

Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



Analisis Sentimen Netizen Terhadap Polusi Udara di DKI Jakarta

Analysis of Netizen Sentiment Towards Air Pollution in DKI Jakarta

Geonaldo Fauzan Akbar¹, Estu Sinduningrum^{2*}

^{1,2} Universitas Muhammadiyah prof dr HAMKA

Jl. Tanah Merdeka No.6, Pasar Rebo, Jakarta Timur, Daerah Khusus Ibukota
Jakarta. email: ¹geonaldofauzan6@gmail.com, ²estu.ningrum@uhamka.ac.id

ABSTRAK

Penelitian ini bermaksud untuk melakukan analisis terhadap sentimen netizen terkait masalah polusi udara di DKI Jakarta, yang tengah menjadi perbincangan di kalangan netizen baik di wilayah Jakarta maupun di luar Jakarta, seiring memburuknya kualitas udara di kota tersebut yang semakin tidak sehat jika terus dikonsumsi oleh penduduk Jakarta. Metode yang digunakan dalam penelitian ini adalah algoritma Naive Bayes, tujuan dari penelitian ini yaitu untuk mengetahui komentar positif dan negatif netizen yang diungkapkan melalui media sosial, dengan memanfaatkan beberapa platform seperti YouTube, TikTok, dan Threads. Data yang berhasil diperoleh yaitu 2106 data positif dan 495 data negatif. Setelah dilakukan eksperimen, diperoleh hasil akurasi 88.07%. Hasil penelitian ini menunjukkan bahwa algoritma Naive Bayes dapat digunakan untuk mengklasifikasikan sentimen komentar netizen terhadap polusi udara di daerah DKI Jakarta ini dengan tingkat akurasi yang cukup tinggi. Hasil ini dapat berguna pemerintah dan warga Jakarta juga dalam mencari solusi untuk masalah polusi ini. Dengan demikian, penelitian ini dapat menjadi dasar bagi pemahaman yang lebih menyeluruh terkait respons masyarakat terhadap isu lingkungan di kawasan perkotaan Jakarta.

Kata Kunci: *Polusi Udara, Algoritma Naive Bayes, Analisis Sentimen, Platform media sosial*

ABSTRACT

This research intends to analyze netizen sentiment regarding the problem of air pollution in DKI Jakarta, which is currently being discussed among netizens both in the Jakarta area and outside Jakarta, in line with the worsening air quality in the city which is becoming increasingly unhealthy if it continues to be consumed by Jakarta residents. The method used in this research is the Naive Bayes algorithm, the aim of this research is to determine the positive and negative comments of netizens expressed via social media, by utilizing several platforms such as YouTube, TikTok, and Threads. The data that was successfully obtained was 2106 positive data and 495 negative data. After conducting experiments, an accuracy result of 88.07%

*Penulis Korespondensi:
email: estu.ningrum@uhamka.ac.id

was obtained. The results of this research show that the Naïve Bayes algorithm can be used to classify the sentiment of netizen comments regarding air pollution in the DKI Jakarta area with a fairly high level of accuracy. These results can be useful for the government and citizens of Jakarta as well in finding solutions to this pollution problem. Thus, this research can become the basis for a more comprehensive understanding of community responses to environmental issues in the urban area of Jakarta.

Keywords: *Air Pollution, Naive Bayes Algorithm, Sentiment Analysis, Social media platforms*

1. PENDAHULUAN

Polusi udara yang terjadi di daerah DKI Jakarta membuat netizen memberikan komentar positif dan negatif pada sosial media karena lambatnya solusi dari pemerintah untuk mengatasi permasalahan ini karena udara yang terjadi di daerah DKI Jakarta ini udara terjadi ketika unsur-unsur berbahaya masuk atau bercampur dalam atmosfer, menyebabkan kerusakan lingkungan dan menurunkan kualitas udara di sekitarnya. Kerusakan lingkungan yang diakibatkan oleh kebakaran hutan memiliki dampak negatif yang signifikan. Salah satu dampak tersebut adalah polusi udara, yang tidak hanya memengaruhi satu wilayah tetapi juga merambat ke negara-negara lain, terutama mereka yang berdekatan dengan Indonesia [1].

