

**KOMPARASI ANALISIS SENTIMEN PUBLIK DENGAN
MODEL *LONG SHORT-TERM MEMORY* (LSTM) DAN
*INDONESIA BIDIRECTIONAL ENCODER
REPRESENTATIONS FROM TRANSFORMERS* (INDOBERT)
(STUDI KASUS : PROGRAM MSIB)**



SKRIPSI

Diajukan untuk memenuhi sebagian syarat untuk memperoleh gelar Sarjana Pendidikan pada Program Studi Pendidikan Sistem dan Teknologi Informasi

Alya Sahrani

2102985

**PROGRAM STUDI S1
PENDIDIKAN SISTEM DAN TEKNOLOGI INFORMASI
KAMPUS UPI DI PURWAKARTA
UNIVERSITAS PENDIDIKAN INDONESIA
2025**

LEMBAR HAK CIPTA

KOMPARASI ANALISIS SENTIMEN PUBLIK DENGAN MODEL *LONG SHORT-TERM MEMORY* (LSTM) DAN *INDONESIA BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS* (INDOBERT) (STUDI KASUS : PROGRAM MSIB)

Oleh

Alya Sahrani

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Pendidikan Sistem dan Teknologi Informasi Program Studi Pendidikan Sistem dan Teknologi Informasi Kampus Daerah Purwakarta

© Alya Sahrani
Universitas Pendidikan Indonesia

Agustus 2025

Hak Cipta dilindungi Undang-Undang. Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan cetak ulang, fotokopi, atau cara lainnya tanpa izin dari penulis.

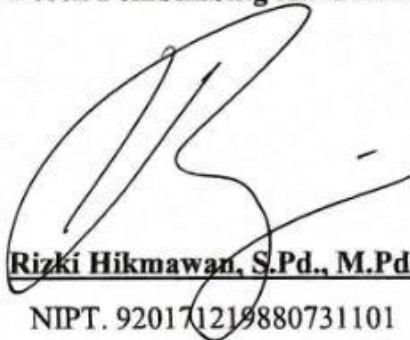
LEMBAR PENGESAHAN

**KOMPARASI ANALISIS SENTIMEN PUBLIK DENGAN MODEL *LONG SHORT-TERM MEMORY* (LSTM) DAN *INDONESIA BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS* (INDOBERT)
(STUDI KASUS : PROGRAM MSIB)**

ALYA SAHRANI

2102985

**Disetujui dan disahkan oleh:
Dosen Pembimbing Akademik**



Rizki Hikmawan, S.Pd., M.Pd.
NIPT. 920171219880731101

**Mengetahui,
Ketua Program Studi
Pendidikan Sistem dan Teknologi Informasi**



Ir. Nuur Wachid Abdul Majid, S.Pd., M.Pd., IPM., ASEAN Eng.
NIP. 920171219910625101

**KOMPARASI ANALISIS SENTIMEN PUBLIK DENGAN MODEL *LONG SHORT-TERM MEMORY* (LSTM) DAN *INDONESIA BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS* (INDOBERT)
(STUDI KASUS : PROGRAM MSIB)**

Alya Sahrani

2102985

ABSTRAK

Penelitian ini bertujuan membandingkan performa model *Long Short-Term Memory* (LSTM) dan *Indonesian Bidirectional Encoder Representations from Transformers* (IndoBERT) dalam analisis sentimen opini publik mengenai Program Magang dan Studi Independen Bersertifikat (MSIB). Di era digital, media sosial menjadi sumber utama opini publik, namun teks yang dihasilkan cenderung tidak terstruktur, penuh sarkasme, *code switching*, serta bahasa informal sehingga menimbulkan tantangan bagi model konvensional. Penelitian ini menggunakan pendekatan kuantitatif deskriptif komparatif dengan data opini publik yang dikumpulkan melalui crawling dari media sosial X sebanyak 3.593 data. Setelah *preprocessing*, distribusi sentimen menunjukkan 61,5% negatif dan 38,5% positif. Model LSTM dan IndoBERT kemudian dilatih dan dievaluasi menggunakan metrik klasifikasi. Hasil pengujian menunjukkan LSTM mencapai akurasi 85%, presisi 91%, *recall* 83%, dan *F1-score* 87%. Sementara itu, IndoBERT secara signifikan unggul dengan akurasi 95%, presisi 95%, *recall* 97%, dan *F1-score* 96%. Hal ini membuktikan IndoBERT lebih efektif dalam menangani teks kompleks khas media sosial Indonesia.

