

**PERBANDINGAN METODE NAIVE BAYES DAN K-NEAREST
NEIGHBOR PADA KLASIFIKASI KUALITAS UDARA DI DKI JAKARTA**

PROYEK TUGAS AKHIR

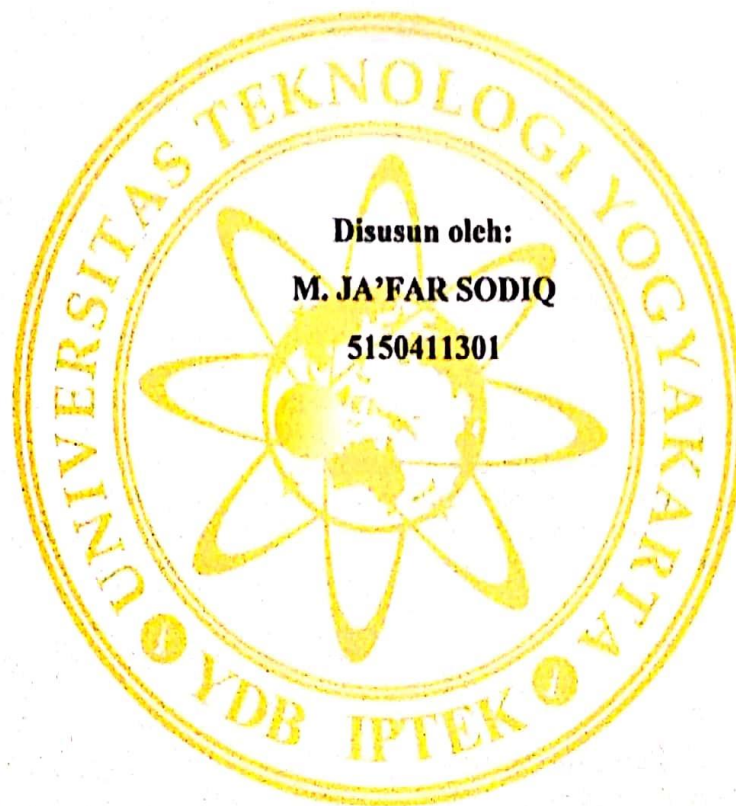


Disusun oleh
M. JA'FAR SODIQ
5150411301

Kepada
PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN ELEKTRO
UNIVERSITAS TEKNOLOGI YOGYAKARTA
2019

NASKAH PUBLIKASI

**PERBANDINGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR PADA
KLASIFIKASI KUALITAS UDARA DI DKI JAKARTA**



Dosen Pembimbing

Dr. Enny Itje Sela, S.Si., M.Kom.

Tanggal 8-12-2019

Perbandingan Metode Naive Bayes dan K-Nearest Neighbor Pada Klasifikasi Kualitas Udara di DKI Jakarta

M. Ja'far Sodik, Enny Itje Sela

*Program Studi Informatika, Fakultas Teknologi Informasi dan Elektro
Universitas Teknologi Yogyakarta
Jl. Ringroad Utara Jombor Sleman Yogyakarta
E-mail : m.jafar.sodik9@gmail.com*

ABSTRAK

Pencemaran udara adalah masuknya atau dimasukkannya zat, energi, dan atau komponen lain kedalam udara oleh kegiatan manusia sehingga mutu udara turun sampai ke tingkat tertentu yang menyebabkan atau mempengaruhi kesehatan manusia. Pencemaran udara dapat ditimbulkan oleh sumber – sumber alami maupun dari kegiatan manusia seperti aktivitas pabrik hingga aktifitas kendaraan bermotor. Indeks Standar Pencemaran Udara (ISPU) adalah angka yang tidak mempunyai satuan yang menggambarkan kondisi kualitas udara ambien di lokasi dan waktu tertentu yang dikeluarkan oleh kementerian lingkungan hidup dan kehutanan. Penetapan ISPU ini mempertimbangkan tingkat mutu udara terhadap kesehatan manusia, hewan, tumbuhan, bangunan dan nilai estetika. Indeks Standar Pencemaran Udara (ISPU) ditetapkan berdasarkan 5 pencemar, yakni karbon monoksida (CO), sulfur dioksida (SO₂), nitrogen dioksida (NO₂), ozon permukaan (O₃) dan partikel debu (PM₁₀). Algoritma Naive Bayes merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma Naive Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Pada penelitian ini penulis membandingkan keakuratan antara algoritma Naive Bayes dan K-Nearest Neighbor dalam pengklasifikasian kualitas udara berdasarkan indeks standar pencemaran udara (ISPU). Penelitian ini menghasilkan akurasi Naive Bayes sebesar 91.862 % dan akurasi K-Nearest Neighbor sebesar 97.3396 %.

Kata kunci : Pencemaran Udara, ISPU, Naive Bayes, KNN

1. PENDAHULUAN

Pencemaran udara adalah masuknya atau dimasukkannya zat, energi, dan atau komponen lain kedalam udara oleh kegiatan manusia sehingga mutu udara turun sampai ke tingkat tertentu yang menyebabkan atau mempengaruhi kesehatan manusia. Pencemaran udara dapat ditimbulkan oleh sumber – sumber alami maupun dari kegiatan manusia seperti aktivitas pabrik hingga aktifitas kendaraan bermotor.

