

# 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 related Dataset that was used in previous blog for this demonstration.

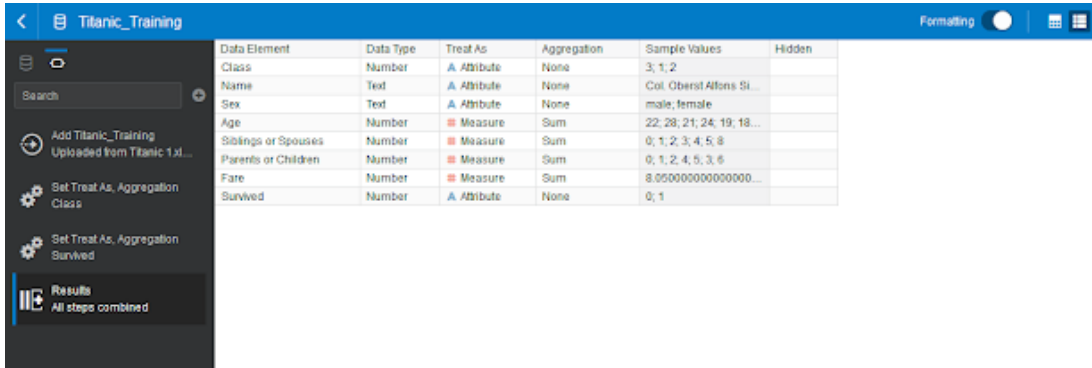
### Case Study

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

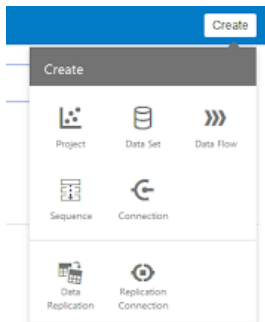
Let us upload first Dataset to train the model. (Titanic\_Training Dataset)



Data Element	Data Type	Treat As	Aggregation	Sample Values	Hidden
Class	Number	A. Attribute	None	3; 1; 2	
Name	Text	A. Attribute	None	Col. Oberst Alfons Si...	
Sex	Text	A. Attribute	None	male, female	
Age	Number	M. Measure	Sum	22; 28; 21; 24; 19; 18...	
Siblings or Spouses	Number	M. Measure	Sum	0; 1; 2; 3; 4; 5; 8	
Parents or Children	Number	M. Measure	Sum	0; 1; 2; 4; 5; 3; 6	
Fare	Number	M. Measure	Sum	8.000000000000000...	
Survived	Number	A. Attribute	None	0; 1	

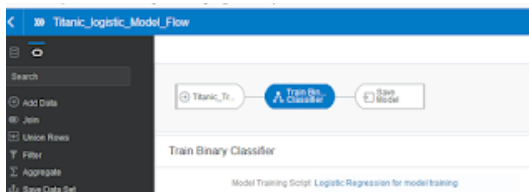
## Step 2: Creating Logistic Regression Model

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'

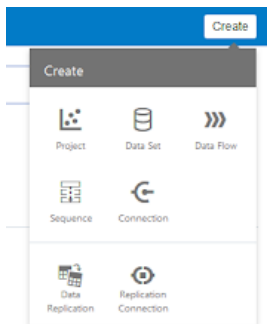
Select Model				
<div>Search <input type="text"/></div>				
Type	Name	Target	Author	Source
	Titanic_Logistic_Model	Survived	Admin	Oracle Analytics Model

### 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.

Titanic_Applied						
Set Treat As, Aggregation						
Data Element	Data Type	Treat As	Aggregation	Sample Values	Hidden	
Class Travelled	Number	A Attribute	None	3; 1; 2		
Full Name	Text	A Attribute	None	Dr. Alice (Famham...		
Gender	Text	A Attribute	None	male; female		
Age of Person	Number	M Measure	Sum	18; 39; 20; 23; 25; ...		
Siblings or Spouses	Number	M Measure	Sum	0; 1; 4; 8; 2; 3		
Parents or Children	Number	M Measure	Sum	0; 1; 2; 3; 5		
Fare	Number	M Measure	Sum	13.000000000000...		
Survived	Number	A Attribute	None	0; 1		

Let us create a new Data Flow apply the Model.



