SDU ENERGY INFORMATICS

Malene Christiane Nielsen
mniel@mmmi.sdu.dk

Model Report: RandomForestClassifier

09/12/2024

Introduction

This is an automated report for the Classification on Red Wine Quality data; the RandomForestClassifier 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 for classifying the Quality of red wine

Model setup

The model has been used for the following forecast purposes:

- one_step

The model has been optimized using the following hyperparameters:

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

And with the following search algorithms:

- grid
- random

The used performance measure is the accuracy measure.

Malene Christiane Nielsen
mniel@mmmi.sdu.dk

Dataset setup

The baseline dataset used for these forecasts is
the 'Wine Quality Dataset' dataset: *'Wine Quality Dataset for red wine'.*

The test size used for the forecasts is 0.2.

- Dataset 1

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

Malene Christiane Nielsen
mniel@mmmi.sdu.dk

Results: RandomForestClassifier

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 r2 score is chosen, and the three best and worst predictions are visualized in a line plot. Furthermore, three steps of the forecasts are plotted.

Malene Christiane Nielsen
mniel@mmmi.sdu.dk

Best one_step forecast over time