[2] R. J. Suhatril, B. Mutiara, A. Suhendra, and I. M. Wiryana, “Strategi Klasifikasi Melalui Pembobotansimilaritas Berbasis Semantic Network,” Universitas Gunadarma, 2015.

[3] Sang-Bum Kim, Kyoung-Soo Han, Hae-Chang Rim, and Sung Hyon Myaeng, “Some Effective Techniques for Naive Bayes Text Classification,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 11, pp. 1457–1466, Nov. 2006, doi:10.1109/tkde.2006.180.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *ArXiv*, vol. abs/1810.0, 2019.

[5] A. M. and S. Salama, “Discovering Performance Evaluation Features of faculty Members using Data Mining Techniques to Support Decision Making,” International Journal of Computer Applications, vol. 178, no. 49, pp. 25–29, Sep. 2019, doi: 10.5120/ijca2019919417.

[6] A. M. Dai and Q. V Le, “Semi-supervised Sequence Learning,” In Advances in Neural Information Processing Systems (pp. 3079–3087). Montreal, Canada: Curran Associates, Inc, 2015.

[7] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, “Improving language understanding by generative pre-training,” Tech. Rep., 2018.

[8] J. Howard and S. Ruder, “Universal language model fine-tuning for text classification,” 2018, arXiv:1801.06146. [Online]. Available: http://arxiv.org/abs/1801.06146

[9] M. E. Peters *et al.*, “Deep Contextualized Word Representations,”. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT’18): Vol 1, 2018, Jun 1-6, New Orleans, LA, USA. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018: 2227-2237.

[10] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, “A large annotated corpus for learning natural language inference,” 2015, arXiv:1508.05326

[11] A. Williams, N. Nangia, and S. R. Bowman, “A broad-coverage challenge corpus for sentence understanding through inference,” 2017, arXiv:1704.05426. [Online]. Available: http://arxiv.org/abs/1704.05426

[12] Dolan, W. B., & Brockett, C. (2005). Automatically constructing a corpus of sentential paraphrases.InProc. of the 3rd Int. Workshop on Paraphrasing, pp. 9–16, Jeju island, Korea, 2005

[13] Tjong Kim Sang, E. and Fien De Meulder. “Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition.” Conference on Computational Natural Language Learning, 2003.

[14] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, “SQuAD: 100,000+ Questions for Machine Comprehension of Text,” Association for Computational Linguistics (ACL), 2016.

[15] Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., et al., “Attention is All you Need,” Proceedings.neurips.cc, [Online]. Available: https://proceedings.neurips. cc/paper/7181-attention-is-all-you-need.2023

[16] P. F. Brown, V. J. Della Pietra, P. V de Souza, J. C. Lai, and R. L. Mercer, “Class-Based n-gram Models of Natural Language,” *Comput. Linguist.*, vol. 18, pp. 467–479, 1992.

[17] R. K. Ando and T. Zhang, “A Framework for Learning Predictive Structures from Multiple Tasks and Unlabeled Data,” *J. Mach. Learn. Res.*, vol. 6, pp. 1817–1853, 2005.

[18] J. Blitzer, R. T. Mcdonald and F. Pereira, "Domain adaptation with structural correspondence learning", Proc. Conf. Empirical Methods Natural Lang. Process., pp. 22-23, Jul. 2007.

[19] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean: Distributed Representations of Words and Phrases and their Compositionality, Advances in Neural Information Processing Systems 26, pp.3111--3119, 2013.

[20] J. Pennington, R. Socher, and C. Manning, ‘‘GloVe: Global vectors for word representation,’’ in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP). Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. [Online]. Available: https://aclanthology.org/D14-1162

[21] J. Turian, L. Ratinov and Y. Bengio, Word representations: A simple and general method for semi-supervised learning, 2010.

[22] A. Mnih and G.E. Hinton, "A Scalable Hierarchical Distributed Language Model", Proceedings of Neural Information Processing Systems 21 (NIPS 2008), pp. 1-8, 2008.

[23] Kiros R, Zhu Y K, Salakhutdinov R, Zemel R, Torralba A, Urtasun R, Fidler S. Skip-thought vectors. arXiv: 1506.06726, 2015. https://arxiv.org/abs/1506.06726, June 2017.

[24] L. Logeswaran and H. Lee, “An efficient framework for learning sentence representations,” *ArXiv*, vol. abs/1803.0, 2018.

[25] Q. V. Le and T. Mikolov, "Distributed Representations of Sentences and Documents", 1st Workshop on Representation Learning for NLP, 2015.

[26] Y. Jernite, S. R. Bowman, and D. A. Sontag, “Discourse-Based Objectives for Fast Unsupervised Sentence Representation Learning,” *ArXiv*, vol. abs/1705.0, 2017.

[27] F. Hill, K. Cho, and A. Korhonen, “Learning Distributed Representations of Sentences from Unlabelled Data,” arXiv preprint arXiv:1602.03483, 2016.

[28] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multi task learning", Proc. of ICML, 2008.

[29] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, “Glue: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding,” *ArXiv*, vol. abs/1804.0, 2018.

[30] Z. Qu, X. Song, S. Zheng, X. Wang, X. Song, and Z. Li, “Improved Bayes Method Based on TF-IDF Feature and Grade Factor Feature for Chinese Information Classification,” 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Jan. 2018, doi: 10.1109/bigcomp.2018.00124.

[31] R. Socher, J. Wu, A. Perelygin, C.D. Manning, J. Chuang et al., "Recursive deep models for semantic compositionality over a sentiment treebank", EMNLP, 2013. Association for Computational Linguistics, Seattle, 2013.

[32] B. Walek and V. Fojtik, “A hybrid recommender system for recommending relevant movies using an expert system,” Expert Systems with Applications, vol. 158, p. 113452, Nov. 2020, doi:10.1016/j.eswa.2020.113452.

[33] H. Saif, M. Fernández, Y. He, and H. Alani, “On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter,” The Open University: Milton Keynes, UK, 2014.