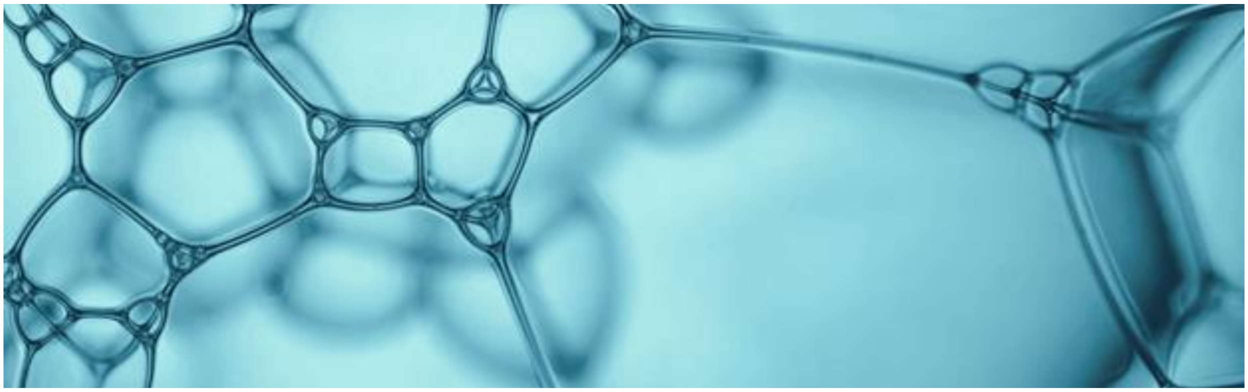


DATA MINING PROJECT

Comprehensive Forecasting System with User Interface for Multiple Sectors

DS-N

FATIMA AZFAR (20i-2658)

RAFFAY AHMED (21i-1674)

ARIMA Model Report:

Purpose:

To forecast future points in a series by exploiting dependencies among the data points.

Methodology:

Data was made stationary through differencing, verified by the Augmented Dickey-Fuller test.

Model parameters (p, d, q) were chosen based on the ACF and PACF.

Results:

Provided reliable short-term forecasts, particularly effective for non-seasonal patterns.

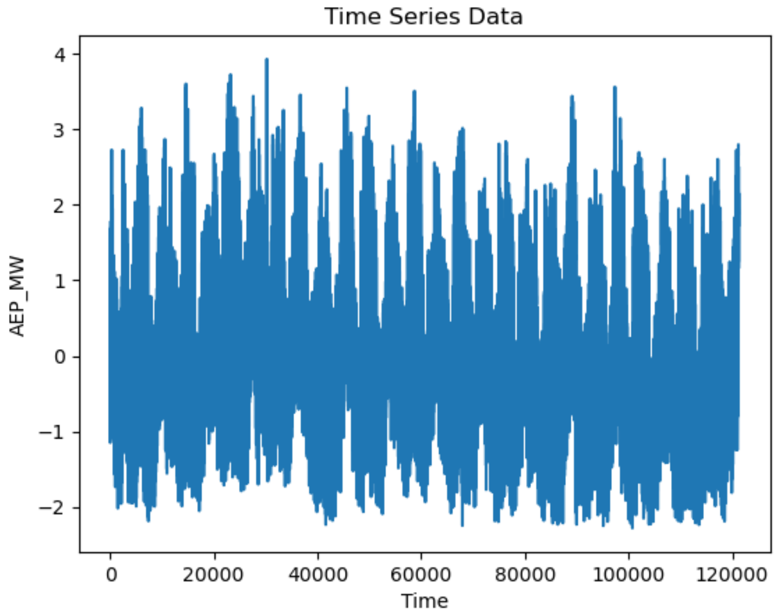
Evaluation:

Evaluated using MSE, showcasing solid performance in capturing trends without overfitting.

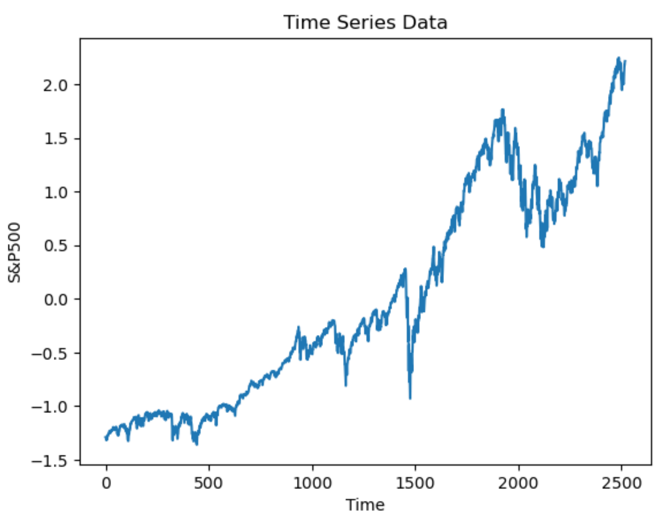
Visualization:

Time series plots with overlaid forecasts to visualize fit and accuracy.

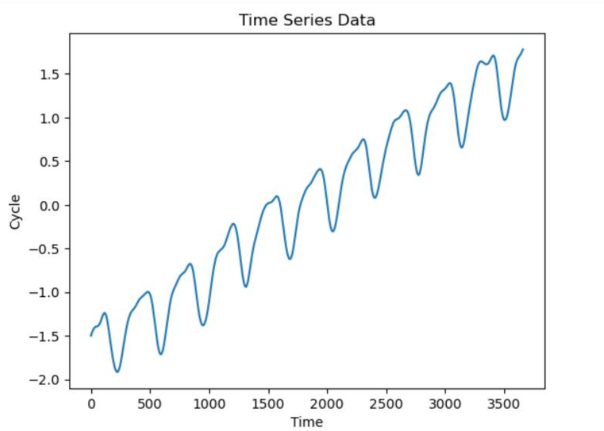
AEP DATASET:



SP_500:



ENERGY CONSUMPTION:



ANN (Artificial Neural Network):

Purpose:

To model complex and non-linear relationships in data that other models might not capture.

Methodology:

Used layers of neurons with activation functions, optimized using backpropagation (typically Adam).

Results:

Excels in datasets where relationships between inputs and outputs are non-linear.

