

Model Report: RandomForest

19/12/2024

Introduction

This is an automated report for the Experiment with tree models on traffic volume dataset; the RandomForest model.

This report will first introduce the model setup, including the hyperparameters and search algorithms used. Hereafter the base dataset will be described, and the differently created training datasets will be listed. After that, the results for the different forecast types will be presented, and the best results will be shown in plots.

Experiment description:

Experiment with tree models on traffic volume dataset

Model setup

The model has been used for the following forecast purposes:

- ☐ one_step
- ☐ multistep
- ☐ recursive

The model has 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:

- ☐ grid
- ☐ random

The used performance measure is the r^2 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
- 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']

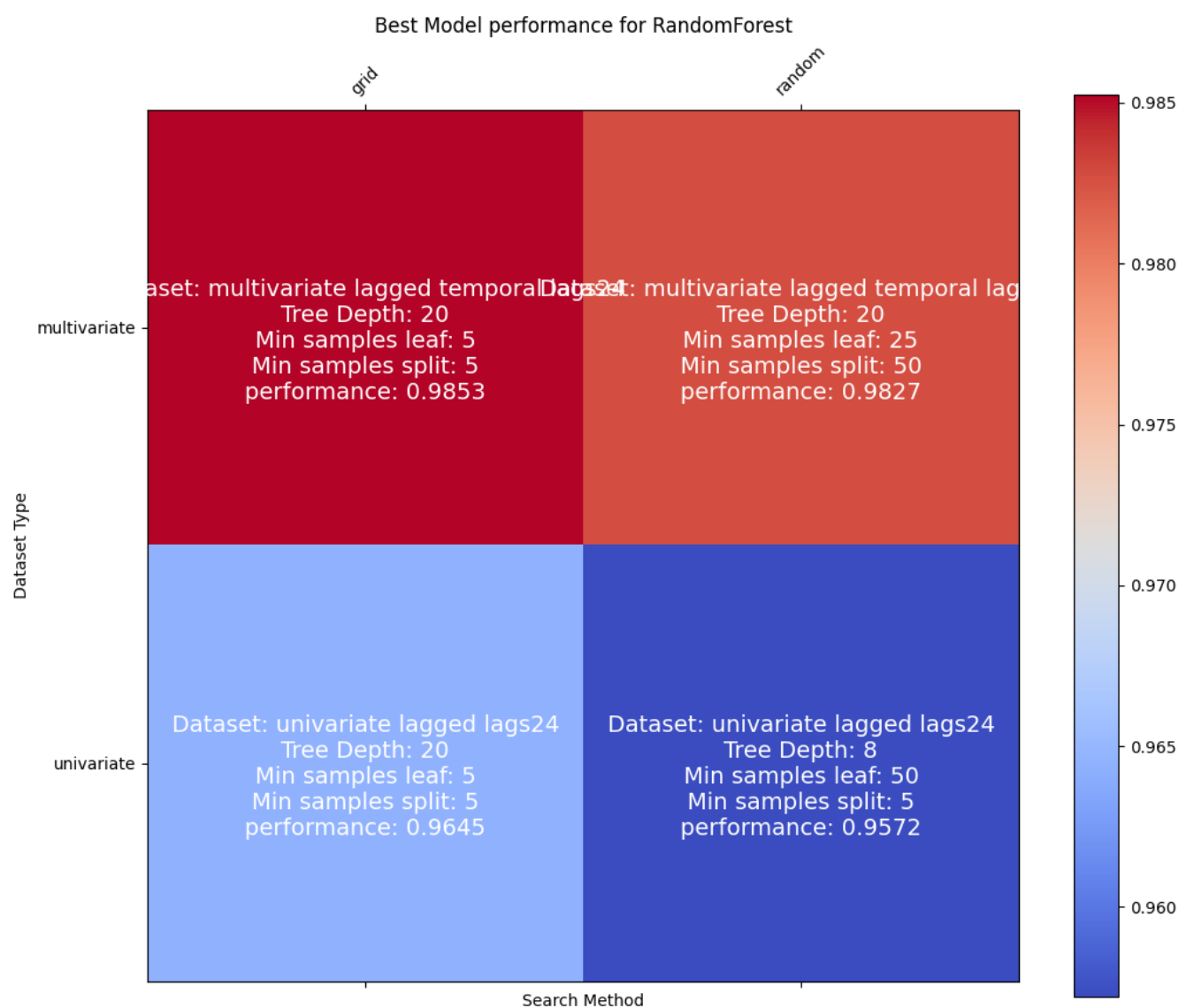
Results: RandomForest

The presentation of the results follows this system: For each prediction type, the best and worst results for each combination of search method and dataset type are presented in heat plots along with the corresponding model setup.

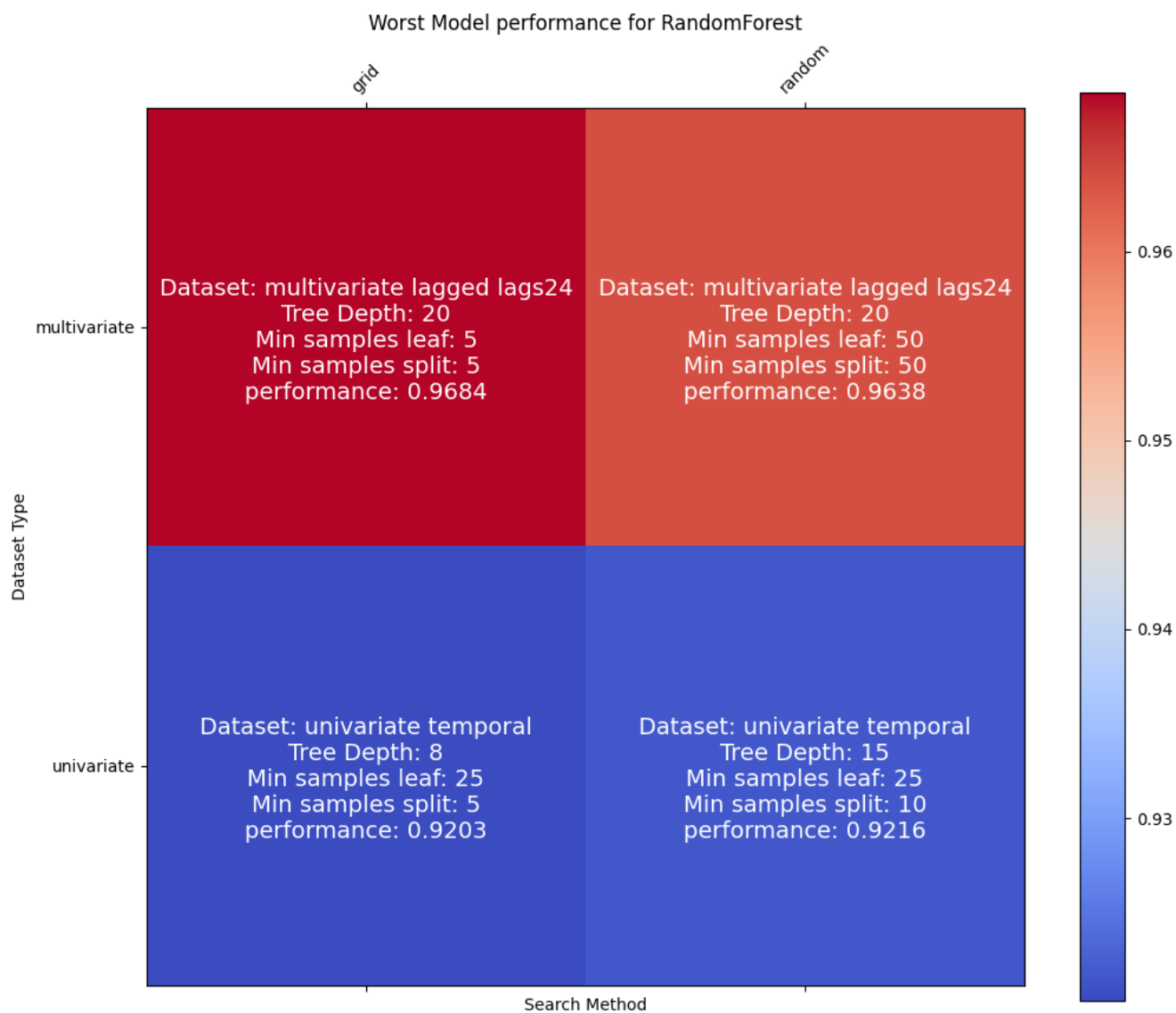
- Then, if the prediction type is one-step forecasts, the best prediction over time is visualized in a line plot.
- If the prediction type is a multi-step forecast, either direct or recursive, the model with the average best r^2 score is chosen, and the three best and worst predictions are visualized in a line plot. Furthermore, three steps of the forecasts are plotted.

