SDU ENERGY INFORMATICS

Malene Christiane Nielsen
mniel@mmmi.sdu.dk

Collected Experiment Report: DecisionTree - RandomForest - GradientBoosting .

15/02/2025

Introduction

This is an automated report for the Experiment with tree models on traffic volume dataset; the following models have been analyzed:

- DecisionTree
- RandomForest
- GradientBoosting

Experiment description:

Experiment with tree models on traffic volume 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 'Metro Interstate Traffic Volume with hourly features and holiday markings.' dataset: *'Metro Interstate Traffic Volume with hourly features and holiday markings.'.*

The test size used for the forecasts is 0.2.

- Dataset 1

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

- Dataset 2

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

- Dataset 3

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

- Dataset 4

  - name: multivariate_lagged_temporal
  - dataset_type: multivariate

Malene Christiane Nielsen
mniel@mmmi.sdu.dk

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

○ Dataset 5

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

○ Dataset 6

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

Results

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

SDU ENERGY INFORMATICS

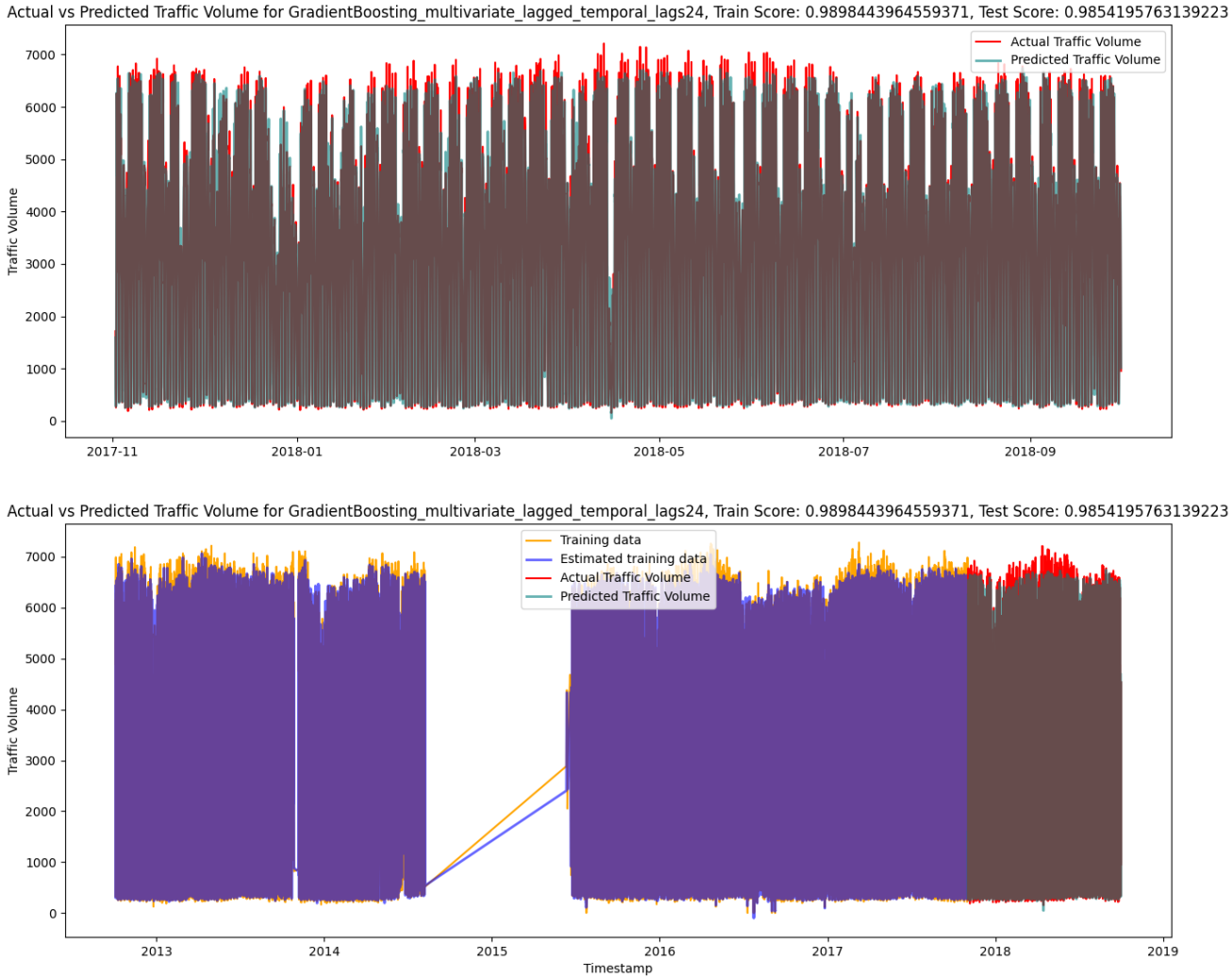
Malene Christiane Nielsen
mniel@mmmi.sdu.dk

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_temporal_lags24 dataset.