Definisi polusi udara oleh Organisasi Kesehatan Dunia (WHO) menyatakan bahwa polusi udara merujuk pada pencemaran lingkungan, baik di dalam (lingkungan dalam ruangan) maupun di luar (lingkungan ambient), oleh zat-zat kimia, fisik, atau biologis yang dapat mengubah karakteristik alami atmosfer. Polusi udara terdiri dari kombinasi partikel dan gas yang dapat mencapai tingkat konsentrasi berbahaya, baik dalam konteks lingkungan luar maupun dalam ruangan. Contoh-contoh polutan umum mencakup jelaga, asap, jamur, serbuk sari, metana, dan karbon monoksida. Kulit sebagai organ yang berkontak langsung dengan polusi udara menjadi sasaran dari berbagai pemicu stres lingkungan [2].

Permasalahan polusi udara kini menjadi isu lingkungan yang mendesak di seluruh dunia. Peningkatan aktivitas industri, transportasi, dan penggunaan energi fosil telah mengakibatkan pelepasan emisi gas beracun dan partikel ke atmosfer, menyebabkan dampak negatif pada kualitas udara yang kita hirup sehari-hari. Polusi udara tidak hanya berdampak buruk pada lingkungan, tetapi juga memiliki konsekuensi serius terhadap kesehatan manusia [3].

Penelitian ini juga pernah dilakukan oleh Riska Indrayana dan Mukhammad Solikhin dengan mengangkat salah satu topik yang dibahas pada sosial media twitter mengenai kendaraan listrik, motor listrik, dan mobil listrik. Kendaraan listrik ini menuai banyak komentar dari masyarakat yang menimbulkan pro dan kontra disosial media twitter dengan menggunakan metode Naïve Bayes, analisis sentimen dan teks mining. Melalui analisis data dari 67% tweet dengan sentimen positif, 25% tweet dengan sentimen menentang dan 8% tweet dengan sentimen netral, Dari hasil analisis sentimen dari media sosial Twitter dengan menggunakan aplikasi RapidMiner terlihat hasil tingginya sentimen positif menandakan bahwa masyarakat tertarik dan setuju dengan perkembangan dan peralihan kendaraan bahan bakar fosil ke kendaraan listrik. Hasil analisis sentimen ini memberikan manfaat penting bagi masyarakat karena memungkinkan mereka untuk mencoba beralih dari kendaraan bahan bakar fosil ke kendaraan listrik sehingga dampaknya dapat mengurangi polusi udara yang terjadi akhir-akhir ini

2. DASAR/TINJAUAN TEORI

2.1 Polusi Udara

Faktor utama penyebab polusi udara di perkotaan adalah kemajuan industri yang mengarah pada penggunaan mesin dan alat transportasi berat. Untuk mengatasi masalah polusi udara di sekitar pabrik, peningkatan jumlah vegetasi atau penerapan penghijauan dengan menanam pohon-pohon dapat menjadi solusi.

2.2 Algoritma Naïve Bayes

Naïve Bayes merupakan salah satu teknik yang dapat digunakan untuk melakukan klasifikasi data. Naïve Bayes adalah suatu metode klasifikasi probabilitas yang sederhana, di mana perhitungan probabilitasnya melibatkan penjumlahan frekuensi dan kombinasi nilai dari dataset yang diterapkan sebagai referensi, yang umumnya disimpan dalam bentuk aturan, sebagai langkah-langkah prosedural untuk pemecahan masalah [4].