Results for the one_step forecast.

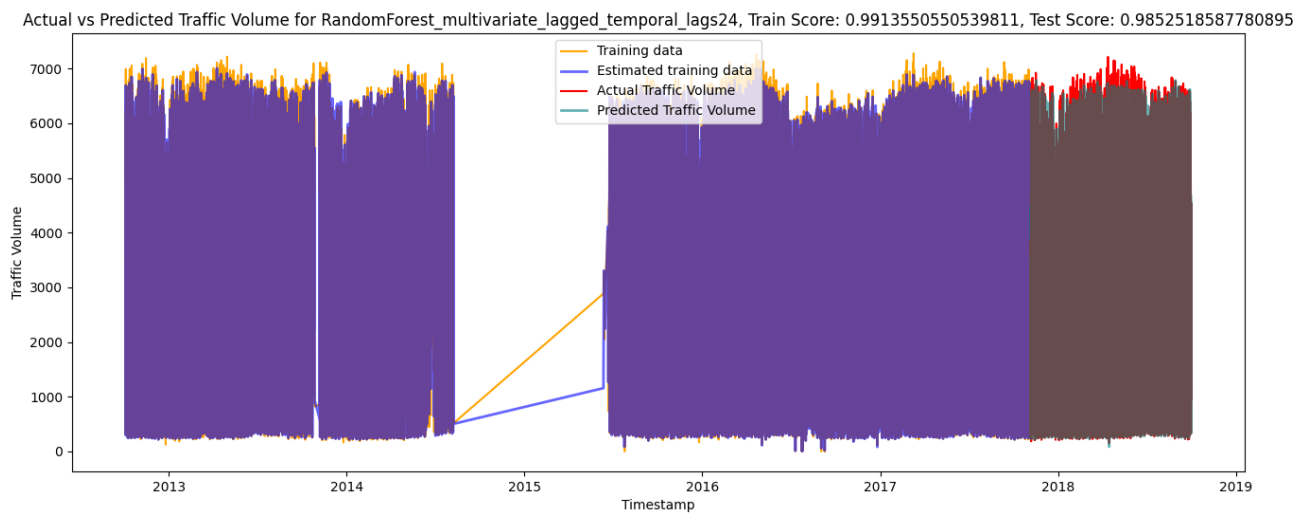
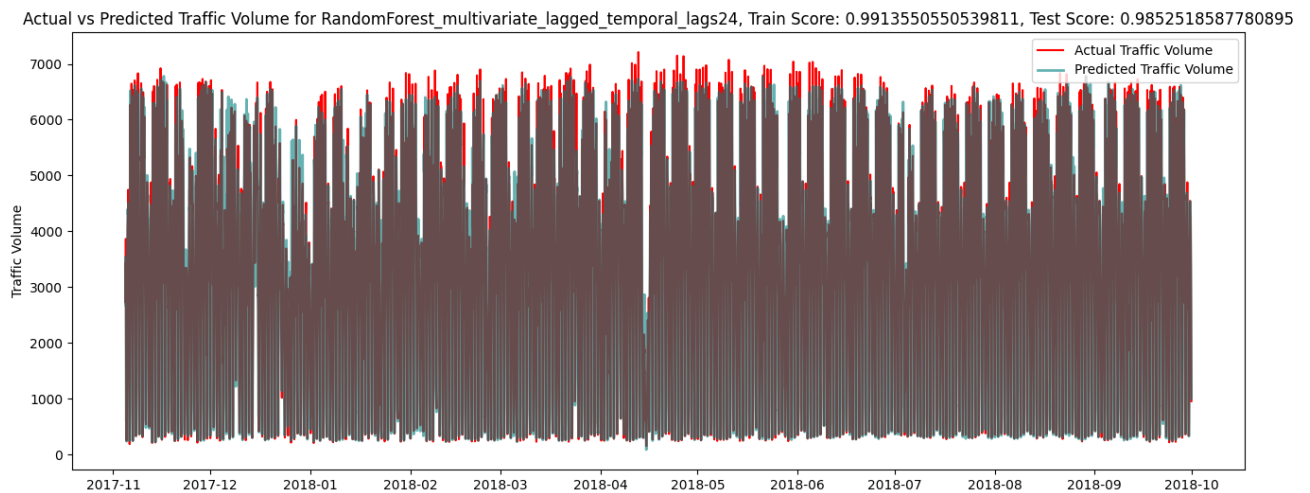
The best results (one_step) for the different setup combinations are as follows:



The worst results (one_step) for the different setup combinations are as follows:

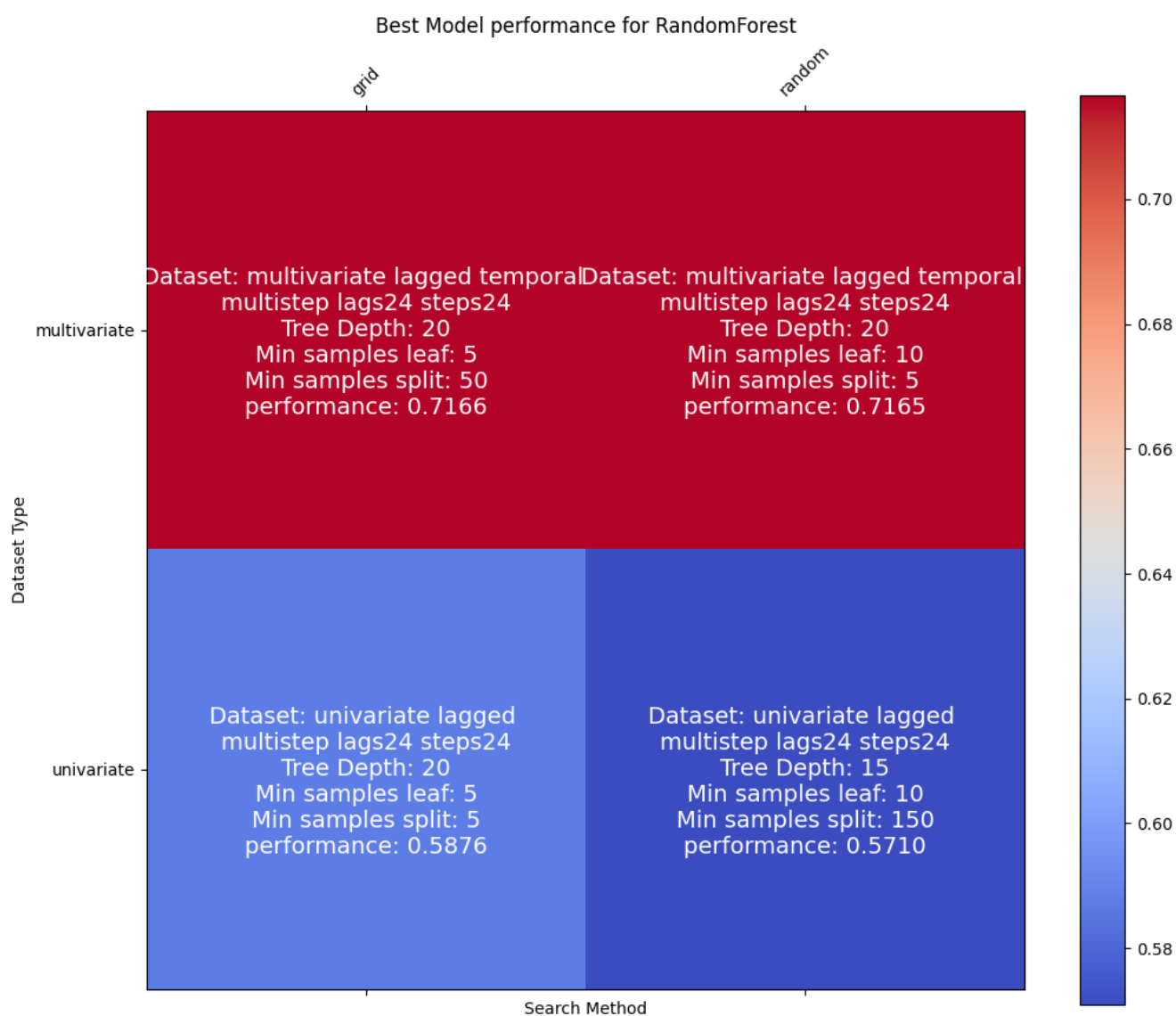


Best one_step forecast over time

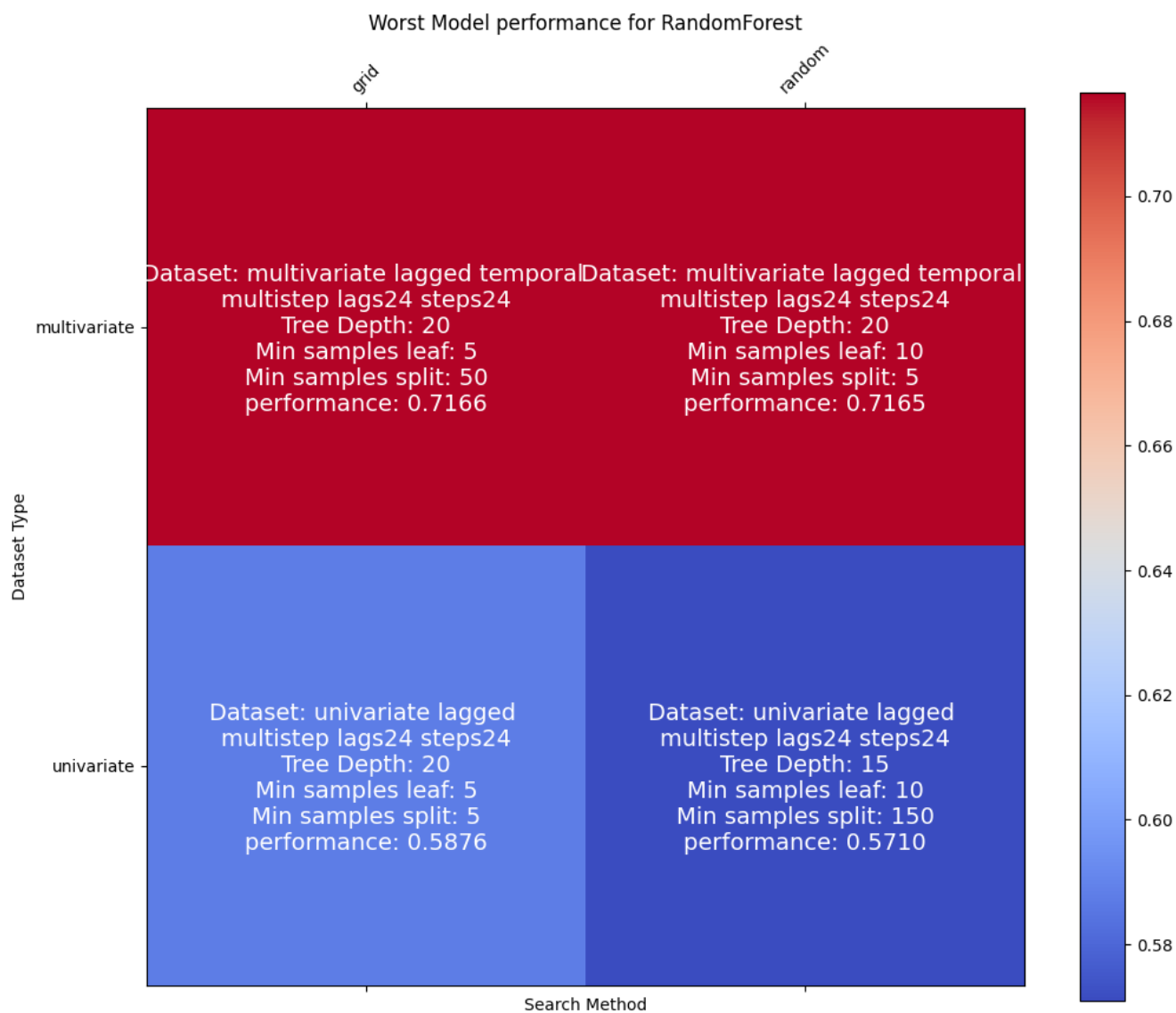


Results for the multistep forecast.

