

OEM Supply Chain Visibility Dashboard Documentation

1 Overview

The Original Equipment Manufacturers (OEM) Supply Chain Visibility Dashboard is designed to enable the real-time supply chain visibility through data analytics. The prediction model is built as the analytic tool based on the historical transactional data. Furthermore, the prediction model can be used to infer the future commitments and to improve the prediction results of current transactional information. In general, the dashboard has the following capabilities:

- View the raw transactional data information.
- Build different prediction models with advanced capabilities and parallel mode.
- Visualize the prediction results in multiple interactive methods and presentation formats.
- Make prediction of future potential transaction with high confidence.

2 Data

2.1 Original Dataset

The historical transactional dataset for analyzing is a purchase order (PO) dataset provided by a major OEM in the US. It contains 53,105 closed POs and 3,533 open POs. The time span is from April 2016 to May 2017. We choose 7 predictors including categorical and continuous variables to predict the PO delivery time in days. We also incorporate an advanced variable selection method to reduce the levels of categorical variables in order to shorten the model training time without loss of model capabilities in term of the prediction performance. Table 1 shows the procedure of the variable selection method. The levels of categorical variables are reduced substantially. But the prediction performance is still good enough.

Table 1: Levels in each categorical variable during different steps of variable selection method

Variable names	Original number of levels	After level selection and combination (Step 1)	After using PCA for dimension reduction (Step 4)
Vendor	881	108	15
Material Number	4,447	1,447	15
Plant ID	43	17	15
Material Group	1,596	398	15

Three different machine learning models with separate parameter settings are investigated upon the PO dataset. The results are measured using absolute prediction errors in days compared with the supplier estimates. Table 2 shows the prediction results, and the numbers with best performance are showed in bold.

Table 2: PO delivery time prediction errors

Approach	Absolute prediction error (in days)				
	25 th percentile	50 th percentile	75 th percentile	90 th percentile	95 th percentile
Supplier estimates	2.00	6.00	13.00	27.00	41.00
Linear regression	3.62	7.64	13.25	23.02	35.93
RF	2.63	5.89	11.25	20.67	31.70
QRF conditional mean	2.38	5.36	10.10	18.01	26.45
QRF conditional median	1.00	4.00	8.50	17.00	27.00
QRF conditional 1 st quantile	1.00	4.25	11.00	23.00	35.00
QRF conditional 3 rd quantile	2.00	6.00	12.00	20.75	29.00
Linear regression	2.90	6.28	12.06	21.65	32.62
RF	2.28	5.14	9.98	18.40	28.08
QRF conditional mean	2.17	4.87	9.37	17.16	25.54
QRF conditional median	1.00	4.00	8.00	16.00	26.00
QRF conditional 1 st quantile	1.00	4.00	10.00	22.00	34.00
QRF conditional 3 rd quantile	2.00	6.00	11.00	19.50	28.25

2.2 Dummy dataset

In the actual deliverable software of the dashboard, we use a substituting dummy dataset to demonstrate the capability of the visibility tool. Please note that the results and performance are not the same for the dummy dataset and the original dataset. They are substantially better in our actual system. Regarding the dummy dataset, it has fewer rows, 925 records in total, for open POs and close POs. The column names and types are maintained. Finally, some sensitive information is anonymized.

3 Available Features

3.1 Database Access View

In this view, the users can specify the database and available tables within it to be used by the training process. After choosing the database, an authentication step is needed. The users can also view the specific table by choosing the column names in the multiple selection box. The number of rows to be displayed is adjustable, and the users can search for useful information in the table. Finally, the users need to specify the response variable and predictor variables. At the top of the page, all current information is displayed to help the users keep track of the dashboard status. Figure 1 shows a screen shot of this page.

Key functions: Specify the database and table, view the table, specify the response variable and predictor variables.

Figure 1: Database Access view

OEM Supply Chain Visibility Dashboard

Database Access

Prediction Model Training

Current Predictions

Future Predictions

You have selected: PO

You have selected: Purchase Order Data

Predictor Variables: Vendor.Num, Material.Number, Plnt, Material.Group, ABC.Indicator, Pdt, Quantity

Response variable: PO.Cycle.Time

Please choose the database: PO

Please select table: Purchase Order Data

Confirm View Select

Variables to show in the table below

Purchasing.Document Item Vendor Vendor.Num Vendor.Name

Material.Number Plnt Short.Text Material.Description

Material.Group ABC.Indicator

Response Variable

PO.Cycle.Time

Select

Predictor Variables

Vendor.Num Material.Number Plnt Material.Group ABC.Indicator Pdt Quantity

Select

Show 10 entries

Search:

