

# Penggunaan Algoritma Naive Bayes dalam klasifikasi Pengaruh Pencemaran Udara

Amri Wicahyo<sup>1\*</sup>, Ahmad Pudoli<sup>2</sup>, Dewi Kusumaningsih<sup>3</sup>, Noripansyah<sup>4</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Universitas Budi Luhur, Indonesia

<sup>4</sup>Program Studi Magister Ilmu Komputer, Universitas Budi Luhur, Indonesia

Email: <sup>1</sup>1311502643@student.budiluhur.ac.id, <sup>2</sup>ahmad.pudoli@budiluhur.ac.id,

<sup>3</sup>dewi.kusumaningsih@budiluhur.ac.id, <sup>4</sup>1611602051@student.budiluhur.ac.id

---

## INFORMASI ARTIKEL

### Histori artikel:

Naskah masuk, 20 Juni 2021

Direvisi, 24 Juni 2021

Diiterima, 24 Juni 2021

---

## ABSTRAK

*Abstract- one of the factors used to identify environmental problems and living things is information on the effect of air pollution. In overcoming this problem, there are 20 types of classification of the effect of air pollutants on the decision of the Environmental Impact Management Agency (BAPEDAL) for the third attachment regarding the Effect of the Air Pollutant Standard Index for Each Pollutant Parameter. Referring to the problem, this study uses the DKI Jakarta dataset source which collects data from the Air Quality Monitoring Station (SPKU) and uses the Air Pollution Effect Classification application which aims to obtain information on the effect of air pollution by conducting a classification process based on past data from the dataset in 2018 to 2020. The application can run the Knowledge Discovery in Database (KDD) process including data mining or data mining using the classification method with the Naive Bayes algorithm. In this application there is a dataset processing feature into a training data to make the dataset accurate as a determining variable in the Naive Bayes classification process. So, the result of this application is that it can give the results of the classification of the effect of air pollution according to past data. For the testing process, 129 lines of testing data against 4061 lines of training data resulted in an accuracy of 96% of the classification of the effects of air pollution. Thus the conclusion of this study, can provide information on the effect of air pollution that is useful for living things and the environment with a fairly high accuracy of past data..*

### Kata Kunci:

*Klasifikasi, naïve bayes, pencemaran, udara*

**Abstrak-** salah satu faktor yang digunakan untuk mengidentifikasi masalah lingkungan dan makhluk hidup adalah informasi pengaruh pencemaran udara. Dalam mengatasi masalah tersebut, terdapat 20 jenis klasifikasi pengaruh pencemar udara pada keputusan Badan Pengendalian Dampak Lingkungan (BAPEDAL) untuk lampiran ketiga mengenai Pengaruh Indeks Standar Pencemar Udara Untuk Setiap Parameter Pencemar. Merujuk kepada masalah, penelitian ini menggunakan sumber dataset DKI Jakarta yang melakukan pengumpulan data dari Stasiun Pemantau Kualitas Udara (SPKU) serta menggunakan aplikasi Klasifikasi Pengaruh Pencemaran Udara yang bertujuan untuk mendapatkan informasi pengaruh pencemaran udara dengan melakukan proses klasifikasi berdasarkan data masa lalu dari dataset di tahun 2018 sampai dengan 2020. Aplikasi tersebut dapat menjalankan proses Knowledge Discovery in Database (KDD) termasuk melakukan penambangan data atau data mining menggunakan metode klasifikasi dengan algoritma Naive Bayes. Pada aplikasi tersebut terdapat fitur pengolahan dataset menjadi sebuah data training untuk menjadikan dataset tersebut akurat sebagai variabel penentu dalam melakukan proses klasifikasi Naive Bayes. Maka, hasil dari aplikasi tersebut yaitu dapat memberikan hasil klasifikasi pengaruh pencemaran udara sesuai dengan data masa lalu. Untuk proses pengujian 129 baris data testing terhadap 4061 baris data training menghasilkan akurasi klasifikasi pengaruh pencemaran udara sebesar 96%.

