

DAFTAR PUSTAKA

- Adnan, K., & Akbar, R. (2019). An analytical study of information extraction from unstructured and multidimensional big data. *Journal of Big Data*, 6(1), 1–38. <https://doi.org/10.1186/S40537-019-0254-8/FIGURES/5>
- Ali, A. H., Yaseen, M. G., Aljanabi, M., & Abed, S. A. (2023). Transfer Learning: A New Promising Techniques. *Mesopotamian Journal of Big Data*, 2023, 29–30. <https://doi.org/10.58496/MJBD/2023/004>
- Almutairi, A. (2024). Newspaper elements detection and newspaper pages categorization using CNNs and transformers. *International Journal on Document Analysis and Recognition (IJDAR)*, 28(2), 225–237. <https://doi.org/10.1007/S10032-024-00503-9>
- Anwar, M. T., Wijoyo, S. H., & Putra, H. W. N. (2024). Implementasi Metode TextRank dan Named Entity Recognition Untuk Ekstraksi Kata Kunci Pada Media Online Berita. *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi*, 5(1), 34–41. <https://doi.org/10.25126/JUSTSI.V5I1.401>
- Ariyani, R. (2023). LITERATURE REVIEW INFORMATION EXTRACTION PADA TEKS BIOMEDICAL DENGAN PENDEKATAN MACHINE LEARNING DAN NAMED ENTITY RECOGNITION. *Jurnal Ilmiah Teknik*, 2(1), 83–88. <https://doi.org/10.56127/JUIT.V2I1.1727>
- Barot, M. P., Prajapati, M. G., & Patel, R. R. (2019). Review Paper for Types of Data in Big Data and Text Mining. *International Journal of Engineering Research & Technology*, 8(10). <https://doi.org/10.17577/IJERTV8IS100028>
- Christianto, D., Siswanto, E., & Chaniago, R. (2015). Penggunaan Named Entity Recognition dan Artificial Intelligence Markup Language untuk Penerapan Chatbot Berbasis Teks. *Jurnal Telematika*, 10(2), 61–68. <https://doi.org/10.61769/TELEMATIKA.V10I2.130>
- Cui, L., & Zhang, Y. (2019). Hierarchically-Refined Label Attention Network for Sequence Labeling. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 4115–4128. <https://doi.org/10.18653/V1/D19-1422>
- Darpito, M. N., Firdausy, K., & Fadlil, A. (2025). Perbandingan Unjuk Kerja Library Optical Character Recognition (OCR) dalam Pengenalan Teks pada Dokumen Digital. *Jurnal Informatika Polinema*, 11(3), 273–282. <https://doi.org/10.33795/JIP.V11I3.7025>
- Dirgantara, M. Y. S., Fauzi, M. A., & Perdana, R. S. (2018). Penerapan Named Entity Recognition Untuk Mengenali Fitur Produk Pada E-commerce Menggunakan Rule Template Dan Hidden Markov Model. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3912–3920. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2791>
- Fredriksson, L. (2023). *Classification of invoices using a 2D NLP approach : A comparison between methods for invoice information extraction for the purpose of classification. TRITA-EECS-EX*. <https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-332438>
- Gautama, T. K., Hendrik, A., & Hendaya, R. (2016). Pengenalan Objek pada Computer Vision dengan Pencocokan Fitur Menggunakan Algoritma SIFT Studi Kasus: Deteksi Penyakit Kulit Sederhana. *Jurnal Teknik Informatika Dan Sistem Informasi*, 2(3). <https://doi.org/10.28932/JUTISI.V2I3.647>
- Gomez-Rodriguez, M., Gummadi, K. P., & Schölkopf, B. (2014). Quantifying Information Overload in Social Media and its Impact on Social Contagions. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, 170–179.

