

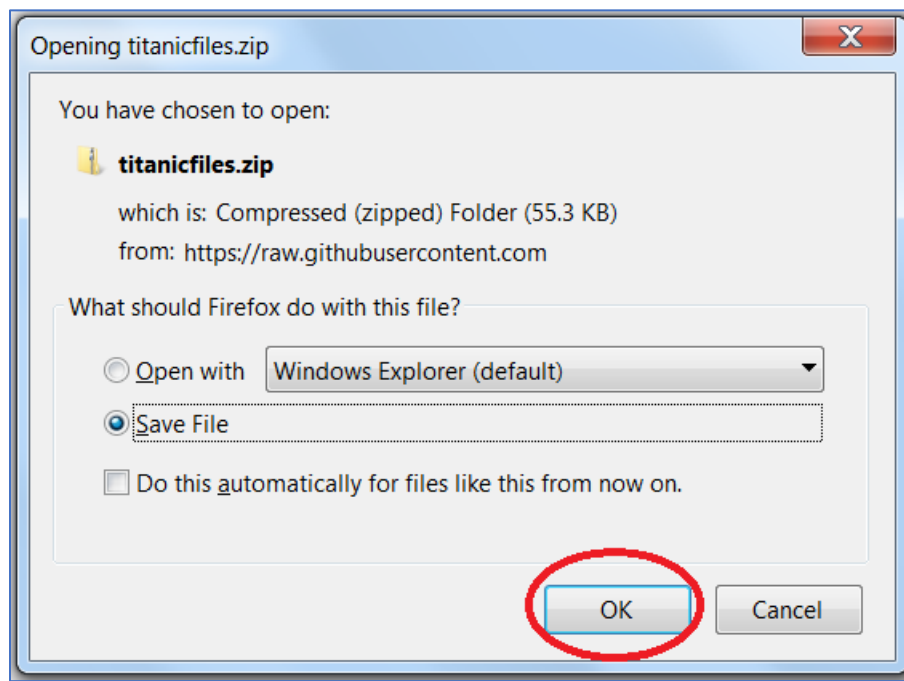
Watson Machine Learning Overview

This lab will introduce the Watson Machine Learning capability using the Titanic dataset. The lab will consist of the following steps:

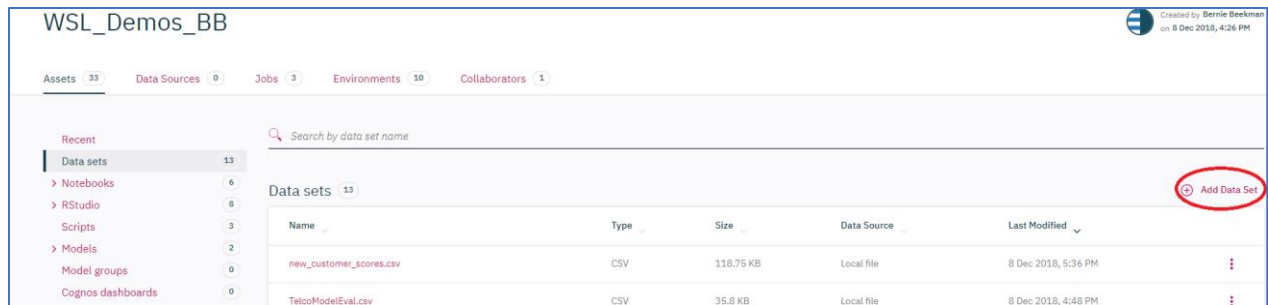
1. Adding a data asset to the DSXL project
2. Creating a Model to predict whether a person would survive
3. Testing the Model