Copyright © 2019 LPPM - STMIK IKMI Cirebon  
This is an open access article under the CC-BY license

---

**Penulis Korespondensi:**

**Dewi Kusumaningsih**

Program Studi Teknik Informatika,  
Universitas Budi Luhur

Jl. Ciledug Raya, Kelurahan Petukangan Utara, Kecamatan Pesanggrahan, Jakarta Selatan

Email: [dewi.kusumaningsih@budiluhur.ac.id](mailto:dewi.kusumaningsih@budiluhur.ac.id)

---

## 1. Pendahuluan

Dalam proses pencemaran udara terdapat komponen atau zat lain yang masuk ke dalam udara, komponen tersebut yaitu Partikulat (PM10), Sulfur Dioksida (SO<sub>2</sub>), Karbon Monoksida (CO), Ozon (O<sub>3</sub>), dan Nitrogen Dioksida (NO<sub>2</sub>) dapat menjadi indikator parameter tingkat baik atau buruk udara pada suatu daerah termasuk di DKI Jakarta. Indikator parameter di atas termasuk dalam aspek pemantauan pada Stasiun Pemantauan Kualitas Udara (SPKU) yang tersebar untuk lima wilayah DKI Jakarta yaitu di Bundaran HI untuk Jakarta Pusat, Kelapa Gading untuk Jakarta Utara, Jagakarsa untuk Jakarta Selatan, Lubang Buaya untuk Jakarta Timur, dan Kebon Jeruk untuk Jakarta Barat.

Pada data hasil pemantauan di beberapa SPKU DKI Jakarta telah dapat memberikan informasi kategori baik, sehat, tidak sehat, sangat tidak sehat, dan berbahaya. Akan tetapi data hasil pemantauan tersebut belum dapat menampilkan informasi berupa klasifikasi pengaruh terhadap indikator parameter pencemaran udara terutama pada kondisi parameter masing-masing bernilai sama.

Untuk itu pada penelitian ini bertujuan dapat menyelesaikan masalah dalam memberikan informasi berupa klasifikasi pengaruh terhadap indikator parameter berbasis dekstop. Proses pengklasifikasian pengaruh tersebut menggunakan metode algoritma Naive Bayes dengan hasil klasifikasi berupa tidak ada efek, sedikit berbau, luka pada beberapa spesies tumbuhan akibat kombinasi dengan SO<sub>2</sub>, luka pada beberapa spesies tumbuhan akibat kombinasi dengan O<sub>3</sub>, perubahan kimia darah tetapi tidak terdeteksi, berbau, luka pada beberapa spesies tumbuhan, terjadi penurunan jarak pandang, peningkatan pada kardiovaskular pada perokok yang sakit jantung, bau dan kehilangan warna, penurunan kemampuan pada atlet yang berlatih keras, bau dan meningkatnya kerusakan tanaman, jarak pandang turun dan terjadi pengotoran debu di manamana, meningkatnya kardiovaskular pada orang bukan perokok yang berpenyakit jantung, meningkatnya sensitivitas pasien yang berpenyakit asma dan bronkitis, dan olahraga ringan mengakibatkan pengaruh pernapasan pada pasien yang berpenyakit paru-paru kronis.

Berdasarkan latar belakang masalah di atas, maka perumusan masalah yang dihasilkan adalah Apa hasil klasifikasi pengaruh pencemaran udara berdasarkan indikator parameter pencemaran udara, serta metode apa yang cocok dalam proses untuk mendapatkan hasil klasifikasi pengaruh pencemaran udara serta hasil akurasi terhadap hasil klasifikasi.

Berdasarkan perumusan masalah di atas, maka batasan masalah untuk Klasifikasi Pengaruh Pencemaran Udara Menggunakan Algoritma Naive Bayes Berbasis Dekstop adalah sebagai berikut: a. Aplikasi ini hanya untuk Klasifikasi Pengaruh Pencemaran Udara. b. Aplikasi ini menggunakan sumber data dari data.jakarta.go.id. c. Data yang menjadi acuan yaitu dari bulan Januari tahun 2018 hingga bulan Juni tahun 2020. d. Software menggunakan bahasa pemrograman Java dan berbasis Dekstop

## 2. Metode

### 2.1 Klasifikasi

Klasifikasi merupakan suatu proses pengelompokan suatu data menjadi beberapa kelompok yang memiliki keterkaitan dalam ruang lingkup masalah yang ada dengan tujuan akhir untuk menemukan model atau fungsi yang menggambarkan dan membedakan tiap kelompok data sehingga dapat memprediksikan suatu kelompok data dari data yang belum terdapat pada beberapa kelompok data. Menurut [1].

