

DAFTAR PUSTAKA

- Ahmadian, H., Abidin, T. F., Riza, H., & Muchtar, K. (2024). Hybrid Models for Emotion Classification and Sentiment Analysis in Indonesian Language. *Applied Computational Intelligence and Soft Computing*, 2024(1). <https://doi.org/10.1155/2024/2826773>
- Aliyah, N. E., Sholikah, R. W., Firdausi, H., Ciptaningtyas, H. T., & Sabilla, I. A. (2025). Enhancing Automated Essay Scoring in Bahasa Indonesia with IndoBERT and IndoSBERT. *2025 International Conference on Smart Computing, IoT and Machine Learning (SIML)*, 1–7. <https://doi.org/10.1109/SIML65326.2025.11080721>
- Altalhan, M., Algarni, A., & Turki-Hadj Alouane, M. (2025). Imbalanced Data Problem in Machine Learning: A Review. *IEEE Access*, 13, 13686–13699. <https://doi.org/10.1109/ACCESS.2025.3531662>
- Andersen, Ø. E., Yuan, Z., Watson, R., & Cheung, K. Y. F. (2021). Benefits of Alternative Evaluation Methods for Automated Essay Scoring. *Proceedings of the 14th International Conference on Educational Data Mining*, 856–864. <https://eric.ed.gov/?id=ED615497>
- Anthony, P., & Zhou, J. (2025). Leveraging Transformer with Self-Attention for Multi-Label Emotion Classification in Crisis Tweets. *Informatics*, 12(4), 114. <https://doi.org/10.3390/informatics12040114>
- Atkinson, J., & Palma, D. (2025). An LLM-based hybrid approach for enhanced automated essay scoring. *Scientific Reports*, 15(1), 14551. <https://doi.org/10.1038/s41598-025-87862-3>
- Aulamo, M., Virpioja, S., & Tiedemann, J. (2020). OpusFilter: A Configurable Parallel Corpus Filtering Toolbox. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 150–156. <https://doi.org/10.18653/v1/2020.acl-demos.20>
- Azhari, A., Santoso, A., Ratna, A. A. P., & Prestiliano, J. (2024). Optimization of AES using BERT and BiLSTM for Grading the Online Exams. *International Journal of Intelligent Engineering & Systems*, 17(5). <https://doi.org/10.22266/ijies2024.1031.31>
- Bertolini, R., Finch, S. J., & Nehm, R. H. (2022). Quantifying variability in predictions of student performance: Examining the impact of bootstrap resampling in data pipelines. *Computers and Education: Artificial Intelligence*, 3, 100067. <https://doi.org/10.1016/j.caeai.2022.100067>
- Bestgen, Y. (2022). Please, Don't Forget the Difference and the Confidence Interval when Seeking for the State-of-the-Art Status. Dalam N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, & S. Piperidis (Ed.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (hlm. 5956–5962). European Language Resources Association. <https://aclanthology.org/2022.lrec-1.640/>
- Bui, N. M., & Barrot, J. S. (2025). ChatGPT as an automated essay scoring tool in the writing classrooms: how it compares with human scoring. *Education and Information Technologies*, 30(2), 2041–2058. <https://doi.org/10.1007/s10639-024-12891-w>

- Cahyo, P. W., Aesy, U. S., Setianto, W. A., & Sulaiman, T. (2025). A Novel Named Entity Recognition approach of Indonesian fake news using part of speech and BERT model on presidential election. *International Journal of Information Management Data Insights*, 5(2), 100354. <https://doi.org/10.1016/j.jjime.2025.100354>
- Chakravarty, A., Brenchley, M., Breakspear, T., Lewin, I., & Huang, Y. (2025). Enhancing Marker Scoring Accuracy through Ordinal Confidence Modelling in Educational Assessments. *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)*, 1498–1507. <https://doi.org/10.18653/v1/2025.acl-industry.106>
- Chamidah, N., Yulianti, E., & Budi, I. (2023). Evaluating the Impact of Sentence Tokenization on Indonesian Automated Essay Scoring Using Pretrained Sentence Embeddings. *Revue d'Intelligence Artificielle*, 37(5), 1101–1108. <https://doi.org/10.18280/ria.370502>
- Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: latest research, applications and future directions. *Artificial Intelligence Review*, 57(6), 137. <https://doi.org/10.1007/s10462-024-10759-6>
