

LSTM Code Explanation

A. MODELING

1. Look at the “**LSTM MODEL**” file.
This script requires the following libraries: pandas, numpy, matplotlib, tensorflow.keras, sklearn.
These functions are imported at the beginning of the file.
2. Lines 6–13: The Excel file “**GHGFactors.xlsx**” is read.
 - 1st row: years (1993–2022)
 - 2nd row: GHG emission data (target variable)
 - 3rd–5th rows: independent variables (Fossil, Renewables, Electricity Demand)
3. Line 15: The features (f1, f2, f3) are stacked and combined with the target variable to form a single dataset (data).
4. Line 18: MinMaxScaler() normalizes all values between 0 and 1 to improve numerical stability during training.
5. Lines 21–28: The create_sequences() function generates time-based data windows.
 - Each input (X) includes the previous 5 time steps.
 - The target (y) is the next GHG value.
The sequence length (time_steps = 5) determines the model’s temporal memory.
6. Lines 31–40: The LSTM model is defined as follows:
 - Layer 1: LSTM(64, activation="elu", return_sequences=True)
 - Layer 2: Dropout(0.2)
 - Layer 3: LSTM(32, activation="elu")
 - Output: Dense(1)
Optimizer: Adam(learning_rate=0.001); Loss: MSE.
The model is trained for 100 epochs with a batch size of 8.
7. Line 41: Model training starts with model.fit(X, y, epochs=100, batch_size=8, verbose=0).
All hyperparameters were optimized according to the minimum error rate observed during training.

8. Lines 43–49: After training, the model produces predicted GHG values (predicted_scaled_all), which are inverse-transformed to real GHG units (predicted_all).
9. Lines 50–54: Actual GHG values (y_actual_all) are also inverse-transformed for comparison.
10. Line 56: The error rate is calculated as the mean absolute percentage difference between actual and predicted GHG values and printed in the console as the model's overall accuracy.

B. IMPACT FACTOR

1. Lines 99–115: Factor (feature importance) analysis is performed after model training.
Independent variables (f1, f2, f3) are normalized again using MinMaxScaler().
2. Line 103: A Linear Regression model is fit as
$$\text{GHG} = \beta_0 + \beta_1 \cdot \text{Fossil} + \beta_2 \cdot \text{Renewables} + \beta_3 \cdot \text{Demand}.$$
3. Line 104: The coefficients ($\beta_1, \beta_2, \beta_3$) represent each factor's direction and influence strength.
Positive values indicate an increasing effect on GHG emissions, while negative values indicate a decreasing effect.
4. Lines 106–115:
 - A bar chart displays each factor's effect.
 - Green bars indicate positive effects, red bars indicate negative ones.
 - Coefficient values are printed above the bars for clarity.
5. This regression-based analysis provides a simple and model-agnostic interpretability of the trained LSTM model.

C. PREDICTION

1. Lines 57–97: The future prediction (forecasting) process begins, where the model predicts upcoming GHG values using its previous outputs.
2. Lines 59–65: The array ghg_only (predicted GHG values) is reshaped and used to train a secondary LSTM model (ghg_model) with the same configuration.
This model captures the long-term GHG trend patterns.
3. Line 69: The number of forecast years is defined as n_future = 12, corresponding to 2030 as the final prediction year.

4. Lines 70–80: The model uses the last five GHG values (`current_input`) and predicts each following year iteratively.
Each new prediction is added to the sequence, enabling continuous chained forecasting.
5. Line 83: The predicted future GHG values (`future_predictions`) are inverse-transformed to obtain the actual emission estimates.
6. Lines 86–94:
 - A plot is generated showing Actual GHG, Model Prediction (1993–2018), and Future Forecast (2019–2030).
 - Blue: actual data, Green: fitted values, Red: future forecasts.
 - The chart includes labeled axes, grid, and legend.
7. Lines 117–129: Results are summarized into two tables:
 - Past (1993–2018): Actual vs. Predicted GHG
 - Future (2019–2030): Predicted GHG onlyBoth tables are combined and printed as the 1993–2030 forecast summary.
8. The model parameters and forecast structure were optimized according to the lowest error rate, ensuring reliable prediction performance for future GHG emissions.