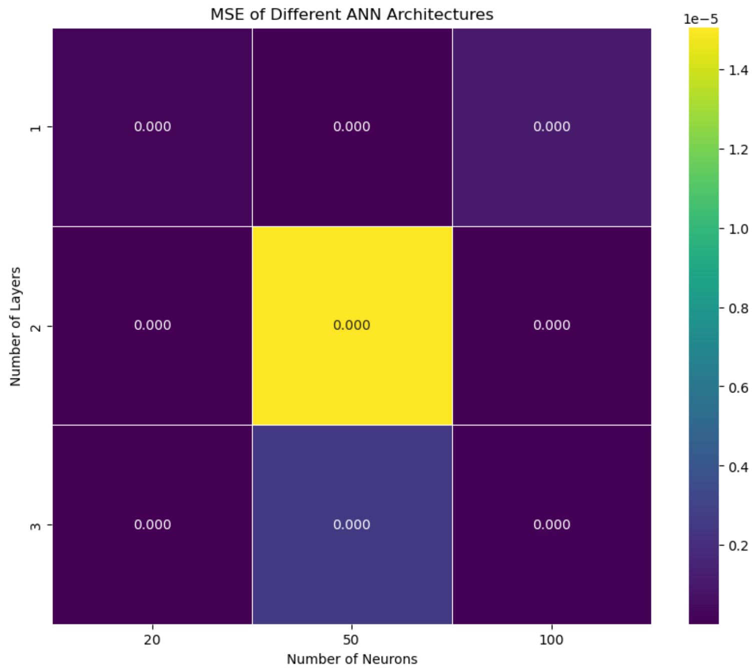
Evaluation:

Performance typically assessed through MSE or MAE on validation datasets.

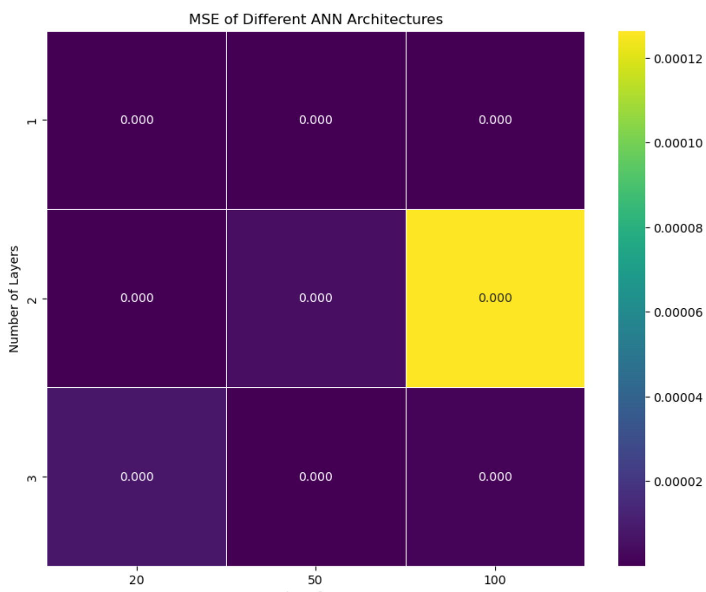
Visualization:

Comparison plots of actual vs. predicted values to assess model accuracy.

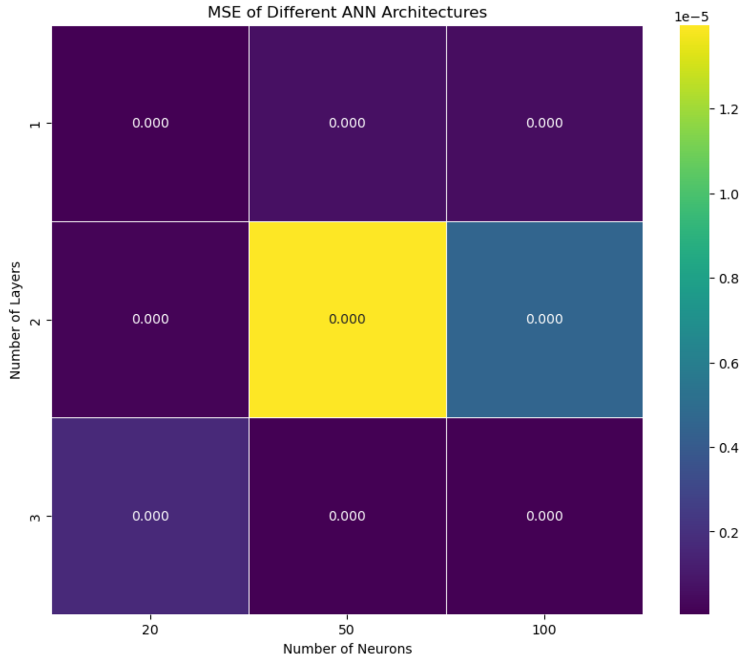
AEP DATASET:



SP_500 DATASET:



ENERGY CONSUMPTION:



SARIMA (Seasonal ARIMA):

Purpose:

Extends ARIMA for datasets with seasonality, providing forecasts that consider both seasonal and non-seasonal fluctuations.

Methodology:

Involves identifying seasonal differencing (D, P, Q, m) alongside non-seasonal components to model the data accurately.

Results:

Highly effective for datasets with clear seasonal patterns and trends.

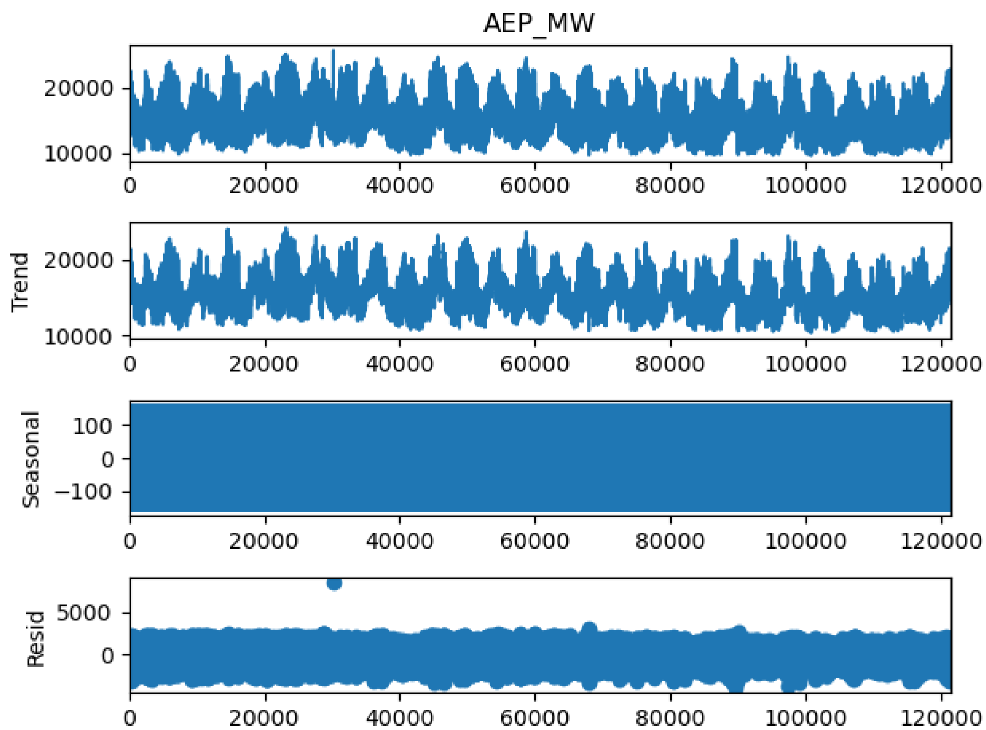
Evaluation:

Analyzed based on how well the seasonal components are captured, with MSE as a common metric.