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. <https://doi.org/10.1037/h0026256>
- Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heafield, K., Heffernan, K., Kalbassi, E., Lam, J., Licht, D., Maillard, J., Sun, A., Wang, S., Wenzek, G., Youngblood, A., Akula, B., Barrault, L., Gonzalez, G. M., Hansanti, P., Hoffman, J., ... Wang, J. (2022). *No Language Left Behind: Scaling Human-Centered Machine Translation*. <https://arxiv.org/abs/2207.04672>
- Crossley, S. A., Tian, Y., Baffour, P., Franklin, A., Benner, M., & Boser, U. (2024). A large-scale corpus for assessing written argumentation: PERSUADE 2.0. *Assessing Writing*, 61, 100865. <https://doi.org/10.1016/j.asw.2024.100865>
- Cui, Y., Jia, M., Lin, T.-Y., Song, Y., & Belongie, S. (2019). Class-Balanced Loss Based on Effective Number of Samples. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9260–9269. <https://doi.org/10.1109/CVPR.2019.00949>
- Dai, X., Chalkidis, I., Darkner, S., & Elliott, D. (2022). Revisiting Transformer-based Models for Long Document Classification. *Findings of the Association for Computational Linguistics: EMNLP 2022*, 7212–7230. <https://doi.org/10.18653/v1/2022.findings-emnlp.534>
- Dave, E., & Chowanda, A. (2024). IPerFEX-2023: Indonesian personal financial entity extraction using indoBERT-BiGRU-CRF model. *Journal of Big Data*, 11(1), 139. <https://doi.org/10.1186/s40537-024-00987-6>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Dalam J. Burstein, C. Doran, & T. Solorio (Ed.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (hlm. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
- Dhini, B. F., Girsang, A. S., Sufandi, U. U., & Kurniawati, H. (2023). Automatic essay scoring for discussion forum in online learning based on semantic and

- keyword similarities. *Asian Association of Open Universities Journal*, 18(3), 262–278. <https://doi.org/10.1108/AAOUJ-02-2023-0027>
- Do, H., Kim, Y., & Lee, G. (2024). Autoregressive Score Generation for Multi-trait Essay Scoring. *Findings of the Association for Computational Linguistics: EACL 2024*, 1659–1666. <https://doi.org/10.18653/v1/2024.findings-eacl.115>
- Doewes, A., Kurdhi, N. A., & Saxena, A. (2023). Evaluating Quadratic Weighted Kappa as the Standard Performance Metric for Automated Essay Scoring. Dalam M. Feng, T. Kãuser, & P. Talukdar (Ed.), *Proceedings of the 16th International Conference on Educational Data Mining* (hlm. 103–113). International Educational Data Mining Society. <https://doi.org/10.5281/zenodo.8115784>
- Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429246593>
- Elmassry, A. M., Zaki, N., Alsheikh, N., & Mediani, M. (2025). A Systematic Review of Pretrained Models in Automated Essay Scoring. *IEEE Access*, 13, 121902–121917. <https://doi.org/10.1109/ACCESS.2025.3584784>
- Faseeh, M., Jaleel, A., Iqbal, N., Ghani, A., Abdusalomov, A., Mehmood, A., & Cho, Y.-I. (2024). Hybrid Approach to Automated Essay Scoring: Integrating Deep Learning Embeddings with Handcrafted Linguistic Features for Improved Accuracy. *Mathematics*, 12(21), 3416. <https://doi.org/10.3390/math12213416>
- Feng, W., & Deng, D. (2021). Align: Aligning All-Pair Near-Duplicate Passages in Long Texts. *Proceedings of the 2021 International Conference on Management of Data*, 541–553. <https://doi.org/10.1145/3448016.3457548>
- Fernando, A., de Silva, N., Velyuthan, M., Rathnayake, C., & Ranathunga, S. (2025). *Improving the quality of Web-mined Parallel Corpora of Low-Resource Languages using Debiasing Heuristics*. <https://arxiv.org/abs/2502.19074>
- Fianti, M. I., Ellianawati, & Subali, B. (2024). Development of Structured Essay Questions to Measure the Student Higher-order Thinking Skills at Senior High School 2 Purworejo. *KnE Social Sciences*. <https://doi.org/10.18502/kss.v9i8.15638>
