



JURNAL SISTEM INFORMASI DAN TEKNOLOGI (S I N T E K)

Situs Jurnal

<https://sintek.stmikku.ac.id/index.php/home>



EKSPLORASI SOCIAL MEDIA MINING DALAM ANALITIK BIG DATA: METODE DAN IMPLEMENTASI

Lukas Umbu Zogara

Program Studi Teknologi Informasi, Universitas Utpadaka Swastika
Metropolis Town Square Lt. 2, Jl. Hartono Raya Modern, Kel. Kelapa Indah.

Email: lukasumbuzogara68@gmail.com

ABSTRAK

Perkembangan big data terus menunjukkan peningkatan yang signifikan seiring dengan kemajuan teknologi Internet of Things (IoT). Salah satu perangkat yang paling umum digunakan dalam kehidupan sehari-hari adalah ponsel, terutama jenis smartphone. Melalui aplikasi media sosial yang tersedia di dalamnya, smartphone menjadi sumber utama penghasil data dalam jumlah besar setiap detik. Data yang dihasilkan ini banyak dimanfaatkan dalam berbagai penelitian yang berkaitan dengan analitik big data. Teknik analisis yang digunakan pun bervariasi tergantung pada jenis data dan tujuan studi. Penelitian ini bertujuan untuk mengulas konsep analitik big data, khususnya pada ranah social media mining, beserta teknik-teknik analisis yang digunakan. Pendekatan yang digunakan adalah studi literatur terhadap jurnal ilmiah dan buku yang relevan dengan topik ini. Berdasarkan hasil kajian, ditemukan bahwa media sosial berperan penting dalam menyuplai data yang dapat dimanfaatkan dalam pengambilan keputusan melalui analitik big data. Setiap teknik analitik memiliki keunggulan dan keterbatasan tersendiri, sehingga integrasi beberapa metode dianjurkan guna meningkatkan akurasi hasil analisis. Penelitian ini diharapkan dapat memperluas pemahaman pembaca yang tertarik pada bidang social media mining.

Kata Kunci: *Big Data Analytics, Social Media Mining, IoT, Analitik Data, Keputusan.*

1. PENDAHULUAN

Fenomena big data telah menjadi terkenal sejak tahun 2011 dan terus berkembang dengan kecepatan yang dipercepat setiap tahun [1]. Perkembangan ini terutama didorong oleh digitalisasi yang sedang berlangsung dan proliferasi Internet of Things (IoT), yang menghasilkan volume data yang substansial secara per detik [2]. Saat ini, perangkat IoT yang paling umum mencakup smartphone, tablet, laptop, dan komputer pribadi, yang secara kolektif menghasilkan triliunan titik data dalam berbagai format seperti teks, gambar, audio, dan video [2]. Kontributor signifikan terhadap asal mula data besar adalah platform media sosial, termasuk Twitter, Facebook, Instagram, dan WeChat, yang secara kolektif menampung miliaran pengguna di seluruh

dunia [3]. Media sosial berfungsi tidak hanya sebagai platform untuk komunikasi tetapi juga sebagai generator data yang dapat dimanfaatkan untuk segudang aplikasi, termasuk analisis tren, pengambilan keputusan berdasarkan informasi, dan identifikasi real-time dari kejadian tertentu [4].

Berbagai studi sebelumnya menunjukkan penerapan analisis big data dari media sosial dalam beragam konteks. Misalnya, data dari platform Flickr digunakan untuk mengklasifikasi tutupan lahan menggunakan metode Artificial Neural Networks (ANNs) [5]. Selain itu, media sosial juga dimanfaatkan dalam sistem peringatan dini terhadap bencana dan penilaian dampaknya [6].

Namun, studi-studi terdahulu masih menghadapi berbagai tantangan, seperti keterbatasan akurasi analisis serta kesulitan dalam menangani volume data besar dari media sosial. Oleh karena itu, penelitian ini bertujuan untuk mengulas dan mengevaluasi beragam teknik social media mining dalam konteks analitik big data, serta mengidentifikasi metode yang paling efektif untuk mengekstraksi informasi dari media sosial.

