Article on Evaluation Project

# Project Name - Temperature Forecast Prediction

# Project Description:

#### This data is for the purpose of bias correction of next-day maximum and minimum air temperatures forecast of the LDAPS model operated by the Korea Meteorological Administration over Seoul, South Korea. This data consists of summer data from 2013 to 2017. The input data is largely composed of the LDAPS model's next-day forecast data, in-situ maximum and minimum temperatures of present-day, and geographic auxiliary variables. There are two outputs (i.e. next-day maximum and minimum air temperatures) in this data. Hindcast validation was conducted for the period from 2015 to 2017.

# Attribute Information:

1. station - used weather station number: 1 to 25
2. Date - Present day: yyyy-mm-dd ('2013-06-30' to '2017-08-30')
3. Present\_Tmax - Maximum air temperature between 0 and 21 h on the present day (Â°C): 20 to 37.6
4. Present\_Tmin - Minimum air temperature between 0 and 21 h on the present day (Â°C): 11.3 to 29.9
5. LDAPS\_RHmin - LDAPS model forecast of next-day minimum relative humidity (%): 19.8 to 98.5
6. LDAPS\_RHmax - LDAPS model forecast of next-day maximum relative humidity (%): 58.9 to 100
7. LDAPS\_Tmax\_lapse - LDAPS model forecast of next-day maximum air temperature applied lapse rate (Â°C): 17.6 to 38.5
8. LDAPS\_Tmin\_lapse - LDAPS model forecast of next-day minimum air temperature applied lapse rate (Â°C): 14.3 to 29.6
9. LDAPS\_WS - LDAPS model forecast of next-day average wind speed (m/s): 2.9 to 21.9
10. LDAPS\_LH - LDAPS model forecast of next-day average latent heat flux (W/m2): -13.6 to 213.4
11. LDAPS\_CC1 - LDAPS model forecast of next-day 1st 6-hour split average cloud cover (0-5 h) (%): 0 to 0.97
12. LDAPS\_CC2 - LDAPS model forecast of next-day 2nd 6-hour split average cloud cover (6-11 h) (%): 0 to 0.97
13. LDAPS\_CC3 - LDAPS model forecast of next-day 3rd 6-hour split average cloud cover (12-17 h) (%): 0 to 0.98
14. LDAPS\_CC4 - LDAPS model forecast of next-day 4th 6-hour split average cloud cover (18-23 h) (%): 0 to 0.97
15. LDAPS\_PPT1 - LDAPS model forecast of next-day 1st 6-hour split average precipitation (0-5 h) (%): 0 to 23.7
16. LDAPS\_PPT2 - LDAPS model forecast of next-day 2nd 6-hour split average precipitation (6-11 h) (%): 0 to 21.6
17. LDAPS\_PPT3 - LDAPS model forecast of next-day 3rd 6-hour split average precipitation (12-17 h) (%): 0 to 15.8
18. LDAPS\_PPT4 - LDAPS model forecast of next-day 4th 6-hour split average precipitation (18-23 h) (%): 0 to 16.7
19. lat - Latitude (Â°): 37.456 to 37.645
20. lon - Longitude (Â°): 126.826 to 127.135
21. DEM - Elevation (m): 12.4 to 212.3
22. Slope - Slope (Â°): 0.1 to 5.2
23. Solar radiation - Daily incoming solar radiation (wh/m2): 4329.5 to 5992.9
24. Next\_Tmax - The next-day maximum air temperature (Â°C): 17.4 to 38.9
25. Next\_Tmin - The next-day minimum air temperature (Â°C): 11.3 to 29.8T

You have to build separate models that can predict the minimum temperature for the next day and the maximum temperature for the next day based on the details provided in the dataset.

# ABSTRACT:

This project aims to correct biases in the next-day maximum and minimum air temperature forecasts provided by the LDAPS model, operated by the Korea Meteorological Administration, for Seoul, South Korea. The dataset spans the summer months from 2013 to 2017 and includes various input features such as current-day temperatures, LDAPS model forecast variables, and geographical auxiliary variables. The objective is to develop predictive models that can accurately forecast the next-day maximum and minimum temperatures.

# INTRODUCTION:

Accurate temperature forecasting is vital for various sectors, including agriculture and energy management. However, forecast models often exhibit biases, resulting in inaccurate predictions. Bias correction techniques are designed to mitigate these errors and enhance the reliability of temperature forecasts. This project aims to correct biases in the next-day maximum and minimum air temperature forecasts produced by the LDAPS model.

