) \approx 0.0245 \times 0.811 \approx 0.0199$$

Perhitungan Likelihood untuk polutan $\text{PM}_{2.5}$:

$$\mu_{\text{PM}_{2.5} | \text{SANGAT TIDAK SEHAT}} = 11.50, \quad \sigma_{\text{PM}_{2.5} | \text{SANGAT TIDAK SEHAT}} = 14.85$$

$$P(\text{PM}_{2.5} = 18 | \text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi}(14.85)^2} \times e^{-\frac{(18-11.50)^2}{2(14.85)^2}}$$

$$(18 - 11.50)^2 = 6.50^2 = 42.25$$

$$\text{Eksponen} = -\frac{42.25}{2 \times (14.85)^2} = -\frac{42.25}{440.30} \approx -0.096$$

$$e^{-0.096} \approx 0.908$$

$$\frac{1}{\sqrt{2\pi}(14.85)^2} = \frac{1}{\sqrt{2\pi \times 220.55}} \approx \frac{1}{37.32} \approx 0.0268$$

$$P(\text{PM}_{2.5} = 18 | \text{SANGAT TIDAK SEHAT}) \approx 0.0268 \times 0.908 \approx 0.0243$$

Perhitungan Likelihood untuk polutan SO₂:

$$\mu_{\text{SO}_2 | \text{SANGAT TIDAK SEHAT}} = 32.50, \quad \sigma_{\text{SO}_2 | \text{SANGAT TIDAK SEHAT}} = 38.89$$

$$P(\text{SO}_2 = 100 | \text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(38.89)^2}} \times e^{-\frac{(100-32.50)^2}{2(38.89)^2}}$$

$$(100 - 32.50)^2 = 67.50^2 = 4556.25$$

$$\text{Eksponen} = -\frac{4556.25}{2 \times (38.89)^2} = -\frac{4556.25}{3023.92} \approx -1.507$$

$$e^{-1.507} \approx 0.221$$

$$\frac{1}{\sqrt{2\pi(38.89)^2}} = \frac{1}{\sqrt{2\pi \times 1512.78}} \approx \frac{1}{97.68} \approx 0.0102$$

$$P(\text{SO}_2 = 100 | \text{SANGAT TIDAK SEHAT}) \approx 0.0102 \times 0.221 \approx 0.00225$$

Perhitungan Likelihood untuk polutan CO:

$$\mu_{\text{CO} | \text{SANGAT TIDAK SEHAT}} = 289.00, \quad \sigma_{\text{CO} | \text{SANGAT TIDAK SEHAT}} = 7.07$$

$$P(\text{CO} = 1500 | \text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(7.07)^2}} \times e^{-\frac{(1500-289.00)^2}{2(7.07)^2}}$$

$$(1500 - 289.00)^2 = 1211.00^2 = 1465841.00$$

$$\text{Eksponen} = -\frac{1465841.00}{2 \times (7.07)^2} = -\frac{1465841.00}{99.97} \approx -14665.84$$

$$e^{-14665.84} \approx 0 \quad (\text{Teramat sangat kecil hingga mendekati } 0)$$

$$\frac{1}{\sqrt{2\pi(7.07)^2}} = \frac{1}{\sqrt{2\pi \times 49.98}} \approx \frac{1}{17.75} \approx 0.0563$$

$$P(\text{CO} = 1500 | \text{SANGAT TIDAK SEHAT}) \approx 0$$

Perhitungan Likelihood untuk polutan O₃:

$$\mu_{\text{O}_3 | \text{SANGAT TIDAK SEHAT}} = 8.50, \quad \sigma_{\text{O}_3 | \text{SANGAT TIDAK SEHAT}} = 0.71$$

$$P(\text{O}_3 = 50 | \text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(0.71)^2}} \times e^{-\frac{(50-8.50)^2}{2(0.71)^2}}$$

$$(50 - 8.50)^2 = 41.50^2 = 1722.25$$

$$\text{Eksponen} = -\frac{1722.25}{2 \times (0.71)^2} = -\frac{1722.25}{1.0082} \approx -1708.09$$

$$e^{-1708.09} \approx 0 \quad (\text{Teramat sangat kecil hingga mendekati } 0)$$

$$\frac{1}{\sqrt{2\pi(0.71)^2}} = \frac{1}{\sqrt{2\pi \times 0.5041}} \approx \frac{1}{1.78} \approx 0.5618$$

