

Identifikasi Pengaruh Kualitas Udara Terhadap Pasien Covid-19 dengan Algoritma Naïve Bayes, K-NN dan SVM (Kasus Kota Jakarta dan Singapura)

Andrian Roland Kenedy Pakpahan¹, Alz Dany Wowor²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Indonesia

E-mail: andrianp920@gmail.com¹, alzdanny.wowor@uksw.edu²

Abstract

The COVID-19 pandemic has emerged as a significant global outbreak, seriously impacting public health. This study examines the influence of air quality on COVID-19 patients in Singapore and Jakarta, utilizing Naive Bayes, Support Vector Machine (SVM), and K-Nearest Neighbor (K-NN) algorithms for data analysis. The findings indicate that better air quality in Singapore contributes to a reduction in disease severity, with 100% accuracy in predicting positive cases and hospitalizations using the SVM algorithm. Additionally, Naive Bayes successfully predicted 100% of mortality cases at a 90/10 ratio, demonstrating that even with relatively good air quality, its impact on patient severity remains significant. Conversely, in Jakarta, where air quality is poorer, there is variability in accuracy results. The K-NN algorithm achieved 100% accuracy for positive cases and isolation, while the SVM only reached 46.15% accuracy for positive case predictions. These findings underscore the significant influence of air quality on the severity and mortality of COVID-19 patients. In countries with better air quality, the impact on severity appears more manageable, whereas in those with poorer air quality, patient severity tends to be higher. This research aims to encourage the Indonesian public to pay closer attention to air quality for public health.

Keywords: Air Quality, Covid-19, Naïve Bayes, Support Vector Machine, k-NN.

Abstrak

Pandemi COVID-19 telah menjadi salah satu wabah global yang signifikan, berdampak serius terhadap kesehatan masyarakat. Penelitian ini mengkaji pengaruh kualitas udara terhadap pasien COVID-19 di Singapura dan Jakarta, menggunakan algoritma Naive Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbor (K-NN). Hasil penelitian menunjukkan bahwa kualitas udara yang lebih baik di Singapura berkontribusi pada pengurangan keparahan penyakit, dengan akurasi 100% dalam memprediksi kasus positif dan perawatan menggunakan SVM. Naive Bayes juga berhasil memprediksi 100% kasus kematian pada rasio 90/10, menunjukkan bahwa meskipun kualitas udara relatif baik, pengaruhnya terhadap keparahan pasien tetap signifikan. Sebaliknya, di Jakarta, yang memiliki kualitas udara lebih buruk, terdapat variasi hasil akurasi. Algoritma K-NN mencapai 100% akurasi pada kasus positif dan isolasi, sementara SVM hanya mencapai 46.15% untuk prediksi kasus positif. Temuan ini menggarisbawahi pengaruh signifikan kualitas udara terhadap tingkat keparahan dan kematian pasien COVID-19. Di negara dengan kualitas udara yang lebih baik, dampak keparahan terlihat lebih terkendali, sedangkan di negara dengan kualitas udara buruk, keparahan pasien cenderung lebih tinggi. Penelitian ini diharapkan mendorong masyarakat Indonesia untuk memperhatikan kualitas udara demi kesehatan masyarakat.

Kata Kunci: Kualitas Udara, Covid-19, Naïve Bayes, Support Vector Machine, k-NN.

1. Pendahuluan

Pandemi COVID-19 yang dimulai pada akhir tahun 2019 telah memberikan dampak besar pada seluruh lapisan masyarakat global [1]. Kasus covid-19 menyebar dengan sangat cepat [2], hingga januari 27 Januari 2020 pemerintah china mengatakan bahwa pneumonia adalah tipe baru dari corona virus atau COVID-19 [3]. Dalam menghadapi pandemi ini, negara-negara di dunia, termasuk dua negara di Asia Tenggara, Indonesia dan Singapura, menghadapi tantangan serupa namun memiliki pilihan dan strategi yang berbeda [4]. Singapura, sebuah negara kota kecil dengan koordinat geografis sekitar 1,3521° LU dan 103,8198° BT, terletak di ujung selatan Semenanjung Malaysia. Negara ini dipisahkan dari Indonesia pada sekitar 1° LU–11° LS dan 95°–141° BB oleh Selat Singapura, yang lebarnya kurang dari 50 km pada titik terdekatnya. Indonesia, negara kepulauan terbesar di dunia, terdiri dari lebih dari 17.000 pulau yang membentang dari Sabang di barat hingga Merauke di timur [5].

