

Exercise Sheet 2: Tree-based Methods

December 15, 2022

- 1) Random Forest
 - a) Predict the demand of all product-location-date combinations in `test.csv` using a Random Forest, e.g., from *scikit-learn*. Repeat the evaluations with this model.
 - b) Estimate the importances of the different features in your model averaged over the training.
- 2) Gradient Boosting
 - a) Predict the demand of all product-location-date combinations in `test.csv` using a Gradient Boosting method, e.g., from *scikit-learn*. Repeat the evaluations with this model.
 - b) Use one of the popular Gradient Boosting implementations like LGBM (e.g., `HistGradBoost` from *scikit-learn*) or XGBoost (python package *xgboost*). Repeat the evaluations with this model.
- 3) Use one of the two methods LIME or SHAP (python packages *lime* and *shap*) to go beyond feature importances averaged over the training and explain a bunch of individual predictions in terms of influences of the different features.