Malene Christiane Nielsen
mniel@mmmi.sdu.dk

Collected Experiment Report: DecisionTree - RandomForest - GradientBoosting .

18/12/2024

Introduction

This is an automated report for the Test on Tetouan City consumption data; the following models have been analyzed:

- DecisionTree
- RandomForest
- GradientBoosting

Experiment description:

First experiment on the dataset with Tetouan City dataset.

Model setup

The models have been used for the following forecast purposes:

- one_step
- multistep
- recursive

The models have been optimized using the following hyperparameters:

- max_depth: [2, 4, 6, 8, 10, 15, 20]
- criterion: ['squared_error']
- random_state: [42]
- min_samples_split: [5, 10, 50, 150, 200, 250]
- min_samples_leaf: [5, 10, 25, 50, 100]
- scaler: [None, StandardScaler(), MinMaxScaler(), RobustScaler(), PowerTransformer()]

And with the following search algorithms:

-

Malene Christiane Nielsen
mniel@mmmi.sdu.dk

grid

- random

The used performance measure is the r2 measure.

Dataset setup

The baseline dataset used for these forecasts is the 'Power consumption of Tetouan City between 2017-01-01 and 2017-12-31 in 10 minutes resolution.' dataset: *'Power consumption of Tetouan City between 2017-01-01 and 2017-12-31 in 10 minutes resolution. Has the weather data for temperature, humidity, wind speed and diffuse flows, as well as the consumption data for 3 different zones in the city. For simplicity, we will only use one of the zones.'.*

The test size used for the forecasts is 0.2.

- Dataset 1

  - name: univariate_temporal
  - dataset_type: univariate
  - prediction_type: one_step
  - components: ['one_step_target', 'temporal_features']

- Dataset 2

  - name: univariate_lagged
  - dataset_type: univariate
  - prediction_type: one_step
  - components: ['one_step_target', 'lagged_target']

- Dataset 3

  - name: univariate_temporal_and_lagged
  - dataset_type: univariate
  - prediction_type: one_step
  - components: ['one_step_target', 'temporal_features', 'lagged_target']

- Dataset 4

  - name: univariate_multistep_temporal

Malene Christiane Nielsen
mniel@mmmi.sdu.dk

- dataset_type: univariate
- prediction_type: multistep
- components: ['multistep_target', 'temporal_features']

○ Dataset 5

- name: univariate_multistep_lagged
- dataset_type: univariate
- prediction_type: multistep
- components: ['multistep_target', 'lagged_target']

○ Dataset 6

- name: univariate_multistep_temporal_and_lagged
- dataset_type: univariate
- prediction_type: multistep
- components: ['multistep_target', 'temporal_features', 'lagged_target']

○ Dataset 7

- name: multivariate_temporal
- dataset_type: multivariate
- prediction_type: one_step
- components: ['one_step_target', 'feature_columns', 'temporal_features']

○ Dataset 8

- name: multivariate_lagged
- dataset_type: multivariate
- prediction_type: one_step
- components: ['one_step_target', 'feature_columns', 'lagged_target']

○ Dataset 9

- name: multivariate_temporal_and_lagged

Malene Christiane Nielsen
mniel@mmmi.sdu.dk

- dataset_type: multivariate
- prediction_type: one_step
- components: ['one_step_target', 'feature_columns', 'temporal_features', 'lagged_target']

○ Dataset 10

- name: multivariate_multistep_temporal
- dataset_type: multivariate
- prediction_type: multistep
- components: ['multistep_target', 'feature_columns', 'temporal_features']

○ Dataset 11

- name: multivariate_multistep_lagged
- dataset_type: multivariate
- prediction_type: multistep
- components: ['multistep_target', 'feature_columns', 'lagged_target']

○ Dataset 12

- name: multivariate_multistep_temporal_and_lagged
- dataset_type: multivariate
- prediction_type: multistep
- components: ['multistep_target', 'feature_columns', 'temporal_features', 'lagged_target']

Results

For the models; DecisionTree, RandomForest, GradientBoosting, the following models and datasets yielded the best results.