Step 1: Adding a Data Asset to the project

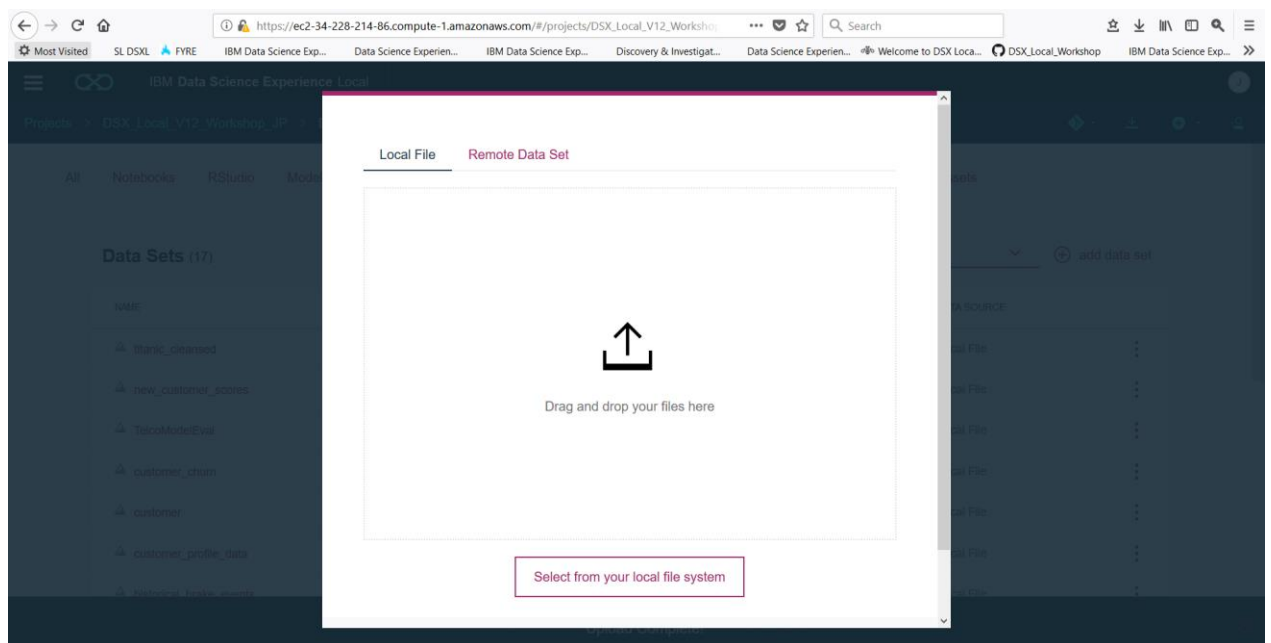
1. Download the Titanic data file from the following location by clicking on the link [Titanic Data](#) and following the instructions below.
2. Click on the **OK** button in the pop-up dialog.



3. Navigate to the directory where the file has been downloaded. Unzip the titanicfiles.zip file. There should be three files (1) titanic_cleansed.csv, (2) titanic.csv, and (3) titanicr.csv. You will use the **titanic_cleansed.csv** for this lab.
4. In your Watson Studio Local project go to **Data Sets** and select **add data set**

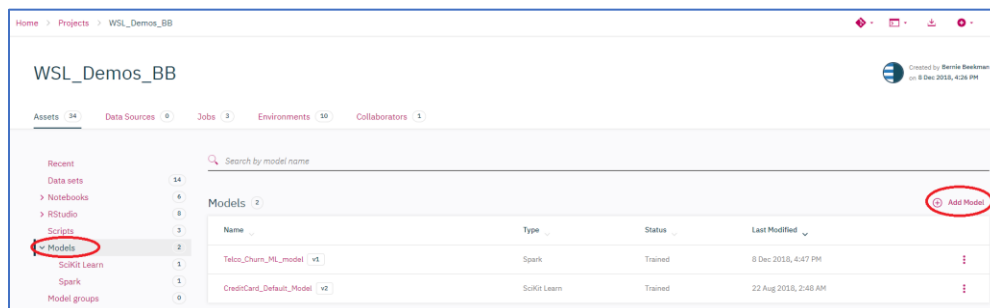


5. Browse or drag the **titanic_cleansed.csv** file



Step 2: Create a Model to predict survival

1. Select the **Models** link and **Add Model**.



3. Enter a model **Name** (eg Titanic), optionally a **Description**, select **model type** of **Machine Learning** and select **Method** of **Manual**. Click on **Create**.

Projects > WSL_Demos_BB > Add Model

Create Model

Blank From File

Name *
Titanic

Description
Model description

Model type *
☒ Machine Learning ☐ Decision Optimization

Method *
☐ Automatic ☒ Manual

Automatic: Prepare my data and create a model automatically.

