Team ID: PTID-CDS-JAN22-1265

Project ID: PRCL-0027 Forecasting Spare parts inventory

Members:

1. Atharvdesh Pande

2. Nazeef Shaikh

3. Lukmanul Hakeem M

4. Sahil

5. Sumit Jangale

Project Analysis:

The project goal of this work is to build a predictive model for spare-parts inventory forecast. As dataset is has many output class names of spare-parts, it seems to be complicated to build model including all spare-parts categories. Therefore we had build individual forecasting models only for top 3 spare-parts category using Auto Regressive (AR) statistical estimator. These models will forecast spare-part count or quantity required for a given day. Engine oil forecasting model have AIC score 1659.68, Chain Lubrication forecasting model have AIC score 1675.19 and General Service forecasting model have AIC score 1466.41.

Workflow:

For this project we use data from a given database using SQL. This dataset is somewhat complicated to build a model upon it. So we tried with two approaches to achieve project goal. In first approach we tried to build a multi-class classifier model for the top seven mostly repeated spare-parts categories. In second approach we tried to build three time-series forecast models for top three mostly repeated categories. For the two approaches, prepared different sets of dataset accordingly from the loaded dataset.

In first approach, dataset consists of top seven classes were used. Tried with different scikit-learn, ensemble learning and ANN techniques by using several feature preprocessing and feature selection methods. But none of these give a good performance score. All we got 0.25 accuracy score.

Since first approach is bad, tried to do second approach. Here we use time-series dataset of top three classes where variables are date and counts (or quantity). For each dataset, tried with AR, ARIMA, SARIMA and FB-Prophet. All models are not good enough as the dataset doesn’t have any significant correlation coefficient at the lower lags.. But AR model with higher order that have lower AIC and BIC measures able to generate new output (counts of spare-part) with given dates. So we choose Auto Regressive time-series model for the three classes.

NOTE: Uploaded documents only contain notebook, project summary and main dataset files. It is because of maximum file number constraint (5 files). Rest of required dataset files for approach 1 and 2, and other files can be generated by running scripts in the notebook.

Files:

* ‘service\_data.csv’: dataset csv file created by saving data from database.
* ‘part1\_data\_loading\_data\_preparation.ipynb’: This file include data loading from database and preparing data for Approach 1 and Approach 2.
* ‘part2\_approach1\_classification\_model.ipynb’: Implementing multi-class classification.
* ‘part2\_approach2\_AR\_ARIMA\_SARIMA\_FBProphet.ipynb’: implementing time-series modeling.