Comparison of Average Temperature Prediction Based on XGBoost and Prophet Models

Xiang Zhou*

Department of Physics, King's College London, London, United Kingdom

*Corresponding author: k21068750@kcl.ac.uk

Abstract. Temperature prediction is extremely important in various fields. With this in mind, this study utilized two prominent models, XGBoost and Prophet, known for their efficacy in time series forecasting, to perform daily average temperature predictions based on historical data. XGBoost model enhances the Gradient Boosted Decision Tree method and the Prophet model leverages an additive model that combines seasonal and holiday features. The dataset comprised historical data (1995-2020) of daily average temperatures from major cities, serving as an ideal time series subject for forecasting. This study trained both models on the initial 80% of the data from four cities and used them to forecast the remaining period. This research employed Root Mean Square Error (RMSE) analysis to evaluate the performance of the models. The results indicated that the Prophet model surpassed the XGBoost model in two out of the four cities and matched its performance in the other two.

Keywords: Temperature prediction; XGBoost; Prophet.

1. Introduction

Weather forecasting is the technology to predict atmospheric conditions with given spatiotemporal data. It has been a crucial issue for human society since the beginning of civilization and has a significant impact on human production and development. The modern numerical weather forecast was first proposed by Lewis Fry Richardson who in his book "Weather Prediction By Numerical Process" set out a detailed algorithm of numerical weather prediction [1]. Current mainstream weather forecasting is based on mathematical and physical models, following Richardson's methodology to solve atmospheric motions. Another popular field in recent years is the application of machine learning. With a fast-growing number of studies published aiming to use the method in meteorology and climatology, it is believed that artificial intelligence and machine learning will be important techniques for monitoring and forecasting weather conditions [2]. There are plenty of works so far that has shown great advance in machine learning in weather forecasting subfields. In 2018, Cohen's team showcased an instance of unsupervised learning that outperformed the accuracy of cutting-edge dynamical systems in making seasonal predictions. In 2018, Cohen's team showcased an instance of unsupervised learning that outperformed the accuracy of cutting-edge dynamical systems in making seasonal predictions [3]. In 2020, Kraemer conducted an in-depth analysis of the use of machine learning for predicting solar power consumption in large-scale Internet of Things (IoT) systems [4]. Besides success in long-term forecasts, machine learning is also widely applied in short-term predictions. In 2019, Potočnik conducted a systematic investigation into the short-term prediction of indoor temperatures for time horizons ranging from 1 to 12 hours [5]. Meanwhile, Rahman developed an innovative real-time rainfall prediction system for smart cities using a fusion technique based on machine learning [6]. Barque also made efforts to enhance the accuracy of wind predictions by incorporating feedback from relevant models on the topic [7].

Recognising the important role that machine learning plays in meteorological forecasting, one decided to use one of the most fundamental parameters, the regional mean temperature, as the forecasting target for this study. The chaotic nature of atmospheric motion makes conventional modelling predictions far more computationally intensive than machine learning predictions, while the periodic and lag features of historical data on temperatures make machine learning predictions superior. Focusing on historical data to forecast greatly reduces computational requirements and

provides predictive data with sufficient accuracy both in short- and long-term forecasts. This paper intends to compare two commonly used time-series prediction models and investigate their specific performance by learning to predict on the same dataset.

This study chose two specific models for comparison: the XGBoost and Facebook Prophet models. The Prophet model is a time-series prediction model open-sourced by Facebook in 2017 [8] based on the additive model output interval estimation, which is very inclusive of seasonality, noises, and missing values. The XGBoost, on the other hand, combines gradient boosting and decision trees and is modified for higher speed and performance. In the method section of this paper, this study will describe the data sources used, as well as the specific parameters of the models. Error metrics will be presented for the evaluation of the model performances. In the result section, the prediction results and corresponding error analyses of different models will be shown based on the same training and testing datasets, and the performance of both models will be discussed.