Indeks Standar Pencemaran Udara (ISPU) adalah angka yang tidak mempunyai satuan yang menggambarkan kondisi kualitas udara ambien di lokasi dan waktu tertentu yang dikeluarkan oleh kementerian lingkungan hidup dan kehutanan. Indeks Standar Pencemaran Udara (ISPU) memberikan laporan kualitas udara kepada masyarakat untuk menerangkan seberapa bersih atau tercemarnya kualitas udara dan bagaimana dampaknya terhadap kesehatan setelah menghirup udara tersebut selama

beberapa jam/hari/bulan. Penetapan ISPU ini mempertimbangkan tingkat mutu udara terhadap kesehatan manusia, hewan, tumbuhan, bangunan dan nilai estetika. Indeks Standar Pencemaran Udara (ISPU) ditetapkan berdasarkan 5 pencemar, yakni karbon monoksida (CO), sulfur dioksida (SO₂), nitrogen dioksida (NO₂), ozon (O₃) dan partikel debu (PM₁₀).

Peningkatan nilai indeks standar pencemara udara sangat berpengaruh pada kesehatan manusia, itu dibuktikan pada hubungan antara peningkatan nilai ISPU akibat kebakaran hutan dengan kejadian gangguan pernafasan di kota pekanbaru pada tahun 2015. Selama periode september sampai dengan oktober 2015 terjadi kenaikan ISPU akibat kebakaran hutan dan lahan di kota pekanbaru melebihi ambang batas aman bagi kesehatan. Penurunan kualitas udara menyebabkan peningkatan kunjungan pasien akibat gangguan pada saluran pernapasan (ispa, asma,

pneumonia). Kecenderungan peningkatan ISPU di kota Pekanbaru mengikuti kecenderungan peningkatan kasus ISPA, asma, dan pneumonia di hari yang sama [1].

Algoritma *Naive Bayes* merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma *Naive Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.

Algoritma *K-Nearest Neighbor* (KNN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Algoritma *K-Nearest Neighbor* (KNN) termasuk dalam supervised learning, dimana hasil query instance yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN.

Pada penelitian ini penulis akan membandingkan keakuratan antara algoritma *Naive Bayes* dan *K-Nearest Neighbor* dalam pengklasifikasian kualitas udara berdasarkan indeks standar pencemaran udara (ISPU).

2. LANDASAN TEORI

2.1 *Naive Bayes*

[4] *Naive Bayes* merupakan sebuah pengklasifikasian *probabilistic* sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma ini menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variable kelas.

[5] Algoritma *Naive Bayes* bekerja berdasarkan probabilitas yang dimiliki data latih. Algoritma *Naive Bayes* sangat optimal jika digunakan dalam dataset yang tidak terlalu banyak dan memiliki fitur atau parameter yang sedikit.

Adapun, langkah-langkah algoritma *Naive Bayes* adalah:

- Langkah 1 : Kelompokkan data latih berdasarkan kelas yang sudah ada.
 Langkah 2 : Hitung *mean* dan *standar deviasi* untuk setiap parameter.

$$mean = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$S = \sqrt{\frac{\sum (x - mean)^2}{n - 1}}$$

Dimana:

x = nilai sampel.

n = jumlah seluruh sampel.

- Langkah 3 : Hitung probabilitas untuk setiap kelas.

$$P(A) = \frac{x_A}{n}$$

Dimana:

P(A) = Probabilitas untuk kelas A.

X_A = Jumlah data A.

n = Jumlah seluruh data.

- Langkah 4 : Hitung probabilitas untuk setiap parameter data menggunakan distribusi *gaussian*.

$$P(H = x | C = j) = \frac{1}{\sqrt{2 \times \pi \times S^2}} \times e^{-\frac{(x - mean)^2}{2 \times S^2}}$$

Dimana:

P(H = x | C = j) = Probabilitas parameter H dengan nilai x dan kelas j.

S = *Standar Deviasi*.

x = Nilai data pada data uji.

e = *exponent* (2,7182).

- Langkah 5 : Hitung probabilitas *likelihood* untuk setiap kelas.

$$P(C = j) = P(H_1 | C_j) \times P(H_2 | C_j) \times \dots \times P(H_n | C_j)$$

Dimana:

P(C = j) = Probabilitas *likelihood* dengan kelas j.

P(H₁ | C = j) = Probabilitas parameter H dengan kelas j.

- Langkah 6 : Hitung probabilitas akhir untuk setiap kelas.

$$P(j) = \frac{P(C = j)}{\sum P(C = n)}$$

Dimana:

P(j) = Probabilitas akhir kelas j.

P(C = j) = probabilitas *likelihood* kelas j.

∑P(C = n) = Jumlah probabilitas *likelihood* semua kelas.

- Langkah 7 : Kelas yang memiliki probabilitas akhir paling mendekati 1 adalah hasil dari klasifikasi.

2.2 K – Nearest Neighbor

[3] *K-Nearest Neighbor* (KNN) adalah sebuah metode pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada.

[2] Algoritma *K-Nearest Neighbor* adalah algoritma untuk mengklasifikasi objek baru berdasarkan atribut dan training *samples*. Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori yang ada.

Adapun, langkah-langkah algoritma *K – Nearest Neighbor* adalah:

Langkah 1 : Tentukan nilai parameter K (jumlah tetangga terdekat yang akan digunakan).

Langkah 2 : Hitung jarak antara data uji dengan setiap data latih. Beberapa metode pencarian jarak diantaranya:

1. *Euclidean Distance*

$$\text{distance}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

2. *Manhattan Distance*

$$\text{distance}(x, y) = \sum_{i=1}^n |x_i - y_i|$$

3. *Minkowski Distance*

$$\text{distance}(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

Dimana:

x = data latih.

y = data uji.

p = lamda

Langkah 3 : Urutkan hasil perhitungan jarak mulai dari yang paling kecil (*ascending*).

