ANALISIS SENTIMEN DAN PEMODELAN TOPIK PADA *POST* TENTANG MEREK TEKNOLOGI DI MEDIA SOSIAL X MENGGUNAKAN *FINE-TUNING* INDOBERT DAN BERTOPIC

SKRIPSI

Diajukan Untuk Memenuhi Sebagian dari
Syarat Memperoleh Gelar Sarjana Komputer
Program Studi Ilmu Komputer

oleh
Muhammad Rayhan Nur
2100192

PROGRAM STUDI ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
2025

ANALISIS SENTIMEN DAN PEMODELAN TOPIK PADA *POST* TENTANG MEREK TEKNOLOGI DI MEDIA SOSIAL X MENGGUNAKAN *FINE-TUNING* INDOBERT DAN BERTOPIC

Oleh
Muhammad Rayhan Nur
2100192

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer di Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

Universitas Pendidikan Indonesia
Juli 2025

LEMBAR PENGESAHAN

MUHAMMAD RAYHAN NUR

2100192

ANALISIS SENTIMEN DAN PEMODELAN TOPIK PADA *POST* TENTANG MEREK TEKNOLOGI DI MEDIA SOSIAL X MENGGUNAKAN *FINE-TUNING* INDOBERT DAN BERTOPIC

DISETUJUI DAN DISAHKAN OLEH TIM PEMBIMBING:

Pembimbing 1,

Dr. Yudi Wibisono, M.T.

NIP. 197507072003121003

Pembimbing 2,

Dr. Rani Megasari, M.T.

NIP. 198705242014042002

Mengetahui,

Ketua Program Studi Ilmu Komputer

Dr. Muhammad Nursalman, M.T.

NIP. 197909292006041002

ANALISIS SENTIMEN DAN PEMODELAN TOPIK PADA *POST* TENTANG MEREK TEKNOLOGI DI MEDIA SOSIAL X MENGGUNAKAN *FINE-TUNING* INDOBERT DAN BERTOPIC

Oleh

Muhammad Rayhan Nur – m.rayhan.nur@upi.edu

2100192

ABSTRAK

Media sosial telah menjadi wadah bagi konsumen untuk menyampaikan persepsi dan opini. Opini yang beredar tersebut berpotensi menjadi sumber data yang berharga bagi *brand*, termasuk Xiaomi, dalam memahami persepsi publik terhadap produk mereka. Penelitian ini bertujuan untuk menganalisis sentimen dan mengidentifikasi topik diskusi pada unggahan (*post*) mengenai merek teknologi Xiaomi di platform X (sebelumnya Twitter) dengan pendekatan berbasis Transformer. Dua metode utama yang digunakan adalah *fine-tuning* IndoBERT untuk model klasifikasi sentimen dan BERTopic untuk pemodelan topik. Data yang berhasil dikumpulkan sebanyak 10.130 *post* dari bulan Mei 2023 hingga Mei 2025 yang dilanjutkan menuju tahapan praproses serta pelabelan. Model klasifikasi dilatih dengan berbagai kombinasi konfigurasi *hyperparameter*, dengan hasil pengujian terbaik menghasilkan nilai *accuracy* 79,8%, *precision* 73,0%, *recall* 67,7%, dan *f1-score* (*macro*) sebesar 0,699. Distribusi sentimen dalam data menunjukkan dominasi sentimen netral, sedangkan BERTopic berhasil menghasilkan 16 *cluster* topik dengan rata-rata nilai *coherence* (C_v) sebesar 0,5437. Topik paling dominan dengan jumlah anggota *cluster* terbanyak membahas mengenai produk Xiaomi *Series* dan Poco. Topik dengan persentase sentimen negatif tertinggi berkaitan dengan layanan *service center*, dan sentimen positif tertinggi mengenai produk komputer tablet (*tab*) Xiaomi. Penggabungan hasil analisis sentimen dan pemodelan topik memberikan pemahaman yang lebih mendalam terhadap isu yang dibicarakan serta persepsi konsumen sekaligus memberikan wawasan yang relevan bagi perusahaan untuk mengidentifikasi kekuatan dan potensi peningkatan yang dapat dilakukan.

Kata Kunci: Media Sosial X, Analisis Sentimen, Pemodelan Topik, IndoBERT, BERTopic

SENTIMENT ANALYSIS AND TOPIC MODELING ON POSTS ABOUT TECHNOLOGY BRAND ON SOCIAL MEDIA X USING FINE-TUNED INDOBERT AND BERTOPIC

Compiled By

Muhammad Rayhan Nur – m.rayhan.nur@upi.edu

2100192

ABSTRACT

Social media has become a platform for consumers to express their perceptions and opinions. These circulating opinions have the potential to be a valuable source of data for brands, including Xiaomi, in understanding public perception of their products. This study aims to analyze sentiment and identify discussion topics in posts about the Xiaomi technology brand on platform X (formerly Twitter) using a Transformer-based approach. Two main methods used are fine-tuning IndoBERT for sentiment classification models and BERTopic for topic modeling. Data collected amounted to 10,130 posts from May 2023 to May 2025, which proceeded to the preprocessing and labeling stages. The classification model was trained with various combinations of hyperparameter configurations, with the best test results producing an accuracy of 79.8%, a precision of 73.0%, a recall of 67.7%, and an f1-score (macro) of 0.699. The distribution of sentiment in the data shows a dominance of neutral sentiment, while BERTopic successfully generated 16 topic clusters with an average coherence value (C_v) of 0.5437. The most dominant topic, with the largest number of cluster members, discussed Xiaomi Series and Poco products. The topic with the highest percentage of negative sentiment related to service center services, and the highest positive sentiment concerned Xiaomi tablet computers (tabs). Combining sentiment analysis and topic modeling results provides a deeper understanding of the issues discussed and consumer perceptions, while providing relevant insights for the company to identify strengths and potential improvements.

