

DAFTAR PUSTAKA

- A. Mullen, L., Benoit, K., Keyes, O., Selivanov, D., dan Arnold, J. (2018). Fast, Consistent Tokenization of Natural Language Text. *Journal of Open Source Software*, 3(23), 655. <https://doi.org/10.21105/joss.00655>
- Abbara, S., Hafez, M., Kazzaz, A., Alhothali, A., dan Alsolami, A. (2023). *ALJP: An Arabic Legal Judgment Prediction in Personal Status Cases Using Machine Learning Models*. <http://arxiv.org/abs/2309.00238>
- Allan, J., dan Kumaran, G. (2003). Stemming in the language modeling framework. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, 455–456. <https://doi.org/10.1145/860435.860548>
- Anantathanavit, R., Chongthanakorn, J., Kongsantinart, W., Praiwattana, P., dan Thaipisitukul, T. (2024). Utilizing AI and Natural Language Processing to Predict Supreme Court Decisions in Thailand. *KST 2024 - 16th International Conference on Knowledge and Smart Technology*, 45–50. <https://doi.org/10.1109/KST61284.2024.10499665>
- Ansari, T., Dhillon, H. S., dan Singh, M. (2024). Machine Learning Model to Predict Results of Law Cases. *2024 4th International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies, ICAECT 2024*. <https://doi.org/10.1109/ICAECT60202.2024.10468789>
- Beaufays, F., dan Strope, B. (2013). Language model capitalization. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 6749–6752. <https://doi.org/10.1109/ICASSP.2013.6638968>
- Bojanowski, P., Grave, E., Joulin, A., dan Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135–146.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Carnegie Mellon University Machine Learning Blog. (2020). 4. *Overfitting*.
- Chen, H., Wu, L., Chen, J., Lu, W., dan Ding, J. (2022). A comparative study of automated legal text classification using random forests and deep learning. *Information Processing dan Management*, 59(2), 102798. <https://doi.org/10.1016/j.ipm.2021.102798>
- Chinchor, N. (1992). MUC-4 Evaluation Metrics. *Proceedings of the 4th Conference on Message Understanding*, 22–29.
- Demilie, W. B. (2020). Implemented Stemming Algorithms for Information Retrieval Applications. *Journal of Information Engineering and Applications*. <https://doi.org/10.7176/JIEA/10-3-01>
- Deng, J. (2024). The Data Preprocessing Technique in Machine Learning. *Science and Technology of Engineering, Chemistry and Environmental Protection*, 1(10). <https://doi.org/10.61173/0hfqqz07>
- Effendy, M. (2012). *Teori Hukum dan Aplikasinya*. Referensi.
- Gupta, A., dan Sharma, R. (2023). Overfitting in Machine Learning. In *Machine Learning for Data Science* (pp. 235–254). Springer. https://doi.org/10.1007/978-3-031-39355-6_10

