

DAFTAR PUSTAKA

- [1] R. Tjahyadi, V. C. Mawardi, and N. J. Perdana, “APLIKASI DETEKSI KOMENTAR SPAM YOUTUBE DENGAN METODE SUPPORT VECTOR MACHINE BERBASIS WEB,” *Jurnal Serina Sains, Teknik dan Kedokteran*, vol. 1, no. 2, pp. 237–246, Sep. 2023, doi: 10.24912/jsstk.v1i2.28510.
- [2] A. N. Cahyo, R. Hermawan, M. Avin, D. Wijaya, and A. P. Sari, “Jurnal Restikom : Riset Teknik Informatika dan Komputer Deteksi Teks Promosi Judi Online Pada Kolom Komentar YouTube dengan Metode Regex dan Fuzzy Matching,” vol. 7, no. 2, pp. 176–187, 2025, [Online]. Available: <https://restikom.nusaputra.ac.id>
- [3] U. Abdurohim *et al.*, “IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI KOMENTAR SPAM PADA INSTAGRAM,” 2024. [Online]. Available: <https://www.kaggle.com/>
- [4] S. Jurnal and D. Prasetya Kristiadi, “JURNAL SISTEM INFORMASI DAN TEKNOLOGI (S I N T E K),” *Jurnal Sistem Informasi dan Teknologi (SINTEK)*, [Online]. Available: <https://sintek.stmikku.ac.id/index.php/home>
- [5] K. Kamdan, M. P. Anugrah, M. J. Almutaali, R. Ramdani, and I. L. Kharisma, “Performance Analysis of IndoBERT for Detection of Online Gambling Promotion in YouTube Comments,” *MDPI AG*, Sep. 2025, p. 66. doi: 10.3390/engproc2025107066.
- [6] F. Jauhari, M. Riza, R. G. Guntara, and M. R. Nugraha, “Implementasi Algoritma Naive Bayes untuk Filtrasi Spam Komentar Judi Online pada YouTube,” 2025. [Online]. Available: <https://ejournal.upi.edu/index.php/IJDB>
- [7] N. Fikriah Nst, D. Hamdhana, and M. Qamal, “Few-Shot Learning for Classifying Genuine and Bot Comments on YouTube Using Transformer Models,” 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [8] A. Widiyanto, M. Prameswari, and M. A. Latief, “Gambling Comments Detection on YouTube: A Comparative Study of Tree-Based Boosting, LSTM and GRU Models,” *JUTI: Jurnal Ilmiah Teknologi Informasi*, vol. 23, no. 2, pp. 144–160, 2025, doi: 10.12962/j24068535.v23i1.a1305.
- [9] M. Farizi, “PENGELOMPOKAN SPAM BOTNET DENGAN METODE FUZZY HASHING,” 2020.
- [10] A. S. Xiao and Q. Liang, “Spam detection for Youtube video comments using machine learning approaches,” *Machine Learning with Applications*, vol. 16, p. 100550, Jun. 2024, doi: 10.1016/j.mlwa.2024.100550.
- [11] A. Agarwal, P. Nikitha, S. Ramkumar, A. Sinha, P. Maheshwari, and A. S. Saini, “DeepGram: Combining Language Transformer and N-Gram based ML Models for YouTube Spam Comment Detection,” *Journal of Data Science and Intelligent Systems*, Nov. 2023, doi: 10.47852/bonviewjdsis3202966.
- [12] M. N. Venkatramana, C. Jashnavi, J. Jithesh Guptha, P. Chandra Sekhar, and M. Chandra, “Random Tree Classifier: A Machine Learning Spam Comment Detection on YouTube,” *Int J Sci Res Sci Technol*, doi: 10.32628/IJSRST.
- [13] M. Clinton, T. Manullang, A. Z. Rakhman, H. Tantriawan, and A. Setiawan, “Comparative Analysis of CNN, Transformers, and Traditional ML for Classifying Online Gambling Spam Comments in Indonesian,” 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>