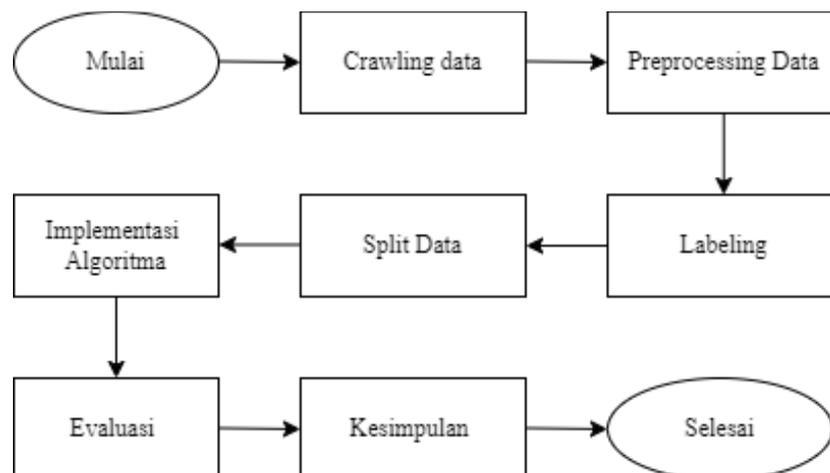
$$P(H|X) = \frac{P(H|X)P(H)}{P(X)}$$

Di mana $P(H|X)$ menyatakan probabilitas hipotesa H berdasarkan kondisi X, dengan X sebagai data pelatihan yang terkait dengan suatu class (label). [5] Dengan informasi yang telah diketahui, H merujuk pada data yang memiliki kelas (label) tertentu, $P(H)$ merupakan hasil probabilitas dari hipotesa X, $P(X)$ adalah hasil probabilitas dari data yang diamati, dan $P(H|X)$ adalah probabilitas kondisional dari X berdasarkan hipotesa H [6].

3. METODE

3.1 Diagram Alur Penelitian

Langkah-langkah yang diterapkan dalam penelitian ini melibatkan beberapa proses, termasuk *Crawling Data*, *Split Data*, *Preprocessing Data*, *Labelling*, *Split Data*. Diagram dibawah ini menunjukkan langkah-langkah yang dilakukan penulis bisa terlihat pada Gambar 1.



Gambar 1. Tahapan penelitian

3.2 Penjelasan Tahapan Penelitian

Pada bagian ini penulis akan menjelaskan proses dan langkah-langkah untuk melakukan Penelitian ini mengikuti langkah-langkah penelitian yang tergambar pada Gambar 1 di atas.

3.2.1 Pengumpulan Data

Pengumpulan data ini mencakup beberapa tahapan yaitu:

- a. Crawling Data
Pengambilan data dengan cara crawler dapat mempersatukan data secara otomatis. Crawler yang diperlukan dapat menyesuaikan dengan karakteristik platform masing-masing [7].
- b. Preprocessing Data
Dalam tahap ini, fungsinya mencakup penghapusan data duplikat, pemeriksaan ulang data yang tidak konsisten, terakhir, koreksi kesalahan yang mungkin terdapat pada data. [8].
- c. Labelling
Pada tahap ini data akan dilakukan pelabelan dengan cara manual oleh peneliti. Dengan pengelompokan ke dalam kategori positif dan negatif pada bagian komentar [9].
- d. Split Data
Pada langkah split data dilakukan melalui aplikasi RapidMiner Studio dan berasal dari langkah sebelumnya yaitu proses labelling yang membagi data menjadi dua bagian, yakni data training dan data testing.
- e. Implementasi Algoritma
implementasi ini adalah bagaimana cara pengolahan data dapat diterapkan di sebuah tools. Maka dari itu penulis menggunakan tools RapidMiner Studio [10].
- f. Evaluasi
Pada Tahap terakhir yaitu evaluasi yang di mana proses ini dilakukan evaluasi dari data yang sudah terbentuk. Evaluasi juga untuk melakukan perhitungan nilai akhir dalam analisis klasifikasi mencakup accuracy, precision, dan recall dengan menggunakan table confusion matrix sebagai referensi, hasil perhitungan tersebut ditampilkan kedalam bentuk table confusion matrix [11].

3.3 Preprocessing

Tahap preprocessing memiliki tujuan untuk mengeliminasi kata-kata yang kurang relevan dalam data, meningkatkan mutu data, serta menghasilkan teks yang lebih optimal pada kumpulan data.

3.4 Confusion Matrix

Confusion matrix adalah sebuah instrumen yang berfungsi untuk menganalisis keefektifan klasifikasi yang telah dilakukan dan juga untuk mengidentifikasi tupel dari kelas yang berbeda [12].

Prediction Class	Actual Class	
	Pred. Positif	Pred. Negatif
Negatif	TP	FN
Positif	FP	TN

- a. True Positives (TP) : Data positif yang berhasil diklasifikasikan dengan tepat oleh classifier.
- b. True Negatives (TN) : Data negatif yang berhasil diklasifikasikan dengan tepat oleh classifier.
- c. False Positives (FP) : Data negatif yang salah diklasifikasikan sebagai positif oleh classifier.

- d. False Negatives (FN) : Data positif yang salah diklasifikasikan sebagai negatif oleh classifier.

Parameter untuk mengukur suatu kinerja dari confusion matrix [13] yaitu :

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

Penjelasan dari TP dapat di sebut True Positive, TN dapat di sebut True Negative, FP dapat di sebut False Positive dan FN dapat di sebut False Negative.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Crawling Data

Pada penelitian ini penulis menggunakan metode *crawling data* dan pemodelan naïve bayes guna untuk mengklasifikasikan data dalam analisis sentiment. Pengumpulan data dilakukan dengan cara manual dari platform YouTube, TikTok kemudian Threads. Dalam analisis ini dua istilah, positif dan negatif, dipakai sebagai acuan untuk menilai sentimen atau respons.

