

Regression with a Tabular Media Campaign Cost Dataset

Playground Series - Season 3, Episode 11

Team#6 :

Алексеев Арсений

Егоров Максим

Dataset



Dataset - Media Campaign Cost Dataset

Train - 360k rows

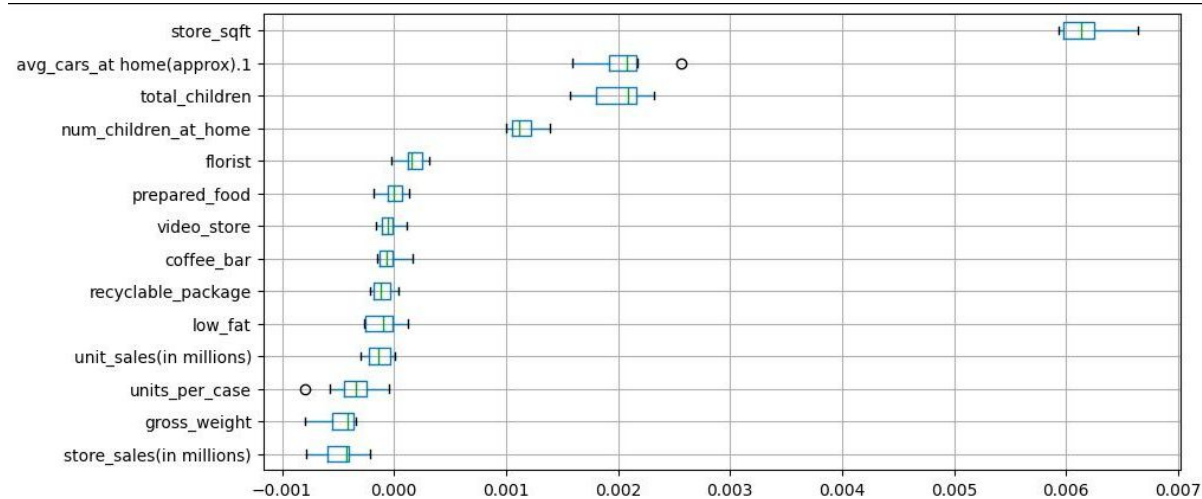
Test - 240k rows

Num_features - 15

Target - "Cost"

All features are numeric

Feature importance with lofo library





Baseline solution

CatBoost, data preprocessing, PolynomialFeatures augmentation



baseline_subm.csv

Complete (after deadline) · 14h ago

0.30121

0.30039





Updated solution

- 1) Pseudo-labeling
- 2) Voting, stacking over LightGBM, XGBoost, CatBoost
- 3) Hyperparameters tuning with automl



automl_filter_pslabel.csv

Complete (after deadline) · 1h ago

0.29668

0.29608

