TIME SERIES WEATHER DATA

Project for IT1244 Artificial Intelligence: Technology and Impact By Team 30 - Lim Shi Ying, Jennifer Liu, Qian Yunhan, Tay Yi Hong

---

1 hour

Our chosen model for predicting relative humidity at the 1-hour horizon is XGBoost. You may run everything from `Decision Tree_XgBoost.ipynb` to view the prediction for 1 hour horizon. The result of its fit on the test data (last 20%) can be found at the end.

6 hours

Our chosen model for predicting relative humidity at the 6-hour horizon is multi-step LSTM. Run the following sections in the LSTM file:
1. Everything under `LSTM modelling functions` until `Trial`
2. Everything under `Multi-step feature prediction`, until `24 hours`, skipping over `feats_model = ...` (the models will be loaded in at the end)
3. Everything under `Run models on test data > 6 hours - Multi-step LSTM`

24 hours

Our chosen model for predicting relative humidity at the 24-hour horizon is XGBoost. You may run everything from `Decision Tree_XgBoost.ipynb` to view the prediction for 1 hour horizon. The result of its fit on the test data (last 20%) can be found at the end.

Acknowledgements

Wiin-Nielsen, A. (1991). The birth of numerical weather prediction. Tellus A: Dynamic Meteorology and Oceanography, 43(4), 36–52. https://doi.org/10.3402/tellusa.v43i4.1193/

Yang, R., Hu, J., Li, Z., et al. (2024). Interpretable machine learning for weather and climate prediction: A review. Atmospheric Environment, 338, 120797. https://doi.org/10.1016/j.atmosenv.2024.120797/

Tyralis, H., & Papacharalampous, G. (2017). Variable Selection in Time Series Forecasting Using Random Forests. Algorithms, 10(4), 114; https://doi.org/10.3390/a10040114/

Zhang, W. Y., Xie, J. F., Wan, G. C., & Tong, M. S. (2021). Single-step and multi-step time series prediction for urban temperature based on LSTM model of TensorFlow. 2021 Photonics & Electromagnetics Research Symposium (PIERS), 1531–1535. https://doi.org/10.1109/PIERS53385.2021.9694882/

How our approach addressed their limitations:

In Study A (Tyralis & Papacharalampous), random forests used raw lags and lacked interpretability-focused pruning. We addressed this by adding time-aware engineered features and applying feature pruning to improve clarity and generalisation.
In Study B (Zhang et al.), the LSTM lacked systematic hyperparameter tuning and baseline ML comparisons. We improved on this by tuning across multiple horizons and benchmarking against tree-based models for interpretability and performance.

Study C (Yang et al.) emphasized the need for interpretable weather ML models. Our project uses tree-based models like RF and XGB to balance performance with explainability through feature importance and pruning.