Channel URL	Name	Comment	Time
http://www.you	@Nasimin-ee9	Buat yang kekeringan kapan?	2023-10-07T15:23:03Z
http://www.you	@fitrikhoriyah9	Dijakarta hujan	2023-09-30T04:59:02Z
http://www.you	@digiplanespo	Kalau memang penyebabnya adalah gedung?	2023-08-28T09:25:50Z
http://www.you	@ardiansahr46	Udah hujan semalem.. Allahu akbar	2023-08-27T21:29:20Z
http://www.you	@jimbojimbo69	banjir nggak nanti?	2023-08-27T12:27:39Z
http://www.you	@tonyhandaya	Pawang hujan di madalika aja di panggil...sun	2023-08-26T15:19:12Z
http://www.you	@JalanTv309	Bro bakar 10000 ban truk biar tambah kaya di	2023-08-26T10:20:06Z
http://www.you	@Ali-fk7ti	Solusinya stop menggunakan kendaraan berb	2023-08-25T20:52:43Z
http://www.you	@afwanizzul93	Cara yg jitu jakarta pabrik dipindah di pedesa	2023-08-25T16:36:11Z

Gambar 2. Hasil Crawling

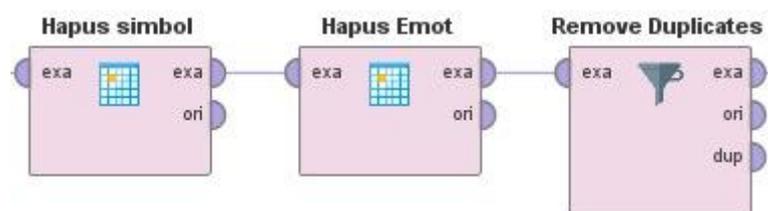
4.2 Preprocessing

Setelah itu, dilakukan proses preprocessing data. Langkah ini bertujuan mengurangi kata-kata yang kurang relevan dalam data, meningkatkan mutu data, dan menghasilkan teks yang lebih optimal pada kumpulan data. Proses ini melibatkan beberapa tahap, antara lain:

- Dalam tahap *Cleansing Data*, langkah awal yang diambil oleh peneliti adalah menggunakan operator *replace*. Selanjutnya, peneliti menyesuaikan parameter *replace* dengan data yang memerlukan pembersihan, seperti menghilangkan data yang duplikat, data yang mengandung simbol atau tanda baca, dan data yang menggunakan emoticon.

Input	Output
Kl memang penyebabnya adalah gedung tinggi sebaiknya di bongkar semua aja gedungnya dan di tanami pohon aja 😊😊	Kl memang penyebabnya adalah gedung tinggi sebaiknya di bongkar semua aja gedungnya dan di tanami pohon aja
Yuk pindah ke mobil listrik yuk biar polusi hilang 😊	Yuk pindah ke mobil listrik yuk biar polusi hilang
asap rokok termasuk ikut menyumbang polusi juga coba untuk stop dulu merokok warga jakarta 😊	asap rokok termasuk ikut menyumbang polusi juga coba untuk stop dulu merokok warga jakarta
Tanam pohon sebanyak banyaknya di wilayah Jakarta 🖱	Tanam pohon sebanyak banyaknya di wilayah Jakarta
terlalu banyak kendaraan karena asap kendaraan penyumbang polusi 😞	Terlalu banyak kendaraan karena asap kendaraan penyumbang polusi
Mending solusinya diberlakukan lg wfh ato Lockdown 🙄	Mending solusinya diberlakukan lg wfh ato Lockdown

Tahapan *Cleansing data* Data akan dihapus setelah diambil secara manual dengan menggunakan aplikasi RapidMiner. Ini mencakup penghapusan data yang memiliki simbol atau tanda baca, data yang mengandung emoticon, dan data yang duplikat. Tahap *Cleansing data* sesuai dengan yang diperlihatkan dalam Gambar 3.



Gambar 3. Proses *Cleansing Data*

- b. Pada tahap Case Folding, langkah yang digunakan untuk mengubah kata atau kalimat ke bentuk huruf kecil dengan menggunakan *lower case text*. Hasil proses *Case Folding* bisa terlihat pada Tabel 2.