Manual: I will prepare my data and select which models to train.

Cancel Create

4. Click on the **titanic_cleansed.csv** and click on **Next**

Projects > WSL_Demos_BB > Titanic

Select data asset

The model builder currently supports CSV files & Remote Data Sets.

NAME	TYPE	SERVICE
<input checked="" type="radio"/> titanic_cleansed.csv	CSV	File System
<input type="radio"/> new_customer_scores.csv	CSV	File System
<input type="radio"/> TelcoModelEval.csv	CSV	File System
<input type="radio"/> customer_churn.csv	CSV	File System
<input type="radio"/> customer.csv	CSV	File System
<input type="radio"/> historical_brake_events.csv	CSV	File System
<input type="radio"/> CUST_HISTORY.csv	CSV	File System

Close Next

5. Select **Add a transformer** to see all available transformers. **Cancel** and use the configured **Auto Data Preparation** transformer. Select **Next**.

Projects > WSL_Demos_BB > Titanic

Select Data

Prepare data set

Prepare

Train

Evaluate

pclass	survived	name	sex	sibsp	parch	ticket	fare	embarked	Age_Bucket
1	1	Allen, Miss. Elisabeth Walton	female	0	0	24160	211.3375	S	3
1	1	Allison, Master. Hudson Trevor	male	1	2	113781	151.55	S	0
1	0	Allison, Miss. Helen Loraine	female	1	2	113781	151.55	S	0
1	0	Allison, Mr. Hudson Joshua Creighton	male	1	2	113781	151.55	S	3
1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	1	2	113781	151.55	S	3
1	1	Anderson, Mr. Harry	male	0	0	19952	26.55	S	4
1	1	Andrews. Miss. Kornelia Theodosia	female	1	0	13502	77.9583	S	4

Configured transformers

Auto Data Preparation Automatic Transformer

Close Previous Next

6. Select **Label Column** to **survived**. This will automatically set **Suggested technique** to **Binary Classification**.

Projects > WSL_Demos_BB > Titanic

Select Data

Select a technique

Column value to predict (Label Col)

survived

Suggested technique.

Binary Classification

Classify new data into defined categories based on existing data. Choose if your label column contains two distinct categories.

Multiclass Classification

Classify new data into defined categories based on existing data. Choose if your label column contains a discrete number of categories.

Regression

Predict values from a continuous set of values. Choose if your label column contains a large number of values.

Validation Split

Train: 60 Test: 20 Holdout: 20

Configured estimators

Add Estimators

Close Previous Next

7. Select **Add Estimators**. Select all estimators and select **Add**.

Select estimator(s)

What type of estimator are you looking for?

Logistic Regression
 Analyzes a data set in which there are one or more independent variables that determine one of two outcomes. Only binary l...

Decision Tree Classifier
 Maps observations about an item (represented in the branches) to conclusions about the item's target value (represented in...

Random Forest Classifier
 Constructs multiple decision trees to produce the label that is a mode of each decision tree. It supports both binary and ...

Gradient Boosted Tree Classifier
 Produces a classification prediction model in the form of an ensemble of decision trees. It only supports binary labels, a...

Cancel
 Add

8. Select Next.

Projects > WSL_Demos_BB > Titanic

Select Data
 Prepare
 Train
 Evaluate

Select a technique

You cannot change label column or model type after adding an estimator.
You must first delete all estimators in order to make changes to these attributes.

Column value to predict (Label Col)
survived

Suggested technique.

Binary Classification
 Classify new data into defined categories based on existing data. Choose if your label column contains two distinct categories.

Multiclass Classification
 Classify new data into defined categories based on existing data. Choose if your label column contains a discrete number of categories.

Regression
 Predict values from a continuous set of values. Choose if your label column contains a large number of values.