- [14] H. Alkaabi, A. K. Jasim, and A. Darroudi, "From Static to Contextual: A Survey of Embedding Advances in NLP," *PERFECT: Journal of Smart Algorithms*, vol. 2, no. 2, pp. 57–66, Jul. 2025, doi: 10.62671/perfect.v2i2.77.
- [15] L. D. Krisnawati, A. W. Mahastama, S.-C. Haw, K.-W. Ng, and P. Naveen, "Indonesian-English Textual Similarity Detection Using Universal Sentence Encoder (USE) and Facebook AI Similarity Search (FAISS)," 2024.
- [16] F. Rifaldy, Y. Sibaroni, and S. S. Prasetyowati, "Effectiveness of Word2Vec and TF-IDF in Sentiment Classification on Online Investment Platforms Using Support Vector Machine," *JIPPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 863–874, Mar. 2025, doi: 10.29100/jipi.v10i2.6055.
- [17] U. N. Jakarta, V. B. Lestari, and C. A. Hutagalung, "J-KOMA Journal of Computer Science and Applications Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes," pp. 22–2025, doi: 10.21009/j.
- [18] Sutriawan, S. Rustad, G. F. Shidik, and Pujiono, "Performance Evaluation of Text Embedding Models for Ambiguity Classification in Indonesian News Corpus: A Comparative Study of TF-IDF, Word2Vec, FastText BERT, and GPT," *Ingenierie des Systemes d'Information*, vol. 30, no. 6, pp. 1469–1482, Jun. 2025, doi: 10.18280/isi.300606.
- [19] M. E. Wibowo, N. Rokhman, and A. Sihabudin, "Combining Multiple Text Representations for Improved Automatic Evaluation of Indonesian Essay Answers," *Scientific Journal of Informatics*, vol. 11, no. 3, pp. 681–692, Aug. 2024, doi: 10.15294/sji.v11i3.9440.
- [20] Ibnu Daqiqil Id, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python.*, 1st ed. Gobah Pekanbaru, 2021.
- [21] Y. Guo, Z. Mustafaoglu, and D. Koundal, "Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms," *Journal of Computational and Cognitive Engineering*, vol. 2, no. 1, pp. 5–9, Feb. 2023, doi: 10.47852/bonviewJCCE2202192.
- [22] M. Sri, L. Sajja, L. Yuktha, N. Esther, and P. Charitha, "ENHANCING SPAM COMMENT DETECTION ON SOCIAL MEDIA WITH EMOJI FEATURE AND POST-COMMENT PAIRS APPROACH USING ENSEMBLE METHODS OF MACHINE LEARNING", doi: 10.62646/ijitce.2024.v15.i3.pp837-845.
- [23] G. Airlangga, "Spam Detection in YouTube Comments Using Deep Learning Models: A Comparative Study of MLP, CNN, LSTM, BiLSTM, GRU, and Attention Mechanisms," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1533–1538, Oct. 2024, doi: 10.57152/malcom.v4i4.1671.
- [24] G. Airlangga, "Spam Detection on YouTube Comments Using Advanced Machine Learning Models: A Comparative Study," *Brilliance: Research of Artificial Intelligence*, vol. 4, no. 2, pp. 500–508, Oct. 2024, doi: 10.47709/brilliance.v4i2.4670.
- [25] K. Kadarsih and D. Pujianto, "APPLICATION OF TRANSFORMER MODEL AND WORD EMBEDDING IN SENTIMENT ANALYSIS OF INDONESIAN E-COMMERCE APPLICATION REVIEW," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 7, no. 3, pp. 720–732, Jul. 2025, doi: 10.47709/cnahpc.v7i3.6354.