Tabel 2. Hasil *Case Folding*

Input	Output
Kl memang penyebabnya adalah gedung tinggi sebaiknya di bongkar semua aja gedungnya dan di tanami pohon aja	kl memang penyebabnya adalah gedung tinggi sebaiknya di bongkar semua aja gedungnya dan di tanami pohon aja
Yuk pindah ke mobil listrik yuk biar polusi hilang	yuk pindah ke mobil listrik yuk biar polusi hilang
asap rokok termasuk ikut menyumbang polusi juga coba untuk stop dulu merokok warga jakarta	asap rokok termasuk ikut menyumbang polusi juga coba untuk stop dulu merokok warga jakarta

Tanam pohon sebanyak banyaknya di wilayah Jakarta	tanam pohon sebanyak banyaknya di wilayah jakarta
terlalu banyak kendaraan karena karena asap kendaraan penyumbang polusi	terlalu banyak kendaraan karena karena asap kendaraan penyumbang polusi
Mending solusinya diberlakukan lg wfh ato Lockdown	mending solusinya diberlakukan lg wfh ato lockdown

- c. Kemudian, langkah penghapusan *Tokenizing* bertujuan untuk memisahkan setiap kata dalam suatu kalimat. Operator yang diterapkan adalah *Tokenizing*. Hasil proses *Tokenizing* bisa terlihat pada Tabel 3.

Tabel 3. Hasil Tokenizing

Input	Output
kl memang penyebabnya adalah gedung tinggi sebaiknya di bongkar semua aja gedungnya dan di tanami pohon aja	kl, memang, penyebabnya, adalah, gedung, tinggi, sebaiknya, di, bongkar, semua, aja, gedungnya, dan, di, tanami, pohon, aja,
Yuk pindah ke mobil listrik yuk biar polusi hilang	yuk, pindah, ke, mobil, listrik, yuk, biar, polusi, hilang,
asap rokok termasuk ikut menyumbang polusi juga coba untuk stop dulu merokok warga jakarta	asap, rokok, termasuk, ikut, menyumbang, polusi, juga, coba, untuk, stop, dulu, merokok, warga, jakarta,
Tanam pohon sebanyak banyaknya di wilayah Jakarta	tanam, pohon, sebanyak, banyaknya, di, wilayah, Jakarta,
terlalu banyak kendaraan karena asap kendaraan penyumbang polusi	terlalu, banyak, kendaraan, karena, asap kendaraan, penyumbang, polusi,
Mending solusinya diberlakukan lg wfh ato Lockdown	mending, solusinya, diberlakukan, lg, wfh, ato, lockdown,

- d. Selanjutnya, proses *stemming* dilakukan untuk merubah kata-kata berimbuhan dan kata sambung menjadi bentuk dasarnya, sehingga dapat dipergunakan pada proses *stemming*, bisa terlihat pada Tabel 4.

Tabel 4. Kata untuk Stemming

Input	Output
kl	kalau
lg	lagi
ato	atau
ikt	ikut
adlh	adalah

Melalui *stemming*, diperoleh hasil dari proses ini yang mencakup transformasi kata-kata dari bahasa kekinian atau bahasa gaul menjadi bentuk dasar, serta pemisahan berdasarkan parameter yang telah ditentukan.. Hasil *Stemming* bisa terlihat pada Tabel 5.

Tabel 5. Hasil Stemming

Input	Output
kl memang penyebabnya adalah gedung tinggi sebaiknya di bongkar semua aja gedungnya dan di tanami pohon aja	kalau, memang, penyebabnya, adalah, gedung, tinggi, sebaiknya, di, bongkar, semua, aja, gedungnya, dan, di, tanami, pohon, aja,
Yuk pindah ke mobil listrik yuk biar polusi hilang	yuk, pindah, ke, mobil, listrik, yuk, biar, polusi, hilang,
asap rokok termasuk ikut menyumbang polusi juga coba untuk stop dulu merokok warga jakarta	asap, rokok, termasuk, ikut, menyumbang, polusi, juga, coba, untuk, stop, dulu, merokok, warga, jakarta,
Tanam pohon sebanyak banyaknya di wilayah Jakarta	ianam, pohon, sebanyak, banyaknya, di, wilayah, Jakarta,
terlalu banyak kendaraan karena asap kendaraan penyumbang polusi	terlalu, banyak, kendaraan, karena, asap kendaraan, penyumbang, polusi,
Mending solusinya diberlakukan lg wfh ato Lockdown	mending, solusinya, diberlakukan, lagi, wfh, atau, lockdown,

- e. Kemudian, pada langkah ini, terdapat hasil dari proses *tokenizing* yang berfokus pada pengambilan kata-kata yang signifikan. Hasil filter stopword bisa terlihat pada Tabel 6.

