



ARTICLE

Identifikasi Pengaruh Kualitas Udara Terhadap Kondisi Pasien Covid-19 dengan Algoritma Naïve Bayes

Identify the Effect of Air Quality on the Condition of Covid-19 Patients with the Naïve Bayes Algorithm

Monica Dias Febriyanti, Ineke Pakereng, dan Alz Danny Wowor *

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Salatiga, Indonesia

*Penulis Korespondensi: alzdanny.wowor@uksw.edu

(Disubmit 23-03-05; Diterima 24-04-16; Dipublikasikan online pada 24-09-05)

Abstrak

Pandemi Covid-19 atau corona virus sangat meresahkan bagi masyarakat Indonesia karena penyebarannya yang sangat mudah, virus ini dapat ditularkan melalui tetesan kecil (droplet) dari mulut maupun hidung penderita covid-19. Pada kasus tertentu virus ini juga bisa menginfeksi pernafasan berat sehingga menyebabkan pneumonia (infeksi paru-paru). Masyarakat yang memiliki daya tahan tubuh lemah sangat mudah tertular virus ini, salah satu penyebab daya tahan tubuh melemah yaitu kualitas udara yang tidak bagus oleh karena itu penelitian ini melakukan identifikasi pengaruh dari kualitas udara terhadap kondisi pasien covid-19 menggunakan algoritma Naïve Bayes. Data diperoleh melalui website corona.jakarta.go.id dan Jakarta OpenData, setelah preprocessing diperoleh 610 data yang dibagi menjadi 80% data training dan 20% data testing. Penggunaan dengan Algoritma Naïve Bayes menunjukkan nilai akurasi yang tinggi yaitu 82,73%. Hasil identifikasi berdasarkan algoritma naïve bayes dalam empat class, class sangat berpengaruh 2,3%, class berpengaruh 62,8%, class kurang berpengaruh 34,5%, dan class tidak berpengaruh 0.4%. Dengan demikian kualitas udara ikut mempengaruhi kondisi pasien covid-19 di wilayah Propinsi DKI Jakarta.

Kata kunci: Covid-19; Kualitas Udara; Naive Bayes.

Abstract

The covid-19 pandemic or corona virus is very troubling for the people of Indonesia because of its very easy spread, this virus can be transmitted through small droplets from the mouth and nose of covid-19 sufferers. In certain cases this virus can also infect severe respiratory tract, causing pneumonia (lung infection). People who have a weak immune system are very susceptible to contracting this virus, one of the causes of a weakened immune system is poor air quality. Therefore, this study identifies the effect of air quality on the condition of covid-19 patients using the Naïve Bayes algorithm. The data was obtained through the corona.jakarta.go.id and Jakarta OpenData websites, after preprocessing 610 data were obtained which were divided into 80% training data and 20% testing data. The use of the Naïve Bayes Algorithm shows a high accuracy value of 82.73%. The results of identification based on the naïve bayes algorithm in four class, class very influential 2.3%, class 62.8% influential, class less effect 34.5%, and class has no effect 0.4%. Thus, air quality also affects the condition of covid-19 patients in the DKI Jakarta Province.

KeyWords: Covid-19; Air Quality; Naive Bayes.

This is an Open Access article - copyright on authors, distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY SA) (<http://creativecommons.org/licenses/by-sa/4.0/>)

How to Cite: M. D. Febriyanti *et al.*, "Identifikasi Pengaruh Kualitas Udara Terhadap Kondisi Pasien Covid-19 dengan Algoritma Naïve Bayes", *JIKO (JURNAL INFORMATIKA DAN KOMPUTER)*, Volume: 8, No.2, Pages 222–229, September 2024, doi: 10.26798/jiko.v8i2.867.

1. Pendahuluan

Pandemi covid-19 ditemukan pertama kali pada tahun 2020 di kota Wuhan, Cina. Di Indonesia kasus pertama kali terdeteksi pada awal maret 2020. Virus ini diketahui menyerang saluran pernapasan dengan gejala peradangan paru-paru (pneumonia), sehingga akan berdampak buruk bagi pasien yang memiliki riwayat masalah pernapasan. Berbagai upaya pemerintah untuk mengurangi penyebaran virus corona melalui proses vaksinasi. Masyarakat yang dapat terkena covid-19 tidak ada batasan usia, namun orang yang berusia lanjut dan atau mempunyai daya tahan tubuh yang lemah akan lebih rentan terpapar virus corona. Di sisi lain, manusia membutuhkan udara yang bersih untuk proses pernapasan, pada kondisi tertentu apabila suatu daerah yang mempunyai kualitas udara yang kurang baik tentunya akan berimplikasi pada saluran pernapasan bagi setiap penduduk di wilayah tersebut.