Salah satu faktor yang dapat memengaruhi respons pandemi adalah kualitas udara [6][7], sebagai tanggapan terhadap ancaman polusi udara, yang membahayakan kesehatan populasi di seluruh dunia, diperlukan intervensi dari berbagai bidang, termasuk kesehatan masyarakat [8], indeks standar yang digunakan oleh pencemaran udara (ISPU) adalah Partikulat (PM10 dan PM2.5), karbon monoksida (CO), sulfur dioksida (SO₂), nitrogen dioksida (NO₂), Ozon (O₃), dan hidrokarbon (HC) [9]. Oleh karena itu, penelitian ini menggunakan metode Naive Bayes, Support Vector Machines (SVM), dan K-Nearest Neighbors (KNN) untuk menganalisis pengaruh kualitas udara terhadap kondisi pasien COVID-19 di kedua negara tersebut. Penelitian ini akan menyelidiki bagaimana variasi kualitas udara dapat berdampak pada tingkat keparahan dan prognosis pasien COVID-19 dan memberikan informasi yang diharapkan dapat mengarah pada pengembangan kebijakan kesehatan masyarakat yang lebih efektif di masa depan [10].

Menurut penelitian sebelumnya, Briandi Try Putra melakukan penelitian terkait “Analisis Tingkat Akurasi Prediksi Gejala COVID-19 Dengan Menggunakan Metode *Logistic Regression* dan *Support Vector Machine*” diperoleh hasil penelitian yang menunjukkan bahwa *Logistic Regression* dan *Support Vector Machine* memiliki akurasi tinggi dalam memprediksi gejala COVID-19 [11]. penelitian yang juga dilakukan oleh Dewi Y. Liliana yang meneliti tentang “Data Mining untuk Prediksi status pasien covid-19 dengan Pengklasifikasian Naive Bayes” menemukan hasil perhitungan probabilitas posterior dan data uji pasien menghasilkan prediksi status isolated karena nilai $P(A_0|x_2) > P(A_1|x_2) > P(A_2|x_2)$ dan hasil akurasi 96,67% (dengan tools RapidMiner) [12], penelitian lainnya yang dilakukan oleh Adinda I. Sang tentang “Analisis data mining untuk klasifikasi data kualitas udara DKI Jakarta menggunakan algoritma *Decision tree* dan *Support Vector Machine*” menemukan model *decision tree* memiliki performa yang lebih baik dibandingkan dengan SVM [13].

Selain itu, studi ini melihat bagaimana polusi udara berdampak pada kemampuan masyarakat untuk mengendalikan wabah penyakit dalam jangka panjang. Selama pandemi COVID-19, penurunan aktivitas industri dan transportasi telah mengubah kualitas udara di beberapa kota besar di seluruh dunia, termasuk Singapura dan Indonesia. Hal ini memberikan peluang langka untuk melacak perubahan kualitas udara dan dampaknya terhadap intensitas dan penyebaran COVID-19 [14]. Penelitian ini bertujuan untuk menemukan hubungan terkait pengaruh kualitas udara terhadap pasien covid-19 terkhususnya untuk kota jakarta dan singapura, serta untuk digunakan sebagai rekomendasi untuk kebijakan kesehatan masyarakat di masa depan yang berkaitan dengan pengelolaan kualitas udara berkelanjutan dan pandemi.

2. Metodologi Penelitian

2.1. Tempat, Waktu, Sampel, Populasi dan Atribut Penelitian

Penelitian ini dilakukan dengan mengumpulkan data melalui metode penelusuran literatur secara online. Sumber data yang digunakan termasuk informasi tentang kualitas

udara dan kasus COVID-19 di Jakarta, Indonesia, dan Singapura. Data mengenai kasus COVID-19 di Jakarta diakses melalui situs web resmi <https://data.jakarta.go.id/>, sementara informasi tentang kualitas udara dan COVID-19 di Singapura diperoleh dari <https://aqicn.org/>. Semua data tersedia dalam format CSV yang dapat diedit dan dianalisis menggunakan Microsoft Excel. Rentang waktu kasus COVID-19 dari tahun 2020 hingga 2022, sedangkan data kualitas udara untuk Singapura mencakup periode dari tahun 2020 hingga 2022. Populasi yang diteliti adalah penduduk Jakarta dan Singapura serta kondisi udara di wilayah tersebut. Sampel untuk penelitian di Jakarta adalah individu yang terkonfirmasi positif COVID-19 dan data kualitas udara yang terukur, dengan total sampel sebanyak 1217 kasus COVID-19 dan 1217 data kualitas udara. Sedangkan untuk Singapura, jumlah sampel adalah sebanyak 717 kasus COVID-19 dan 717 data kualitas udara.