Kata Kunci: Analisis Sentimen, LSTM, IndoBERT, Program MSIB, *Machine Learning*.

**COMPARATIVE ANALYSIS OF PUBLIC SENTIMENT USING LONG
SHORT-TERM MEMORY (LSTM) AND INDONESIA BIDIRECTIONAL
ENCODER REPRESENTATIONS FROM TRANSFORMERS
(INDOBERT) MODELS: A CASE STUDY ON THE MSIB PROGRAM**

Alya Sahrani

2102985

ABSTRACT

This study aims to compare the performance of Long Short-Term Memory (LSTM) and Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT) in sentiment analysis of public opinion regarding the Magang dan Studi Independen Bersertifikat (MSIB) program. In the digital era, social media has become a primary source of public opinion; however, the texts are often unstructured, containing sarcasm, code-switching, and informal language, which pose challenges for conventional models. This research adopts a quantitative descriptive-comparative approach using 3,593 public opinion data collected through crawling from the social media platform X. After preprocessing, sentiment distribution indicated 61.5% negative and 38.5% positive opinions. Both LSTM and IndoBERT models were trained and evaluated using classification metrics. The experimental results revealed that LSTM achieved 85% accuracy, 91% precision, 83% recall, and an F1-score of 87%. In contrast, IndoBERT significantly outperformed LSTM with 95% accuracy, 95% precision, 97% recall, and an F1-score of 96%. These findings demonstrate that IndoBERT is more effective in handling complex text characteristics typical of Indonesian social media.

Keywords: *Sentiment Analysis, LSTM, IndoBERT, MSIB Program, Machine Learning.*

DAFTAR ISI

LEMBAR HAK CIPTA.....	i
LEMBAR PENGESAHAN	ii
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI.....	v
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Penelitian	1
1.2 Rumusan Masalah	4
1.3 Ruang Lingkup Penelitian.....	4
1.4 Tujuan Penelitian	5
1.5 Manfaat penelitian.....	5
BAB II KAJIAN PUSTAKA.....	7
2.1 Analisis Sentimen	7
2.2 Magang dan Studi Independen Bersertifikat (MSIB)	8
2.3 NLP (<i>Natural Language Processing</i>)	8
2.4 <i>Machine Learning</i>	9
2.5 <i>Deep Learning</i>	10
2.6 LSTM (<i>Long Short-term Memory</i>)	12
2.7 BERT (<i>Bidirectional Encoder Representations From Transformers</i>)..	13
2.8 IndoBERT (<i>Indonesian Bidirectional Encoder Representations from Transformer</i>)	15

2.9	Metrik Evaluasi	16
2.10	Penelitian Terdahulu	18
BAB III METODE		21
3.1	Jenis Penelitian.....	21
3.2	Desain Penelitian.....	21
3.3	Prosedur Penelitian.....	22
3.3.1	<i>Data Collection</i>	23
3.3.2	<i>Text Preprocessing</i>	23
3.3.3	<i>Labelling</i>	25
3.3.4	<i>Modelling</i>	25
3.3.5	<i>Model Evaluation</i>	26
3.4	Populasi dan Sampel	26
3.5	Instrumen Penelitian.....	26
3.6	Teknik Pengumpulan Data	27
3.7	Teknik Analisis Data.....	28
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Hasil	30
4.1.1	Karakteristik Data dan Hasil Tahapan Pra-pemodelan	30
4.1.2	Hasil Performa Model LSTM	40
4.1.3	Hasil Performa Model IndoBERT	43
4.1.4	Perbandingan Performa model LSTM dan IndoBERT	47
4.2	Pembahasan.....	49
4.2.1	Analisis Komparatif Model.....	49
4.2.2	Rekomendasi Optimasi Model LSTM	50
BAB V SIMPULAN DAN SARAN		52
4.1	Simpulan	52