$$P(O_3 = 50 | \text{SANGAT TIDAK SEHAT}) \approx 0$$

Perhitungan Likelihood untuk polutan NO₂:

$$\mu_{\text{NO}_2 | \text{SANGAT TIDAK SEHAT}} = 52.50, \quad \sigma_{\text{NO}_2 | \text{SANGAT TIDAK SEHAT}} = 70.00$$

$$P(\text{NO}_2 = 100 | \text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(70.00)^2}} \times e^{-\frac{(100-52.50)^2}{2(70.00)^2}}$$

$$(100 - 52.50)^2 = 47.50^2 = 2256.25$$

$$\text{Eksponen} = -\frac{2256.25}{2 \times (70.00)^2} = -\frac{2256.25}{9800} \approx -0.230$$

$$e^{-0.230} \approx 0.794$$

$$\frac{1}{\sqrt{2\pi(70.00)^2}} = \frac{1}{\sqrt{2\pi \times 4900}} \approx \frac{1}{187.68} \approx 0.00533$$

$$P(\text{NO}_2 = 100 | \text{SANGAT TIDAK SEHAT}) \approx 0.00533 \times 0.794 \approx 0.00423$$

d. Menghitung probabilitas posterior

Rumus probabilitas posterior:

$$P(\text{SANGAT TIDAK SEHAT} | X) = P(\text{SANGAT TIDAK SEHAT}) \times P(\text{PM}_{10}=20 | \text{SANGAT TIDAK SEHAT}) \times P(\text{PM}_{2.5}=18 | \text{SANGAT TIDAK SEHAT}) \times P(\text{SO}_2=100 | \text{SANGAT TIDAK SEHAT}) \times P(\text{CO}=1500 | \text{SANGAT TIDAK SEHAT}) \times P(\text{O}_3=50 | \text{SANGAT TIDAK SEHAT}) \times P(\text{NO}_2=100 | \text{SANGAT TIDAK SEHAT})$$

Hasilnya adalah:

$$P(\text{SANGAT TIDAK SEHAT}|X) \approx 0.02 \times 0.0199 \times 0.0243 \times 0.00226 \times 0 \times 0 \times 0.004$$

23

4.3.4 Kategori Kualitas Udara Sedang

- a. Mendefinisikan Rata-Rata dan Standar Deviasi

Jenis Polutan	Rata-Rata(μ)	Standar Deviasi(σ)
PM ₁₀	37,00	15,24
PM _{2.5}	26,70	23,12
SO ₂	53,60	55,10
CO	29,60	40,23
O ₃	0,10	0,32
NO ₂	42,50	43,29

Tabel 4.4 Rata-Rata dan standar Deviasi setiap polutan untuk Kategori Sedang

- b. Menghitung probabilitas prior

$$P(\text{BAIK}) = \frac{10}{100} = 0,10$$

- c. Menghitung Likelihood

Perhitungan Likelihood untuk polutan PM₁₀:

$$\mu_{\text{PM}_{10} | \text{SEDANG}} = 37.00, \quad \sigma_{\text{PM}_{10} | \text{SEDANG}} = 15.24$$

$$P(\text{PM}_{10} = 20 | \text{SEDANG}) = \frac{1}{\sqrt{2\pi(15.24)^2}} \times e^{-\frac{(20-37.00)^2}{2(15.24)^2}}$$

$$(20 - 37.00)^2 = (-17.00)^2 = 289$$

$$\text{Eksponen} = -\frac{289}{2 \times (15.24)^2} = -\frac{289}{464.45} \approx -0.622$$

$$e^{-0.622} \approx 0.537$$

$$\frac{1}{\sqrt{2\pi(15.24)^2}} = \frac{1}{\sqrt{2\pi \times 232.26}} \approx \frac{1}{38.22} \approx 0.0262$$

$$P(\text{PM}_{10} = 20|\text{SEDANG}) \approx 0.0262 \times 0.537 \approx 0.0141$$