2. Data and Method

The dataset used in this paper is the Daily Temperature of Major Cities available on Kaggle [9]. It can be observed from Fig. 1 that the regional data has strong seasonal fluctuations, and does not have an obvious overall trend, with outliers concentrated in the troughs and peaks. Fig. 2 depicts the overall temperature distribution of the dataset for one city (Beijing). It can be observed that the standard deviation for a given city is relatively larger than the standard deviation for the whole dataset.

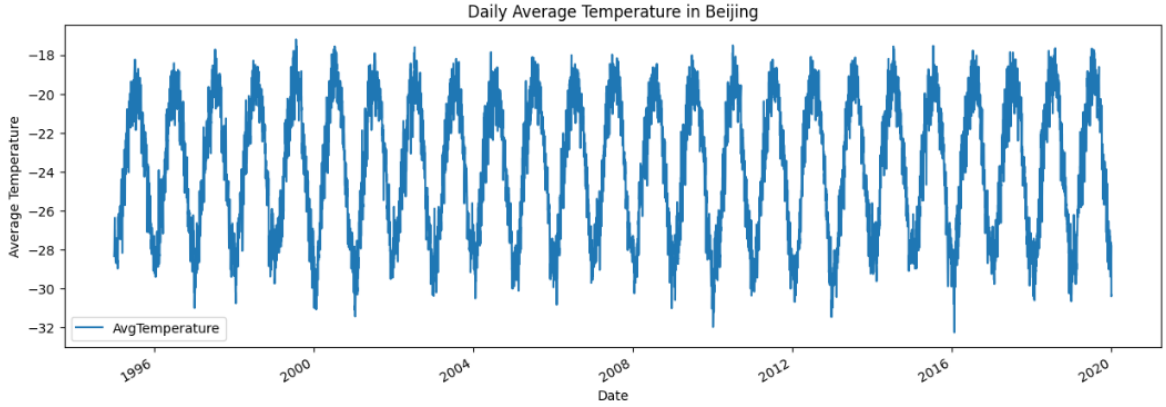


Fig. 1 Beijing daily average temperature series from the main dataset.

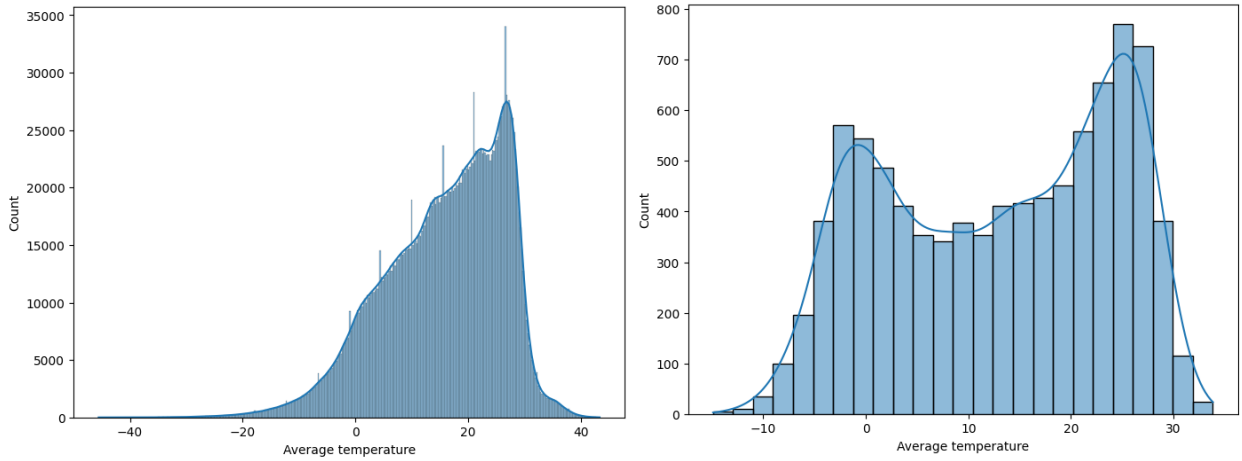


Fig. 2 Average Temperature Histogram of the whole dataset and for Beijing.

