

A Novel AutoML Solution with NePS

Ömer Emre POLAT, Barbaros INAK
Dropouts

Modality 1/2

Motivation

Our main motivation was to create a versatile one click AutoML solution that can adapt to any tabular dataset using 8 different state-of-the-art regression models.

NePS was chosen as an hyperparameter optimizer because it was the option which we had experience in exercises and it was transparent with the source code.

Used Methods

Bayesian Hyperparameter Optimization

Successive Halving

IQR Variable Scale Outlier Detection

Standardization

- Week 1
- Week 2
- Week 3
- Week 4
- Week 5
- Week 6
- Week 7
- Week 8
- Week 9
- Week 10
- Bonus
- Literature

Resources Used

For development:

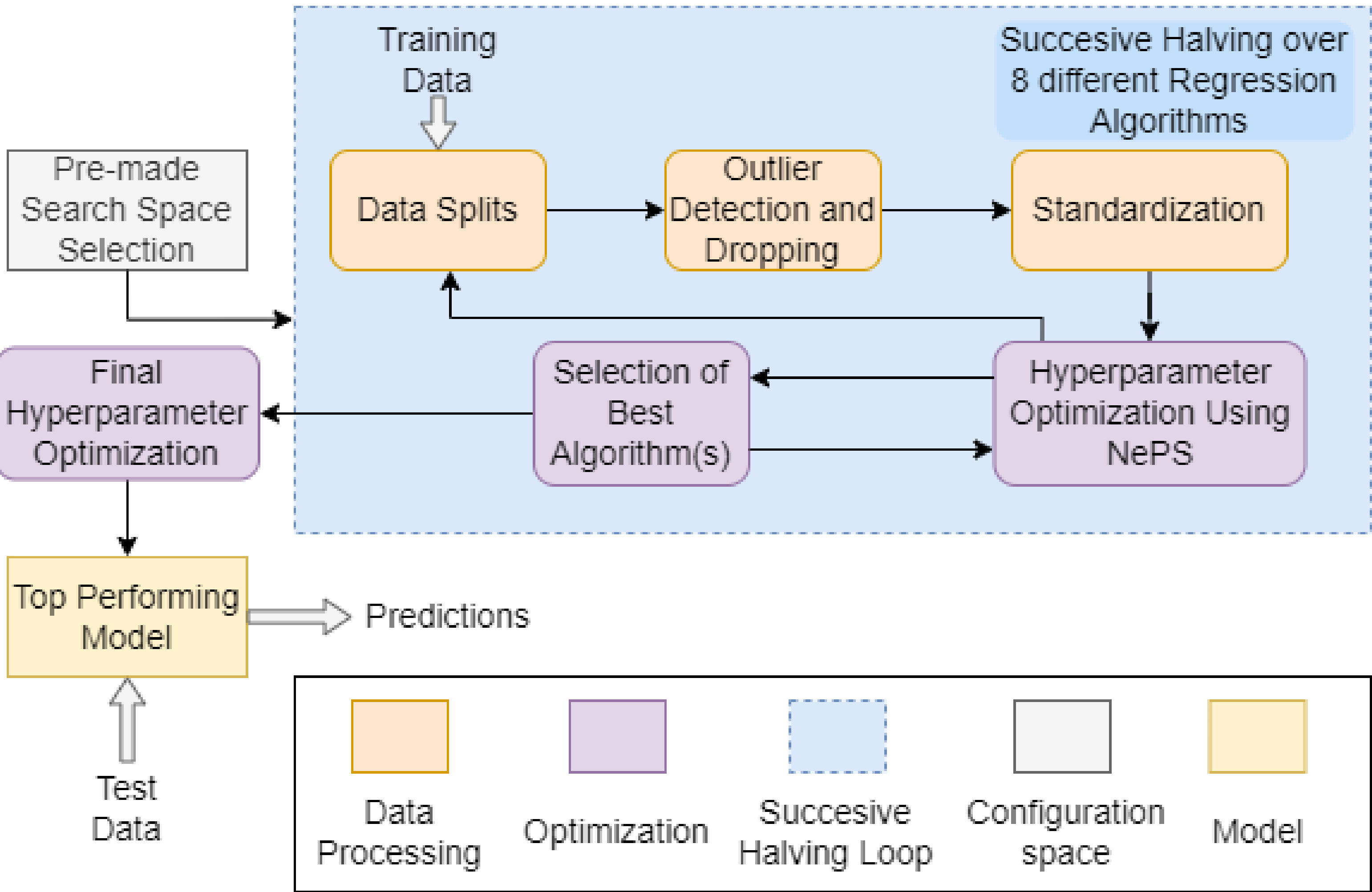
- 1 Ryzen 9 5900HX (Laptop)
- Total compute estimate: 50 CPU-hours

Workforce:

- 2 full week on average

Our Approach

Pipeline



Details

The pipeline consists of a main successive halving loop which works on 8 different regression algorithms such as:

- Sklearn and XGB Random Forest
- Gradient Boost
- Ada Boost
- MLP
- Bayesian Ridge
- Elastic Net
- Dart Boost

Optimization run on the given configuration space, the IQR scale and the validation split ratio. The resulting best model is then saved to be used for predictions.

Empirical Results

y_prop Dataset	Bike Sharing Dataset	Brazilian Houses Dataset	Exam Dataset
Validation 0.1229	Validation 0.9423	Validation 0.9997	Validation 0.9016
Test 0.1135	Test 0.9755	Test 0.997	Test
Chosen Algorithm: XGB Random Forest	Chosen Algorithm: SKlearn Random Forest	Chosen Algorithm: SKlearn Random Forest	Chosen Algorithm: Gradient Boost
For the y_prop dataset, our AutoML solution found the XGB Random Forest as the most optimal algorithm.	For the Bike Sharing Dataset, our AutoML solution found the Gradient Boost as the most optimal algorithm.	For the Brazilian Houses Dataset, our AutoML solution found the MLP as the most optimal algorithm.	For the Exam Dataset, our AutoML solution found the Dart Boost as the most optimal algorithm.

Number of queries for test score generation: 5