Perhitungan Likelihood untuk polutan $\text{PM}_{2.5}$:

$$\mu_{\text{PM}_{2.5}|\text{SEDANG}} = 26.70, \quad \sigma_{\text{PM}_{2.5}|\text{SEDANG}} = 23.12$$

$$P(\text{PM}_{2.5} = 18|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(23.12)^2}} \times e^{-\frac{(18-26.70)^2}{2(23.12)^2}}$$

$$(18 - 26.70)^2 = (-8.70)^2 = 75.69$$

$$\text{Eksponen} = -\frac{75.69}{2 \times (23.12)^2} = -\frac{75.69}{1067.36} \approx -0.071$$

$$e^{-0.071} \approx 0.932$$

$$\frac{1}{\sqrt{2\pi(23.12)^2}} = \frac{1}{\sqrt{2\pi \times 534.54}} \approx \frac{1}{58.14} \approx 0.0172$$

$$P(\text{PM}_{2.5} = 18|\text{SEDANG}) \approx 0.0172 \times 0.932 \approx 0.0160$$

Perhitungan Likelihood untuk polutan SO_2 :

$$\mu_{\text{SO}_2|\text{SEDANG}} = 53.60, \quad \sigma_{\text{SO}_2|\text{SEDANG}} = 55.10$$

$$P(\text{SO}_2 = 100|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(55.10)^2}} \times e^{-\frac{(100-53.60)^2}{2(55.10)^2}}$$

$$(100 - 53.60)^2 = 46.40^2 = 2153.66$$

$$\text{Eksponen} = -\frac{2153.66}{2 \times (55.10)^2} = -\frac{2153.66}{6060.61} \approx -0.355$$

$$e^{-0.355} \approx 0.701$$

$$\frac{1}{\sqrt{2\pi(55.10)^2}} = \frac{1}{\sqrt{2\pi \times 3036.01}} \approx \frac{1}{138.78} \approx 0.0072$$

$$P(\text{SO}_2 = 100|\text{SEDANG}) \approx 0.0072 \times 0.701 \approx 0.00505$$

Perhitungan Likelihood untuk polutan CO :

$$\mu_{\text{CO}|\text{SEDANG}} = 29.60, \quad \sigma_{\text{CO}|\text{SEDANG}} = 40.23$$

$$P(\text{CO} = 1500|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(40.23)^2}} \times e^{-\frac{(1500-29.60)^2}{2(40.23)^2}}$$

$$(1500 - 29.60)^2 = 1470.40^2 = 2169256.16$$

$$\text{Eksponen} = -\frac{2169256.16}{2 \times (40.23)^2} = -\frac{2169256.16}{3244.39} \approx -668.72$$

$$e^{-668.72} \approx 0 \quad (\text{Teramat sangat kecil hingga mendekati } 0)$$

$$\frac{1}{\sqrt{2\pi(40.23)^2}} = \frac{1}{\sqrt{2\pi \times 1618.47}} \approx \frac{1}{100.95} \approx 0.00991$$

$$P(\text{CO} = 1500|\text{SEDANG}) \approx 0 \quad (\text{Nilai yang sangat kecil mendekati } 0)$$

Perhitungan Likelihood untuk polutan O₃:

$$\mu_{\text{O}_3|\text{SEDANG}} = 0.10, \quad \sigma_{\text{O}_3|\text{SEDANG}} = 0.32$$

$$P(\text{O}_3 = 50|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(0.32)^2}} \times e^{-\frac{(50-0.10)^2}{2(0.32)^2}}$$

$$(50 - 0.10)^2 = 49.90^2 = 2490.01$$

$$\text{Eksponen} = -\frac{2490.01}{2 \times (0.32)^2} = -\frac{2490.01}{0.2048} \approx -12155.37$$

$$e^{-12155.37} \approx 0 \quad (\text{Teramat sangat kecil hingga mendekati } 0)$$

$$\frac{1}{\sqrt{2\pi(0.32)^2}} = \frac{1}{\sqrt{2\pi \times 0.1024}} \approx \frac{1}{0.254} \approx 3.937$$

$$P(\text{O}_3 = 50|\text{SEDANG}) \approx 0 \quad (\text{Nilai yang sangat kecil mendekati } 0)$$

Perhitungan Likelihood untuk polutan NO₂:

$$\mu_{\text{NO}_2|\text{SEDANG}} = 42.50, \quad \sigma_{\text{NO}_2|\text{SEDANG}} = 43.29$$

$$P(\text{NO}_2 = 100|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(43.29)^2}} \times e^{-\frac{(100-42.50)^2}{2(43.29)^2}}$$

$$(100 - 42.50)^2 = 57.50^2 = 3306.25$$

$$\text{Eksponen} = -\frac{3306.25}{2 \times (43.29)^2} = -\frac{3306.25}{3745.47} \approx -0.883$$

$$e^{-0.883} \approx 0.414$$

$$\frac{1}{\sqrt{2\pi(43.29)^2}} = \frac{1}{\sqrt{2\pi \times 1873.99}} \approx \frac{1}{108.67} \approx 0.0092$$

$$P(\text{NO}_2 = 100|\text{SEDANG}) \approx 0.0092 \times 0.414 \approx 0.00381$$

