Analytics Vidhya Job-a-thon

- April 2022

Dileep Kumar
Senior Machine Learning Engineer, Quantiphi Analytics

Objective

The main objective of the problem is to develop the machine learning approach to forecast the demand of car rentals on an hourly basis.

Evaluation metric

The evaluation metric for this hackathon is RMSE score.

Approach

1. Data Ingestion: After downloading the data from Analytics Vidhya, read using pandas
2. Feature engineering using date columns and demand columns
3. Checking correlation of feature vector with target vector
4. Selecting XGboost with default hyperparameter as base model
5. Trying out different ML models.
6. Creating stacked ensemble model
7. Hyperparameter tuning using grid search for each model used in stacked ensemble model
8. Checking performance on test data using analytics vidhya portal

Which Data-preprocessing / Feature Engineering ideas really worked? How did you discover them?

1. Firstly as a feature engineering, created extra features like dayofweek, quarter, month, monthofquarter, dayofyear, dayofmonth, weekofyear, year, along with hour and time shifted demand columns.
2. Tested correlation of each feature column with demand column, time shifted demand columns were highly correlated with demand, so excluded from feature vector to avoid overfitting.

What does your final model look like? How did you reach it?

1. Tested different ML models starting with XGBoost as base model, mean_squared_error as loss function
2. LSTM model is tested with expectation of overcoming of xgboost model but it didn’t perform as per expectation.
3. Tried Random forest, XGboost with custom hyperparameter, and LGBM which outperformed previous implementation, however, it didn’t give much benefit.
4. Tried stacking XGboost, Random Forest, LGBM, and LSTM. It outperformed all other previous models
5. I optimized each model’s hyperparameter using gridsearch and final RMSE on test data was