“Klasifikasi adalah pengelompokan data atau objek baru ke dalam kelas atau label berdasarkan atribut-atribut tertentu”. Berdasarkan peneliti lain mengemukakan sebagai berikut: “Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui, dalam mencapai tujuan tersebut klasifikasi membentuk suatu model yang mampu membedakan data dalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu, dengan model yang berupa pohon keputusan, atau formula matematis” [2].

## 2.2 Algoritma Naive Bayes

Algoritma Naive Bayes secara umum merupakan berasal dari teorema Bayes yang memiliki pengertian suatu proses prediksi peluang untuk masa depan berdasarkan pengalaman masa lalu. Algoritma Naive Bayes sendiri merupakan salah satu teknik pengklasifikasian terhadap data yang bersifat kuantitatif dan diskrit dengan menggunakan metode probabilitas dan statistik untuk mendapatkan hasil hipotesis sebagai informasi baru dalam mengambil keputusan. Menurut peneliti lain Mustafa, mengemukakan sebagai berikut: “Algoritma Naive Bayes Classifier (NBC) dapat mengolah data kuantitatif dan data diskrit yang hanya memerlukan sejumlah kecil data pelatihan untuk perhitungan estimasi peluang yang dibutuhkan untuk klasifikasi”[3].

“Metode NBC di gunakan sebagai alat pengambilan keputusan untuk memperbaharui suatu informasi” [4]. Untuk dapat menjelaskan algoritma Naive Bayes dalam menjalankan proses klasifikasi, berikut ini persamaan teorema Bayes (1):

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (1)$$

Dengan keterangan P adalah probabilitas. X adalah data dengan kelas yang belum teridentifikasi. H adalah hipotesis data X yang merupakan suatu kelas spesifik. P(H) adalah probabilitas hipotesis H. P(X) adalah probabilitas dari X. P(H|X) adalah probabilitas hipotesis H berdasarkan kondisi X. P(X|H) adalah probabilitas X berdasarkan kondisi tersebut

## 2.3 Indeks Parameter standar pengukuran

Dalam menentukan dampak atau pengaruh pencemaran udara terdapat indeks parameter sebagai standar pengukuran dampak atau pengaruh pencemaran udara. Terdapat lima indeks parameter sebagai standar pengukuran yaitu Partikulat (PM10), Sulfur Dioksida (SO<sub>2</sub>), Karbon Monoksida (CO), Ozon (O<sub>3</sub>), dan Nitrogen Dioksida (NO<sub>2</sub>).

## 2.4 Studi Literatur

Dalam penelitian Kurniawan & Kriestanto mengemukakan sebagai berikut: “Dalam menyelesaikan suatu masalah kurang hati-hati atau kurang cermat dalam proses analisis pemberian kredit yang menyebabkan kredit bermasalah atau kredit macet, dapat menggunakan metode penelitian penerapan algoritma Naive Bayes dalam melakukan proses klasifikasi dan prediksi. Ruang lingkup masalah yaitu menggunakan data survei Baitul Maal wat Tamwil (BMT) Beringharjo dan data nasabah BMT Beringharjo serta menggunakan bahasa pemrograman berbasis Jawa. Penelitian ini bertujuan sebagai alternatif dan alat bantu dalam

memprediksikan risiko kelayakan kredit yang memperkirakan layak atau tidak jika pemohon atau nasabah untuk diberikan kredit. Dalam penelitian ini data masukan menggunakan data pengajuan kredit yaitu jaminan, total penghasilan, pinjaman lain, total pengeluaran, nominal kredit, status rumah, dan tempo pengembalian. Pengklasifikasian data dan hasil proses adalah layak atau tidak layak. Data yang digunakan sebanyak 250 data. Kesimpulan dari penelitian ini yaitu penerapan algoritma data mining menggunakan Naive Bayes dapat dilakukan untuk memprediksi kelayakan kredit pada BMT Beringharjo Yogyakarta. Akurasi data yang diperoleh dalam penelitian ini adalah sebesar 92,5% dengan error sebesar 7,5% dari 160 data yang digunakan untuk training dan 40 data untuk testing. Semakin banyak kriteria yang digunakan dalam proses prediksi maka akurasi data akan semakin tinggi” [5].