The initial stage of the research is to collect and understand the data. For this paper, average temperature data from Kaggle was used. The dataset was later converted into understandable form and predictive analysis was performed. Finally, this study will evaluate the models by statistical method. This is the process of gathering data and preprocessing raw data. The dataset used was examined and amended for missing and duplicate values. The Gradient Boosting Decision Tree

(GBDT) is a model that enhances its predictions through the concept of boosting integration. It employs a forward distribution algorithm for eager learning during the training phase. In each iteration, it learns a Classification and Regression Tree (CART) to fit the residuals between the predictions of the preceding t-1 trees and the actual values of the training samples. EXtreme Gradient Boost (XGBoost) shares the same principles as GBDT, but substantially optimises the latter, including second-order Taylor expansion to optimize the loss function, regularization terms to simplify the model, and Blocks storage structure to parallelise the computation [10].

The Prophet model, developed by Facebook (Taylor and Letham), is based on an addition model consisting of growth, seasonality, and holiday features. The following terms show the structure of the model:

$$Y = g(t) + s(t) + h(t) + \epsilon(t) \tag{1}$$

Here, g is the growth function, in which changepoints are automatically selected for piecewise-linear trends. s is the seasonality composed by Fourier terms of relevant periods. Holiday terms represented by h are dummy variables and ϵ is the white noise error term. It can be easily seen from the structure that the model has the best performance on time series with strong seasonality. Root Mean Square Error (RMSE) was applied to evaluate the accuracy of each model in this paper, given by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

3. Results and Discussion

In this paper, multiple representative cities were chosen as prediction training data. For instance, comparing Fig. 1 and Fig. 3, the temperature in Sydney is more unstable and the anomalies during the high-temperature intervals are larger. Hence, expanding the training set to multiple cities has diverse data and acts as a way to compare the performance of the two models more comprehensively.

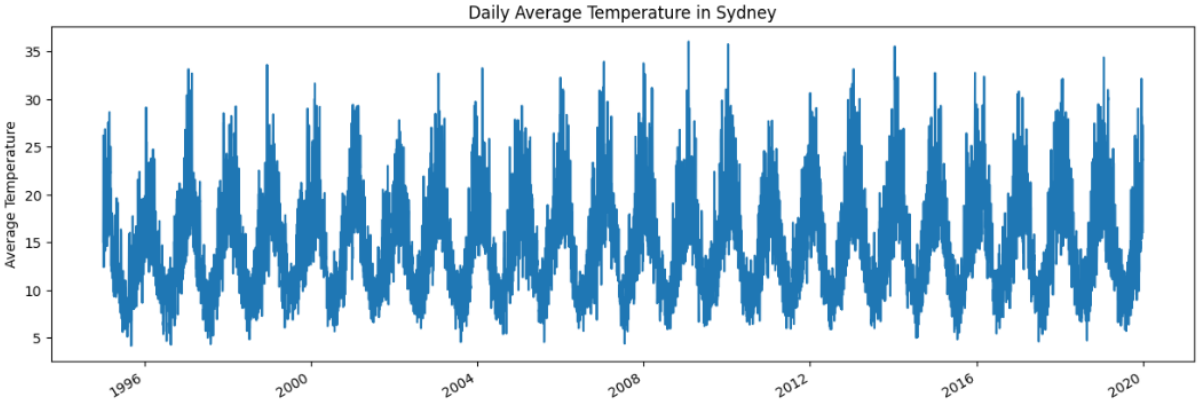


Fig. 3 Sydney daily average temperature series.

Fig. 4 shows the prediction done by the XGBoost model. For each of the four forecasts above, the model selected the first 80% of the daily temperature data from 1995 to 2020 as the training set and the last 20% as the test set. It can be observed from Fig. 4 and Table 1. that all four predictions incorporate the seasonal feature well. The temperature fluctuation and trend in each season are reflected precisely consistent with the main trend of the test data. However, during peaks and troughs in temperature, when the raw data shows a lot of high and low temperature fluctuations and noises, the XGBoost model fails to perfectly represent the intense temperature changes, but instead shows relatively smooth change points. The following is the error analysis of each prediction result.