	Purchasing.Document	Item	Vendor	Vendor.Num	Vendor.Name	Material.Number	Plnt	Short.Text	Material.Description	Material.Group	ABC.Indicator
1	4502719166	10	87405	223099	FUIZOuplialqaaazaA,Ad	6090981	3288	UUAAOEZUZ P601 OU SU			A
2	4502719166	20	87405	223099	FUIZOuplialqaaazaA,Ad	6094996	3288	YAEYOOF0 4K37 (IYI) OAV OO			B
3	4502719166	30	87405	223099	FUIZOuplialqaaazaA,Ad	6094996	3288	YAEYOOF0 4K37 (IYI) OAV OO			B
4	4502719166	40	87405	223099	FUIZOuplialqaaazaA,Ad	6094996	3288	YAEYOOF0 4K37 (IYI) OAV OO			B

Basic operations:

1. Use the dropdown selection box to choose the desired database to use.
2. Press confirm to submit your selection and put in your name and password for authentication.
3. After successful authentication, the table selection box will pop up and you should use it to identify the table you want to analyze.
4. The two selections, View and Select, are available for the selected table. Click on View first and the data table will show up with further options.
5. Use “Variable to show in the table below” box to select multiple variables that you want to view in the table below. You can order the table by different column and do advanced search use the search box in the upper right corner with the data table area. Additionally, you can select how many rows to be shown in the data table.
6. After you have decided the table to use, click on Select to confirm your selection.
7. Finally, you should use the response variable selection box and the predictor variables selection box to specify the response and predictors in your prediction model. Note that you are forced to select only one response and any number of predictors.

3.2 Prediction Model Training View

The users can select the training model to use and all the parameters used by the model. Additionally, the users can choose whether to use parallel training mode and variable selection method. If the variable selection method is selected, the users need to specify number of principal components to be extracted using principal component analysis method and number

of levels to be kept in each categorical variable. After training the model, the detailed information about the model is shown on the top of the page. To let the users keep track of the training process, a progress bar on the bottom right will indicate the model training progress. Key functions: Specify the parameters of the model, choose whether to use parallel training mode and variable selection method or not, show training progress.

Figure 2: Model Training View

OEM Supply Chain Visibility Dashboard

Database Access

Prediction Model Training

Current Predictions

Future Predictions

Current table in use: Purchase Order Data from PO

Predicting PO.Cycle.Time using Vendor.Num, Material.Number, Plnt, Material.Group, ABC.Indicator, Pdt, Quantity as predictors.

Current model in use: Quantile Regression Forests with variable selection in parallel training mode.

Training model to use: Quantile Regression Forests

Number of Trees: 120

Number of PCs: 5

Number of Variables sampled at each split: 2

Number of Levels kept in categorical variables: 15

Minimum size of terminal nodes: 5

Using variable selection method

Using parallel training mode

Build the model

Model training in progress This may take a while...

Basic operations:

1. In this view, you will train the model and set different parameters. It is suggested to use the default settings. First select the desired prediction models to train in the dropdown selection box.
2. Specify whether to use variable selection method and parallel training mode.
3. For Random Forest and Quantile Regression Forest, set the number of trees to grow, number of variables sampled at each split and minimum size of terminal nodes.
4. For variable selection method, set the number of Principle Components to be exacted from the dataset and number of levels kept in each categorical variable.
5. Finally, click on build the model to train it.