The best score for the one_step forecasting is 0.9854195763139223.

Best GradientBoosting forecast over time



SDU ENERGY INFORMATICS

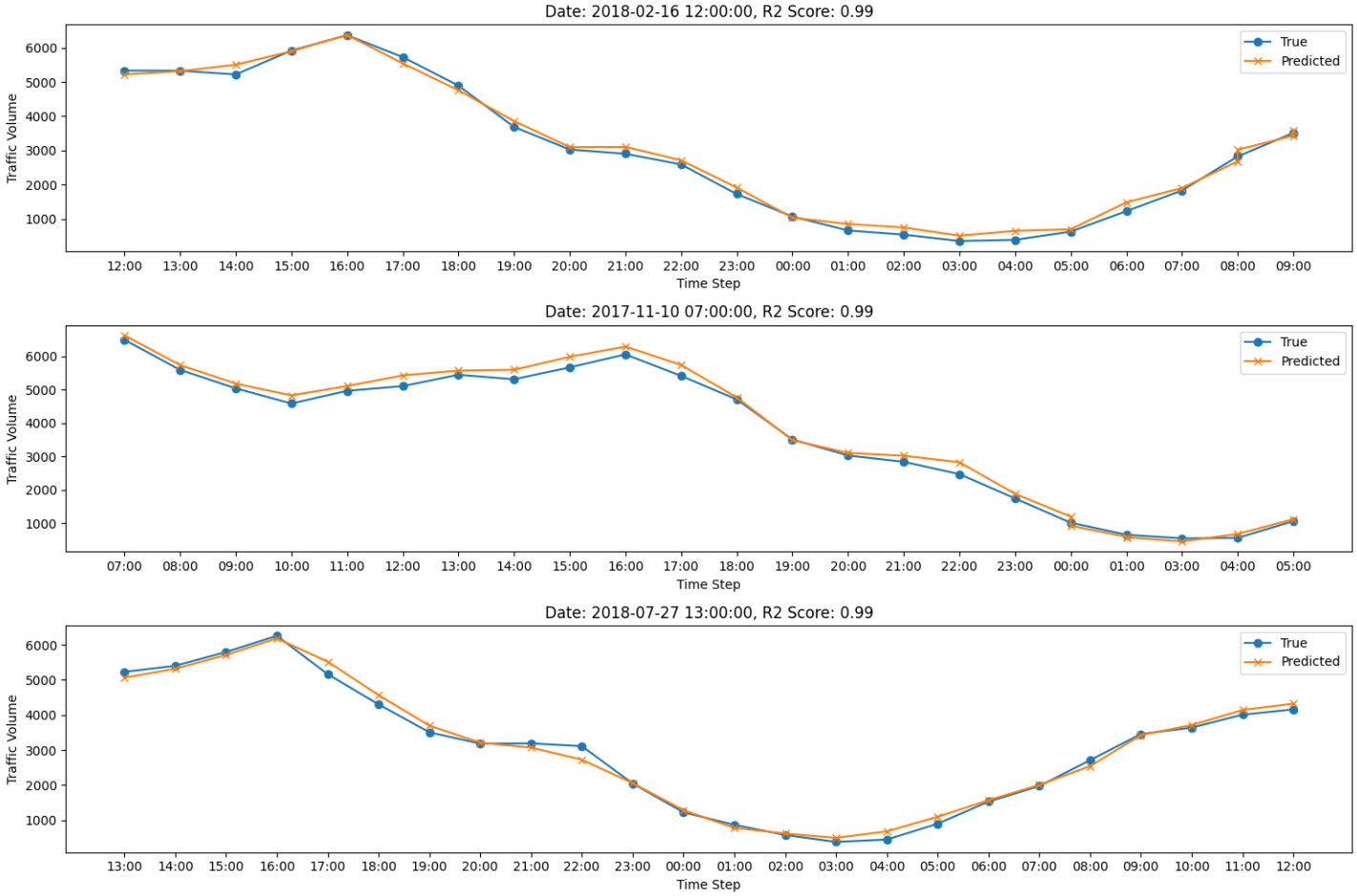
Malene Christiane Nielsen
mniel@mmmi.sdu.dk

The best model for multistep forecasting.

