

DAFTAR PUSTAKA

- Agustina, V., & Herliana, A. (2025). Analisis Sentimen Publik atas Kebijakan Efisiensi Anggaran 2025 dengan Text Mining dan Natural Language Processing. *JURNAL MEDIA INFORMATIKA (JUMIN)*, 6(3), 2182–2194.
- Airen, S., & Agrawal, J. (2021). Movie Recommender System Using K-Nearest Neighbors Variants. *National Academy Science Letters*, 45(1), 75–82. <https://doi.org/10.1007/s40009-021-01051-0>
- Artetxe, M., Ruder, S., Yogatama, D., Labaka, G., & Agirre, E. (2020). A Call for More Rigor in Unsupervised Cross-lingual Learning. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, 7375–7388. <https://doi.org/10.18653/v1/2020.acl-main.658>
- Artstein, R., & Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34(4), 555–596. <https://doi.org/10.1162/coli.07-034-R2>
- Arustashvili, M., & Balog, K. (2026). SciNUP: Natural Language User Interest Profiles for Scientific Literature Recommendation. *Lecture Notes in Computer Science*, 16486 LNCS, 432–446. https://doi.org/10.1007/978-3-032-21321-1_51
- Ayoade, O. B., Shahrestani, S., & Ruan, C. (2025). Machine Learning and Deep Learning Approaches for Predicting Diabetes Progression: A Comparative Analysis. *Electronics (Switzerland)*, 14(13). <https://doi.org/10.3390/electronics14132583>
- Basha, M. J., Vijayakumar, S., Jayashankari, J., Alawadi, A. H., & Durdona, P. (2023). Advancements in Natural Language Processing for Text Understanding. *E3S Web of Conferences*, 399. <https://doi.org/10.1051/e3sconf/202339904031>
- Basile, V. (2020). It's the end of the gold standard as we know it. On the impact of pre-aggregation on the evaluation of highly subjective tasks. *CEUR Workshop Proceedings*, 2776, 31–40.
- Boppana, V., & Sandhya, P. (2021). Web crawling based context aware recommender system using optimized deep recurrent neural network. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00534-7>
- Chavhan, S., Raghuwanshi, M. M., & Dharmik, R. C. (2021). Information Retrieval using Machine learning for Ranking: A Review. *Journal of Physics: Conference Series*, 1913(1). <https://doi.org/10.1088/1742-6596/1913/1/012150>
- Dani, A. H., Puspaningrum, E. Y., & Mumpuni, R. (2024). Studi Performa TF-IDF dan Word2Vec Pada Analisis Sentimen Cyberbullying. *Jurnal Teknik Informatika dan Terapan*, 2(2), 94–106. <https://doi.org/10.62951/router.v2i2.76>
- Darmawan, I., Maulana, M., Gunawan, R., & Widiyasono, N. (2022). Evaluating

- Web Scraping Performance Using XPath, CSS Selector, Regular Expression, and HTML DOM With Multiprocessing Technical Applications. *International Journal on Informatics Visualization*, 6(4), 904–910. <https://doi.org/10.30630/joiv.6.4.1525>
- Dettori, J. R., & Norvell, D. C. (2020). Kappa and Beyond: Is There Agreement? *Global Spine Journal*, 10(4), 499–501. <https://doi.org/10.1177/2192568220911648>
- Diouf, R., Birregah, B., Sarr, E. N., Bousso, M., Sall, O., & Mbaye, S. N. (2019). Web Scraping: State-of-the-Art and Areas of Application. *IEEE International Conference on Big Data*, 6040–6042.
- Eminagaoglu, M. (2020). A new similarity measure for vector space models in text classification and information retrieval. *Journal of Information Science*, : 1–14. <https://doi.org/10.1177/0165551520968055>
- Engl, F., & Herrmann, F. (2023). Machine Learning based Approach on Employee Attrition Prediction with an Emphasize on predicting Leaving Reasons. *Anwendungen und Konzepte der Wirtschaftsinformatik*, 18(18), 11. <https://doi.org/10.26034/lu.akwi.2023.4488>
- Fajriansyah, M., Adikara, P. P., & Widodo, A. W. (2021). Sistem Rekomendasi Film menggunakan Content- based Filtering. *J-PTIHK (Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer)*, 5(6), 2188–2199.
- Firdaus, A., & Firdaus, W. I. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). *Jurnal JUPITER*, 13(1), 66–78.
- Flores, V. A., Permatasari, P. A., & Jasa, L. (2020). Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword. *Majalah Ilmiah Teknologi Elektro*, 19(2), 157–162. <https://doi.org/10.24843/MITE.2020.v19i02.P06>
- Glauber, R., & Loula, A. (2019). *Collaborative Filtering vs . Content-Based Filtering : differences and similarities.*
- Hambarde, K. A., & Proenca, H. (2023). Information Retrieval: Recent Advances and beyond. *IEEE Access*, 11(July), 76581–76604. <https://doi.org/10.1109/ACCESS.2023.3295776>
- Indriyanto, I., & Sumitra, I. D. (2019). Measuring the Level of Plagiarism of Thesis using Vector Space Model and Cosine Similarity Methods. *IOP Conference Series: Materials Science and Engineering*, 022111. <https://doi.org/10.1088/1757-899X/662/2/022111>
- Jadon, A., & Patil, A. (2024). A Comprehensive Survey of Evaluation Techniques for Recommendation Systems. *arXiv preprint*. https://doi.org/10.1007/978-3-031-71484-9_25
- Ji, Y., Sun, A., Zhang, J., & Li, C. (2023). A Critical Study on Data Leakage in Recommender System Offline Evaluation. *ACM Transactions on Information*