***Keywords**: Social Media X, Sentiment Analysis, Topic Modeling, IndoBERT, BERTopic*

DAFTAR ISI

LEMBAR PERNYATAAN KEASLIAN iii
KATA PENGANTAR iv
UCAPAN TERIMA KASIH v
ABSTRAK vii
ABSTRACT viii
DAFTAR ISI ix
DAFTAR TABEL xii
DAFTAR GAMBAR xiv
BAB I PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 5
1.4 Manfaat Penelitian 5
1.5 Batasan Masalah 6
1.6 Sistematika Penulisan 6
BAB II KAJIAN PUSTAKA 8
2.1 *Electronic Word of Mouth (e-WOM)* 8
2.2 Media Sosial X (Twitter) 10
2.3 *Natural Language Processing* 16
2.4 Analisis Sentimen 18
2.5 Pemodelan Topik 25
2.6 Transformer 28
2.7 *Bidirectional Encoder Representations from Transformers* (BERT) 35
2.7.1 *Pre-Training* dan *Fine-Tuning* 37
2.7.2 IndoBERT 39
2.8 BERTopic 40
2.9 *Dataset* 46
2.10 Metode Evaluasi 48
2.11 Penelitian Terkait 55
BAB III METODOLOGI PENELITIAN 58

3.1 Desain Penelitian 58
3.2 Alat dan Bahan Penelitian 62
3.2.1 Alat Penelitian 62
3.2.2 Bahan Penelitian 62
BAB IV HASIL DAN PEMBAHASAN 63
4.1 Pengumpulan Data 63
4.2 Perancangan Model Komputasi 64
4.3 Pembentukan *Dataset* 65
4.3.1 Penanganan Data Duplikat 65
4.3.2 Pemilihan Data 65
4.3.3 Praproses Teks 68
4.3.3.1 *Case Folding* 68
4.3.3.2 *Punctuation Removal* 68
4.3.3.3 *Slang Normalization* 68
4.3.3.4 *Stopwords Removal* 69
4.3.4 Pelabelan Data Sentimen 69
4.4 Skenario Eksperimen 72
4.4.1 Skenario Eksperimen Pembangunan Model Klasifikasi 72
4.4.1.1 Skenario Penggunaan Data Input 72
4.4.1.2 Skenario Optimasi *Hyperparameter* 72
4.4.2 Skenario Eksperimen Pemodelan Topik 73
4.5 Eksperimen 73
4.5.1 Eksperimen Pembangunan Model Klasifikasi 73
4.5.1.1 Pembagian Data 74
4.5.1.2 Proses *Fine-tuning* Model IndoBERT 75
4.5.1.3 Hasil dan Pembahasan Eksperimen Pembangunan Model Relevansi 77
4.5.1.4 Hasil dan Pembahasan Eksperimen Pembangunan Model Klasifikasi Sentimen 79
4.5.2 Eksperimen Pemodelan Topik 87
4.5.2.1 Konfigurasi *Hyperparameter* 87
4.5.2.2 Hasil dan Pembahasan Eksperimen Pemodelan Topik 87

4.6 Analisis Lanjutan 93
4.6.1 Analisis Sentimen per Bulan 93
4.6.2 Analisis Distribusi Sentimen per *Cluster* Topik 95
4.7 Pembuatan *Interface* 97
BAB V KESIMPULAN DAN SARAN 100
5.1 Kesimpulan 100
5.2 Saran 101
DAFTAR PUSTAKA 102

