**ADTA-5560 Recurrent Neural Networks for Sequence Data**

**by**

Sai Satish Pallapolu, 11753831

**Project Proposal**

**Forecasting Electricity Demand Using RNNs**

**Purpose and Story**

Electricity demand forecasting is a critical application of time series analysis, particularly in the context of modern energy grids that increasingly rely on renewable energy sources. The intermittent nature of renewables like solar and wind energy makes it essential to accurately predict electricity demand to ensure grid stability, optimize energy production, and reduce costs. This project will focus on forecasting hourly electricity demand based on historical consumption data. The insights gained will help grid operators make informed decisions about energy distribution and storage, ultimately contributing to a more sustainable and efficient energy system.

**Research Questions**

1. How accurately can we forecast hourly electricity demand using historical data?
2. What are the key temporal patterns (e.g., daily, weekly, seasonal) in electricity consumption?
3. How can external factors like temperature and holidays improve the forecasting model?

Shape

**Data Set Summary**

**Data Source**

The data set is sourced from the **U.S. Energy Information Administration (EIA)**, specifically the **BALANCE** files for the years 2022–2025. These files contain hourly electricity balance data, including demand, generation, and interchange metrics, for various U.S. regions. The data is original, reliable, and qualifies as a time series, making it suitable for forecasting.

**Link to Data Source**

* [EIA Electricity Data](https://www.eia.gov/electricity/gridmonitor/sixMonthFiles/)

**Variables**

The dataset includes the following key variables:

1. **Balancing Authority**: The region or entity responsible for grid balancing (e.g., AECI).
2. **Data Date**: The date of the observation.
3. **Hour Number**: The hour of the day (1–24).
4. **Local Time at End of Hour**: The local timestamp for the end of the hour.
5. **UTC Time at End of Hour**: The UTC timestamp for the end of the hour.
6. **Demand Forecast (MW)**: The forecasted electricity demand for the hour.
7. **Demand (MW)**: The actual electricity demand for the hour (target variable).
8. **Net Generation (MW)**: The total electricity generated in the region.
9. **Total Interchange (MW)**: The net electricity interchange (imports/exports) for the region.
10. **Net Generation by Fuel Type**: Detailed breakdown of electricity generation by fuel source (e.g., coal, natural gas, nuclear, solar, wind, etc.).
11. **Region**: The geographic region (e.g., MIDW for Midwest).

**Additional Features (to be incorporated)**

* **Temperature (°F)**: Average hourly temperature for the region (external feature, to be sourced from a weather API or dataset).
* **Day of Week**: Categorical variable indicating the day of the week (e.g., Monday, Tuesday).
* **Holiday Indicator**: Binary variable indicating whether the day is a holiday (1) or not (0).

**Data Size**

* The data set contains hourly electricity demand data for the past ~3 years (2022–2025), resulting in approximately **43,800 data points per region**. This size is sufficient for in-depth statistical analysis and model training.

**Reason for Selecting This Data Set**

* The data set is original and reliable, sourced directly from a government agency.
* It includes hourly granularity, making it ideal for capturing temporal patterns, which is essential for forecasting.
* The inclusion of generation and interchange metrics provides additional context for understanding demand patterns.