

DAFTAR PUSTAKA

- Banerjee, S., & Lavie, A. (2005). *METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments*.
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization Yoshua Bengio. In *Journal of Machine Learning Research* (Vol. 13). <http://scikit-learn.sourceforge.net>.
- Brookhart, S. M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy and Practice*, 20(1), 69–90. <https://doi.org/10.1080/0969594X.2012.703170>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language Models are Few-Shot Learners*. <http://arxiv.org/abs/2005.14165>
- Burrows, S., Gurevych, I., & Stein, B. (2015). The eras and trends of automatic short answer grading. In *International Journal of Artificial Intelligence in Education* (Vol. 25, Number 1, pp. 60–117). Springer New York LLC. <https://doi.org/10.1007/s40593-014-0026-8>
- Chaudhari, R., & Patel, M. (2024). Deep Learning in Automated Short Answer Grading: A Comprehensive Review. *ITM Web of Conferences*, 65, 03003. <https://doi.org/10.1051/itmconf/20246503003>
- Condor, A., Litster, M., & Pardos, Z. (2021). Automatic short answer grading with SBERT on out-of-sample questions. In *Proceedings of the 14th International Conference on Educational Data Mining (EDM 2021)* (pp. 430-435). International Educational Data Mining Society.
- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <https://github.com/tensorflow/tensor2tensor>

- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- G. Khalf, A., Nabil, E., H. Gomaa, W., Benrhouma, O., & M. El-Mandouh, A. (2026). Leveraging Transformers and LLMs for Automated Grading and Feedback Generation Using a Novel Dataset. *Data*, 11(3). <https://doi.org/10.3390/data11030057>
- Galhardi, L. B., & Brancher, J. D. (2018). Machine learning approach for automatic short answer grading: A systematic review. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11238 LNAI, 380–391. https://doi.org/10.1007/978-3-030-03928-8_31
- Gamit, N. C., & Panchal, S. (2024). Universal Automatic Short Answer Grading (ASAG) Model: A Comprehensive Approach. In *Journal of Information Systems Engineering and Management* (Vol. 2024, Number 4s). <https://jisem-journal.com/>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- Kakarla, S., Borchers, C., Thomas, D., Bhushan, S., & Koedinger, K. R. (2025). *Comparing Few-Shot Prompting of GPT-4 LLMs with BERT Classifiers for Open-Response Assessment in Tutor Equity Training*. <http://arxiv.org/abs/2501.06658>
- Kannan, N., Sivasubramanian, S., Kaliappan, M., Vimal, S., & Suresh, A. (2019). Predictive big data analytic on demonetization data using support vector machine. *Cluster Computing*, 22, 14709–14720. <https://doi.org/10.1007/s10586-018-2384-8>
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
- Kusumaningrum, R., Wiguna, I. M. K. A., Damanhuri, R., Sutikno, S., Khadijah, K., Sugiharto, A., & Pramayoga, A. S. (2025). *Integrating Human Feedback and*

Unsupervised Fine-Tuning in Open-Domain Short-Answer Grading: A Scalable HITL Architecture. 1–9. <https://doi.org/10.1109/rasse64831.2025.11315287>

Mohler, M., Bunescu, R., & Mihalcea, R. (2011). *Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments.* <http://nlp.stanford.edu/software/dependencies>

Ndukwe, I. G., Amadi, C. E., Nkomo, L. M., & Daniel, B. K. (2020). Automatic Grading System Using Sentence-BERT Network. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12164 LNAI, 224–227. https://doi.org/10.1007/978-3-030-52240-7_41

OpenAI. (2026). GPT-5.4 dan GPT-5.4-nano. OpenAI API Documentation. <https://developers.openai.com/api/docs/models>

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. In *IEEE Transactions on Knowledge and Data Engineering* (Vol. 22, Number 10, pp. 1345–1359). <https://doi.org/10.1109/TKDE.2009.191>

Pilicita, A., & Barra, E. (2025). LLMs in Education: Evaluation GPT and BERT Models in Student Comment Classification. *Multimodal Technologies and Interaction*, 9(5). <https://doi.org/10.3390/mti9050044>

Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivè, V., Beygelzimer, A., D'alché-Buc, F., Paris, T., Larochelle, H., Florence D'alché-Buc, E., & Fox, H. L. (2021). Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program). In *Journal of Machine Learning Research* (Vol. 22).

Pinto, G., Cardoso-Pereira, I., Ribeiro, D. M., Lucena, D., de Souza, A., & Gama, K. (2023). *Large Language Models for Education: Grading Open-Ended Questions Using ChatGPT.* <http://arxiv.org/abs/2307.16696>

Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained Models for Natural Language Processing: A Survey. *Science China technological sciences*, 63(10), 1872-1897.

- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners*. <https://github.com/codelucas/newspaper>
- Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. <https://github.com/UKPLab/>
- Roy, S., Narahari, Y., & Deshmukh, O. D. (2015). *A Perspective on Computer Assisted Assessment Techniques for Short Free-text Answers*. https://www.experience.com/alumnus/article?channel_id=education&source_
- Schneider, J., Schenk, B., & Niklaus, C. (2023). *Towards LLM-based Autograding for Short Textual Answers*. arXiv preprint arXiv:2309.11508.
- Sennrich, R., Haddow, B., & Birch, A. (2016). *Improving Neural Machine Translation Models with Monolingual Data*. <http://arxiv.org/abs/1511.06709>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0197-0>
- Tay, Y., Dehghani, M., Bahri, D., & Metzler, D. (2022). *Efficient Transformers: A Survey*. <http://arxiv.org/abs/2009.06732>
- Torrey, L., & Shavlik, J. (2010). Transfer Learning. In E. Soria, J. Martin, R. Magdalena, M. Martinez, & A. Serrano (Eds.), *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques* (pp. 242-264). IGI Global.
- Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing and Management*, 50(1), 104–112. <https://doi.org/10.1016/j.ipm.2013.08.006>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. <http://arxiv.org/abs/1706.03762>
- Wang, K., Reimers, N., & Gurevych, I. (2021). *TSDAE: Using Transformer-based Sequential Denoising Auto-Encoder for Unsupervised Sentence Embedding Learning*. <https://github.com/>

- Waspada, I., Bahtiar, N., & Wibowo, A. (2019). Clustering student behavior based on quiz activities on moodle LMS to discover the relation with a final exam score. *Journal of Physics: Conference Series*, 1217(1). <https://doi.org/10.1088/1742-6596/1217/1/012118>
- Wei, J., & Zou, K. (2019). *EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks*. <http://arxiv.org/abs/1901.11196>
- Wijanto, M. C., & Yong, H. S. (2024). Combining Balancing Dataset and SentenceTransformers to Improve Short Answer Grading Performance. *Applied Sciences (Switzerland)*, 14(11). <https://doi.org/10.3390/app14114532>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. Le, Gugger, S., ... Rush, A. M. (2020). *HuggingFace's Transformers: State-of-the-art Natural Language Processing*. <http://arxiv.org/abs/1910.03771>
- Xia, M., Jiang, Z., Wang, H., Zhao, J., Hu, T., & Chen, G. (2025). *Ensembling Prompting Strategies for Zero-Shot Hierarchical Text Classification with Large Language Models*.