Number of Rows: 7752

Number of Columns: 25

# Methodology:

**Data Preprocessing:** Perform data cleaning, handle missing values, and encode categorical variables if any.

**Exploratory Data Analysis:** Conduct exploratory data analysis to understand the distribution of variables, identify outliers, and analyze correlations.

**Feature Scaling:** Scale the numerical features to ensure uniformity and prevent dominance of certain variables.

**Feature Selection:** Use techniques like variance inflation factor (VIF) to identify and remove multicollinear features.

**Model Development:** Train machine learning models to predict the next-day maximum and minimum temperatures. Use algorithms like linear regression, random forest, and gradient boosting.

**Model Evaluation:** Evaluate the performance of models using metrics such as mean squared error, mean absolute error, and R-squared.

**Bias Correction:** Apply bias correction techniques to adjust the model predictions and minimize errors.

**Model Comparison:** Compare the performance of biased and bias-corrected models to assess the effectiveness of bias correction.

# **Exploratory Data Analysis of Temperature Forecast Dataset**

# Introduction:

Exploratory Data Analysis (EDA) is crucial for uncovering patterns and relationships within a dataset. In this article, we conduct EDA on a temperature forecast dataset, focusing on maximum and minimum temperatures, as well as various meteorological variables. Through univariate analysis, bivariate analysis, statistical analysis, and correlation analysis, we aim to understand the dataset's characteristics and identify potential trends or patterns.

# Univariate Analysis:

- The maximum temperature ranges from 21°C to 38°C, while the minimum temperature ranges from 14°C to 30°C.

- The ratio of average temperatures appears consistent across all years.

# Bivariate Analysis:

- We observe variations between present maximum (Tmax) and minimum (Tmin) temperatures.

- Yearly variations in maximum and minimum temperatures are not evident.

# Statistical Analysis:

- Statistical measures suggest the presence of outliers, indicated by discrepancies between mean and 50th percentile values.

# Correlation Between Variables:

- Present\_Tmax and Present\_Tmin exhibit a strong correlation with the target variables.

- LDAPS\_Tmax\_lapse and LDAPS\_Tmin\_lapse also show significant correlations with the target variables.

- LDAPS cloud cover and precipitation variables demonstrate negative correlations with the target variables.

# Skewness in Datasets:

- Present\_Tmax, Present\_Tmin, LDAPS\_RHmax, LDAPS\_Tmax\_lapse, and LDAPS\_Tmin\_lapse exhibit left skewness.

- LDAPS\_WS, LDAPS\_LH, and other LDAPS variables related to cloud cover and precipitation display right skewness.

# Data Preprocessing:

We split the dataset into feature variables, consisting of meteorological parameters, and target variables, representing next-day minimum and maximum temperatures. Afterward, we perform feature scaling to normalize the data distribution and prevent dominance of certain features during model training.

# Model Training:

We train the models in two steps, first for predicting next-day minimum temperature (Next\_Tmin) and then for predicting next-day maximum temperature (Next\_Tmax). We utilize various regression algorithms, including RandomForestRegressor, DecisionTreeRegressor, ExtraTreesRegressor, BaggingRegressor, AdaBoostRegressor, GradientBoostingRegressor, and XGBRegressor. Each model is trained and evaluated using cross-validation and performance metrics such as mean squared error and R-squared.

# Hyperparameter Tuning:

To optimize the model's performance, we perform hyperparameter tuning using GridSearchCV. This technique systematically searches for the best combination of hyperparameters within predefined ranges. By fine-tuning the models' parameters, we aim to enhance their predictive accuracy and generalization capabilities.

# Model Evaluation:

We evaluate the trained models using cross-validation scores and R-squared values to assess their performance on unseen data. The models with the highest cross-validation scores and R-squared values are considered optimal for predicting next-day temperatures.

# Prediction and Model Deployment:

Finally, we save the best-performing model for next-day temperature prediction based on the results of hyperparameter tuning. This model can be deployed in production environments to make real-time predictions of next-day temperatures, enabling informed decision-making in various domains.

# Conclusion:

Through comprehensive exploratory data analysis, we have gained valuable insights into the temperature forecast dataset. The analysis highlights the range and distribution of temperature variables, identifies potential outliers, and explores correlations between variables. These findings lay the foundation for further analysis and model development, providing essential guidance for predictive modeling tasks related to temperature forecasting.

This project aims to improve the accuracy of next day & present maximum and minimum air temperature forecasts by correcting biases in the predictions generated by the LDAPS model. By implementing bias correction techniques and evaluating model performance, we seek to enhance the reliability of temperature forecasts, which can have significant implications for various sectors including agriculture, energy, and public safety.

###### =====================Thanks====================