

**PERBANDINGAN KINERJA MODEL INDOBERT DAN
INDOBERTWEET PADA DETEKSI EMOSI
TWEET BAHASA INDONESIA**

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun oleh:

MOCH. HILU MAULIDY

NIM 09020622034

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
2026**

PERNYATAAN KEASLIAN

PERNYATAAN KEASLIAN

Saya yang bertandatangan dibawah ini,

Nama : Moch. Hilu Maulidy

NIM : 09020622034

Prodi : Sistem Informasi

Angkatan : 2022

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: "PERBANDINGAN KINERJA MODEL INDOBERT DAN INDOBERTWEET PADA DETEKSI EMOSI TWEET BAHASA INDONESIA". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 23 Juni 2026

Yang menyatakan,



(Moch. Hilu Maulidy)

NIM. 09020622034

LEMBAR PERSETUJUAN PEMBIMBING

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi Oleh:

NAMA : MOCH HILU MAULIDY

NIM : 09020622034

JUDUL : PERBANDINGAN KINERJA MODEL INDOBERT DAN
INDOBERTWEET PADA DETEKSI EMOSI TWEET BAHASA
INDONESIA

Ini telah diperiksa dan disetujui untuk diujikan.

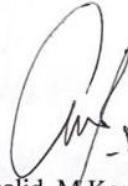
Surabaya, 04 Juni 2026

Dosen Pembimbing 1



(Dr. Eng. Anang Kunaefi, M.Kom)
NIP. 197911132014031001

Dosen Pembimbing 2



(Khalid, M.Kom)
NIP. 197906092014031002

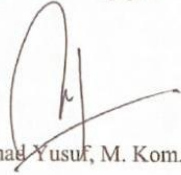
LEMBAR PENGESAHAN TIM PENGUJI

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Moch. Hilu Maulidy ini telah
dipertahankan di depan tim penguji skripsi
di Surabaya, 15 Juni 2026

Mengesahkan,
Dewan penguji

Dosen Penguji I



(Ahmad Yusuf, M. Kom.)

NIP. 199001202014031003

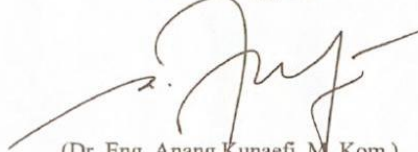
Dosen Penguji II



(Dwi Rolliawati, M.T.)

NIP. 197909272014032001

Dosen Penguji III



(Dr. Eng. Anang Kunaefi, M. Kom.)

NIP. 197911132014031001

Dosen Penguji IV



(Khalid, M. Kom.)

NIP. 197906092014031002

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UTN Suran Ampel Surabaya



(Ascep Saopul Hamdani, M.Pd.)

NIP. 196507312000031002

**LEMBAR PENGESAHAN PERSETUJUAN PUBLIKASI KARYA TULIS
ILMIAH UNTUK KEPENTINGAN AKADEMIS**



**KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN**

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax. 031-8413300
E-Mail: perpus@uinsby.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini,
saya:

Nama : Moch. Hilu Maulidy
NIM : 09020622034
Fakultas/Jurusan : Sains dan Teknologi/Sistem Informasi
E-mail address : emperormh2@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan
UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

☒ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :

“PERBANDINGAN KINERJA MODEL INDOBERT DAN INDOBERTWEET PADA
DETEKSI EMOSI TWEET BAHASA INDONESIA”

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini
Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan,
mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan
menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk
kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama
saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN
Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak
Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 26 Juni 2026

Penulis

(Moch. Hilu Maulidy)

ABSTRAK

PERBANDINGAN KINERJA MODEL INDOBERT DAN INDOBERTTWEET PADA DETEKSI EMOSI TWEET BAHASA INDONESIA

Oleh:

Moch. Hilu Maulidy

Deteksi emosi pada tweet berbahasa Indonesia merupakan salah satu tugas *Natural Language Processing* yang masih menghadapi permasalahan ketidakseimbangan distribusi data (*class imbalance*) sehingga dapat memengaruhi performa model klasifikasi, serta masih terbatasnya penelitian yang membandingkan kinerja IndoBERT dan IndoBERTtweet dengan penerapan teknik *Random Undersampling*. Oleh karena itu, penelitian ini bertujuan untuk menerapkan model IndoBERT dan IndoBERTtweet dalam deteksi emosi pada tweet berbahasa Indonesia, menganalisis pengaruh *Random Undersampling*, serta membandingkan kinerja kedua model. Hasil pengujian menunjukkan bahwa IndoBERT memperoleh *accuracy* sebesar 77,5% pada dataset original dan meningkat 1,2% menjadi 78,7% setelah diterapkan *undersampling*. Sementara itu, IndoBERTtweet memperoleh *accuracy* sebesar 78,6% pada dataset original dan meningkat 2,1% menjadi 80,7% pada dataset hasil *undersampling*, dengan nilai *f1-score* tertinggi sebesar 80,8%. Hasil tersebut menunjukkan bahwa teknik *Random Undersampling* mampu meningkatkan performa model dalam mendeteksi emosi serta mengurangi pengaruh ketidakseimbangan data. Selain itu, IndoBERTtweet menghasilkan performa yang lebih baik dibandingkan IndoBERT karena lebih mampu memahami karakteristik bahasa yang digunakan pada media sosial, sehingga kombinasi IndoBERTtweet dan *Random Undersampling* menjadi pendekatan terbaik dalam penelitian ini.

Kata Kunci: Deteksi Emosi, IndoBERT, IndoBERTtweet, Media Sosial, *RandomUndersampling*

ABSTRACT

PERFORMANCE COMPARISON OF INDOBERT AND INDOBERTTWEET MODELS FOR EMOTION DETECTION IN INDONESIAN LANGUAGE TWEETS

By:
Moch. Hilu Maulidy

Emotion detection in Indonesian-language tweets is one of the Natural Language Processing (NLP) tasks that still faces the challenge of class imbalance, which can affect the performance of classification models. In addition, studies comparing the performance of IndoBERT and IndoBERTweet with the implementation of the Random Undersampling technique remain limited. Therefore, this study aims to implement IndoBERT and IndoBERTweet for emotion detection in Indonesian tweets, analyze the effect of Random Undersampling, and compare the performance of both models. The experimental results show that IndoBERT achieved an accuracy of 77.5% on the original dataset and improved by 1.2% to 78.7% after applying undersampling. Meanwhile, IndoBERTweet achieved an accuracy of 78.6% on the original dataset and improved by 2.1% to 80.7% on the undersampled dataset, with the highest F1-score of 80.8%. These results indicate that the Random Undersampling technique can improve model performance in emotion detection while reducing the impact of data imbalance. Furthermore, IndoBERTweet outperformed IndoBERT because it is better suited to understanding the linguistic characteristics commonly found in social media text. Therefore, the combination of IndoBERTweet and Random Undersampling proved to be the most effective approach for emotion detection in Indonesian-language tweets.

Keywords : Emotion Detection, IndoBERT, IndoBERTweet, Random Undersampling, Social Media

DAFTAR ISI

PERNYATAAN KEASLIAN	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PENGESAHAN TIM PENGUJI.....	iii
LEMBAR PENGESAHAN PERSETUJUAN PUBLIKASI KARYA TULIS ILMIAH UNTUK KEPENTINGAN AKADEMIS	iv
MOTTO	v
UCAPAN TERIMAKASIH	vi
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah	5
1.6 Sistematika Penulisan	5
BAB II KAJIAN PUSTAKA.....	7
2.1 Tinjauan Penelitian Terdahulu.....	7
2.2 Dasar Teori	11
2.2.1 Emosi.....	11
2.2.2 <i>Natural Language Processing</i>	13
2.2.3 <i>Transformer</i>	15
2.2.4 BERT	17
2.2.5 IndoBERT	21
2.2.6 IndoBERTweet	24
2.2.7 <i>Exploratory Data Analysis</i>	26
2.2.8 <i>Data Balancing</i>	27
2.2.9 Evaluasi Model.....	33
2.3 Media Sosial Twitter.....	36
2.4 Integrasi Keilmuan.....	37

BAB III METODOLOGI PENELITIAN	40
3.1 Tahapan Penelitian.....	40
3.1.1 Studi Literatur.....	40
3.1.2 Pengumpulan Data	41
3.1.3 <i>Exploratory Data Analysis</i>	41
3.1.4 Penanganan Data Tidak Seimbang (<i>Data Balancing</i>).....	42
3.1.5 Pembagian Data (<i>Split Data</i>).....	42
3.1.6 Pemodelan	43
3.1.7 Evaluasi Model.....	45
BAB IV HASIL DAN PEMBAHASAN	47
4.1 Pengumpulan Data.....	47
4.2 <i>Exploratory Data Analysis</i> (EDA).....	48
4.3 Penanganan Data Tidak Seimbang (<i>Data Balancing</i>).....	49
4.4 Pembagian Data (<i>Split Data</i>)	50
4.5 Evaluasi Model	51
4.5.1 IndoBERT	52
4.5.2 Indobertweet.....	59
4.5.3 Perbandingan Model IndoBERT dan IndoBERTweet	66
4.6 Analisis Perbandingan dengan Penelitian Sebelumnya.....	71
BAB V KESIMPULAN DAN SARAN	74
5.1 Kesimpulan	74
5.2 Saran	75
DAFTAR PUSTAKA.....	76

DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu	7
Tabel 2. 2 Tabel Confusion Matrix	34
Tabel 3. 1 Distribusi Data Berdasarkan Kelas Emosi (Riccosan et al., 2022).....	42
Tabel 3. 2 Skenario Eksperimen	43
Tabel 4. 1 Distribusi Data	47
Tabel 4. 2 Contoh data yang telah dikumpulkan	47
Tabel 4. 3 Distribusi Data Sebelum dan Sesudah <i>undersampling</i>	50
Tabel 4. 4 Distribusi Label setelah <i>Split Data</i>	51
Tabel 4. 5 Hasil <i>Evaluation Matrix</i> IndoBERT Dataset <i>Original</i>	52
Tabel 4. 6 Hasil <i>Evaluation Matrix</i> Indobert Dataset <i>Undersampling</i>	56
Tabel 4. 7 Hasil <i>Evaluation Matrix</i> IndoBERTweet Dataset <i>Original</i>	59
Tabel 4. 8 Hasil <i>Evaluation Matrix</i> IndoBERTweet Dataset <i>Undersampling</i>	63
Tabel 4. 9 Perbandingan Model IndoBERT dan IndoBERTweet.....	66
Tabel 4. 10 Pengaruh <i>Undersampling</i> terhadap Model	68
Tabel 4.11 Perbandingan Hyperparameter.....	72
Tabel 4.12 Perbandingan Hasil Penelitian.....	72

UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur <i>Transformer</i>	16
Gambar 2. 2 Ilustrasi Arsitektur BERT (<i>Base & Large</i>)	18
Gambar 2. 3 Representasi <i>Input</i> Pada BERT	19
Gambar 2. 4 Ilustrasi <i>Pre-Training</i> Dan <i>Fine-Tuning</i> BERT	21
Gambar 2. 5 Arsitektur Indobert untuk Klasifikasi Teks.....	22
Gambar 2. 6 Ilustrasi Teknik <i>Oversampling</i>	28
Gambar 2. 7 Ilustrasi Teknik <i>Undersampling</i>	30
 Gambar 3. 1 Diagram Alur Penelitian.....	 40
 Gambar 4. 1 <i>Dataset Creation Flow</i>	 48
Gambar 4. 2 Presentase Distribusi Data Pada Tiap Label Emosi	49
Gambar 4. 3 Grafik <i>Training and Validation loss</i> IndoBERT Dataset <i>Original</i> ..	54
Gambar 4. 4 <i>Confusion Matrix</i> IndoBERT Dataset <i>Original</i>	55
Gambar 4. 5 Grafik <i>Training and Validation loss</i> IndoBERT Dataset <i>Undersampling</i>	57
Gambar 4. 6 <i>Confusion Matrix</i> IndoBERT Dataset <i>Undersampling</i>	58
Gambar 4. 7 Grafik <i>Training and Validation loss</i> IndoBERTweet Dataset <i>Original</i>	61
Gambar 4. 8 <i>Confusion Matrix</i> IndoBERTweet Dataset <i>Original</i>	62
Gambar 4. 9 Grafik <i>Training and Validation loss</i> IndoBERTweet Dataset <i>Undersampling</i>	64
Gambar 4. 10 <i>Confusion Matrix</i> IndoBERTweet Dataset <i>Undersampling</i>	65

UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR PUSTAKA

- Ahmad, K.R; Amir, L. S. H. (2025). Pengaruh Media Sosial Terhadap Pola Komunikasi Generasi Z Dalam Interaksi Sosial. *Jurnal Dinamika Sosial Dan Sains*, 2(1), 412–416. <https://doi.org/10.60145/jdss.v2i1.114>
- Devi, D., Biswas, S. K., & Purkayastha, B. (2020). A Review on Solution to Class Imbalance Problem: Undersampling Approaches. *2020 International Conference on Computational Performance Evaluation, ComPE 2020*, 626–631. <https://doi.org/10.1109/ComPE49325.2020.9200087>
- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <https://github.com/tensorflow/tensor2tensor>
- 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, 4171–4186.
- Dey, S., Wasif, S., Tonmoy, D. S., & Sultana, S. (2020). A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews. *International Conference on Contemporary Computing and Applications*, 1–4.
- Geni, L., Yulianti, E., & Sensuse, D. I. (2023a). Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using Bert Language Models. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, 9(3), 746–757. <https://doi.org/10.26555/jiteki.v9i3.26490>
- Geni, L., Yulianti, E., & Sensuse, D. I. (2023b). Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using Bert Language Models. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, 9(3), 746–757. <https://doi.org/10.26555/jiteki.v9i3.26490>
- Gu, S., Wang, F., Patel, N. P., Bourgeois, J. A., & Huang, J. H. (2019). A model for basic emotions using observations of behavior in *Drosophila*. *Frontiers in Psychology*, 10(APR), 1–13. <https://doi.org/10.3389/fpsyg.2019.00781>
- Hairani, H., Widiyaningtyas, T., & Prasetya, D. D. (2024). Addressing Class