The best model for multistep forecasting is the GradientBoosting model.
The model has been trained on the multivariate_lagged_temporal_multistep_lags24_steps24 dataset.

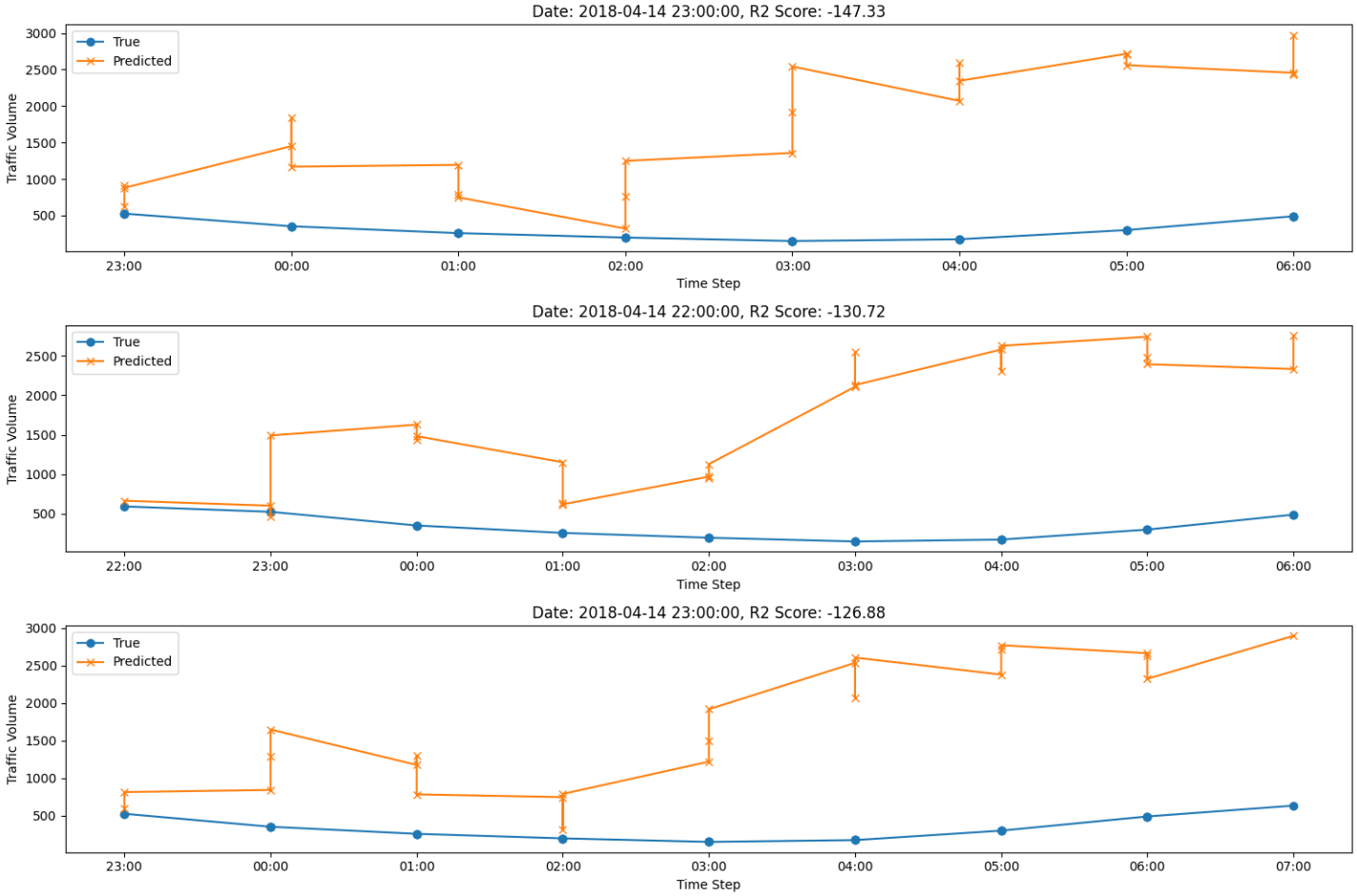
The best score for the multistep forecasting is 0.7093092946852239.

Best predicted days for GradientBoosting.