DAFTAR TABEL

Tabel 2.1 Contoh Baris Data pada *Dataset* SmSA IndoNLU 47
Tabel 2.2 Contoh Baris *Dataset Colloquial Indonesian Lexicon* 47
Tabel 2.3 Hasil Jendela Proses *Sliding Window* dengan Ukuran 3....................... 53
Tabel 4.1 Contoh Data *Post* 63
Tabel 4.2 Hasil Proses Penanganan Data Duplikat... 65
Tabel 4.3 Contoh Hasil Proses Pemilihan Data 66
Tabel 4.4 Contoh Hasil Proses *Case Folding* 68
Tabel 4.5 Contoh Hasil Proses *Punctuation Removal* 68
Tabel 4.6 Contoh Hasil Proses *Slang Normalization* 69
Tabel 4.7 Contoh Hasil Proses *Stopwords Removal* 69
Tabel 4.8 Contoh Kesalahan Pelabelan Data oleh Model Bantu......................... 71
Tabel 4.9 Tabel Skenario Kombinasi *Hyperparameter* Eksperimen Model Klasifikasi 72
Tabel 4.10 Distribusi Kelas Relevansi tiap Subset.. 74
Tabel 4.11 Distribusi Kelas Sentimen tiap Subset 74
Tabel 4.12 Contoh Hasil Proses Tokenisasi... 75
Tabel 4.13 Contoh Hasil Proses *Encoding* 76
Tabel 4.14 Contoh Hasil Proses *Vectorization*... 76
Tabel 4.15 Hasil Evaluasi Eksperimen Model Relevansi Tanpa Praproses Teks Tahap *Stopwords Removal* 77
Tabel 4.16 Hasil Evaluasi Eksperimen Model Relevansi Melalui Praproses Teks Tahap *Stopwords Removal* 78
Tabel 4.17 Hasil Evaluasi Model Klasifikasi Relevansi Terbaik pada Dua Skenario Input Teks 79
Tabel 4.18 Hasil Evaluasi Eksperimen Model Klasifikasi Sentimen Tanpa Praproses Teks Tahap *Stopwords Removal* 80
Tabel 4.19 Hasil Evaluasi Eksperimen Model Klasifikasi Sentimen Melalui Praproses Teks Tahap *Stopwords Removal* 80

Tabel 4.20 Hasil Evaluasi Model Klasifikasi Sentimen Terbaik pada Dua Skenario Input Teks 81
Tabel 4.21 *F1-score* per Kelas Model Klasifikasi Sentimen Terbaik pada Dua Skenario Input Teks 83
Tabel 4.22 Contoh Kesalahan Prediksi Label Sentimen oleh Model Terbaik Skenario Data Input Tanpa *Stopwords Removal* 85
Tabel 4.23 Perbandingan Hasil Evaluasi dengan Model *Deep Learning* Lain 86
Tabel 4.24 Konfigurasi *Hyperparameter* BERTopic Terbaik 88
Tabel 4.25 Hasil Pemodelan Topik Menggunakan BERTopic dengan Konfigurasi *Hyperparameter* Terbaik 88
Tabel 4.26 Lima *Cluster* Topik Terdominan 90
Tabel 4.27 Lima *Cluster* Topik dengan Nilai *C_v* Tertinggi 92
Tabel 4.28 Contoh *Post* Sentimen Positif pada Bulan Maret 93
Tabel 4.29 Contoh *Post* Sentimen Negatif pada Bulan Agustus 94
Tabel 4.30 Contoh *Post* Pada *Cluster* Topik dengan Persentase Sentimen Netral Tertinggi 96
Tabel 4.31 Contoh *Post* Pada Topik dengan Persentase Sentimen Negatif Tertinggi 96
Tabel 4.32 Contoh *Post* Pada Topik dengan Persentase Sentimen Positif Tertinggi 97

DAFTAR GAMBAR

Gambar 1.1 Contoh Opini Konsumen pada *Post* di X 2
Gambar 2.1 Laporan Data Total Pengguna Potensial X di Indonesia 2025 (We Are Social, 2025) 11
Gambar 2.2 Contoh *Post* dan *Reply* di X... 12
Gambar 2.3 Contoh *Quote Post* di X.. 12
Gambar 2.4 Contoh Likes di X.. 13
Gambar 2.5 Fitur *Direct Messages* di X (Ochai Emmanuel, 2024) 13
Gambar 2.6 Fitur *Bookmark Folders* di X (Sumber: *Post* @X di X, 2023)....... 14
Gambar 2.7 Contoh *List* di X.. 14
Gambar 2.8 Fitur *Communities* di X 15
Gambar 2.9 Contoh *Post* Ulasan Konsumen di X.. 16
Gambar 2.10 Tipe-tipe Pendekatan dalam Analisis Sentimen (Yaakub dkk., 2019) 20
Gambar 2.11 Contoh Hasil ABSA (Alshaikh dkk., 2024)................................. 23
Gambar 2.12 Ilustrasi Proses Pemodelan Topik (botpenguin.com, 2024)........... 25
Gambar 2.13 Konsep LDA untuk Pemodelan Topik (Blei, 2012)...................... 26
Gambar 2.14 Contoh Visualisasi Hasil Pemodelan Topik (C. Li & Hu, 2025)... 27
Gambar 2.15 Contoh Visualisasi Hasil Pendekatan Topic-based Sentiment Analysis (C. Li & Hu, 2025).. 28
Gambar 2.16 Arsitektur Model Transformer (Vaswani dkk., 2017)................... 29
Gambar 2.17 Ilustrasi Matriks Hasil Perhitungan *Positional Encoding* (Saeed, 2023) 31
Gambar 2.18 Ilustrasi self-attention mechanism yang memfokuskan pada frasa “the animal” ketika memproses kata “it” (Alammar & Grootendorst, 2024)....... 32
Gambar 2.19 Arsitektur Detil Model Transformer: Komponen *Self-Attention Mechanism* (Weng, 2018) 33
Gambar 2.20 Representasi Input Model BERT (Özçift dkk., 2021) 36
Gambar 2.21 Arsitektur dalam Pelatihan Model BERT (Devlin dkk., 2019)...... 37
Gambar 2.22 Ilustrasi *Masked Language Modeling* (Mittal & Modi, 2021) 37