SDU ENERGY INFORMATICS

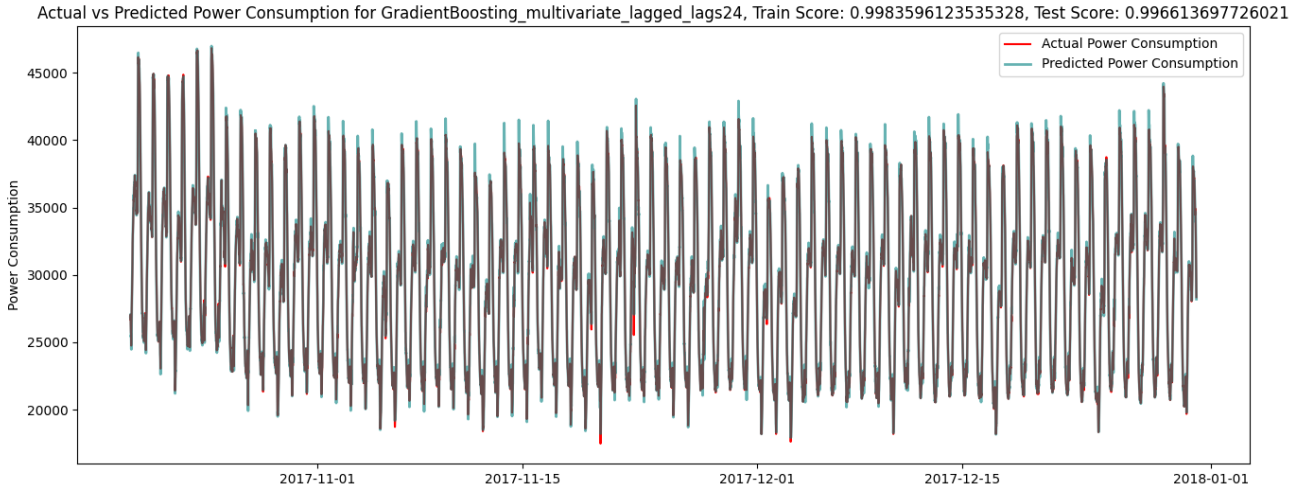
Malene Christiane Nielsen
mniel@mmmi.sdu.dk

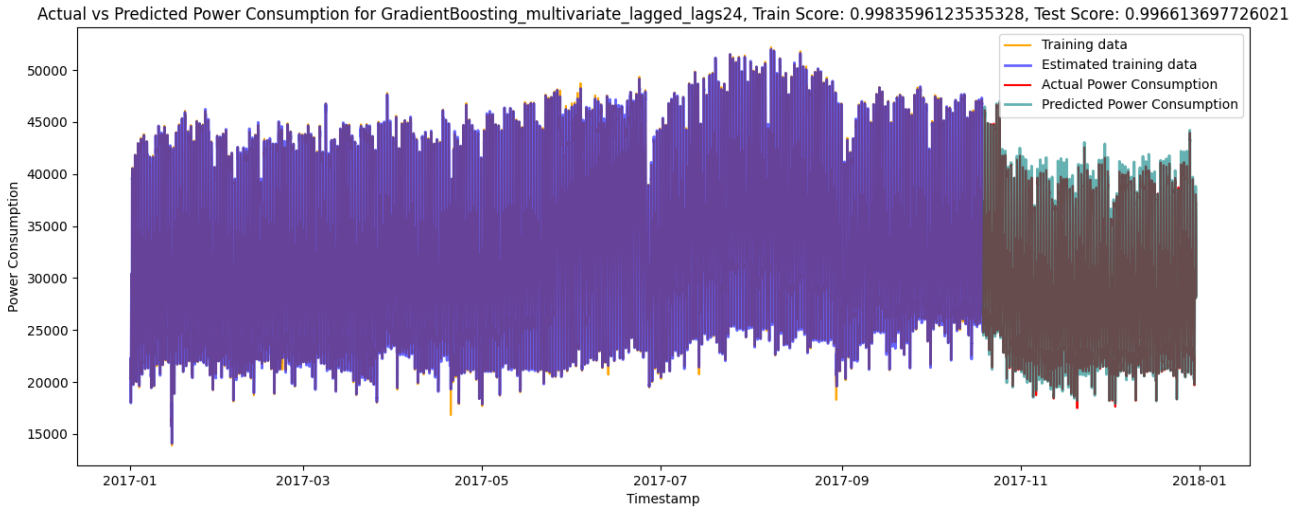
The best model for one_step forecasting.