Malene Christiane Nielsen
mniel@mmmi.sdu.dk

Worst predicted days for GradientBoosting.

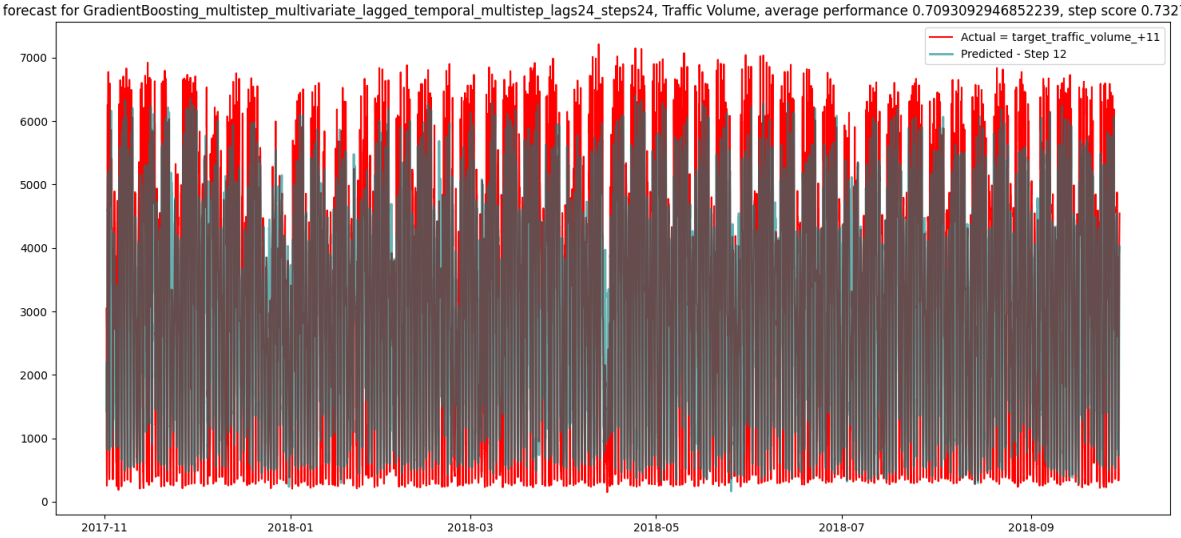


Malene Christiane Nielsen
mniel@mmmi.sdu.dk

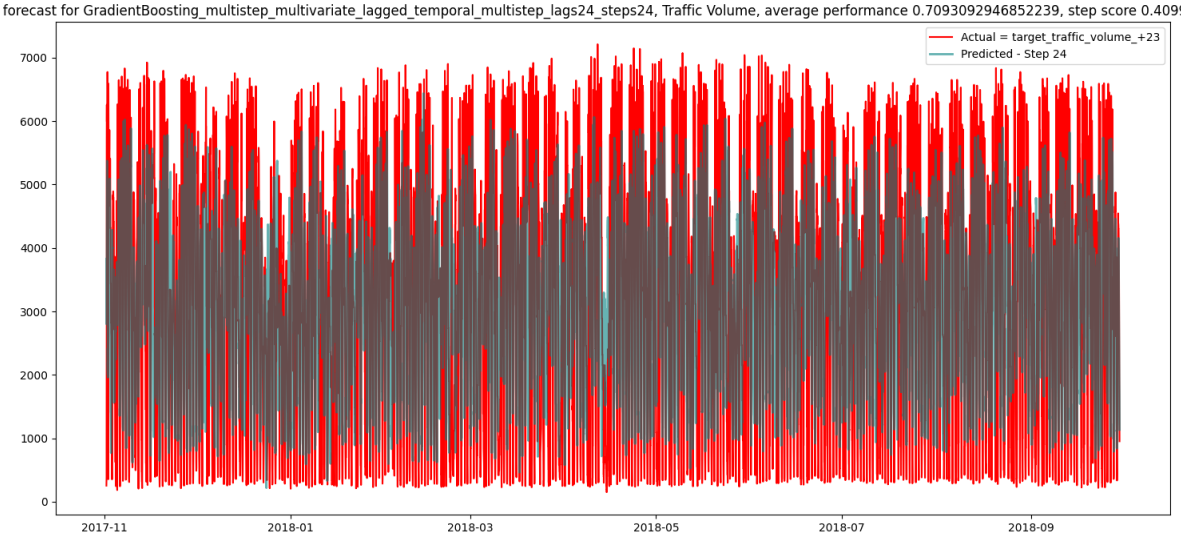
Steps plots for GradientBoosting forecasts over time