- He, Y., Jiang, F., Chu, X., & Li, P. (2022). Automated Chinese Essay Scoring from Multiple Traits. Dalam N. Calzolari, C.-R. Huang, H. Kim, J. Pustejovsky, L. Wanner, K.-S. Choi, P.-M. Ryu, H.-H. Chen, L. Donatelli, H. Ji, S. Kurohashi, P. Paggio, N. Xue, S. Kim, Y. Hahm, Z. He, T. K. Lee, E. Santus, F. Bond, & S.-H. Na (Ed.), *Proceedings of the 29th International Conference on Computational Linguistics* (hlm. 3007–3016). International Committee on Computational Linguistics. <https://aclanthology.org/2022.coling-1.266/>
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, 15(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Ifenthaler, D. (2022). Automated Essay Scoring Systems. Dalam *Handbook of Open, Distance and Digital Education* (hlm. 1–15). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0351-9_59-1
- Jazuli, A., Widowati, & Kusumaningrum, R. (2024). Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian

- Student Feedback. *Applied Sciences*, 15(1), 172. <https://doi.org/10.3390/app15010172>
- Jiang, Y. (2023). Team QUST at SemEval-2023 Task 3: A Comprehensive Study of Monolingual and Multilingual Approaches for Detecting Online News Genre, Framing and Persuasion Techniques. *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, 300–306. <https://doi.org/10.18653/v1/2023.semeval-1.40>
- Jong, Y.-J., Kim, Y.-J., & Ri, O.-C. (2022). Improving Performance of Automated Essay Scoring by Using Back-Translation Essays and Adjusted Scores. *Mathematical Problems in Engineering*, 2022, 1–10. <https://doi.org/10.1155/2022/6906587>
- Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 100804. <https://doi.org/10.1016/j.patter.2023.100804>
- Khan, L., Qazi, A., Chang, H.-T., Alhajlah, M., & Mahmood, A. (2025). Empowering Urdu sentiment analysis: an attention-based stacked CNN-Bi-LSTM DNN with multilingual BERT. *Complex & Intelligent Systems*, 11(1), 10. <https://doi.org/10.1007/s40747-024-01631-9>
- Koehn, P. (2004). Statistical Significance Tests for Machine Translation Evaluation. Dalam D. Lin & D. Wu (Ed.), *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (hlm. 388–395). Association for Computational Linguistics. <https://aclanthology.org/W04-3250/>
- Kumar, R., Mathias, S., Saha, S., & Bhattacharyya, P. (2022). Many Hands Make Light Work: Using Essay Traits to Automatically Score Essays. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1485–1495. <https://doi.org/10.18653/v1/2022.naacl-main.106>
- Kundu, A., & Barbosa, D. (2024). *Are Large Language Models Good Essay Graders?* <https://arxiv.org/abs/2409.13120>
- Kusuma, J. S., Halim, K., Pranoto, E. J. P., Kanigoro, B., & Irwansyah, E. (2022). Automated Essay Scoring Using Machine Learning. *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, 1–5. <https://doi.org/10.1109/ICORIS56080.2022.10031338>
- Levy-Feldman, I. (2025). The Role of Assessment in Improving Education and Promoting Educational Equity. *Education Sciences*, 15(2), 224. <https://doi.org/10.3390/educsci15020224>
- Li, F., Xi, X., Cui, Z., Li, D., & Zeng, W. (2023). Automatic Essay Scoring Method Based on Multi-Scale Features. *Applied Sciences*, 13(11), 6775. <https://doi.org/10.3390/app13116775>
- Li, S., & Ng, V. (2024). Automated Essay Scoring: A Reflection on the State of the Art. Dalam Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Ed.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (hlm. 17876–17888). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.991>
- Liew, P. Y., & Tan, I. K. T. (2024). On Automated Essay Grading using Large Language Models. *Proceedings of the 2024 8th International Conference on*

- Computer Science and Artificial Intelligence*, 204–211.
<https://doi.org/10.1145/3709026.3709030>
- Lim, C. T., Bong, C. H., Wong, W. S., & Lee, N. K. (2021). A Comprehensive Review of Automated Essay Scoring (AES) Research and Development. *Pertanika Journal of Science and Technology*, 29(3).