Tabel 1. Atribut Covid-19 yang digunakan pada Singapura dan Kota Jakarta

No	Variabel	Detail Penggunaan
1	Meninggal	Digunakan Atribut
2	Isolasi	Digunakan Atribut
3	Dirumah Sakit	Digunakan Atribut
4	Positif	Digunakan Atribut

Pada Tabel 1 merupakan Kategori yang diperoleh dari masing-masing data yang telah dilakukan cleaning sehingga dapat memenuhi kriteria dan dapat digunakan untuk melakukan penelitian ini.

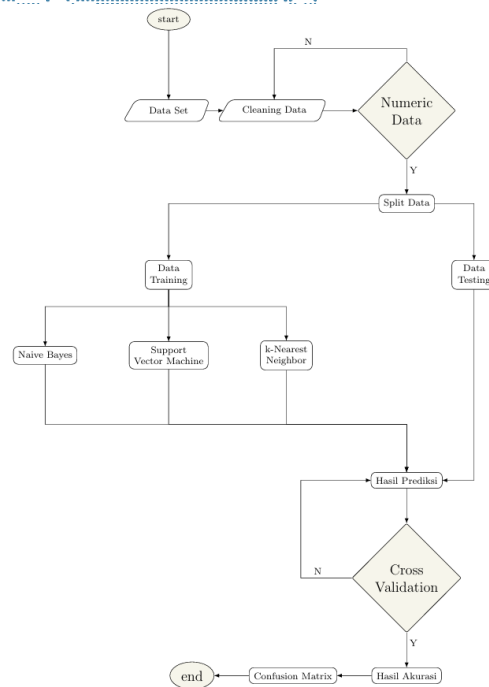
Tabel 2. Atribut Kualitas Udara yang digunakan pada Singapura dan Kota Jakarta

No	Variabel	Detail Penggunaan
1	PM2.5	Digunakan Atribut
2	PM10	Digunakan Atribut
3	SO ₂	Digunakan Atribut
4	CO	Digunakan Atribut
5	O ₃	Digunakan Atribut
6	NO ₃	Digunakan Atribut

Pada Tabel 2 diperlihatkan data Kualitas Udara pada setiap lokasi yang sudah di pilih untuk dapat digunakan sebagai variabel penelitian pada penelitian ini.

2.2. Schema Penelitian

Gambar 1 adalah schema dari penelitian ini, schema penelitian ini dilakukan menggunakan tools Rappid Miner. Data yang sudah di dapatkan kemudian dilakukan cleaning data untuk memisahkan data angka dan data yang bukan angka, kemudian dilakukan split data untuk membagi antara data training dan data testing, data training dan data testing yang digunakan pada penelitian ini antara lain 70/30,75/25,80/20,85/15, dan 90/10. Data training dan data testing yang sudah dibagi kemudian dilakukan pengujian menggunakan algoritma yang sudah dipilih antara lain adalah Naive Bayes, Support Vector Machine dan k-NN yang kemudian menghasilkan hasil prediksi, kemudian dilakukan Cross Validation untuk mencegah kesalahan dalam melakukan penelitian, sehingga mendapatkan hasil akurasi dan Confusion Matrix.



Gambar 1. Schema Penelitian

3. Hasil dan Pembahasan

3.1. Implementasi Algoritma

Pada tahapan ini, diberikan hasil penelitian perbandingan dari 3 model yang sudah di pilih, klasifikasi ini dilakukan pada dua kota yang berbeda, yaitu Singapura dan Kota Jakarta, Pengujian ini dilakukan menggunakan perangkat lunak RappidMiner untuk membantu pengolahan data yang sudah di peroleh. Dalam proses penelitian, pengolahan data yang dilakukan menggunakan RappidMiner menggunakan operator split data untuk membagi data ke dalam data testing dan training, data testing dan training dibagi menjadi beberapa rasio yaitu 70:30, 80:20, dan 90:10. Data yang sudah dikumpulkan kemudian disajikan dalam bentuk tabel dan grafik untuk memudahkan interpretasi dan analisis.