- [26] C. J. L. Tobing, IGN Lanang Wijayakusuma, and Luh Putu Ida Harini, “Perbandingan Kinerja IndoBERT dan MBERT Untuk Deteksi Berita Hoaks Politik dalam Bahasa Indonesia,” *JST (Jurnal Sains dan Teknologi)*, vol. 14, no. 1, pp. 114–123, May 2025, doi: 10.23887/jstundiksha.v14i1.92126.
- [27] S. M. Isa, G. Nico, and M. Permana, “INDOBERT FOR INDONESIAN FAKE NEWS DETECTION,” *ICIC Express Letters*, vol. 16, no. 3, pp. 289–297, Mar. 2022, doi: 10.24507/icicel.16.03.289.
- [28] M. Diky Fadhilahsyah Ramadhan, F. Rakhmat Umbara, and R. Ilyas, “Klasifikasi Sentimen Kebohongan Berita Menggunakan Metode Indobert,” *Jurnal Sosial dan Sains (SOSAINS)*, vol. 5, no. 8, 2025.
- [29] N. P. I. Maharani, Y. Yustiawan, F. C. Rochim, and A. Purwarianti, “Domain-Specific Language Model Post-Training for Indonesian Financial NLP,” Oct. 2023, [Online]. Available: <http://arxiv.org/abs/2310.09736>
- [30] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” Nov. 2020, [Online]. Available: <http://arxiv.org/abs/2011.00677>
- [31] M. B. M. Amin *et al.*, “Deteksi Spam Berbahasa Indonesia Berbasis Teks Menggunakan Model Bert,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1291–1302, Dec. 2024, doi: 10.25126/jtiik.2024118121.
- [32] G. Soliman, H. Zaki, and M. Kilany, “A comparative analysis of encoder only and decoder only models for challenging LLM-generated STEM MCQs using a self-evaluation approach,” *Natural Language Processing Journal*, vol. 10, p. 100131, Mar. 2025, doi: 10.1016/j.nlp.2025.100131.
- [33] M. B. M. Amin *et al.*, “Deteksi Spam Berbahasa Indonesia Berbasis Teks Menggunakan Model Bert,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1291–1302, Dec. 2024, doi: 10.25126/jtiik.2024118121.
- [34] Y. Liang and M. Li, “A deep learning model for prediction of lysine crotonylation sites by fusing multi-features based on multi-head self-attention mechanism,” *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-04058-5.
- [35] G. H. Setiawan *et al.*, “Topic Clustering of Student Complaints Based on Semantic Meaning Using the indoBERT and K-Means Models,” 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [36] W. Liang and Y. Liang, “BPDec: Unveiling the Potential of Masked Language Modeling Decoder in BERT pretraining,” Dec. 2024, [Online]. Available: <http://arxiv.org/abs/2401.15861>
- [37] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, “Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN,” *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [38] C. Shaw, P. LaCasse, and L. Champagne, “Exploring emotion classification of Indonesian tweets using large scale transfer learning via IndoBERT,” *Soc Netw Anal Min*, vol. 15, no. 1, Dec. 2025, doi: 10.1007/s13278-025-01439-6.
- [39] A. Durap, “Explainable deep learning techniques for wind speed forecasting in coastal areas: Integrating model configuration, regularization, early stopping, and SHAP analysis,” *Neural Comput Appl*, vol. 37, no. 25, pp. 21219–21257, Sep. 2025, doi: 10.1007/s00521-025-11433-w.
- [40] I. Markoulidakis and G. Markoulidakis, “Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance

- Analysis,” *Technologies (Basel)*, vol. 12, no. 7, Jul. 2024, doi: 10.3390/technologies12070113.
- [41] A. Andryani, A. N. Salim, T. Sutabri, and K. Kunci, “Deteksi Email Spam dan Non-Spam Berdasarkan Isi Konten Menggunakan Metode K-Nearest Neighbor dan Support Vector Machine,” vol. 6, no. 2, 2024, doi: 10.46799/syntax.
- [42] M. Sitarz, “Extending F1 metric, probabilistic approach,” Oct. 2022, doi: 10.54364/AAIML.2023.1161.
- [43] G. Girinoto, D. A. Alwan, G. A. N. Gde K.T. D, O. G. Nabila, A. Arizal, and D. F. Priambodo, “Implementasi Deteksi Judul Berita Clickbait Berbahasa Indonesia dengan pre-trained model Multilingual BERT Pada Aplikasi Berbasis Chrome Extension,” *Jurnal Ilmiah SINUS*, vol. 20, no. 2, p. 25, Jul. 2022, doi: 10.30646/sinus.v20i2.624.
- [44] M. F. Rizal, “Optimalisasi Klasifikasi Disposisi Pengaduan Masyarakat melalui Kanal LAPOR Menggunakan Algoritma Naïve Bayes dan Integrasi Ekstensi Chrome Mu’izza ’Ainur Rasyida,” *Integrative Perspectives of Social and Science Journal*, vol. 2, no. 3, p. 4115, 2025.
- [45] M. Aljabri *et al.*, “Kashif: A Chrome Extension for Classifying Arabic Content on Web Pages Using Machine Learning,” *Applied Sciences (Switzerland)*, vol. 14, no. 20, Oct. 2024, doi: 10.3390/app14209222.
- [46] S. Essahli, O. Sarsar, A. Bentajer, A. Motii, and I. Fouad, “FakeZero: Real-Time, Privacy-Preserving Misinformation Detection for Facebook and X,” Nov. 2025, [Online]. Available: <http://arxiv.org/abs/2510.25932>
- [47] M. Raheem, N. Fathima Abubacker, and D. Wiyani, “SENTIMENT-BASED RECOMMENDATION FOR ONLINE SHOPPING,” *J Theor Appl Inf Technol*, vol. 15, no. 9, 2024, [Online]. Available: www.jatit.org