<https://doi.org/10.47836/pjst.29.3.27>
- Liu, X., & Wang, Y. (2025). Color Design of Home Based on Computer Cloud Image Processing Technology. *International Journal of High Speed Electronics and Systems*. <https://doi.org/10.1142/S0129156425402773>
- Loshchilov, I., & Hutter, F. (2019). *Decoupled Weight Decay Regularization*.
<https://arxiv.org/abs/1711.05101>
- Lubis, F. F., Mutaqin, M., Putri, A., Waskita, D., Sulistyaningtyas, T., Arman, A. A., & Rosmansyah, Y. (2021). Automated Short-Answer Grading using Semantic Similarity based on Word Embedding. *International Journal of Technology*, 12(3), 571. <https://doi.org/10.14716/ijtech.v12i3.4651>
- Ludwig, S., Mayer, C., Hansen, C., Eilers, K., & Brandt, S. (2021). Automated Essay Scoring Using Transformer Models. *Psych*, 3(4), 897–915.
<https://doi.org/10.3390/psych3040056>
- Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1), 91–99. <https://doi.org/10.1016/j.gltp.2022.04.020>
- Mayfield, E., & Black, A. W. (2020). Should You Fine-Tune BERT for Automated Essay Scoring? Dalam J. Burstein, E. Kochmar, C. Leacock, N. Madnani, I. Pilán, H. Yannakoudakis, & T. Zesch (Ed.), *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications* (hlm. 151–162). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/2020.bea-1.15>
- Mu, Y., Jin, M., Song, X., & Aletras, N. (2024). Enhancing Data Quality through Simple De-duplication: Navigating Responsible Computational Social Science Research. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 12477–12492.
<https://doi.org/10.18653/v1/2024.emnlp-main.694>
- Nie, Y. (2025). Automated essay scoring with SBERT embeddings and LSTM-Attention networks. *PeerJ Computer Science*, 11, e2634.
<https://doi.org/10.7717/peerj-cs.2634>
- Parmigiani, D., Nicchia, E., Murgia, E., & Ingersoll, M. (2024). Formative assessment in higher education: an exploratory study within programs for professionals in education. *Frontiers in Education*, 9.
<https://doi.org/10.3389/educ.2024.1366215>
- Rainio, O., Teuvo, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14(1), 6086.
<https://doi.org/10.1038/s41598-024-56706-x>
- Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: a systematic literature review. *Artificial Intelligence Review*, 55(3), 2495–2527.
<https://doi.org/10.1007/s10462-021-10068-2>
- Ramnarain-Seetohul, V., Bassoo, V., & Rosunally, Y. (2022). Similarity measures in automated essay scoring systems: A ten-year review. *Education and*

- Information Technologies*, 27(4), 5573–5604. <https://doi.org/10.1007/s10639-021-10838-z>
- Sarica, S., & Luo, J. (2021). Stopwords in technical language processing. *PLOS ONE*, 16(8), e0254937. <https://doi.org/10.1371/journal.pone.0254937>
- Sechidis, K., Tsoumakas, G., & Vlahavas, I. (2011). *On the Stratification of Multi-label Data* (hlm. 145–158). https://doi.org/10.1007/978-3-642-23808-6_10
- Sethi, A., & Singh, K. (2022). Natural Language Processing based Automated Essay Scoring with Parameter-Efficient Transformer Approach. *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, 749–756. <https://doi.org/10.1109/ICCMC53470.2022.9753760>
- Shaw, C., LaCasse, P., & Champagne, L. (2025). Exploring emotion classification of Indonesian tweets using large scale transfer learning via IndoBERT. *Social Network Analysis and Mining*, 15(1), 22. <https://doi.org/10.1007/s13278-025-01439-6>
- Shibata, T., & Uto, M. (2022). Analytic Automated Essay Scoring Based on Deep Neural Networks Integrating Multidimensional Item Response Theory. Dalam N. Calzolari, C.-R. Huang, H. Kim, J. Pustejovsky, L. Wanner, K.-S. Choi, P.-M. Ryu, H.-H. Chen, L. Donatelli, H. Ji, S. Kurohashi, P. Paggio, N. Xue, S. Kim, Y. Hahm, Z. He, T. K. Lee, E. Santus, F. Bond, & S.-H. Na (Ed.), *Proceedings of the 29th International Conference on Computational Linguistics* (hlm. 2917–2926). International Committee on Computational Linguistics. <https://aclanthology.org/2022.coling-1.257/>