4.2	Saran.....	53
DAFTAR PUSTAKA.....		54
LAMPIRAN.....		59

DAFTAR TABEL

Tabel 2. 1 Tabel <i>Confusion Matrix</i>	16
Tabel 2. 2 Penelitian Terdahulu	18
Tabel 3. 1 Tabel Indikator.....	28
Tabel 4. 1 Hasil Pengujian Model LSTM	42
Tabel 4. 2 Hasil Pengujian Model IndoBERT	46
Tabel 4. 3 Perbandingan Hasil Pengujian Model LSTM dan IndoBERT	47

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur LSTM	12
Gambar 2. 2 Arsitektur BERT	13
Gambar 3. 1 Flowchart Prosedur Penelitian	22
Gambar 4. 1 Hasil Pengumpulan Data melalui Teknik Crawling.....	31
Gambar 4. 2 Karakteristik Data	32
Gambar 4. 3 Hasil <i>Data Cleaning</i>	34
Gambar 4. 4 Hasil <i>Case Folding</i>	35
Gambar 4. 5 Hasil <i>Data Normalization</i>	35
Gambar 4. 6 Hasil <i>Stopword Removal</i>	36
Gambar 4. 7 Hasil <i>Tokenized</i>	37
Gambar 4. 8 Hasil <i>Stemming</i>	38
Gambar 4. 9 Proporsi Kategori Sentimen	39
Gambar 4. 10 Kurva Akurasi Pelatihan dan Validasi Model LSTM.....	40
Gambar 4. 11 <i>Confusion Matrix</i> Pengujian.....	41
Gambar 4. 12 Kurva Akurasi Pelatihan dan Validasi Model IndoBERT	44
Gambar 4. 13 <i>Confusion Matrix</i> Pengujian Model IndoBERT.....	45

DAFTAR LAMPIRAN

Lampiran 1. 1 Surat Keputusan Pembimbing Skripsi.....	59
Lampiran 1. 2 Hasil Pengecekan Plagiarisme.....	62
Lampiran 1. 3 Kartu Bimbingan Skripsi.....	63
Lampiran 2. 1 <i>Source Code</i> Model LSTM.....	64
Lampiran 2. 2 Akurasi dan <i>Loss</i> Model LSTM per <i>Epoch</i>	67
Lampiran 2. 3 <i>Source Code</i> Model IndoBERT.....	68
Lampiran 2. 4 Akurasi dan <i>Loss</i> Model IndoBERT per <i>Epoch</i>	71

