Homework 2

1. Use the dataset from UCI Machine Learning Repository: "Individual household electric power consumption" for performing time series analysis.

https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption

2. Perform Exploratory Data Analysis (EDA) of the dataset:

- Visualize time series trends
- Check for seasonality and cyclical patterns
- Analyze distribution of power consumption
- Identify and handle missing values or outliers

3. Implement a linear regression model to predict power consumption for the last three time periods:

- Split the data into training and testing sets
- Prepare features (consider lag variables, time-based features)
- Train the model and make predictions

4. Evaluate the linear regression model using appropriate metrics:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R-squared (R^2) value

5. Implement a Recurrent Neural Network (RNN) for power consumption prediction:

- Preprocess data for RNN input
- Design and train the RNN model
- Make predictions and visualize results
- Compare performance metrics with linear regression

6. Implement Long Short-Term Memory (LSTM) for power consumption prediction:

- Preprocess data for LSTM input
- Design and train the LSTM model
- Make predictions and visualize results
- Compare performance metrics with previous models

7. Implement an LSTM model with an Attention layer for power consumption prediction:
- Design and train the LSTM model with Attention
- Make predictions and visualize results
- Compare performance metrics with previous models
- Analyze the Attention weights to interpret model focus

8. Data augmentation experiment:
- Modify up to 10% of the dataset to potentially improve prediction results
- Retrain and evaluate all three models (RNN, LSTM, LSTM with Attention)
- Compare the impact of data changes on each model's performance

9. Data reduction experiment:
- Remove up to 10% of the data randomly
- Retrain and evaluate all three models
- Compare how data reduction affects each model's performance

10. Data resolution experiment:
- Reduce the time resolution of the data by 50% (e.g., from minute-level to 2-minute intervals)
- Retrain and evaluate all three models
- Analyze how changes in data resolution impact each model's performance

11. Conclusion and insights:
- Summarize findings from all experiments
- Discuss which model performed best under different conditions
- Provide insights on the dataset's characteristics and their impact on model performance