Time Series Forecasting of Electricity Consumption

## Overview

This project aims to forecast the hourly electricity consumption for different buildings on the Michigan State University (MSU) campus. Electricity consumption data is recorded every hour, and our goal is to predict future usage for various forecast windows (1 hour, 10 hours, 24 hours, etc.).

We have tried multiple approaches:

- **Naive Forecasting** for 1-hour ahead predictions.

- **ARIMA (AutoRegressive Integrated Moving Average)** for 10-hour and 24-hour ahead predictions, with parameters found using an *auto ARIMA* approach (e.g., p, d, q) = (1, 1, 2)).

- **Prophet (Facebook Prophet)** to explore the forecasting capabilities on a longer time horizon with additional regressor and seasonality options.

So far:

- The Naive model often outperforms other methods for 1-hour ahead predictions.

- ARIMA provides moderate results for longer forecast windows.

- Prophet experiments have allowed us to incorporate multiple seasonalities, holiday regressors (e.g., academic breaks), and custom features such as a business hours flag. These experiments include iterative forecasting, cross-validation, and rolling window experiments to determine the optimal amount of recent data for training.

## Repository Contents

1. **NaiveForecasting.ipynb**

Demonstrates the Naive approach for 1-hour ahead predictions and compares predictions with actual consumption values.

2. **ARIMA-10hours.ipynb**

Explores ARIMA models for 10-hour ahead predictions using auto ARIMA to select optimal parameters. Plots and evaluation metrics are provided.

3. **Prophet\_Experiments.ipynb**

- **Iterative Forecasting with Prophet:**

Uses an iterative forecasting approach where the model is retrained regularly, forecasting 24 hours ahead.

- **Cross-Validation and Rolling Window Experiments**:

Implements Prophet’s cross-validation features to simulate retraining the model on a fixed-length rolling window (e.g., 30, 60, 90, 180, or 365 days). These experiments help to identify the impact of training window size on forecast performance using metrics such as MAE and MAPE.

- **Regressor Optimization:**

Studies the effect of adding additional regressors—such as academic break periods (summer break, winter break), weekly seasonality, and a business hours flag—on forecast accuracy. Early results suggest that while the academic break regressor increased the error (MAPE) on the test set, the default weekly and business hours settings performed comparably to the baseline.

4**. Figures/Plots**

- Plots showing actual vs. predicted consumption values for the test set, comparing the different methods (Naive, ARIMA, and Prophet) for various forecasting horizons.

## Data Description

- **Source**: Hourly electricity consumption data from MSU campus buildings.

- **Frequency**: Hourly.

- **Features**:

- **Timestamp (ds**): Date/Time of recording.

- **Consumption** **(y)**: Electricity usage in appropriate units (e.g., MWh).

- (**Optional)** External regressors such as weather data, occupancy, and academic calendar data may also be incorporated.

## Project Workflow

1. **Data Preprocessing**

- Load the raw data and handle missing values.

- Convert timestamps to a uniform date-time format.

- Split the data into training and test sets (typically using a time-based split).

2. **Exploratory Data Analysis (EDA)**

- Visualize trends, seasonality, and outliers.

- Check for daily, weekly, or monthly patterns.

- Identify correlations with external factors if applicable (e.g., weather).

3. **Modeling**

- **Naive Forecasting**:

Uses the last observed value as the forecast for the next step.

- **ARIMA**:

Uses Auto ARIMA to select the optimal `(p, d, q)` parameters for 10-hour and 24-hour forecasts.

- **Prophet**:

- Initial experiments implement iterative forecasting, where the model is periodically retrained to forecast 24 hours ahead.

- Cross-validation is performed using fixed-length rolling windows to evaluate performance (MAE, MAPE) over various training window sizes.

- Additional regressors (academic break schedules, weekly seasonality, business hours) are tested. While academic break regressors increased MAPE, the other features maintained comparable performance with the default settings.

4. **Results**

- Visualizations comparing predicted vs. actual consumption.

- Summary of forecast accuracy metrics (e.g., MAE, RMSE, MAPE) for each modeling approach.

- Iterative improvement through experiments with Prophet has provided insights into the optimal training window size and the impact of various regressors, guiding future model refinements.

5. **Future Steps**

- Continue to refine the Prophet model by testing different fixed rolling windows and hyperparameter tuning (e.g., changepoint and seasonality priors).

- Explore additional regressors (such as detailed weather or occupancy data) to further enhance predictive performance.

- Investigate ensemble methods combining the strengths of ARIMA, Prophet, and machine learning techniques.

## Still need to gather Prophet model results and add it here.