Artikel ini merupakan hasil tinjauan dari sejumlah penelitian yang berkaitan dengan social media mining dalam analitik big data. Penulisan ini bersifat orisinal, dikembangkan dari studi-studi sebelumnya namun dengan pendekatan dan sudut pandang yang berbeda.

2. LANDASAN TEORI

2.1 Big Data dan Internet of Things (IoT)

Big data adalah kumpulan data berukuran sangat besar dan kompleks yang memerlukan metode khusus untuk pengelolaannya. Seiring berkembangnya teknologi IoT, volume data yang dihasilkan meningkat secara drastis. IoT memungkinkan perangkat untuk saling terhubung dan bertukar informasi dalam skala besar, menghasilkan data yang sangat berguna untuk berbagai analisis, termasuk analitik big data.

2.2 Social Media Mining

Social media mining merupakan proses untuk menggali pola dan informasi dari data yang dihasilkan pengguna media sosial. Proses ini melibatkan berbagai metode seperti analisis sentimen, analisis jaringan sosial (network analysis), dan text mining untuk mengidentifikasi tren serta perilaku pengguna.

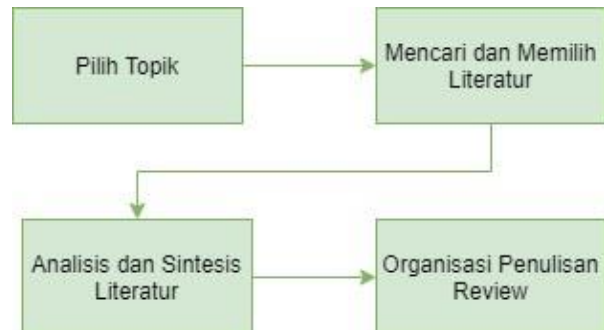
2.3. Big data dari media social.

Big data dari media sosial telah dimanfaatkan mining dan text mining digunakan untuk memetakan aktivitas rekreasi satwa liar [7]. dalam berbagai bidang. Misalnya, data Dalam konteks keamanan, social media mining juga diterapkan untuk mendeteksi potensi aktivitas terorisme dengan menganalisis karakteristik tweet berbahasa Arab [8].

Dengan memahami konsep-konsep dasar seperti big data, IoT, dan social media mining, artikel ini bertujuan meninjau berbagai pendekatan analitik big data yang

digunakan dalam social media mining guna memperoleh informasi yang lebih presisi dan aplikatif.

3. METODOLOGI PENELITIAN



Gambar 2.1 Metodologi Penelitian

Proses penelitian ini dilakukan melalui empat tahapan utama, sebagaimana ditampilkan dalam Gambar 2.1. Setiap tahapan dirancang untuk memastikan bahwa peninjauan literatur dilakukan secara sistematis dan terstruktur. Berikut adalah penjelasan dari masing-masing tahap: Berikut ini adalah penjelasan tahapannya: (1) Langkah awal adalah menentukan topik utama yang akan dikaji, yaitu social media mining dalam konteks analitik big data. Topik ini dipilih karena relevan dengan perkembangan teknologi informasi dan kebutuhan pengambilan keputusan berbasis data; (2) Literatur yang dikumpulkan merupakan referensi yang relevan dengan topik yang telah ditentukan. Literatur tersebut diperoleh dari berbagai sumber terpercaya seperti jurnal ilmiah, prosiding konferensi, dan artikel akademik lainnya yang membahas penerapan social media mining dan big data analytics; (3) Pada tahap ini dilakukan telaah mendalam terhadap isi dari literatur yang telah dipilih. Peneliti mengidentifikasi poin-poin penting dari masing-masing studi, termasuk pendekatan, metode, dan hasil temuan. Informasi tersebut kemudian disintesis untuk memperoleh gambaran menyeluruh mengenai perkembangan dan tantangan dalam topik yang diteliti; (4) Tahap akhir adalah menyusun kerangka penulisan artikel review berdasarkan hasil analisis dan sintesis. Penulisan dilakukan secara sistematis agar alur pembahasan mudah dipahami dan dapat memberikan kontribusi ilmiah yang signifikan bagi pembaca.