Tabel 6. Hasil Filter Stopword

Input	Output
kalau memang penyebabnya adalah gedung tinggi sebaiknya di bongkar semua aja gedungnya dan di tanami pohon aja	kalau, memang, penyebabnya, adalah, gedung, tinggi, sebaiknya, di, bongkar, semua, aja, gedungnya, dan, di, tanami, pohon, aja,
Yuk pindah ke mobil listrik yuk biar polusi hilang	yuk, pindah, ke, mobil, listrik, yuk, biar, polusi, hilang,
asap rokok termasuk ikut menyumbang polusi juga coba untuk stop dulu merokok warga jakarta	asap, rokok, termasuk, ikut, menyumbang, polusi, juga, coba, untuk, stop, dulu, merokok, warga, jakarta,
Tanam pohon sebanyak banyaknya di wilayah Jakarta	tanam, pohon, sebanyak, banyaknya, di, wilayah, Jakarta,
terlalu banyak kendaraan karena asap kendaraan penyumbang polusi	terlalu, banyak, kendaraan, karena, asap kendaraan, penyumbang, polusi,
Mending solusinya diberlakukan lagi wfh atau Lockdown	mending, solusinya, diberlakukan, lagi, wfh, atau, lockdown,

- f. Dalam langkah ini, proses *Filter Token (By Length)* bertujuan untuk mengeliminasi kata-kata yang telah ditetapkan, yakni kalimat yang terdiri dari kurang dari 3 huruf. Hasil *filter token (by length)* bisa terlihat pada Tabel 7.

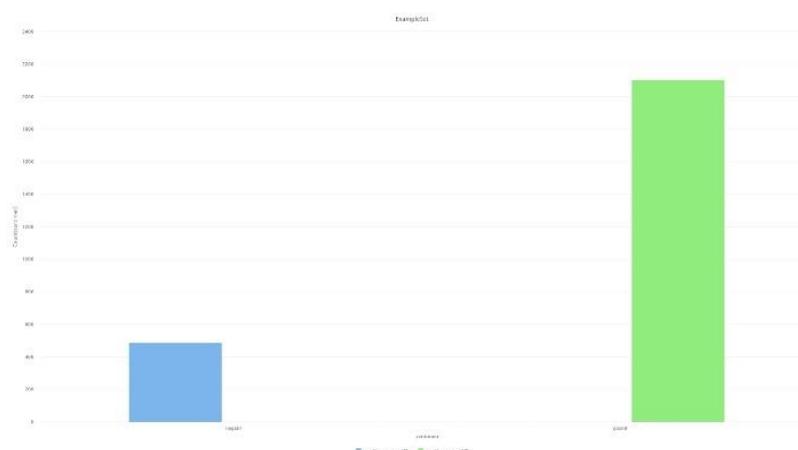
Tabel 7. Hasil Filter Token (By Length)

Input	Output
kalau, memang, penyebabnya, adalah, gedung, tinggi, sebaiknya, di, bongkar, semua, aja, gedungnya, dan, di, tanami, pohon, aja,	kalau memang penyebabnya adalah gedung tinggi sebaiknya di bongkar semua gedungnya di tanami pohon
yuk, pindah, ke, mobil, listrik, yuk, biar, polusi, hilang,	pindah ke mobil listrik biar polusi hilang
asap, rokok, termasuk, ikut, menyumbang, polusi, juga, coba, untuk, stop, dulu, merokok, warga, jakarta	asap rokok termasuk ikut menyumbang polusi juga coba untuk stop dulu merokok warga jakarta
tanam, pohon, sebanyak, banyaknya, di, wilayah, Jakarta	tanam pohon sebanyak banyaknya di wilayah Jakarta
terlalu, banyak, kendaraan, karena, asap kendaraan, penyumbang, polusi,	terlalu banyak kendaraan karena asap kendaraan penyumbang polusi
mending, solusinya, diberlakukan, lagi, wfh, atau, lockdown	mending solusinya diberlakukan lagi atau lockdown

4.3 Labelling

Dalam tahap ini, terdapat empat langkah dalam mengklasifikasikan data. Pertama labelling untuk memberikan label pada data yang akan dianalisis adalah langkah pertama. Setelah itu, dilakukan pemisahan data dengan proses Split Data menjadi dua bagian, yakni Data Training (data latih) dan Data Testing (data uji). Kemudian, implementasi algoritma dilakukan dengan menggunakan operator Naïve Bayes. Langkah terakhir melibatkan proses Evaluasi, dengan menambahkan grafik dan wordcloud sebagai output akhir dari pengujian data.