- <https://doi.org/10.1609/icwsm.v8i1.14549>
- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: an Overview*. <https://arxiv.org/abs/2008.05756v1>
- Hadinata, K. C., Navastara, D. A., & Fabroyir, H. (2022). Ekstraksi Informasi pada Dokumen Teks Menggunakan Metode Named-Entity Recognition untuk Sistem Autofill Formulir Lowongan SIM Magang MyITS StudentConnect. *Jurnal Teknik ITS*, 11(1), 501478. <https://doi.org/10.12962/J23373539.V11I1.83074>
- Hasan, K., Utami, A., Eni, S., Izzah, N., & Ramadhana, S. C. (2023). KOMUNIKASI DI ERA DIGITAL: ANALISIS MEDIA KONVENSIONAL VS NEW MEDIA PADA KALANGAN MAHASISWA ILMU KOMUNIKASI UNIVERSITAS MALIKUSSALEH ANGKATAN 2021. *Jurnal Komunikasi Pemberdayaan*, 2(1), 56–63. <https://doi.org/10.47431/JKP.V2I1.302>
- Hendrycks, D., & Gimpel, K. (2016). *Gaussian Error Linear Units (GELUs)*. <https://arxiv.org/abs/1606.08415v5>
- Ho, Y., & Wookey, S. (2020). The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. *IEEE Access*, 8, 4806–4813. <https://doi.org/10.1109/ACCESS.2019.2962617>
- Hosseini, H., Xiao, B., & Poovendran, R. (2017). Google’s Cloud Vision API Is Not Robust To Noise. *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017, 2017-December*, 101–105. <https://doi.org/10.1109/ICMLA.2017.0-172>
- Huang, Y., Lv, T., Cui, L., Lu, Y., & Wei, F. (2022). LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking. *MM 2022 - Proceedings of the 30th ACM International Conference on Multimedia*, 4083–4091. <https://doi.org/10.1145/3503161.3548112>
- Hussain, M., Bird, J. J., & Faria, D. R. (2019). A Study on CNN Transfer Learning for Image Classification. *Advances in Intelligent Systems and Computing*, 840, 191–202. https://doi.org/10.1007/978-3-319-97982-3_16
- Jiang, J., Huang, W., Zhang, M., Suzuki, T., & Nie, L. (2024). Unveil Benign Overfitting for Transformer in Vision: Training Dynamics, Convergence, and Generalization. *Advances in Neural Information Processing Systems*, 37, 135464–135625. <https://doi.org/10.52202/079017-4306>
- Jiang, X., Wang, W., Tian, S., Wang, H., Lookman, T., & Su, Y. (2025). Applications of natural language processing and large language models in materials discovery. *Npj Computational Materials* 2025 11:1, 11(1), 79-. <https://doi.org/10.1038/s41524-025-01554-0>
- Kaffah, L. S., Setiawan, D., & Waluyo, E. (2023). PEMANFAATAN MEDIA CETAK POSTER DALAM PEMBELAJARAN BAHASA INDONESIA DENGAN MENGGUNAKAN APLIKASI CANVA DI KELAS V SD. *Jurnal Ilmiah Wahana Pendidikan*, 9(16), 482–492. <https://doi.org/10.5281/ZENODO.8246435>
- Keraghel, I., Morbieu, S., & Nadif, M. (2024). *Recent Advances in Named Entity Recognition: A Comprehensive Survey and Comparative Study*. <https://arxiv.org/abs/2401.10825v3>
- Khandelwal, A., Kar, A., Chikka, V. R., & Karlapalem, K. (2022). Biomedical NER using Novel Schema and Distant Supervision. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 155–160. <https://doi.org/10.18653/V1/2022.BIONLP-1.15>
- Kristanto, A. (2016). *Media Pembelajaran* (1st ed.). Penerbit Bintang Sutabaya.
- Kumalija, E., & Nakamoto, Y. (2022). Performance evaluation of automatic speech

- recognition systems on integrated noise-network distorted speech. *Frontiers in Signal Processing*, 2, 999457. <https://doi.org/10.3389/FRSIP.2022.999457/BIBTEX>
- Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural Architectures for Named Entity Recognition. *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, 260–270. <https://doi.org/10.18653/V1/N16-1030>
- Marshall, C. C., McCown, F., & Nelson, M. L. (2007). Evaluating Personal Archiving Strategies for Internet-based Information. *Archiving 2007 - Final Program and Proceedings*, 151–156. <https://arxiv.org/abs/0704.3647v1>
- Morris, A. C., Maier, V., & Green, P. (2004). From WER and RIL to MER and WIL: improved evaluation measures for connected speech recognition. *8th International Conference on Spoken Language Processing, ICSLP 2004*, 2765–2768. <https://doi.org/10.21437/INTERSPEECH.2004-668>
- Mosbach, M., Andriushchenko, M., & Klakow, D. (2020). On the Stability of Fine-tuning BERT: Misconceptions, Explanations, and Strong Baselines. *International Conference on Learning Representations*.