Bustami dalam melakukan penelitian mengemukakan sebagai berikut: “Terdapat sangat banyak nasabah perusahaan asuransi menunggak dalam membayar premi yang merupakan pendapatan asuransi. Penelitian ini menggunakan algoritma Naive Bayes dalam melakukan klasifikasi dengan tujuan untuk memberikan informasi tentang teknik klasifikasi data nasabah yang akan bergabung dalam kelompok lancar, kelompok kurang lancar atau tidak lancar dalam membayar premi. Dalam penelitian ini data masukan menggunakan data calon nasabah. Proses klasifikasi untuk pembayaran lancar, pembayaran kurang lancar, dan pembayaran tidak lancar. Kesimpulan dari penelitian ini yaitu sistem klasifikasi data nasabah ini digunakan untuk menampilkan informasi klasifikasi lancar, kurang lancar atau tidak lancarnya calon nasabah dalam membayar premi asuransi dengan menggunakan algoritma Naive Bayes. Algoritma Naive Bayes didukung oleh ilmu Probabilitas dan ilmu statistika khususnya dalam penggunaan data petunjuk untuk mendukung keputusan pengklasifikasian. Pada algoritma Naive Bayes, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain. Variabel penentu yang digunakan dalam penelitian ini adalah jenis kelamin, usia, status, pekerjaan, penghasilan per tahun, masa pembayaran asuransi, dan cara pembayaran asuransi”[6].

Penelitian selanjutnya, Nofriansyah et al. (2016) mengemukakan sebagai berikut: “Perusahaan provider kartu internet ingin selalu meluncurkan kartu internet terbaru tanpa memikirkan kartu internet tersebut akan laku atau tidak di pasaran. Metode dalam penelitian ini yaitu penggunaan algoritma Naive Bayes dalam melakukan prediksi peluang masa depan berdasarkan pengalaman masa sebelumnya dengan

menggunakan data transaksi penjualan kartu perusahaan provider kartu internet XL yang pengumpulan data menggunakan metode studi literatur, observasi, dan wawancara. Penggunaan data masukan menggunakan data kasus yang berjumlah 20 baris dan proses klasifikasi yang menjelaskan pelanggan minat atau tidak. Kesimpulan dari penelitian ini yaitu sistem klasifikasi data kartu internet ini digunakan untuk menampilkan informasi klasifikasi minat atau tidak minat pada kartu internet yang baru di luncurkan dengan menggunakan algoritma Naive Bayes. Algoritma Naive Bayes sangat cocok diterapkan dalam memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga memudahkan perusahaan untuk memprediksi peminatan masyarakat terhadap kartu internet yang baru diluncurkan. Dengan mengetahui minat atau tidak minat kartu internet yang baru diluncurkan, akan meminimalkan kerugian pada perusahaan. Perusahaan juga akan lebih selektif dalam meluncurkan produk baru”[7]. Pada penyelesaian penelitian Mardi (2017) mengemukakan sebagai berikut: “Penghapusan database yang tidak bermanfaat dalam jangka waktu tertentu merupakan tindakan yang tidak baik karena informasi dari database tersebut dapat menghasilkan informasi baru. Tujuan dari penelitian ini yaitu untuk mengambil informasi yang terdapat dalam database dengan proses data mining dengan menggunakan data masukan berupa data klasifikasi risiko yang kemudian terdapat pengklasifikasian risiko kredit dalam bentuk nasabah bermasalah kredit atau tidak. Kesimpulan dari penelitian ini yaitu sumber data merupakan database yang di dalamnya terdapat informasi yang bisa diambil dan dimanfaatkan untuk kepentingan bisnis dan penelitian. Algoritma yang dapat dilakukan untuk membuat pohon keputusan salah satunya adalah algoritma C4.5”[8]. Selain itu berdasarkan penulis lain manfaat data mining yaitu “Dapat digunakan untuk mendapatkan penemuan pengetahuan dan atau informasi di dalam database” [9]. Dalam penelitian Anggraini et al. (2019) mengungkapkan sebagai berikut: “Semakin meningkat penyebaran informasi melalui blog menyebabkan perlu pengklasifikasian professional blogger atau bukan. Metode penelitian ini yaitu penggunaan salah satu algoritma data mining yaitu metode Naive Bayes menggunakan teknik klasifikasi dengan model deskriptif dan prediktif. Tujuan penelitian ini yaitu untuk melakukan klasifikasi data professional blogger atau bukan dengan algoritma Naive Bayes. Penggunaan data masukan berupa data survei 100 responden yang kemudian terdapat pengklasifikasian data menjadi Blogger Professional (BP) dan Blogger Musiman (BM). Kesimpulan dari penelitian ini yaitu dari hasil akurasi pengolahan data blogger