The best model for one_step forecasting is the GradientBoosting model.
The model has been trained on the multivariate_lagged_lags24 dataset.

The best score for the one_step forecasting is 0.996613697726021.

Best GradientBoosting forecast over time





SDU ENERGY INFORMATICS

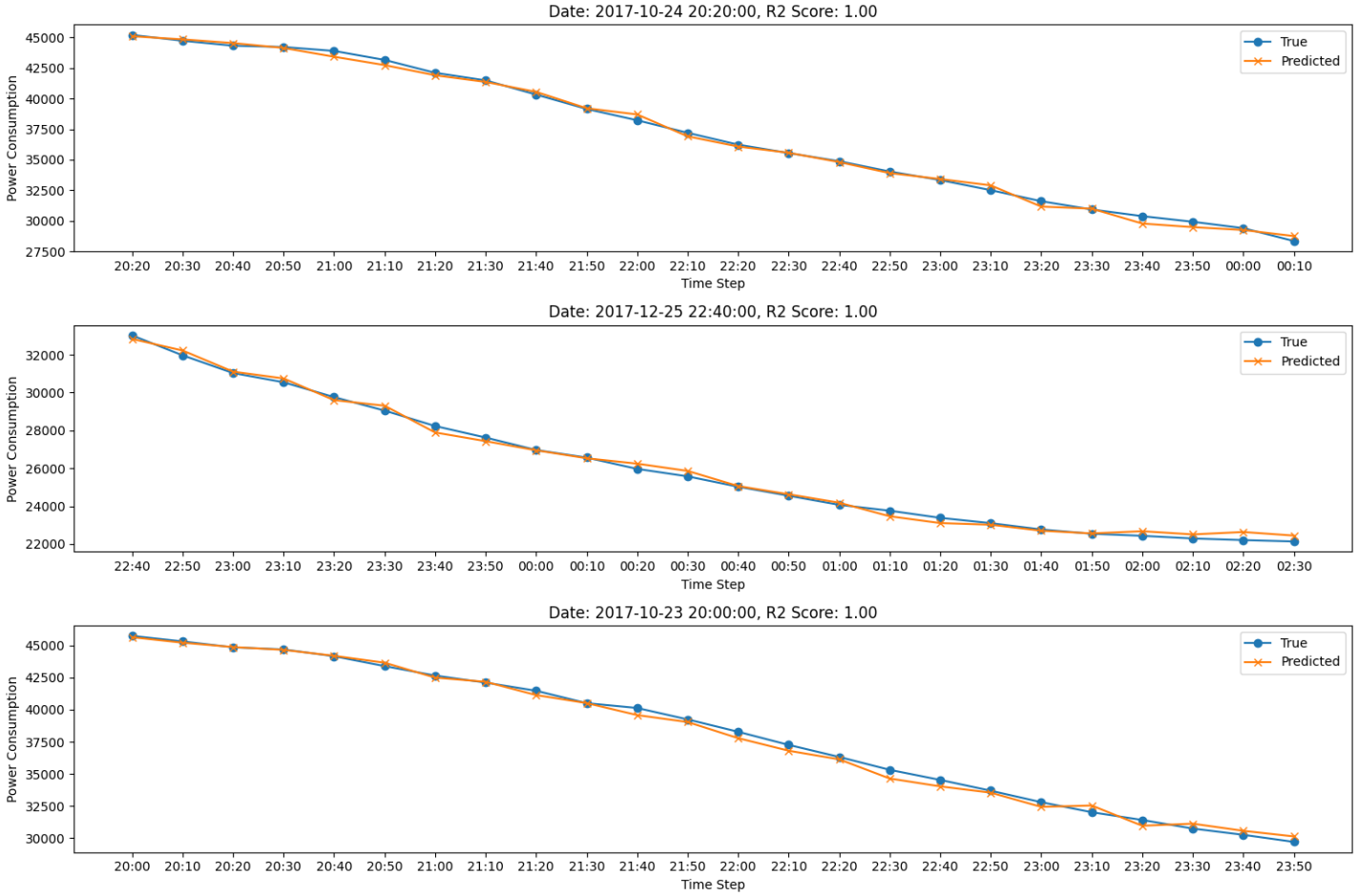
Malene Christiane Nielsen
mniel@mmmi.sdu.dk

The best model for multistep forecasting.