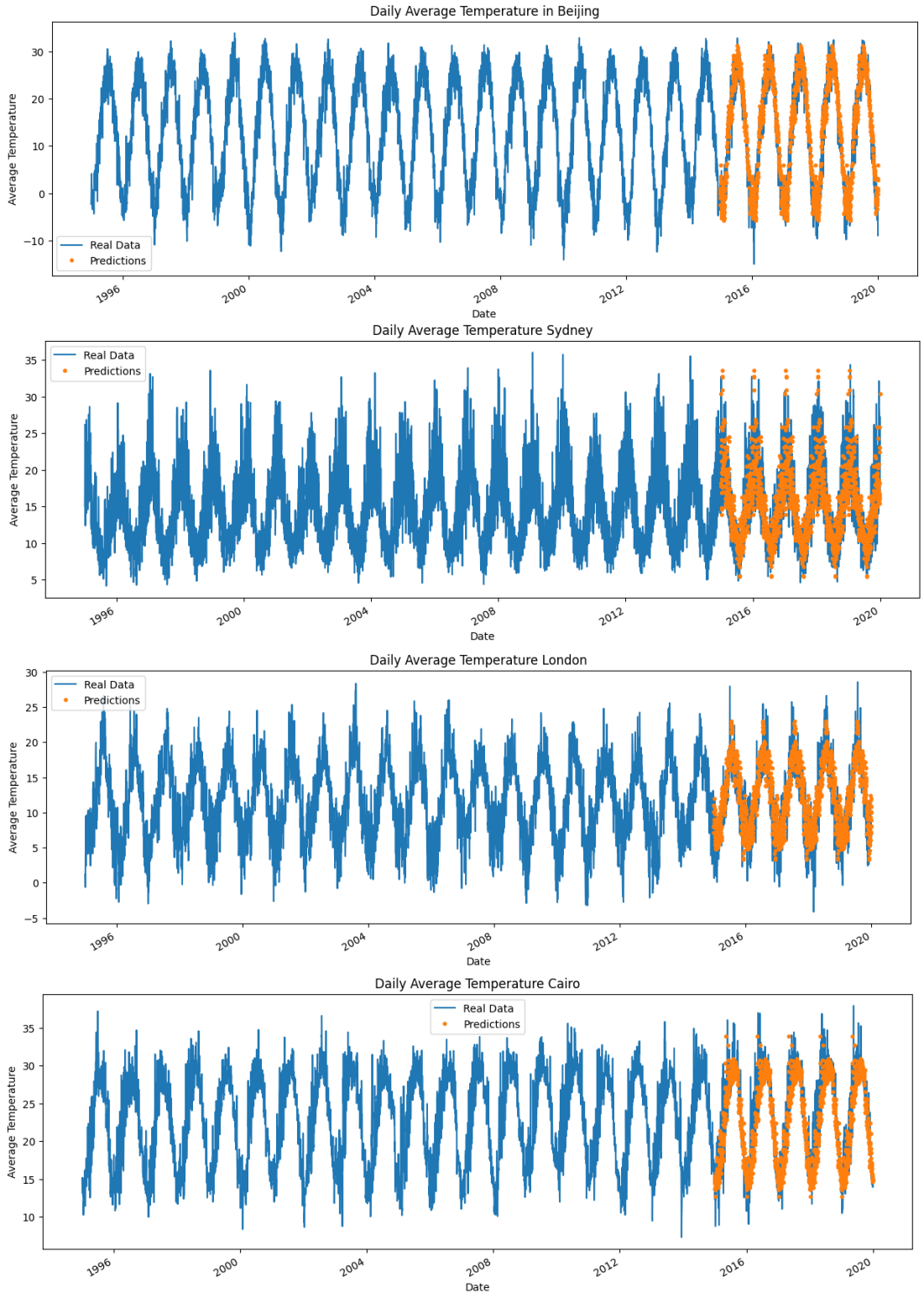


Fig. 4 The prediction results of XGBoost on the testing dataset for Beijing, Sydney, London, and Cairo.

