Creator **Olusanmi Hundogan** Created **Jul 30, 2023, 19:48** Last updated **Aug 08, 2023, 12:49**

Files

- experiments/experiment_optuna_metric.py
- notebooks/analyze_optuna_metric.py

Motivation

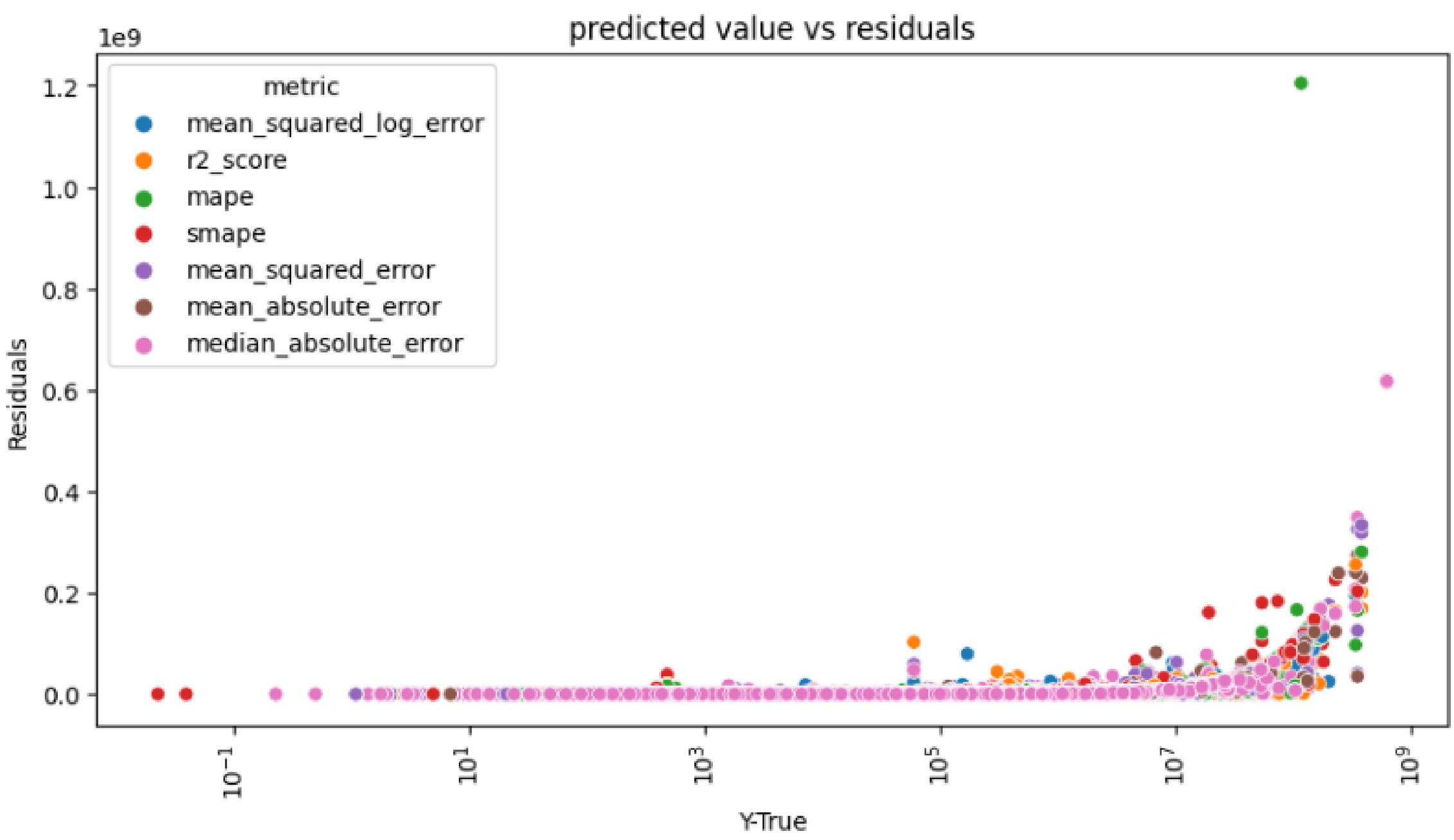
To use optuna it is necessary to choose an optimization objective for the hyperparameter optimization process. There are several metrics to choose from. This experiment investigates which of them yields the best results.

Design

This experiment runs X repetitions of training an XGB model using different hyperparameter optimization metrics. Each repetition has a different data sample. For each metric the models are hyperparameter optimized on the data sample. For the analysis we gather the residuals.

Results and Insights

The results show that most metrics behave similar. It is difficult to distinguish visually which of the metrics performed the best.



When looking at the median, mean and sum of the residuals, no metric strictly dominates the other metrics. Only two metrics appear to be the best in at least one category.

The sum of MSE appears to be just x10.000 of the mean. The reason is unclear but it could be that a large outlier residual dominates the other ones. In other words, if one divides

10 000 000 + 0.01 + 0.001 + 0.1 + 0.01 + 0.0001 + 0.01 + 0.1 + 0.1 + 0.01 by 10 instances, then the average will be like dividing close to 1 Mio.

	median	mean	sum
mape	43010.390625	2664558	26645580000
mean_absolute_error	39100.970391	2300540	23005400000
mean_squared_error	41726.169922	**2268697**	**22686970000**
mean_squared_log_error	43635.478516	2424838	24248380000
median_absolute_error	**36455.137695**	2364672	23646720000
r2_score	41956.496826	2353687	23536870000
smape	43101.539062	2705603	27056030000

Decision

We will use the median_absolute_error as a metric. The median for the residuals are the lowest for this metric. Meaning that 50% of the residuals are below this number.

We reject the Mean Squared Error, because the mean is susceptible to outliers.