













- [10] CHO, K., VAN MERRIËNBOER, B., BAHDANAU, D., AND BENGIO, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259* (2014).
- [11] CONNEAU, A., KHANDLWAL, K., GOYAL, N., CHAUDHARY, V., WENZKE, G., GUZMÁN, F., GRAVE, E., OTT, M., ZETTLEMOYER, L., AND STOYANOV, V. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116* (2019).
- [12] CONNEAU, A., AND LAMPLE, G. Cross-lingual language model pretraining. *Advances in neural information processing systems 32* (2019).
- [13] DEVLIN, J., CHANG, M.-W., LEE, K., AND TOUTANOVA, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [14] GAANOUN, K., NAIRA, A. M., ALLAK, A., AND BENELALLAM, I. Darijabert: a step forward in nlp for the written moroccan dialect. *Research-Square* (2023). <https://doi.org/10.21203/rs.3.rs-2560653/v1>.
- [15] GANI, M. O., AYYASAMY, R. K., SANGODIAH, A., AND FUI, Y. T. Ustw vs. stw: A comparative analysis for exam question classification based on bloom’s taxonomy. *MENDEL 28*, 2 (2022), 25–40.
- [16] GAROUANI, M., AND KHARROUBI, J. Mac: an open and free moroccan arabic corpus for sentiment analysis. In *The Proceedings of the International Conference on Smart City Applications* (2021), Springer, pp. 849–858.
- [17] HOCHREITER, S., AND SCHMIDHUBER, J. Long short-term memory. *Neural computation 9*, 8 (1997), 1735–1780.
- [18] INOUE, G., ALHAFNI, B., BAIMUKAN, N., BOUAMOR, H., AND HABASH, N. The interplay of variant, size, and task type in arabic pre-trained language models. *arXiv preprint arXiv:2103.06678* (2021).
- [19] KINGMA, D. P., AND BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [20] LAN, Z., CHEN, M., GOODMAN, S., GIMPEL, K., SHARMA, P., AND SORICUT, R. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942* (2019).
- [21] LIU, Y., OTT, M., GOYAL, N., DU, J., JOSHI, M., CHEN, D., LEVY, O., LEWIS, M., ZETTLEMOYER, L., AND STOYANOV, V. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).
- [22] RUDER, S. Why you should do nlp beyond english. <https://ruder.io/nlp-beyond-english> [Retrieved April 18, 2023].
- [23] SAMIH, Y., AND MAIER, W. An arabic-moroccan darija code-switched corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)* (2016), pp. 4170–4175.
- [24] SANH, V., DEBUT, L., CHAUMOND, J., AND WOLF, T. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108* (2019).
- [25] SENNRICH, R., HADDOW, B., AND BIRCH, A. Das hunderttage-stadion: Entstehungsgeschichte des bad nauheimer kunstestadions unter colonel paul r. knight. *Acl* (2016), 1715–1725.
- [26] TACHICART, R., BOUZOUBAA, K., AND JAAFAR, H. Lexical differences and similarities between moroccan dialect and arabic. In *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)* (2016), IEEE, pp. 331–337.
- [27] VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, L., AND POLOSUKHIN, I. Attention is all you need. *Advances in neural information processing systems 30* (2017).
- [28] WIJAYA, M. C. The classification of documents in malay and indonesian using the naive bayesian method uses words and phrases as a training set. *MENDEL 26*, 2 (2020), 23–28.
- [29] WOLF, T., DEBUT, L., SANH, V., CHAUMOND, J., DELANGUE, C., MOI, A., CISTAC, P., RAULT, T., LOUF, R., FUNTOWICZ, M., ET AL. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations* (2020), pp. 38–45.
- [30] WU, Y., SCHUSTER, M., CHEN, Z., LE, Q. V., NOROUZI, M., MACHEREY, W., KRIKUN, M., CAO, Y., GAO, Q., MACHEREY, K., ET AL. Google’s neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144* (2016).
- [31] YANG, Z., DAI, Z., YANG, Y., CARBONELL, J., SALAKHUTDINOV, R. R., AND LE, Q. V. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems 32* (2019).
- [32] ZAHIR, J. Iadd: An integrated arabic dialect identification dataset. *Data in Brief 40* (2022), 107777.