Analytics

Oracle Analytics Cloud & Server

Applying Logistic Regression Model in Oracle Analytics

Laxminag Mamillapalli | April 28, 2021

Oracle Analytics Cloud (AC) enables data analysts to train Machine Learning (ML) Models and score their datasets. These models are usually applied on new datasets to make predictions. In previous blogs (Logistic Regression: Understanding model hyper-parameters), we demonstrated how Logistic Regression Model can be trained and how its hyper-parameters can be tuned further to improve accuracy. In this post, we will explore how to apply the logistic regression model on a new dataset. We will use same Titanic reated Dataset that was used in previous blog for this demonstration.

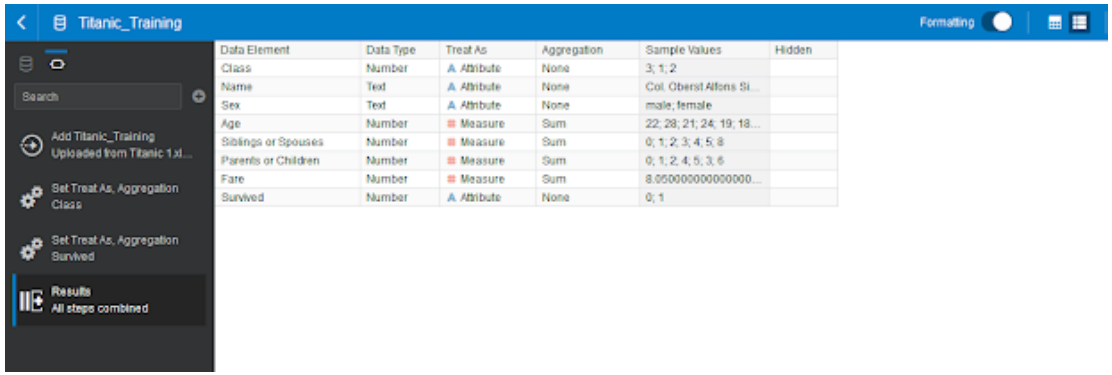
Case Study

Titanic Dataset is split into two sets. First set contains a list of 750 of Titanic passengers and some of their details. This dataset is named as Titanic_Training. The second dataset is call Titanic_Prediction which has 137 Records. Each row represents one person in both the datasets. We first create a Logistic Regression Model

that predicts if the passenger survived the disaster or not. In this post we apply this the model on a new dataset.

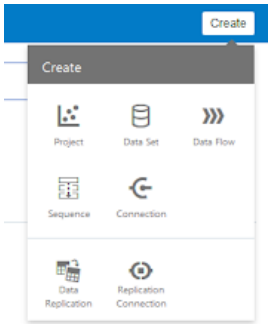
Step 1: Generation of Model

Let us upload first Dataset to train the model. (Titanic_Training Dataset)



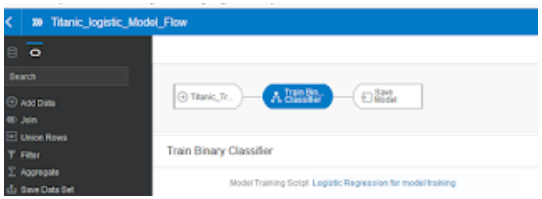
Step 2: Creating Logistic Regression Model

Let's create a Data Flow for that

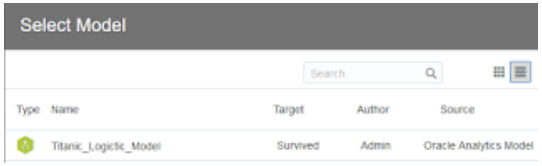


Train Binary Classifier Model

If you like to learn more about Hyper Parameter Tuning and Model Generation please refer to the previous blog.

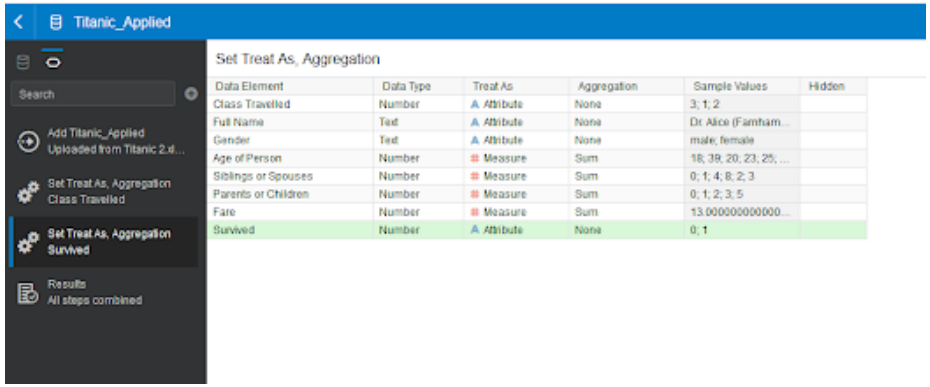


Save the Model as 'Titanic Logistic Model'

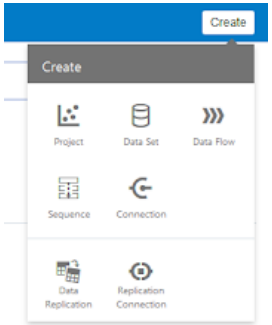


Step 3 : Applying the Model

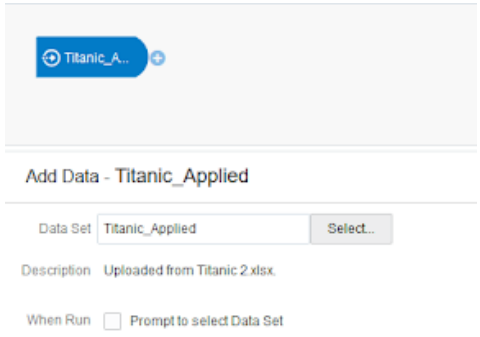
Let us upload the new dataset on which our model is applied. We require same/similar column for the model to be applied. The column names can be different. In the case below, the column names are different for demonstration purpose.