Gambar 2.23 Ilustrasi Proses *Next Sentence Prediction* (Bozuyla & Özçift, 2022) 38
Gambar 2.24 Tahapan Proses dalam BERTopic (Grootendorst, 2022) 40
Gambar 2.25 Arsitektur Sentence-BERT (SBERT) (Reimers & Gurevych, 2019) 41
Gambar 2.26 Contoh Penerapan UMAP 42
Gambar 2.27 Perbandingan Hasil *Clustering* Algoritma *K-Means* dengan HBDSCAN (Berba, 2020) 44
Gambar 2.28 Contoh Visualisasi Hasil *Topic Modeling* dengan BERTopic (Grootendorst, 2025) 46
Gambar 2.29 Contoh *Sliding Window* dengan Ukuran 3 51
Gambar 3.1 Diagram Alur Penelitian 58
Gambar 4.1 Model Komputasi Analisis Sentimen dan Pemodelan Topik........... 64
Gambar 4.2 Distribusi Jumlah Data Berdasarkan Kelas Relevansi..................... 67
Gambar 4.3 Distribusi Jumlah Data Berdasarkan Kelas Sentimen dalam *Dataset* 71
Gambar 4.4 *Confusion Matrix* Model Sentimen Terbaik (Input *Stopwords* Masih Ada*)* 82
Gambar 4.5 *Confusion Matrix* Model Sentimen Terbaik (Input Stopwords Tidak Ada*)* 82
Gambar 4.6 Evaluasi Kurva ROC Untuk Model Terbaik Skenario Input Dengan *Stopwords Removal* 84
Gambar 4.7 Grafik Jumlah *Post* per Bulan untuk 5 *Cluster* Topik Terdominan 90
Gambar 4.8 Berita Perilisan Produk Xiaomi *Series* pada Maret 2025 91
Gambar 4.9 Distribusi Persentase Sentimen pada *Dataset* per Bulan................. 93
Gambar 4.10 Distribusi Sentimen per *Cluster* Topik 95
Gambar 4.11 Tampilan Awal *Interface* Aplikasi... 98
Gambar 4.12 Tampilan Hasil Visualisasi Distribusi Sentimen 99
Gambar 4.13 Tampilan Hasil Visualisasi Distribusi Sentimen per *Cluster* Topik99
Gambar 4.14 Tampilan Tabel Daftar *Cluster* Hasil Pemodelan Topik 99

DAFTAR PUSTAKA

Adiguna, M. A. (2023). BRAND REPUTATION MONITORING SYSTEM BASED ON SENTIMENT ANALYSIS USING THE K-NEAREST NEIGHBOR METHOD. *JUPIK: Jurnal Penelitian Ilmu komputer*, *1*(1).

Adilazuarda, M. F., Cahyawijaya, S., Winata, G. I., Fung, P., & Purwarianti, A. (2024). *IndoRobusta: Towards Robustness Against Diverse Code-Mixed Indonesian Local Languages*. 25–34. https://doi.org/10.18653/v1/2022.sumeval-1.5

Akbari, M., Foroudi, P., Fashami, R. Z., Mahavarpour, N., & Khodayari, M. (2022). Let us talk about something: The evolution of e-WOM from the past to the future. *Journal of Business Research*, *149*, 663–689.

Alammar, J., & Grootendorst, M. (2024). *Hands-on large language models: language understanding and generation*. " O'Reilly Media, Inc."

Alharbi, M., Yin, J., & Wang, H. (2023). Surveying the landscape: compound methods for aspect-based sentiment analysis. *Australasian Database Conference*, 101–115.

Ali, T., Omar, B., & Soulaimane, K. (2022). Analyzing tourism reviews using an LDA topic-based sentiment analysis approach. *MethodsX*, *9*(November), 101894. https://doi.org/10.1016/j.mex.2022.101894

Alotaibi, A., & Nadeem, F. (2025). An Unsupervised Integrated Framework for Arabic Aspect-Based Sentiment Analysis and Abstractive Text Summarization of Traffic Services Using Transformer Models. *Smart Cities*, *8*(2). https://doi.org/10.3390/smartcities8020062

Alshaikh, K. A., Almatrafi, O. A., & Abushark, Y. B. (2024). BERT-Based Model for Aspect-Based Sentiment Analysis for Analyzing Arabic Open-Ended Survey Responses: A Case Study. *IEEE Access*, *12*, 2288–2302. https://doi.org/10.1109/ACCESS.2023.3348342

Ambiya, M. Z., & Anggriani, S. (2018). Bahasa Indonesia dalam Situs Jejaring Sosial (Kajian Perkembangan Ragam Tidak Baku dalam Dunia Maya). *Prosiding Seminar Nasional Kajian Bahasa, Sastra, dan Pengajarannya V*, 277–281.