- Imbalance of Health Data: a Systematic Literature Review on Modified Synthetic Minority Oversampling Technique (SMOTE) Strategies. *International Journal on Informatics Visualization*, 8(3), 1310–1318. <https://doi.org/10.62527/joiv.8.3.2283>
- Hasanah, U., Soleh, A. M., & Sadik, K. (2024). Effect of *Random Under sampling*, *Oversampling*, and *SMOTE* on the Performance of Cardiovascular Disease Prediction Models. *Jurnal Matematika, Statistika Dan Komputasi*, 21(1), 88–102. <https://doi.org/10.20956/j.v21i1.35552>
- He, H. (2009). Learning from imbalanced data. *Studies in Computational Intelligence*, 807(9), 81–110. https://doi.org/10.1007/978-3-030-04663-7_4
- Indrawati, A., Subagyo, H., Sihombing, A., Wagiyah, & Afandi, S. (2020). Analyzing the Impact of Resampling Method for Imbalanced Data Text in Indonesian. *BACA: Jurnal Dokumentasi Dan Informasi*, 42(2), 133–141.
- Jannah, Y. A. N., & Prasetyo, R. B. (2022). Analisis Sentimen dan Emosi Publik pada Awal Pandemi COVID-19 Berdasarkan Data Twitter dengan Pendekatan Berbasis Leksikon. *Seminar Nasional Official Statistics*, 2022(1), 597–608. <https://doi.org/10.34123/semnasoffstat.v2022i1.1483>
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*. <https://doi.org/10.1186/s40537-019-0192-5>
- Khairani. (2024). *Perbandingan Model Indobert Dan*. https://cdn.ar-raniry.ac.id/aps-test/teknologi_informasi/Pembimbing_di_USK_a.n_Hendri_Tugas_Akhir_Ulfia_Khairani_1908107010068.pdf
- Khairani, U., Mutiawani, V., & Ahmadian, H. (2024). Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(4), 887–894. <https://doi.org/10.25126/jtiik.1148315>
- Khairunnisa, S., Adiwijaya, A., & Faraby, S. Al. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *Jurnal Media Informatika Budidarma*, 5(2), 406. <https://doi.org/10.30865/mib.v5i2.2835>
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural Language

Processing : State of The Art , Current Trends and Challenges Department of Computer Science and Engineering Accendere Knowledge Management Services Pvt . Ltd ., India Abstract. *Sentiment Analysis Has Become One of the Most Profound Research Areas with the Increasing Growth of Social Media on Web. Nowadays, Millions of Users Exchange Their Views, Ideas, Expressions, Feelings, Opinions on Social Media like Twitter and Facebook.* Se, (Figure 1).

- Koto, F., Lau, J. H., & Baldwin, T. (2021). INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization. *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 10660–10668. <https://doi.org/10.18653/v1/2021.emnlp-main.833>
- Krawczyk, B. (2016). Learning from imbalanced data : open challenges and future directions. *Progress in Artificial Intelligence*, 5(4), 221–232. <https://doi.org/10.1007/s13748-016-0094-0>
- Mailoa, F. F. (2021). Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia. *Journal of Information Systems for Public Health*, 6(1), 44. <https://doi.org/10.22146/jisph.44455>
- Masaling, N. A. P., Siswanto, R. R., & Girsang, A. S. (2024a). Indonesian Tweet Emotion Detection Using IndoBERT. *Proceedings of 2024 International Conference on Information Management and Technology, ICIMTech 2024*, 478–482. <https://doi.org/10.1109/ICIMTech63123.2024.10780847>
- Masaling, N. A. P., Siswanto, R. R., & Girsang, A. S. (2024b). Indonesian Tweet Emotion Detection Using IndoBERT. *Proceedings of 2024 International Conference on Information Management and Technology, ICIMTech 2024*, 478–482. <https://doi.org/10.1109/ICIMTech63123.2024.10780847>
- Ningsih, S. (2025). *Deteksi Risiko Depresi Pada Pengguna Media Sosial Indonesia Menggunakan Indobertweet*. 1(1), 21–28.
- Nugraha, K. A. (2021). Analisis Sentimen Berbasis Emoticon pada Komentar Instagram Bahasa Indonesia Menggunakan Naïve Bayes. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(3), 715–721. <https://doi.org/10.28932/jutisi.v7i3.4094>

