Creator **Olusanmi Hundogan** Created **Aug 02, 2023, 15:49** Last updated **Aug 08, 2023, 12:52**

Files

- experiments/experiment_MMA_n_trials_n_startup_trials.py
- notebooks/analyze_MMA_n_trials_n_startuptrials.py

Motivation

The hyperparameter optimizers used for the MMA estimator, both for classification and regression tasks, are based on Optuna studies. The creation of these studies is based on a sampler, which requires the "n_startup_trials" parameter (the number of random iterations at the begining - more details about this here: https://optuna.readthedocs.io/en/stable/reference/samplers/generated/optuna.samplers.TPESampler.html). Then, at the optimization stage of the study, the "n_trials" parameter is used to specify the full number of iterations.

The "n_trials" and "n_startup_trials" should have optimal values as well, in order to make use of Optuna's full potential for the optimization tasks of our model.

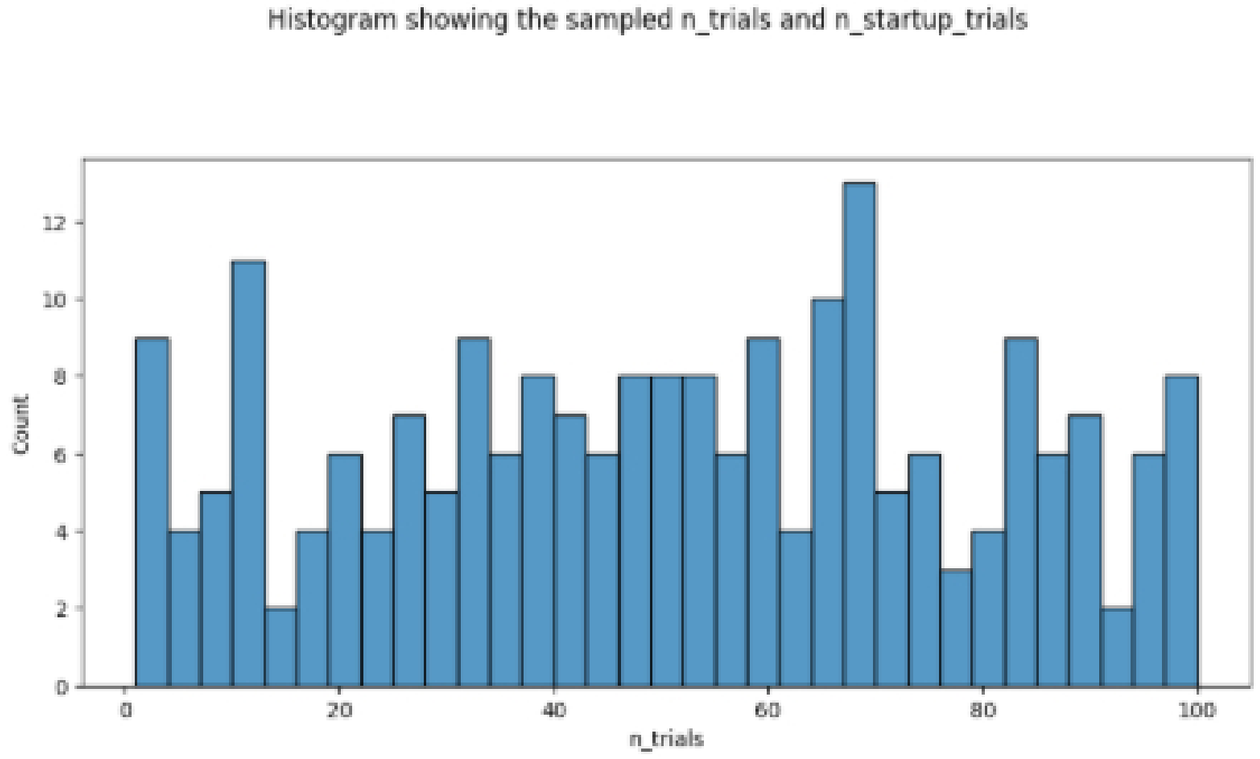
Design

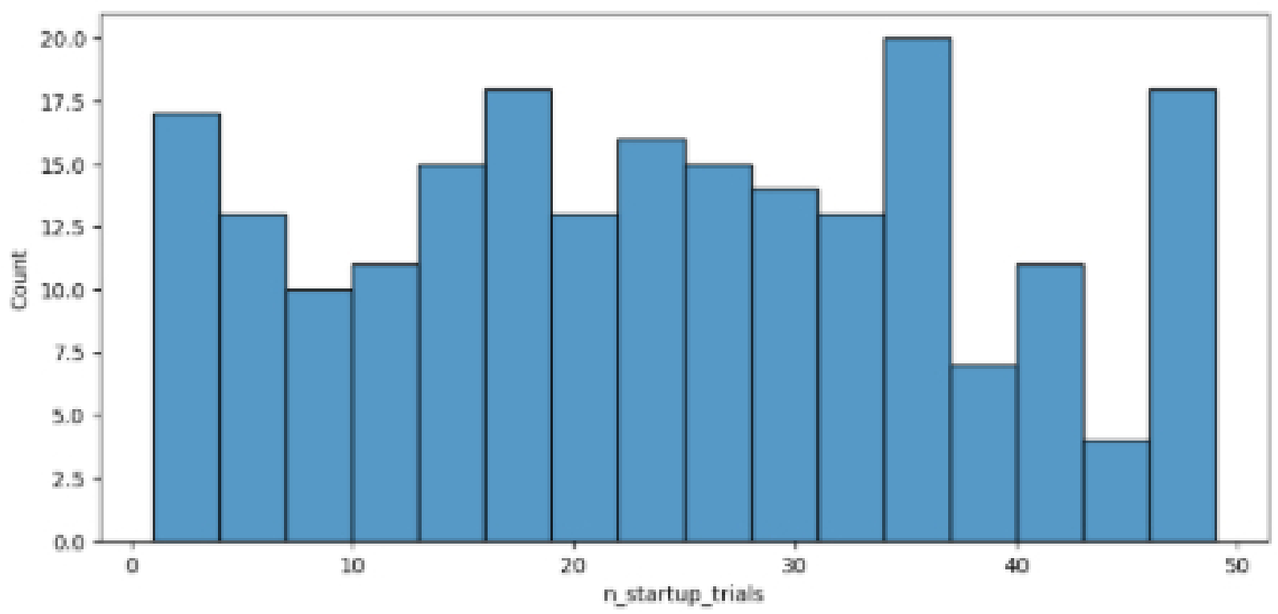
There are 30 data splits of the full dataset, on each split we have 10 iterations of the experiment. In each iteration we define a new pipeline that uses the MMA estimator, for which we choose the values for "n_trials" and "n_startup_trials" randomly from intervals 1 - 100, respectively 1 - 50. The results were computed only for scope 1.

We ran this experiment in multiple batches since it took a very long time, and got approximately 200 results.

Results and Insight

Below we can see that the random selection of values for the trials parameters was uniform enough to be able to make a proper analysis of the results.