The best model for multistep forecasting is the GradientBoosting model.
The model has been trained on the multivariate_multistep_temporal_and_lagged_lags24_steps24 dataset.

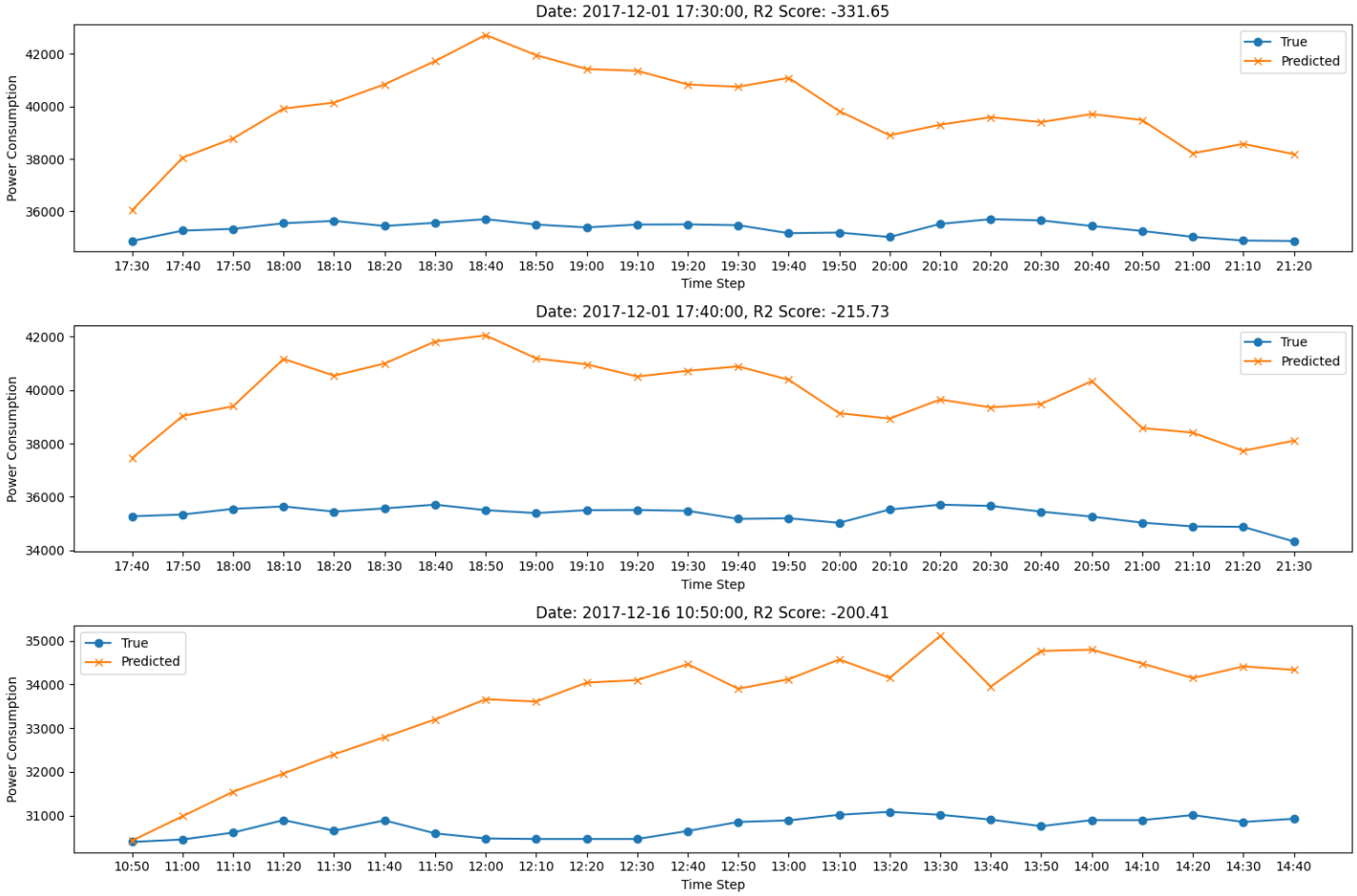
The best score for the multistep forecasting is 0.9052468783883622.