4. HASIL DAN PEMBAHASAN

4.1 Penggunaan Media Sosial Dalam Analitik Big Data

Tabel 1. Penelitian-Penelitian Big Data yang Menggunakan Media Sosial

| No | Penulis | Tujuan | Metode |
|----|------------------------|---|--|
| 1 | Hu et al. [9]. | Mengevaluasi sentimen pengguna lokal dari data geo-tweet secara spasial dan temporal | VADER, LISA, dan LDA (Latent Dirichlet Allocation) |
| 2 | Kolahka j et al. [10]] | Menyusun rekomendasi paket wisata berdasarkan data multidimensi dan geo-tagged photos | PrefixSpan, DBSCAN |
| 3 | Park et al. [11] | Menganalisis mobilitas wisatawan internasional ke Korea melalui data ponsel | DBSCAN, SPADE |
| 4 | ElQadi et al. [5] | Membangun framework untuk klasifikasi tutupan lahan dari citra media sosial | Artificial Neural Networks (ANNs) |
| 5 | Yan et al. [12] | Menggabungkan data media sosial dan lokasi geografis untuk menilai dampak bencana | SentiWordNet 3.0 |
| 6 | Yan, dkk.[12] | Mengkaji pemulihan pascabencana di lokasi wisata menggunakan sentimen geo-tagged | PatternAnalyzer, LDA |
| 7 | Hou et al. [13] | Mengeksplorasi perhatian publik terkait COVID-19 melalui media sosial | LDA, Baidu Sentiment Analysis |

| | | | |
|----|--------------------------|--|---|
| 8 | Monkman et al. [7] | Mengembangkan pendekatan untuk mengidentifikasi aktivitas rekreasi satwa liar | Text and Data Mining, leksikon berbasis ulasan & ahli |
| 9 | Alhalabi et al. [8] | Menyediakan sistem pendeteksi perilaku teroris melalui karakteristik dinamis tweet | Algoritma kecerdasan buatan yang dikembangkan sendiri |
| 10 | Abdul-Rahman et al. [14] | Membangun framework analitik Twitter untuk perencanaan kota berkelanjutan | LDA, VADER |

Berdasarkan ringkasan dalam Tabel 4.1, dapat disimpulkan bahwa media sosial memiliki peran yang signifikan dalam berbagai bidang yang berkaitan dengan analitik big data, seperti kesehatan [15], [13], perencanaan dan pembangunan kota [14], [5], [9], pariwisata [11], [10], [12], keamanan dan sosial [8], mitigasi serta analisis dampak bencana alam [5], [16], hingga konservasi lingkungan dan satwa liar [7]. Beragamnya bidang ini menunjukkan bahwa informasi yang diperoleh dari media sosial mampu memberikan kontribusi nyata dalam pengambilan keputusan di berbagai sektor. Metode yang digunakan pun beragam, mulai dari algoritma yang sudah dikenal luas hingga metode baru yang dikembangkan secara khusus untuk tujuan tertentu [8]. Beberapa penelitian memanfaatkan pendekatan standar seperti LDA, VADER, dan DBSCAN, sedangkan lainnya menciptakan sistem atau framework baru yang disesuaikan dengan karakteristik data. Keanekaragaman ini mengindikasikan bahwa efektivitas metode sangat tergantung pada konteks permasalahan dan jenis data yang dianalisis. Dengan demikian, pemahaman mendalam tentang teknik social media mining menjadi penting agar hasil analisis dari media sosial dapat diolah menjadi informasi yang akurat dan bermanfaat dalam konteks big data.