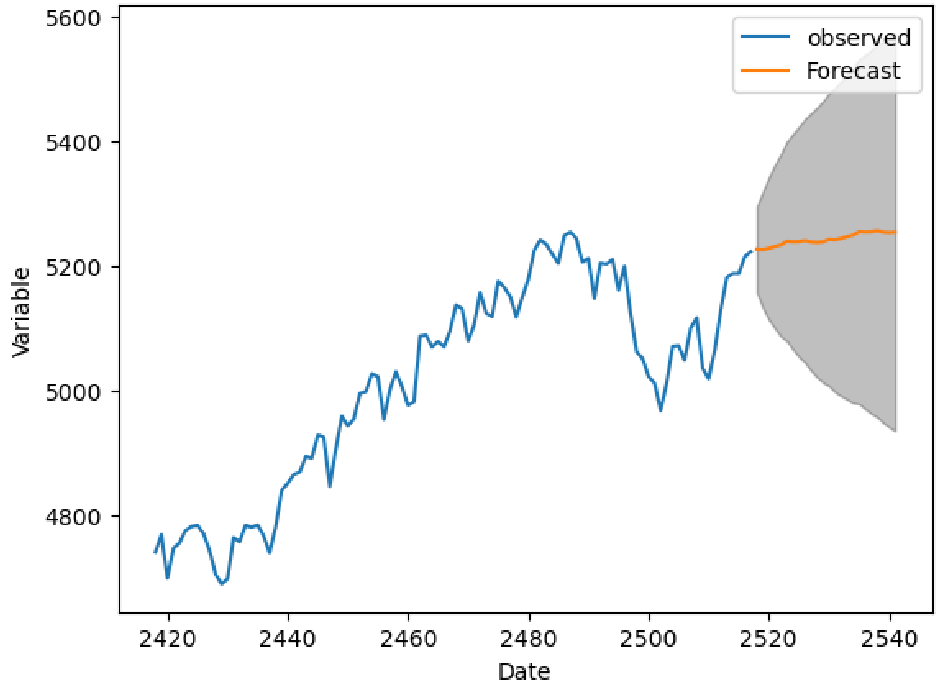
Visualization:

Plots showing seasonal adjustments and forecast against actual data.

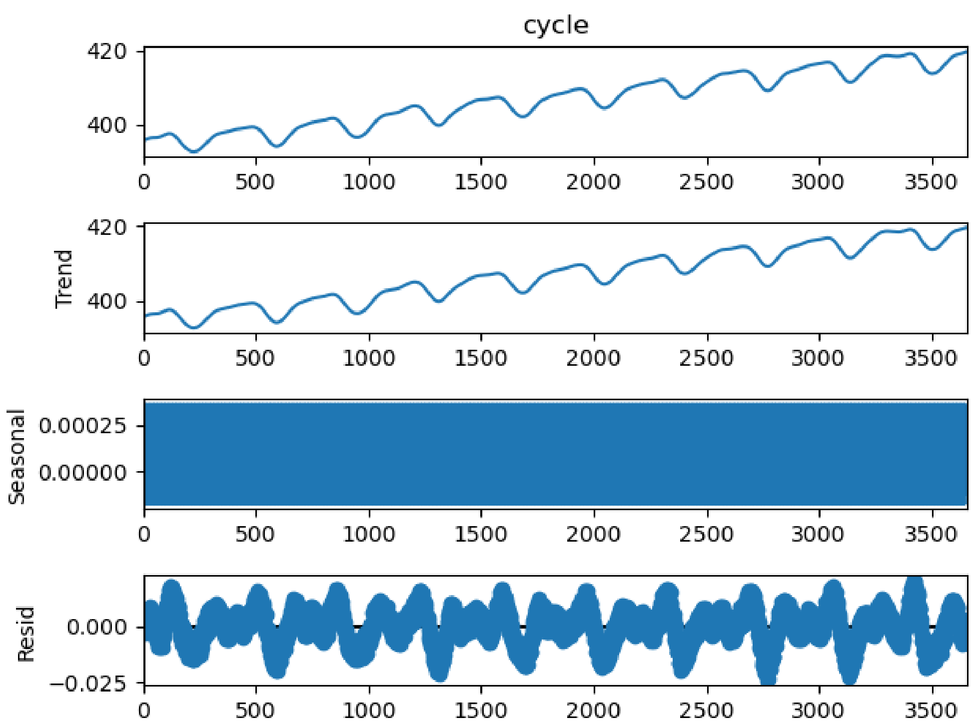
AEP DATASET:



SP_500:



ENERGY CONSUMPTION:



ETS (Exponential Smoothing State Space Model):

Purpose:

To forecast time series data by estimating trends and seasonal components with exponential smoothing.

Methodology:

Models' data with combinations of error, trend, and seasonal components in an additive or multiplicative manner.

Results:

Effective for data with trends and seasonalities, adapting well to changes in trend and seasonality over time.

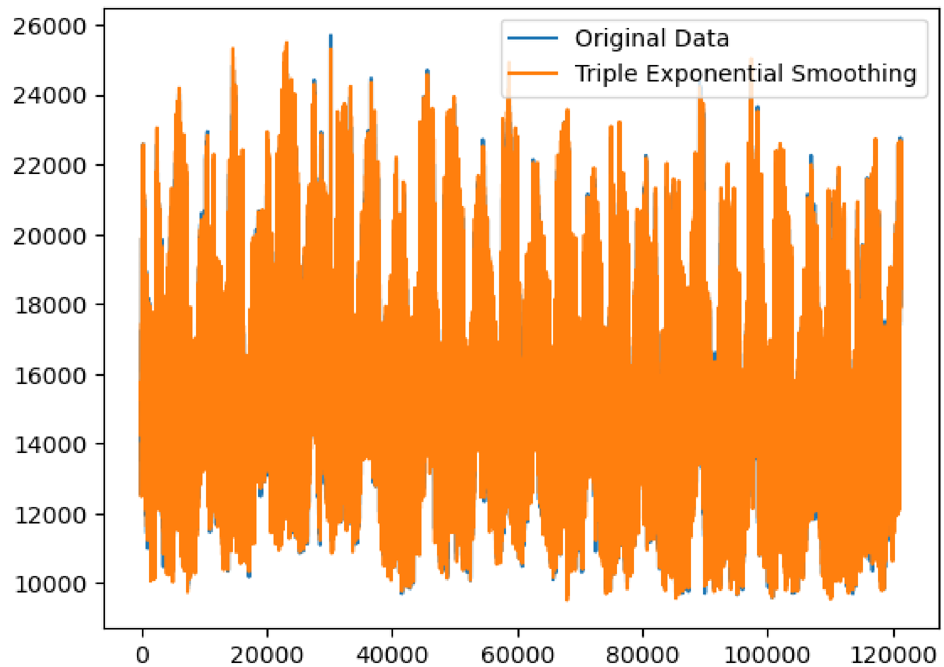
Evaluation:

Model fit assessed using AIC, BIC, and graphical residuals analysis.

