

DAFTAR PUSTAKA

Alakukku, L. (2022). Domain-Specific Boilerplate Removal from Web Pages with Entropy and Clustering. *Tesis, Aalto University*.

Alarte, J., & Silva, J. (2021). Page-Level main content extraction from heterogeneous webpages. *ACM Transactions on Knowledge Discovery From Data*, 15(6), 1-105. <https://doi.org/10.1145/3451168>

Bevendorff, J., Gupta, S., Kiesel, J., & Stein, B. (2023). An Empirical Comparison of Web Content Extraction Algorithms. *Proceedings of the 46th ACM SIGIR Conference*. <https://doi.org/10.1145/3539618.3591920>

Desai, M., & Shah, M. (2020). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clinical eHealth*, 4, 1–11. <https://doi.org/10.1016/j.ceh.2020.11.002>

G. Jung, S. Han, H. Kim, K. Kim and J. Cha, "Extracting the Main Content of Web Pages Using the First Impression Area," in *IEEE Access*, vol. 10, pp. 129958-129969, 2022, doi: 10.1109/ACCESS.2022.3229080.

Hamayel, M. J., & Owda, A. Y. (2021). A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms. *AI*, 2(4), 477–496. <https://doi.org/10.3390/ai2040030>

Ifty, Al Amin. (2023). Introduction to the Perceptron and Its Applications. <https://doi.org/10.13140/RG.2.2.10439.10404>

Ismail, A., & Kuppusamy, K. (2019). Web accessibility investigation and identification of major issues of higher education websites with statistical measures: A case study of college websites. *Journal of King Saud University - Computer and Information Sciences*, 34(3), 901–911. <https://doi.org/10.1016/j.jksuci.2019.03.011>

Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(11), 1-15.

Kohlschütter, C., Fankhauser, P., & Nejdl, W. (2010). Boilerplate Detection using Shallow Text Features. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM '10)* (pp. 441–450). ACM. <https://doi.org/10.1145/1718487.1718542>

Leonhardt, J., Anand, A., & Khosla, M. (2020). Boilerplate Removal using a Neural Sequence Labeling Model. *Companion Proceedings of the Web Conference 2018*, 226–229. <https://doi.org/10.1145/3366424.3383547>

Lindemann, B., Maschler, B., Sahlab, N., & Weyrich, M. (2021). A survey on anomaly detection for technical systems using LSTM networks. *Computers in Industry*, 131, 103498. <https://doi.org/10.1016/j.compind.2021.103498>

Muehlethaler, C., & Albert, R. (2021). Collecting data on textiles from the internet using web crawling and web scraping tools. *Forensic Science International*, 322, 110753. <https://doi.org/10.1016/j.forsciint.2021.110753>

M. Fernández-Pichel, & M. Prada-Corral. (2024). An Unsupervised Perplexity-based Method for Boilerplate Removal. *Natural Language Engineering*, 30(1). <https://doi.org/10.1017/S1351324923000049>.

M. Dampfhofer, T. Mesquida, and J. Kummert, "Backpropagation-based learning techniques for deep spiking neural networks: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

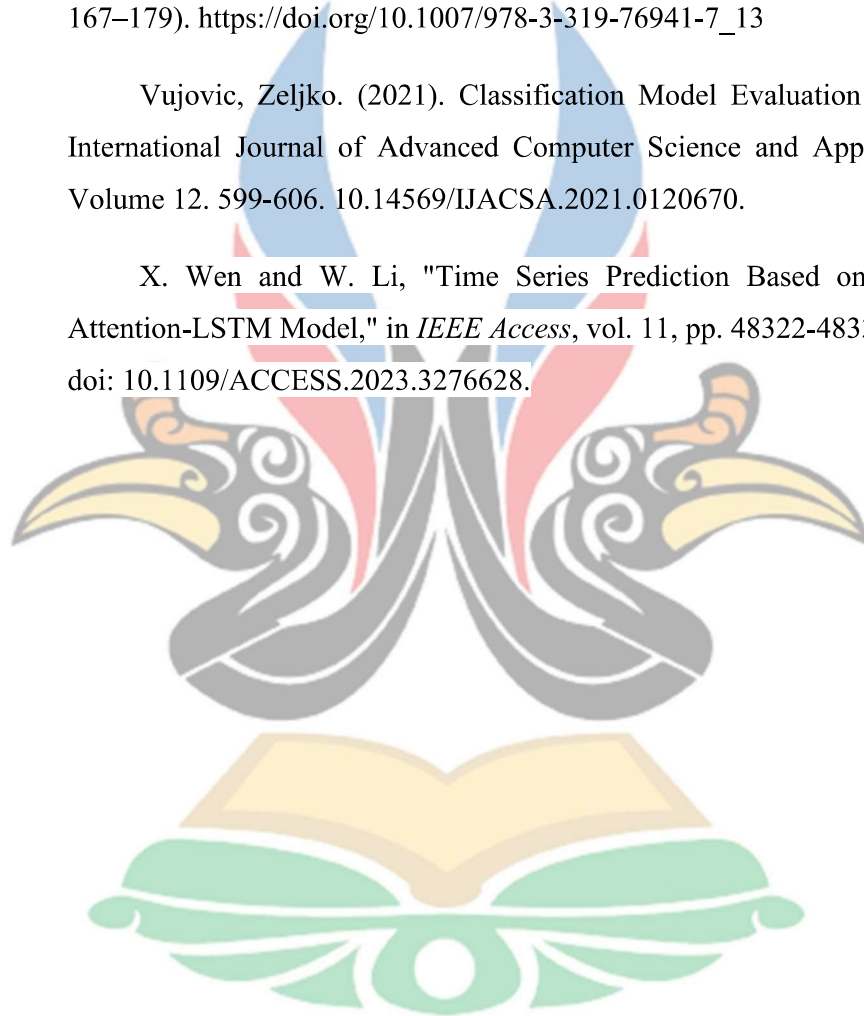
Sharifani, K., & Amini, M. (2023). Machine Learning and Deep Learning: A Review of Methods and Applications. *World Information Technology and Engineering Journal*, 10(7), 3897-3904. Available at SSRN: <https://ssrn.com/abstract=4458723>