4.2 Teknik Sosial Media Mining

Social media mining merupakan bagian dari proses data mining yang bertujuan untuk menggali informasi serta pola-pola tersembunyi dari data yang dihasilkan oleh pengguna media sosial. Proses ini sangat penting mengingat data yang tersedia tidak hanya dalam jumlah besar, tetapi juga bersifat tidak terstruktur dan sangat bervariasi, seperti teks, gambar, atau metadata lokasi. Data mining sendiri didefinisikan sebagai suatu metode untuk mengekstraksi pengetahuan dan relasi yang berguna dari kumpulan data berukuran sangat besar secara otomatis [17]. Proses ini memungkinkan ditemukannya hubungan-hubungan tersembunyi antar data yang nantinya dapat digunakan sebagai dasar dalam pengambilan keputusan strategis [20]. Keberhasilan data mining banyak didukung oleh bidang ilmu lainnya, seperti jaringan saraf tiruan (neural networks), basis data gambar, pengenalan pola, pemrosesan sinyal, dan analisis data spasial [18]. Beberapa karakteristik utama dari data mining antara lain adalah kemampuannya untuk menemukan pola tersembunyi yang sebelumnya tidak diketahui, penerapannya pada volume data yang sangat besar agar hasilnya dapat dipercaya, serta perannya yang penting dalam mendukung pengambilan keputusan kritis [18]. Dalam konteks media sosial, proses mining menjadi lebih kompleks karena data yang dihasilkan sangat dinamis, tidak terstruktur, dan kadang sensitif. Namun demikian, karakteristik ini justru menjadikan media sosial sebagai sumber data yang menarik bagi banyak pemangku kepentingan, baik di sektor kesehatan, pariwisata, pendidikan, keamanan, dan lainnya [21]. Hal ini juga didorong oleh sifat komunikasi di media sosial yang bersifat real-time, yang memungkinkan pengambilan keputusan dilakukan secara cepat dan berbasis pada informasi terkini. Untuk memperoleh hasil analisis yang optimal, proses data mining umumnya mengikuti enam tahap utama, yaitu: (1) pembersihan data (data cleaning) untuk menghilangkan informasi yang tidak relevan atau rusak; (2) integrasi data (integration) dari berbagai sumber; (3) seleksi data (data selection) sesuai kebutuhan analisis; (4) transformasi data (data transformation) menjadi format yang sesuai; (5) penambangan data (mining) dengan algoritma tertentu; dan (6) evaluasi hasil (evaluation) untuk menilai keakuratan dan

kegunaan pola yang ditemukan [17]. Secara umum, data mining adalah bagian inti dari proses Knowledge Discovery in Databases (KDD), yang mencakup berbagai langkah mulai dari seleksi, praproses, transformasi, penambangan, hingga evaluasi data untuk memperoleh informasi yang bernilai [19]. Dalam praktiknya, terdapat beberapa algoritma yang sering digunakan dalam analisis data dari media sosial, di antaranya:

Naive Bayes, sebuah algoritma klasifikasi sederhana yang banyak digunakan karena efisiensinya, terutama untuk dataset kecil [20]. Namun, pada skala data yang sangat besar seperti 20.000 tweet atau lebih, performa Naive Bayes cenderung menurun, sehingga diperlukan kombinasi dengan algoritma lain untuk meningkatkan akurasi [21].

Random Forest, metode ansambel yang menggabungkan banyak pohon keputusan untuk menentukan klasifikasi data baru. Meskipun ditujukan untuk meningkatkan akurasi, dalam beberapa studi, Random Forest menunjukkan performa yang lebih rendah dibandingkan metode lain seperti K-NN dan Naive Bayes, sehingga disarankan untuk dikombinasikan dengan pendekatan lain [21], [22].

K-Nearest Neighbor (K-NN), yang bekerja dengan menentukan sejumlah k tetangga terdekat dari data yang sedang dianalisis. Teknik ini sangat baik dalam hal presisi, meskipun nilai recall-nya lebih rendah dibandingkan Naive Bayes. Tingkat akurasi K-NN meningkat seiring bertambahnya jumlah fitur dalam data [21], [22].

Adaboost, metode pembelajaran ansambel yang meningkatkan akurasi secara bertahap dengan menyesuaikan bobot pada data yang salah klasifikasi dalam setiap iterasi. Dengan karakteristik tersebut, Adaboost terbukti mampu mengungguli algoritma lain dalam hal akurasi, terutama ketika digunakan dalam konteks pemrosesan teks [21], [22].