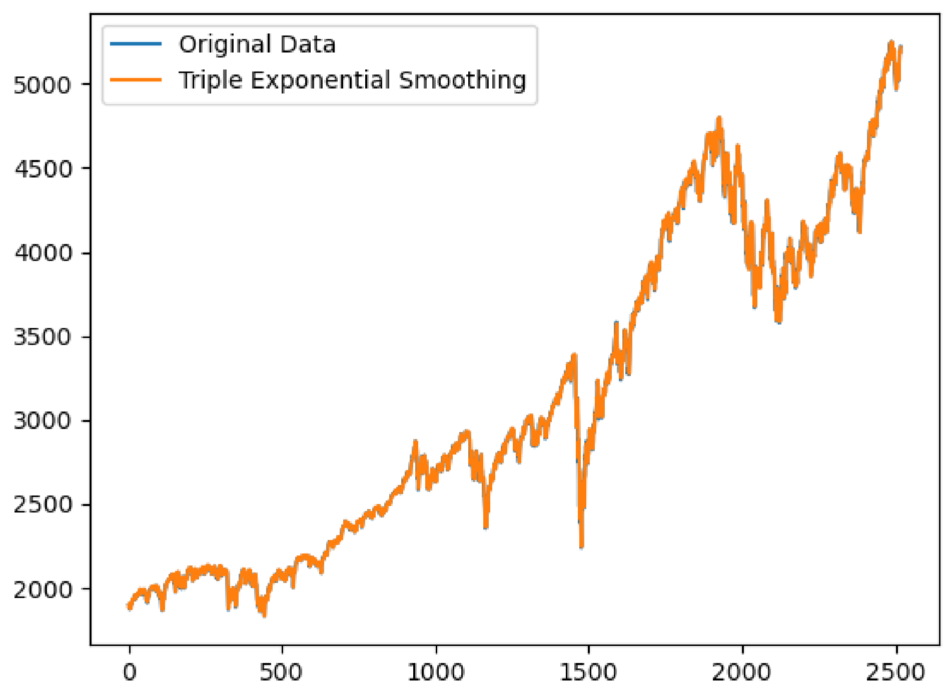
Visualization:

Time series plots showing the level, trend, and seasonal components of the model.

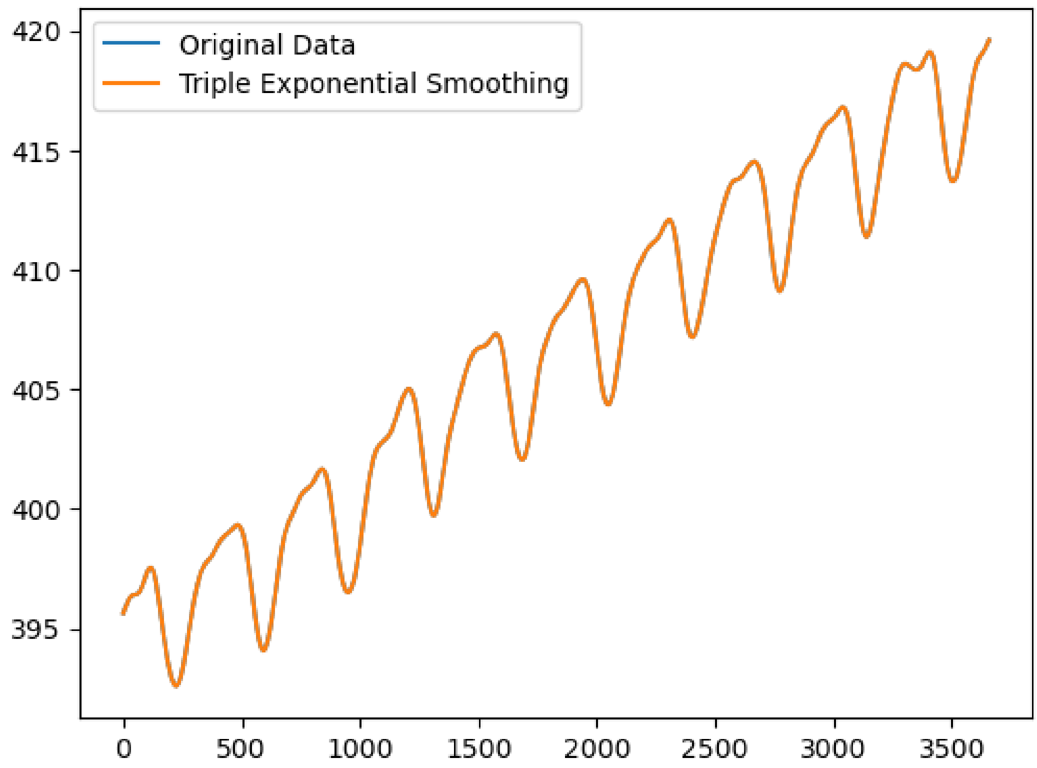
AEP DATASET:


Original Data
Triple Exponential Smoothing
26000
24000
22000
20000
18000
16000
14000
12000
10000
0
20000
40000
60000
80000
100000
120000

SP_500:


Original Data
Triple Exponential Smoothing
5000
4500
4000
3500
3000
2500
2000
0
500
1000
1500
2000
2500

ENERGY CONSUMPTION:



Prophet Report:

Purpose:

Designed for forecasting time series data with strong seasonal effects and historical holidays.

Methodology:

Utilizes a decomposable time series model with three main model components: trend, seasonality, and holidays.

Results:

Provides robust forecasts even with missing data and large outliers.

