

# models

November 17, 2025

## 0.0.1 Model Comparison

```
[1]: %load_ext autoreload
      %autoreload 2
```

```
[2]: from mld1_hw3.preprocessing import DataLoader
      from mld1_hw3.feature_engineering import build_feature_engineering_pipeline
      from mld1_hw3.experiment import Experiment, ExperimentConfig

      from xgboost import XGBRegressor
      from sklearn.ensemble import RandomForestRegressor
      from catboost import CatBoostRegressor
      import pandas as pd
```

```
[3]: df_train, df_test = DataLoader("../dataset").load()

      X_train = df_train.copy()
      y_train = X_train.pop("Price")
      X_test = df_test.drop(columns=["Price"])
```

### XGBoost

```
[4]: xgb_exp = Experiment(
      ExperimentConfig(
          name="xg-boost",
          pipeline=build_feature_engineering_pipeline(XGBRegressor())
      )
  )

  xgb_exp_result = xgb_exp.run(X_train, y_train, X_test)
```

```
[Experiment: xg-boost]
Cross-validating (5-folds)...
CV score: 0.1372 ± 0.0116
Training on full training set...
Creating submission on test set...
Submission created: artifacts/experiment-results/xg-boost.csv
Experiment complete
```

## Random Forest

```
[5]: random_forest_exp = Experiment(  
    ExperimentConfig(  
        name="random-forest",  
        pipeline=build_feature_engineering_pipeline(RandomForestRegressor()),  
    )  
)  
  
random_forest_exp_result = random_forest_exp.run(X_train, y_train, X_test)
```

```
[Experiment: random-forest]  
Cross-validating (5-folds)...  
CV score: 0.1527 ± 0.0146  
Training on full training set...  
Creating submission on test set...  
Submission created: artifacts/experiment-results/random-forest.csv  
Experiment complete
```

## CatBoost

```
[6]: cat_boost_exp = Experiment(  
    ExperimentConfig(  
        name="cat-boost",  
        pipeline=build_feature_engineering_pipeline(  
            CatBoostRegressor(  
                silent=True,  
                train_dir="./artifacts/catboost",  
                loss_function="MAPE",  
            )  
        ),  
    )  
)  
  
cat_boost_exp_result = cat_boost_exp.run(X_train, y_train, X_test)
```

```
[Experiment: cat-boost]  
Cross-validating (5-folds)...  
CV score: 0.1396 ± 0.0164  
Training on full training set...  
Creating submission on test set...  
Submission created: artifacts/experiment-results/cat-boost.csv  
Experiment complete
```

---

```
[7]: pd.DataFrame(  
    {  
        "Model": ["XGBoost", "Random Forest", "CatBoost"],
```

```

    "MAPE": [
        xgb_exp_result.cv_score,
        random_forest_exp_result.cv_score,
        cat_boost_exp_result.cv_score,
    ],
    "MAPE std": [
        xgb_exp_result.cv_std,
        random_forest_exp_result.cv_std,
        cat_boost_exp_result.cv_std,
    ],
},
).set_index("Model").sort_values(by="MAPE")

```

```

[7]:
      MAPE  MAPE std
Model
XGBoost    0.137150  0.011620
CatBoost    0.139573  0.016407
Random Forest 0.152714  0.014628

```