Singrodia, V., Mitra, A., & Paul, S. (2019). A Review on Web Scraping and its Applications. *2019 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 1-6. <https://doi.org/10.1109/ICCCI.2019.8821809>

Vogels, T., Ganea, O., & Eickhoff, C. (2018). Web2Text: Deep structured boilerplate removal. In *Lecture notes in computer science* (pp. 167–179). https://doi.org/10.1007/978-3-319-76941-7_13

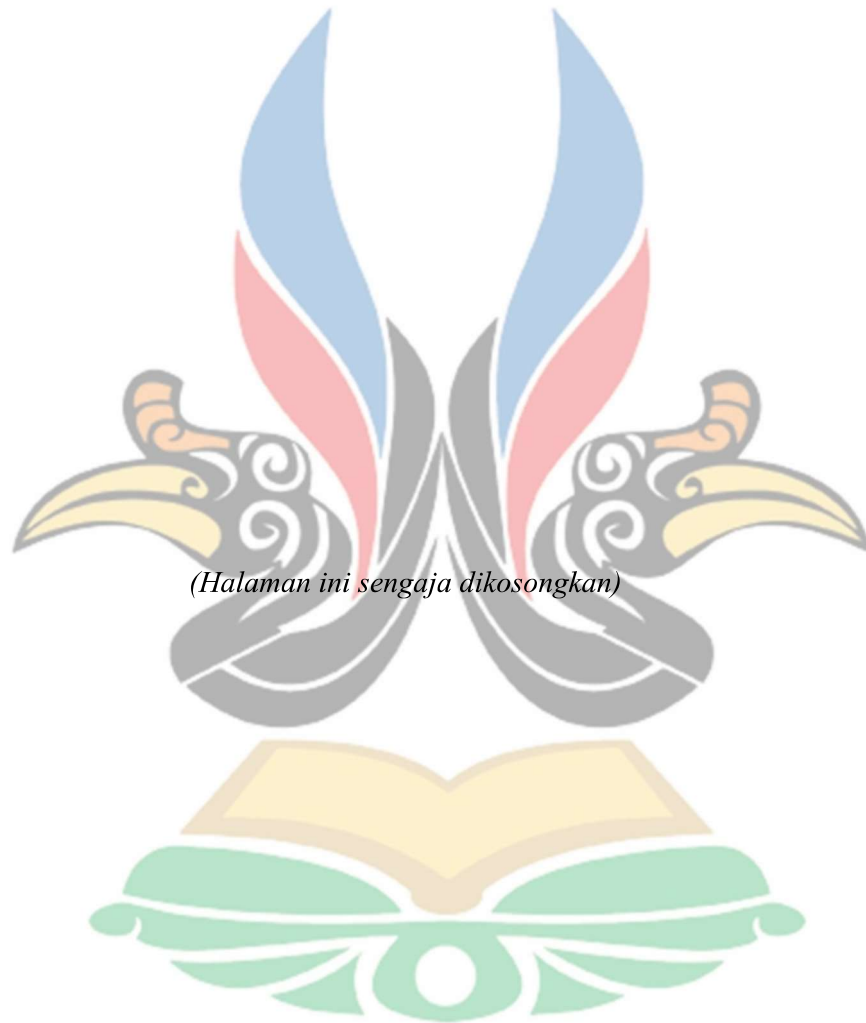
Vujovic, Zeljko. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*. Volume 12. 599-606. 10.14569/IJACSA.2021.0120670.

X. Wen and W. Li, "Time Series Prediction Based on LSTM-Attention-LSTM Model," in *IEEE Access*, vol. 11, pp. 48322-48331, 2023, doi: 10.1109/ACCESS.2023.3276628.



www.itk.ac.id

www.itk.ac.id



(Halaman ini sengaja dikosongkan)

www.itk.ac.id