4.3 Perbandingan Metode Social Media Mining

Tabel 4.2 Perbandingan Metode Social Media Mining

| Judul Penelitian | Metode | Accuracy |
|--|---|----------|
| <i>Supervised Learning for Suicidal Ideation</i> | Random Forest dengan berbagai jenis fitur | 96,38% |

| | | |
|--|---------------------------------------|-----|
| <i>Detection in Online User Content [23]</i> | | |
| <i>Text Mining and Determinants of Sentiments: Twitter Usage by Media in Uganda [24]</i> | Random Forest dan Logistic Regression | 72% |
| Analisis Sentimen Data Twitter tentang Masalah Obesitas di Indonesia [25] | Naïve Bayes Classifier | 94% |

Dari tabel di atas, dapat dilihat bahwa meskipun metode yang digunakan pada beberapa studi terlihat serupa, hasil akurasi dapat sangat bervariasi. Sebagai contoh, penggunaan Random Forest dalam studi deteksi ide bunuh diri menghasilkan akurasi yang sangat tinggi (96,38%) [23], namun kombinasi Random Forest dengan Logistic Regression dalam konteks lain justru menunjukkan penurunan akurasi yang cukup signifikan (72%) [24]. Sementara itu, algoritma Naïve Bayes yang dikenal sederhana justru mampu menghasilkan akurasi tinggi (94%) dalam studi mengenai sentimen obesitas di Twitter [25]. Hal ini menunjukkan bahwa pemilihan dan kombinasi metode serta fitur yang tepat sangat memengaruhi efektivitas model, sehingga penting untuk mempertimbangkan konteks data dan tujuan analisis secara menyeluruh.

5. PENUTUP

Media sosial memiliki dampak besar dalam menyediakan data yang mencerminkan berbagai fenomena sosial, lingkungan, dan kondisi global. Melalui proses social media mining, informasi bernilai dapat diekstraksi dari kumpulan data tekstual yang sangat besar, dan hasilnya dapat dimanfaatkan dalam berbagai aspek analitik big data, seperti kesehatan, pariwisata, keamanan, hingga perencanaan kota. Sejumlah metode telah dikembangkan dan digunakan untuk mendukung proses ini, masing-masing dengan kelebihan dan keterbatasannya. Oleh karena itu, pemilihan metode yang tepat serta kemungkinan mengombinasikan beberapa algoritma menjadi strategi penting untuk meningkatkan akurasi dan efektivitas analisis. Hasil perbandingan

menunjukkan bahwa integrasi metode atau pemanfaatan pendekatan yang lebih adaptif dapat menghasilkan performa yang lebih tinggi. Kesimpulannya, social media mining merupakan pendekatan strategis dalam memaksimalkan potensi data media sosial untuk mendukung pengambilan keputusan berbasis big data di berbagai sektor.

DAFTAR PUSTAKA

- [1] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, Apr. 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.
- [2] M. Batty *et al.*, "Smart cities of the future," *Eur. Phys. J. Spec. Top.*, vol. 214, no. 1, pp. 481–518, Dec. 2012, doi: 10.1140/epjst/e2012-01703-3.
- [3] M. A. Carlos, M. Nogueira, and R. J. Machado, "Analysis of dengue outbreaks using big data analytics and social networks," in *Proc. 4th Int. Conf. Systems and Informatics (ICSAI)*, 2017, vol. 2018-Jan., pp. 1592–1597, doi: 10.1109/ICSAI.2017.8248538.
- [4] Anonymous, "Sentiment Analysis On Social Media Big Data With Multiple Tweet Words," *Int. J. Innov. Technol. Explor. Eng.*, doi: 10.35940/ijitee.J9684.0881019.
- [5] M. M. ElQadi, M. Lesiv, A. G. Dyer, and A. Dorin, "Computer vision-enhanced selection of geo-tagged photos on social network sites for land cover classification," *Environ. Model. Softw.*, vol. 128, Mar. 2020, Art. no. 104696, doi: 10.1016/j.envsoft.2020.104696.
- [6] D. Wu and Y. Cui, "Disaster early warning and damage assessment analysis using social media data and geo-location information," *Decis. Support Syst.*, vol. 111, pp. 48–59, 2018, doi: 10.1016/j.dss.2018.04.005.
- [7] G. G. Monkman, M. J. Kaiser, and K. Hyder, "Text and data mining of social media to map wildlife recreation activity," *Biol. Conserv.*, vol. 228, pp. 89–99, 2018, doi: 10.1016/j.biocon.2018.10.010.
- [8] W. Alhalabi *et al.*, "Social mining for terroristic behavior detection through Arabic tweets characterization," *Future Gener. Comput. Syst.*, vol. 116, pp. 132–144, 2021, doi: 10.1016/j.future.2020.10.027.
- [9] T. Hu, B. She, L. Duan, H. Yue, and J. Clunis, "A systematic spatial and temporal sentiment analysis on geo-tweets," *IEEE Access*, vol. 8,