- Harahap, M. Y. (2005). *Pembahasan Permasalahan dan Penerapan KUHAP: Pemeriksaan Sidang Pengadilan, Banding, Kasasi, dan Peninjauan Kembali*. Sinar Grafika.
- Hassler, M., dan Fliedl, G. (2006). Text preparation through extended tokenization. *Data Mining VII: Data, Text and Web Mining and Their Business Applications*, 13–21. <https://doi.org/10.2495/DATA060021>
- Hastie, T., Tibshirani, R., dan Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- Jurafsky, D., dan Martin, J. H. (2020). *Speech and Language Processing*. Pearson.
- K., J., dan R., J. (2016). Stop-Word Removal Algorithm and its Implementation for Sanskrit Language. *International Journal of Computer Applications*, 150(2), 15–17. <https://doi.org/10.5120/ijca2016911462>
- Khan, N., Bakht, M. P., Khan, M. J., Samad, A., dan Sahar, G. (2019). Spotting Urdu Stop Words By Zipf's Statistical Approach. *2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*, 1–5. <https://doi.org/10.1109/MACS48846.2019.9024817>
- Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. In *Informatica* (Vol. 31).
- Ladani, D. J., dan Desai, N. P. (2020). Stopword Identification and Removal Techniques on TC and IR applications: A Survey. *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 466–472. <https://doi.org/10.1109/ICACCS48705.2020.9074166>
- Liaw, A., dan Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Lim, J. P., Tan, S., Choo, D., dan Lauw, H. W. (2025). *A partition cover approach to tokenization*. <http://arxiv.org/abs/2501.06246>
- Lind, E. A., dan Tyler, T. R. (1988). *The social psychology of procedural justice*. Springer Science dan Business Media.
- Madrakhimov, S., Makharov, K., dan Lolaev, M. (2021). *Data preprocessing on input*. 030003. <https://doi.org/10.1063/5.0058132>
- Malik, V., Sanjay, R., Nigam, S. K., Ghosh, K., Guha, S. K., Bhattacharya, A., dan Modi, A. (2021). *ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation*. <http://arxiv.org/abs/2105.13562>
- Manning, C. D., Raghavan, P., dan Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- Mikolov, T., Chen, K., Corrado, G., dan Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *1st International Conference on Learning Representations (ICLR 2013), Workshop Track Proceedings*. <http://arxiv.org/abs/1301.3781>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., dan Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26.
- Mumcuoğlu, E., Öztürk, C. E., Ozaktas, H. M., dan Koç, A. (2021). Natural language processing in law: Prediction of outcomes in the higher courts of Turkey. *Information Processing and Management*, 58(5). <https://doi.org/10.1016/j.ipm.2021.102684>

- Mustari, N., Sen, S. K., Banik, A., Mehedi, M. H. K., dan Rasel, A. A. (2023). Techniques to Estimate the Status of Legal Proceedings Considering Sequential Text Data. *2023 International Conference on Emerging Smart Computing and Informatics, ESCI 2023*. <https://doi.org/10.1109/ESCI56872.2023.10099995>
- National Center for Biotechnology Information. (2021). *Overfitting and Underfitting in Machine Learning*.
- Noguti, M. Y., Vellasques, E., dan Oliveira, L. S. (2020). *Legal Document Classification: An Application to Law Area Prediction of Petitions to Public Prosecution Service*. <https://doi.org/10.1109/IJCNN48605.2020.9207211>
- Nuranti, E. Q., Yulianti, E., dan Husin, H. S. (2022). Predicting the Category and the Length of Punishment in Indonesian Courts Based on Previous Court Decision Documents. *Computers, 11(6)*. <https://doi.org/10.3390/computers11060088>
- Pavani, B. V., Mahitha, D., Prabhakar, P., dan Pati, P. B. (2024). Identifying Sentiment in Legal Case Judgments using Random Forest Classifier. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 1–5. <https://doi.org/10.1109/I2CT61223.2024.10544116>
- Pidishetti, R. V., dan Chinta, U. (2025). Breaking the Fake: A CNN-BiLSTM Approach for Fake News Detection on Social Media. *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)*, 01118–01124. <https://doi.org/10.1109/CCWC62904.2025.10903759>
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14(3)*, 130–137.
- Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies, 2(1)*, 37–63.
- Pramudita, D. A., dan Santoso, A. (2020). Analisis putusan pengadilan melalui text mining: Studi pada Direktori Putusan Mahkamah Agung. *Jurnal Hukum Dan Pembangunan, 50(3)*, 723–744. <https://doi.org/10.21143/jhp.vol50.no3.2675>
- Pratiwi, F. S. (2023). *Data Jumlah Kejahatan di Indonesia pada 2023*. <https://dataindonesia.id/varia/detail/data-jumlah-kejahatan-di-indonesia-pada-2023>
- Radbruch, G. (1973). *Legal Philosophy*. Oxford University Press.
- Rahardjo, S. (2010). *Hukum dalam jagat ketertiban*. UKI Press.
- Rahman, A., dan Ahmed, S. (2023). Overfitting in Machine Learning: Causes, Consequences, and Remedies. *Journal of Information Systems Engineering dan Management, 8(2)*, 2072. <https://www.jisem-journal.com/index.php/journal/article/download/2072/798/3321>
- Rai, A., dan Borah, S. (2021). *Study of Various Methods for Tokenization* (pp. 193–200). https://doi.org/10.1007/978-981-15-6198-6_18
- Sage, A. (2018). *Random forest robustness, variable importance, and tree aggregation* [Iowa State University, Digital Repository]. <https://doi.org/10.31274/etd-180810-6083>
- Schmidt, C. W., Reddy, V., Zhang, H., Alameddine, A., Uzan, O., Pinter, Y., dan Tanner, C. (2024). *Tokenization Is More Than Compression*. <http://arxiv.org/abs/2402.18376>