Amplayo, R. K., Lee, S., & Song, M. (2018). Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis. *Information Sciences*, *454*, 200–215.

Anoop, V. S., & Asharaf, S. (2019). Aspect-oriented sentiment analysis: A topic modeling-powered approach. *Journal of Intelligent Systems*, *29*(1), 1166–1178.

Anule, A. P., Hambali, M. A., & Ifeoma, C. J. (2022). *Review of sentiment analysis: An hybrid approach*.

Barde, B. V., & Bainwad, A. M. (2017). An overview of topic modeling methods and tools. *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 745–750. https://ieeexplore.ieee.org/abstract/document/8250563/

Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I. W. H., Ng, L. G., Ginhoux, F., & Newell, E. W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*, *37*(1), 38–47. https://doi.org/10.1038/nbt.4314

Berba, P. (2020). *Understanding HDBSCAN and Density-Based Clustering*. https://pberba.github.io/stats/2020/01/17/hdbscan/

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning* (1 ed.). Springer.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

botpenguin.com. (2024). *Topic Modeling: Techniques and Applications*. https://botpenguin.com/glossary/topic-modeling

Bozuyla, M., & Özçift, A. (2022). Developing a fake news identification model with advanced deep languagetransformers for turkish covid-19 misinformation data. *Turkish Journal of Electrical Engineering and Computer Sciences*, *30*(3), 908–926.

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7819 LNAI*(PART 2), 160–172.

https://doi.org/10.1007/978-3-642-37456-2_14

Çeliktuğ, M. F. (2018). Twitter sentiment analysis, 3-way classification: positive, negative or neutral? *2018 IEEE international conference on big data (big data)*, 2098–2103. https://ieeexplore.ieee.org/abstract/document/8621970

Chitty-Venkata, K. T., Mittal, S., Emani, M., Vishwanath, V., & Somani, A. K. (2023). A survey of techniques for optimizing transformer inference. *Journal of Systems Architecture*, *144*, 102990.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Dang, C. N., Moreno-García, M. N., & De La Prieta, F. (2021). Hybrid Deep Learning Models for Sentiment Analysis. *Complexity*, *2021*. https://doi.org/10.1155/2021/9986920

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*(Mlm), 4171–4186. https://aclanthology.org/N19-1423/

Elov, B. B., Khamroeva, S. M., & Xusainova, Z. Y. (2023). The pipeline processing of NLP. *E3S Web of Conferences*, *413*, 3011.

Erkan, I., & Evans, C. (2016). The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in human behavior*, *61*, 47–55.

Fanni, S. C., Febi, M., Aghakhanyan, G., & Neri, E. (2023). Natural language processing. In *Introduction to artificial intelligence* (hal. 87–99). Springer.

Garrido-Merchan, E. C., Gozalo-Brizuela, R., & Gonzalez-Carvajal, S. (2023). Comparing BERT against traditional machine learning models in text classification. *Journal of Computational and Cognitive Engineering*, *2*(4), 352–356.

Geetha, M. P., & Karthika Renuka, D. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. *International Journal of Intelligent Networks*, *2*(June), 64–69.

https://doi.org/10.1016/j.ijin.2021.06.005

Giachanou, A., & Crestani, F. (2016). Opinion retrieval in Twitter using stylistic variations. *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, 1077–1079. https://dl.acm.org/doi/abs/10.1145/2851613.2851922

González-Carvajal, S., & Garrido-Merchán, E. C. (2020). Comparing BERT against traditional machine learning text classification. *arXiv preprint arXiv:2005.13012*.

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. http://arxiv.org/abs/2203.05794

Grootendorst, M. (2025). *BERTopic*. https://maartengr.github.io/BERTopic/index.html

Gulati, A., Qin, J., Chiu, C. C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Wu, Y., & Pang, R. (2020). Conformer: Convolution-augmented transformer for speech recognition. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2020-Octob*, 5036–5040. https://doi.org/10.21437/Interspeech.2020-3015

Gunasekaran, K. P. (2023). Exploring sentiment analysis techniques in natural language processing: A comprehensive review. *arXiv preprint arXiv:2305.14842*. https://arxiv.org/abs/2305.14842

Gupta, N., & Agrawal, R. (2020). Application and techniques of opinion mining. In *Hybrid Computational Intelligence: Challenges and Applications*. INC. https://doi.org/10.1016/B978-0-12-818699-2.00001-9

Haq, B., Daudpota, S. M., Imran, A. S., Kastrati, Z., & Noor, W. (2023). A Semi-Supervised Approach for Aspect Category Detection and Aspect Term Extraction from Opinionated Text. *Computers, Materials and Continua*, *77*(1), 115–137. https://doi.org/10.32604/cmc.2023.040638

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, *18*(1), 38–52.

Hoang, M., Bihorac, O. A., & Rouces, J. (2019). Aspect-Based Sentiment Analysis Using BERT. *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, 187–196.

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*, *30*(4), 330–338. https://doi.org/10.1016/j.jksues.2016.04.002

Inaam ul Haq, M., Li, Q., & Hou, J. (2022). Analyzing the research trends of IoT using topic modeling. *The Computer Journal*, *65*(10), 2589–2609.

