

ANALISIS SENTIMEN TERHADAP MOBIL LISTRIK DI INDONESIA PADA TWITTER: PENERAPAN NAÏVE BAYES CLASSIFIER UNTUK MEMAHAMI OPINI PUBLIK

Assiva Nurul Huzna¹, Indah Nurhayati, Alda Eva Saputri³, dan Mohammad Qomarul Huda⁴

^{1,2,3,4}Sistem Informasi Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta

assiva.huzna20@mhs.uinjkt.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen terhadap mobil listrik di Indonesia melalui Twitter dengan menerapkan Naïve Bayes Classifier guna memahami opini publik di media sosial. Data yang dikumpulkan mencakup 223 sentimen negatif dan 162 sentimen positif terkait mobil listrik. Metode klasifikasi Naïve Bayes digunakan untuk mengidentifikasi dan mengelompokkan sentimen-sentimen tersebut. Hasil penelitian menunjukkan bahwa opini publik cenderung bervariasi, dengan 48% sentimen negatif dan 52% sentimen positif. Faktor-faktor seperti harga, infrastruktur pengisian, dan performa kendaraan menjadi fokus utama dalam pembentukan opini. Implikasi temuan ini dapat memberikan wawasan bagi pemangku kepentingan industri mobil listrik di Indonesia untuk mengidentifikasi area peningkatan dan pengembangan guna memperkuat penerimaan masyarakat terhadap teknologi ramah lingkungan ini.

Kata Kunci: analisis sentiment; mobil listrik; *naïve bayes classifier*; *twitter*.

Abstract

This study aims to analyze the sentiment towards electric cars in Indonesia via Twitter by applying Naïve Bayes Classifier to understand public opinion on social media. The data collected includes 223 negative sentiments and 162 positive sentiments related to electric cars. The Naïve Bayes classification method was used to identify and categorize the sentiments. The results show that public opinion tends to vary, with 48% negative sentiments and 52% positive sentiments. Factors such as price, charging infrastructure, and vehicle performance are the main focus in opinion formation. The implications of these findings can provide insights for stakeholders of the electric car industry in Indonesia to identify areas of improvement and development to strengthen public acceptance of this environmentally friendly technology.

Keywords: *electric car; sentiment analysis; twitter; naïve bayes classifier.*

1. Pendahuluan

Pengguna *platform* komunikasi sosial terlibat dalam berbagai bentuk ekspresi. Mereka secara aktif berpartisipasi dalam penciptaan kendaraan atau ide baru, serta apresiasi dan

dukungan terhadap kendaraan, individu, dan layanan yang menawarkan dukungan kepada orang lain (Agüero-Torales et al., 2019). *Twitter*, sebagai *platform* media sosial yang sangat populer dengan miliaran pengguna aktif setiap bulan, memberikan ruang luas

bagi pengguna untuk berinteraksi global. Popularitasnya membuat pemantauan opini publik menjadi krusial untuk bisnis dan pemerintahan (Xu et al., 2022).

Jumlah pengguna *Twitter* terus meningkat dengan stabil. Menurut studi (We Are Social, 2023) pada awal tahun 2023, terjadi kenaikan sebanyak 120 juta pengguna dari 2022 hingga Januari 2023 mencapai total 556 juta pengguna di seluruh dunia (Kemp, 2023). Di Indonesia, *Twitter* juga memperlihatkan keberhasilannya dengan 24 juta pengguna pada awal tahun 2023, mencakup sekitar 8,7 persen dari total populasi negara ini.

Pemanfaatan data *Twitter* untuk analisis teks telah mendapatkan popularitas karena jumlah karakter yang dibatasi (280) dan meluasnya penggunaan *Twitter* sebagai platform bagi individu dari berbagai latar belakang untuk menyuarakan pendapat mereka tentang berbagai topik. Melalui analisis tweet dan balasan, dimungkinkan untuk mengidentifikasi keadaan emosional dan sentimen yang diungkapkan oleh pengguna, membentuk jaringan emosional yang saling berhubungan. Jaringan ini memungkinkan untuk mendeteksi emosi positif dan negatif. Dengan memeriksa tweet yang dibuat pengguna, menjadi jelas bagaimana *influencer* berperan dalam membentuk persepsi emosional di seluruh jaringan (Sailunaz & Alhaji, 2019).