- d. Menghitung probabilitas posterior

Rumus probabilitas posterior:

$$P(\text{SEDANG}|X) = P(\text{SEDANG}) \times P(\text{PM}_{10}=20|\text{SEDANG}) \times P(\text{PM}_{2.5}=18|\text{SEDANG}) \times P(\text{SO}_2=100|\text{SEDANG}) \times P(\text{CO}=1500|\text{SEDANG}) \times P(\text{O}_3=50|\text{SEDANG}) \times P(\text{NO}_2=100|\text{SEDANG})$$

Hasilnya adalah:

$$P(\text{SEDANG}|X) = 0.10 \times 0.0141 \times 0.0160 \times 0.00505 \times 0 \times 0 \times 0.00381$$

4.3.5 Kategori Kualitas Udara Tidak Sehat

- a. Mendefinisikan Rata-Rata dan Standar Deviasi

Jenis Polutan	Rata-Rata(μ)	Standar Deviasi(σ)
PM ₁₀	23,75	7,21
PM _{2.5}	4,25	7,30
SO ₂	35,75	63,55
CO	101,25	89,75
O ₃	7,00	14,21
NO ₂	89,13	92,44

Tabel 4.5 Rata-Rata dan standar Deviasi setiap polutan untuk Kategori Tidak Sehat

b. Menghitung probabilitas prior

$$P(\text{TIDAK SEHAT}) = \frac{8}{100} = 0,08$$

c. Menghitung Likelihood

Perhitungan Likelihood untuk polutan PM₁₀:

$$\mu_{\text{PM}_{10} | \text{TIDAK SEHAT}} = 23.75, \quad \sigma_{\text{PM}_{10} | \text{TIDAK SEHAT}} = 7.21$$

$$P(\text{PM}_{10} = 20 | \text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(7.21)^2}} \times e^{-\frac{(20-23.75)^2}{2(7.21)^2}}$$

$$(20 - 23.75)^2 = (-3.75)^2 = 14.06$$

$$\text{Eksponen} = -\frac{14.06}{2 \times (7.21)^2} = -\frac{14.06}{104.05} \approx -0.135$$

$$e^{-0.135} \approx 0.874$$

$$\frac{1}{\sqrt{2\pi(7.21)^2}} = \frac{1}{\sqrt{2\pi \times 52.00}} \approx \frac{1}{18.09} \approx 0.0553$$

$$P(\text{PM}_{10} = 20 | \text{TIDAK SEHAT}) \approx 0.0553 \times 0.874 \approx 0.0483$$

Perhitungan Likelihood untuk polutan PM_{2.5}:

$$\mu_{\text{PM}_{2.5} | \text{TIDAK SEHAT}} = 4.25, \quad \sigma_{\text{PM}_{2.5} | \text{TIDAK SEHAT}} = 7.30$$

$$P(\text{PM}_{2.5} = 18 | \text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(7.30)^2}} \times e^{-\frac{(18-4.25)^2}{2(7.30)^2}}$$

$$(18 - 4.25)^2 = 13.75^2 = 189.06$$

$$\text{Eksponen} = -\frac{189.06}{2 \times (7.30)^2} = -\frac{189.06}{106.73} \approx -1.772$$

$$e^{-1.772} \approx 0.170$$

$$\frac{1}{\sqrt{2\pi(7.30)^2}} = \frac{1}{\sqrt{2\pi \times 53.29}} \approx \frac{1}{18.34} \approx 0.0545$$

$$P(\text{PM}_{2.5} = 18 | \text{TIDAK SEHAT}) \approx 0.0545 \times 0.170 \approx 0.00927$$