- Nguyen, N. D., Tan, W., Du, L., Buntine, W., Beare, R., & Chen, C. (2023). Low-Resource Named Entity Recognition: Can One-vs-All AUC Maximization Help? *Proceedings - IEEE International Conference on Data Mining, ICDM*, 1241–1246. <https://doi.org/10.1109/ICDM58522.2023.00155>
- Okurowski, M. E. (1993). *Information Extraction Overview*. 117–121. <https://doi.org/10.3115/1119149.1119164>
- Pendit, I. K. D. (2019). STRATEGI PROMOSI MELALUI MEDIA DESAIN POSTER DALAM PENCITRAAN PUBLIK. *Stilistika : Jurnal Pendidikan Bahasa Dan Seni*, 7(2), 192–210. <https://doi.org/10.59672/STILISTIKA.V7I2.283>
- Pinasti, W., & Suadaa, L. H. (2024). Named Entity Recognition in Statistical Dataset Search Queries. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 13(3), 171–177. <https://doi.org/10.22146/JNTETI.V13I3.11580>
- Puspitarani, Y. (2021). Indonesian Information Extraction : Challenges and Opportunities. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(1), 421–429. <https://doi.org/10.35957/JATISI.V8I1.710>
- Rachmad, D. S. (2020). Review Named Entity Recognition dengan Menggunakan Machine Learning. *Jurnal Sains Dan Informatika*, 6(1), 28–33. <https://doi.org/10.34128/JSI.V6I1.204>
- Rana, C., & Shokeen, J. (2018). A review on the dynamics of social recommender systems. *International Journal of Web Engineering and Technology*, 13(3), 255. <https://doi.org/10.1504/IJWET.2018.10016164>
- Rifani, R., Bijaksana, A. M., & Asror, I. (2019). Named Entity Recognition for an Indonesian Based Language Tweet using Multinomial Naive Bayes Classifier. *Indonesian Journal on Computing (Indo-JC)*, 4(2), 119–126. <https://doi.org/10.34818/INDOJC.2019.4.2.330>
- Rosidy, A. S., Akhriza, T. M., & Husni, M. (2020). Kombinasi metode NER-OCR untuk meningkatkan efisiensi pengambilan informasi di poster berbahasa Indonesia. *Jurnal Teknologi Dan Sistem Komputer*, 8(4), 263–269. <https://doi.org/10.14710/JTSISKOM.2020.13686>
- Sadaiyandi, J., Arumugam, P., Sangaiah, A. K., & Zhang, C. (2023). Stratified Sampling-Based Deep Learning Approach to Increase Prediction Accuracy of Unbalanced Dataset. *Electronics* 2023, Vol. 12, Page 4423, 12(21), 4423.

- <https://doi.org/10.3390/ELECTRONICS12214423>
- Sarawagi, S. (2008). Information Extraction. *Foundations and Trends in Databases*, 1(3), 261–377. <https://doi.org/10.1561/19000000003>
- Sari, C. P., Mafruhah, O. R., Fajria, R. N., & Meta, A. (2019). Evaluasi Pelayanan Resep Berdasarkan Pelaksanaan Standar Kefarmasian di Apotek Tempat Praktik Kerja Profesi Apoteker (PKPA) Kota Yogyakarta. *Jurnal Pharmascience*, 6(1), 18–29. <https://doi.org/10.20527/JPS.V6I1.6071>
- Seni, J., Fsr, M., Yogyakarta, I., & Putri, N. D. (2019). KAJIAN SEMIOTIKPOSTER “BALI TOLAK REKLAMASI” KARYA ALIT AMBARA. *Journal of Contemporary Indonesian Art*, 5(1), 24–41. <https://doi.org/10.24821/JOCIA.V5I1.2522>
- SETIAWAN, M. R. R. (2022). *Ekstraksi Informasi Beasiswa dari Media Sosial menggunakan BiLSTM-CRF*. <https://repository.telkomuniversity.ac.id/home/catalog/id/181388/slug/ekstraksi-informasi-beasiswa-dari-media-sosial-menggunakan-bilstm-crf.html>
- Shi, E., Wang, Y., Zhang, H., Du, L., Han, S., Zhang, D., & Sun, H. (2023). Towards Efficient Fine-Tuning of Pre-trained Code Models: An Experimental Study and Beyond. *ISSTA 2023 - Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis*, 39–51. <https://doi.org/10.1145/3597926.3598036>
- Sugiarta, N. P. A. S. A., & Sanjaya, N. A. (2021). Location Named-Entity Recognition using Rule-Based Approach for Balinese Texts. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 9(3), 435–442. <https://doi.org/10.24843/JLK.2021.V09.I03.P15>
- Sumartono, S., & Astuti, H. (2018). PENGGUNAAN POSTER SEBAGAI MEDIA KOMUNIKASI KESEHATAN. *KOMUNIKOLOGI: Jurnal Ilmiah Ilmu Komunikasi*, 15(1). <https://doi.org/10.47007/JKOMU.V15I1.187>
- Sumarudin, M., & Syafrullah, M. (2024). Named Entity Recognition In Electronic Medical Records Based On Hybrid Neural Network And Transformer. *Eduvest - Journal of Universal Studies*, 4(6), 5263–5279. <https://doi.org/10.59188/EDUVEST.V4I6.1473>
- Sunik, B. (2013). Theory of meaningful information, background, excerpts and meaning representation. *Artificial Intelligence Research*, 2(3), 102. <https://doi.org/10.5430/AIR.V2N3P102>
- Thammarak, K., Kongkla, P., Sirisathitkul, Y., & Intakosum, S. (2022). Comparative analysis of Tesseract and Google Cloud Vision for Thai vehicle registration certificate. *International Journal of Electrical and Computer Engineering*, 12(2), 1849–1858. <https://doi.org/10.11591/IJECE.V12I2.PP1849-1858>
- Utami, A. E., Nurhayati, O. D., & Martono, K. T. (2016). Aplikasi Penerjemah Bahasa Inggris – Indonesia dengan Optical Character Recognition Berbasis Android. *Jurnal Teknologi Dan Sistem Komputer*, 4(1), 167–177. <https://doi.org/10.14710/JTSISKOM.4.1.2016.167-177>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, 30.