The best results (multistep) for the different setup combinations are as follows:

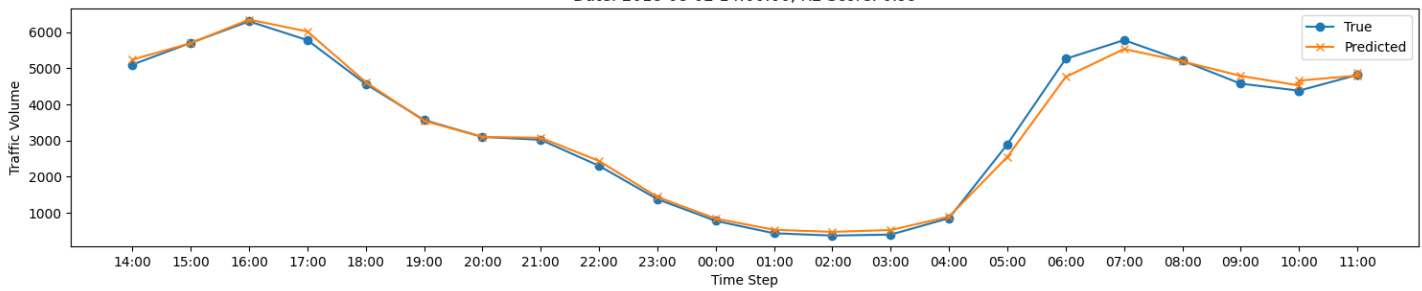


The worst results (multistep) for the different setup combinations are as follows:

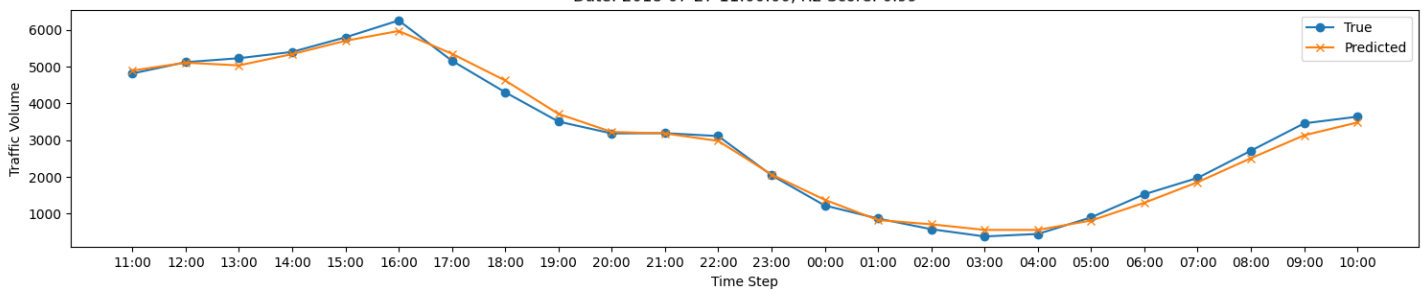


Best predicted days for RandomForest.

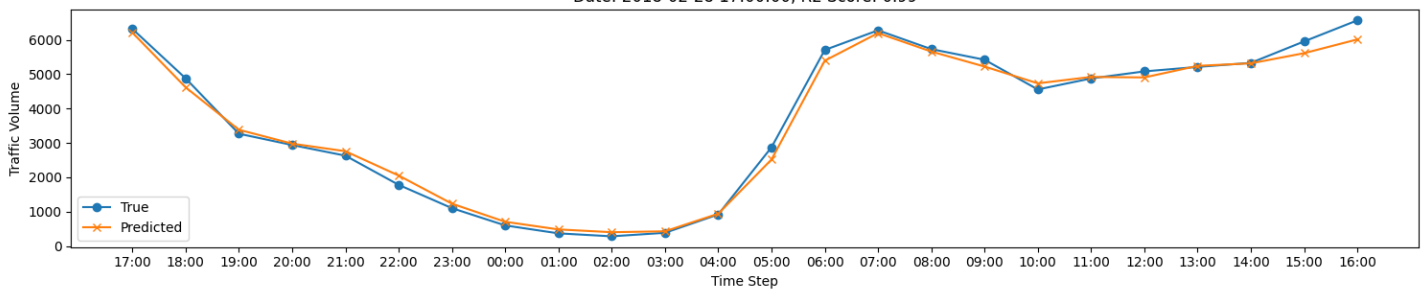
Date: 2018-08-02 14:00:00, R2 Score: 0.99



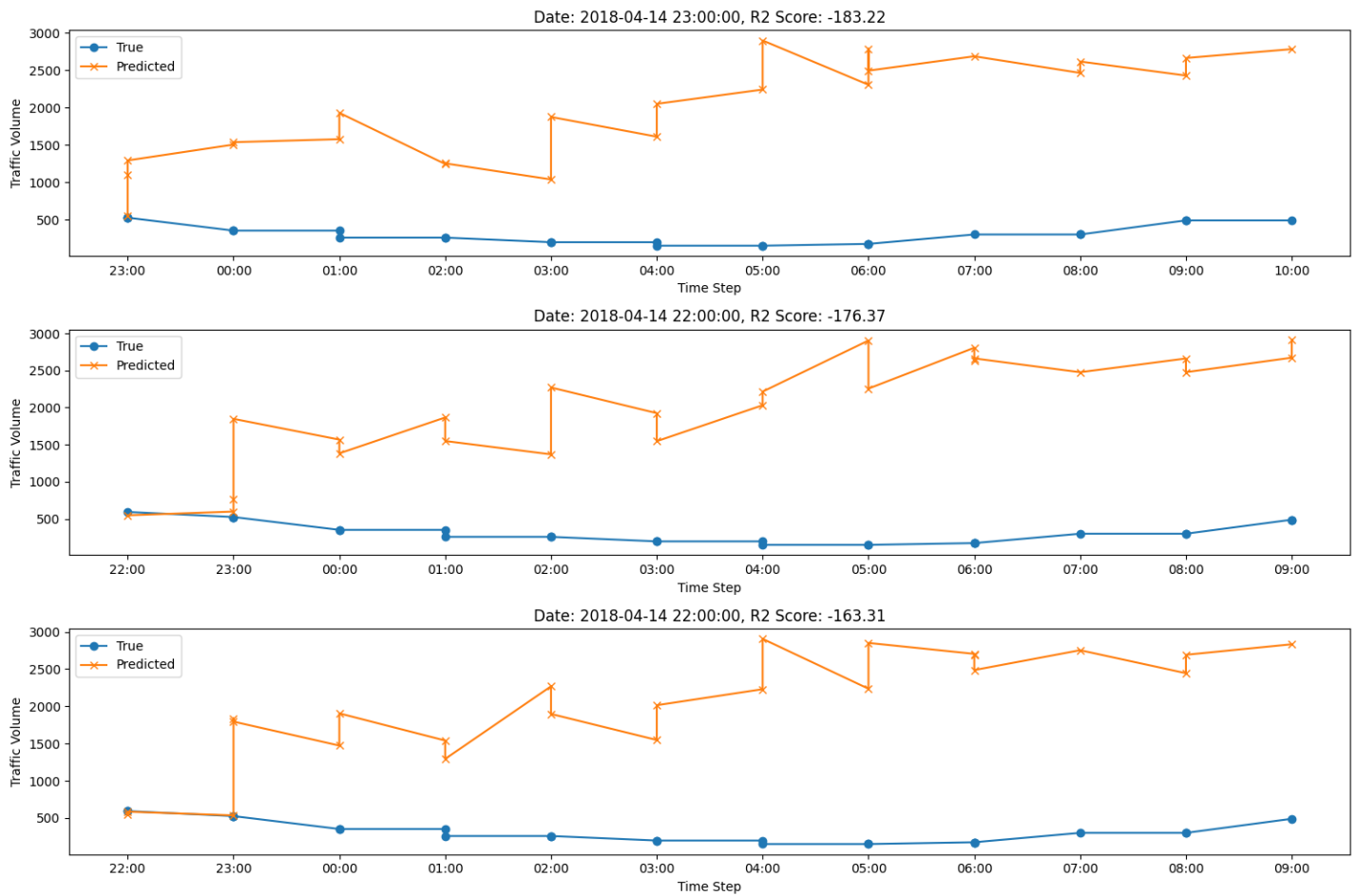
Date: 2018-07-27 11:00:00, R2 Score: 0.99