- Prusa, J., Khoshgoftaar, T. M., Dittman, D. J., & Napolitano, A. (2015). Using *Random Undersampling to Alleviate Class Imbalance on Tweet Sentiment Data. Proceedings - 2015 IEEE 16th International Conference on Information Reuse and Integration, IRI 2015*, 197–202. <https://doi.org/10.1109/IRI.2015.39>
- Putra Perssela, R., Mahendra, R., & Rahmadiani, W. (2022). Pemanfaatan Media Sosial Untuk Efektivitas Komunikasi. *Jurnal Ilmiah Mahasiswa Kuliah Kerja Nyata (JIMAKUKERTA)*, 2(3), 650–656. <https://doi.org/10.36085/jimakukerta.v2i3.4525>
- Qadrini, L., Hikmah, H., & Megasari, M. (2022). Oversampling, Undersampling, Smote SVM dan *Random Forest* pada Klasifikasi Penerima Bidikmisi Sejava Timur Tahun 2017. *Journal of Computer System and Informatics (JoSYC)*, 3(4), 386–391. <https://doi.org/10.47065/josyc.v3i4.2154>
- Radhi, M., Ryan, D., Sitompul, H., Sinurat, S. H., & Indra, E. (2021). ANALISIS BIG DATA DENGAN METODE EXPLORATORY DATA ANALYSIS (EDA) DAN METODE VISUALISASI MENGGUNAKAN JUPYTER NOTEBOOK. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima*, 4(2), 2–6.
- Rahma, I. A., & Suadaa, L. H. (2023). Penerapan Text Augmentation untuk Mengatasi Data yang Tidak Seimbang pada Klasifikasi Teks Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(6), 1329–1340. <https://doi.org/10.25126/jtiik.2023107325>
- Riccosan, Saputra, K. E., Pratama, G. D., & Chowanda, A. (2022). Emotion dataset from Indonesian public opinion. *Data in Brief*, 43, 0–5. <https://doi.org/10.1016/j.dib.2022.108465>
- Rizky, A. S., & Hidayat, E. Y. (2025). Klasifikasi Emosi Pada Teks Bahasa Indonesia Menggunakan IndoBERT. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 10(4), 2963–2970.
- Rossi, A., Lestari, T., Setya Perdana, R., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12), 1718–1724. <http://j-ptiik.ub.ac.id>

- Sailunaz, K., & Alhadj, R. (2019). Emotion and sentiment analysis from Twitter text. *Journal of Computational Science*, 36. <https://doi.org/10.1016/j.jocs.2019.05.009>
- Samudra, G., Anggraeni, A. W., Jamroni, A. R. B., Sarif, A., & Wiyanto, W. (2025). Efektivitas Teknik SMOTE Dalam Meningkatkan Performa Naïve Bayes Deteksi Gangguan Kecemasan Mahasiswa. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 12(3). <https://doi.org/10.35957/jatisi.v12i3.12197>
- Saripuddin, M., Suliman, A., Syarmila Sameon, S., & Jorgensen, B. N. (2021). Random Undersampling on Imbalance Time Series Data for Anomaly Detection. *ACM International Conference Proceeding Series*, (September 2021), 151–156. <https://doi.org/10.1145/3490725.3490748>
- 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, 2017-Decem(Nips)*, 5999–6009. <https://doi.org/10.1201/9781003561460-19>
- Wijaya, W., & Budi, S. (2025). Exploratory Data Analysis Pada Dataset Ulasan Exploratory Data Analysis on McDonald ' s Customer Review Dataset in Australia. *Journal of Smart Technology and Engineering*, 01(01), 58–70.
- 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. *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, ACL-IJCNLP 2020*, 843–857. <https://doi.org/10.18653/v1/2020.aacl-main.85>
- Wongvorachan, T., He, S., & Bulut, O. (2023). A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. *Information (Switzerland)*, 14(1). <https://doi.org/10.3390/info14010054>
- Zulfa, I., & Winarko, E. (2017). Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 11(2), 187. <https://doi.org/10.22146/ijccs.24716>