Perhitungan Likelihood untuk polutan SO₂:

$$\mu_{\text{SO}_2 | \text{TIDAK SEHAT}} = 35.75, \quad \sigma_{\text{SO}_2 | \text{TIDAK SEHAT}} = 63.55$$

$$P(\text{SO}_2 = 100 | \text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(63.55)^2}} \times e^{-\frac{(100-35.75)^2}{2(63.55)^2}}$$

$$(100 - 35.75)^2 = 64.25^2 = 4128.06$$

$$\text{Eksponen} = -\frac{4128.06}{2 \times (63.55)^2} = -\frac{4128.06}{8061.78} \approx -0.512$$

$$e^{-0.512} \approx 0.599$$

$$\frac{1}{\sqrt{2\pi(63.55)^2}} = \frac{1}{\sqrt{2\pi \times 4038.60}} \approx \frac{1}{159.33} \approx 0.00627$$

$$P(\text{SO}_2 = 100 | \text{TIDAK SEHAT}) \approx 0.00627 \times 0.599 \approx 0.00375$$

Perhitungan Likelihood untuk polutan CO:

$$\mu_{\text{CO} | \text{TIDAK SEHAT}} = 101.25, \quad \sigma_{\text{CO} | \text{TIDAK SEHAT}} = 89.75$$

$$P(\text{CO} = 1500 | \text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(89.75)^2}} \times e^{-\frac{(1500-101.25)^2}{2(89.75)^2}}$$

$$(1500 - 101.25)^2 = 1398.75^2 = 1956506.56$$

$$\text{Eksponen} = -\frac{1956506.56}{2 \times (89.75)^2} = -\frac{1956506.56}{16100.06} \approx -121.54$$

$$e^{-121.54} \approx 0 \quad (\text{Teramat sangat kecil hingga mendekati } 0)$$

$$\frac{1}{\sqrt{2\pi(89.75)^2}} = \frac{1}{\sqrt{2\pi \times 8056.56}} \approx \frac{1}{224.88} \approx 0.00445$$

$$P(\text{CO} = 1500 | \text{TIDAK SEHAT}) \approx 0$$

Perhitungan Likelihood untuk polutan O₃:

$$\mu_{O_3 | \text{TIDAK SEHAT}} = 7.00, \quad \sigma_{O_3 | \text{TIDAK SEHAT}} = 14.21$$

$$P(O_3 = 50 | \text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(14.21)^2}} \times e^{-\frac{(50-7.00)^2}{2(14.21)^2}}$$

$$(50 - 7.00)^2 = 43.00^2 = 1849.00$$

$$\text{Eksponen} = -\frac{1849.00}{2 \times (14.21)^2} = -\frac{1849.00}{403.91} \approx -4.576$$

$$e^{-4.576} \approx 0.0103$$

$$\frac{1}{\sqrt{2\pi(14.21)^2}} = \frac{1}{\sqrt{2\pi \times 201.92}} \approx \frac{1}{35.67} \approx 0.028$$

$$P(O_3 = 50 | \text{TIDAK SEHAT}) \approx 0.028 \times 0.0103 \approx 0.00029$$

Perhitungan Likelihood untuk polutan NO₂:

$$\mu_{NO_2 | \text{TIDAK SEHAT}} = 89.13, \quad \sigma_{NO_2 | \text{TIDAK SEHAT}} = 92.44$$

$$P(NO_2 = 100 | \text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(92.44)^2}} \times e^{-\frac{(100-89.13)^2}{2(92.44)^2}}$$

$$(100 - 89.13)^2 = 10.87^2 = 118.12$$

$$\text{Eksponen} = -\frac{118.12}{2 \times (92.44)^2} = -\frac{118.12}{17088.96} \approx -0.00691$$

$$e^{-0.00691} \approx 0.9931$$

$$\frac{1}{\sqrt{2\pi(92.44)^2}} = \frac{1}{\sqrt{2\pi \times 8545.44}} \approx \frac{1}{231.34} \approx 0.00432$$

$$P(NO_2 = 100 | \text{TIDAK SEHAT}) \approx 0.00432 \times 0.9931 \approx 0.00429$$

d. Menghitung probabilitas posterior

Rumus probabilitas posterior:

$$P(\text{TIDAK SEHAT} | X) = P(\text{TIDAK SEHAT}) \times P(\text{PM}_{10}=20 | \text{TIDAK SEHAT}) \times P$$

$$(\text{PM}_{2.5}=18 | \text{TIDAK SEHAT}) \times P(\text{SO}_2=100 | \text{TIDAK SEHAT}) \times P(\text{CO}=1500 | \text{TIDAK SEHAT}) \times P(\text{O}_3=50 | \text{TIDAK SEHAT}) \times P(\text{NO}_2=100 | \text{TIDAK SEHAT})$$