Date: 2018-02-28 17:00:00, R2 Score: 0.99

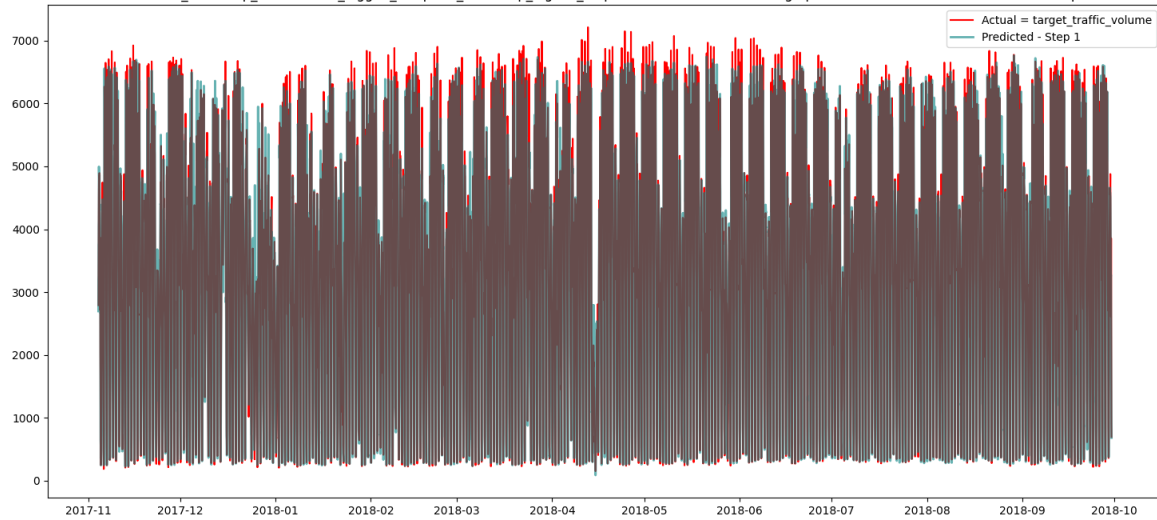


Worst predicted days for RandomForest.

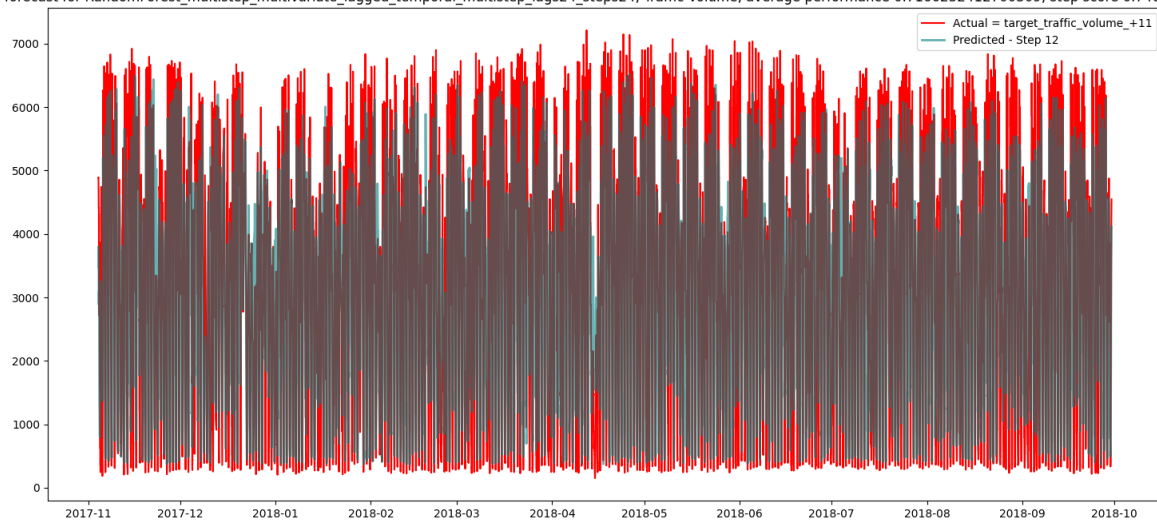


Steps plots for RandomForest forecasts over time

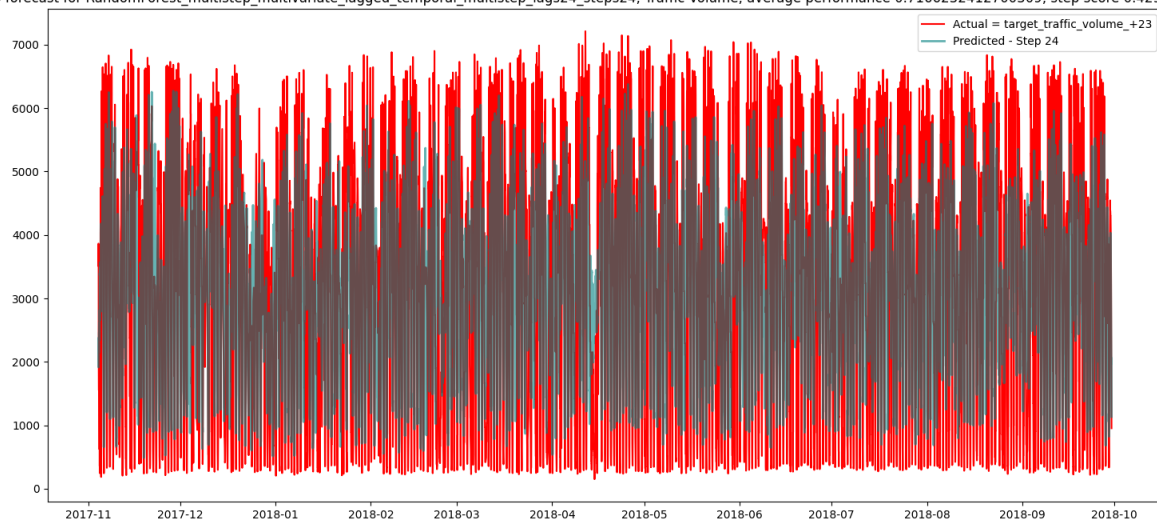
Multistep forecast for RandomForest_multistep_multivariate_lagged_temporal_multistep_lags24_steps24, Traffic Volume, average performance 0.7166232412700309, step score 0.9835439286909851