Best predicted days for GradientBoosting.



Malene Christiane Nielsen
mniel@mmmi.sdu.dk

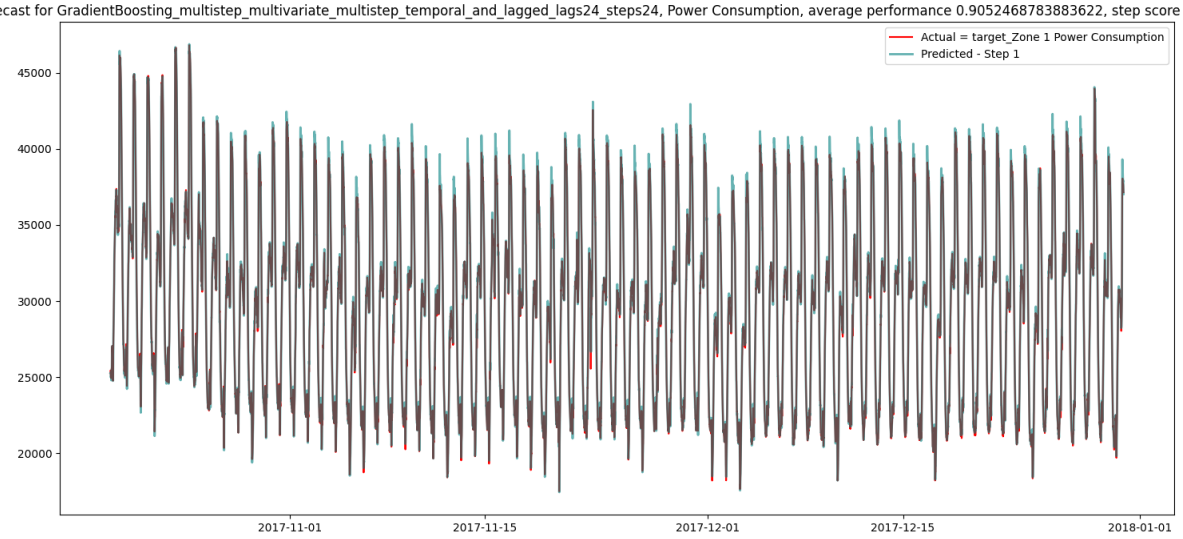
Worst predicted days for GradientBoosting.



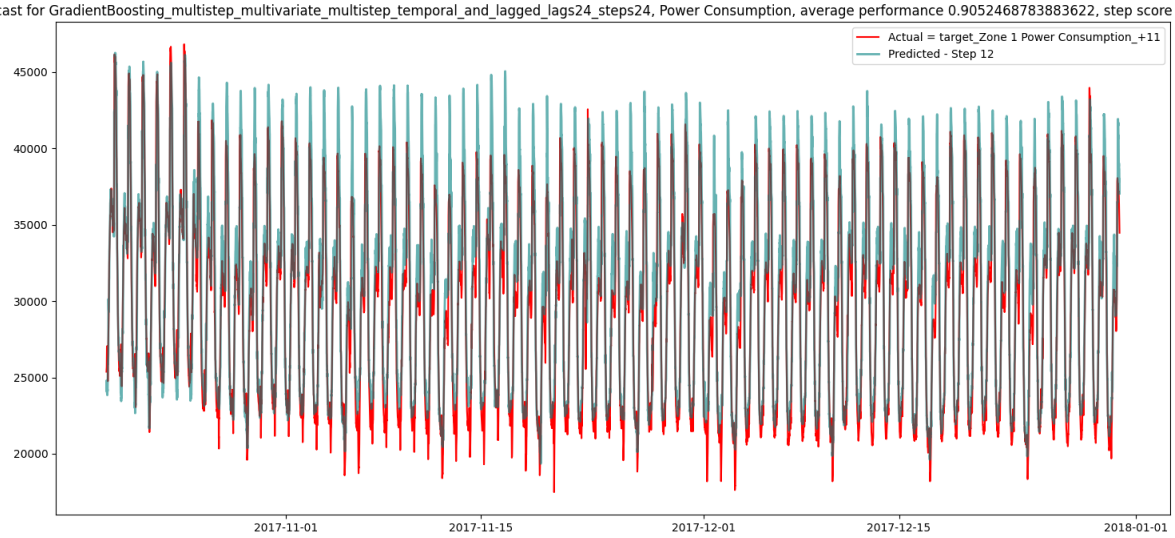
Malene Christiane Nielsen
mniel@mmmi.sdu.dk