Indriati, R. (2024). Analisis Sentimen Undang-Undang Cipta Kerja pada Twitter. In *Journal GEEJ* (Vol. 7, Nomor 2). PT. Literasi Nusantara Abadi Grup.

Jayadianti, H., Kaswidjanti, W., Utomo, A. T., Saifullah, S., Dwiyanto, F. A., & Drezewski, R. (2022). Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN. *ILKOM Jurnal Ilmiah*, *14*(3), 348–354. https://doi.org/10.33096/ilkom.v14i3.1505.348-354

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2018). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. In *Multimedia Tools and Applications* (Vol. 78, Nomor 11). https://doi.org/10.1007/s11042-018-6894-4

Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* (3rd ed.). https://web.stanford.edu/~jurafsky/slp3/

Kaur, A., & Wallace, J. R. (2024). Moving Beyond LDA: A Comparison of Unsupervised Topic Modelling Techniques for Qualitative Data Analysis of Online Communities. *arXiv preprint arXiv:2412.14486*.

Khalilur, M., & Khan, R. (2021). *Electronic Word of Mouth (E-WOM) for Apparel Merchandising*. *November*. https://www.researchgate.net/publication/356568298

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in Vision: A Survey. *ACM Comput. Surv.*, *54*(10s). https://doi.org/10.1145/3505244

Kim, E., Sung, Y., & Kang, H. (2014). Brand followers' retweeting behavior on Twitter: How brand relationships influence brand electronic word-of-mouth. *Computers in Human Behavior*, *37*, 18–25.

Kusumawati, N., Suhandi, A. T., & Purwanegara, M. S. (2024). Why do customers decide to complain on social media? And how to produce customer forgiveness. *Review of Integrative Business and Economics Research*, *13*(3), 235–254. http://gmp-riber.com/uploads/3/4/9/8/34980536/riber_13-3_17_s23-084_235-254.pdf

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th international conference on World wide web*, 591–600.

Levine, Y., Wies, N., Sharir, O., Cohen, N., & Shashua, A. (2022). *Chapter 7 - Tensors for deep learning theory: Analyzing deep learning architectures via tensorization* (Y. B. T.-T. for D. P. Liu (ed.); hal. 215–248). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-824447-0.00013-3

Li, C., & Hu, X. (2025). Medical Artificial Intelligence in Scholarly and Public Perspective : BERTopic-Based Analysis of Topic-Sentiment Collaborative Mining. *Data Science and Informetrics*. https://doi.org/10.1016/j.dsim.2025.05.001

Li, D. (2024). Applying Topic Modeling to Literary Analysis : A Review. *Digital Studies in Language and Literature*, *1*, 113–141. https://doi.org/10.1515/dsll-2024-0010

Mahfudiyah, N., & Alamsyah, A. (2023). Understanding User Perception of Ride-Hailing Services Sentiment Analysis and Topic Modelling using IndoBERT and BERTopic. *2023 International Conference on Digital Business and Technology Management, ICONDBTM 2023*, *July*, 1–6. https://doi.org/10.1109/ICONDBTM59210.2023.10327320

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093–1113.

Mite-Baidal, K., Delgado-Vera, C., Solís-Avilés, E., Espinoza, A. H., Ortiz-Zambrano, J., & Varela-Tapia, E. (2018). Sentiment analysis in education domain: A systematic literature review. *Communications in Computer and*

Information Science, *883*, 285–297. https://doi.org/10.1007/978-3-030-00940-3_21

Mittal, A., & Modi, A. (2021). ReCAM@ IITK at SemEval-2021 task 4: BERT and ALBERT based ensemble for abstract word prediction. *arXiv preprint arXiv:2104.01563*.

Moore, S. G., & Lafreniere, K. C. (2020). How online word-of-mouth impacts receivers. *Consumer Psychology Review*, *3*(1), 34–59.

Movahedi, F., Padman, R., & Antaki, J. F. (2020). Limitations of ROC on imbalanced data: Evaluation of LVAD mortality risk scores. arXiv 2020. *arXiv preprint arXiv:2010.1625*.

Murthy, D. (2024). Sociology of Twitter/X: Trends, Challenges, and Future Research Directions. *Annual Review of Sociology*, *50*(1), 169–190. https://doi.org/10.1146/annurev-soc-031021-035658

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, *11*(1). https://doi.org/10.1007/s13278-021-00776-6

Noh, J. S., Jang, H. Y., & Jeon, S. S. (2021). *The Influence of e-WOM Information Characteristics on Learning Trust and e-WOM Intention Among Online Learning*. *8*(2), 134–150. https://doi.org/10.33168/LISS.2021.0208

Ocal, A. (2024). Perceptions of the Future of AI on Social Media: A Topic Modeling and Sentiment Analysis Approach. *IEEE Access*, *12*(December). https://doi.org/10.1109/ACCESS.2024.3510526

Ochai Emmanuel. (2024). *Apa yang dimaksud dengan DM di Twitter? Pesan Instan di Aplikasi X*. tweeteraser.com. https://www.tweeteraser.com/id/resources/what-is-a-dm-on-twitter-instant-messaging-on-the-x-app/

Özçift, A., Akarsu, K., Yumuk, F., & Söylemez, C. (2021). Advancing natural language processing (NLP) applications of morphologically rich languages with bidirectional encoder representations from transformers (BERT): an empirical case study for Turkish. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, *62*(2), 226–238.