DAFTAR PUSTAKA

- ADAM, A. (2023). *Analisis Sentimen Peserta Magang Studi Independen Bersertifikat (Msib) Kemendikbud Pada Twitter Menggunakan Naïve Bayes*. <https://repository.nusaputra.ac.id/id/eprint/912/>
- Bahaweres, R. B., Agustian, F., Hermadi, I., Suroso, A. I., & Arkeman, Y. (2020). Software defect prediction using neural network based smote. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2020-Octob(June 2021), 71–76. <https://doi.org/10.23919/EECSI50503.2020.9251874>
- CAHYA, N. (2023). *ANALISIS SENTIMEN MENGENAI PROGRAM MSIB PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER*. 87(1,2), 149–200, <https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/167638/341506.pdf?sequence=1&isAllowed=y%0Ahttps://repositorio.ufsm.br/bitstream/handle/1/8314/LOEBLEIN%2C%20LUCINEIA%20CARLA.pdf?sequence=1&isAllowed=y%0Ahttps://antigo.mdr.gov.br/saneamento/prooes>
- Chairunissa, D., Anisah, A., & Rahmayanti, H. (2024). Membentuk Kesiapan Kerja Mahasiswa Melalui Program Magang. *Journal of Engineering Education and Pedagogy*, 2(1), 1–7. <https://doi.org/10.56855/jeep.v2i1.683>
- Chollet, F. (2022). Deep learning with Python (2a ed) - François Chollet - Manning, outubro 2021, 504 pp. *Interações: Sociedade e as Novas Modernidades*, 42, 113–115. <https://doi.org/10.31211/interacoes.n42.2022.r1>
- Cohen, V. (2010). Natural language processing. *Encyclopedia of Systems Biology*, 2(3), 352–357. <https://doi.org/10.1002/wics.76>
- Delimayanti, M. K., Sari, R., Laya, M., Faisal, M. R., & Pahrul, P. (2021). Pemanfaatan Metode Multiclass-SVM pada Model Klasifikasi Pesan Bencana Banjir di Twitter. *Edu Komputika Journal*, 8(1), 39–47. <https://doi.org/10.15294/edukomputika.v8i1.47858>
- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Naacl-Hlt 2019, Mlm*, 4171–4186. <https://aclanthology.org/N19-1423.pdf>
- Elzayady, H., Badran, K. M., & Salama, G. I. (2020). Arabic Opinion Mining Using Combined CNN - LSTM Models. *International Journal of Intelligent Systems and Applications*, 12(4), 25–36. <https://doi.org/10.5815/ijisa.2020.04.03>
- Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., & Hochreiter, S. (2021). Rainfall-runoff prediction at multiple timescales with a single Long Short-

- Term Memory network. *Hydrology and Earth System Sciences*, 25(4), 2045–2062. <https://doi.org/10,5194/hess-25-2045-2021>
- Hafidhigbal, N. (2024). Implementasi model indobert untuk klasifikasi jenis tindak pidana korupsi berbasis web application. In *Repository.Uinjkt.Ac.Id*. [https://repository.uinjkt.ac.id/dspace/handle/123456789/79800%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/79800/1/NAUFALDI HAFIDHIGBAL-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/79800%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/79800/1/NAUFALDI%20HAFIDHIGBAL-FST.pdf)
- Huljannah, M. (2025). *Analisis Performansi Hate Comments pada Learning Rate 10-1-10-3 dengan Dataset dari X*. 12(2), 2418–2425.
- Husni, N., Dodi Vionanda, Nur Leli, & Syafriandi Syafriandi. (2025). Analisis Sentimen Program MSIB pada Aplikasi X (Twitter) Menggunakan Algoritma Naïve Bayes. *UNP Journal of Statistics and Data Science*, 3(2), 189–196. <https://doi.org/10,24036/ujsds/vol3-iss2/361>
- 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>
- Joseph, T. (2024). Natural Language Processing (NLP) for Sentiment Analysis in Social Media. *International Journal of Computing and Engineering*, 6(2), 35–48. <https://doi.org/10,47941/ijce.2135>
- Juarto, B., & Yulianto. (2023). Indonesian News Classification Using IndoBert. *International Journal of Intelligent Systems and Applications in Engineering*, 11(2), 454–460,
- Kansara, D., & Sawant, V. (2020). *Comparison of Traditional Machine Learning and Deep Learning Approaches for Sentiment Analysis*. 365–377. https://doi.org/10,1007/978-981-15-3242-9_35
- Kholifatullah, B. A. H., & Prihanto, A. (2023). Penerapan Metode Long Short Term Memory Untuk Klasifikasi Pada Hate Speech. *Journal of Informatics and Computer Science (JINACS)*, 04, 292–297. <https://doi.org/10,26740/jinacs.v4n03.p292-297>
- Khotimah, K., Yudistira, F., & Ardiansyah, M. (2024). Efisiensi Deep learning untuk Analisis Data dan Pengambilan Keputusan. *Jurnal Insan Peduli Pendidikan (JIPENDIK)*, 2(2), 79–82.
- Kusnadi, Putra, S. (n.d.). Transfer Learning in Indonesian NLP: The Case of IndoBERT. *International Journal of Artificial Intelligence and Applications*.
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10,1038/nature14539>