Penduduk yang berada di daerah dengan kualitas udara buruk perlu tentunya juga mempunyai risiko yang lebih tinggi bila dibandingkan dengan daerah yang mempunyai kualitas udara yang baik. Hal tersebut merupakan konsekuensi logis dengan lingkungan yang sehat, terkait dengan itu virus covid-19 langsung berimplikasi dengan saluran pernapasan, sehingga setiap orang yang tidak memiliki masalah medis tertentu juga dapat mempunyai risiko terkena covid-19.

Tabel 1. Penelitian Terkait Covid-19 dan Kualitas Udara

Peneliti	Judul Penelitian	Masalah Penelitian	Hasil Penelitian
D. Y. Liliansa, H. Maulana, & A. Setiawan	Data Mining untuk Prediksi Status Pasien Covid-19 dengan Pengklasifikasi Naive Bayes	Prediksi status pasien covid-19 untuk menentukan isolasi, sembuh dan meninggal	Probabilitas posterior data uji pasien menghasilkan prediksi status isolated karena $P(A0 x2) > P(A1 x2) > P(A2 x2)$ dan hasil akurasi 96,67% [1]
F. Fathonah, & Asti Herlina	Penerapan <i>Text Mining</i> Analisis Sentimen mengenai Vaksin Menggunakan Metode Naive Bayes	Mencari pendapat positif dan negatif mengenai vaksin covid-19	Naive Bayes dapat mengelompokkan data dengan optimal sehingga menghasilkan analisis sentimen terbaik dengan akurasi 100% [2].
Hasnan Afif, Kusri, & M.R. Arief	Prediksi Performa Mahasiswa Menggunakan Algoritma Naive Bayes Classifier	Melakukan uji coba untuk klasifikasi kelulusan mahasiswa menggunakan algoritma Naive Bayes	Pengolahan variabel terbaik dengan nilai akurasi 96,15% dan pengolahan seluruh variabel menghasilkan nilai akurasi 93,93% [3].
O. P. Barus, & Anton Tehja	Prediksi Kesembuhan Pasien Covid-19 di Indonesia Melalui Terapi Menggunakan Metode Naive Bayes	Mencari terapi yang tepat dan akurasi metode Naive Bayes dalam sukses/gagal untuk terapi pasien covid-19	Perhitungan presisi, sensitifitas dan akurasi menghasilkan prediksi yang sama dengan hasil aktualnya dari data testing menghasilkan tingkat akurasi 96,51% [4].
D. Pramana, & Mustakim	Prediksi Status Penanganan Pasien covid-19 dengan Algoritma Naive Bayes Classifier di Provinsi Riau	Mencari penanganan yang tepat untuk pasien covid-19 dengan Algoritma Naive Bayes Classifier	Klasifikasi penanganan pasien covid-19 memberikan hasil yang diukur dalam nilai presisi, recall dan akurasi yang memiliki nilai tinggi [5].
A.I. Sang, E. Sutoyo, & I. Darmawan	Analisis <i>Data Mining</i> untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma <i>Decision Tree</i> dan <i>Support Vector Machine</i>	Membandingkan <i>decision tree</i> dan <i>Support Vector Machine</i> untuk analisis kualitas udara di DKI Jakarta	Algoritma <i>Decision Tree</i> memiliki performa yang lebih baik dari algoritma SVM dalam menentukan klasifikasi kualitas udara di Provinsi DKI Jakarta [6].

Tabel 1 adalah penelitian terkait dengan covid-19 dan kualitas udara yang telah dilakukan sebelumnya. Setiap penelitian yang dilakukan lebih melihat bagaimana metode *Naive Bayes* dalam mengklasifikasi dan prediksi dengan kualitas udara maupun masalah covid-19, berbeda dengan penelitian yang dilakukan saat ini, di mana algoritma *Naive Bayes* digunakan untuk identifikasi apakah kualitas udara memberikan pengaruh terhadap kondisi pasien covid-19. Masalah ini dipandang perlu untuk dikaji lebih mendalam, sehingga dapat memberikan informasi bagi instansi terkait dalam proses penanggulangan dan upaya pen-

cegahan pada pasien yang terpapar dapat dipersiapkan dengan baik. Algoritma yang digunakan adalah *Naïve Bayes*, karena dapat memprediksi peluang di masa depan berdasarkan pengalamannya.