Di tengah pertumbuhan pesat populasi dan urbanisasi yang berkelanjutan di Jakarta, Indonesia, permasalahan polusi udara telah menjadi salah satu tantangan utama yang mempengaruhi kualitas hidup warga kota (World Bank Group, 2020). Jakarta, sebagai salah satu kota megapolitan terpadat di dunia, telah terkena dampak serius dari polusi udara, terutama yang disebabkan oleh emisi dari kendaraan bermotor (The Jakarta Post, 2023). Meskipun langkah-langkah telah diambil oleh pemerintah daerah untuk mengatasi polusi udara, seperti regulasi emisi dan promosi transportasi berkelanjutan, tantangan yang dihadapi tetap signifikan.

Sementara itu, industri otomotif global telah menyaksikan transformasi yang signifikan menuju kendaraan ramah lingkungan, terutama dengan perkembangan mobil listrik. Mobil listrik adalah alternatif

yang menjanjikan untuk mengurangi tingkat emisi gas buang dan mengatasi permasalahan polusi yang terus berkembang di kota-kota besar seperti Jakarta (Electric Vehicles Initiative, 2019). Namun, adopsi mobil listrik di Indonesia masih dalam tahap awal, dan pandangan serta reaksi masyarakat terhadap kendaraan ini dapat berpengaruh pada perkembangan mereka di pasar domestik (We Are Social, 2023).

Ketika kita melihat perkembangan teknologi mobil listrik dan kondisi polusi di Jakarta, muncul pertanyaan penting: "Bagaimana sentimen masyarakat di Indonesia terkait mobil listrik, khususnya di tengah tantangan polusi udara yang serius di Jakarta?" Pertanyaan ini menggarisbawahi pentingnya memahami pandangan publik tentang mobil listrik dan bagaimana pandangan tersebut dapat mempengaruhi adopsi mereka di Indonesia.

Dalam konteks ini, analisis sentimen menjadi alat yang relevan dan kuat untuk memahami dinamika opini masyarakat terhadap mobil listrik di Indonesia, serta potensi peran mereka dalam mengatasi masalah polusi udara. Dengan menggabungkan teknologi analisis sentimen seperti *Naïve Bayes Classifier* (NBC) dengan data dari media sosial *Twitter*, kita dapat menggali lebih dalam dalam pandangan publik yang berkaitan dengan mobil listrik, yang pada gilirannya dapat membantu pemangku kepentingan, termasuk pemerintah dan produsen otomotif, dalam merancang kebijakan yang lebih efektif dan strategi pemasaran yang relevan (Gandomi & Haider, 2015).

Dalam kerangka ini, penelitian ini akan mengkaji sentimen masyarakat Indonesia terhadap mobil listrik melalui media sosial *Twitter*, dan akan menggunakan NBC untuk mengklasifikasikan sentimen dalam konteks ini. Penelitian ini bertujuan untuk memberikan wawasan yang lebih baik tentang pandangan publik dan bagaimana pandangan tersebut dapat berperan dalam mempromosikan mobilitas berkelanjutan, yang pada akhirnya dapat mengurangi dampak polusi udara di Jakarta dan kota-kota besar lainnya di Indonesia.

Analisis sentimen telah menjadi metode yang populer untuk menggali opini dan sikap pengguna terhadap berbagai topik, termasuk kendaraan dan layanan di media social (Shaik et al., 2023). Melalui analisis sentimen, dapat diketahui apakah tanggapan pengguna terhadap suatu hal seperti penggunaan mobil listrik cenderung positif, negatif, atau netral. Informasi ini sangat berharga bagi pengembang kendaraan dalam mengidentifikasi kekuatan dan kelemahan kendaraan mereka, serta mengidentifikasi area untuk perbaikan. Dengan memahami sentimen pengguna, pengembang dapat mengambil tindakan yang tepat untuk meningkatkan pengalaman pengguna dan meningkatkan popularitas kendaraan. Analisis sentimen merupakan salah satu teknik penelitian yang paling populer untuk memahami pendapat dan perasaan orang berdasarkan data teks (Leelawat et al., 2022). Selain itu, *Naïve Bayes Classifier* (NBC) adalah metode pembelajaran mesin yang telah terbukti efektif dalam berbagai tugas klasifikasi teks, termasuk analisis sentimen.