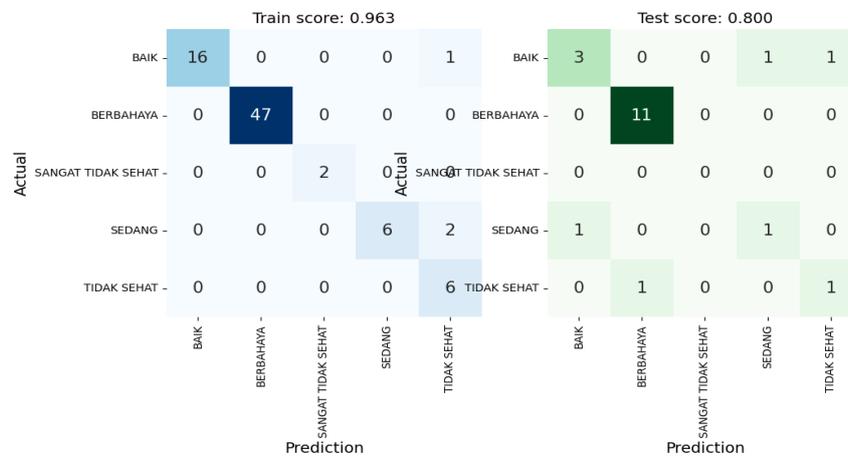
$$\times P(\text{NO}_2=100 | \text{TIDAK SEHAT})$$

Hasilnya adalah:

$$P(\text{TIDAK SEHAT}|X) \approx 0.08 \times 0.0483 \times 0.00927 \times 0.00375 \times 0 \times 0.00029 \times 0.00429$$

4.4 Confusion Matrix

Pada tahap ini, dilakukan pengujian performa algoritma yang telah diimplementasikan dengan memanfaatkan confusion matrix. Gambar 4.5 menampilkan hasil confusion matrix untuk algoritma “naïve bayes”.



Gambar 4.5 Confusion Mtrix

BAB V

KESIMPULAN & SARAN

5.1 Kesimpulan

1. Penelitian ini berhasil mengembangkan suatu sistem yang efisien dan efektif dalam mengkategorikan kualitas udara di Kawasan Industri Medan (KIM). Sistem ini menggunakan 100 data yang tersedia untuk memberikan hasil klasifikasi yang akurat, yang dapat digunakan sebagai acuan dalam memantau dan mengelola kualitas udara di wilayah tersebut. Implementasi sistem ini membuktikan bahwa teknologi dapat menjadi alat yang penting dalam mendukung upaya pelestarian lingkungan, khususnya di kawasan industri.
2. Hasil dari penelitian ini memperlihatkan bahwa metode Naïve Bayes terbukti efektif untuk mengklasifikasikan kualitas udara di Kawasan Industri Medan (KIM). Metode ini mampu mengolah data dengan baik dan memberikan prediksi dengan variabel baik, berbahaya, sangat tidak sehat, sedang, tidak sehat yang akurat terkait kategori kualitas udara. Akurasi yang dicapai oleh metode ini mengindikasikan bahwa metode Naïve Bayes merupakan sebuah perangkat yang kuat dan terpercaya untuk aplikasi sejenis di bidang lain yang memerlukan klasifikasi berbasis data.

5.2 Saran

Berdasarkan penelitian ini, beberapa saran yang dapat diberikan adalah untuk mempertimbangkan penggunaan algoritma Naïve Bayes dalam pengawasan dan pemantauan tingkat polusi udara di kota Jakarta. Algoritma ini dapat memberikan kontribusi dalam mengidentifikasi potensi pencemaran udara. Disarankan untuk melakukan studi lebih lanjut guna membandingkan performa algoritma Naïve

Bayes dengan algoritma lainnya, serta menguji model ini dengan dataset untuk memperoleh hasil yang lebih komprehensif dan valid. Implementasi sistem pengawasan kualitas udara secara langsung di area Kawasan Industri Medan dapat menjadi langkah lanjutan, yang memungkinkan pengambilan keputusan lebih cepat untuk mitigasi pencemaran.