3.3 Current Predictions View

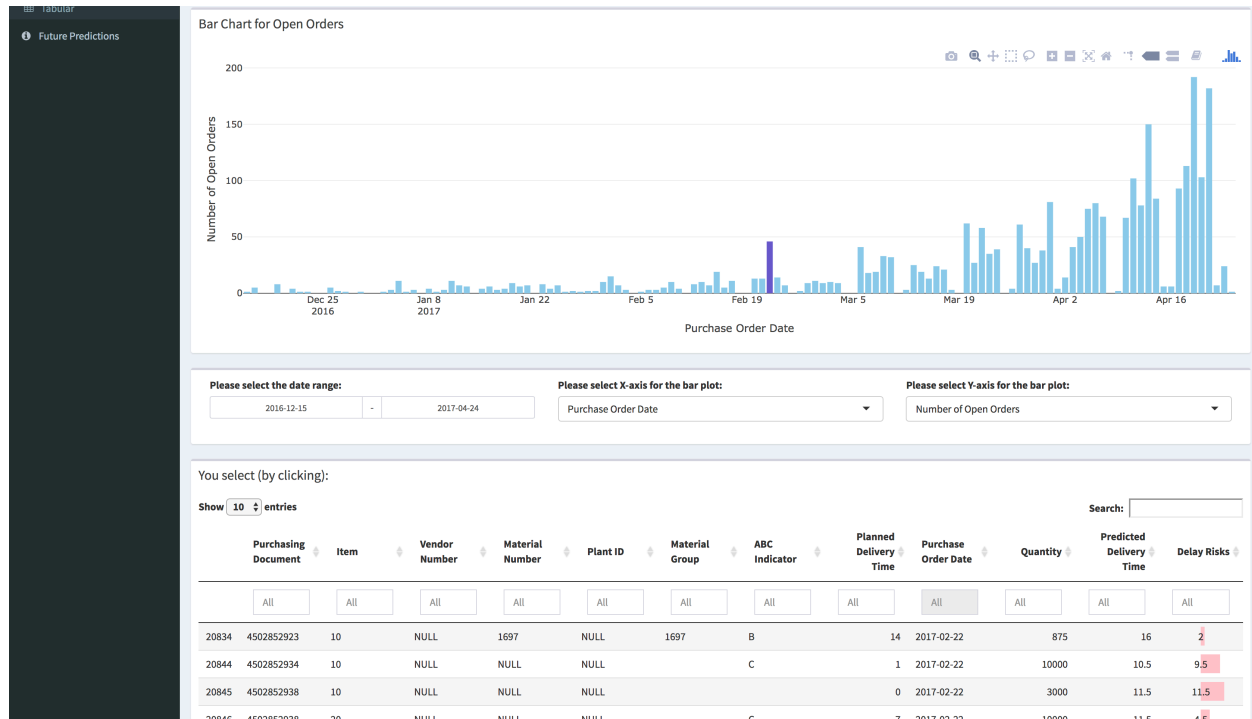
3.3.1 Delivery Time Predictions Page

A bar chart for open orders is shown on this page. The users can hover on each bar using their mouse, and more information about that bar will pop up as hover info. Additionally, the users can click on each bar. The bar will change color and the detailed information will be shown in

the data table below. The users can select different time range, x-axis and y-axis for the bar plot.

Key functions: View current open orders, select parameters for the bar plot, multiple interactive methods.

Figure 3: Delivery Time Prediction view



Basic operations:

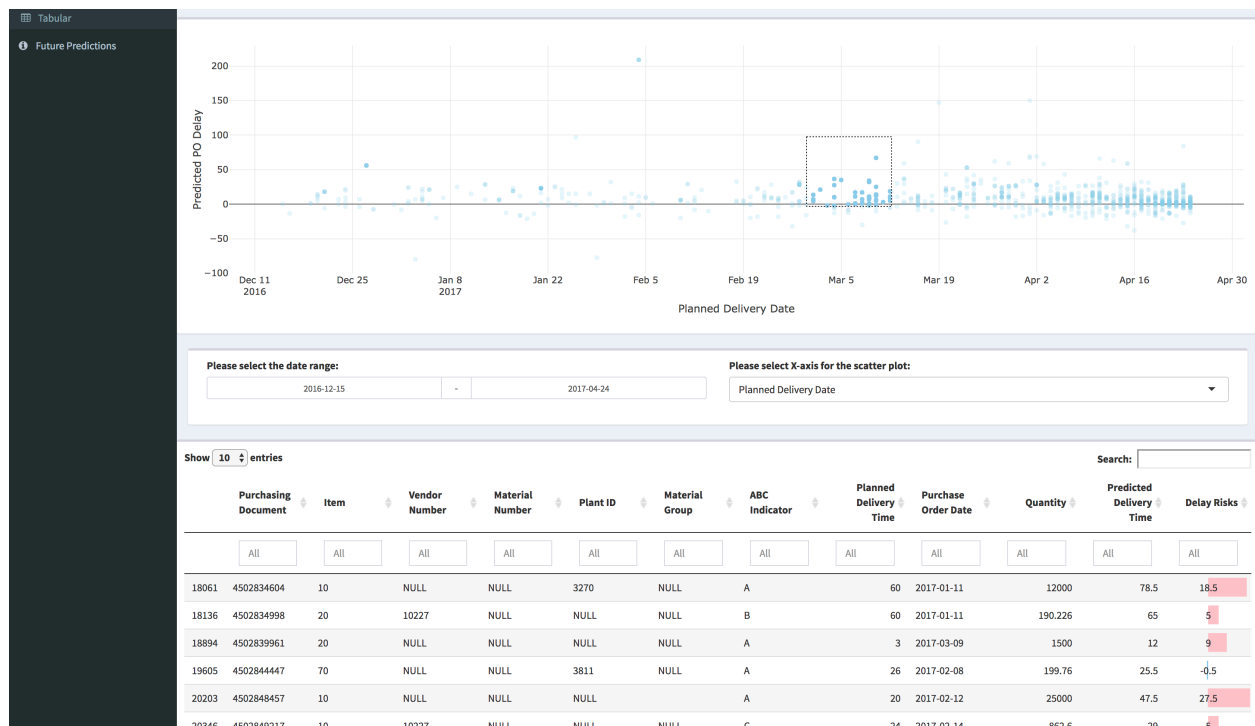
1. Select the date range you want to view.
2. Select the x-axis and y-axis using the two dropdown selection boxes.
3. Click on a bar, then the detailed information will show up in the table below.
4. Sort or search using the table advanced options.