- Siino, M., Tinnirello, I., & La Cascia, M. (2024). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. *Information Systems*, 121, 102342. <https://doi.org/10.1016/j.is.2023.102342>
- Supryadi, Pan, L., & Xiong, D. (2024). An Empirical Study on the Robustness of Massively Multilingual Neural Machine Translation. Dalam N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Ed.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (hlm. 1086–1097). ELRA and ICCL. <https://aclanthology.org/2024.lrec-main.97/>
- Tang, X., Chen, H., Lin, D., & Li, K. (2024). Harnessing LLMs for multi-dimensional writing assessment: Reliability and alignment with human judgments. *Heliyon*, 10(14), e34262. <https://doi.org/10.1016/j.heliyon.2024.e34262>
- Tohka, J., & van Gils, M. (2021). Evaluation of machine learning algorithms for health and wellness applications: A tutorial. *Computers in Biology and Medicine*, 132, 104324. <https://doi.org/10.1016/j.compbiomed.2021.104324>
- Uto, M., Xie, Y., & Ueno, M. (2020). Neural Automated Essay Scoring Incorporating Handcrafted Features. Dalam D. Scott, N. Bel, & C. Zong (Ed.), *Proceedings of the 28th International Conference on Computational Linguistics* (hlm. 6077–6088). International Committee on Computational Linguistics. <https://doi.org/10.18653/v1/2020.coling-main.535>
- Uyar, A. C., & Büyükahıska, D. (2025). Artificial intelligence as an automated essay scoring tool: A focus on ChatGPT. *International Journal of Assessment Tools in Education*, 12(1), 20–32. <https://doi.org/10.21449/ijate.1517994>

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need*. <https://arxiv.org/abs/1706.03762>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Werner de Vargas, V., Schneider Aranda, J. A., dos Santos Costa, R., da Silva Pereira, P. R., & Victória Barbosa, J. L. (2023). Imbalanced data preprocessing techniques for machine learning: a systematic mapping study. *Knowledge and Information Systems*, *65*(1), 31–57. <https://doi.org/10.1007/s10115-022-01772-8>
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 843–857. <https://doi.org/10.18653/v1/2020.aacl-main.85>
- Yakubu, M. A., Sain, Z. H., Lawal, U. S., & Budiman, S. A. (2024). Application of Artificial Intelligence [AI] for Automated Essay Grading in Nigerian Schools. *Indonesian Journal on Education (IJoEd)*, *1*(2), 93–103. <https://doi.org/https://doi.org/10.70437/vnve7j92>
- Yefferson, D. Y., Lawijaya, V., & Girsang, A. S. (2024). Hybrid model: IndoBERT and long short-term memory for detecting Indonesian hoax news. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *13*(2), 1913. <https://doi.org/10.11591/ijai.v13.i2.pp1913-1924>
- Yu, C., Sun, W., Xiong, Q., Gao, J., & Qu, G. (2022). Double-Balanced Loss for Imbalanced Colorectal Lesion Classification. *Computational and Mathematical Methods in Medicine*, *2022*, 1–14. <https://doi.org/10.1155/2022/1691075>
- Yulita, I. N., Wijaya, V., Rosadi, R., Sarathan, I., Djuyandi, Y., & Prabuwno, A. S. (2023). Analysis of Government Policy Sentiment Regarding Vacation during the COVID-19 Pandemic Using the Bidirectional Encoder Representation from Transformers (BERT). *Data*, *8*(3), 46. <https://doi.org/10.3390/data8030046>
- Zaffran, M., Dieuleveut, A., Josse, J., & Romano, Y. (2023). Conformal Prediction with Missing Values. Dalam A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, & J. Scarlett (Ed.), *Proceedings of the 40th International Conference on Machine Learning* (Vol. 202, hlm. 40578–40604). PMLR. <https://proceedings.mlr.press/v202/zaffran23a.html>