Multistep forecast for GradientBoosting_multistep_multivariate_lagged_temporal_multistep_lags24_steps24, Traffic Volume, average performance 0.7093092946852239, step score 0.9847141856002299

Multistep forecast for GradientBoosting_multistep_multivariate_lagged_temporal_multistep_lags24_steps24, Traffic Volume, average performance 0.7093092946852239, step score 0.7327763363378932



Multistep forecast for GradientBoosting_multistep_multivariate_lagged_temporal_multistep_lags24_steps24, Traffic Volume, average performance 0.7093092946852239, step score 0.4099227007759745



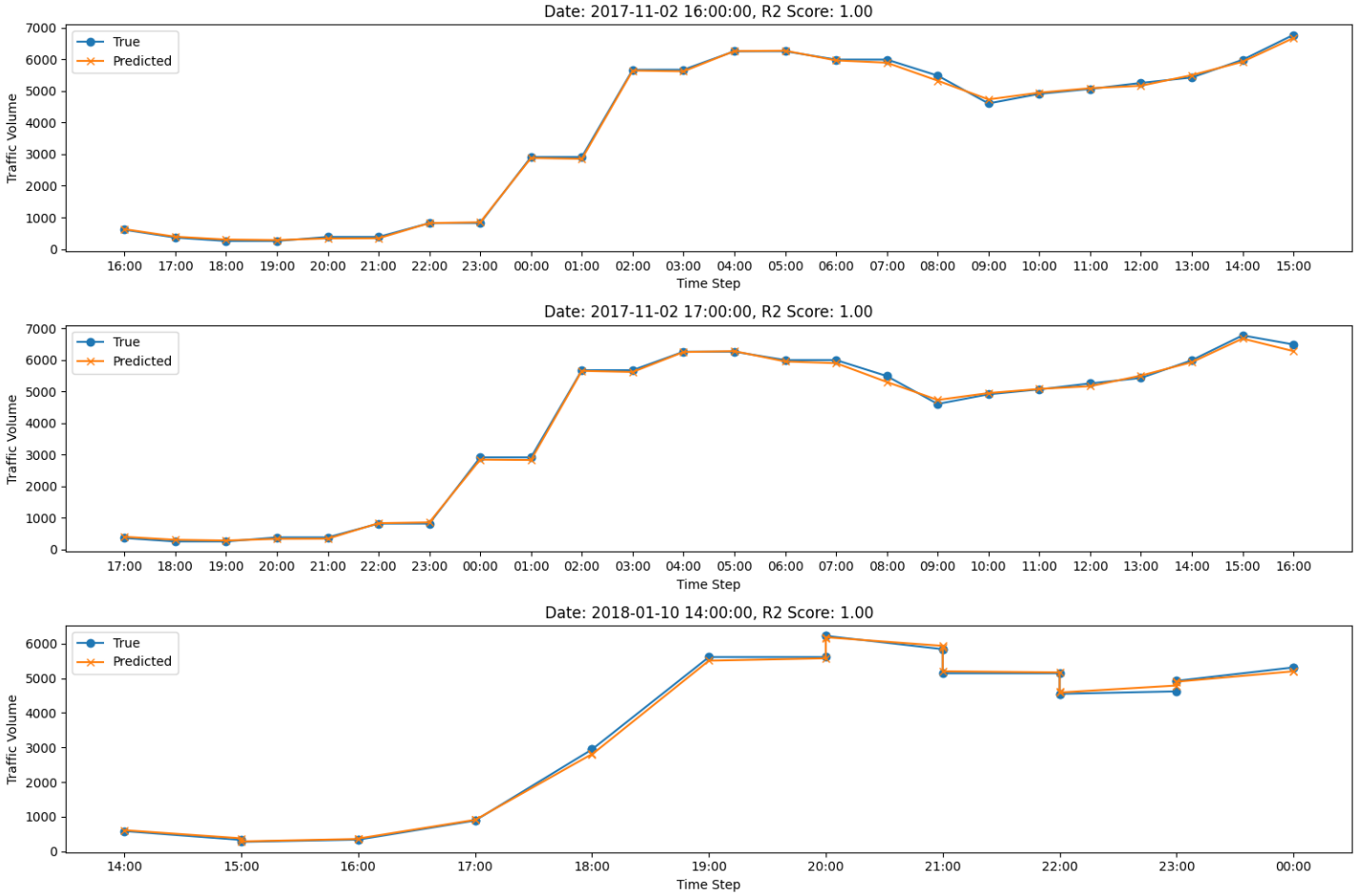
Malene Christiane Nielsen
mniel@mmmi.sdu.dk

The best model for recursive forecasting.