3.1.1. Implementasi Algoritma Naïve Bayes

Pada Tabel 3, terlihat hasil akurasi dari penerapan algoritma pada data dari Singapura dengan rasio data 70%:30%, 75%:25%, 80%:20%, 85%:15%, dan 90%:10%. Atribut 'Positif' mencapai akurasi tertinggi pada rasio 90%:10% dengan nilai 94,44%, sementara nilai terendah berada pada rasio 75%:25% sebesar 93,02%. Atribut 'Meninggal' mendapatkan akurasi tertinggi, dengan nilai sempurna 100% pada rasio 90%:10%, dan nilai terendah tercatat pada rasio 80%:20% sebesar 97,20%. Untuk atribut 'Isolasi', akurasi tertinggi terdapat pada rasio 70%:30% dengan nilai 94,88%, sedangkan akurasi terendah terlihat pada rasio 90%:10% sebesar 90,14%. Atribut 'Dirumah sakit' menunjukkan akurasi tertinggi pada rasio 85%:15% dengan nilai 98,60%, dan terendah pada rasio 90%:10% sebesar 95,83%.

Tabel 3. Hasil Akurasi Data Singapura

No	Atribut	70/30	75/25	80/20	85/15	90/10
1	Positif	94.42%	93.02%	93.71%	94.41%	94.44%
2	Meninggal	97.69%	97.69%	97.20%	99.30%	100.00%
3	Isolasi	94.88%	91.63%	90.97%	92.36%	90.14%
4	Dirumah Sakit	96.28%	96.28%	97.90%	98.60%	95.83%

Pada Tabel 4, terlihat hasil akurasi dari penerapan algoritma pada data dari Indonesia dengan rasio data 70%:30%, 75%:25%, 80%:20%, 85%:15%, dan 90%:10%. Atribut 'Positif' menunjukkan variasi akurasi, dengan nilai tertinggi pada rasio 80%:20% sebesar 96,30%, dan 9 nilai terendah pada rasio 85%:15% sebesar 81,48%. Atribut 'Meninggal' mendapatkan nilai akurasi sempurna 100% pada semua rasio data. Sementara itu, atribut 'Isolasi' mencapai akurasi tertinggi pada rasio 80%:20% dengan nilai 85,71%, dan nilai terendah pada rasio 75%:25% sebesar 73,17%. Atribut 'Dirumah sakit' memiliki akurasi tertinggi pada rasio 70%:30% sebesar 80,00%, dengan akurasi terendah pada rasio 90%:10% sebesar 57,14%.

Tabel 4. Hasil Akurasi Data Indonesia

No	Atribut	70/30	75/25	80/20	85/15	90/10
1	Positif	90.00%	82.50%	96.30%	81.48%	84.62%
2	Meninggal	100.00%	100.00%	100.00%	100.00%	100.00%
3	Isolasi	75.61%	73.17%	85.71%	82.14%	84.62%
4	Dirumah Sakit	80.00%	75.00%	77.78%	70.37%	57.14%

3.1.2. Implementasi Algoritma K-NN

Pada Tabel 5, terlihat hasil akurasi dari penerapan algoritma pada data dari Singapura dengan rasio data 70%:30%, 75%:25%, 80%:20%, 85%:15%, dan 90%:10%. Atribut 'Positif' mencapai akurasi tertinggi pada rasio 75%:25% dengan nilai 95.53%, sementara akurasi terendah tercatat pada rasio 90%:10% sebesar 93.06%. Atribut 'Meninggal' menunjukkan hasil akurasi tertinggi pada rasio 70%:30% sebesar 84.72%, dan akurasi terendah pada rasio 90%:10% sebesar 81.94%. Untuk atribut 'Isolasi', akurasi sempurna 100.00% tercatat pada rasio 70%:30% dan 90%:10%, sementara akurasi terendah terjadi pada rasio 75%:25% sebesar 98.31%. Atribut 'Dirumah sakit' memiliki akurasi tertinggi pada rasio 80%:20% dengan nilai 99.30%, dan akurasi terendah pada rasio 85%:15% sebesar 97.17%.