Selain itu mempunyai kemampuan memprediksi dengan atribut-atribut yang tidak saling berkaitan dari masing-masing kejadian [7]. Model *Cross-validation* (CV) digunakan untuk mengevaluasi kinerja algoritma, dimana data dipisahkan menjadi data proses pembelajaran (data training) dan data testing. Model atau algoritma dilatih oleh subset pembelajaran dan divalidasi [8].

Kualitas udara di beberapa daerah Indonesia memburuk, terutama Propinsi DKI Jakarta. Berdasarkan data DKI, menjadi propinsi di Indonesia yang mempunyai kasus covid-19 tertinggi dan juga sekaligus mempunyai kualitas udara yang kurang baik. Selain itu Propinsi DKI Jakarta menjadi kota tersibuk di Indonesia, tentunya aktifitas masyarakat juga sangat sibuk dan berlipikasi pada penggunaan kendaraan bermotor yang menjadi salah satu faktor yang menyebabkan menurunnya kualitas udara. Penelitian [6] melakukan kajian analisis data mining untuk klasifikasi data kualitas udara DKI Jakarta, menunjukkan bahwa kualitas udara berada dalam kategori 'tidak sehat'. Hal ini menjadi masalah tersendiri, bagi Propinsi DKI dalam penanganan wabah covid-19. Masalah ini yang mendorong penelitian ini dilakukan, sehingga dapat mengidentifikasi pengaruh kualitas udara terhadap kondisi pasien covid-19, secara khusus di DKI Jakarta. Harapannya penelitian ini dapat memberikan informasi bagi masyarakat maupun instansi terkait untuk dapat mengambil keputusan yang tepat dalam penanganan covid-19.

2. Metode Penelitian

2.1 Tempat, Waktu, Sampel dan Populasi Penelitian

Dataset yang digunakan adalah data kualitas udara dan pasien covid-19 yang berada di Jakarta. Data diambil dari situscorona.jakarta.go.id dan situs Jakarta open data atau dapat diakses melalui laman <https://data.jakarta.go.id/>. Data untuk pasien covid-19 digunakan dari bulan maret 2020 hingga desember 2021 dan data kualitas udara didapatkan sejumlah 611 data dengan rentang waktu dari bulan maret 2020 hingga oktober 2021. Populasi dalam penelitian ini adalah masyarakat yang bertempat tinggal di Jakarta dan kondisi kualitas udara. Sedangkan sample yang diambil untuk masyarakat Jakarta adalah orang-orang yang terdeteksi terjangkit covid-19 dan kualitas udara yang parameternya dapat diukur. Total sampel sebanyak 655 data pasien covid-19 dan 611 data kualitas udara.

2.2 Atribut Penelitian

Penelitian ini menggunakan lima belas atribut yang secara umum hanya terdapat dua variabel utama yaitu data pasien dan kualitas udara. Atribut secara lengkap diberikan pada Tabel 2. Fakta rill yang diperoleh adalah satuan berbeda dari dari kedua variabel utama. Satuan untuk variabel data pasien adalah orang, sedangkan untuk variabel kualitas udara mempunyai satuan $\mu\text{g}/\text{m}^3$. Sehingga proses ekuivalensi perlu dilakukan sehingga asas ke-lazim-an dari proses perhitungan dan interpretasi dapat dilakukan atau digunakan sebagaimana mestinya.

Tabel 2. Atribut Penelitian

No	Variabel	No	Variabel	No	Variabel
1	Meninggal	6	Positif Aktif	11	SO ₂
2	Sembuh	7	Tanpa Gejala	12	CO
3	Isolasi Mandiri	8	Bergejala	13	O ₃
4	Masih Perawatan	9	pm10	14	NO ₂
5	Total Pasien	10	pm25	15	MAX

Untuk proses ekuivalensi, setiap atribut kualitas udara dikonversi ke rata-rata kualitas udara per orang, sehingga menjadi sebanding dengan variabel data pasien dan layak untuk di gabung dalam proses penelitian. Rata-rata kualitas udara per orang diperoleh dari perbandingan antara total kualitas udara dengan jumlah penduduk pada wilayah tertentu. Secara umum dapat diberikan pada Persamaan (1).