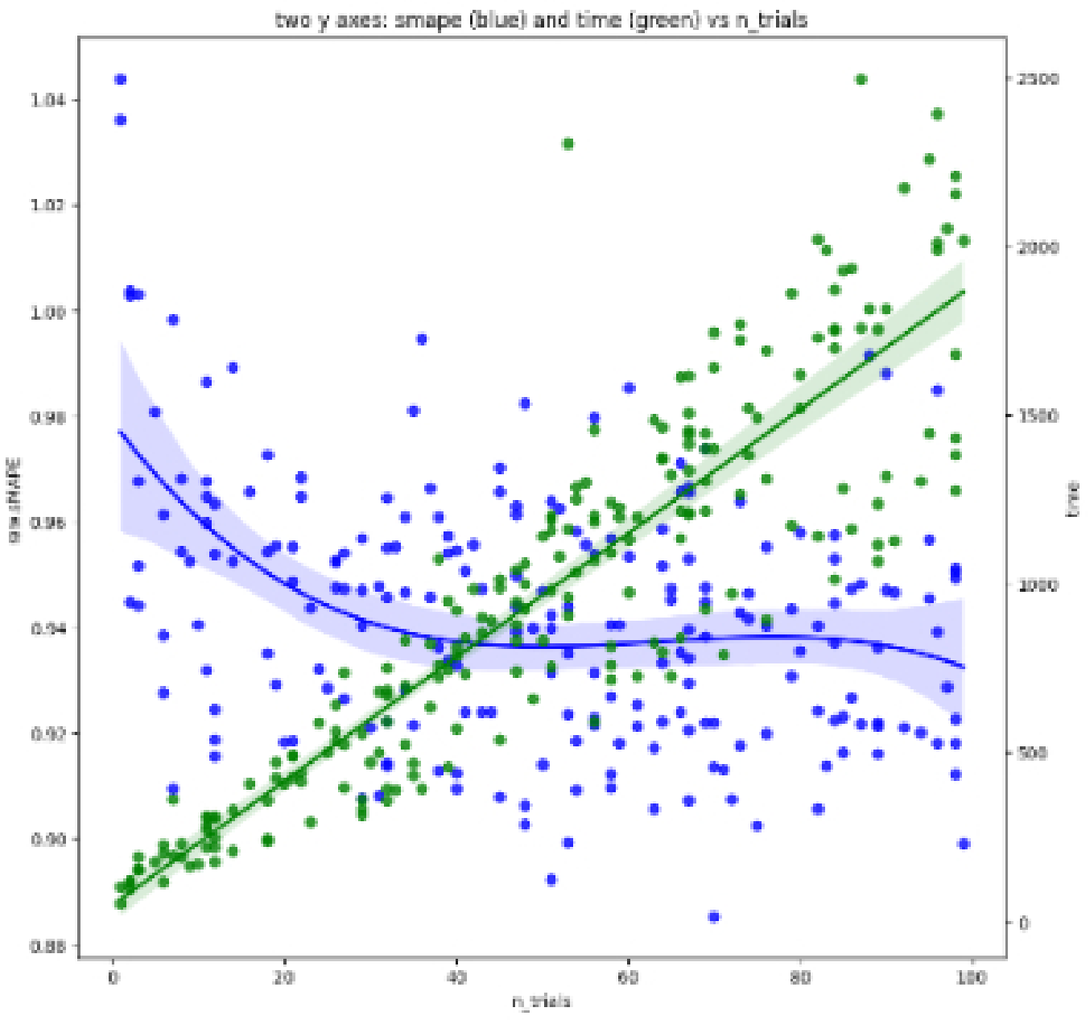


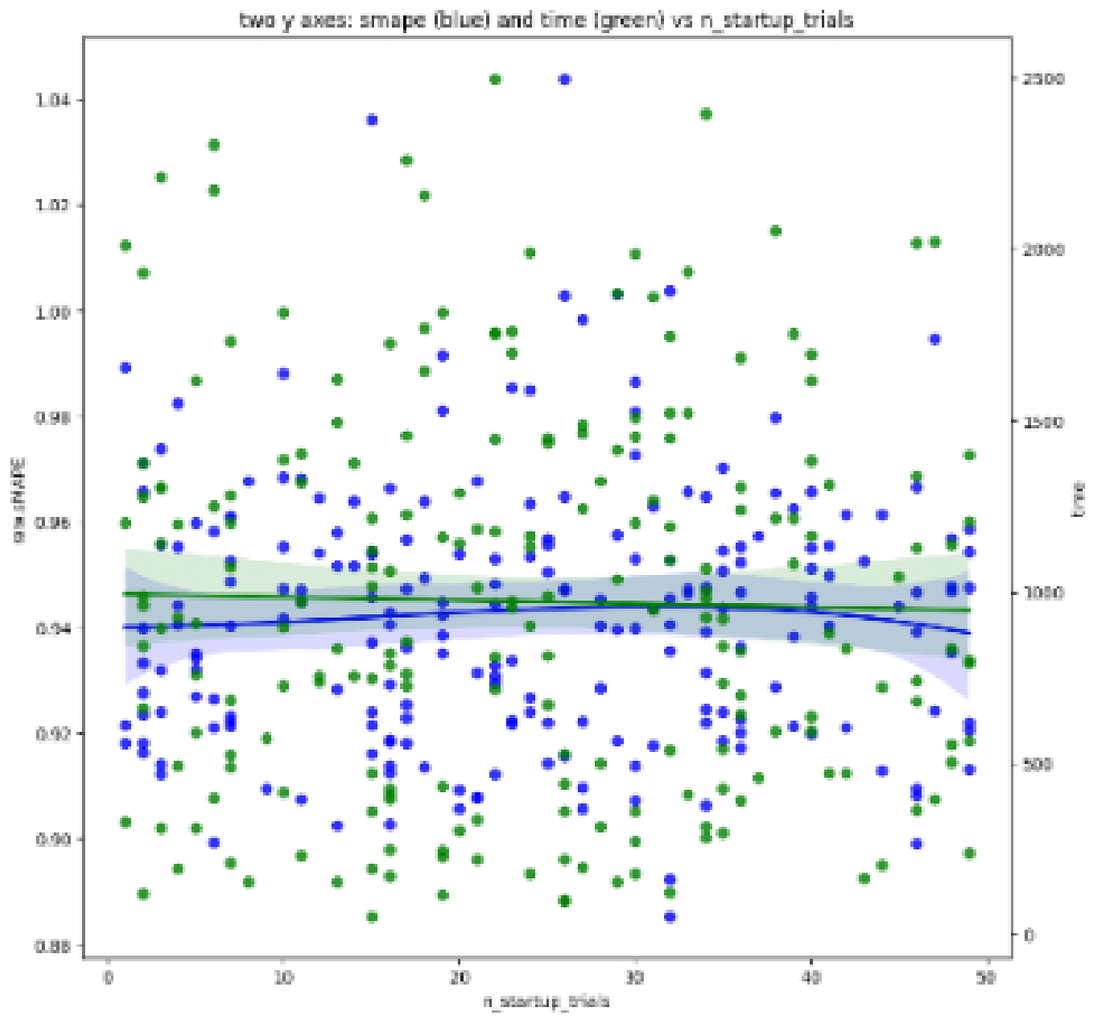


The choice of the "n_trials" parameter has a big impact on the performance of the model and on the runtime, while "n_startup_trials" barely makes any changes to them.

It makes sense to see smaller runtimes for iterations with fewer trials, but the goal was to find a small enough value of "n_trials" which does not increase the sMAPEscore. The plots below show that the best point of tradeoff is at 40 n_trials and 20 n_startup_trials

The sharp drop at the end of the sMAPEcurve is likely an artifact from curve fitting with a polinomial of degree 3.





These heatmaps present the same results, but with the sMAPE score aggregated (left shows the mean, right shows the median). Between the lightest (best sMAPE) areas, the one which also corresponds to a reasonable runtime is the bin of 40 - 60 n_trials and 20 - 30 n_startup_trials, so we can safely choose the lower boundaries of these intervals.

Decision

We will use 40 n_trials and 20 n_startup_trials from now on, as these values bring us as close as possible to both goals, performance and runtime.