

References

- [1] S. Youn and D. McLeod, "A Comparative Study for Email Classification," in *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, 2007, pp. 387–391. doi: 10.1007/978-1-4020-6264-3_67.
- [2] P. Li, J. Li, and Q. Zhu, "An Approach to Hierarchical Email Categorization Based on ME," in *Natural Language Processing and Information Systems*, Berlin, Heidelberg, 2007, pp. 25–34. doi: 10.1007/978-3-540-73351-5_3.
- [3] S. Maldonado and G. L'Huillier, "SVM-Based Feature Selection and Classification for Email Filtering," in *Pattern Recognition - Applications and Methods*, Berlin, Heidelberg, 2013, pp. 135–148. doi: 10.1007/978-3-642-36530-0_11.
- [4] M. Ikonomakis, S. Kotsiantis, and V. Tampakas, "Text Classification Using Machine Learning Techniques," p. 10.
- [5] D. K. Gupta and S. Goyal, "Email Classification into Relevant Category Using Neural Networks," *ArXiv180203971 Cs*, Feb. 2018, Accessed: Feb. 17, 2020. [Online]. Available: <http://arxiv.org/abs/1802.03971>
- [6] A. Sallab and M. Rashwan, "E-mail classification using deep networks," *J. Theor. Appl. Inf. Technol. JATIT*, vol. 37, Mar. 2012.
- [7] P. K. Pushp and M. M. Srivastava, "Train Once, Test Anywhere: Zero-Shot Learning for Text Classification," *ArXiv171205972 Cs*, Dec. 2017, Accessed: Sep. 03, 2020. [Online]. Available: <http://arxiv.org/abs/1712.05972>
- [8] B. Agarwal and N. Mittal, "Text Classification Using Machine Learning Methods-A Survey," in *Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012*, New Delhi, 2014, pp. 701–709. doi: 10.1007/978-81-322-1602-5_75.
- [9] C. E. Perez, "The Paradigm Shift of Self-Supervised Learning," *Medium*, Jun. 13, 2019. <https://medium.com/intuitionmachine/the-paradigm-shift-of-self-supervised-learning-744a6819ce08> (accessed Feb. 17, 2020).
- [10] T. Silva, "Self-Supervised Learning and the Quest for Reducing Labeled Data in Deep Learning," *Medium*, Jan. 14, 2020. <https://towardsdatascience.com/self-supervised-learning-and-the-quest-for-reducing-labeled-data-in-deep-learning-db59a563e25b> (accessed Feb. 17, 2020).
- [11] M. F. A. Hady and F. Schwenker, "Semi-supervised Learning," in *Handbook on Neural Information Processing*, M. Bianchini, M. Maggini, and L. C. Jain, Eds. Berlin, Heidelberg: Springer, 2013, pp. 215–239. doi: 10.1007/978-3-642-36657-4_7.
- [12] P. West, A. Holtzman, J. Buys, and Y. Choi, "BottleSum: Unsupervised and Self-supervised Sentence Summarization using the Information Bottleneck Principle," *ArXiv190907405 Cs*, Sep. 2019, Accessed: Feb. 19, 2020. [Online]. Available: <http://arxiv.org/abs/1909.07405>
- [13] H. Wang *et al.*, "Self-Supervised Learning for Contextualized Extractive Summarization," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, Jul. 2019, pp. 2221–2227. doi: 10.18653/v1/P19-1214.
- [14] M. K. Dalal and M. A. Zaveri, "Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews," *Applied Computational Intelligence and Soft Computing*, 2013. <https://www.hindawi.com/journals/acisc/2013/910706/> (accessed Feb. 19, 2020).
- [15] E. T. Matsubara, M. C. Monard, and G. E. A. P. A. Batista, "Multi-view Semi-supervised Learning: An Approach to Obtain Different Views from Text Datasets," in

- Proceedings of the 2005 conference on Advances in Logic Based Intelligent Systems: Selected Papers of LAPTEC 2005*, NLD, May 2005, pp. 97–104.
- [16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *ArXiv181004805 Cs*, May 2019, Accessed: Aug. 18, 2020. [Online]. Available: <http://arxiv.org/abs/1810.04805>
 - [17] M. E. Peters *et al.*, “Deep contextualized word representations,” *ArXiv180205365 Cs*, Mar. 2018, Accessed: Aug. 18, 2020. [Online]. Available: <http://arxiv.org/abs/1802.05365>
 - [18] J. E. van Engelen and H. H. Hoos, “A survey on semi-supervised learning,” *Mach. Learn.*, Nov. 2019, doi: 10.1007/s10994-019-05855-6.
 - [19] A. Blum and T. Mitchell, “Combining Labeled and Unlabeled Data with Co-Training,” *Proc. Annu. ACM Conf. Comput. Learn. Theory*, Oct. 2000, doi: 10.1145/279943.279962.
 - [20] Y. Liu *et al.*, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” *ArXiv190711692 Cs*, Jul. 2019, Accessed: May 07, 2021. [Online]. Available: <http://arxiv.org/abs/1907.11692>
 - [21] L. Xue *et al.*, “mT5: A massively multilingual pre-trained text-to-text transformer,” *ArXiv201011934 Cs*, Mar. 2021, Accessed: May 07, 2021. [Online]. Available: <http://arxiv.org/abs/2010.11934>
 - [22] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, “Improving Language Understanding by Generative Pre-Training,” p. 12.
 - [23] Y. Peng, S. Yan, and Z. Lu, “Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets,” *ArXiv190605474 Cs*, Jun. 2019, Accessed: May 07, 2021. [Online]. Available: <http://arxiv.org/abs/1906.05474>
 - [24] P. Liu and T.-S. Moh, “Content Based Spam E-mail Filtering,” in *2016 International Conference on Collaboration Technologies and Systems (CTS)*, Oct. 2016, pp. 218–224. doi: 10.1109/CTS.2016.0052.
 - [25] “AG News Classification Dataset.” <https://kaggle.com/amananandrai/ag-news-classification-dataset> (accessed May 10, 2021).