Poongodi, M., Nguyen, T. N., Hamdi, M., & Cengiz, K. (2021). Global cryptocurrency trend prediction using social media. *Information Processing & Management*, *58*(6), 102708.

Purwitasari, A. (2024). Word Of Mouth Sebagai Alat Pemasaran Efektif: Tinjauan Literature Empiris. *Jurnal Akademik Ekonomi Dan Manajemen*, *1*(4), 493–507.

Puspita, E., Shiddieq, D. F., & Roji, F. F. (2024). Pemodelan Topik pada Media Berita Online Menggunakan Latent Dirichlet Allocation (Studi Kasus Merek Somethinc): Topic Modeling on Online News Media Using Latent Diriclet Allocation (Case Study Somethinc Brand). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(2), 481–489.

Putri, N. R., Rizan, M., & Febrilia, I. (2024). Pengaruh Electronic Word of Mouth terhadap Niat Pembelian Ulang dengan Citra Merek dan Kepercayaan Merek sebagai Variabel Mediasi pada Pengguna E-commerce di Jakarta. *Jurnal Masharif Al-Syariah: Jurnal Ekonomi Dan Perbankan Syariah*, *9*(3).

Qi, J., Zhang, Z., Jeon, S., & Zhou, Y. (2016). Mining customer requirements from online reviews: A product improvement perspective. *Information & Management*, *53*(8), 951–963. https://www.sciencedirect.com/science/article/pii/S0378720616300581

Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining*, *13*(1), 1–14. https://doi.org/10.1007/s13278-023-01030-x

Rahman, A. U., Alsenani, Y., Zafar, A., Ullah, K., Rabie, K., & Shongwe, T. (2024). Enhancing heart disease prediction using a self-attention-based transformer model. *Scientific Reports*, *14*(1), 514.

Ramadhan, A., Hidayanto, A. N., Evik, C. S., Rizkiandini, N., Rahimullah, N. A., Muthiah, R. H., Anggreainy, M. S., & Phusavat, K. (2022). Factors affecting the continuation to use and e-WOM intention of online library resources by university students: A study in Indonesia. *The Journal of Academic Librarianship*, *48*(6), 102592.

Rani, S., & Kumar, M. (2021). Topic modeling and its applications in materials science and engineering. *Materials Today: Proceedings*, *45*, 5591–5596.

https://doi.org/10.1016/j.matpr.2021.02.313

Rantung, V. P. (2023). *Teknik-Teknik Pemrosesan Bahasa Alami (NLP)*. Lakeisha.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 3982–3992. https://doi.org/10.18653/v1/d19-1410

Rijcken, E. (2023). *C_v Topic Coherence Explained*. https://towardsdatascience.com/cᵥ-topic-coherence-explained-fc70e2a85227/

Riyanto, A. D. (2025). *Hootsuite (We are Social): Data Digital Indonesia 2025*. https://andi.link/hootsuite-we-are-social-data-digital-indonesia-2025/

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *Proceedings of the eighth ACM international conference on Web search and data mining*, 399–408. https://dl.acm.org/doi/abs/10.1145/2684822.2685324

Rothman, D., & Gulli, A. (2021). *Transformers for natural language processing* (Vol. 267). Packt Publishing.

Saeed, M. (2023). *A Gentle Introduction to Positional Encoding in Transformer Models, Part 1*. machinelearningmastery.com. https://machinelearningmastery.com/a-gentle-introduction-to-positional-encoding-in-transformer-models-part-1/

Salsabila, N., Winatmoko, Y., Septiandri, A., & Jamal, A. (2018). *Colloquial Indonesian Lexicon*. https://doi.org/10.1109/IALP.2018.8629151

Sarayar, M. C. L., Soepeno, D., & Raintung, M. C. (2021). PENGARUH E-WOM, HARGA DAN KUALITAS PRODUCT TERHADAP MINAT BELI â€ œFOLCIS PUDDINGâ€ MANADO. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi*, *9*(4), 294–303.

Setiana, R., & Marlien, R. . (2021). Niat Beli Ulang: E-Wom, Celebrity Endorse Dan Citra Merek. *Proceeding SENDIU*, 978–979.

Sharifani, K., & Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. *World Information Technology and Engineering Journal*, *10*(07), 3897–3904.

Sibitenda, H., Member, G. S., Diattara, A. W. A., & Ba, C. (2024). *Extracting Semantic Topics About Development in Africa From Social Media*. *12*(October), 142343–142359.

Sihombing, Y. O., & Situmorang, N. V. (2024). Prediksi Sentimen Pada Teks Media Sosial Corporate University Menggunakan RoBERTa. *Prosiding PITNAS Widyaiswara*, *1*, 302–316.