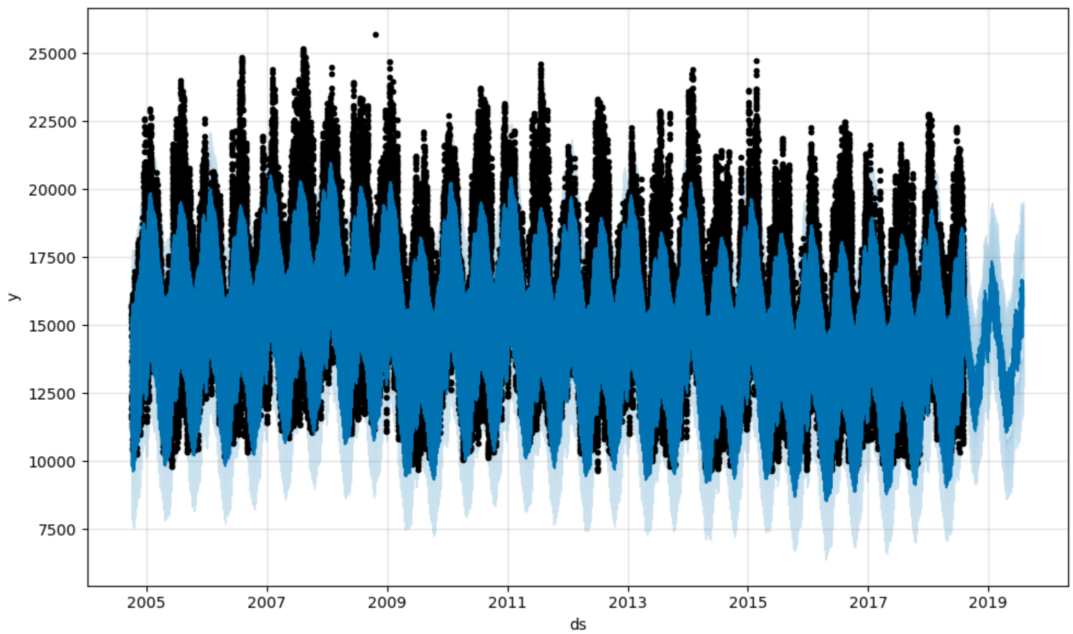
Evaluation:

Typically uses cross-validation to measure forecast error using metrics like MSE.

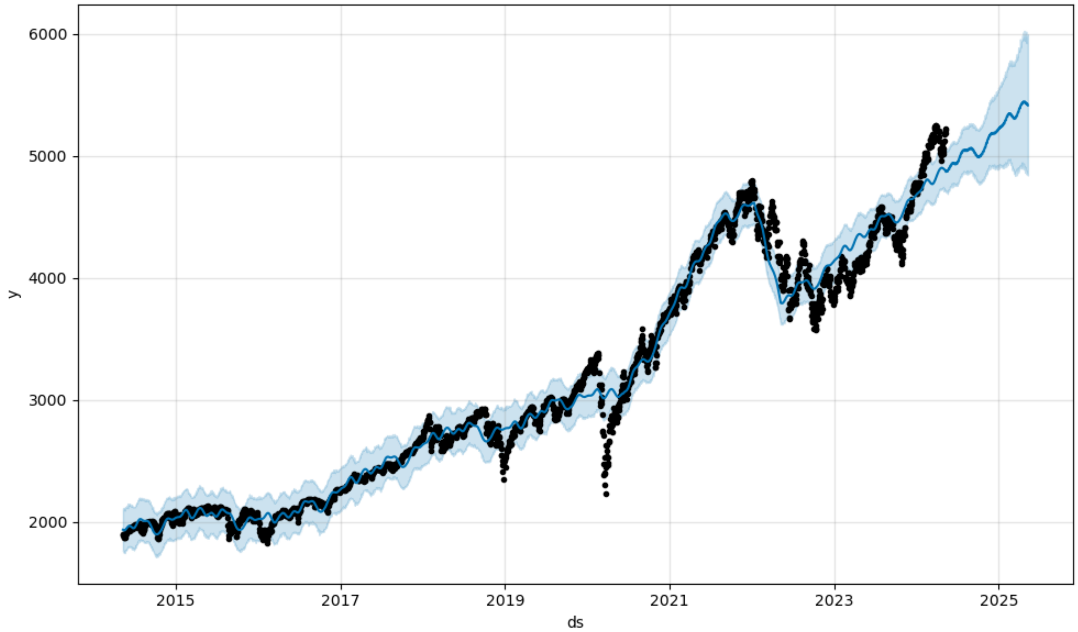
Visualization:

Component plots showing trends, yearly seasonality, and effects of holidays.

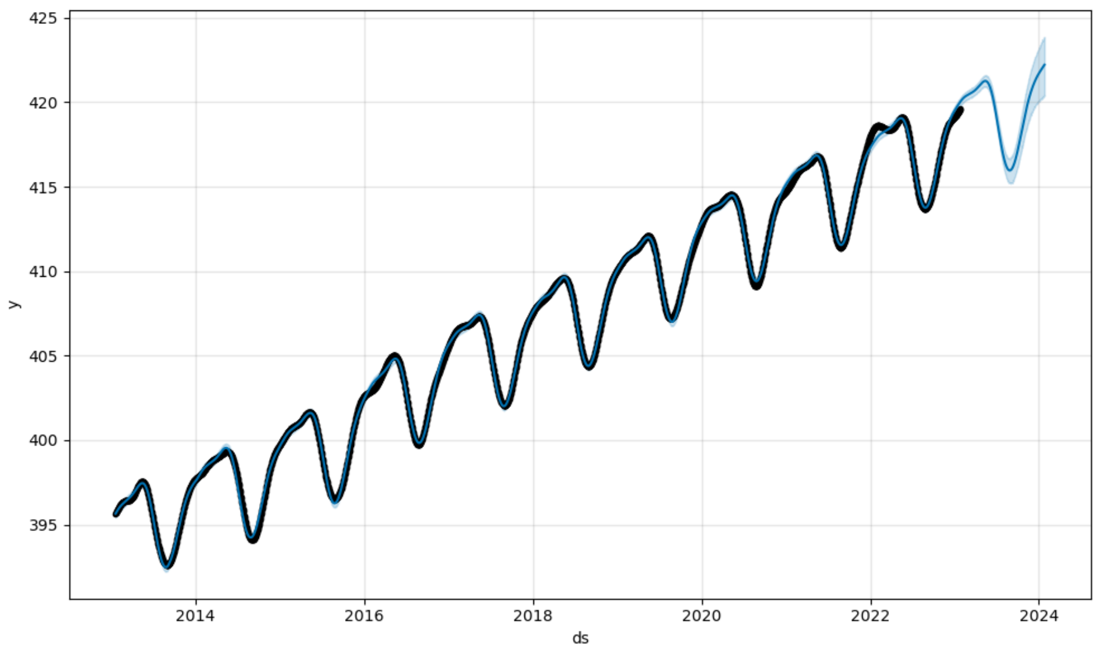
AEP DATASET:



SP_500:



ENERGY CONSUMPTION:



SVR (Support Vector Regression):

Purpose:

Applies SVM concepts to regression problems, focusing on fitting the error within a certain threshold.

Methodology:

Kernel function (linear, polynomial, RBF) used to fit data. Tuning of C (regularization) and gamma (kernel coefficient) is crucial.

Results:

Performs well with both linear and non-linear data, especially with appropriate kernel choice.