Steps plots for GradientBoosting forecasts over time

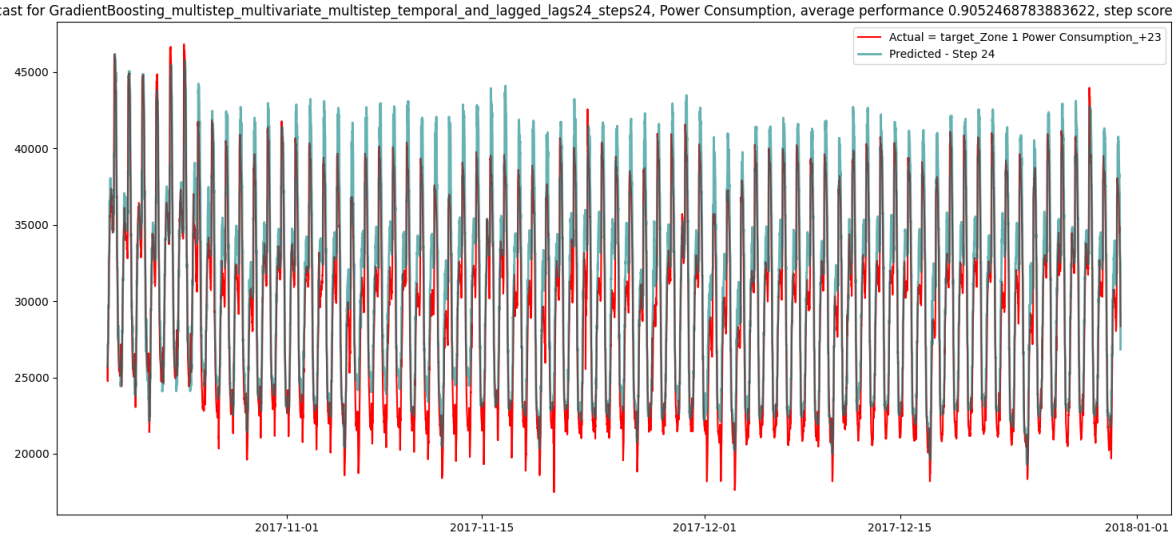
Multistep forecast for GradientBoosting_multistep_multivariate_multistep_temporal_and_lagged_lags24_steps24, Power Consumption, average performance 0.9052468783883622, step score 0.99643247609192



Multistep forecast for GradientBoosting_multistep_multivariate_multistep_temporal_and_lagged_lags24_steps24, Power Consumption, average performance 0.9052468783883622, step score 0.887086196756694



Multistep forecast for GradientBoosting_multistep_multivariate_multistep_temporal_and_lagged_lags24_steps24, Power Consumption, average performance 0.9052468783883622, step score 0.854576682626919



