# Homework: Some Scientific Articles

Here are some articles that I find interesting, useful, well-explained, or a combination of these.

1. Hyperopt-Sklearn: Automatic Hyperparameter Configuration for Scikit-Learn; Komer, B et al., 2014. <https://proceedings.scipy.org/articles/Majora-14bd3278-006> (introduces the sklearn API for hyperopt; ML-related; includes GitHub repo)
2. XGBoost: A Scalable Tree Boosting System; Chen, T. and Guestrin, C.; 2016. <https://arxiv.org/abs/1603.02754> (the XGBoost paper)
3. Efficient and Robust Automated Machine Learning; Feurer, M. and Klein, A, 2015. <https://www.semanticscholar.org/paper/Efficient-and-Robust-Automated-Machine-Learning-Feurer-Klein/775a4e375cc79b53b94e37fa3eedff481823e4a6> (probably the most popular modern method to do auto ML)
4. Generative Nowcasting of Marine Fog Visibility in the Grand Banks area and Sable Island in Canada; Gultepe, E. et al., 2024; <https://arxiv.org/abs/2402.06800> (has some deep learning, but is well-written as an intro to advanced image processing techniques)
5. Low Level Feature Extraction for Cilia Segmentation; Zain, M et al., 2023. <https://proceedings.scipy.org/articles/majora-212e5952-026> (biology and image processing; includes GitHub repo; has some deep learning)
6. Organ aging signatures in the plasma proteome track health and disease; Oh, H. et al., 2023. <https://www.nature.com/articles/s41586-023-06802-1> (studies aging but focuses heavily on data analysis and ML, does not require a deep understanding of the domain)
7. A new hybrid deep learning model for monthly oil prices forecasting; Guon, K. and Gong, X.; 2023. <https://www.sciencedirect.com/science/article/abs/pii/S0140988323006345> (has a bit of deep learning; not hard to reproduce but doesn’t provide source code)
8. Climate change attention and carbon futures return prediction; Gong, X. and Li, M.; 2023. <https://onlinelibrary.wiley.com/doi/10.1002/fut.22443> (by the authors of the above article)
9. Fare revenue forecast in public transport: A comparative case study; Krembsler, J. et al., 2024. <https://www.sciencedirect.com/science/article/pii/S0739885924000404?via%3Dihub> (revenue analysis seems to be very common for ML, this paper uses very simple models and not so simple datasets)
10. Extreme Learning Tree; Akusok, A. et al.; 2019; <https://arxiv.org/abs/1912.09087> (interesting approach towards a replacement for isolation forests / ensembles for anomaly detection)
11. The impact of imbalanced training data on machine learning for author name disambiguation; Kim, J. and Kim, J.; 2018. <https://arxiv.org/abs/1808.00525> (quite interesting text processing and visualization)
12. m-arcsinh: An Efficient and Reliable Function for SVM and MLP in scikit-learn; Parisi, L.; 2020. <https://arxiv.org/abs/2009.07530> (an interesting kernel function, and a good overview of the theory and maths)
13. Pre-symptomatic detection of COVID-19 from smartwatch data; Mishra, T. et al., 2020. <https://www.nature.com/articles/s41551-020-00640-6> (the title and journal are self-explanatory; one of the best articles on and about COVID-19)
14. Document-level sentiment classification using hybrid machine learning approach; Tripathy, A. et al., 2017. <https://link.springer.com/article/10.1007/s10115-017-1055-z> (focuses on feature selection on many types of text data; uses standard datasets)
15. Classification vs regression in overparameterized regimes: Does the loss function matter? Muthukumar, V. et al., 2020; <https://arxiv.org/abs/2005.08054> (generalizes very well to neural networks)
16. Optimised one-class classification performance; Lenz, O. et al., 2022; <https://arxiv.org/abs/2102.02618> (a useful approach to semi-supervised learning and outlier detection)
17. Spatial Decompositions for Large Scale SVMs; Thomann, P. et al., 2018; <https://arxiv.org/abs/1612.00374> (presents an approach to make SVMs work with many examples – the authors claim 10 million or more)
18. Causal structure learning from time series: Large regression coefficients may predict causal links better in practice than small p-values; Weichwald, S. et al., 2020; <https://arxiv.org/abs/2002.09573> (improving interpretability in time-series models)