Langkah 4 : Ambil kelompok jarak terkecil sesuai dengan nilai parameter K yang sudah ditentukan.

Langkah 5 : Ambil hasil kategori mayoritas dari pengelompokkan tadi.

2.3 Pencemaran Udara

Menurut Keputusan Menteri Kesehatan Republik Indonesia Nomor 1407 Tahun 2002, pencemaran udara adalah masuknya atau dimasukkannya zat, energi, dan

atau komponen lain kedalam udara oleh kegiatan manusia sehingga mutu udara turun sampai ke tingkat tertentu yang menyebabkan atau mempengaruhi kesehatan manusia.

2.4 Indeks Standar Pencemaran Udara

Menurut Keputusan Menteri Negara Lingkungan Hidup Nomor KEP-45 Tahun 1997, indeks standar pencemaran udara adalah angka yang tidak mempunyai satuan yang menggambarkan kondisi kualitas udara ambien di lokasi dan waktu tertentu yang didasarkan kepada dampak terhadap kesehatan manusia, nilai estetika dan makhluk hidup lainnya. Parameter yang digunakan dalam perhitungan indeks standar pencemaran udara ada 5, diantaranya:

a. Partikel Debu (PM10)

Partikel debu (PM) adalah istilah untuk partikel padat atau cair yang ditemukan di udara. Partikel dengan ukuran besar atau cukup padat biasanya disebut asap, sedangkan partikel yang sangat kecil dapat dilihat dengan bantuan mikroskop *electron*.

b. *Sulfur Dioksida* (SO₂)

Pencemaran oleh *sulfur oksida* terutama disebabkan oleh dua komponen *sulfur* berbentuk gas yang tidak berwarna, yaitu *sulfur dioksida* (SO₂) dan *sulfur trioksida* (SO₃) kemudian keduanya menjadi *sulfur oksida* (SO_x). *Sulfur Dioksida* memiliki karakteristik bau yang tajam dan tidak mudah terbakar diudara, sedangkan *sulfur trioksida* merupakan komponen yang tidak reaktif.

c. *Karbon Monoksida* (CO)

Karbon monoksida merupakan senyawa yang tidak berbau, tidak berasa dan pada suhu udara normal berbentuk gas yang tidak berwarna. Tidak seperti senyawa lain, CO mempunyai potensi bersifat racun yang berbahaya karena dapat menghalangi *hemoglobin* mengangkut *Oksigen* yang dibutuhkan oleh tubuh manusia. Hal ini disebabkan karena *hemoglobin* lebih mudah mengikat *Karbon Monoksida* dibandingkan *Oksigen*.

d. *Ozon* (O₃)

Ozon (O₃) merupakan bentuk *oksigen* yang sangat reaktif. *Ozon* memiliki ciri berwarna biru pucat dan memiliki bau yang menyengat. *Ozon* menjadi berbahaya ketika muncul di permukaan tanah dengan konsentrasi diatan 50 ppm. *Ozon* menjadi berbahaya karena *Ozon* merupakan oksidan yang sangat kuat, sehingga ketika terhirup *Ozon* dapat merusak jaringan *mukosa* dan pernapasan manusia, binatang dan tumbuhan.

e. *Nitrogen Dioksida* (NO₂)

Nitrogen Dioksida (NO₂) memiliki karakteristik bau yang sangat menyengat dan berwarna merah kecoklatan. Sifat racun yang dimiliki *Nitrogen Dioksida* empat kali lebih kuat dibandingkan *Nitrogen Oksida* (NO). Organ tubuh yang paling peka terhadap pencemaran gas NO₂ yakni paru – paru. Paru – paru yang terkontaminasi gas NO₂ akan membengkak sehingga penderita sulit bernapas yang dapat menyebabkan kematian.

Nilai dari indeks standar pencemaran udara memiliki kategori tingkat pencemaran udara. Kategori tersebut yang digunakan untuk menentukan tingkat pencemaran udara yang terdeteksi. Nilai dan kategori tingkat pencemaran udara dapat dilihat pada tabel 1.

Tabel 1: Nilai dan Kategori Pencemaran Udara

Nilai ISPU	Kategori
1 – 50	Baik
51 – 100	Sedang
101 – 199	Tidak Sehat
200 – 299	Sangat Tidak Sehat
300 – 500	Berbahaya

3. METODOLOGI PENELITIAN

3.1 Data Penelitian

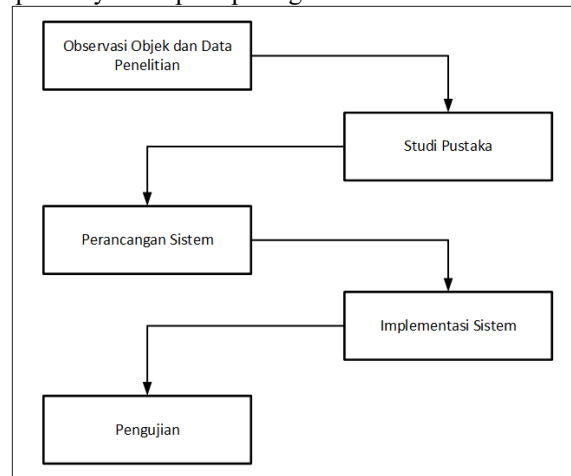
Data yang digunakan dalam penelitian ini adalah data Indeks Standar Pencemaran Udara DKI Jakarta tahun 2017 dan tahun 2018 yang didapatkan dari website <https://data.jakarta.go.id>. Informasi mengenai data dapat dilihat pada tabel 2.