Multistep forecast for RandomForest_multistep_multivariate_lagged_temporal_multistep_lags24_steps24, Traffic Volume, average performance 0.7166232412700309, step score 0.7401884063041231

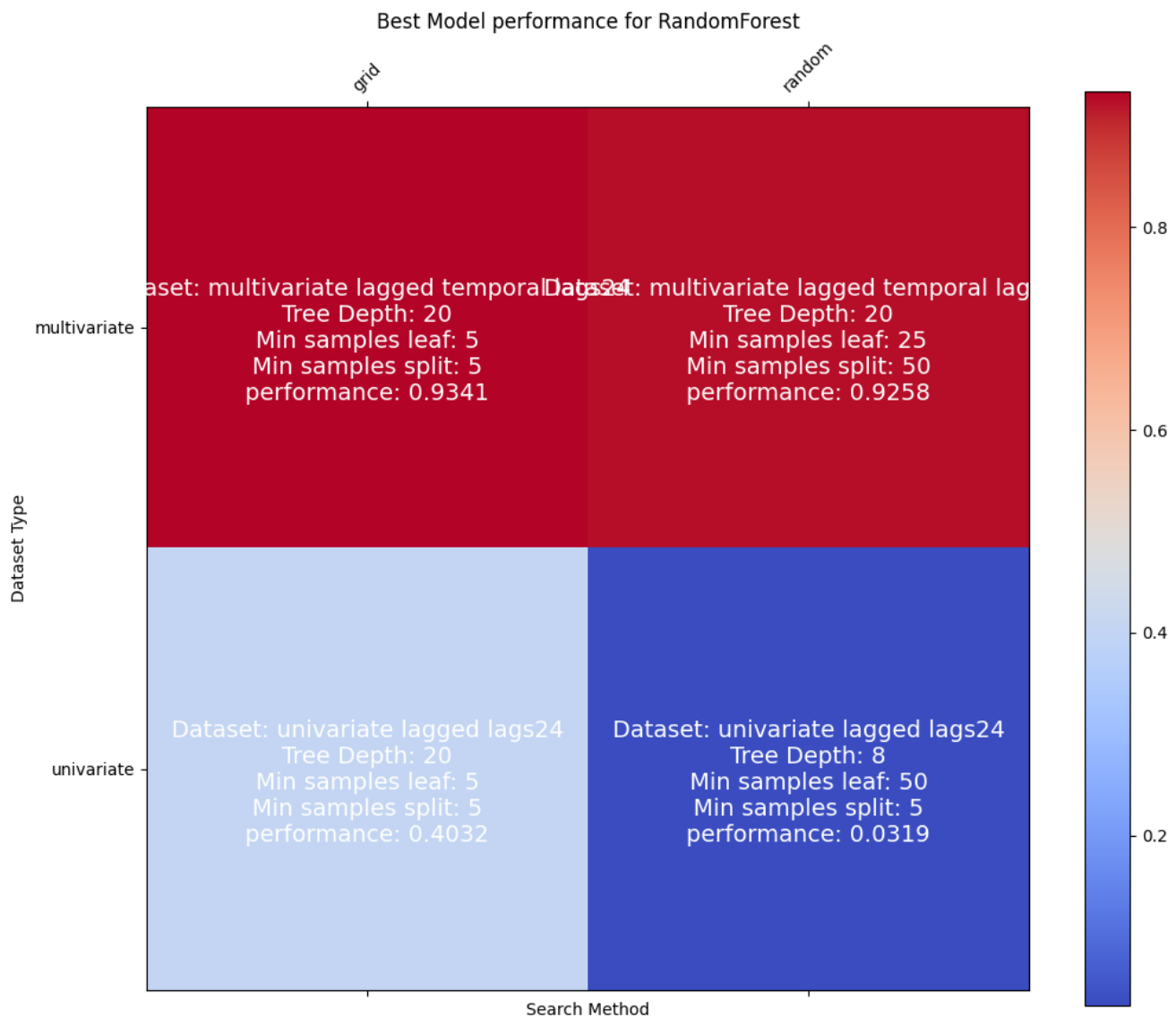


Multistep forecast for RandomForest_multistep_multivariate_lagged_temporal_multistep_lags24_steps24, Traffic Volume, average performance 0.7166232412700309, step score 0.4296072464226528