- pp. 8658–8667, 2020, doi: 10.1109/ACCESS.2019.2961100.
- [10] M. Kolahkaj, A. Harounabadi, A. Nikravanshalmani, and R. Chinipardaz, "A hybrid context-aware approach for e-tourism package recommendation," *Electron. Commer. Res. Appl.*, vol. 42, Feb. 2020, Art. no. 100978, doi: 10.1016/j.elerap.2020.100978.
- [11] S. Park, Y. Xu, L. Jiang, Z. Chen, and S. Huang, "Spatial structures of tourism destinations: A trajectory data mining approach leveraging mobile big data," *Ann. Tour. Res.*, vol. 84, Jan. 2020, Art. no. 102973, doi: 10.1016/j.annals.2020.102973.
- [12] Y. Yan, J. Chen, and Z. Wang, "Mining public sentiments and perspectives from geotagged social media data," *Appl. Geogr.*, vol. 123, Feb. 2020, Art. no. 102306, doi: 10.1016/j.apgeog.2020.102306.
- [13] K. Hou, T. Hou, and L. Cai, "Public attention about COVID-19 on social media: An investigation based on data mining and text analysis," *Pers. Individ. Dif.*, vol. 175, Sep. 2020, Art. no. 110701, doi: 10.1016/j.paid.2021.110701.
- [14] M. Abdul-Rahman, E. H. W. Chan, M. S. Wong, V. E. Irekponor, and M. O. Abdul-Rahman, "A framework to simplify pre-processing location-based social media big data for sustainable urban planning and management," *Cities*, vol. 109, Sep. 2021, Art. no. 102986, doi: 10.1016/j.cities.2020.102986.
- [15] K. G. Blumenthal *et al.*, "Mining social media data to assess the risk of skin and soft tissue infections," *J. Allergy Clin. Immunol.*, vol. 144, no. 1, pp. 129–134, Jul. 2019, doi: 10.1016/j.jaci.2019.01.029.
- [16] J. A. de Bruijn, H. de Moel, B. Jongman, J. Wagemaker, and J. C. J. H. Aerts, "TAGGS: Grouping Tweets to Improve Global Geoparsing for Disaster Response," *J. Geovisualization Spat. Anal.*, vol. 2, no. 1, 2018, doi: 10.1007/s41651-017-0010-6.
- [17] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, 2nd ed. Hoboken, NJ: Wiley, 2014.
- [18] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 2nd ed. [Online]. Available: https://www.researchgate.net/publication/262562891_Data_Mining_Concepts_and_Techniques_2nd_Edition. [Accessed: May 22, 2021].
- [19] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "Knowledge Discovery and Data Mining: Towards a Unifying Framework," in *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining*, 1996.
- [20] M. Allahyari *et al.*, "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques," *arXiv preprint*, arXiv:1707.02919, Jul. 2017.
- [21] N. Siyam, O. Alqaryouti, and S. Abdallah, "Mining government tweets to identify and predict citizens engagement," *Technol. Soc.*, vol. 60, 2020, Art. no. 101211, doi: 10.1016/j.techsoc.2019.101211.
- [22] C. C. Aggarwal, "An Introduction to Social Network Data Analytics," in *Social Network Data Analytics*, Springer US, 2011, pp. 1–15.
- [23] S. Ji, C. P. Yu, S. F. Fung, S. Pan, and G. Long, "Supervised learning for suicidal ideation detection in online user content," *Complexity*, vol. 2018, 2018, Art. no. 6157249, doi: 10.1155/2018/6157249.
- [24] F. Namugera, R. Wesonga, and P. Jehopio, "Text mining and determinants of sentiments: Twitter social media usage by traditional media houses in Uganda," *Comput. Soc. Netw.*, vol. 6, no. 1, 2019, doi: 10.1186/s40649-019-0063-4.
- [25] F. F. Mailo *et al.*, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia," *J. Sist. Inf. Kesehat. Masy.*, vol. 4, no. 1, pp. 28–36, 2019.