

Explainable Automated Machine Learning

Project 1

Kristjan Laid, Mattias Väli, Evely Kirsiaed

Tartu Ülikool

Digits dataset

- Classification task for hand written digits
- 10 classes
- Sample total at 1797

Base Models

Created base models and measured accuracy with 5 fold cross validation

- KNeighborsClassifier - 94.4%
- SVC - 94.9%
- RandomForestClassifier - 93.9%
- GradientBoostingClassifier - 92.2%
- LogisticRegression - 91.9%
- XGBClassifier - 91.8%
- RidgeClassifier - 88.8%
- NearestCentroid - 85.6%
- BernoulliNB - 85.3%
- DecisionTreeClassifier - 79.0%
- GaussianNB - 76.9%

Hyperpot with Random Search

Calculating the best possible hyperparameters using RandomForest and Random Search

- n estimators from 30 to 250 with steps of 5
- max depths from 2 to 20 with steps of 1
- n estimators from 1 to 50 with steps of 1
- n estimators from 1 to 50 with steps of
- Best depth - 14 ,min samples - 44, minimum split - 31, n estimators - 210

Cross validated accuracy fell to 87% with all parameters used
Only calculating n estimators raised the initial accuracy to 94,1%

Hyperpot with Grid Search

Used 704 combinations

This value was chosen as the combination that had a sensible calculation time with improved accuracy after validations

- n estimators from 30 to 250 with steps of 20
- max depths from 2 to 20 with steps of 5
- n estimators from 1 to 20 with steps of 5
- n estimators from 1 to 20 with steps of 5
- Best depth - 14 ,min samples - 44, minimum split - 31, n estimators - 210

Final results lowered the accuracy to 90,3%

Only using n estimator and max depth produced a increase in accuracy of 94,6%

Bayesian optimization

To validate performance we also tried the same dataset with Bayesian optimization:

Final 5 fold cross-validation score was 95,2%