- Mariel, W. C. F., Mariyah, S., & Pramana, S. (2018). Sentiment analysis: A comparison of deep learning neural network algorithm with SVM and naïve Bayes for Indonesian text. *Journal of Physics: Conference Series*, 971(1). <https://doi.org/10.1088/1742-6596/971/1/012049>
- Mudding, A. A., & Arifin A Abd Karim. (2022). Analisis Sentimen Menggunakan Algoritma Lstm Pada Media Sosial. *Jurnal Publikasi Ilmu Komputer Dan Multimedia*, 1(3), 181–187. <https://doi.org/10.55606/jupikom.v1i3.517>
- Naquitasia, R. (2022). *Analisis Sentimen Berbasis Aspek Pada Ulasan Tempat Wisata Halal Dengan Deep Learning*. <https://dspace.uui.ac.id/handle/123456789/39202%0Ahttps://dspace.uui.ac.id/bitstream/handle/123456789/39202/18523214.pdf?sequence=1>
- Natzir, S. M., & Jatiprasetya, H. (2025). Prediksi Harga Cryptocurrency Xlm Menggunakan Metode Deep Learning Lstm Dan Gru. *HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi*, 16(1), 49–58. <https://doi.org/10.52972/hoa.vol16no1.p49-58>
- Nisa, R., Amriza, S., & Supriyadi, D. (2021). Komparasi Metode. *Penelitian Ilmu Dan Teknologi Komputer*, 13(2), 130–139.
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711.
- Nurmaini, S., Tondas, A. E., Darmawahyuni, A., Rachmatullah, M. N., Effendi, J., Firdaus, F., & Tutuko, B. (2021). Electrocardiogram signal classification for automated delineation using bidirectional long short-term memory. *Informatics in Medicine Unlocked*, 22, 100507. <https://doi.org/10.1016/j.imu.2020.100507>
- Pradana, L. S. (2024). Analisis Sentimen Masyarakat Media Sosial Twitter terhadap Kinerja Penjabat Gubernur DKI Jakarta Menggunakan model IndoBERT. *Universitas Islam Negeri Syarif Hidayatullah*.
- Prakash, T. N., & Aloysius, A. (2021). Textual Sentiment Analysis using Lexicon Based Approaches. *Annals of the Romanian Society for Cell Biology*, 25(December), 9878–9885. <http://annalsofrscb.ro/index.php/journal/article/view/3734>
- Purnamasari, D. (2023). *Pengantar metode analisis sentimen* (Gunadarma, Ed.; Vol. 1).
- Qadri, M. (2020). Pengaruh Media Sosial Dalam Membangun Opini Publik. *Qaumiyyah: Jurnal Hukum Tata Negara*, 1(1), 49–63. <https://doi.org/10.24239/qaumiyyah.v1i1.4>