Tabel 2: Data ISPU DKI Jakarta

Nama File	Jumlah Data
ISPU-SPKU-JANUARI-DESEMBER-2017.csv	1587
ISPU-di-SKPU-Bulan-Januari.csv	133
ISPU-di-SKPU-Bulan-Februari-edited.csv	115
ISPU-di-SPKU-Bulan-Maret-edited.csv	143
ISPU-di-SKPU-Bulan-April.csv	124
ISPU-di-SPKU-Bulan-Mei.csv	127
ISPU-di-SPKU-Bulan-Juni.csv	133
ISPU-di-SPKU-Bulan-Juli.csv	131
ISPU-di-SKPU-Bulan-Agustus.csv	144
ISPU-di-SPKU-Bulan-September.csv	145
ISPU-di-SPKU-Bulan-Oktober.csv	136
ISPU-di-SPKU-Bulan-November.csv	140
ISPU-SPKU-Bulan-Desember-Tahun-2018.csv	127

3.2 Metode Penelitian

Metode penelitian adalah langkah yang dimiliki dan dilakukan oleh peneliti dalam rangka untuk mengumpulkan informasi atau data serta melakukan investigasi pada data yang telah didapatkan tersebut. Metode penelitian memberikan gambaran rancangan penelitian yang meliputi antara lain: prosedur dan langkah-langkah yang harus ditempuh, waktu penelitian, sumber data, dan dengan langkah apa data-data tersebut diperoleh dan selanjutnya diolah dan dianalisis. Metode penelitian yang dilakukan penulis dalam penelitian yang dilakukan untuk membangun aplikasi yaitu seperti pada gambar 1.



Gambar 1: Metode Penelitian

3.2.1 Observasi Objek dan Data Penelitian

Observasi yang dilakukan untuk penelitian perbandingan metode *Naive Bayes* dan *K-Nearest Neighbor* pada klasifikasi kualitas udara di DKI Jakarta adalah dengan melakukan pengamatan informasi status kualitas udara di situs KLHK, melakukan pembelajaran mengenai cara perhitungan indeks standar pencemaran udara dan pengambilan data indeks standar pencemaran udara DKI Jakarta tahun 2017 – 2018 di website <https://data.jakarta.go.id>.

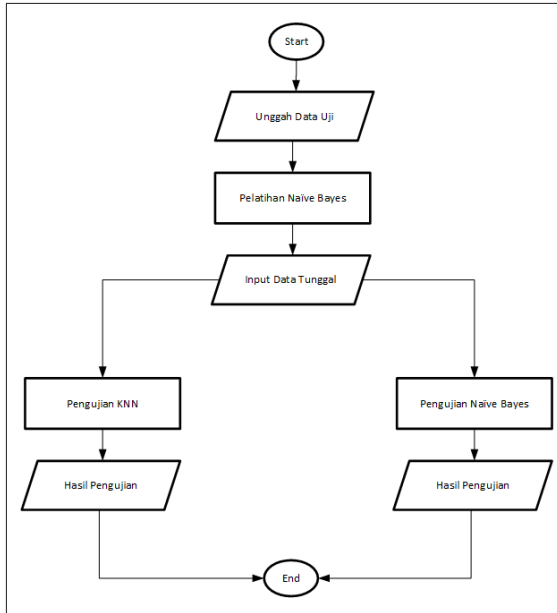
3.2.2 Studi Pustaka

Studi pustaka dilakukan untuk mencari rujukan dalam penelitian perbandingan metode *Naive Bayes* dan *K-Nearest Neighbor* pada klasifikasi kualitas udara di DKI Jakarta yakni dengan melakukan pencarian data indeks standar pencemaran udara beserta cara pengklasifikasiannya dan melakukan pencarian penelitian – penelitian sebelumnya yang memiliki objek yang sama.

3.2.3 Perancangan Sistem

Perancangan sistem merupakan proses membuat gambaran sistem yang akan dibangun. Gambaran

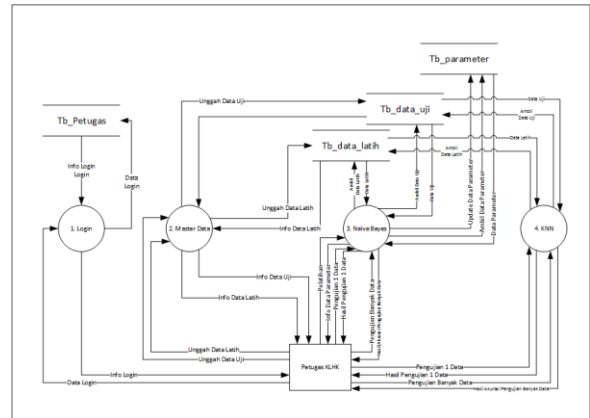
sistem ditampilkan dengan *flowchart* dan diagram alur data (DAD). *Flowchart* rancangan sistem dapat dilihat pada gambar 2 dan diagram alur data rancangan sistem dapat dilihat pada gambar 3.



