Open-Ended Lab Report

Design a model to Forecast in the CityLearn challenge.

Asad Khan
Reg No: 21JZELE0442

Instructor: Sir Irshad Ullah
June 2025

University of Engineering and Technology, Peshawar – Jalozai Campus
Course: EE-271 Machine Learning Lab

Problem Statement

The objective of this open-ended lab is to develop a machine learning model to forecast energy consumption based on historical time-series data using Long Short-Term Memory (LSTM) neural networks. The goal is to predict the next hour's energy usage by learning from previous time steps, capturing both temporal and categorical influences on demand.

Dataset Description

The dataset consists of synthetic smart-building-like data with the following features:

- **Timestamp** – Date and time in string format
- **EnergyConsumption** – Target variable (continuous)
- **Temperature, Humidity, Occupancy, RenewableEnergy** – Numeric features
- **HVACUsage, LightingUsage, DayOfWeek, Holiday** – Categorical features

Methodology

1. Data Preprocessing

- **Timestamp Conversion:** Converted datetime strings into UNIX float time (seconds since epoch).

- **One-Hot Encoding:** All categorical variables (e.g., HVACUsage, DayOfWeek) were transformed into binary dummy variables using `pandas.get_dummies()`.
- **Normalization:** All numerical columns were scaled to [0,1] using MinMaxScaler.

2. Sequence Generation

A rolling window of 24 time steps (representing hours) was used to predict the 25^{th} hour's energy consumption.

3. LSTM Model Architecture

The LSTM model was built using the following architecture:

- LSTM (64 units, return_sequences=True)
- Dropout (0.2)
- LSTM (32 units, return_sequences=False)
- Dropout (0.2)
- Dense (16 units, ReLU activation)
- Dense (1 unit, linear output)

Compiled with Adam optimizer and Mean Squared Error (MSE) as the loss function.

Results and Evaluation

- **Training Loss:** Decreased and stabilized around 0.0023 over 20 epochs.
- **Evaluation Metric:** Mean Squared Error (MSE) was used to measure prediction performance:
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$
- **Visualization:** A plot of predicted vs actual energy values showed good alignment, confirming model performance.

Conclusion

The LSTM-based approach accurately forecasted future energy consumption using a combination of numeric, temporal, and categorical features. One-hot encoding and timestamp conversion enabled the model to effectively learn patterns and trends in the data.

Future Work

- Integrate weather forecast features for multi-day prediction.
- Use GRU or Transformer-based models for comparison.
- Deploy the model on real-time smart building data.

Instructor Comments

Instructor Name: Sir Irshad Ullah