- Ramadhan, G. A. (2023). Analisis Sentimen Ulasan Aplikasi Ruangguru dengan Algoritma Long Short Term Memory. In *Repository.Uinjkt.Ac.Id*. <https://repository.uinjkt.ac.id/dspace/handle/123456789/71510>
- Ramadhan, N. R., & Hendrastuty, N. (2024). Perbandingan Algoritma Naïve Bayes dan LSTM untuk Analisis Sentimen Terhadap Opini Masyarakat Tentang Sandwich Generation. *Building of Informatics, Technology and Science (BITS)*, 6(3), 1677–1687. <https://doi.org/10.47065/bits.v6i3.6385>
- Saptadi, N. T. S., Kristiawan, H., Nugroho, A. Y., Rahayu, N., Suwarmiyati, Waseso, B., Intan, I., Khairunnas, Martono, Saputra, P. Y., Sutriawan, Soekarman, Mahatma, K., Yudianto, I., Soleh, O., Sutoyo, Muh. N., Siswoyo, B., & Aliyah. (2025). *Deep Learning Teori, Algoritma, dan Aplikasi* (Issue March).
- Sefuloh, F. (2023). Analisis Sentimen Masyarakat Terhadap Program Magang Bersertifikat Pada Media Sosial Twitter Menggunakan Klasifikasi Naïve Bayes Classifier (NBC). *Doctoral Dissertation, Universitas Muhammadiyah Purwokerto*.
- Setiawan, W. (2021). *Deep Learning menggunakan Convolutional Neural Network: Teori dan Aplikasi* (1st ed.). Media Nusa Creative (MNC Publishing).
- Simon, H. A. (1995). Artificial intelligence: an empirical science. *Artificial Intelligence*, 77(1), 95–127. [https://doi.org/10.1016/0004-3702\(95\)00039-H](https://doi.org/10.1016/0004-3702(95)00039-H)
- Sitohang, N. (2023). Jurnal Sains Informatika Terapan (JSIT). *Penerapan Data Mining Untuk Peringatan Dini Banjir Menggunakan Metode Klastering K-Means*, 2(1), 16–20,
- Srinivas, A. C. M. V., Satyanarayana, Ch., Divakar, Ch., & Sirisha, K. P. (2021). Sentiment Analysis using Neural Network and LSTM. *IOP Conference Series: Materials Science and Engineering*, 1074(1), 012007. <https://doi.org/10.1088/1757-899x/1074/1/012007>
- Suahati, A., Nurrahman, A., & Rukmana, O. (2022). Penggunaan Jaringan Syaraf Tiruan – Backpropagation dalam Memprediksi Jumlah Mahasiswa Baru. *Jurnal Media Teknik Dan Sistem Industri*, 6. <https://doi.org/10.35194/jmtsi.v6i1.1589>
- Syah, M. P., Wardani, A. P., & Idhom, M. (2025). *Perbandingan Representasi Teks Tf-Idf Dan Bert Terhadap Akurasi*. 5(1), 47–59.
- Syamsuadi, A., Sepriyani, H., Endrini, S., & Febriani, A. (2022). Implementasi Kebijakan Merdeka Belajar Kampus Merdeka di Universitas Abdurrah pada Program Magang Mahasiswa. *Edukatif: Jurnal Ilmu Pendidikan*, 4(1), 1341–1348. <https://doi.org/10.31004/edukatif.v4i1.2203>

- Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavaf, N., & Fox, E. A. (2021). *Natural Language Processing Advancements By Deep Learning: A Survey*. 1–23. <http://arxiv.org/abs/2003.01200>
- Vidya Chandradev, I Made Agus Dwi Suarjaya, & I Putu Agung Bayupati. (2023). Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT. *Jurnal Buana Informatika*, 14(02), 107–116. <https://doi.org/10.24002/jbi.v14i02.7244>
- Witanto, K. S., Sanjaya ER, N. A., Karyawati, A. E., Kadyanan, I. G. A. G. A., Suhartana, I. K. G., & Astuti, L. G. (2022). Implementasi LSTM Pada Analisis Sentimen Review Film Menggunakan Adam Dan RMSprop Optimizer. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 10(4), 351. <https://doi.org/10.24843/jlk.2022.v10,i04.p05>
- Zhao, Z., Liu, W., & Wang, K. (2023). Research on sentiment analysis method of opinion mining based on multi-model fusion transfer learning. *Journal of Big Data*, 10(1). <https://doi.org/10.1186/s40537-023-00837-x>