Tabel 5. Hasil Akurasi Data Singapura

No	Atribut	70/30	75/25	80/20	85/15	90/10
1	Positif	95.35%	95.53%	93.01%	95.33%	93.06%
2	Meninggal	84.72%	82.02%	82.52%	83.18%	81.94%
3	Isolasi	100.00%	98.31%	99.31%	99.07%	100.00%
4	Dirumah Sakit	99.07%	98.88%	99.30%	97.17%	98.61%

Pada Tabel 6, terlihat hasil akurasi dari penerapan algoritma pada data dari Indonesia dengan rasio data 70%:30%, 75%:25%, 80%:20%, 85%:15%, dan 90%:10%. Atribut 'Positif' menunjukkan akurasi sempurna sebesar 100.00% pada hampir semua rasio, kecuali pada rasio 75%:25%, di mana akurasi sedikit menurun menjadi 97.06%. Atribut 'Meninggal' mencapai akurasi sempurna 100.00% di semua rasio data. Atribut 'Isolasi' menunjukkan akurasi tertinggi pada rasio 90%:10% dengan nilai 100.00%, dan terendah pada rasio 75%:25% sebesar 82.35%. Sementara itu, atribut 'Dirumah sakit' memiliki akurasi tertinggi pada rasio 70%:30% dengan nilai 85.00%, dan terendah pada rasio 75%:25% sebesar 75.76%.

Tabel 6. Hasil Akurasi Data Indonesia

No	Atribut	70/30	75/25	80/20	85/15	90/10
1	Positif	100.00%	97.06%	100.00%	100.00%	100.00%
2	Meninggal	100.00%	100.00%	100.00%	100.00%	100.00%
3	Isolasi	92.68%	82.35%	96.43%	95.24%	100.00%
4	Dirumah Sakit	85.00%	75.76%	77.78%	75.00%	78.57%

3.1.3. Implementasi Algoritma SVM

Pada Tabel 7, terlihat hasil akurasi dari penerapan algoritma pada data Indonesia dengan rasio data 70%:30%, 75%:25%, 80%:20%, 85%:15%, dan 90%:10%. Atribut 'Positif' menunjukkan akurasi tertinggi sebesar 46.15% pada rasio 90%:10% dan terendah 42.86% pada rasio 85%:15%. Atribut 'Meninggal' mencapai akurasi tertinggi 97.50% pada rasio 70%:30% dan terendah 92.86% pada rasio 90%:10%. Atribut 'Isolasi' menunjukkan akurasi tertinggi pada rasio 85%:15% dengan nilai 66.67%, dan terendah pada rasio 70%:30% sebesar 48.78%. Sementara itu, atribut 'Dirumah sakit' memiliki akurasi tertinggi pada rasio 85%:15% dengan nilai 85.00%, dan terendah pada rasio 80%:20% sebesar 62.96%.

Tabel 7. Hasil Akurasi Data Singapura

No	Atribut	70/30	75/25	80/20	85/15	90/10
1	Positif	97.21%	93.85%	95.80%	99.07%	98.61%
2	Meninggal	82.87%	83.71%	83.92%	81.31%	84.72%
3	Isolasi	76.74%	79.78%	79.86%	78.50%	85.92%
4	Dirumah Sakit	98.14%	100.00%	99.30%	98.11%	100.00%

Pada Tabel 8, terlihat hasil akurasi dari penerapan algoritma pada data dari Indonesia dengan berbagai rasio data. Atribut 'Meninggal' memiliki akurasi tertinggi di semua rasio data, namun menunjukkan sedikit penurunan akurasi seiring dengan peningkatan proporsi data latih. 'Isolasi' dan 'Dirumah Sakit' menunjukkan peningkatan akurasi yang cukup signifikan seiring bertambahnya proporsi data latih. 'Positif' menunjukkan sedikit fluktuasi akurasi, dengan penurunan pada 80%:20% dan kemudian peningkatan pada 90%:10%.

Tabel 8. Hasil Akurasi Data Indonesia

No	Atribut	70/30	75/25	80/20	85/15	90/10
1	Positif	45.00%	44.12%	44.44%	42.86%	46.15%
2	Meninggal	97.50%	97.06%	96.43%	95.00%	92.86%
3	Isolasi	48.78%	52.94%	53.57%	66.67%	61.54%
4	Dirumah Sakit	75.00%	69.70%	62.96%	85.00%	71.43%

3.2. Tabel Confusion Matrix

Tabel 9. Menyajikan Hasil Confusion Matrix model Naive Bayes dalam mengklasifikasikan data terkait "meninggal". Model ini menunjukkan performa yang sangat baik dengan presisi sempurna 100,00% pada kelas "SR" dan "R". Kelas "S" memiliki presisi sebesar 72,73%, sementara kelas "T" menunjukkan presisi 83,33%. Secara keseluruhan, recall tertinggi terdapat pada kelas "SR" sebesar 100,00%, diikuti oleh kelas "R" dengan 96,30%, kelas "S" dengan 88,89%, dan kelas "T" dengan recall sebesar 83,33%.

