

DAFTAR PUSTAKA

- Alasadi, S. A., & Bhaya, W. S. (2017). Review of data *preprocessing* techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16), 4102-4107.
- Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di X. *Jurnal Tekno Kompak*, 14(2), 86-91.
- Amirullah, F., & Alam, S. (2023). Analisis Sentimen Terhadap Kinerja KPU Menjelang Pemilu 2024 Berdasarkan Opini X Menggunakan Naïve Bayes. *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, 2(3), 69-76.
- Arifiandy, R., & Fahmi, H. (2024). Comparison Email Spam detection vectorizing using *Bag of Word*, TFIDF and Word2Vec in Multinomial Naïve Bayes. *Syntax: Jurnal Informatika*, 13(01), 01-14.
- Babalola, O., Ojokoh, B., & Boyinbode, O. (2024). Comprehensive Evaluation of *LDA*, NMF, and *BERTopic*'s Performance on News Headline *Topic Modeling*. *Journal of Computing Theories and Applications*, 2(2), 268-289.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Buchmüller, A., Kant, G., Weisser, C., Säfken, B., Kis-Katos, K., & Kneib, T. (2022). Twitmo: a X data *Topic Modeling* and visualization package for R. *arXiv preprint arXiv:2207.11236*.
- Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2020). *Topic Modeling* in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8, 439-453.
- Djunaidi, K., Kusuma, D. T., Ningrum, R. F., Siswipraptini, P. C., & Murad, D. F. (2025). Big Data Analytics of Knowledge and Skill Sets for Web Development Using Latent Dirichlet Allocation and Clustering Analysis. *International Journal of Advanced Computer Science & Applications*, 16(1).
- Egger, R., & Yu, J. (2022). A *Topic Modeling* comparison between *LDA*, *nmf*, *top2vec*, and *BERTopic* to demystify X posts. *Frontiers in sociology*, 7, 886498.
- Fahlevvi, M. R., & Azhari, S. N. (2022). *Topic Modeling* on online news. Portal using Latent Dirichlet Allocation (*LDA*). *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 16(4), 335-344.
- Gan, L., Yang, T., Huang, Y., Yang, B., Luo, Y. Y., Richard, L. W. C., & Guo, D. (2023, October). Experimental comparison of three *Topic Modeling* methods with *LDA*, *Top2Vec* and *BERTopic*.

In *International Symposium on Artificial Intelligence and Robotics* (pp. 376-391). Singapore: Springer Nature Singapore.

Goud, Y. S., Tripathi, M. K., Shinde, A., Bongulwar, D., Tingare, B., & Angadi, S. Advanced Techniques in Cancer Prediction: A Review of Data Mining and Machine Learning Approaches.

Grootendorst, M. (2022). *BERTopic: Neural Topic Modeling with a class-based TF-IDF procedure*. arXiv. *arXiv preprint arXiv:2203.05794*.

Grootendorst, M. (2022). *BERTopic: Neural Topic Modeling with a class-based TF-IDF procedure*. *arXiv preprint arXiv:2203.05794*.

HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of *preprocessing* on text classification using a bag-of-words representation. *PloS one*, *15*(5), e0232525.

Hikmah, F. N., Basuki, S., & Azhar, Y. (2020). Deteksi Topik Tentang Tokoh Publik Politik Menggunakan Latent Dirichlet Allocation (*LDA*). *Jurnal Repositor*, *2*(4).

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), 1527-1554.

Junaidi, S., Devegi, M., & Kurniawan, H. (2023). Pelatihan Pengolahan dan Visualisasi Data Penduduk menggunakan Python. *ADMA: Jurnal Pengabdian dan Pemberdayaan Masyarakat*, *4*(1), 151-162.

Kadhim, A. I. (2018). An evaluation of *preprocessing* techniques for text classification. *International Journal of Computer Science and Information Security (IJCSIS)*, *16*(6), 22-32.

