

## REFERENSI

- [1] Y. Zhang and Z. Teng, *Natural language processing : A machine learning perspective*. 2021.
- [2] J. Eisenstein, *Introduction to Natural Language Processing. Adaptive Computation and Machine Learning serie*. MIT Press, 2019.
- [3] “An Introduction to Natural Language Processing (NLP) | Built In.” <https://builtin.com/data-science/introduction-nlp> (accessed Nov. 13, 2021).
- [4] D. B. Setyohadi and F. A. Kristiawan, “Preprocessing Iterative Partitioning Filter Algorithm,” 2017.
- [5] “Dasar Text Preprocessing dengan Python | by Kuncahyo Setyo Nugroho | Medium.” <https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe> (accessed Nov. 18, 2021).
- [6] J. Resti and F. Selva Jumeilah, “Terbit online pada laman web jurnal : <http://jurnal.iaii.or.id> Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian,” 2017. [Online]. Available: <http://jurnal.iaii.or.id>
- [7] F. Rahutomo, A. Retno, and T. H. Ririd, “Evaluasi Daftar Stopword Bahasa Indonesia,” vol. 6, no. 1, pp. 41–48, 2019, doi: 10.25126/jtiik.201861226.
- [8] O. Rezalina, “Perbandingan Algoritma Stemming Nazief & Adriani, Porter dan Arifin Setiono untuk Dokumen Teks Bahasa Indonesia,” 2016.
- [9] B. Zaman, “Modifikasi Algoritma Porter untuk Stemming Pada Kata Bahasa Indonesia,” 2014.
- [10] M. N. Saadah, R. W. Atmagi, D. S. Rahayu, and A. Z. Arifin, “Sistem Temu Kembali Dokumen Teks dengan Pembobotan Tf-Idf Dan LCS,” *Jurnal Ilmiah Teknologi Informasi*, vol. 11, pp. 17–20, 2013, doi: <http://dx.doi.org/10.12962/j24068535.v11i1.a16>.
- [11] V. Amrizal, “Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) dan Cosine Similarity pada Sistem Temu Kembali Informasi untuk Mengetahui Syarah Hadits Berbasis Web,” *Jurnal Teknik Informatika*, vol. 11, no. 2, pp. 149–164, Nov. 2018, doi: 10.15408/jti.v11i2.8623.
- [12] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” pp. 757–770, Nov. 2020, doi: 10.48550/arxiv.2011.00677.
- [13] “indolem/indobert-base-uncased · Hugging Face.” <https://huggingface.co/indolem/indobert-base-uncased> (accessed Jul. 12, 2022).
- [14] S. W. Iriananda, M. A. Muslim, and H. S. Dachlan, “Identifikasi Kemiripan Teks Menggunakan Class Indexing Based dan Cosine Similarity Untuk Klasifikasi Dokumen Pengaduan,” *Matics*, vol. 10, no. 2, p. 30, Mar. 2019, doi: 10.18860/mat.v10i2.5327.
- [15] “Overview of Text Similarity Metrics in Python | by Sanket Gupta | Towards Data Science.” <https://towardsdatascience.com/overview-of-text-similarity-metrics-3397c4601f50> (accessed Nov. 18, 2021).
- [16] S. R. Wardhana, D. R. Yunianto, A. Z. Arifin, and D. Purwitasari, “Pembobotan Kata Berbasis Preferensi Dan Hubungan Semantik Pada Dokumen Fiqih Berbahasa Arab,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2, pp. 132–137, Jul. 2015, doi: 10.25126/JTIK.201522146.
- [17] F. Jáñez-Martino, E. Fidalgo, S. González-Martínez, and J. Velasco-Mata, “Classification of Spam Emails through Hierarchical Clustering and Supervised Learning”, Accessed: Jul. 10, 2022. [Online]. Available: <https://talosintelligence.com/reputation>

- [18] F. Alzami, E. D. Udayanti, D. P. Prabowo, and R. A. Megantara, "Document Preprocessing with TF-IDF to Improve the Polarity Classification Performance of Unstructured Sentiment Analysis," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, pp. 235–242, Aug. 2020, doi: 10.22219/kinetik.v5i3.1066.
- [19] S. Maheshwary and H. Misra, "Matching Resumes to Jobs via Deep Siamese Network," 2018, doi: 10.1145/3184558.3186942.
- [20] A. Marpaung, R. Rismala, and H. Nurrahmi, "Hate Speech Detection in Indonesian Twitter Texts using Bidirectional Gated Recurrent Unit," *KST 2021 - 2021 13th International Conference Knowledge and Smart Technology*, pp. 186–190, Jan. 2021, doi: 10.1109/KST51265.2021.9415760.
- [21] Y. Shen and J. Liu, "Comparison of Text Sentiment Analysis based on Bert and Word2vec," *2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer, ICFTIC 2021*, pp. 144–147, 2021, doi: 10.1109/ICFTIC54370.2021.9647258.
- [22] E. Popova and V. Spitsyn, "Sentiment Analysis of Short Russian Texts Using BERT and Word2Vec Embeddings".
- [23] J. Han, M. Kamber, and J. Pei, "Getting to Know Your Data," *Data Mining*, pp. 39–82, 2012, doi: 10.1016/B978-0-12-381479-1.00002-2.