Tabel 9. Confusion Matrix Meninggal Singapura k-NN

	true SR	true R	true S	true T	true ST	class precision
pred. SR	167	0	0	0	0	100.00%
pred. R	0	26	0	0	0	100.00%
pred. S	0	1	8	2	0	72.73%
pred. T	0	0	1	10	1	83.33%
pred. ST	0	0	0	0	0	0.00%
class recall	100.00%	96.30%	88.89%	83.33%	0.00%	

Tabel 10 menyajikan hasil evaluasi model K-NN dalam mengklasifikasikan data terkait "Meninggal" di Indonesia. Model ini menunjukkan performa yang baik dengan presisi tertinggi sebesar 85,57% untuk kelas "SR". Kelas lainnya memiliki presisi yang bervariasi, dengan nilai antara 33,33% hingga 84,62%. Secara keseluruhan, model K-NN

menunjukkan kemampuan yang cukup baik dalam membedakan kelas-kelas pada data terkait "Meninggal" di Indonesia, dengan rata-rata recall sebesar 50,87%.

Tabel 10. Confusion Matrix Meninggal Singapura k-NN

	true SR	true R	true S	true T	true ST	class precision
pred. SR	166	26	2	0	0	85.57%
pred. R	1	1	1	0	0	33.33%
pred. S	0	0	5	1	0	83.33%
pred. T	0	0	1	11	1	84.62%
pred. ST	0	0	0	0	0	0.00%
class recall	99.40%	3.70%	55.56%	91.67%	0.00%	

Tabel 11 menyajikan hasil evaluasi model SVM dalam mengklasifikasikan data terkait "Meninggal" di Indonesia. Model ini menunjukkan performa yang cukup baik dengan presisi sebesar 85,20% untuk kelas "SR". Kelas lainnya memiliki presisi yang lebih rendah, berkisar antara 50,00% hingga 83,33%. Secara keseluruhan, model SVM menunjukkan kemampuan yang cukup baik dalam membedakan kelas-kelas pada data terkait "Meninggal" di Indonesia, dengan rata-rata recall sebesar 41,91%.

Tabel 11. Confusion Matrix Meninggal Singapura k-NN

	true SR	true R	true S	true T	true ST	class precision
pred. SR	166	26	2	0	0	85.57%
pred. R	1	1	1	0	0	33.33%
pred. S	0	0	5	1	0	83.33%
pred. T	0	0	1	11	1	84.62%
pred. ST	0	0	0	0	0	0.00%
class recall	99.40%	3.70%	55.56%	91.67%	0.00%	

Tabel 12 menyajikan hasil evaluasi model NaiveBayes dalam mengklasifikasikan data terkait "Meninggal" di Indonesia. Model ini menunjukkan performa yang sempurna dengan presisi 100,00% untuk kelas "SR" dan "T". Secara keseluruhan model Naive Bayes menunjukkan kemampuan yang baik dalam membedakan kelas-kelas pada data terkait "Meninggal" di Indonesia, dengan rata-rata recall 66,67%.

Tabel 12. Confusion Matrix meniggal Indoensia Naive Bayes

	true SR	true T	true ST	class precision
pred. SR	38	0	0	100.00%
pred. T	0	2	0	100.00%
pred. ST	0	0	0	0.00%
class recall	100.00%	100.00%	0.00%	

Tabel 13 menyajikan hasil evaluasi model K-Nearest Neighbors (KNN) dalam mengklasifikasikan data terkait "Meninggal" di Indonesia. Model ini menunjukkan performa yang sempurna dengan presisi 100,00% untuk kelas "SR" dan "T". Secara keseluruhan model KNN menunjukkan kemampuan yang baik dalam membedakan kelas-kelas pada data terkait "Meninggal" di Indonesia, dengan rata-rata recall 66,67%.