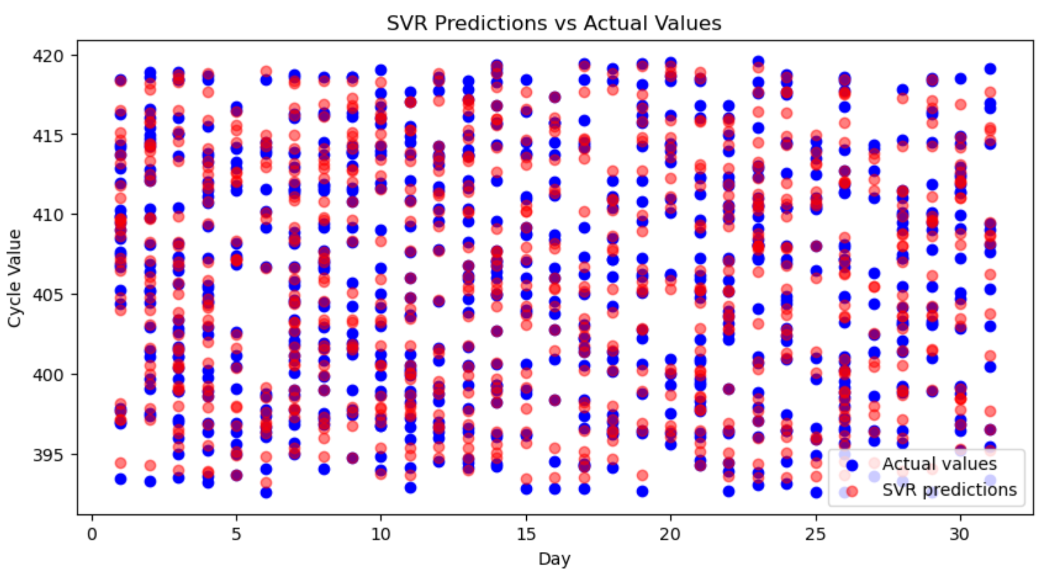
Evaluation:

Evaluated using MSE or MAE, and model selection often involves cross-validation.

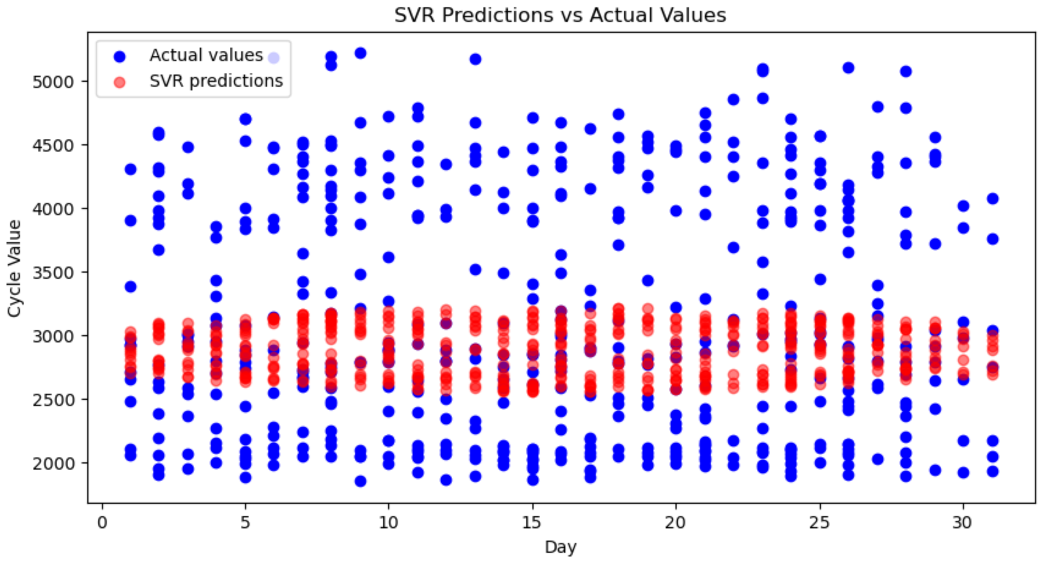
Visualization:

Scatter plots of actual vs. predicted values with support vector boundaries if feasible.

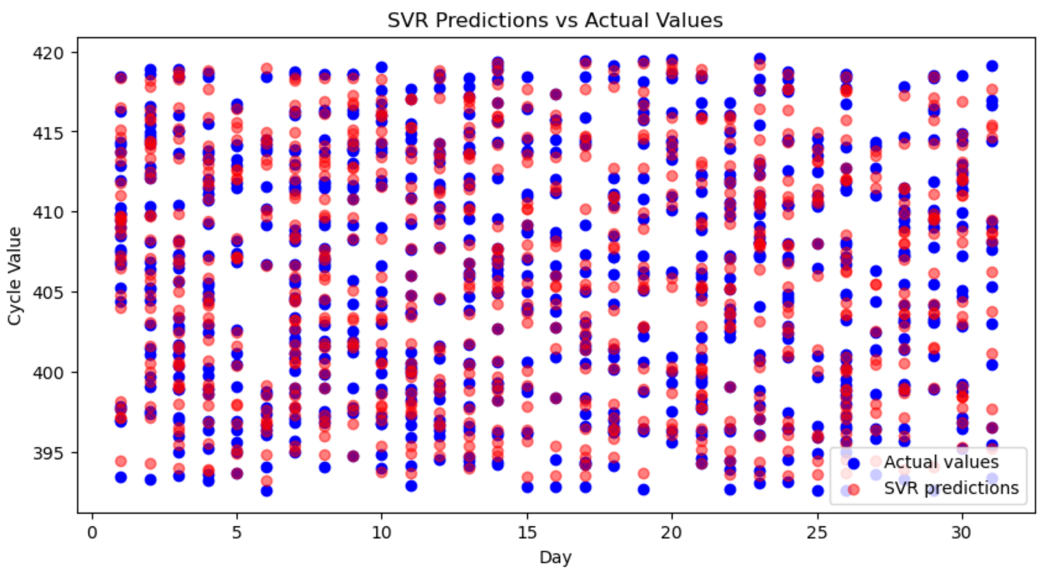
AEP DATASET:



SP_500 DATASET:



ENERGY CONSUMPTION DATASET:



LSTM (Long Short-Term Memory):

Purpose:

To capture long-term dependencies in sequence prediction problems, suitable for complex time series datasets.

Methodology:

Utilizes LSTM units in a RNN architecture, which helps in learning dependencies in data sequences that standard RNNs fail to capture.

Results:

Particularly effective for predictions where past information is crucial, like stock prices or weather conditions.