3.3.2 Delay Risk Identification Page

A scatter plot shown delay risks is used to help the users to identify the most risky open orders. Delay risk is defined as the absolute difference between predicted delivery time and planned delivery time. The users can specify the date range and x-axis of the plot. Similar to the bar plot, the user can hover on each dot to view more information and select a rectangular area to show detailed information in the data table below.

Key functions: View delay risks, select parameters for the bar plot, multiple interactive methods.

Figure 4: Delay Risk Identification view



Basic operations:

1. Select date range you want to view.
2. Select the x-axis for the scatter plot above.
3. Select a rectangular area using the cursor in the scatter plot, then the detailed information of the selected area will pop up in the data table below.
4. Sort or search using the table advanced options.

3.3.3 Tabular Page

The users can view open purchase orders and closed purchase orders using this page. The detailed information of each PO will be shown in the table format. The users can search and specify different parameters for the table.

Key functions: View open/closed POs.

Figure 5: Data Table view

OEM Supply Chain Visibility Dashboard

Database Access

Prediction Model Training

Current Predictions

Delivery Time Predictions

Delay Risk Identification

Tabular

Future Predictions

Current table in use: podata from PO

Predicting PO.Cycle.Time using Vendor.Num, Material.Number, Plnt, Material.Group, ABC.Indicator, Pdt, Quantity as predictors.

Current model in use: Quantile Regression Forests with variable selection in parallel training mode.

Closed POs

Show 10 entries

Search:

	Purchasing Document	Item	Vendor Number	Material Number	Plant ID	Material Group	ABC Indicator	Planned Delivery Time	Actual Delivery Time	Purchase Order Date	Delivery Date	Quantity	Predicted Delivery Time
	All	A	A	AI	AI	AI	AI	AI	AI	AI	AI	AI	AI
1	4502719166	10	223099	NULL	3288		A	12	37	Apr 24, 2016	May 31, 2016	700	35.0613220178749
2	4502719166	20	223099	NULL	3288		B	0	37	Apr 24, 2016	May 31, 2016	2000	36.8897694800476
3	4502719166	30	223099	NULL	3288		B	0	37	Apr 24, 2016	May 31, 2016	2000	35.9561008138369
4	4502719166	40	223099	NULL	3288		B	0	37	Apr 24, 2016	May 31, 2016	2000	35.5572728877187
5	4502719166	50	223099	NULL	3288		B	0	37	Apr 24, 2016	May 31, 2016	2000	36.5227489600341
6	4502719166	60	223099	NULL	3288		B	0	37	Apr 24, 2016	May 31, 2016	2000	35.799350066277
7	4502719166	70	223099	NULL	3288		B	0	37	Apr 24, 2016	May 31, 2016	2000	37.0780626014493
8	4502719166	80	223099	NULL	3288		A	5	37	Apr 24, 2016	May 31, 2016	2000	36.0719372126134
9	4502719166	90	223099	NULL	3288		A	5	37	Apr 24, 2016	May 31, 2016	2000	37.8172391696757
10	4502719166	100	223099	NULL	3288		A	5	37	Apr 24, 2016	May 31, 2016	2000	36.8632778174184

Showing 1 to 10 of 43,276 entries

Previous 1 2 3 4 5 ... 4328 Next

Basic operations:

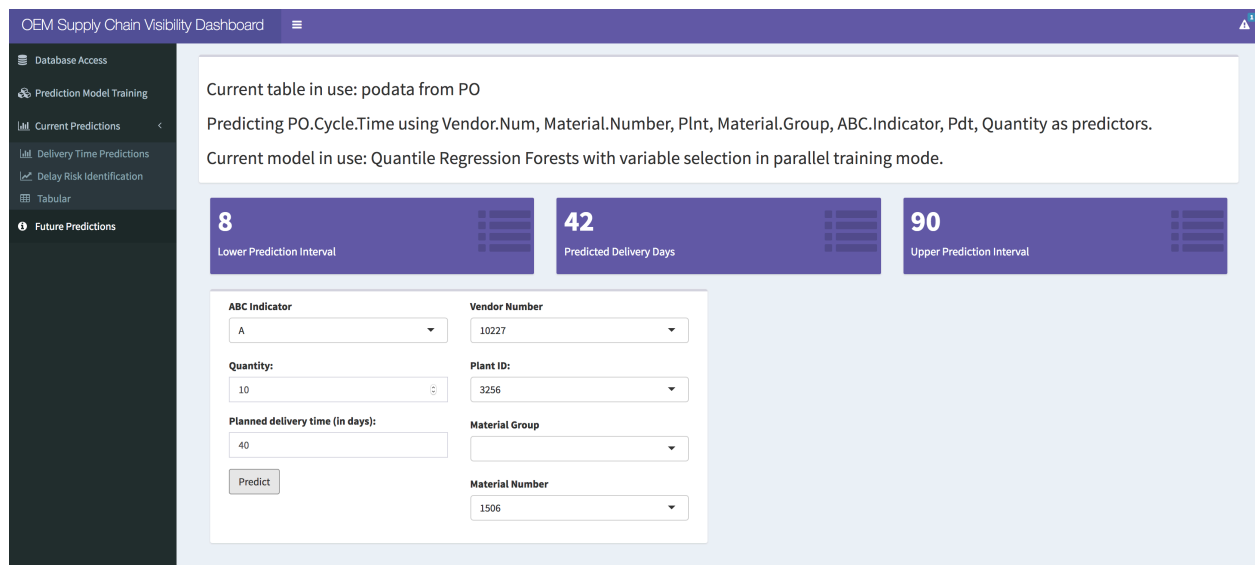
1. Select Closed POs or Open POs using the dropdown menu.
2. Sort or search using the table advanced options.

3.4 Future Prediction View

This view is to help users make predictions on future purchase orders. The users can specify the order details using the input area. The inputs correspond to the parameters chosen in the model training view. The results will be shown as a confidence interval and a prediction value.

Key functions: Make predictions on future purchase orders.

Figure 6: Future Prediction view



Basic operations:

1. Based on the model, different parameters needed to be set using the selection box.
2. Select each purchase order settings using the dropdown menus.
3. Click on predict, then the results will show up in the top boxes.

4 Implementation Detail

The installation steps are in the README file. Please follow the steps in the documentation to install the whole package to your local machine. The whole software resides in a docker container. Hence please make sure your local machine has docker installed.

4.1 MySQL Database

The software uses MySQL database to store all the data. Please make sure your local machine is compatible with MySQL server. The current example software has a test MySQL database installed. If you want to use your own database, please import your data to MySQL database and configure the corresponding connection in the web files which will be explained in detail later.

4.2 R Files

4.2.1 global.R file

This file contains login information of the database and Javascript functions for rendering the plots. Please change the database settings to connect to your local database correspondingly. However, do not change the rendering functions without specific needs.

4.2.2 ui.R

This file contains all the front-end visualization functions. Header menu is defined in header function which includes title and tile width, etc. Side bar menu is defined in sidebar function

which defines all the tabs and subtabs of the dashboard. Finally, the body function defines each page layout individually.

4.2.3 server.R

All the major front-end and back-end interactions are defined in this file.

Line 2-74 define the database authentication function. If login successfully, the following operations will be granted.

Line 140-405 define the core functions of model training process. Please do not change this part of codes without special assistant. The built-in models are Linear Regression, Random Forest and Quantile Regression Forest. The default model is already set to QRF predicting conditional median using parallel training mode. If you would like to change the default model, please refer to the ui.R file. Note that the parallel training mode require multi-thread programming. Please make sure your operation system has multi-core and multi-thread support.

Line 466-522 define the bar plot of current prediction visualization. Line 554-600 define the scatter plot of current prediction visualization. These functions have already defined the whole visualization process. Please do not change this part without special assistant.

Line 632-681 define the future prediction tab. The model used in the prediction function is already defined in the model training function. If you want to change the width of confidence interval, please modify 'what' parameter in line 643 and line 652. To predict other quantiles, please also modify this parameter.