Gambar 2: Rancangan Sistem

Langkah – langkah yang ada pada perancangan siste antara lain :

- 1) Memasukkan data latih dilakukan dengan mengunggah data indeks standar pencemaran udara dengan format *.csv. Dalam proses ini juga dilakukan *preprocessing* data. *Preprocessing* data dilakukan dengan menghapus data yang memiliki kategori “tidak ada data” dan yang memiliki fitur (pm10, so2, co, o3, no2) yang selain angka.
- 2) Pelatihan *Naive Bayes* dilakukan setelah data latih berhasil diunggah. Pada pelatihan *Naive Bayes*, dihitung nilai *mean* dan *standar deviasi* setiap parameter untuk setiap kategori.
- 3) Memasukkan data uji. Proses unggah data uji secara garis besar sama dengan proses unggah data latih, perbedaannya hanya pada media penyimpanannya. Jika data latih disimpan pada tabel data latih, maka data uji disimpan pada tabel data uji.
- 4) Pengujian *Naive Bayes*. Pada proses pengujian *Naive Bayes* dilakukan 2 macam pengujian yakni pengujian dengan 1 data dan pengujian dengan banyak data.
- 5) Pengujian *K-Nearest Neighbor*. Pada proses pengujian *K-Nearest Neighbor* dilakukan 2 macam pengujian yakni pengujian dengan 1 data dan pengujian dengan banyak data.



Gambar 3: Diagram Alur Data

Pada Gambar 3, dijelaskan bahwa ada 4 buah data storage, atau tabel yaitu: tabel petugas, tabel data latih, tabel data uji dan tabel parameter. Terdapat pula 4 buat proses yaitu proses login, proses master data, proses *Naive Bayes* dan proses KNN. Entitas yang terlibat dalam sistem ada 1, yaitu petugas KLHK.

3.2.4 Implementasi Sistem

Implementasi merupakan membuat sistem dari hasil perancangan. Dalam implementasi dilakukan dengan menggunakan bahasa pemrograman PHP, Javascript dan basisdata MySQL.

3.2.5 Pengujian

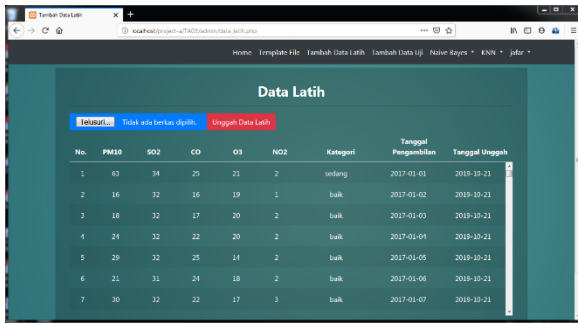
Sistem yang dibuat kemudian dilakukan pengujian untuk mendapatkan akurasi dari setiap metode. Pada metode *Naive Bayes* dilakukan pelatihan, pengujian dengan data uji dan pengujian dengan data acak. Sedangkan pada *K-Nearest Neighbor* dilakukan pengujian dengan data uji dan data acak dengan menggunakan 3 metode pencarian jarak yang berbeda dan 5 nilai K yang berbeda.

4. HASIL DAN PEMBAHASAN

4.1 Implementasi Program

1. Halaman Tambah Data Latih

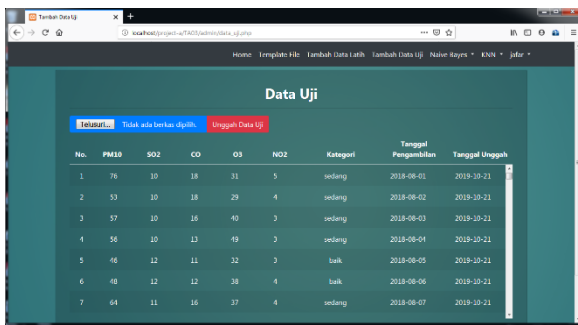
Halaman tambah data latih adalah halaman dimana petugas KLHK menambahkan data latih kedalam sistem. Adapaun cara menambahkannya yakni petugas KLHK mengunggah data ISPU dalam format file *.csv, implementasi halaman tambah data latih dapat dilihat pada gambar 4.



Gambar 4: Halaman Tambah Data Latih

2. Halaman Tambah Data Uji

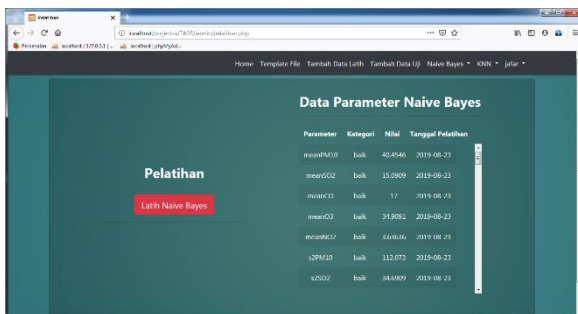
Halaman tambah data uji adalah halaman dimana petugas KLHK menambahkan data uji kedalam sistem. Adapaun cara menambahkannya yakni petugas KLHK mengunggah data ISPU dalam format file *.csv, implementasi halaman tambah data uji dapat dilihat pada gambar 5.



Gambar 5: Halaman Tambah Data Uji

3. Halaman Pelatihan *Naïve Bayes*

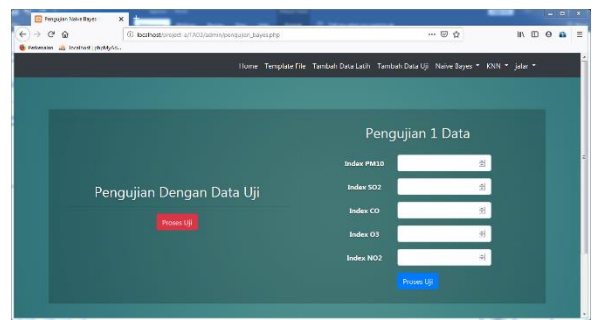
Halaman pelatihan naïve bayes adalah halaman dimana petugas KLHK melakukan pelatihan pada sistem untuk mendapatkan data parameter yang dibutuhkan dalam perhitungan naïve bayes. Implementasi halaman pelatihan naïve bayes dapat dilihat pada gambar 6.