menggunakan metode Naive Bayes adalah baik. Hasil klasifikasi dari data blogger dengan metode Naive Bayes membagi 2 kelas klasifikasi BP yaitu kelas yes dan kelas no. Untuk nilai kelas yes yaitu 0,680 dan nilai kelas no yaitu 0,320. Dari hasil pengolahan data dapat diketahui bahwa tingkat akurasi pengklasifikasian data blogger mencapai 86,67%”[10].

## 2.5 Data Penelitian

Dalam menentukan pengaruh pencemaran udara terdapat Keputusan Kepala Badan Pengendalian Dampak Lingkungan KABAPEDAL yang mengatur standar pencemaran udara dalam bentuk indeks parameter yaitu Partikulat (PM10) yang merupakan partikel udara berukuran lebih kecil dari 10 mikrometer, *Sulfur Dioksida* (SO<sub>2</sub>) yang merupakan suatu parameter bersumber dengan menggunakan alat *ion chromatography* dengan pengukuran metode *Passive Gas*, Karbon Monoksida (CO) yang merupakan suatu gas bersumber dari pembakaran tidak sempurna, *Ozon* (O<sub>3</sub>) yang merupakan parameter dengan pengukuran menggunakan alat *ozone analyzer* dengan metode *UV Photometric*, dan *Nitrogen Dioksida* (NO<sub>2</sub>) yang merupakan parameter dengan pengukuran menggunakan alat *passive sampler* metode *Passive Gas*.

Berdasarkan standar pencemaran tersebut di atas, terdapat dataset berukuran besar yang bersumber dari pengumpulan dataset DKI Jakarta berdasarkan laporan Stasiun Pemantauan Kualitas Udara (SPKU) Bundaran Hotel Indonesia (HI), Kelapa Gading, Jagakarsa, Lubang Buaya, dan Kemayoran.

Tabel 1. Dataset

tanggal	stasiun	pm10	so2	co	o3	no2	max	critical	kategori
01/01/2020	DKI1 (Bundaran HI)	30	20	10	32	9	32	O3	BAIK
02/01/2020	DKI1 (Bundaran HI)	27	22	12	29	8	29	O3	BAIK
03/01/2020	DKI1 (Bundaran HI)	39	22	14	32	10	39	PM10	BAIK
04/01/2020	DKI1 (Bundaran HI)	41	---	27	19	18	41	PM10	BAIK
05/01/2020	DKI1 (Bundaran HI)	35	22	12	31	9	35	PM10	BAIK
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
30/01/2020	DKI5 (Kebon Jeruk) Jakarta Barat	27	19	16	61	8	61	O3	SEDANG
31/01/2020	DKI5 (Kebon Jeruk) Jakarta Barat	46	16	24	44	7	46	PM10	BAIK

## 2.6 Penerapan Metode

Berikut ini penerapan metode dalam melakukan proses KDD terhadap dataset SPKU dari

bulan Januari tahun 2018 hingga bulan Juni tahun 2020:

Data selection atau seleksi data merupakan tahapan pemilihan data pada dataset yang berkaitan dengan klasifikasi pengaruh pencemaran udara sesuai dengan indikator standar parameter pencemaran udara. Berdasarkan tabel dataset di atas, pemilihan data yang berkaitan dengan klasifikasi pengaruh pencemaran udara yaitu data PM10, SO<sub>2</sub>, CO, O<sub>3</sub>, dan NO<sub>2</sub> serta menambahkan atribut pengaruh sebagai wadah data klasifikasi berdasarkan standar pencemaran udara.

Tabel 2. Data Selection

pm10	so2	co	o3	no2	pengaruh
30	20	10	32	9	
27	22	12	29	8	
39	22	14	32	10	
41	---	27	19	18	
35	22	12	31	9	
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
27	19	16	61	8	
46	16	24	44	7	

### 2.7 Rancangan Dataset

Berikut ini adalah spesifikasi tabel yang digunakan pada dataset pengaruh pencemaran udara berdasarkan SPKU, data training, dan data testing.