Permasalahan yang diidentifikasi dalam penelitian ini terfokus pada pengaruh topik polusi terhadap sentimen masyarakat terkait kendaraan Mobil Listrik yang menjadi perbincangan populer di *Twitter*. Tweet yang dihasilkan oleh pengguna memiliki potensi mempengaruhi persepsi individu terhadap Mobil Listrik. Meskipun Mobil Listrik sedang mengalami perkembangan dan mendapatkan tanggapan positif dan negatif di Indonesia, belum ada analisis sentimen yang dilakukan untuk mengevaluasi dampak keseluruhan perkembangan Mobil Listrik di negara ini.

Rumusan masalah penelitian ini adalah sejalan dengan latar belakang, di mana peneliti mencoba menjawab dua pertanyaan utama yaitu bagaimana analisis sentimen terhadap kendaraan mobil listrik dilakukan dengan menggunakan metode *Naïve Bayes Classifier*, dan bagaimana tingkat akurasi hasil analisis sentimen tersebut dengan metode *Naïve Bayes Classifier*.

Berdasarkan uraian diatas, penelitian ini akan menggagas "Analisis Sentimen Pada Media Sosial *Twitter* Terhadap Kendaraan Mobil Listrik Menggunakan dan *Naïve Bayes Classifier* (NBC)". Dengan penelitian ini

diharapkan penulis akan memperoleh penilaian akurasi yang tepat dalam menentukan sentimen masyarakat terkait dengan tweet positif dan negatif mengenai kendaraan mobil listrik.

2. Metode Penelitian

Metode pendekatan yang digunakan dalam penelitian ini adalah kuantitatif. Hal ini terlihat dari penggunaan metode *Naïve Bayes Classifier* untuk menganalisis sentimen masyarakat terhadap mobil listrik di Indonesia. *Naïve Bayes Classifier* adalah metode klasifikasi teks yang sederhana namun sangat akurat dan efektif. Metode ini menggunakan probabilitas untuk menentukan kelas dari suatu teks.

2.1. Evaluasi Performa Kualifikasi

Dalam evaluasi kinerja klasifikasi, pendekatan yang umum digunakan adalah perhitungan akurasi, presisi, *recall*, dan F-measure. Akurasi mengukur sejauh mana sentimen yang benar teridentifikasi secara akurat dan ditentukan dengan membagi jumlah sentimen yang teridentifikasi dengan benar dengan total data pengujian, sebagaimana ditentukan dalam Persamaan (1).

$$akurasi = \frac{jumlah\ sentimen\ benar}{jumlah\ data\ tes} \times 100\% \quad (1)$$

Selain akurasi, metrik lain yang digunakan untuk menilai kinerja klasifikasi mencakup presisi, perolehan, dan pengukuran-F. Presisi mengukur proporsi data relevan yang diidentifikasi di antara semua data yang ditempatkan. Persamaan (2) menguraikan perhitungan presisi.

$$precision = \frac{true\ positive}{true\ positive + false\ positive} \quad (2)$$

Recall mengukur sejauh mana materi relevan diidentifikasi dari total materi relevan. Perhitungan *recall*, seperti dijelaskan pada Persamaan (3), melibatkan pembagian data yang benar yang bernilai positif dengan jumlah data yang benar yang bernilai positif dan data yang salah yang bernilai negatif. Yang terakhir ini berasal dari nilai-nilai selain *true positive* di setiap kelas.

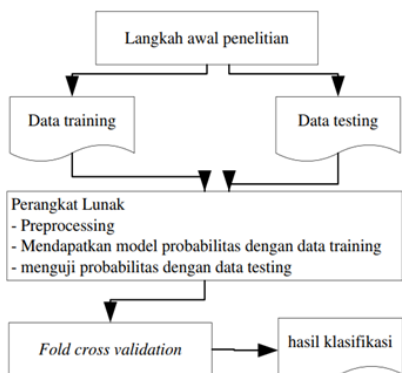
$$recall = \frac{true\ positive}{true\ positive + false\ negative} \quad (3)$$

Pengukuran F berfungsi sebagai metrik

terpadu yang menilai keberhasilan pengambilan dengan menggabungkan perolehan dan presisi. Ini dihitung dengan mengalikan presisi dan perolehan, membagi hasilnya dengan jumlah presisi dan perolehan, lalu mengalikan hasilnya dengan dua. Perhitungan F-measure ditentukan dengan menggunakan Persamaan (4).

$$f - measure = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

Untuk lebih jelasnya alur penelitian dapat dilihat pada Gambar 1.