The best model for recursive forecasting is the GradientBoosting model.
The model has been trained on the multivariate_lagged_temporal_lags24 dataset.

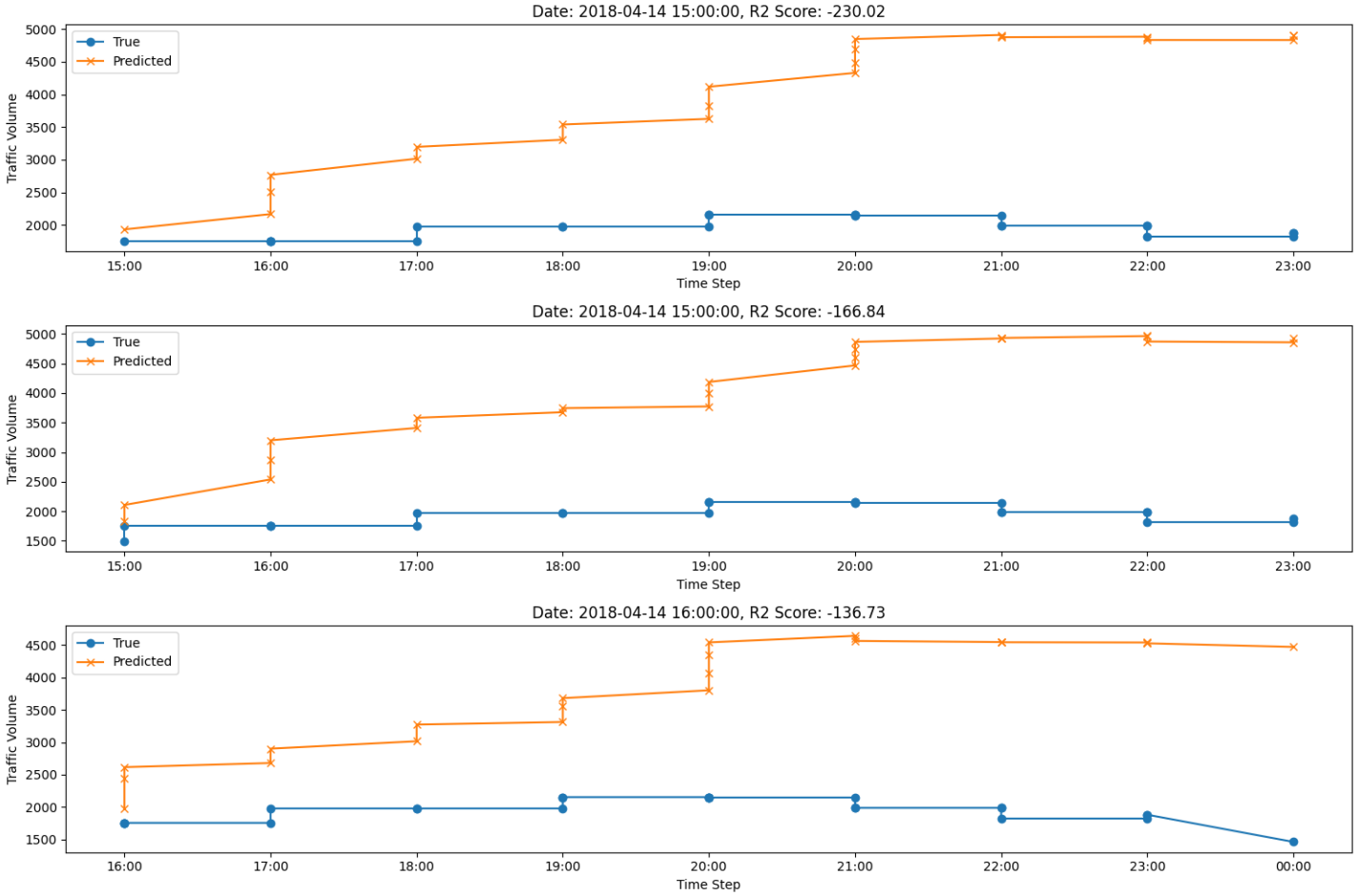
The best score for the recursive forecasting is 0.9429717020006919.

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