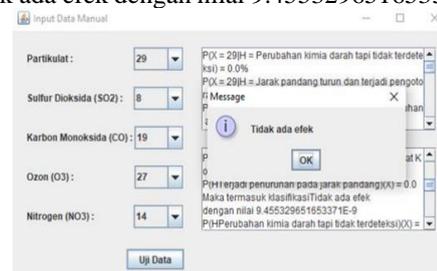
Tabel 3. Spesifikasi Dataset

No.	Nama Kolom	Keterangan
1.	Tanggal	Tanggal dari dataset
2.	Stasiun	Nama stasiun dari dataset
3.	PM10	Nilai PM10 dari dataset
4.	SO <sub>2</sub>	Nilai SO <sub>2</sub> dari dataset
5.	CO	Nilai CO dari dataset
6.	O <sub>3</sub>	Nilai O <sub>3</sub> dari dataset
7.	NO <sub>2</sub>	Nilai NO <sub>2</sub> dari dataset
8.	Max	Nilai tertinggi dari 5 parameter
9.	Critical	Nama dari parameter nilai tertinggi
10.	Category	Nama kategori dari rentang nilai parameter

### 3. Hasil dan Pembahasan

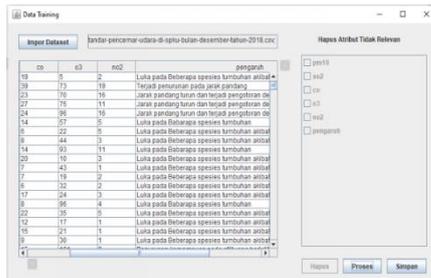
Untuk penjelasan tahapan metode klasifikasi Naive Bayes sebagai berikut: Menghitung Probabilitas Hipotesis (P(H)) pada data training P(H = perubahan kimia darah tapi tidak terdeteksi) = 8.59% P(H = jarak pandang turun dan terjadi pengotoran debu di mana-mana) = 0.25% P(H = luka

pada beberapa spesies tumbuhan akibat kombinasi dengan o<sub>3</sub>) = 0.25% P(H = penurunan kemampuan pada atlet yang berlatih keras) = 3.11% P(H = luka pada beberapa spesies tumbuhan) = 0.12% P(H = peningkatan pada kardiovaskular pada perokok yang sakit jantung) = 2.37% P(H = luka pada beberapa spesies tumbuhan) = 36.61% P(H = tidak ada efek) = 16.69% P(H = luka pada beberapa spesies tumbuhan akibat kombinasi dengan so<sub>2</sub>) = 8.22% P(H = terjadi penurunan pada jarak pandang) = 23.79%. Menghitung Probabilitas Hipotesis tiap kolom (P(X|H)) sesuai dengan masukan sampel data testing P(X = 29|H = Perubahan kimia darah tapi tidak terdeteksi) = 0.0% P(X = 29|H = Jarak pandang turun dan terjadi pengotoran debu di manamana) = 0.0% P(X = 29|H = Luka pada Beberapa spesies tumbuhan akibat kombinasi dengan O<sub>3</sub>) = 0.0% P(X = 29|H = Penurunan kemampuan pada atlet yang berlatih keras) = 0.0% P(X = 29|H = Luka pada Beberapa spesies tumbuhan) = 0.0% P(X = 29|H = Peningkatan pada kardiovaskular pada perokok yang sakit jantung) = 0.0% P(X = 29|H = Luka pada Beberapa spesies tumbuhan) = 0.0% P(X = 29|H = Tidak ada efek) = 1.4925373134328357% P(X = 29|H = Luka pada Beberapa spesies tumbuhan akibat Kombinasi dengan SO<sub>2</sub>) = 1.5151515151515151%. Menghitung Perkalian P(H) dengan P(X|H) P(H = Perubahan kimia darah tapi tidak terdeteksi)(X) = 0.0 P(H = Jarak pandang turun dan terjadi pengotoran debu di manamana)(X) = 0.0 P(H = Luka pada Beberapa spesies tumbuhan akibat kombinasi dengan O<sub>3</sub>)(X) = 0.0 P(H = Penurunan kemampuan pada atlet yang berlatih keras)(X) = 0.0 P(H = Luka pada Beberapa spesies tumbuhan)(X) = 0.0 P(H = Peningkatan pada kardiovaskular pada perokok yang sakit jantung)(X) = 0.0 P(H = Luka pada Beberapa spesies tumbuhan)(X) = 0.0 P(H = Tidak ada efek)(X) = 9.455329651653371E-9 P(H = Luka pada Beberapa spesies tumbuhan akibat Kombinasi dengan SO<sub>2</sub>)(X) = 0.0 P(H = Terjadi penurunan pada jarak pandang)(X) = 0.0 Maka termasuk klasifikasi Tidak ada efek dengan nilai 9.455329651653371E-9



Gambar 1. Hasil Sample Uji Coba.