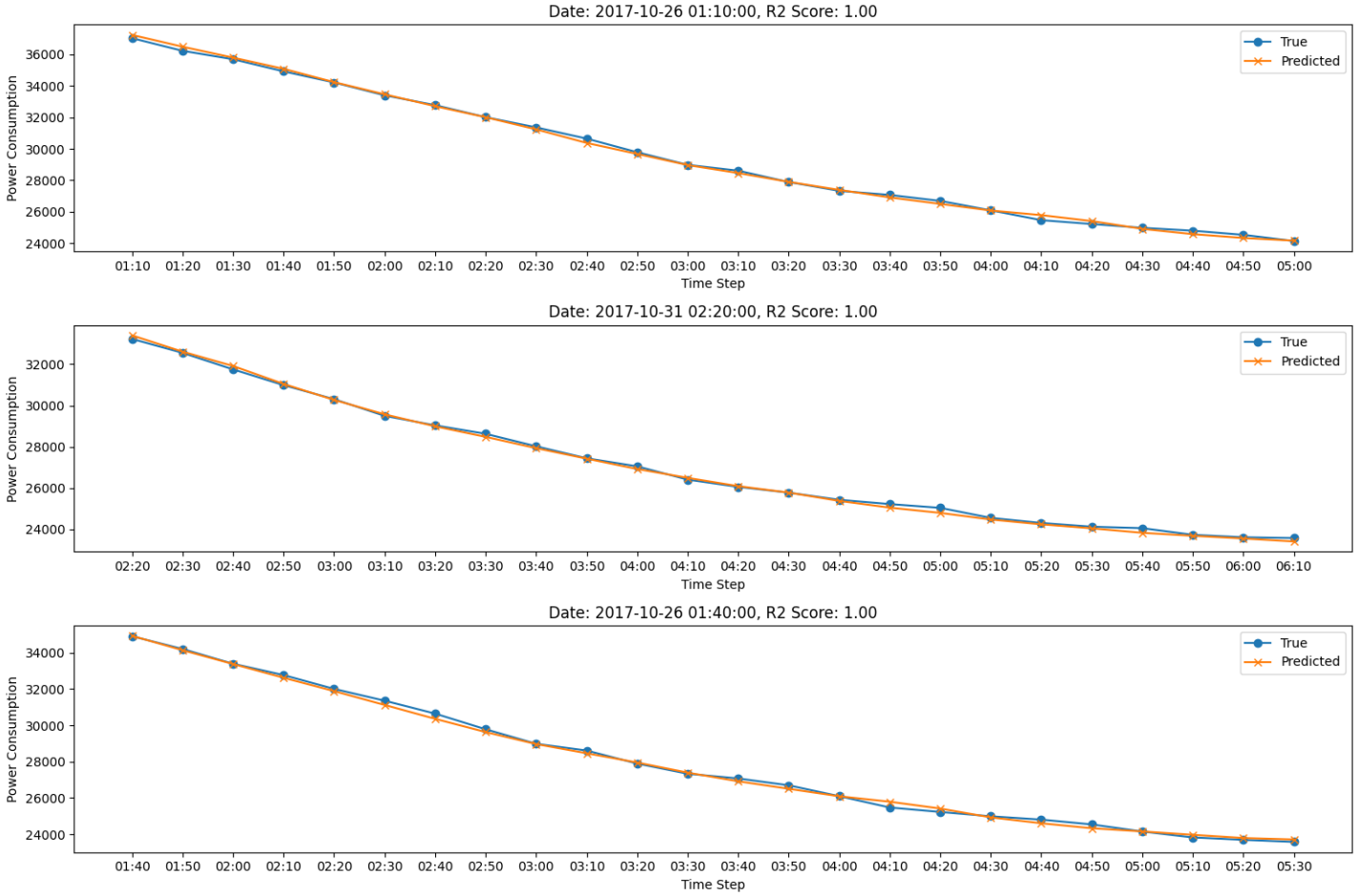
Malene Christiane Nielsen
mniel@mmmi.sdu.dk

The best model for recursive forecasting.

The best model for recursive forecasting is the GradientBoosting model.
The model has been trained on the multivariate_temporal_and_lagged_lags24 dataset.