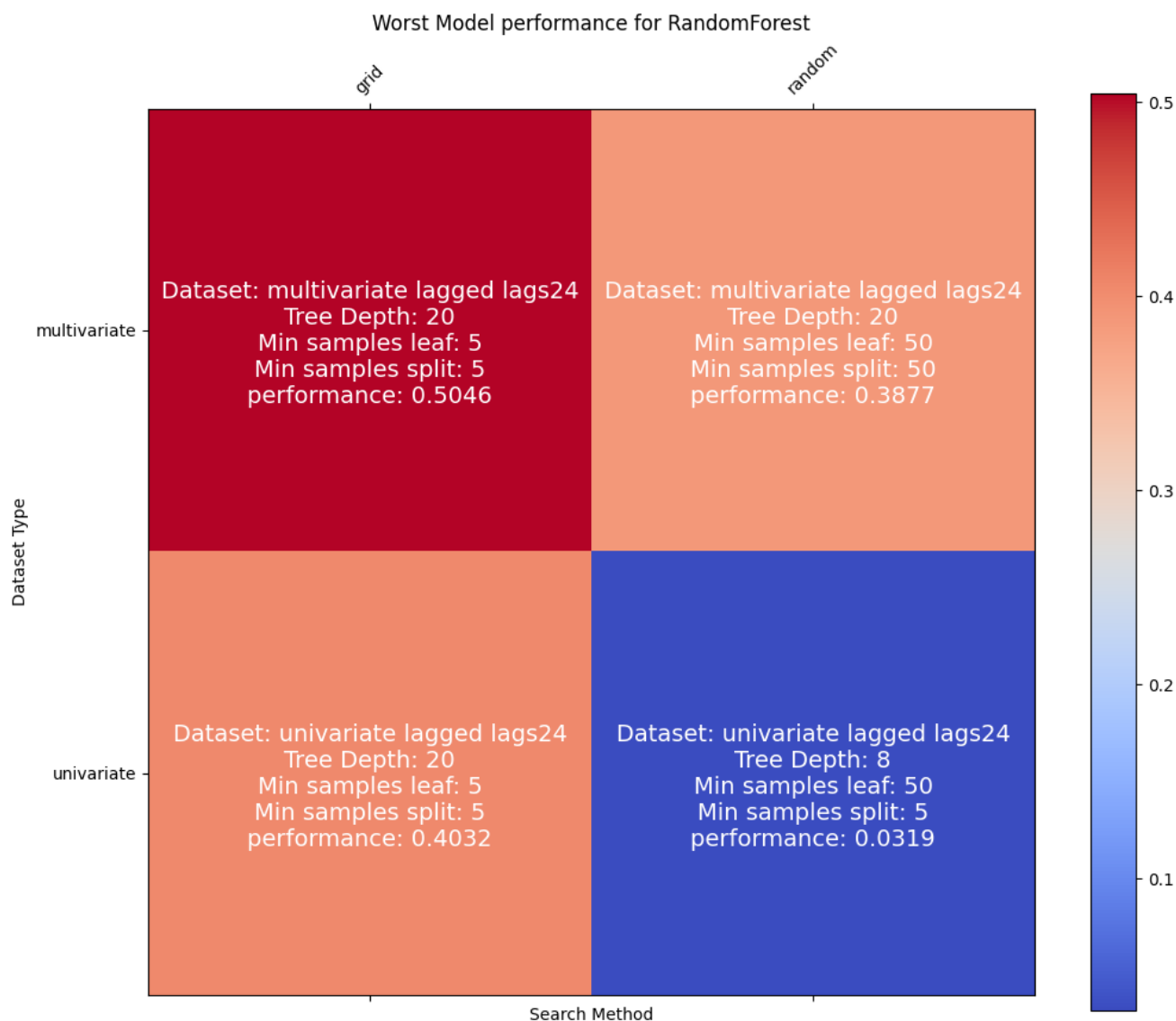


Results for the recursive forecast.

The best results (recursive) for the different setup combinations are as follows:

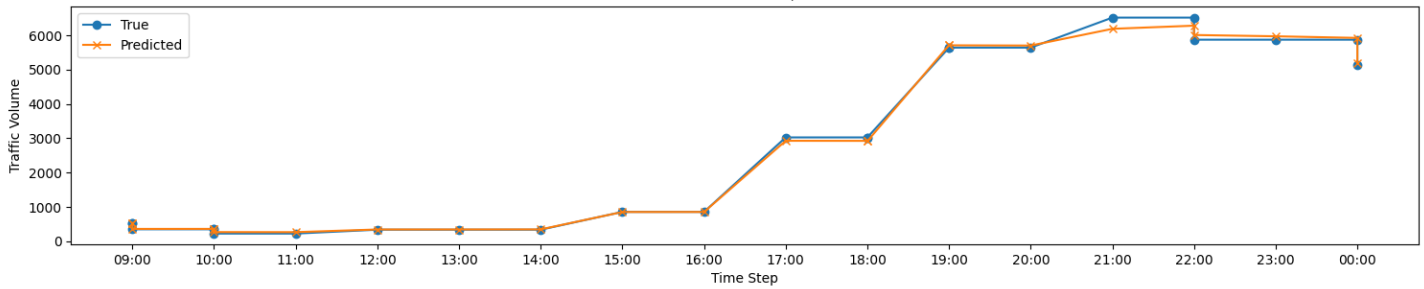


The worst results (recursive) for the different setup combinations are as follows:

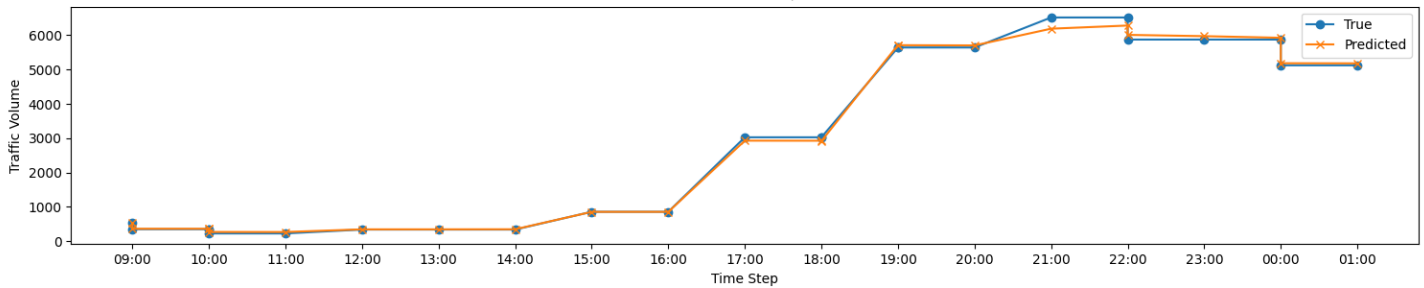


Best predicted days for RandomForest.

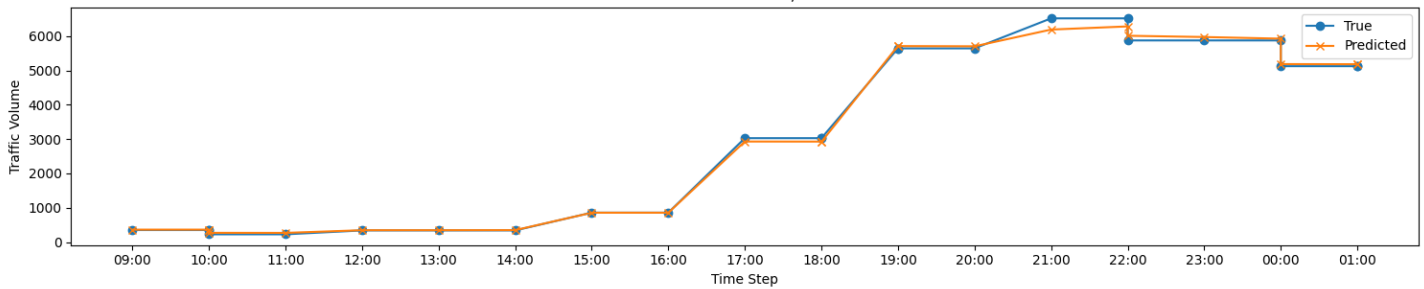
Date: 2018-09-18 09:00:00, R2 Score: 1.00