Tabel 13. Confusion Matrix meninggal Indonesia k-NN

	true SR	true T	true ST	class precision
pred. SR	38	0	0	100.00%
pred. T	0	2	0	100.00%
pred. ST	0	0	0	0.00%
class recall	100.00%	100.00%	0.00%	

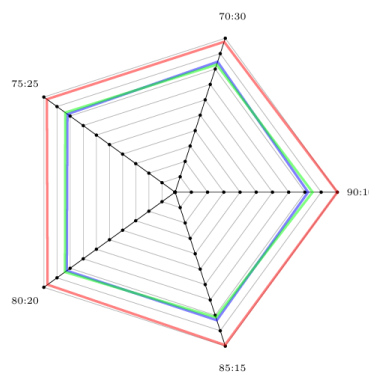
Tabel 14 menyajikan hasil evaluasi model Support Vector Machine (SVM) dalam mengklasifikasikan data terkait "Meninggal" di Indonesia. Model ini menunjukkan performa yang sangat baik dengan presisi 97,44% untuk kelas "SR" dan presisi sempurna 100,00% untuk kelas "T". Secara keseluruhan model SVM menunjukkan kemampuan yang baik dalam membedakan kelas-kelas pada data terkait "Meninggal" di Indonesia, dengan rata-rata recall 50,00%.

Tabel 14. Confusion Matrix meninggal Indonesia SVM

	true SR	true T	true ST	class precision
pred. SR	38	1	0	97.44%
pred. T	0	1	0	100.00%
pred. ST	0	0	0	0.00%
class recall	100.00%	50.00%	0.00%	

3.3. Komparasi Algoritma

Pada Gambar 2, terlihat perbandingan kinerja tiga algoritma dalam memprediksi hubungan antara kualitas udara dan kesembuhan di Singapura. Naive Bayes (garis merah) menunjukkan performa terbaik dengan akurasi yang konsisten tinggi, mencapai 100% pada rasio 90:10. K-Nearest Neighbors (K-NN, garis biru) menunjukkan variasi akurasi antara 81% hingga 85%, sementara Support Vector Machine (SVM, garis hijau) stabil di sekitar 82% hingga 84%.



Gambar 2. Komparasi Algoritma Meninggal Singapura

Pada Gambar 3 menunjukkan perbandingan akurasi dari tiga algoritma dalam memprediksi hubungan antara variabel yang berkaitan dengan kematian di Indonesia. Pada grafik ini, algoritma Naive Bayes (garis merah) menunjukkan performa yang sangat baik, dengan akurasi yang mencapai 100% pada semua rasio data yang diuji (70:30, 75:25, 80:20, 85:15, dan 90:10). Di sisi lain, algoritma K-Nearest Neighbors (K-NN, garis biru) juga menunjukkan akurasi sempurna sebesar 100% pada rasio-rasio yang sama. Namun, pada rasio 80:20, ada sedikit penurunan performa, meskipun tetap berada pada tingkat yang sangat tinggi. Sementara itu, algoritma Support Vector Machine (SVM, garis hijau) menunjukkan hasil yang sedikit lebih rendah dengan akurasi berkisar antara 92.86% hingga 97.50%.

Berdasarkan hasil pengujian terhadap dua negara dengan kondisi geografis dan demografis yang berbeda, yaitu Singapura (Asia Tenggara) dan Indonesia (Asia Selatan), untuk melihat hubungan antara kualitas udara dengan tingkat keparahan pasien Covid-19, digunakan beberapa algoritma klasifikasi seperti Naive Bayes, K-Nearest Neighbor (k-NN), dan Support Vector Machine (SVM). Pengaruh kualitas udara di Singapura, yang dikenal memiliki kualitas udara lebih baik dibandingkan Indonesia, memperlihatkan hasil yang lebih konsisten pada tingkat keparahan pasien. Ini terlihat dari akurasi model yang tinggi pada kasus pasien "Positif" dan "Dirumah Sakit" menggunakan algoritma SVM dengan akurasi mencapai 100%, serta Naive Bayes yang berhasil memprediksi 100%

kasus "Meninggal" pada rasio 90/10. Hal ini mengindikasikan bahwa kualitas udara di Singapura tetap mempengaruhi tingkat keparahan pasien, meskipun kualitas udaranya relatif lebih baik. Sebaliknya, di Indonesia yang memiliki kualitas udara lebih rendah, terdapat variasi hasil akurasi di mana algoritma k-NN menunjukkan akurasi terbaik pada sebagian besar atribut, seperti "Positif" dan "Isolasi" dengan akurasi mencapai 100%. Namun, algoritma SVM memberikan hasil yang lebih rendah terutama pada prediksi kasus "Positif", dengan akurasi hanya 46.15%. Hal ini menunjukkan bahwa kualitas udara yang buruk di Indonesia memiliki dampak signifikan terhadap tingkat keparahan pasien Covid-19, terutama pada pasien yang membutuhkan perawatan intensif.