- Wahyuni, N. M. S., & Sanjaya, N. A. (2021). Rule-based Named Entity Recognition (NER) to Determine Time Expression for Balinese Text Document. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 9(4), 555–562. <https://doi.org/10.24843/JLK.2021.V09.I04.P14>
- Walker, J., Fujii, Y., & Popat, A. C. (2018). A web-based ocr service for documents. *Proceedings of the 13th IAPR International Workshop on Document Analysis Systems (DAS), Vienna, Austria, 1*.

- Wulandari, D. W., Adikara, P. P., & Adinugroho, S. (2018). Named Entity Recognition (NER) Pada Dokumen Biologi Menggunakan Rule Based dan Naive Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 4555–4563. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2988>
- Ye, W., Li, B., Xie, R., Sheng, Z., Chen, L., & Zhang, S. (2019). Exploiting Entity BIO Tag Embeddings and Multi-task Learning for Relation Extraction with Imbalanced Data. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 1351–1360. <https://doi.org/10.18653/V1/P19-1130>
- Yunmar, R. A. (2020). Ontology Learning pada Teks Tidak Terstruktur. *Electrician Jurnal Rekayasa Dan Teknologi Elektro*, 14(2), 46–51. <https://doi.org/10.23960/ELC.V14N2.2139>
- Zhang, Z., & Chen, A. L. P. (2022). Biomedical named entity recognition with the combined feature attention and fully-shared multi-task learning. *BMC Bioinformatics*, 23(1), 1–21. <https://doi.org/10.1186/S12859-022-04994-3/TABLES/8>
- Zhao, S., Qiu, Z., & He, Y. (2021). Transfer learning strategy for plastic pollution detection in soil: Calibration transfer from high-throughput HSI system to NIR sensor. *Chemosphere*, 272, 129908. <https://doi.org/10.1016/J.CHEMOSPHERE.2021.129908>
- Zheng, H., Wang, S., Huang, L., & Tech, V. (2024). A Comprehensive Survey on Document-Level Information Extraction. *Proceedings of the Workshop on the Future of Event Detection (FuturED)*, 58–72. <https://doi.org/10.18653/V1/2024.FUTURED-1.6>
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109(1), 43–76. <https://doi.org/10.1109/JPROC.2020.3004555>
- Julianto, S. (2022). Penerapan Optical Character Recognition (OCR) untuk penjadwalan acara dengan Google Cloud Vision API algoritma Convolution Neural Network (CNN) (Skripsi sarjana, Universitas Pembangunan Nasional “Veteran” Yogyakarta).
- Putri, D. B. (2020). Optical Character Recognition (OCR) untuk deteksi kartu identitas menggunakan Mobile Vision berbasis Android (Skripsi sarjana, Universitas Pembangunan Nasional “Veteran” Yogyakarta).
- Google Cloud. (2025). *Detect text in images*. Diakses 19 Juni 2025, dari <https://cloud.google.com/vision/docs/ocr>
- Eventkampus. (2025). Seminar Nasional Kewirausahaan 2018. Diakses 15 Juni 2025, dari <https://eventkampus.com/event/detail/1830/seminar-nasional-kewirausahaan-2018>
- Wikipedia. (2025). Universitas Pembangunan Nasional Veteran Yogyakarta. Diakses 6 September 2025, dari https://id.wikipedia.org/wiki/Universitas_Pembangunan_Nasional_Veteran_Yogyakarta
- Medium. (2025). Review: Semi-Supervised Sequence Tagging with Bidirectional Language Models (TagLM). Diakses 7 September 2025, dari <https://sh-tsang.medium.com/review-semi-supervised-sequence-tagging-with-bidirectional-language-models-taglm-15485e98d15>
- Dicoding. (2025). *Dicoding Event*. Diakses 12 November 2025, dari <https://www.dicoding.com/events>