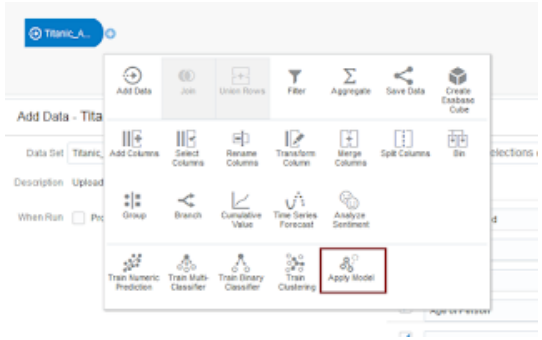
Let us create a new Data Flow apply the Model.



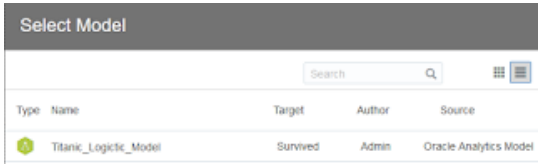
Add the Dataset (Titanic_Applied) to the Data Flow.



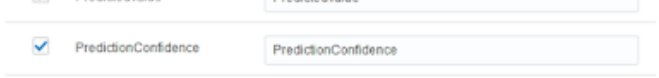
Select the Apply Model



Select the Model that was previously generated.



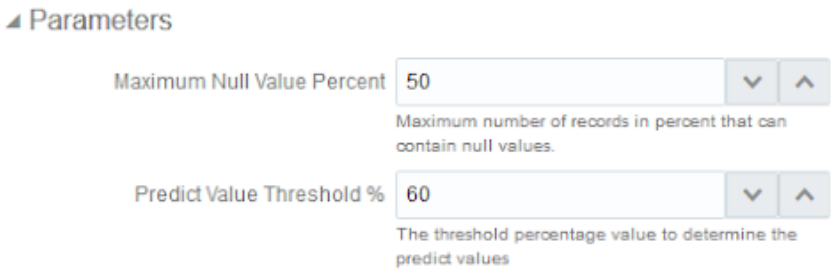
After adding the dataset, specify the outputs required in the final dataset. In this case we are choosing both 'PredictedValue' and 'PredictedConfidence'. You can deselect or change column names as required.



Modify the Parameters to suit your needs. In this case You can adjust 2 parameters:

1. Maximum Null Value Percent: This helps you to apply the model only if the maximum number of records that have Null Values are less than 50%.

2. Predict Value Threshold %: The cut off value (60%) at which your model is going to classify is specified in this parameter.



Once you have specified the Model, check if the columns match the training data. In this case, OAC could not match the training columns - 'Class', 'Sex' and 'Age'. Match the corresponding columns.

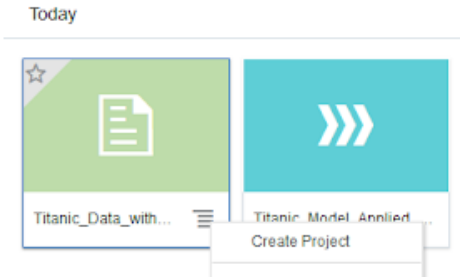
Before matching columns:

After matching columns:

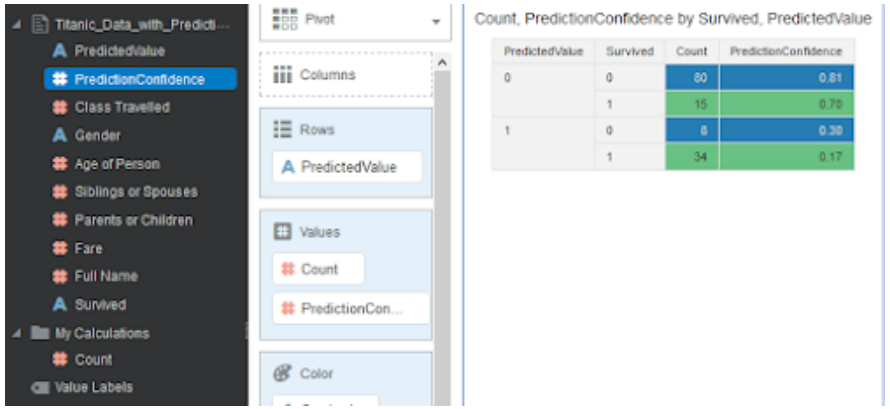
Save the Data (Titanic_Data_with_Predictions) and Run the Workflow.

Step 4: Working with the predictions

Go the main page and Open the newly saved dataset.



You can use the predictions which are shown as columns in the newly created dataset. Let us create a simple pivot with PredictedValue, Survived, Count and PredictionConfidence to check the accuracy of the predictions. In our case Model accuracy is around 83.2% (114/137)



Conclusion

OAC allows you to apply logistic regression model to any number of datasets.The model building and model applying process via dataflows is simple and intuitive process. It is easy to apply the model over a dataset, and then visualize the output as per your requirements.

Thank you for reading our blog!

Laxminag Mamillapalli

Previous Post

Create a Private Access Channel in Oracle Analytics

Peter Monteiro | 3 min read

Next Post

Migrating File-based Data from Oracle Analytics Server to Oracle Analytics Cloud

Amarpreet Nagra | 1 min read