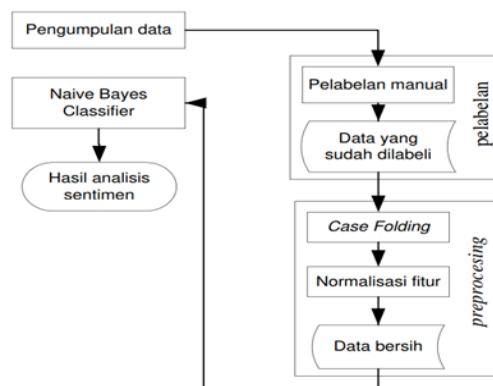


Gambar 1. Alur Evaluasi Performa Kualifikasi

2.2. Arsitektur Sistem

Sistem ini terdiri dari empat bagian: pengambilan data, preprocessing, pelabelan, dan klasifikasi data. Data input adalah tweet opini tentang mobil listrik berbahasa Indonesia, diperoleh melalui web scraping dengan mencari hashtag #judul mobil listrik di *search.x.com*. Pengumpulan data berlangsung dari 10 Januari 2023 hingga 10 September 2023, dengan total 500 *tweet*.

Data ini dibagi menjadi dua sentimen: positif untuk komentar yang mendukung mobil listrik dan negatif untuk komentar yang kritik. Desain arsitektur sistem dapat dilihat pada Gambar 1.



Gambar 2. Desain Arsitektur Sistem

2.3. Pemodelan Data

Pemodelan data adalah langkah penting dalam menguji keakuratan sistem prediksi, seperti model klasifikasi sentimen pada *tweet*. Proses training melibatkan penggunaan data training yang sudah dilabeli dan di-preprocess. Hasilnya disimpan sebagai model data dalam format *pickle*. Sistem hanya mempertimbangkan kata-kata dengan 3 karakter atau lebih dalam proses training.

Pada tahap testing, data baru digunakan tanpa perlu dilabeli, tetapi tetap di-preproses. Sistem mengekstrak fitur kata dan mengklasifikasikannya sesuai dengan model yang sudah dibuat, menghasilkan output berupa sentimen positif atau negatif. *Naïve Bayes Classifier*.

2.4. Naïve Bayes Classifier

Dalam proses penerapan Naïve Bayes Classifier untuk analisis sentimen *tweet* menggunakan NLTK, langkah awal yang dilakukan adalah menyiapkan data pelatihan dan pengujian. Secara khusus, daftar tupel dibuat, bersama dengan daftar fitur kata, untuk membangun pengklasifikasi. Untuk melatih pengklasifikasi, digunakan 200 *tweet* positif dan 200 *tweet* negatif dari file *positif.txt* dan *negatif.txt*. Fase pengujian melibatkan pengorganisasian data pengujian untuk mengevaluasi akurasi pengklasifikasi.

Kumpulan data positif dan negatif diubah menjadi daftar tupel, di mana kata-kata diubah menjadi huruf kecil dan kata-kata yang memiliki kurang dari tiga huruf dikecualikan. Daftar fitur kata dibuat, berisi kata-kata unik yang disusun

berdasarkan frekuensi. Berbagai fungsi digunakan untuk menampilkan fitur kata, mengekstraksi kata dari tweet, dan menentukan frekuensinya.

Pengklasifikasi dikembangkan dengan mengidentifikasi fitur-fitur yang relevan, dan ekstraksi fitur digunakan sebagai kamus untuk mengidentifikasi kata-kata dalam tweet. Pengklasifikasi Naïve Bayes menggunakan probabilitas sebelumnya dan kontribusi fitur untuk klasifikasi tweet. Model probabilitas dihitung menggunakan metode seperti kemungkinan maksimum atau kesamaan tertinggi dalam pembelajaran yang diawasi. Dalam konteks klasifikasi, penghitungan berkisar pada penentuan $P(H|X)$, yang mewakili probabilitas kebenaran hipotesis (valid) berdasarkan data sampel yang diamati.

$$P(H|X) = \frac{p(X|H)p(H)}{p(x)} \tag{5}$$

Keterangan:

X = Data sample dengan klas (label) yang tidak diketahui

H = Hipotesa bahwa X adalah data dengan klas (label) C

$P(H|X)$ = Peluang bahwa hipotesa benar (valid) untuk data sampel X yang diamati

$P(X|H)$ = Peluang data sample X, bila diasumsikan bahwa hipotesa benar (valid).