- Systems*, 41(3). <https://doi.org/10.1145/3569930>
- Kanwal, S., Nawaz, S., & Malik, M. (2021). A Review of Text-Based Recommendation Systems. *IEEE Access*, 9. <https://doi.org/10.1109/ACCESS.2021.3059312>
- Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and its Applications*, 13(3), 144–168. <https://doi.org/10.15849/ijasca.211128.11>
- Ko, H., Lee, S., & Park, Y. (2022). A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics*, 11(1), 141. <https://doi.org/10.3390/electronics11010141>
- Kumar, N., Gupta, M., Sharma, D., & Ofori, I. (2022). Technical Job Recommendation System Using APIs and Web Crawling. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/7797548>
- Kurniawan, H. H., Lukman, W. S., Fredyana, R., & Ibrahim, M. A. (2024). Movie Recommendation System: A Comparison of Content-Based and Collaborative Filtering. *Procedia Computer Science*, 245, 860–868. <https://doi.org/10.1016/j.procs.2024.10.313>
- Laili, A. N., Adikara, P. P., & Adinugroho, S. (2019). Rekomendasi Film Berdasarkan Sinopsis Menggunakan Metode Word2Vec. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(6), 6035–6043.
- Le, H. H. (2024). *Data Leakage in Sequential Recommender Systems Evaluation*. 68.
- Leonardelli, E., Menini, S., Aprosio, A. P., Guerini, M., & Tonelli, S. (2021). Agreeing to Disagree: Annotating Offensive Language Datasets with Annotators' Disagreement. *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 10528–10539. <https://doi.org/10.18653/v1/2021.emnlp-main.822>
- Li, M., Gao, Q., & Yu, T. (2023). Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters. *BMC Cancer*, 23(1), 1–5. <https://doi.org/10.1186/s12885-023-11325-z>
- Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based Recommender Systems: State of the Art and Trends. In F. Ricci, L. Rokach, & B. Shapira (Ed.), *Recommender Systems Handbook* (hal. 73–105). Springer. https://doi.org/10.1007/978-0-387-85820-3_3
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. In *arXiv preprint*.
- Natarajan, S., Vairavasundaram, S., Natarajan, S., & Gandomi, A. H. (2020). Resolving data sparsity and cold start problem in collaborative filtering recommender system using Linked Open Data. *Expert Systems with Applications*, 149, 113248. <https://doi.org/10.1016/j.eswa.2020.113248>