Siregar, A. I. (2024). Studi Literature: Hubungan E-WOM dan Keputusan Pembelian. *Jurnal Ilmu Multidisplin*, *3*(1), 50–58. https://doi.org/10.38035/jim.v3i1.547

Siswanto, A. T., Muhima, R. R., & Wardhana, S. R. (2022). Analisa Sentimen Review Produk Handphone Pada Situs Amazon Menggunakan Pendekatan Lexicon Berdasarkan Sentiwordnet. *KERNEL: Jurnal Riset Inovasi Bidang Informatika dan Pendidikan Informatika*, *3*(1), 14–19. https://doi.org/10.31284/j.kernel.2022.v3i1.1928

Srivastava, A., & Kumar, V. (2021). Hotel attributes and overall customer satisfaction: What did COVID-19 change? *Tourism Management Perspectives*, *40*, 100867.

Sulistyo, D., Ahda, F., & Fitria, V. A. (2021). Epistomologi dalam Natural Language Processing. *Jurnal Inovasi Teknologi dan Edukasi Teknik*, *1*(9), 652–664. https://doi.org/10.17977/um068v1i92021p652-664

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, *27*.

Tala, F. (2003). *A study of stemming effects on information retrieval in Bahasa Indonesia*.

Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Applied Sciences (Switzerland)*, *13*(7). https://doi.org/10.3390/app13074550

Thiodorus, G., Prasetia, A., Ardhani, L. A., & Yudistira, N. (2021). Klasifikasi citra makanan/non makanan menggunakan metode Transfer Learning dengan model Residual Network. *Teknologi: Jurnal Ilmiah Sistem Informasi*, *11*(2), 74–83.

Ting, K. M. (2011). Confusion Matrix. In C. Sammut & G. I. Webb (Ed.), *Encyclopedia of Machine Learning* (hal. 209). Springer US. https://doi.org/10.1007/978-0-387-30164-8_157

Titiana, T., & Bangkalang, D. H. (2023). Analisis Dan Penerapan Topic Modeling Pada Judul Tugas Akhir Mahasiswa Menggunakan Metode Latent Dirichlet Allocation (LDA). *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, *8*(4), 1275–1287.

Umbara, F. W. (2021). User Generated Content di Media Sosial Sebagai Strategi Promosi Bisnis. *Jurnal Manajemen Strategi dan Aplikasi Bisnis*, *4*(2), 572–581. http://ejournal.imperiuminstitute.org/index.php/JMSAB/article/view/366

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582. https://doi.org/https://doi.org/10.1016/j.is.2020.101582

Wang, L., & Carvalho, L. (2024). Multiclass ROC. *arXiv preprint arXiv:2404.13147*.

Wayasti, R. A., & Surjandari, I. (2018). Mining Customer Opinion for Topic Modeling Purpose: Case Study of Ride-Hailing Service Provider. *2018 6th International Conference on Information and Communication Technology (ICoICT)*, 305–309. https://ieeexplore.ieee.org/abstract/document/8528751

Weng, L. (2018). *Attention? Attention!* lilianweng.github.io. https://lilianweng.github.io/posts/2018-06-24-attention/

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. In K.-F. Wong, K. Knight, & H. Wu (Ed.), *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing* (hal. 843–857). Association for

Computational Linguistics. https://aclanthology.org/2020.aacl-main.85

Wu, C., Wu, F., Huang, Y., & Xie, X. (2023). Personalized news recommendation: Methods and challenges. *ACM Transactions on Information Systems*, *41*(1), 1–50.

X Help Center. (2025a). *About Bookmarks*. https://help.x.com/en/using-x/bookmarks

X Help Center. (2025b). *About Direct Messages*. https://help.x.com/en/using-x/direct-messages

X Help Center. (2025c). *Communities on X*. https://help.x.com/en/using-x/communities

X Help Center. (2025d). *How to like a post*. https://help.x.com/en/using-x/liking-posts-and-moments

X Help Center. (2025e). *How to use X Lists*. https://help.x.com/en/using-x/x-lists

X Help Center. (2025f). *New user FAQ*. https://help.x.com/en/resources/new-user-faq

X Help Center. (2025g). *Repost FAQs*. https://help.x.com/en/using-x/repost-faqs

Yaakub, M. R., Latiffi, M. I. A., & Zaabar, L. S. (2019). A review on sentiment analysis techniques and applications. *IOP conference series: materials science and engineering*, *551*(1), 12070.

Zaini, A., Patrajaya, R., Noor, W., & Iksannudin, M. (2024). Optimalisasi Penggunaan Media Sosial dalam Meningkatkan Efektivitas dan Aksesibilitas Penyebaran Informasi di KUA Kecamatan Jekan Raya. *Jurnal Ilmiah Pengabdian Kepada Masyarakat (Nyiur-Dimas)*, *4*(2), 86–95. https://ejournal.iain-manado.ac.id/index.php/nyiur/article/view/1165

Zhao, J., He, H., Zhao, X., & Lin, J. (2022). Modeling and simulation of microblog-based public health emergency-associated public opinion communication. *Information Processing & Management*, *59*(2), 102846. https://www.sciencedirect.com/science/article/pii/S0306457321003186

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI conference on artificial intelligence*, *35*(12), 11106–11115.