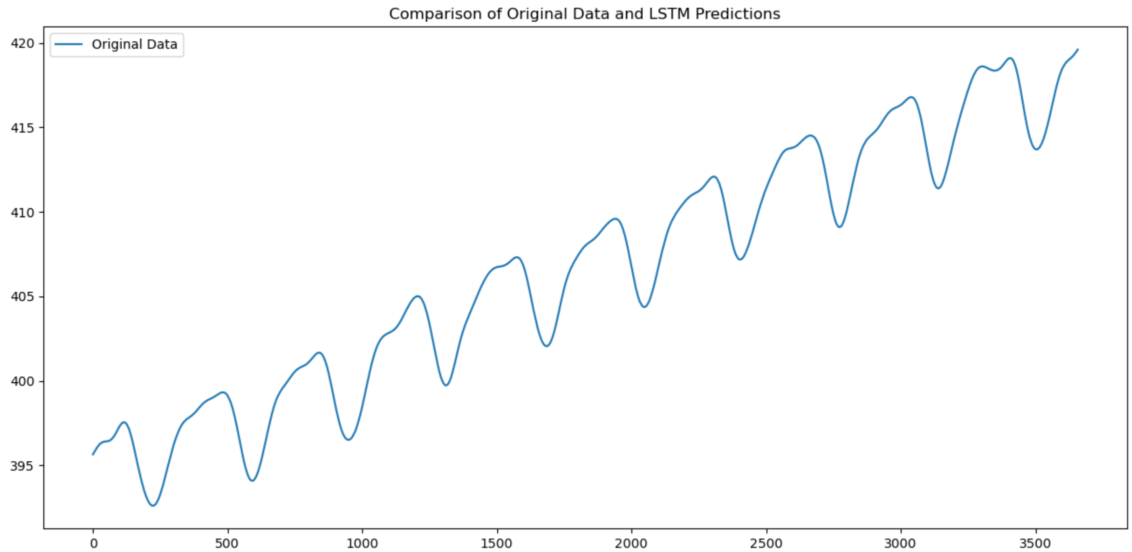
Evaluation:

Performance generally evaluated using sequence prediction accuracy metrics like MSE.

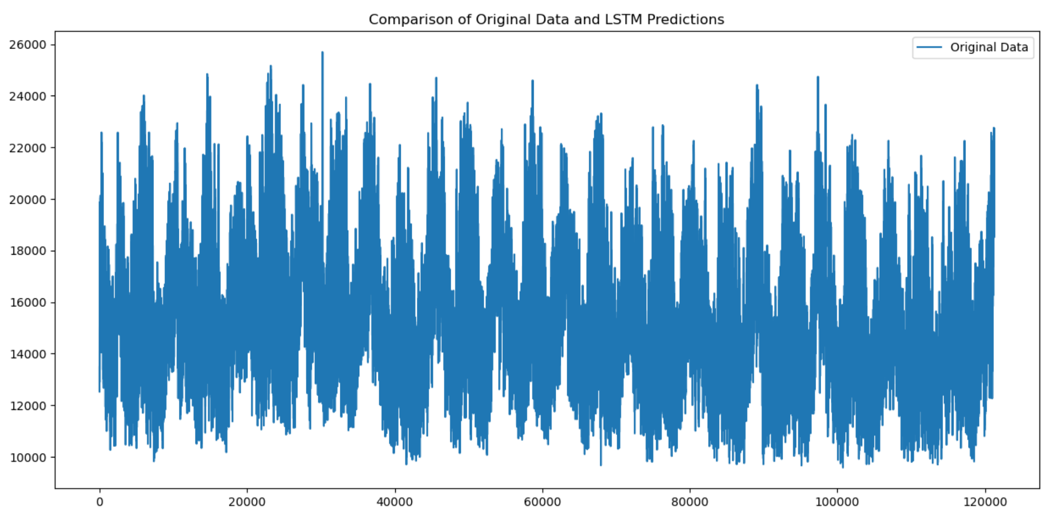
Visualization:

Sequence plots show predicted sequences against actual sequences, highlighting the long-term dependencies captured by the model.

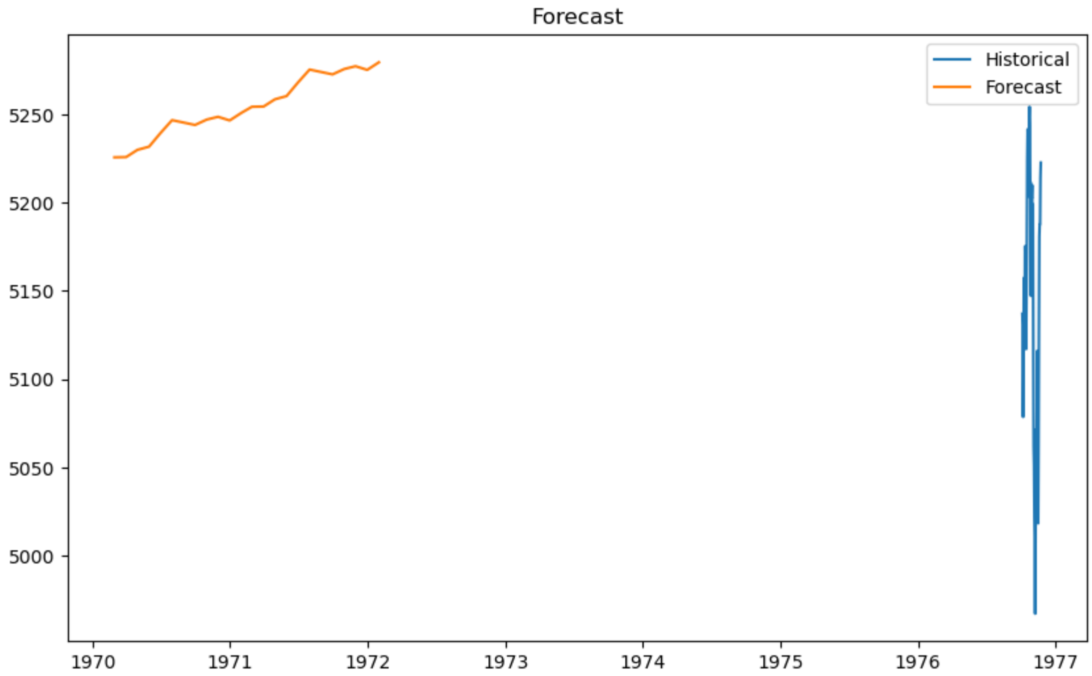
Energy Consumption:



AEP:



S&P500_Index:



Hybrid ARIMA-ANN Model:

Purpose:

Combines ARIMA's ability to model linear relationships and ANN's capability to capture non-linear patterns.

Methodology:

Uses ARIMA model predictions and ANN to model the residuals of the ARIMA predictions.

Results:

The hybrid approach generally provides more accurate predictions by leveraging strengths of both linear and non-linear models.