Gambar 2 dari pengujian tombol proses pada form data training setelah dilakukan seleksi, pra-pemrosesan, dan transformasi.



Gambar 2. pengujian tombol proses pada form data training setelah dilakukan seleksi, pra-pemrosesan, dan transformasi.

Pada tahap pengujian performa aplikasi dengan menggunakan Confusion Matrix untuk menghasilkan nilai Accuracy, Precision dan Recall. Pengujian tersebut menggunakan dataset yang sudah melalui tahapan-tahapan yang diperlukan dalam data mining, kemudian data tersebut dipecah menjadi 70% Data Training dan 30% Data Testing. Tabel Confusion Matrix dapat dilihat pada Tabel 3.

Tabel 3. Tabel Confusion Matrix

Confusion Matrix		Prediction			
		Baik	Sedang	Tidak Sehat	Sangat Tidak Sehat
Actual	Baik	1	0	0	0
	Sedang	0	55	0	0
	Tidak Sehat	0	0	29	0
	Sangat Tidak Sehat	0	2	0	0

#### 4. Kesimpulan

Hasil dari penelitian ini bahwa Aplikasi Klasifikasi Pengaruh Pencemaran Udara berhasil menentukan klasifikasi pengaruh pencemaran udara menggunakan metode Naive Bayes terhadap dataset bulan Januari tahun 2018 hingga bulan Juni tahun 2020. Hasil klasifikasi pengaruh pencemaran udara mendapatkan nilai akurasi sebesar 96%. Hasil akurasi cukup baik untuk menentukan 129 data testing terhadap data training berkisar yang 4000 baris data, sehingga peneliti menyimpulkan metode Naive Bayes baik dalam melakukan klasifikasi pengaruh pencemaran udara.

#### Daftar Pustaka

[1] Septianto, R. H. (2015). Diagnosa Penyakit Tanaman Kopi Arabika dengan Metode Modified K-Nearest Neighbor. *Jurnal Mahasiswa PTIIK UB*, 6–12.

[2] Nofriansyah, D., Erwansyah, K., & Ramadhan, M. (2016). Penerapan Data Mining dengan Algoritma Naive Bayes Classifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL ( Studi Kasus di. *SAINTIKOM*, 15, 81–92.

[3] Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2018). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Creative Information Technology Journal*. <https://doi.org/10.24076/citec.2017v4i2.106>

[4] Sarwo. (2016). HIBRID METHOD MENGGUNAKAN DATA MINING DAN NAIVE BAYES MODEL UNTUK PREDIKSI STUDI KASUS KERUSAKAN LAMPU EFEK. *Jurnal Teknologi Pelita Bangsa*.

[5] Kurniawan, D. A., & Kristanto, D. (2016). PENERAPAN NAIVE BAYES UNTUK PREDIKSI KELAYAKAN KREDIT. *JIKO (Jurnal Informatika Dan Komputer)*. <https://doi.org/10.26798/jiko.2016.v1i1.10>

[6] Bustami. (2014). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *Jurnal Informatika Ahmad Dahlan*, 8(1), 102632. <https://doi.org/10.26555/jifo.v8i1.a2086>

[7] Nofriansyah, D., Erwansyah, K., & Ramadhan, M. (2016). Penerapan Data Mining dengan Algoritma Naive Bayes Classifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL ( Studi Kasus di. *SAINTIKOM*, 15, 81–92.

[8] Mardi, Y. (2017). *Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5*. Edik Informatika.

[9] Amelia, M. winny, Lumenta, A. S. ., & Jacobus, A. (2017). Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naive Bayes. *Jurnal Teknik Informatika*. <https://doi.org/10.35793/jti.11.1.2017.17652>

[10] Amelia, M. winny, Lumenta, A. S. ., & Jacobus, A. (2017). Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naive Bayes. *Jurnal Teknik Informatika*. <https://doi.org/10.35793/jti.11.1.2017.17652>