Mercedes-Benz Greener Manufacturing

Predict Car testing

Content

1. **Exploratory Visualization**
2. **Data Cleaning**
3. **Feature Engineering**
4. **Modelling & Evaluation**
5. **Ensemble Methods**
6. **Stacking Technique**
7. **Improve score using dimensionality reduction and Stacking technique**
8. **Apply kernel to Kaggle data**

EDA & finding best performing Kernel

1. **Exploratory Visualization – covered in notebook**

2. **Data Cleaning – covered in notebook**

3. **Feature Engineering**

   Converted some categorical features into numeric values using LabelEncoder

4. **Modelling & Evaluation**

   We choose 13 models and use 5-folds cross-validation to evaluate (LinearRegression,Ridge,Lasso,RandomForestRegressor,GradientBoostingRegressor,SVR,Linear SVR,ElasticNet,SGDRegressor,BayesianRidge,KernelRidge,ExtraTreesRegressor,XGBRegressor,LassoLarsCV) models

5. **Ensemble Methods**

   Average base models according to their weights, picked 6 best R-score models from point [4]

6. **Stacking Technique**

   Combine information from multiple predictive models to generate a new model. Often times the stacked model (also called 2nd-level model) will outperform each of the individual models due its smoothing nature and ability to highlight each base model where it performs best and discredit each base model where it performs poorly. Also, I have combine features generated from stacking and original features.

Using PCA Scaled features

▶ Applied Dimensionality reduction techniques and ran them on XGBoost model

▶ Also used Stacking technique to combine models (LassoLarsCV,KernelRidge,BayesianRidge,Lasso,GradientBoostingRegressor,XGB Regressor) and to generate new model

▶ Using both the models predicated values and the weights generated final score.

▶ Final score was 0.5684

▶ I could improve the score if I can spent more time on hyper tune the stacking models.

Using normal features

- Applied Dimensionality reduction techniques and ran them on XGBoost model

- Also used Stacking technique to combine models (LassoLarsCV,KernelRidge,BayesianRidge,Lasso,GradientBoostingRegressor,XGB Regressor) and to generate new model

- Using both the models predicated values and the weights generated final score.

- Final score was 0.5812

- I could improve the score if I can spent more time on hyper tune the stacking models.

Score improvement between PCA Scaled and normal feature

- The score shows good results with normal feature keeping kernel same.

- So using same kernel approach I will apply same on Kaggle competition data for Mercedes-Benz problem.

Using Kaggle's data Result

▶ Applied Dimensionality reduction techniques and ran them on XGBoost model

▶ Also used Stacking technique to combine models (LinearRegression, Ridge, ElasticNet, BayesianRidge, Lassolcv, Lasso) and to generate new model

▶ Using both the models predicated values and the weights generated final score.

▶ Final score was 0.56784 and rank is 1061 as per public score

▶ I could improve the score if I can spent more time on hyper tune the stacking models.

Submission and Description	Private Score	Public Score	Use for Final Score
stacked-models.csv 3 days ago by Tejas Magia version 1	0.54851	0.56784	☐