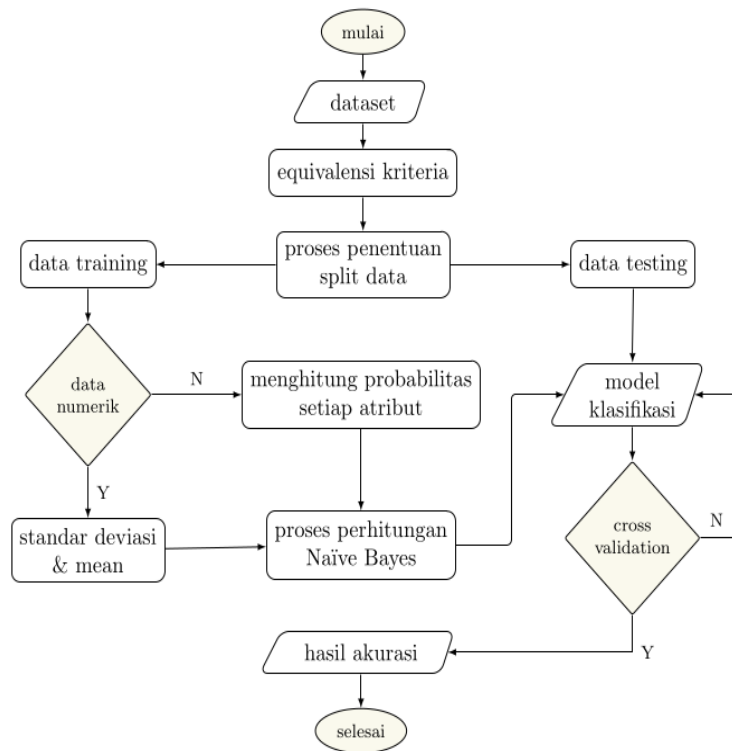
$$r = t/p \quad (1)$$

Sebagai contoh, jika pada suatu wilayah terdapat $1.000\mu\text{g}/\text{m}^3$ kualitas udara dan jumlah penduduknya 10.000 orang, maka rata-rata kualitas udara per orang dapat dihitung dengan cara: Rata-rata kualitas udara per orang = $1.000\mu\text{g}/\text{m}^3 / 10.000$ orang. Rata-rata kualitas udara per orang = $0,1\mu\text{g}/\text{m}^3/\text{orang}$. Sehingga rata-rata kualitas udara per orang adalah $r = 0,1\mu\text{g}/\text{m}^3/\text{orang}$.

2.3 Proses Penelitian

Proses penelitian untuk menentukan pengaruh kualitas udara terhadap kesembuhan pasien covid-19 menggunakan algoritma *Naïve Bayes* secara umum diberikan pada Gambar 1. Langkah awal adalah penentuan variabel, dan kriteria penelitian. Selanjutnya adalah pencarian data penelitian yang disesuaikan dengan kriteria yang telah ditentukan sebelumnya.

Equivalensi kriteria menjadi langkah yang penting, karena kedua variabel pasien covid-19 dan kualitas udara mempunyai satuan yang berbeda. Proses ini menyetarakan kedua satuan sehingga proses perhitungan menjadi relevan untuk dihitung. Persamaan (1) merupakan proses ekuivalensi di mana data setiap kriteria pada variabel kualitas udara disesuaikan dengan satuan setiap orang. Sehingga data kriteria pada variabel kualitas udara memiliki satuan $\mu\text{g}/\text{m}^3/\text{orang}$.



Gambar 1. Diagram Alur Penelitian

Dataset dibagi menjadi dua bagian dengan operator *split* data yaitu 80% menjadi data training dan 20% menjadi data testing. Baca data training dan proses menghitung nilai probabilitas, namun apabila data numerik maka: Hitung nilai *mean* dan standar deviasi dari masing-masing atribut. Cari nilai probabilitas dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut. Hitung Probabilitas setiap kelas. Hitung probabilitas posterior, kemudian ambil nilai probabilitas tertinggi untuk dijadikan hasil klasifikasi. Untuk mendapat estimasi yang baik pada akurasi model maka menggunakan *cross validation* dengan nilai iterasi 10. Menghitung kesesuaian hasil klasifikasi dengan data testing lalu hitung akurasi total.