- Penulis menerapkan sampel labelling data, Grafik dari labelling bisa terlihat di dalam gambar 4.



Gambar 4. Grafik Hasil Labeling

Diperoleh dari total data class positif sejumlah 2106 data dan class negatif sejumlah 495 data. Penggunaan visualisasi wordcloud digunakan untuk mengamati kata-kata yang sering muncul dalam ulasan yang bersifat positif. yang bisa terlihat pada gambar 5.



Gambar 5. Tampilan wordcloud ulasan positif

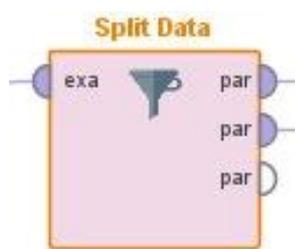
Penggunaan wordcloud visualisasi digunakan untuk mengamati kata-kata yang sering muncul dalam ulasan yang bersifat negatif. yang bisa terlihat pada gambar 6.



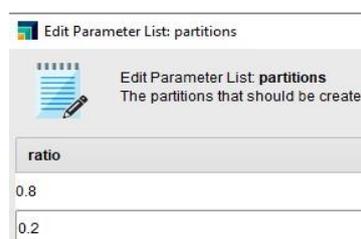
Gambar 6. Tampilan wordcloud ulasan negative

4.4 Split Data

Langkah berikutnya adalah memisahkan data menjadi dua kelompok, yaitu data training (data latih) dan data testing (data uji). Dalam penelitian ini, pembagian data dilakukan dengan menyisihkan 80% atau 0,8 dari total data untuk digunakan sebagai data training, sementara sisanya sebesar 20% atau 0,2 digunakan sebagai data testing, bisa terlihat pada gambar 7 dan 8.



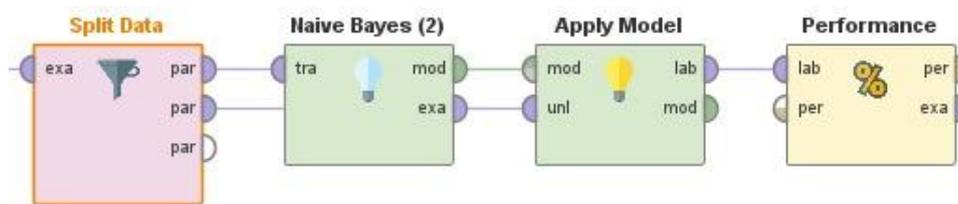
Gambar 7. Split Data



Gambar 8. Operator Split Data

4.5 Implementasi Algoritma Naïve Bayes

Langkah penerapan algoritma klasifikasi Naïve Bayes melibatkan analisis dan klasifikasi data training menggunakan algoritma Naïve Bayes, sementara data testing disesuaikan dengan bentuk pemodelan. Tahapan tersebut bisa diamati pada gambar 9.



Gambar 9. Proses Penerapan Algoritma Naïve Bayes

4.6 Evaluasi

Tahapan terakhir yaitu Evaluasi, Penulis memperoleh hasil analisis klasifikasi melalui penggunaan operator Split Data dan Naïve Bayes. Output dari perhitungan tersebut disajikan dalam bentuk tabel confusion matrix. Berikut ini adalah hasil yang dapat diperlihatkan menggunakan tabel confusion matrix bisa terlihat pada tabel 8.

Tabel 8. Tabel confusion matrix

	True Positif	True Negatif	Class Precision
Pred. positif	1433	0	100.00%
Pred. negatif	247	390	
Class recall	85.30%	100.00%	
Accuracy			88.07%

Diperoleh hasil dengan menggunakan operator Split Data, dengan tingkat akurasi sebesar 88.07%. Untuk kelas positif, ditemukan tingkat Precision sebesar 100% dan Recall sebesar 85.30%, sedangkan untuk kelas negatif, Precision mencapai 61.22% dan Recall mencapai 100%.