- Nawangarsi, R. P., Kusumaningrum, R., & Wibowo, A. (2019). Word2vec for Indonesian sentiment analysis towards hotel reviews: An evaluation study. *Procedia Computer Science*, 157, 360–366. <https://doi.org/10.1016/j.procs.2019.08.178>
- Nguyen, L. V., Nguyen, T.-H., & Jung, J. J. (2020). Content-Based Collaborative Filtering using Word Embedding: A Case Study on Movie Recommendation. *International Conference on Research in Adaptive and Convergent Systems (RACS '20)*, October 2020, 96–100. <https://doi.org/10.1145/3400286.3418253>
- Papadakis, H., Papagrigoriou, A., Kosmas, E., Panagiotakis, C., Markaki, S., & Fragopoulou, P. (2023). Content-Based Recommender Systems Taxonomy. *Foundations of Computing and Decision Sciences*, 48(2), 211–241. <https://doi.org/10.2478/fcds-2023-0009>
- Permana, A. H. J., & Wibowo, A. T. (2023). Movie Recommendation System Based on Synopsis Using Content-Based Filtering with TF-IDF and Cosine Similarity. *International Journal on ICT*, 9(2), 1–14. <https://doi.org/10.21108/ijoiict.v9i2.747>
- Prabowo, Y. D., Marselino, T. L., & Suryawiguna, M. (2019). Pembentukan Vector Space Model Bahasa Indonesia Menggunakan Metode Word to Vector. *Jurnal Buana Informatika*, 10(1), 29–40.
- Putra, M. P. R., & Wardani, K. R. N. (2020). PENERAPAN TEXT MINING DALAM MENGANALISIS KEPRIBADIAN PENGGUNA MEDIA SOSIAL. *JUTIM (Jurnal Teknik Informatika Musirawas)*, 05(01), 63–71.
- Rizquina, A. Z., & Ratnasari, C. I. (2023). Implementasi Web Scraping untuk Pengambilan Data Pada Website E- Commerce. *Jurnal Teknologi dan Sistem Informasi Bisnis (JTEKSIS)*, 5(4), 377–383.
- Romero Meza, L., & D'Urso, G. (2024). Laura Romero Meza Giulio D'Urso. *Psychological Studies*, 69(3), 349–367. <https://doi.org/10.1007/s12646-024-00807-0>
- Saputra, J. M. A., Huizen, L. M., Khoirudin, Firmandaru, A., & Arianto, D. B. (2024). Sistem Rekomendasi Film pada Platform Streaming Menggunakan Metode Content-Based Filtering. *TRANSFORMATIKA*, 22(1), 10–21. <https://doi.org/DOI: 10.26623/transformatika.v22i1.7041>
- Shahmirzadi, O., Lugowski, A., & Younge, K. (2018). *Text Similarity in Vector Space Models: A Comparative Study*.
- Sinha, S., & Sharma, T. (2023). Content-Based Movie Recommendation System : An Enhanced Approach to Personalized Movie Recommendations. *International Journal of Innovative Research in Computer Science and Technology (IJIRCST)*, 11(3), 67–71. <https://doi.org/10.55524/ijircst.2023.11.3.12>
- Tacuri, A., Firmenich, S., Fernández, A., Riva, F., Urbieto, M., & Rossi, G. (2025). Web Scraping by End Users. *IEEE Access*, 13, 205027–205044. <https://doi.org/10.1109/ACCESS.2025.3636662>

- Valcarce, D., Bellogín, A., Parapar, J., & Castells, P. (2020). Assessing ranking metrics in top-N recommendation. In *Information Retrieval Journal* (Vol. 23, Nomor 4). Springer Netherlands. <https://doi.org/10.1007/s10791-020-09377-x>
- Velamentosa, D., & Zuliarso, E. (2025). SISTEM REKOMENDASI FILM MENGGUNAKAN METODE CONTENT-BASED FILTERING. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 2918–2922.
- Wang, J., & Dong, Y. (2020). Measurement of Text Similarity: A Survey. *Information*, 11(9), 421. <https://doi.org/10.3390/info11090421>
- Widayanti, R., Chakim, M. H. R., Lukita, C., Rahardja, U., & Lutfiani, N. (2023). Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering. *Journal of Applied Data Sciences*, 4(3), 289–302. <https://doi.org/10.47738/jads.v4i3.115>
- Xia, H. (2023). Continuous-bag-of-words and Skip-gram for word vector training and text classification. *Journal of Physics: Conference Series*, 2634, 012052. <https://doi.org/10.1088/1742-6596/2634/1/012052>
- Yao, Z. (2023). Review of Movie Recommender Systems Based on Deep Learning. *Proceedings of the International Conference on Language, Communication and Culture (ICLCC 2023)*, 159, 02010. <https://doi.org/10.1051/shsconf/202315902010>
- Zhu, Y., Yuan, H., Wang, S., Liu, J., Liu, W., Deng, C., Chen, H., Liu, Z., Dou, Z., & Wen, J. R. (2025). Large Language Models for Information Retrieval: A Survey. *ACM Transactions on Information Systems*, 44(1). <https://doi.org/10.1145/3748304>