Table 1. RMSE score of the XGBoost predictions.

City	RMSE score
Beijing	3.67
Sydney	4.53
London	3.40
Cairo	2.84

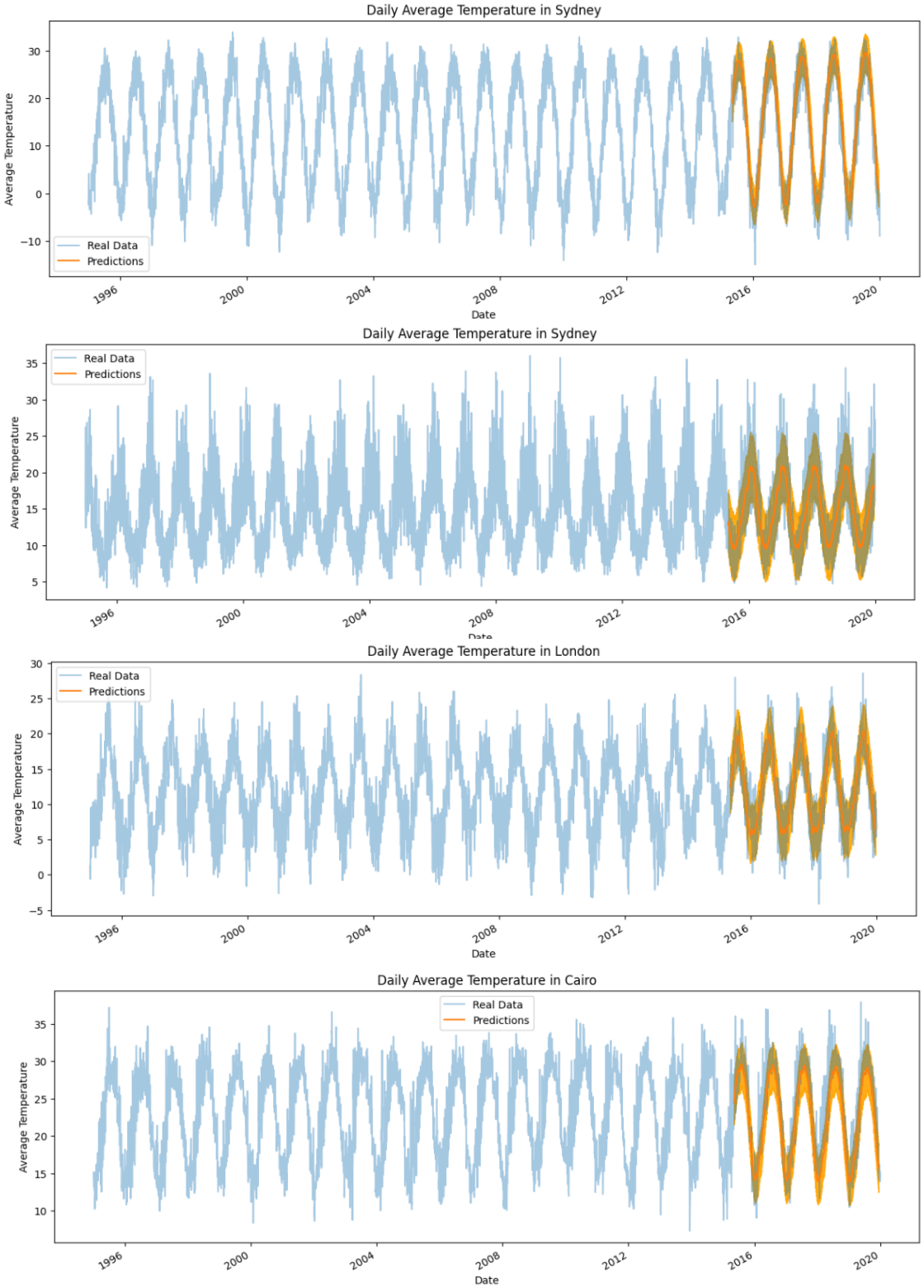


Fig. 5 The prediction results of Prophet on the testing dataset for Beijing, Sydney, London, and Cairo.

