

Abidin, R., Herawati, A., Ekonomi, F., Islam, B., Abdurrahman, U. K. H., & Pekalongan, W. (2024). ANALISIS SENTIMEN PUBLIK TERHADAP KEBIJAKAN PROGRAM TABUNGAN PERUMAHAN RAKYAT (TAPERA). *Journal of Information System and Computer*, 4(1). <https://doi.org/https://doi.org/10.34001/jister.v4i1.1002>

Ansari, G., Kaur, P., & Saxena, C. (2024). Data Augmentation for Improving Explainability of Hate Speech Detection. *Arabian Journal for Science and Engineering*, 49(3), 3609–3621. <https://doi.org/10.1007/s13369-023-08100-4>

Ardiansyah, Widagdo, A. S., Nuresa Qodri, K., Edi Nugroho Saputro, F., Akbar Rizky, N. P., Informasi, T., Kesehatan Dan Teknologi, F., & Muhammadiyah Klaten, U. (2023). Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT. *Jurnal Fasilkom*, 13. <https://doi.org/10.37859/jf.v13i02.5170>

Atmaja, R., & Yustanti, W. (2021). Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers). *JEISBI*, 02, 55–62. <https://doi.org/10.26740/jeisbi.v2i3.41567>

Beddiar, D. R., Jahan, M. S., & Oussalah, M. (2021). Data expansion using back translation and paraphrasing for hate speech detection. *Online Social Networks and Media*, 24. <https://doi.org/10.1016/j.osnem.2021.100153>

Bencke, L., & Moreira, V. P. (2024). Data augmentation strategies to improve text classification: a use case in smart cities. *Language Resources and Evaluation*, 58(2), 659–694. <https://doi.org/10.1007/s10579-023-09685-w>

Cao, Y., Sun, Z., Li, L., & Mo, W. (2022). A Study of Sentiment Analysis Algorithms for Agricultural Product Reviews Based on Improved BERT Model. *Symmetry*, 14(8). <https://doi.org/10.3390/sym14081604>

Chai, C. P. (2022). Comparison of text preprocessing methods. *Natural Language Engineering*, 29(3), 509–553. <https://doi.org/10.1017/S1351324922000213>

Fajri, F., Tutuko, B., Sukemi, S., Studi, P., Informasi, S., Sains, F., Teknologi, D., Kunci, K., & Bert, : (2022). Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter. *JUSIFO (Jurnal Sistem Informasi)*, 8(2), 71–80. <https://doi.org/10.19109/jusifo.v8i2.13885>

Geni, L., Yulianti, E., & Sensuse, D. I. (2023). Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika (JITEKI)*, 9(3), 746–757. <https://doi.org/10.26555/jiteki.v9i3.26490>

Julianti, O. N., Suarna, N., & Prihartono, W. (2024). PENERAPAN NATURAL LANGUAGE PROCESSING PADA ANALISIS SENTIMEN JUDI ONLINE DI MEDIA SOSIAL TWITTER. *Jurnal Mahasiswa Teknik Informatika*, 8(3). <https://doi.org/10.36040/JATI.V8I3.9613>

Koto, F., & Rahmaningtyas, G. Y. (2017, December). InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs. *Proceedings of the*



Optimasi IndoBERT untuk Analisis Sentimen pada Data Tidak Seimbang dengan Augmentasi Data : Studi Kasus Kebijakan TAPERA dan Dataset

Dina Febriana, Dr. Sri Mulyana, M.Kom.

Universitas Gadjah Mada, 2026 | Diunduh dari <http://etd.repository.ugm.ac.id>

[https://doi-](https://doi.org/10.1109/IALP.2017.8300625)

[org.ezproxy.ugm.ac.id/10.1109/IALP.2017.8300625](https://doi.org/10.1109/IALP.2017.8300625)