Add the Dataset (Titanic\_Applied) to the Data Flow.

Titanic\_A...

---

Add Data - Titanic\_Applied

Data Set

Description

When Run ☐ Prompt to select Data Set

Select the Apply Model

Titanic\_A...

---

Add Data - Tita

Data Set

Description

When Run ☐ Prompt to select Data Set

Select the Model that was previously generated.

Select Model				
<input type="text" value="Search"/> <input type="button" value="Q"/> <input type="button" value="Grid"/> <input type="button" value="List"/>				
Type	Name	Target	Author	Source
	Titanic_Logistic_Model	Survived	Admin	Oracle Analytics Model

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.

Model

Outputs

Create	Output	Column Name
<input checked="" type="checkbox"/>	PredictedValue	PredictedValue

☒ PredictionConfidence

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.

#### Parameters

Maximum Null Value Percent     
Maximum number of records in percent that can contain null values.

Predict Value Threshold %     
The threshold percentage value to determine the predict values

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:

Inputs

Model	Input
99 Class	Select a column
ab Sex	Select a column
99 Age	Select a column

Only showing model columns that were not found [Show all columns](#)

After matching columns:

Inputs

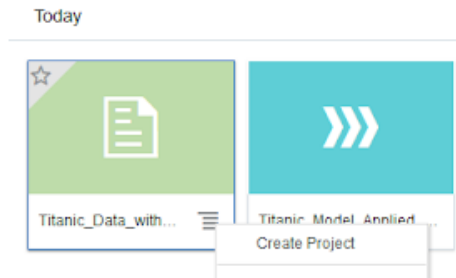
Model	Input
99 Class	Class Travelled
ab Sex	Gender
99 Age	Age of Person

Only showing model columns that were not found [Show all columns](#)

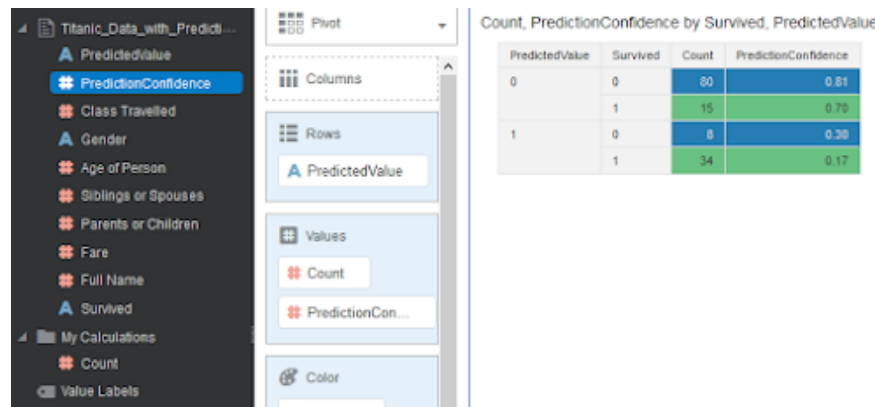
Save the Data (Titanic\_Data\_with\_Predictions) and Run the Workflow.

#### 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)



PredictedValue	Survived	Count	PredictionConfidence
0	0	80	0.81
	1	15	0.79
1	0	8	0.38
	1	34	0.17

#### 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](#)

[Next Post](#)

## Create a Private Access Channel in Oracle Analytics

[Peter Monteiro](#) | 3 min read

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

[Amarpreet Nagra](#) | 1 min read

2021 © Oracle

[Site Map](#)

[Privacy / Do Not Sell My Info](#)

[Cookie Preferences](#)

[Ad Choices](#)

[Careers](#)