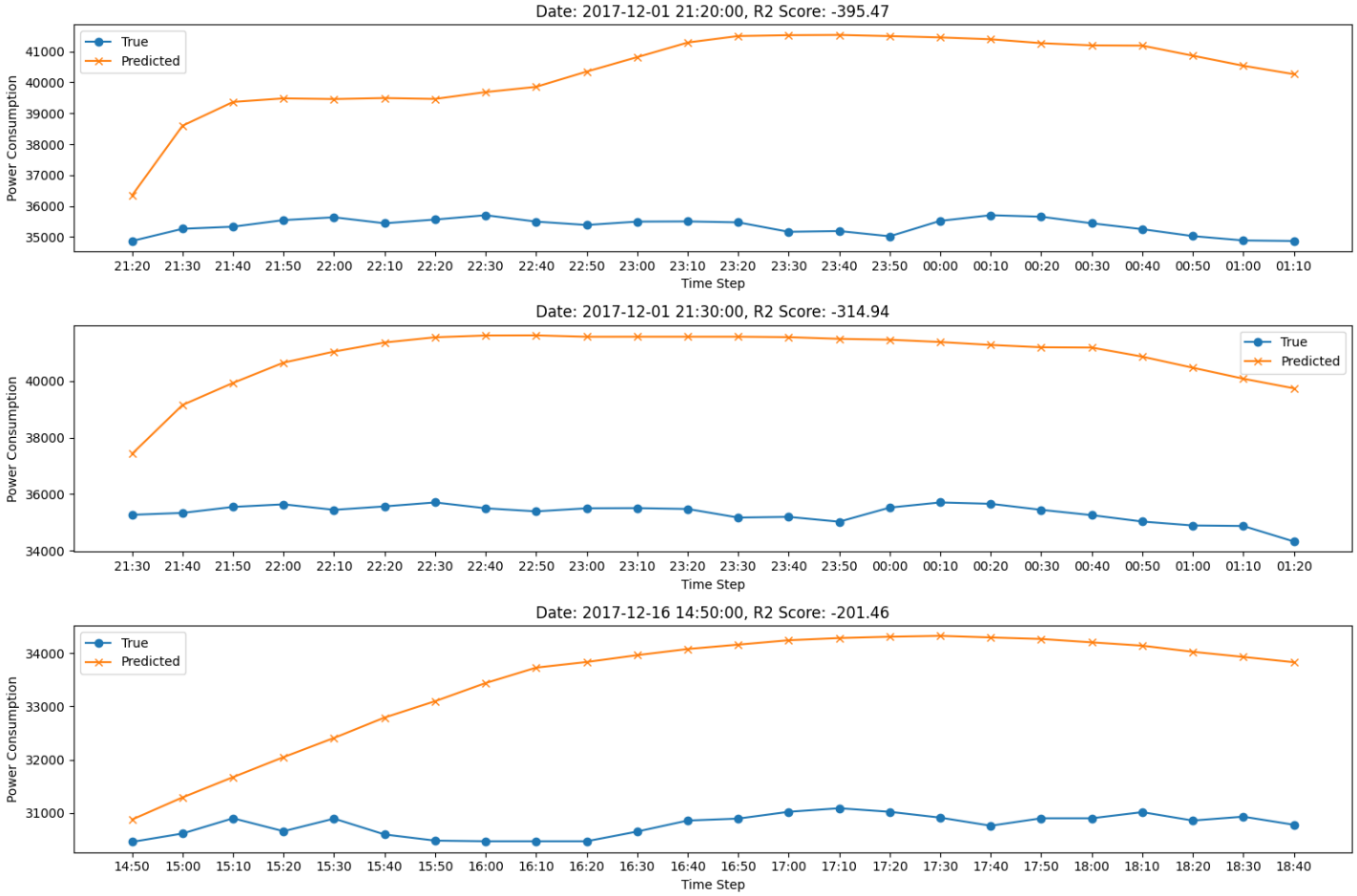
The best score for the recursive forecasting is 0.8559304262069055.

Best predicted days for GradientBoosting.



Malene Christiane Nielsen
mniel@mmmi.sdu.dk

Worst predicted days for GradientBoosting.



Malene Christiane Nielsen
mniel@mmmi.sdu.dk

Steps plots for GradientBoosting forecasts over time