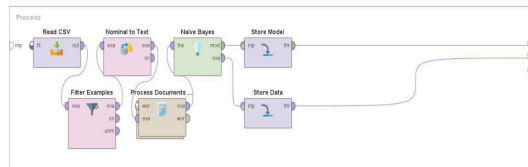
$P(H)$ = Peluang dari hipotesa H

$P(X)$ = Peluang data sample yang diamati

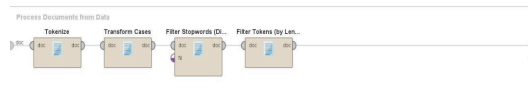
3. Hasil dan Pembahasan

3.1 Process Text

Preprocessing adalah suatu metode dalam data mining yang melibatkan transformasi data mentah ke dalam suatu format yang terstruktur dan dapat dimengerti. Data mentah sering kali tidak lengkap, tidak konsisten, dan mungkin mengandung banyak kesalahan. Penggunaan teknik *preprocessing* untuk membersihkan data dari *noise* dapat meningkatkan kinerja pengklasifikasi dan mempercepat proses klasifikasi, terutama dalam analisis sentimen secara *realtime* [12]. Proses *preprocessing* mencakup beberapa tahapan seperti *transform cases*, *tokenize*, *filter tokens*, *filter stopwords*, dan *stemming* yang dapat dilihat pada Gambar 4.



Gambar. 3. Proses data



Gambar. 4. Process Text Pada RapidMiner

- Transform case*: Proses konversi keseluruhan teks dalam *tweet* ke bentuk standar, yakni dengan merubah hurufnya menjadi huruf kecil atau *lowercase*.
- Tokenize*: Tokenisasi merupakan langkah mengumpulkan semua kata yang muncul dalam teks sekaligus menghilangkan segala tanda baca dan simbol yang tidak bersifat huruf.
- Filters tokens*: Pada tahap ini dilakukan pemilihan token dengan memperhatikan panjangnya, yaitu memilih token yang memiliki panjang setidaknya 4 karakter (huruf), meskipun beberapa di antaranya adalah *stopwords*.
- Filter stopwords*: *Filter stopwords* merupakan langkah penghapusan kata yang dianggap tidak memiliki makna dalam dokumen *tweet* penelitian ini. Tujuannya adalah untuk menyederhanakan proses pemrosesan.
- Stem (stemming)*: *Stemming* merupakan proses untuk menemukan kata dasar dengan mengeliminasi imbuhan seperti awalan, sisipan, akhiran, dan kombinasi awalan dan akhiran dari kata. Dengan melakukan stemming, variasi kata dapat diminimalkan sehingga semua bentuk kata yang berasal dari satu kata dapat direpresentasikan dengan bentuk yang sama.

3.2 Hasil Analisis Sentimen

Setelah menyelesaikan proses data training dan data testing dengan menggunakan algoritma naïve bayes, sistem akan secara otomatis menghitung nilai akurasi berdasarkan data yang ada. Hasil analisis sentimen terhadap mobil listrik dalam penelitian ini dapat dinyatakan dalam bentuk positif dan negatif, yang dapat ditemukan pada Gambar 6 di bawah ini.

Gambar. 5. Hasil analisis sentimen terhadap mobil listrik di Indonesia

Berdasarkan gambar yang tersedia, dapat disimpulkan bahwa penerimaan masyarakat terhadap mobil listrik di Twitter cenderung negatif, dibuktikan dengan 223 tweet negatif.

Negatifnya cuitan tersebut disebabkan oleh pengguna yang cenderung mengutamakan emosi dibandingkan pertimbangan logis saat mengutarakan pendapatnya. Sebaliknya, 162 tweet positif menunjukkan bahwa pengguna lebih cenderung berpikir kritis dan menilai potensi risiko sebelum membagikan komentar positif.