In the Sydney dataset, where the errors are particularly high, sharp fluctuations are observed more frequently than others. Sydney experiences many anomalous high-temperature values during the high-temperature period, which makes XGBoost predictions perform poorly in this period. Although the overall error values for all four predictions are relatively low, XGBoost does not predict fluctuations well in the data at both high and low ends. Fig. 5 shows the prediction done by the Prophet model. Similarly, the data is divided into a training set and a test set by 80%. Unlike the previous

models, the Prophet model gives not only the predicted values but also the error intervals, which are represented in Fig. 5 as orange translucent blocks. Although the upper and lower bounds on the error given by the model are not used in the calculation of the error, the Prophet model still gives relatively smaller error scores in Sydney and London forecasts, as shown in Table 2.

Table 2. RMSE score of the Prophet predictions.

City	RMSE score
Beijing	3.82
Sydney	3.49
London	3.06
Cairo	2.85

With the inclusion of error intervals, Prophet not only figures out the seasonal trends in the data as accurately as the XGBoost model but also covers some of the anomalies in the peaks and valleys, which is particularly noticeable in the Sydney data, where the error scores are about 23 per cent smaller than those of XGBoost in predicting the temperature.

4. Conclusion

Both the XGboost and Prophet models perform well in time series forecasting. The former optimises the GBDT method to analyse the data using greedy learning algorithms, while the latter uses an additive model combining seasonal and holiday features and does well for data with strong such characteristics. In this paper, the historical data (1995-2020) of the daily average temperature of the major cities is used as an ideal time series and forecast subject. Both models learn the first 80% of the data from four cities and then perform the forecast for the remaining period. The remaining 20% of the data was used as the test set to compare with the predicted data for error analysis. By applying the RMSE analysis, the Prophet model outperforms the XGBoost model in two of the four cities and ties in the remaining two. Many potential improvements could be explored in a follow-up study. First, other relevant daily data can be introduced to perform multivariate analysis prediction, such as windspeed and pressure. Feature analysis can be performed to examine those effect weights on the main series, and therefore one can figure out the relationship between them. Alternatively, introducing other base models for comparison would be a viable option, such as ARIMA and SARIMA, linear regression, KNN, etc. Finally, changing the data set so that it shows an upward or downward trend is also a possible way of testing the sensitivity of the model to the general trend of the seasonal data, such as the global warming predictions.

References

[1] Lewis F R. Weather prediction by numerical process. Hardpress Publishing, 2012.

[2] Bcochenk B, Ustrnul Z. Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives. Atmosphere, 2022, 13(2): 180.

[3] Cohen J, Coumou D, Hwang J, et al. S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts Wiley Interdisciplinary Reviews: Climate Change, 2018, 10(2): 10-15.

[4] Kramer F A, Palma D, Braten A E, etc. Operationalizing Solar Energy Predictions for Sustainable, Autonomous IoT Device Management. IEEE Internet of Things Journal, 2021, 1: 1.

[5] Potocnik P, Vidrih B, Kitanovski A, et al. Neural network, ARX, and extreme learning machine models for the short-term prediction of temperature in buildings. Building Simulation, 2019, 12(6): 1077-1093.

[6] Rahman A, Abbas S, Gollapalli M, et al. Rainfall Prediction System Using Machine Learning Fusion for Smart Cities. Sensors, 2022, 22(9): 3504.

[7] Barque M, Martin S, Vianin N, et al. Improving wind power prediction with retraining machine learning algorithms.Jumps, 2018

[8] Taylor S J, Letham B. Forecasting at scale. 2017.

[9] Daily Temperature of Major Cities. Retrieved from: https://www.kaggle.com/datasets/sudalairajkumar/daily-temperature-of-major-cities.

[10] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.