3. Hasil dan Pembahasan

Penelitian ini menggunakan 655 data pasien covid-19 dan juga 611 data kualitas udara yang berada di Provinsi DKI Jakarta. Bagian sebelumnya telah dijelaskan proses ekuivalensi kriteria sehingga proses per-

hitungan menjadi relevan untuk dilakukan. Ekuivalensi dilakukan dengan setiap kriteria pada variabel kualitas udara diubah ke rata-rata kebutuhan setiap orang, seperti yang telah diberikan pada Persamaan (1). Sehingga untuk 611 data kualitas udara sudah disertakan dengan data pasien covid-19.

1. Nilai *Mean* dan Standar Deviasi.

Berdasarkan data analisis yang diambil pada kedua situs resmi corona.jakarta.go.id dan <https://data.jakarta.go.id/> terdapat beberapa data numerik maka terlebih dahulu mencari nilai mean dan standar deviasi. Hasil hitung nilai mean dapat dilihat pada Tabel 3.

Tabel 3. Hasil Perhitungan Nilai Mean

No	Variabel	Tidak Pengaruh	Kurang Pengaruh	Berpengaruh	Sangat Pengaruh
1	Meninggal	0,333333333	6704881	4056864	2172357
2	Sembuh	0	393942,7	225138,9	99288
3	Isolasi Mandiri	0	14131,33	5883632	6629071
4	Masih Perawatan	0,66666667	5825,9	2921081	2653286
5	Total Pasien	1	420604,8	238000,5	110742,7
6	Positif Aktif	0,66666667	19957,23	8804713	9282357
7	Tanpa Gejala	0	5494952	3596117	3434
8	Bergejala	0	8258619	4433436	4372
9	pm_{10}	34,33333333	70,77619	57,67363	31,85714
10	pm_{25}	0	71,61429	30,74151	0
11	SO_2	0	56,95714	45,86684	37,42857
12	CO	22,3333	20,45238	20,98172	31,21429
13	O_3	206,6667	76,61905	61,72324	36,21429
14	NO_2	10,33333	43,4381	31,78329	30,35714
15	MAX	206,6667	119,6857	79,40731	45,28571

Standar Deviasi atau sering disebut simpangan baku merupakan variasi sebaran data dalam sampel, semakin kecil nilai sebarannya berarti variasi nilai data semakin sama jika sebarannya bernilai 0, maka nilai semua datanya adalah sama. Hasil nilai simpangan baku (*standar deviasi*) dapat dilihat pada Tabel 4.

Tabel 4. Hasil Perhitungan *Standar Deviasi*

No	Variabel	Tidak Pengaruh	Kurang Pengaruh	Berpengaruh	Sangat Pengaruh
1	Meninggal	0,57735	4905021	4045367	1270943
2	Sembuh	0	319765,3	261028,8	65112,93
3	Isolasi Mandiri	0	21714,27	5652716	3639655
4	Masih Perawatan	1,154701	7311283	2047261	1651497
5	Total Pasien	1,732051	331771	265705,2	71124,71
6	Positif Aktif	1,154701	28659,56	7221056	4907495
7	Tanpa Gejala	0	8084583	5178818	2433999
8	Bergejala	0	12290,37	4451255	2922384
9	pm_{10}	21,36196	10,52736	13,48571	7,704715
10	pm_{25}	0	58,81777	40,3947	0
11	SO_2	0	23,88996	21,38798	10,61722
12	CO	19,39931	10,11939	10,94332	12,19192
13	O_3	6,027714	35,87535	19,17109	8,577059
14	NO_2	5,033223	32,69913	23,79675	16,02693
15	MAX	6,027714	16,31468	13,39634	2,785362

2. Probabilitas

Nilai probabilitas dari setiap kategori berdasarkan algoritma *Naïve Bayes* secara lengkap diberikan pada

Tabel 5, atau sering disebut peluang atau kemungkinan merupakan suatu nilai yang digunakan untuk mengukur tingkat terjadinya suatu kejadian yang acak.