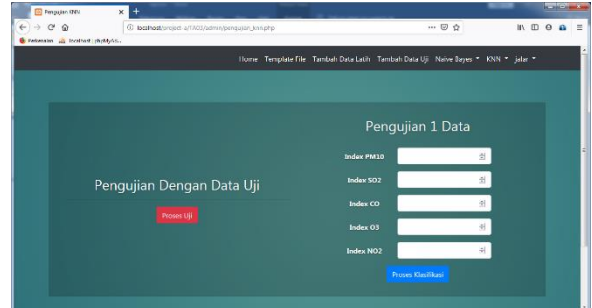
Gambar 6: Halaman Pelatihan *Naïve Bayes*

4. Halaman Pengujian

Halaman pengujian adalah halaman awal dimana petugas KLHK akan melakukan pengujian. Terdapat 2 pilihan pengujian yakni pengujian dengan data uji dan pengujian 1 data. Pada pengujian dengan data uji akan menghasilkan akurasi sistem dalam metode yang dipilih, sedangkan pengujian 1 data akan menghasilkan klasifikasi tingkat kualitas udara berdasarkan data masukan dan metode yang dipilih. Terdapat 2 halaman pengujian, yakni halaman pengujian *Naïve Bayes* dan halaman pengujian KNN. Halaman pengujian *Naïve Bayes* dapat dilihat pada gambar 7 dan halaman pengujian KNN dapat dilihat pada gambar 8.



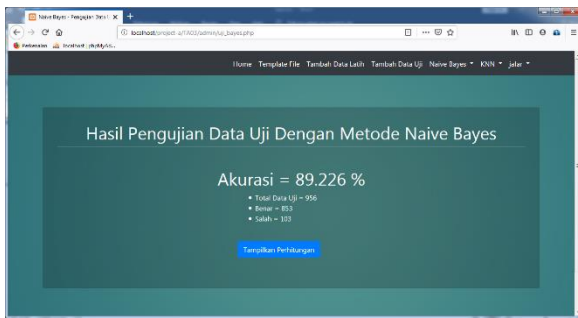
Gambar 7: Halaman Pengujian *Naïve Bayes*



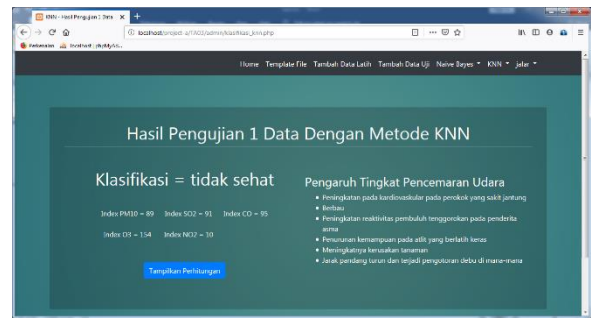
Gambar 8: Halaman Pengujian KNN

5. Halaman Hasil Pengujian Dengan Data Uji

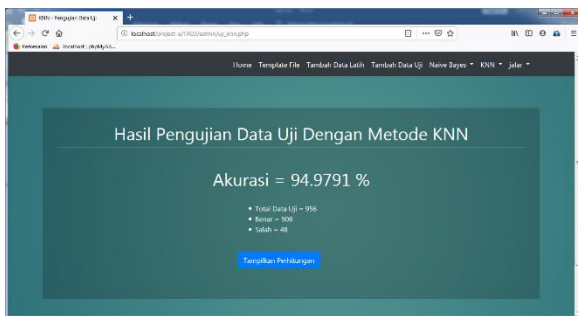
Halaman hasil pengujian dengan data uji adalah halaman dimana ditampilkannya hasil pengujian menggunakan data uji sesuai dengan metode yang dipilih. Terdapat 2 halaman hasil pengujian dengan data uji, yakni halaman *Naïve Bayes* dan halaman KNN. Halaman hasil pengujian dengan data uji menggunakan metode *Naïve Bayes* dapat dilihat pada gambar 9, dan halaman hasil pengujian dengan data uji menggunakan metode KNN dapat dilihat pada gambar 10.



Gambar 9: Halaman Hasil Pengujian Data Uji Dengan *Naive Bayes*

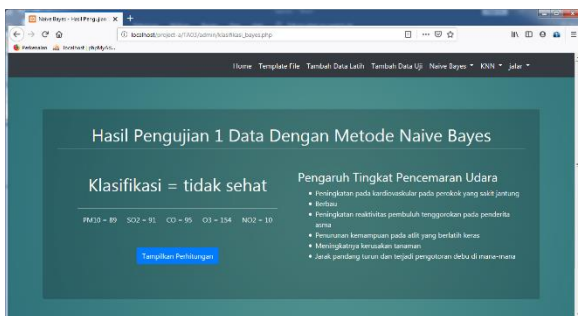


Gambar 12: Halaman Hasil Pengujian 1 Data Dengan KNN



Gambar 10: Halaman Hasil Pengujian Data Uji Dengan KNN

6. Halaman Hasil Pengujian Dengan 1 Data
 Halaman hasil pengujian dengan 1 data adalah halaman dimana ditampilkannya hasil pengujian menggunakan 1 data sesuai dengan metode yang dipilih. Terdapat 2 halaman hasil pengujian dengan 1 data, yakni halaman *Naive Bayes* dan halaman KNN. Halaman hasil pengujian dengan 1 data menggunakan metode *Naive Bayes* dapat dilihat pada gambar 11, dan halaman hasil pengujian dengan 1 data menggunakan metode KNN dapat dilihat pada gambar 12.



