References

[1] The European Commission, *In-depth analysis in support of the commission communication COM(2018) 773*. The European Commission, 2018.

[2] S. AB, "The pathways study: Achieving fossil-free commercial transport by 2050," 151 87 Södertälje, Tech. Rep., May 2018.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN 0387310738

[4] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *ArXiv*, vol. abs/1502.03167, 2015.

[5] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 12 2014.

[6] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774.

[7] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should I trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, 2016, pp. 1135–1144.

[8] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[9] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *ICML*, 1997.

[10] L. Zhang, X. Wang, and L. Qu, "Feature reduction based on analysis of covariance matrix," in *2008 International Symposium on Computer Science and Computational Technology*, vol. 1, 2008. doi: 10.1109/ISCSCT.2008.17 pp. 59–62.

[11] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, 1987. doi: https://doi.org/10.1016/0169-7439(87)80084-9 Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.

[12] R. Liu, N. Yang, X. Ding, and L. Ma, "An unsupervised feature selection algorithm: Laplacian score combined with distance-based entropy measure," in *2009 Third International Symposium on Intelligent Information Technology Application*, vol. 3, 2009. doi: 10.1109/IITA.2009.390 pp. 65–68.

[13] J. Brank, M. Grobelnik, N. Milic-Frayling, and D. Mladenic, "Feature selection using linear support vector machines," Tech. Rep. MSR-TR-2002-63, June 2002.

[14] N. Longford, "A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects," *Biometrika*, vol. 74, pp. 817–827, 1987.

[15] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," in *UAI*, 2011.

[16] W. E. Marcilio and D. M. Eler, "From explanations to feature selection: assessing shap values as feature selection mechanism," *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 340–347, 2020.

[17] D. Fryer, I. Strümke, and H. Nguyen, "Shapley values for feature selection: The good, the bad, and the axioms," *ArXiv*, vol. abs/2102.10936, 2021.

[18] P. Liu, B. Wu, H. Ma, and M. Seok, "Memnas: Memory-efficient neural architecture search with grow-trim learning," in *2020 IEEE/CVF*

Conference on Computer Vision and Pattern Recognition (CVPR), 2020. doi: 10.1109/CVPR42600.2020.00218 pp. 2105–2113.

[19] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[20] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*. Springer, 2009, pp. 1–4.

[21] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, and C. Citro, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015.

Appendices