Tabel 5. Probabilitas Kategori

No	Kategori	Nilai Probabilitas
1	Tidak Pengaruh	0,004918
2	Kurang Pengaruh	0,344262
3	Berpengaruh	0,627869
4	Sangat Pengaruh	0,022951

3. Menghitung Nilai *Distribusi Gaussian*

Distribusi Gaussian merupakan langkah terakhir untuk mengetahui hasil dari data training, atau sebuah model uji data untuk mengambil nilai dari peluang dari data training.

Tabel 6. Data Tes

No	Variabel	Data Uji	Tdk Pengaruh	Kurang Pengaruh	Berpengaruh	Sgt Pengaruh
1	Meninggal	99	0	0,002300630	0,003887643	0,002958423
2	Sembuh	58	0	0,000330464	0,00538542	0,000489629
3	Isolasi Mandiri	223	0	0,002205772	0,003214665	0,001405388
4	Masih Perawatan	691	0	0,003646810	0,004872729	0,004847477
5	Total Pasien	1071	0	0,0003114410	0,000520187	0,000455732
6	Positif Aktif	2	0,190658418	0,001849735	0,002233724	0,000952906
7	Tanpa Gejala	0	0	0,003522704	0,00435714	0,002989722
8	Bergejala	0	0	0,002872034	0,003642225	0,002410723
9	pm10	66	0,028775679	0,110959227	0,089805858	$7,82364 \times 10^{-6}$
10	pm25	0	0	0,024794441	0,046999594	0
11	SO2	28	0	0,039163370	0,060869709	0,082559533
12	CO	25	0,089747791	0,113393596	0,112763347	0,100362002
13	O3	83	$6,44434 \times 10^{-93}$	0,065577142	0,049229889	$4,7129 \times 10^{-8}$
14	NO2	7	0,142842442	0,037506493	0,047559403	0,034466018
15	MAX	83	$6,44434 \times 10^{-93}$	0,007884244	0,105174205	$3,6949 \times 10^{-41}$

Nilai "MAX" diperoleh sebesar 83 dengan memilih probabilitas posterior maksimum dari keempat kelas menghasilkan prediksi kategori kualitas udara "Berpengaruh" dengan nilai $2,70884 \times 10^{-33}$.

3.1 Evaluasi dan Validasi

Confusion Matrix digunakan untuk mengukur tingkat *accuracy*, *precision* dan *recall* yang memanfaatkan kinerja algoritma *Naïve Bayes* untuk mengolah dataset sebanyak 610 data, kemudian pada penelitian ini data dibagi menjadi dua dengan pembagian 8:2 yaitu menjadi data training sejumlah 487 data dan data testing sejumlah 124 data. Hasil secara lengkap diberikan pada Tabel 7.

Hasil akurasi pada Tabel 7 yang didapatkan dari implementasi RapidMiner digunakan untuk mengukur kinerja algoritma *Naïve Bayes*. Hasil analisa yang tercantum pada *confusion matrix* yaitu jumlah data yang diuji adalah 487 data, jumlah data yang di prediksi benar yaitu 393, dan jumlah data yang di prediksi salah yaitu 94. Nilai akurasi yang didapatkan yaitu 82,73% dengan standar Deviasi +/- 7,04%.

Nilai *recall* merupakan nilai yang menunjukkan tingkat keberhasilan untuk mengetahui banyaknya data yang sebenarnya positif atau merupakan tingkat keberhasilan yang diberikan oleh sistem dalam menemukan kembali sebuah informasi. Pada hasil *confusion matrix* pada penelitian ini nilai *recall* yang dihasilkan adalah 92,48%. Nilai *precision* merupakan nilai yang menunjukkan tingkat ketepatan prediksi dalam mengenali data sesuai dengan kelas aslinya. Pada penelitian ini nilai dari precision adalah 82,99%.

Tabel 7. Hasil Confusion Matrix

	tdk pengaruh	kurang pengaruh	pengaruh	sangat pengaruh	class precision
pred. tidak pengaruh	0	0	16	10	0.00%
pred. kurang pengaruh	0	57	283	1	92.44%
pred. berpengaruh	2	110	7	0	82.99%
pred. sangat pengaruh	0	1	0	0	38.46%
class recall	0.00%	65.48%	92.48%	90.91%	

Berdasarkan algoritma *naïve bayes* diperoleh model distribusi untuk setiap label kategori diperoleh hasil dari empat class hasil pengolahan dataset pada penelitian ini menunjukkan nilai class sangat pengaruh 2,3%, class berpengaruh 62,8%, class kurang berpengaruh 34,5%, dan class tidak berpengaruh 0,4%. Hasil ini menunjukkan bahwa kuliatas udah ikut mempengaruhi kondisi pasien covid-19 di wilayah Propinsi DKI Jakarta.