Gambar 3. Komparasi Algoritma Meninggal Indonesia

4. Kesimpulan

Dari hasil tersebut, dapat disimpulkan bahwa kualitas udara memainkan peran penting dalam memperburuk kondisi kesehatan pasien Covid-19. Di negara dengan kualitas udara yang lebih baik, dampak terhadap tingkat keparahan pasien terlihat lebih terkendali, sedangkan di negara dengan kualitas udara yang lebih buruk, tingkat keparahan pasien lebih tinggi. Penelitian ini memberikan wawasan penting bagi masyarakat dan pemerintah terkait untuk terus meningkatkan kualitas udara guna mengurangi risiko kesehatan masyarakat, terutama dalam menghadapi pandemi atau penyakit pernapasan lainnya.

Daftar Pustaka

- [1] World Health Organization, "WHO Coronavirus Disease (COVID-19) Dashboard," <https://www.who.int/emergencies/diseases/novel>. [Online]. Available: <https://www.who.int/emergencies/diseases/novel>
- [2] C. Sohrabi *et al.*, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *Int. J. Surg.*, vol. 76, pp. 71–76, 2020, doi: <https://doi.org/10.1016/j.ijsu.2020.02.034>.
- [3] L. Qun *et al.*, "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, Mar. 2020, doi: [10.1056/NEJMoa2001316](https://doi.org/10.1056/NEJMoa2001316).
- [4] J. Bin Tan, M. J. Cook, P. Logan, L. Rozanova, and A. Wilder-Smith, "Singapore's Pandemic Preparedness: An Overview of the First Wave of COVID-19," *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, Dec. 2020, doi: [10.3390/ijerph18010252](https://doi.org/10.3390/ijerph18010252).
- [5] C. I. Agency, "The World Factbook: Indonesia." [Online]. Available: <https://www.cia.gov/the-world-factbook/>
- [6] M. Borro *et al.*, "Evidence-based considerations exploring relations between sars-cov-2 pandemic and air pollution: Involvement of pm2.5-mediated up-regulation of the viral receptor ace-2," *Int. J. Environ. Res. Public Health*, vol. 17, no. 15, pp. 1–13, 2020, doi: [10.3390/ijerph17155573](https://doi.org/10.3390/ijerph17155573).
- [7] S. Comunian, D. Dongo, C. Milani, and P. Palestini, "Air pollution and covid-19:

- The role of particulate matter in the spread and increase of covid-19's morbidity and mortality," *Int. J. Environ. Res. Public Health*, vol. 17, no. 12, pp. 1–22, 2020, doi: 10.3390/ijerph17124487.
- [8] A.-C. Pinho-Gomes *et al.*, "Air pollution and climate change," *Lancet Planet. Heal.*, vol. 7, no. 9, pp. e727–e728, Sep. 2023, doi: 10.1016/S2542-5196(23)00189-4.
 - [9] C. Copat *et al.*, "The role of air pollution (PM and NO₂) in COVID-19 spread and lethality: A systematic review.," *Environ. Res.*, vol. 191, p. 110129, Dec. 2020, doi: 10.1016/j.envres.2020.110129.
 - [10] X. Wu, R. C. Nethery, M. B. Sabath, D. Braun, and F. Dominici, "Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis," *Sci. Adv.*, vol. 6, no. 45, p. eabd4049, Sep. 2024, doi: 10.1126/sciadv.abd4049.
 - [11] B. T. P. Briandy, E. Yulianingsih, Fatmasari, and Ferdiansyah, "Analisis Tingkat Akurasi Prediksi Gejala COVID - 19 Dengan Menggunakan Metode Logistic Regression dan Support Vector Machine," *J. Fasilkom*, vol. 13, no. 02, pp. 269–278, 2023, doi: 10.37859/jf.v13i02.5629.
 - [12] D. Y. Liliana, H. Maulana, and A. Setiawan, "Data Mining untuk Prediksi Status Pasien Covid-19 dengan Pengklasifikasi Naïve Bayes," *Multinetics*, vol. 7, no. 1, pp. 48–53, 2021, doi: 10.32722/multinetics.v7i1.3786.
 - [13] 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 Data Mining Analysis for Classification of Air Quality Data Dki Jakarta Using Decision Tree Algorithm and Support Vector ," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 8954–8963, 2021.
 - [14] Y. Ogen, "Assessing nitrogen dioxide (NO₂) levels as a contributing factor to coronavirus (COVID-19) fatality," *Sci. Total Environ.*, vol. 726, p. 138605, 2020, doi: <https://doi.org/10.1016/j.scitotenv.2020.138605>.