Energy Production Prediction Using Deep Learning

Objective:

Predict hourly energy production based on given features such as weather conditions, time, and energy source (e.g., solar) using a deep learning approach.

Dataset:

The dataset provided contains the following features:

- Date and Hour: The date and hour of the energy production record.
- Date, StartHour, EndHour: The starting and ending hour of the energy production period.
- Source: The type of renewable energy source (e.g., Solar, Wind).
- Production: The actual energy produced in that hour (target variable).
- dayOfYear: The day of the year (1-365).
- dayName: The name of the day (e.g., Monday, Tuesday).
- monthName: The name of the month (e.g., January, February).

Steps:

1. Data Preprocessing:

- Handle missing values (if any).
- Convert date-related columns into proper datetime objects.
- Create additional time-based features like hour, day, and month if necessary.
- Normalize or standardize features such as production values, if needed.

2. Exploratory Data Analysis (EDA):

- Visualize production trends based on time features like hour, day of the week, and month.
- Investigate relationships between the energy source and production.
- Look for any seasonality patterns in the data.

3. Model Selection and Building:

- Baseline Model: Start by creating a simple linear regression model or a random forest regressor to set a baseline.

- Deep Learning Model: Build a model using either LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Units) to handle sequential data.

- Include Dense layers for feature processing, Dropout layers for regularization, and use ReLU activation.

- The final layer should output a single value representing the predicted energy production for that hour.

4. Model Training and Evaluation:

- Split the data into training and testing sets.

- Train the model using metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), and R^2 score.

- Plot the training and validation losses to monitor for overfitting or underfitting.

5. Hyperparameter Tuning:

- Tune hyperparameters like the number of layers, neurons per layer, batch size, and learning rate using GridSearchCV or RandomSearchCV.

- Compare the performance before and after tuning.

6. Final Submission:

- Submit a Jupyter Notebook or Python script with all your code.

- Include a short report (500 words) summarizing the approach, model architecture, and results.

Dataset Access:

The dataset is provided as 'intermittent-renewables-production-france.csv'. Use it for your analysis.

Evaluation Criteria:

- Preprocessing: Handling of missing data, feature engineering, and normalization.
- EDA: Depth of analysis and insights from visualizations.
- Modeling: Appropriateness of model selection, understanding of time series, and deep learning techniques.
- Evaluation: Use of appropriate metrics and tuning efforts.
- Report Quality: Clarity and explanation of the approach and results.

Deadline:

Submit your work by 11 AM on 20th September 2024.