4. Simpulan

Penggunaan data mulai dari bulan Maret 2020 - Oktober 2021, model ini dibangun menggunakan dataset dengan berbagai variabel berjumlah 655 data, dan proses pre-processing diperoleh 610 data. Berdasarkan dari hasil penelitian dengan algoritma Naïve Bayes diperoleh hasil akurasi 82,73%, nilai recall 92,48%, dan nilai precision 82,99%. Sehingga dapat disimpulkan bahwa algoritma Naïve Bayes dapat diterapkan untuk mengklasifikasi seberapa besar pengaruh kualitas udara terhadap kondisi pasien covid-19 di Jakarta.

Hasil identifikasi berdasarkan algoritma Naïve Bayes dalam empat class, yang secara berturut-turut diperoleh hasil class sangat pengaruh 2,3%, class berpengaruh 62,8%, class kurang berpengaruh 34,5%, dan class tidak berpengaruh 0,4%. Dengan demikian kualitas udara ikut mempengaruhi kondisi pasien covid-19 di wilayah Propinsi DKI Jakarta. Hasil dari penelitian ini juga memberikan informasi pada masysrakat, Pemerintah atau Instansi terkait untuk berusaha menjaga kualitas udara.

Pustaka

- [1] D. Y. Liliana, H. Maulana, and A. Setiawan, "Data mining untuk prediksi status pasien covid-19 dengan pengklasifikasi naïve bayes," *Multinetics*, pp. 48–53, 2021. [Online]. Available: <https://doi.org/10.32722/multinetics.v7i1.3786>.
- [2] F. Fathonah and A. Herliana, "Penerapan text mining analisis sentimen mengenai vaksin covid - 19 menggunakan metode naïve bayes," *Jurnal Sains dan Informatika*, pp. 155–164, 2021. [Online]. Available: <https://doi.org/10.34128/jsi.v7i2.331>.
- [3] H. Afif, Kusriani, and M. R. Arief, "Prediksi performa mahasiswa menggunakan algoritma naive bayes classifier," *Teknomatika*, vol. 11, no. 2, pp. 47–57, 2019.
- [4] O. P. Barus and A. Tehja, "Prediksi kesembuhan pasien covid-19 di indonesia melalui terapi menggunakan metode naïve bayes," *Journal Information System Development (ISD)*, vol. 6, no. 2, pp. 59–66, 2021. [Online]. Available: <https://ejournal.medan.uph.edu/index.php/isd/article/view/460.D>.
- [5] D. Pramana and Mustakim, "Prediksi status penanganan pasien covid-19 dengan algoritma naïve bayes classifier di provinsi riau," *Jurnal Sistem Komputer dan Informatika (JSON)*, pp. 202–208, 2021. [Online]. Available: <https://doi.org/10.30865/json.v3i2.3570>.
- [6] A. I. Sang, E. Sutoyo, and I. Darmawan, "Analisis data mining untuk klasifikasi data kualitas udara dki jakarta menggunakan algoritma decision tree dan support vector machine," *e-Proceeding of Engineering*, vol. 8, no. 5, pp. 8954–8963, 2021.
- [7] R. Amalia, "Penerapan data mining untuk memprediksi hasil kelulusan siswa menggunakan metode naive bayes," *Jurnal Informatika Dan Sistem Informasi*, vol. 6(1), pp. 33–42, 2020.

- [8] F. Tempola, M. Muhammad, and A. Khairan, "Perbandingan klasifikasi antara knn dan naive bayes pada penentuan status gunung berapi dengan k-fold cross validation," *Jurnal Teknologi Informasi dan Ilmu Komputer*, p. 577, 2018. [Online]. Available: <https://doi.org/10.25126/jtiik.201855983>.
- [9] H. Henderi and R. L. Wanda, "Preprocessing data untuk sistem peramalan tingkat kedisiplinan mahasiswa," *ICIT Journal*, pp. 296–308, 2017. [Online]. Available: <https://doi.org/10.33050/icit.v3i2.70>.