- Kusnadi, R., Yusuf, Y., Andriantony, A., Ardian Yaputra, R., & Caintan, M. (2021). ANALISIS SENTIMEN TERHADAP GAME GENSHIN IMPACT MENGGUNAKAN BERT. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 6(2), 122–129. <https://doi.org/10.36341/rabit.v6i2.1765>
- Laurensz, B., & Sedyono, E. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19 (Analysis of Public Sentiment on Vaccination in Efforts to Overcome the Covid-19 Pandemic). *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi* |, 10(2), 118–123. <https://doi.org/10.22146/jnteti.v10i2.1421>
- Li, M., Chen, L., Zhao, J., & Li, Q. (2021). Sentiment analysis of Chinese stock reviews based on BERT model. *Applied Intelligence*. <https://doi.org/10.1007/s10489-020-02101-8/Published>
- Lutfiyanto, M. D., & Setiawan, E. B. (2021). Expansion Feature dengan Word2Vec untuk Analisis Sentimen pada Opini Politik di Twitter dengan Klasifikasi Support Vector Machine, Naïve Bayes, dan Random Forest. *E-Proceeding of Engineering*, 10399–10410.
- Oktariansyah, I. A., Umbara, F. R., & Kasyidi, F. (2024). Klasifikasi Sentimen Untuk Mengetahui Kecenderungan Politik Pengguna X Pada Calon Presiden Indonesia 2024 Menggunakan Metode IndoBert. *Technology and Science (BITS)*, 6(2). <https://doi.org/10.47065/bits.v6i2.5435>
- Pei, Y., Chen, S., Ke, Z., Silamu, W., & Guo, Q. (2022). AB-LaBSE: Uyghur Sentiment Analysis via the Pre-Training Model with BiLSTM. *Applied Sciences (Switzerland)*, 12(3). <https://doi.org/10.3390/app12031182>
- Pellicer, L. F. A. O., Ferreira, T. M., & Costa, A. H. R. (2023). Data augmentation techniques in natural language processing. *Applied Soft Computing*, 132. <https://doi.org/10.1016/j.asoc.2022.109803>
- Pradnyana, G. A., Anggraeni, W., Yuniarno, E. M., & Purnomo, M. H. (2023). Fine-Tuning IndoBERT Model for Big Five Personality Prediction from Indonesian Social Media. *2023 International Seminar on Intelligent Technology and Its Applications: Leveraging Intelligent Systems to Achieve Sustainable Development Goals, ISITIA 2023 - Proceeding*, 93–98. <https://doi.org/10.1109/ISITIA59021.2023.10221074>
- Putri, Faraby, S., & Adiwijaya. (2020). Analisis Sentimen Review Film Berbahasa Inggris Dengan Pendekatan Bidirectional Encoder Representations from Transformers. *Jurnal Teknik Informatika Dan Sistem Informasi*, 6(2), 181–193. <https://doi.org/10.35957/jatisi.v6i2.206>
- Putri, N. A. R., & Ardiansyah. (2023). Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa. *Jurnal Sains Dan Informatika*, 9(2), 136–145. <https://doi.org/10.34128/jsi.v9i2.649>
- Rashid, M. A., & Amirkhani, H. (2021, March 3). The Effect of Using Masked Language Models in Random Textual Data Augmentation. *26th International Computer Conference, Computer Society of Iran, CSICC 2021*. <https://doi.org/10.1109/CSICC52343.2021.9420616>



Optimasi IndoBERT untuk Analisis Sentimen pada Data Tidak Seimbang dengan Augmentasi Data : Studi Kasus Kebijakan TAPERA dan Dataset

Rizki Dina Febriana, Dr. Sri Mulyana, M.Kom.

UNIVERSITAS
GADJAH MADA

Kebijakan Turan Tabungan Perumahan Rakyat (Tapera) pada Platform X Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector Machine. *JEISBI (Journal of Emerging Information System and Business Intelligence)*, 05, 298–306. <https://doi.org/10.26740/jeisbi.v5i3.64074>

Robert Antonius, Zulkarnain, A. R., & Irsyad, H. (2024). Pendekatan TF-IDF, SMOTE, dan SVM dalam Klasifikasi Sentimen Masyarakat terhadap Pemblokiran Judi Online. *Buletin Ilmiah Informatika Teknologi*, 2(3), 115–122. <https://doi.org/10.58369/biit.v2i3.65>

Saputra, A. C., Saragih, A. S., & Ronaldo, D. (2024). Bert & DistillBERT ulasan course. *Jurnal Teknologi Informasi*, 18(02), 160–171. <https://doi.org/https://doi.org/10.47111/jti.v18i2.13278>