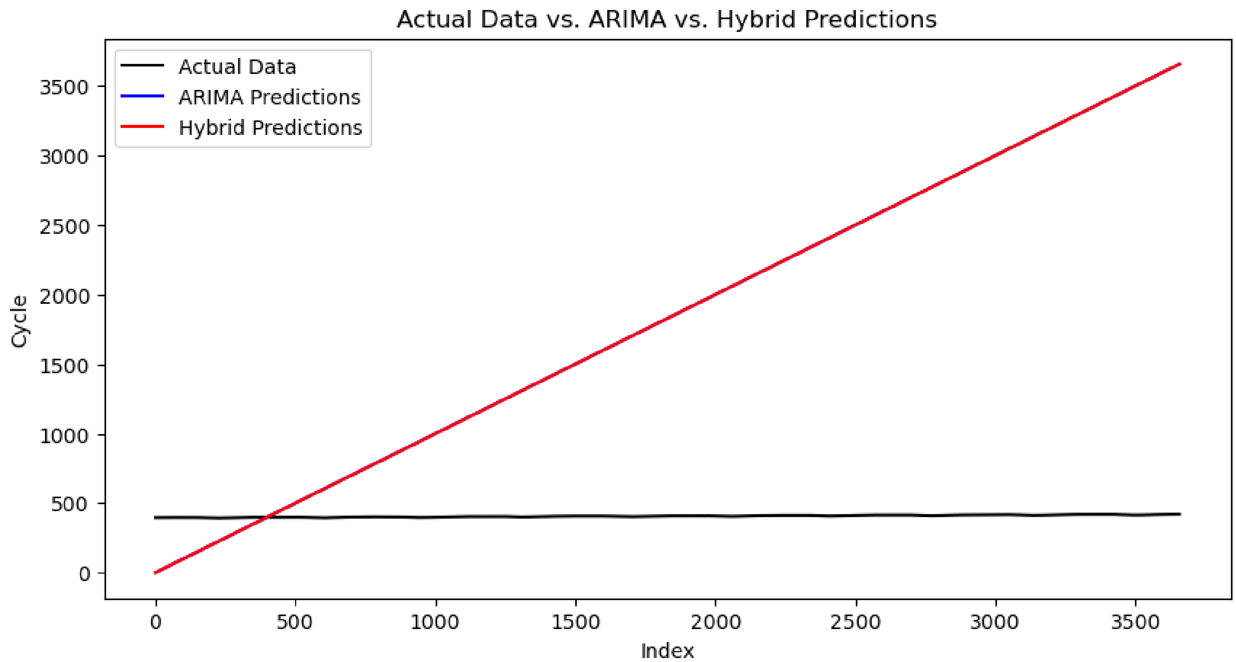
Evaluation:

Evaluated the accuracy of residual modeling and improvement over standalone models.

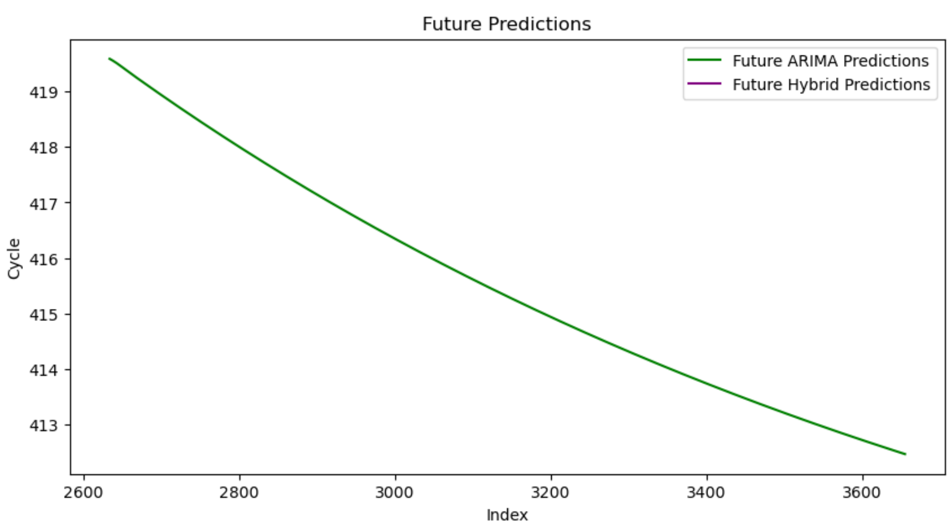
Visualization:

Overlaid plots of ARIMA predictions, ANN adjustments, and the final hybrid prediction to showcase the incremental improvement.

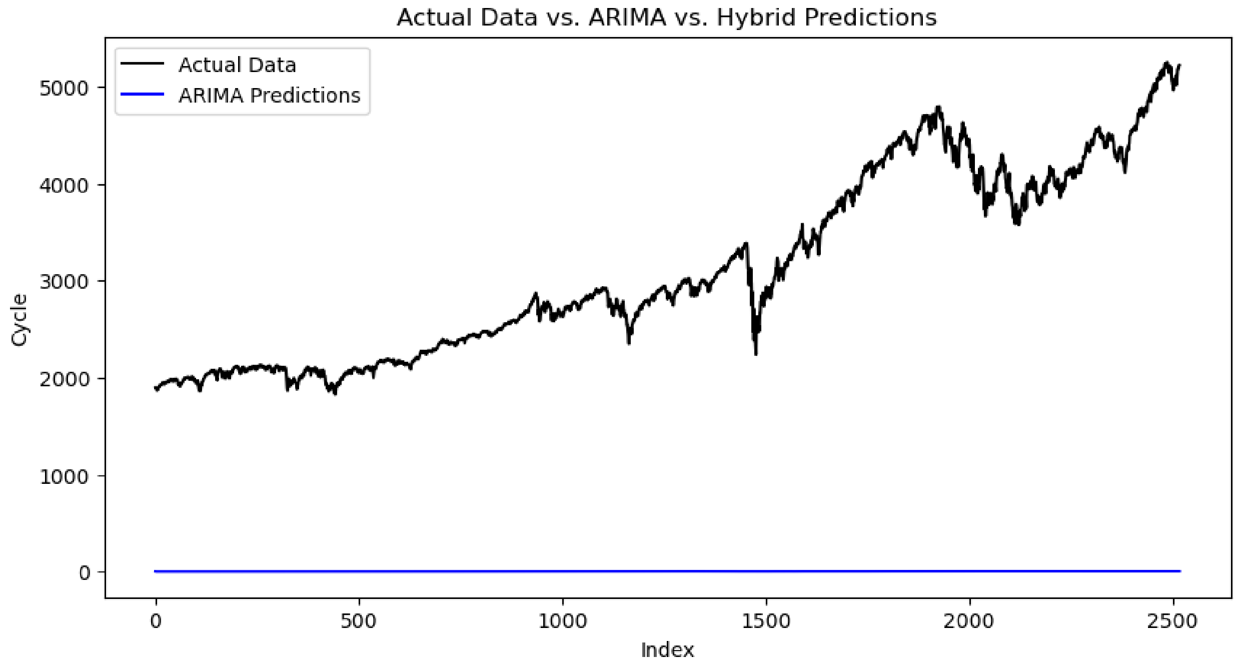
Energy Consumption:



AEP:



S&P500_Index:



Github Link:

https://github.com/Fateemaaa/Comprehensive-Forecasting-System-with-User-Interface-for-Multiple-Sectors.git