Gambar 11: Halaman Hasil Pengujian 1 Data Dengan *Naive Bayes*

4.2 Pelatihan *Naive Bayes*

Pelatihan *Naive Bayes* dilakukan untuk mendapatkan nilai parameter yang dibutuhkan untuk perhitungan *Naive Bayes*. Pada pelatihan *Naive Bayes* dicari nilai *mean* dan *standar deviasi* setiap parameter dan nilai probabilitas setiap kategori. Hasil pelatihan didapat nilai *mean* dan *standar deviasi* untuk setiap parameter dan nilai probabilitas untuk setiap kelas. Hasil pelatihan dapat dilihat pada tabel 3. Pelatihan menggunakan presentase data latih – data uji sebesar 80% - 20%.

Tabel 3: Hasil Pelatihan *Naive Bayes*

No	Parameter	Kategori	Nilai
1	Mean PM10	Baik	28.4719
2	Mean SO2	Baik	20.3371
3	Mean CO	Baik	16.8892
4	Mean O3	Baik	32.0851
5	Mean NO2	Baik	5.78973
6	Standar Deviasi PM10	Baik	148.761
7	Standar Deviasi SO2	Baik	92.465
8	Standar Deviasi CO	Baik	92.9829
9	Standar Deviasi O3	Baik	119.528
10	Standar Deviasi NO2	Baik	18.4493
11	Probabilitas Kategori	Baik	0.2503
12	Mean PM10	Sedang	54.3284
13	Mean SO2	Sedang	25.7934
14	Mean CO	Sedang	19.2456
15	Mean O3	Sedang	66.1253
16	Mean NO2	Sedang	9.61571
17	Standar Deviasi PM10	Sedang	247.371
18	Standar Deviasi SO2	Sedang	178.178

19	Standar Deviasi CO	Sedang	138.287
20	Standar Deviasi O3	Sedang	376.471
21	Standar Deviasi NO2	Sedang	26.9932
22	Probabilitas Kategori	Sedang	0.56719
23	Mean PM10	Tidak Sehat	62.3012
24	Mean SO2	Tidak Sehat	23.4348
25	Mean CO	Tidak Sehat	19.4689
26	Mean O3	Tidak Sehat	125.404
27	Mean NO2	Tidak Sehat	10.5932
28	Standar Deviasi PM10	Tidak Sehat	194.984
29	Standar Deviasi SO2	Tidak Sehat	109.281
30	Standar Deviasi CO	Tidak Sehat	115.371
31	Standar Deviasi O3	Tidak Sehat	562.154
32	Standar Deviasi NO2	Tidak Sehat	31.8122
33	Probabilitas Kategori	Tidak Sehat	0.12956
34	Mean PM10	Sangat Tidak Sehat	69.7273
35	Mean SO2	Sangat Tidak Sehat	23.5455
36	Mean CO	Sangat Tidak Sehat	19.5455
37	Mean O3	Sangat Tidak Sehat	210.364
38	Mean NO2	Sangat Tidak Sehat	9.09091
39	Standar Deviasi PM10	Sangat Tidak Sehat	94.4182
40	Standar Deviasi SO2	Sangat Tidak Sehat	3.47273
41	Standar Deviasi CO	Sangat Tidak Sehat	36.0727
42	Standar Deviasi O3	Sangat Tidak Sehat	93.2545
43	Standar Deviasi NO2	Sangat Tidak Sehat	38.6909
44	Probabilitas Kategori	Sangat Tidak Sehat	0.00481

4.3 Pengujian

4.3.1 Pengujian *Naïve Bayes*

Pengujian metode *Naïve Bayes* dilakukan dengan 4 kali pengujian, setiap pengujian memiliki presentasi data latih dan data uji yang berbeda. Hasil pengujian dapat dilihat pada tabel 4.

Tabel 4: Hasil Pengujian *Naïve Bayes*

No	Data Latih	Data Uji	Akurasi
1	50 %	50 %	89.649 %
2	60 %	40 %	89.028 %
3	70 %	30 %	90.387 %
4	80 %	20 %	91.862 %

Berdasarkan hasil pengujian, dapat dilihat bahwa persentase data latih – data uji 80% - 20% memiliki akurasi yang tertinggi yakni 91.862 %.

4.3.2 Pengujian K – Nearest Neighbor

Pengujian *K-Nearest Neighbor* dilakukan sebanyak 4 kali dengan presentase data latih dan data uji yang berbeda, setiap pengujian menggunakan 5 nilai K yang berbeda dan menggunakan 3 metode pencarian jarak, yakni *euclidian distance*, *manhattan distance* dan *minkowski distance*.