Saputra, F., Wijaya, S., Nurhadryani, Y., & Defina. (2020). Lexicon Addition Effect on Lexicon-Based of Indonesian Sentiment Analysis on Twitter. *Proceedings - 2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020*, 136–141. <https://doi.org/10.1109/ICIMCIS51567.2020.9354269>

Saraswati, N. P. V. D., Yudistira, N., & Adikara, P. P. (2023). Analisis Sentimen terhadap Perundangan Siber pada Twitter menggunakan Algoritma Bidirectional Encoder Representations from Transformer (BERT). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(2), 909–916. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12345>

Sari, A. K., Irsyad, A., Aini, D. N., Islamiyah, & Ginting, S. E. (2024). Analisis Sentimen Twitter Menggunakan Machine Learning untuk Identifikasi Konten Negatif. *Adopsi Teknologi Dan Sistem Informasi (ATASI)*, 3(1), 64–73. <https://doi.org/10.30872/atasi.v3i1.1373>

Sera, E., Hazriani, Mirfan, & Yuyun. (2023). Analisis Sentimen Ulasan Produk di E-Commerce Bukalapak Menggunakan Natural Language Processing Sentiment Analysis of Product Reviews on E-Commerce's Bukalapak Using Natural Language Processing. *Prosiding Seminar Nasional Sistem Informasi Dan Teknologi (SISFOTEK)*.

Seran, Y. B., & Supatman. (2024). ANALISIS SENTIMEN MASYARAKAT TERHADAP KINERJA KERJA PRESIDEN JOKO WIDODO ENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *Jurnal Mahasiswa Teknik Informatika*, 8(4). <https://doi.org/https://doi.org/10.36040/jati.v8i4.10171>

Siino, M., Tinnirello, I., & La Cascia, M. (2024). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. *Information Systems*, 121. <https://doi.org/10.1016/j.is.2023.102342>

Sofi, N., Sulistyorini, T., & Nazaruddin, M. (2023). Analisis Sentimen Masyarakat Pengguna Media Sosial Twitter Terhadap Motogp Mandalika Lombok Menggunakan Metode Bidirectional Encoder Representation From Transformers (BERT). *Jurnal Informasi, Sains, Dan Teknologi*, (1), 2023. <https://doi.org/10.55606/isaintek.v6i1.103>

Sriyanti, Z. A., Kartika, D. S. Y., & Najaf, A. R. E. (2024). IMPLEMENTASI MODEL BERT PADA ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP AKSI BOIKOT PRODUK ISRAEL. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3). <https://doi.org/10.23960/jitet.v12i3.4743>



UNIVERSITAS
GADJAH MADA

Optimasi IndoBERT untuk Analisis Sentimen pada Data Tidak Seimbang dengan Augmentasi Data : Studi Kasus Kebijakan TAPERA dan Dataset

Dina Febriana, Dr. Sri Mulyana, M.Kom.

Sriyanto, A., & Kusfina, K. (2025). Hybrid Deep Learning and USE Algorithm for Essay Scoring: Accuracy and Performance Analysis. *Sinkron*, 9(2), 914–924.

<https://doi.org/10.33395/sinkron.v9i2.14784>

Tang, T., Tang, X., & Yuan, T. (2020). Fine-Tuning BERT for Multi-Label Sentiment Analysis in Unbalanced Code-Switching Text. *IEEE Access*, 8, 193248–193256. <https://doi.org/10.1109/ACCESS.2020.3030468>

Wagh, V., Laddha, S., & Kadam, P. (2024). Detecting plagiarism using Latent Semantic Analysis and Cosine Similarity Approach. *2024 IEEE International Conference on Blockchain and Distributed Systems Security, ICBDS 2024*. <https://doi.org/10.1109/ICBDS61829.2024.10837475>

Wirawan, A., Cahyono, H. D., & Winarno. (2023). Easy Data Augmentation in Sentiment Analysis of Cyberbullying. *2023 6th International Conference on Information and Communications Technology, ICOIACT 2023*, 443–447. <https://doi.org/10.1109/ICOIACT59844.2023.10455817>

Yang, F., Zamzmi, G., Angara, S., Rajaraman, S., Aquilina, A., Xue, Z., Jaeger, S., Papagiannakis, E., & Antani, S. K. (2023). Assessing Inter-Annotator Agreement for Medical Image Segmentation. *IEEE Access*, 11, 21300–21312. <https://doi.org/10.1109/ACCESS.2023.3249759>