- Karmila, S., Arianto, R., & Rezki, R. (2020). Penerapan Fuzzy C-Means Pada Klasterisasi Karakteristik Pengunjung Website Pmb Stt–Pln Untuk Meningkatkan Kepadatan Kunjungan. *Jurnal Informatika dan Komputasi: Media Bahasan, Analisa dan Aplikasi*, 14(1), 8-17.
- Kaur, A., & Wallace, J. R. (2024). Moving Beyond *LDA*: A Comparison of Unsupervised *Topic Modeling* Techniques for Qualitative Data Analysis of Online Communities. *arXiv preprint arXiv:2412.14486*.
- Kukushkin, K., Ryabov, Y., & Borovkov, A. (2022). Digital twins: a systematic literature review based on data analysis and *Topic Modeling*. *Data*, 7(12), 173.
- Kusuma, D. T., Ahmad, N., Ahmad, S. S. S., Sangadji, I. B., & Arvio, Y. (2025). An efficient clustering approach in electrical energy consumption patterns. *Bulletin of Electrical Engineering and Informatics*, 14(2), 1168-1177.
- Medvecki, D., Bašaragin, B., Ljajić, A., & Milošević, N. (2023, March). Multilingual transformer and *BERTopic* for short text *Topic Modeling*: The case of serbian. In *Conference on Information Technology and its Applications* (pp. 161-173). Cham: Springer Nature Switzerland.
- Mendonça, M., & Figueira, Á. (2024). Topic extraction: *BERTopic*'s insight into the 117th congress's Xverse. *Informatics*, 11 (1), 8. *Online: <https://doi.org/10.3390/informatics11010008>*.

- Mimno, D., Wallach, H., Naradowsky, J., Smith, D. A., & McCallum, A. (2009, August). Polylingual topic models. In *Proceedings of the 2009 conference on empirical methods in natural language processing* (pp. 880-889).
- Mramra, W. Y. Y., & Nataliani, Y. (2025). Penentuan bidang unggulan akademik universitas melalui metode topik Linear Discriminant Analysis (LDA). *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, 4(1), 93-105.
- Nurhaliza, A. F. (2024). Penerapan Pemodelan Topik menggunakan Metode Latent Dirichlet Allocation terhadap Pembahasan Pemilu Indonesia tahun 2024 di X. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8(7).
- Nurhalizah, R. S., Ardianto, R., & Purwono, P. (2024). Analisis Supervised dan *Unsupervised learning* pada Machine Learning: Systematic Literature Review. *Jurnal Ilmu Komputer dan Informatika*, 4(1), 61-72.
- Putra, K. T., Hariyadi, M. A., & Crysdiyan, C. (2023). Perbandingan Feature Extraction Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis Svm. *Jurnal Cahaya Mandalika ISSN 2721-4796 (online)*, 3(2), 1449-1463.
- Ramamoorthy, T., Kulothungan, V., & Mappillairaju, B. (2024). *Topic Modeling* and social network analysis approach to explore diabetes discourse on X in India. *Frontiers in artificial intelligence*, 7, 1329185.

- Retnoningsih, E., & Pramudita, R. (2020). Mengenal machine learning dengan teknik supervised dan *unsupervised learning* menggunakan python. *Bina Insani Ict Journal*, 7(2), 156-165.
- Rifai, M. F., Jatnika, H., & Valentino, B. (2019). Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS). *Petir*, 12(2), 520754.
- Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408).
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229.
- Siahaan, A. L. S. (2024). Pengaruh Perkembangan Media Sosial Terhadap Etika Komunikasi dan Demokrasi Pancasila: Perspektif Hukum di Indonesia. *Jurnal Citizenship Virtues*, 4(2), 796-801.
- Siswipraptini, P. C. (2023). Klasifikasi Pekerjaan Bidang Teknologi Informasi Menggunakan Algoritma Cosine Similarity. *KILAT*, 12(1), 38-48.
- Suresha, H. P., & Tiwari, K. K. (2021). *Topic Modeling* and sentiment analysis of electric vehicles of X data. *Asian Journal of Research in Computer Science*, 12(2), 13-29.
- Wahyuni, W., Lestari, T. P., Apriliana, M., & Gumelta, R. (2025). Implementation of *BERTopic* for *Topic Modeling* Analysis of the

Free Nutritious Meal Program Based on YouTube Comments. *Journal of Applied Informatics and Computing*, 9(4), 1964-1971.

Wibowo, I. S., Witanti, A., & Susilawati, I. (2024). Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 11(1).

Wirth, R., & Hipp, J. (2000, April). *CRISP-DM: Towards a standard process model for data mining*. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29-39).

YULISTIANTI, Y. (2024). *implementasi metode Latent Dirichlet Allocation (Lda) dalam topic modeling pada data berita kesehatan* (Doctoral dissertation, Nusa Putra University).