- scikit-learn developers. (2025). *Metrics and Scoring: Quantifying the Quality of Predictions*.
- Sen, P. C., Hajra, M., dan Ghosh, M. (2020). *Supervised Classification Algorithms in Machine Learning: A Survey and Review* (pp. 99–111). https://doi.org/10.1007/978-981-13-7403-6_11
- Shaikh, R. A., Sahu, T. P., dan Anand, V. (2020). Predicting Outcomes of Legal Cases based on Legal Factors using Classifiers. *Procedia Computer Science*, 167, 2393–2402. <https://doi.org/10.1016/j.procs.2020.03.292>
- Sharma, A., Verma, C. K., dan Singh, P. (2025). Enhancing Option Pricing Accuracy in the Indian Market: A CNN-BiLSTM Approach. *Computational Economics*, 65(6), 3751–3778. <https://doi.org/10.1007/s10614-024-10689-z>
- Siddiqua, B., dan Adepu, Dr. A. (2024, November 7). Integrating NLP and Random Forest Algorithms for Clinical Decision Support: Leveraging EHR Data to Improve Patient Care and Operational Efficiency. *Proceedings of the International Conference on Industrial Engineering and Operations Management*. <https://doi.org/10.46254/IN04.20240089>
- Sidharta, B. A. (2008). *Ilmu hukum Indonesia: Upaya pengembangan ilmu hukum sistematik yang responsif terhadap perubahan masyarakat*. Genta Publishing.
- Singh, A., Thakur, N., dan Sharma, A. (2016). A review of supervised machine learning algorithms. *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 1310–1315.
- Sokolova, M., dan Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing dan Management: An International Journal*, 45(4), 427–437.
- Strickson, B., dan De La Iglesia, B. (2020). Legal Judgement Prediction for UK Courts. *ACM International Conference Proceeding Series*, 204–209. <https://doi.org/10.1145/3388176.3388183>
- Taghi-Zadeh, H., Sadreddini, M. H., Diyanati, M. H., dan Rasekh, A. H. (2015). A New Hybrid Stemming Method for Persian Language. *Digital Scholarship in the Humanities*, fqv053. <https://doi.org/10.1093/llc/fqv053>
- Tanya, B. L. (2017). *Teori Hukum: Strategi Tertib Manusia Lintas Ruang dan Generasi*. Genta Publishing.
- Upadhyaya, B., Sharma, K., dan Gurung, S. (2022). *A Survey on Various Stemming Techniques for Hindi and Nepali Language* (pp. 137–142). https://doi.org/10.1007/978-981-16-2911-2_14
- Wilbur, W. J., dan Sirotkin, K. (1992). The automatic identification of stop words. *Journal of Information Science*, 18(1), 45–55. <https://doi.org/10.1177/016555159201800106>
- Zhuang, L., Hu, J., Wang, Q., Wang, Y., Zhang, K., dan Liu, S. (2024). An Improved Lightweight CNN_BiLSTM Model for Network Intrusion Detection Using the UNSW-NB15 Dataset. *2024 2nd International Conference on Computer, Vision and Intelligent Technology (ICCVIT)*, 1–9. <https://doi.org/10.1109/ICCVIT63928.2024.10872444>