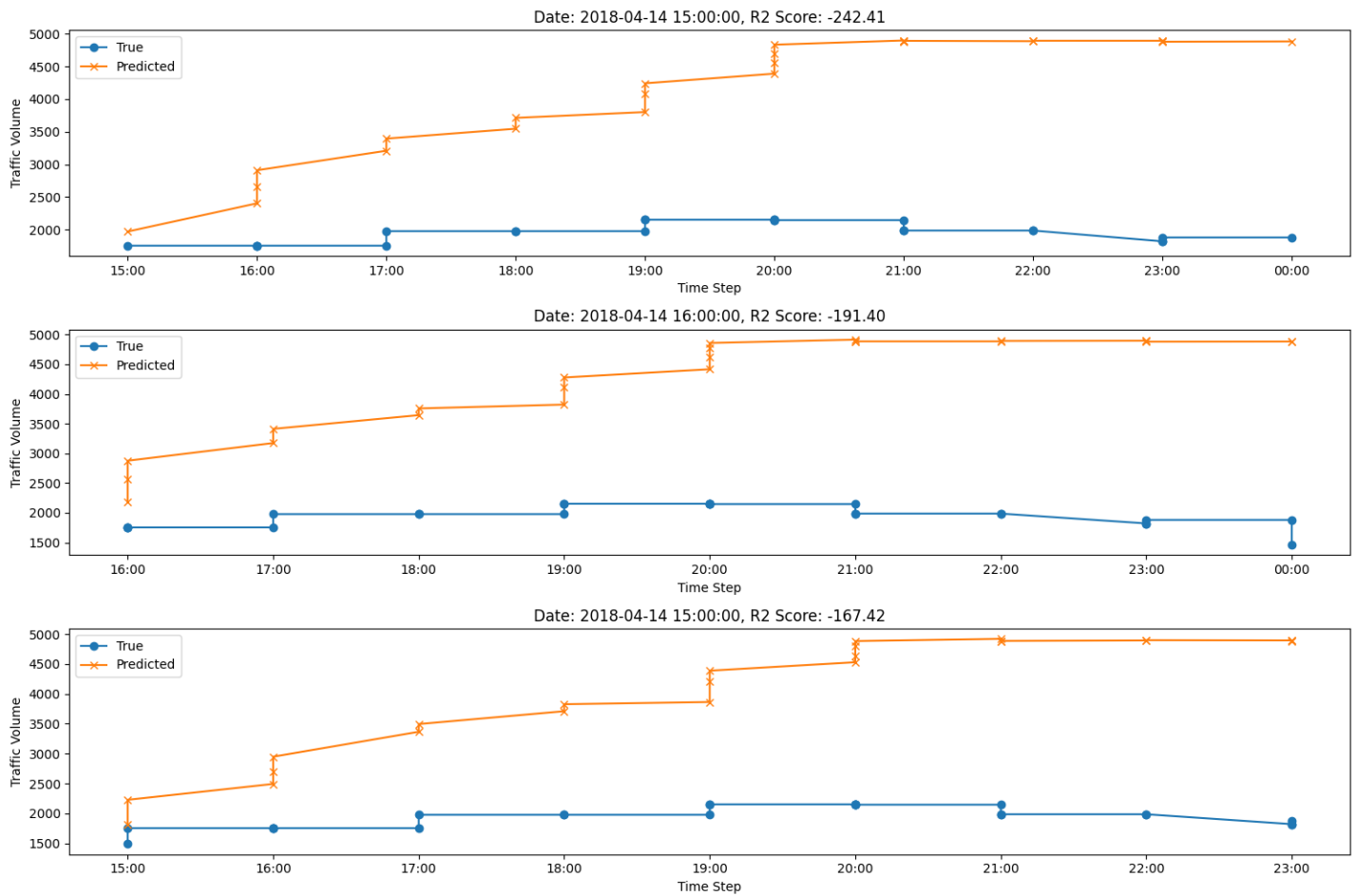
Date: 2018-09-18 09:00:00, R2 Score: 1.00



Date: 2018-09-18 09:00:00, R2 Score: 1.00

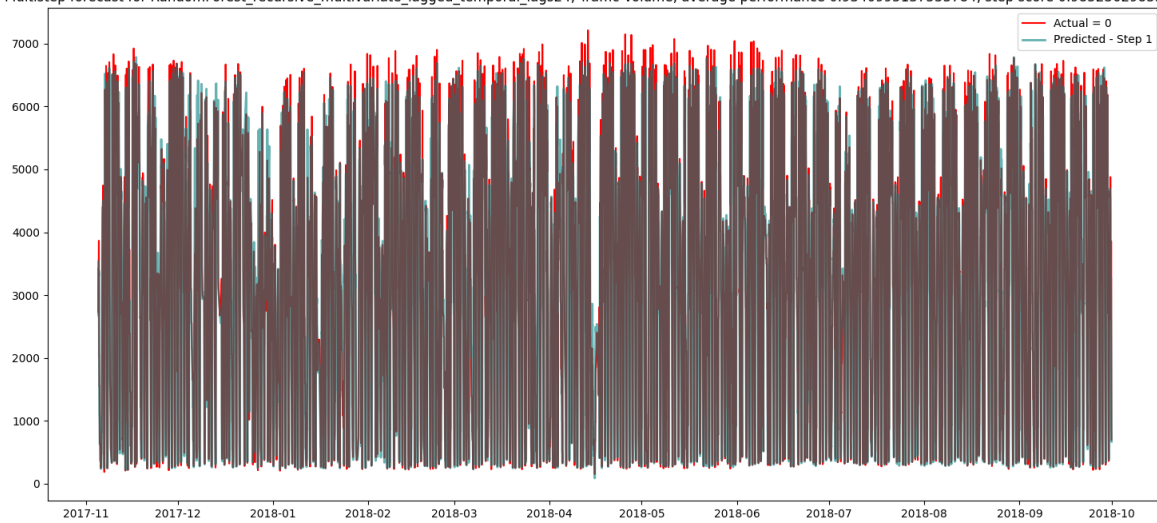


Worst predicted days for RandomForest.

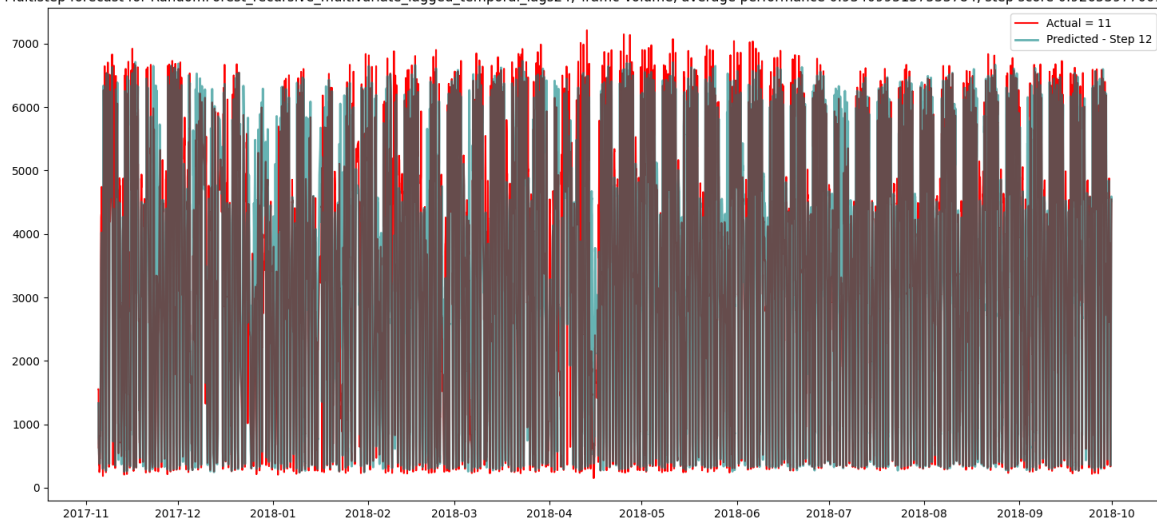


Steps plots for RandomForest forecasts over time

Multistep forecast for RandomForest_recursive_multivariate_lagged_temporal_lags24, Traffic Volume, average performance 0.9340995137533784, step score 0.9852362983365711



Multistep forecast for RandomForest_recursive_multivariate_lagged_temporal_lags24, Traffic Volume, average performance 0.9340995137533784, step score 0.9265397700111615



Multistep forecast for RandomForest_recursive_multivariate_lagged_temporal_lags24, Traffic Volume, average performance 0.9340995137533784, step score 0.920486438626498