DAFTAR PUSTAKA

- Aziiz, Abdul, et al. "Klasifikasi Tingkat Kualitas Udara Dki Jakarta Menggunakan Algoritma Naive Bayes." *EProceedings* 9.3 (2022): 1962-69.
- Qosim, A. L. (2021). *Perbandingan metode klasifikasi Support Vector Machine (SVM) dan Naive Bayes Classifier (NBC) untuk menentukan kualitas udara* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- Kurniawan, M. O. P. (2024). *PERBANDINGAN METODE RANDOM FOREST DAN K-NEAREST NEIGHBOR DALAM KLASIFIKASI KUALITAS UDARA DI YOGYAKARTA* (Doctoral dissertation, Universitas Atma Jaya Yogyakarta).
- Lahal, A. (2021). *Rancang Bangun Alat Monitoring Polusi Udara Berbasis Arduino* (Doctoral dissertation, Prodi Teknik Informatika).
- Maulana, Sandy Andika, Shabrina Husna Batubara, and Wahyu Kurnia Rahman. "Penerapan Metode Naïve Bayes dalam Peramalan Polusi Udara di Kota Jakarta." *Mutiara: Jurnal Penelitian dan Karya Ilmiah* 1.6 (2023): 347-362.
- Maulana, Ali, Ade Irma Purnamasari, and Irfan Ali. "ANALISIS KLASIFIKASI INDEKS KUALITAS UDARA KOTA DI INDONESIA MENGGUNAKAN METODE K-NEAREST NEIGHBOR DAN NAÏVE BAYES." *JATI (Jurnal Mahasiswa Teknik Informatika)* 8.1 (2024): 215-222.

Ramses, J. P. (2020). Implementasi Internet of Things (IOT) Sistem Pendeteksi Kualitas Udara Pa-da Daerah Sekitar Kawasan Industri Medan (KIM) Berbasis NodeMcu.

Sang, Adinda Inez, Edi Sutoyo, and Irfan Darmawan. "Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine." *eProceedings of Engineering* 8.5 (2021).

SARTIKA, ELISA. "KLASIFIKASI DATA KUALITAS UDARA MENGGUNAKAN FUZZY RANDOM FOREST DENGAN BOOSTRAP SAMPLING."

SIAHAAN, Johannes Kristanto, et al. Penerapan Metode Naïve Bayes untuk Menentukan Tingkat Polusi Udara di Kota Medan. 2023.

Simanjuntak, Adelin Yoseva, and Anita Anita. "IMPLEMENTASI DATA MINING MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER UNTUK DATA KENAIKAN PANGKAT DINAS KETENAGAKERJAAN KOTA MEDAN." *Journal of Science and Social Research* 5.1 (2022): 85-91.

Widiawati, F., Kurniawan, R., & Suprapti, T. (2023). KLASIFIKASI DATA TINGKAT KUALITAS UDARA DI TANGERANG SELATAN MENGGUNAKAN ALGORITMA NAIVE BAYES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3739-3745.

DAFTAR LAMPIRAN

Lampiran I: Source Code Sistem Klasifikasi Udara

```
!pip install pandas
!pip install matplotlib
!pip install numpy
!pip install scikit-learn
!pip install keras
!pip install jcopml
!pip install scipy
datalatih = pd.read_excel('Data Latih.xlsx')
datalatih
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import LabelEncoder

from jcopml.pipeline import num_pipe, cat_pipe
X = datalatih.drop(columns="Categori")
y = datalatih.Categori
en=LabelEncoder()
datalatih['Categori']=en.fit_transform(datalatih['Categori'])
datalatih
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
X_train.shape, X_test.shape, y_train.shape, y_test.shape
preprocessor = ColumnTransformer([
    ('numeric', num_pipe(),
 ['PM10', 'PM2.5', 'SO2', 'CO', 'O3', 'NO2']),
    ('categoric', cat_pipe(encoder='onehot'), ['Critical']),
])
from sklearn.naive_bayes import GaussianNB
pipeline = Pipeline([
    ('prep', preprocessor),
    ('algo', GaussianNB())
])
pipeline.fit(X_train, y_train)
pipeline.score(X_train, y_train)
pipeline.score(X_test, y_test)
from jcopml.plot import plot_confusion_matrix
plot_confusion_matrix(X_train, y_train, X_test, y_test,
pipeline)
```

```
X_pred = pd.read_excel('Data Uji.xlsx')
X_pred
pipeline.predict(X_pred)
X_pred["Categori"] = pipeline.predict(X_pred)
X_pred
X_pred.to_excel('Hasil_Klasifikasi.xlsx')
```