Algoritme Naïve Bayes, yang diproses melalui Rapidminer pada 385 tweet, menghasilkan wawasan. Di kelas ingatan positif, seluruh 162 tweet diprediksi dengan tepat sebagai sentimen positif, sehingga menghasilkan tingkat ingatan positif sebenarnya sebesar 100%. Mengenai kelas presisi dalam memprediksi sentimen negatif, 223 tweet teridentifikasi secara akurat sebagai negatif, namun 162 tweet salah diprediksi sebagai sentimen positif. Bagian selanjutnya menyajikan data akurasi yang diperoleh dari proses algoritma *Naïve Bayes* menggunakan Rapidminer:

PerformanceVector

```
PerformanceVector:
accuracy: 100.00%
ConfusionMatrix:
True:  negatif positif
negatif:    223    0
positif:    0    162
precision: 100.00% (positive class: positif)
ConfusionMatrix:
True:  negatif positif
negatif:    223    0
positif:    0    162
recall: 100.00% (positive class: positif)
ConfusionMatrix:
True:  negatif positif
negatif:    223    0
positif:    0    162
AUC (optimistic): 1.000 (positive class: positif)
AUC: 1.000 (positive class: positif)
AUC (pessimistic): 1.000 (positive class: positif)
```

Gambar. 6. Hasil data akurasi dari proses algoritma *naïve bayes* menggunakan *rapidminer*

Berdasarkan data dijelaskan bahwa dari 500 data keseluruhan berdasarkan sentimen masyarakat melalui tweet mobil listrik terdapat 385 data akurat yang merupakan *true negatif* 223, 162 *true positif* dengan tingkat akurasi sebanyak 100%.

4. Kesimpulan

Analisis sentimen pada penelitian ini yaitu untuk menggali mengenai opini dan sikap masyarakat terhadap penggunaan mobil listrik, apakah pengguna mobil listrik tersebut cenderung positif, negatif, atau netral.

Metode yang digunakan untuk analisis sentimen ini yaitu *Naïve Bayes Classifier* (NBC) untuk mengklasifikasikan teks, termasuk hasil analisis sentimen masyarakat terkait kendaraan Mobil Listrik. Hasil data sentimen masyarakat diperoleh dari *Twitter* yang kemudian diolah, dan dibagi menjadi menjadi dua sentimen yaitu, positif untuk komentar yang mendukung mobil listrik dan negatif untuk komentar yang kritik. Berdasarkan hasil perhitungan menggunakan algoritma *naive bayes* terhadap 500 data keseluruhan berdasarkan sentimen masyarakat melalui tweet mobil listrik terdapat 385 data akurat yang merupakan *true negatif* 223, 162 *true positif* dengan tingkat akurasi sebanyak 100%. Dengan memahami sentimen pengguna, pengembang dapat mengambil tindakan yang tepat untuk meningkatkan pengalaman pengguna dan meningkatkan popularitas kendaraan.

Daftar Pustaka

- üero-Torales, M. M., Cobo, M. J., Herrera-Viedma, E., & López-Herrera, A. G. (2019). A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor. *Procedia Computer Science*, *162*, 392–399. <https://doi.org/10.1016/j.procs.2019.12.002>
- lectric Vehicles Initiative. (2019). *Electric Vehicles Initiative – Programmes - IEA*. <https://www.iea.org/programmes/electric-vehicles-initiative>
- ndomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- mp, S. (2023). *Digital 2023: Global Overview Report*. <https://datareportal.com/reports/digital-2023-global-overview-report>
- elawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtum, K.,

- Laosunthara, A., Yudha, A. K., & Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. *Heliyon*, 8(10). <https://doi.org/10.1016/j.heliyon.2022.e10894>
- Sailunaz, K., & Alhaji, R. (2019). Emotion and sentiment analysis from Twitter text. *Journal of Computational Science*, 36, 101003. <https://doi.org/https://doi.org/10.1016/j.joics.2019.05.009>
- Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, 2, 100003. <https://doi.org/https://doi.org/10.1016/j.nlp.2022.100003>
- The Jakarta Post. (2023, December). *Sesak napas di Jakarta*. <https://www.thejakartapost.com/opinion/2023/06/21/breathless-in-jakarta.html?lang=id>
- We Are Social. (2023, December). *Digital 2023*. <https://wearesocial.com/id/blog/2023/01/digital-2023/>
- World Bank Group. (2020, February 11). *The hidden wealth of cities: creating, financing, and managing public spaces*. The World Bank. <https://www.worldbank.org/en/topic/urbandevelopment/publication/the-hidden-wealth-of-cities-creating-financing-and-managing-public-spaces>
- Xu, Q. A., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, 3, 100073. <https://doi.org/doi.org/10.1016/j.dajour.2022.100073>