Tabel 5: Hasil Pengujian KNN Dengan Data Latih 80% dan Data Uji 20%

No	Nilai K	Hasil Akurasi		
		Euclidian	Manhattan	Minkowski
1	1	96.4006 %	96.4006 %	96.4006 %
2	3	96.5571 %	96.5571 %	96.4006 %
3	5	96.8701 %	97.0266 %	96.4006 %
4	7	96.8701 %	97.3396 %	96.2441 %
5	9	96.2441 %	97.3396 %	96.0876 %

Tabel 6: Hasil Pengujian KNN Dengan Data Latih 70% dan Data Uji 30%

No	Nilai K	Hasil Akurasi		
		Euclidian	Manhattan	Minkowski
1	1	95.5068 %	95.1933 %	95.1933 %
2	3	95.1933 %	95.6113 %	95.0888 %
3	5	95.2978 %	95.8203 %	95.0888 %
4	7	95.2978 %	96.0293 %	94.9843 %
5	9	95.5068 %	96.1338 %	94.9843 %

Tabel 7: Hasil Pengujian KNN Dengan Data Latih 60% dan Data Uji 40%

No	Nilai K	Hasil Akurasi		
		Euclidian	Manhattan	Minkowski
1	1	94.3574 %	94.6708 %	94.2790 %
2	3	94.7492 %	95.4545 %	94.9843 %
3	5	95.0627 %	95.9248 %	95.4545 %
4	7	95.5329 %	96.0815 %	95.2194 %
5	9	95.6897 %	96.0815 %	95.6113 %

Tabel 8: Hasil Pengujian KNN Dengan Data Latih 50% dan Data Uji 50%

No	Nilai K	Hasil Akurasi		
		Euclidian	Manhattan	Minkowski
1	1	94.0402 %	94.6675 %	93.8519 %
2	3	95.2321 %	95.7967 %	94.7930 %
3	5	95.1066 %	95.9849 %	95.4203 %
4	7	95.6085 %	96.1731 %	95.2321 %
5	9	96.0477 %	96.9260 %	95.6085 %

Hasil pengujian KNN menggunakan 3 metode pencarian jarak dan 5 nilai k yang berbeda, didapatkan metode pencarian jarak *manhattan distance* dengan nilai K sebesar 7 dan presentase data latih – data uji 80% - 20% memiliki nilai akurasi terbesar, yakni 97.3396 %.

4.4 Pembahasan

Penelitian ini menekankan pada perbandingan metode *Naïve Bayes* dengan *K-Nearest Neighbor* dalam pengklasifikasian tingkat kualitas udara berdasarkan indeks standar pencemaran udara. Data yang digunakan yakni data indeks pencemaran udara DKI Jakarta tahun 2017 sampai tahun 2018 yang didapatkan dari <https://data.jakarta.go.id>. Pada penelitian ini didapatkan bahwa akurasi *Naïve Bayes* ditentukan seberapa banyaknya data latih yang tersedia, semakin banyak data latih akurasi akan semakin bagus. Pada pengujian *Naïve Bayes* dengan menggunakan data latih sebesar 80% dari data yang telah disiapkan didapatkan akurasi sebesar 91.862 %. Akurasi *K-Nearest Neighbor* dipengaruhi oleh metode pencarian jarak dan nilai k yang digunakan, pada penelitian ini akurasi terbaik metode *K-Nearest Neighbor* menggunakan metode pencarian jarak *Manhattan Distance* dan nilai K sebesar 7 mendapatkan akurasi sebesar 97.3396 %. Pada penelitian ini dapat disimpulkan metode *K-Nearest Neighbor* memiliki akurasi yang lebih baik dari metode *Naïve Bayes* dengan catatan data yang digunakan yakni data indeks standar pencemaran udara DKI Jakarta tahun 2017 sampai tahun 2018 dengan presentase 80% data digunakan sebagai data latih dan 20% digunakan sebagai data uji.

5. PENUTUP

5.1. Kesimpulan

Setelah melakukan penelitian ini dapat didapatkan beberapa kesimpulan antara lain:

1. Metode *K-Nearest Neighbor* dengan *Manhattan Distance* dan nilai K sebesar 7 memiliki tingkat akurasi yang lebih baik dibandingkan metode *Naïve Bayes* yakni sebesar 97.3396 % berbanding dengan 91.862 %.
2. Pada metode *Naïve Bayes*, jumlah data latih dan keberagaman data berpengaruh pada akurasi yang dihasilkan.
3. Pada metode *K-Nearest Neighbor*, metode pencarian jarak dan nilai K berpengaruh pada akurasi yang dihasilkan.

5.2. Saran

Berdasarkan kesimpulan diatas, penulis memberikan saran untuk data latih yang digunakan sebaiknya memiliki keberagaman data yang tinggi, sehingga akurasi pada metode *Naïve Bayes* dapat lebih baik lagi.

DAFTAR PUSTAKA

- [1] Hermawan, A., Hananto, M. dan Lasut, D. (2016), *Peningkatan Indeks Standar Pencemaran Udara (ISPU) dan Kejadian Gangguan Saluran Pernapasan di Kota Penkanbaru.*, 76–86.
- [2] Krisandi, N., Prihandono, B. dan Helmi (2013), *Algoritma K - Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT. MINAMAS Kecamatan Parindu, Buletin Ilmiah Math.Stat.Dan Terapannya(Bimaster), 02(1), 33–38.*
- [3] Murdianingsih, Y. dan Lukmana, A. (2017), *SISTEM PENDUKUNG KEPUTUSAN ANALISIS PENGAJUAN KREDIT MOTOR MENGGUNAKAN METODE KNN (K Nearest Neighbor) (Study Kasus pada PT. Federal International Finance (FIF) Cabang Subang) Yuli Murdianingsih *1 , Angga Lukmana #2, Jurnal Teknologi Informasi Dan Komunikasi, 34–62.*
- [4] Patil, T.R. dan Sherekar, S.. (2013), *Performance Analysis of Naive Bayes And J48 Classification Algorithm for Data Classification, International Journal Of Computer Science And Applications, 6(2), 256–261.*
- [5] Sela, E.I. dan Pulungan, R. (2019), *Osteoporosis identification based on the validated trabecular area on digital dental radiographic images, Procedia Computer Science, 157, 282–289.*