5. KESIMPULAN

Melalui analisis sentimen yang dilakukan terhadap polusi udara di DKI Jakarta, dapat diambil kesimpulan bahwa sebagian besar netizen menyampaikan keprihatinan dan kekhawatiran terhadap tingkat pencemaran udara yang tinggi di daerah tersebut. Banyak komentar yang mengekspresikan harapan akan adanya langkah-langkah konkret dan proaktif dari pihak berwenang guna menangani permasalahan ini. Meskipun beberapa netizen menunjukkan rasa frustrasi, secara umum, sentimen netizen mencerminkan kesadaran akan pentingnya penanganan cepat terhadap polusi udara demi kesehatan dan kesejahteraan masyarakat Jakarta. Oleh karena itu, diperlukan kerjasama bersama dan implementasi tindakan nyata untuk mengatasi permasalahan polusi udara di DKI Jakarta. Dari jumlah total data yang diolah, mencapai 2601 data dengan labeling secara manual, dapat disimpulkan bahwa algoritma Naïve Bayes sesuai untuk digunakan dalam analisis sentimen pada topik yang penulis teliti. Hasil evaluasi, termasuk nilai akurasi, presisi, dan recall dalam penelitian ini, dapat dirinci sebagai berikut: Akurasi Naïve Bayes mencapai 88.07%, nilai presisi sebesar 61.22%, dan nilai recall sebesar 85.30%. Hasil penelitian menunjukkan bahwa pembagian data secara manual dengan menggunakan operator Split Data mendapatkan hasil yang baik sehingga data yang diproses tersebut memiliki hasil perhitungan yang kuat.

REFERENSI

- [1] V. R. Kinanta, "Pencemaran Polusi Udara," *Res. Gate*, vol. 2, no. 1, p. 5, 2023.
- [2] M. D. Fernanda, H. T. Sibero, and H. Mutiara, "Polusi Udara dan Permasalahan terhadap Kulit," *Med. Prof. J. Lampung*, vol. 13, no. 1, pp. 66–71, 2023, doi: 10.53089/medula.v13i1.590.
- [3] S. Maharani and W. R. Aryanta, "Dampak Buruk Polusi Udara Bagi Kesehatan Dan Cara Meminimalkan Risikonya," *J. Ecocentrism*, vol. 3, no. 2, pp. 47–58, 2023, doi: 10.36733/jeco.v3i2.7035.
- [4] M. Ridho Handoko, "Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 50–58, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>.
- [5] H. Javadikasgari, E. G. Soltész, and A. M. Gillinov, "Surgery for Atrial Fibrillation," *Atlas of Cardiac Surgical Techniques*. pp. 479–488, 2018, doi: 10.1016/B978-0-323-46294-5.00028-5.
- [6] Y. Yuliyana and A. S. R. M. Sinaga, "Sistem Pakar Diagnosa Penyakit Gigi Menggunakan Metode Naive Bayes," *Fountain Informatics J.*, vol. 4, no. 1, p. 19, 2019, doi: 10.21111/fij.v4i1.3019.
- [7] J. Budiarto, "Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi Covid-19 di Media Sosial dengan Metode Crawling (Requirements Identification for NTB People in pandemic covid-19 at Social Media Using Crawling Method)," vol. 2, no. 4, pp. 244–250, 2021.
- [8] I. Riswanto and R. H. Laluma, "Klasifikasi Kelayakan Pinjaman Pada Koperasi Karyawan Menggunakan Metode Naive Bayes Classifier Berbasis Web," *Infotronik J. Teknol. Inf. dan Elektron.*, vol. 5, no. 1, pp. 11–16, 2020, doi: 10.32897/infotronik.2020.5.1.2.
- [9] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvananda, and I. K. A. G. Wiguna, "Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 205–215, 2022, doi: 10.23887/janapati.v11i3.49450.
- [10] Rayuwati, Husna Gemasih, and Irma Nizar, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- [11] F. Septianingrum, J. H. Jaman, and U. Enri, "Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1431, 2021, doi: 10.30865/mib.v5i4.3260.
- [12] T. A. Sundara, S. Ekaputri, and S. Sotar, "Naive Bayes Classifier untuk Analisis Sentimen Isu Radikalisme," *Pros. SISFOTEK*, vol. 4, no. 1, pp. 93–98, 2020, [Online]. Available: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/159>.
- [13] N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 123–131, 2021, doi: 10.29207/resti.v5i1.2587.