Validation Split

Train: 60 Test: 20 Holdout: 20

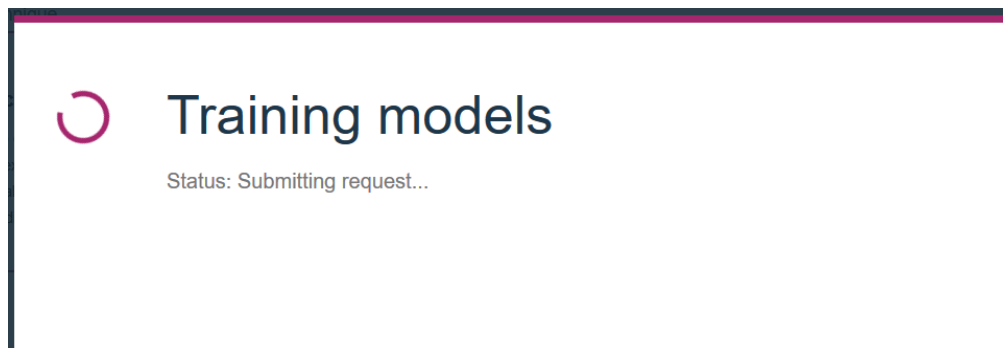
Add Estimators

Configured estimators

- Logistic Regression Not yet trained
- Decision Tree Classifier Not yet trained
- Random Forest Classifier Not yet trained
- Gradient Boosted Tree Classifier Not yet trained

Close
 Previous
 Next

9. Wait for all models to be trained



10. Review model performance. Models are ranked from best to worst performing.

Projects > WSL_Demos_BB > Titanic

Select Data

Prepare

Train

Evaluate

Select model

	ESTIMATOR TYPE	PERFORMANCE	AREA UNDER ROC CURVE	AREA UNDER PR CURVE	LAST VALIDATION	ACTIONS
<input checked="" type="radio"/>	Random Forest Classifier	Fair	0.7636	0.76519	9 Dec 2018, 5:40 PM	...
<input type="radio"/>	Logistic Regression	Fair	0.74916	0.75278	9 Dec 2018, 5:39 PM	...
<input type="radio"/>	Decision Tree Classifier	Fair	0.74597	0.74747	9 Dec 2018, 5:40 PM	...
<input type="radio"/>	Gradient Boosted Tree Classifier	Fair	0.72702	0.73131	9 Dec 2018, 5:41 PM	...

ClosePreviousSave

Step 3: Saving and Testing a Model

We can deploy the model to enable applications to invoke it via an API call. This is a Web Service deployment or Online deployment.

1. Select the **Save** button for the model you wish to deploy

Projects > WSL_Demos_BB > Titanic

Select Data

Prepare

Train

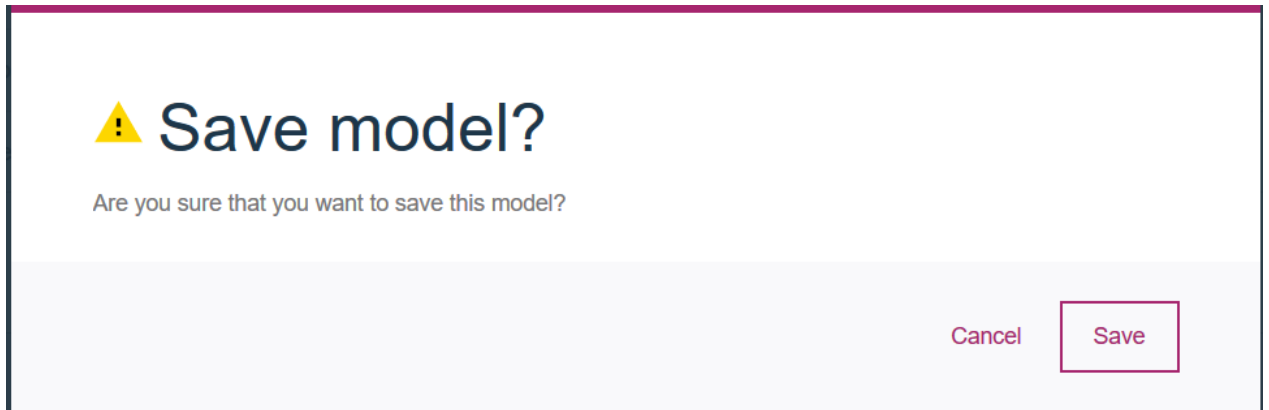
Evaluate

Select model

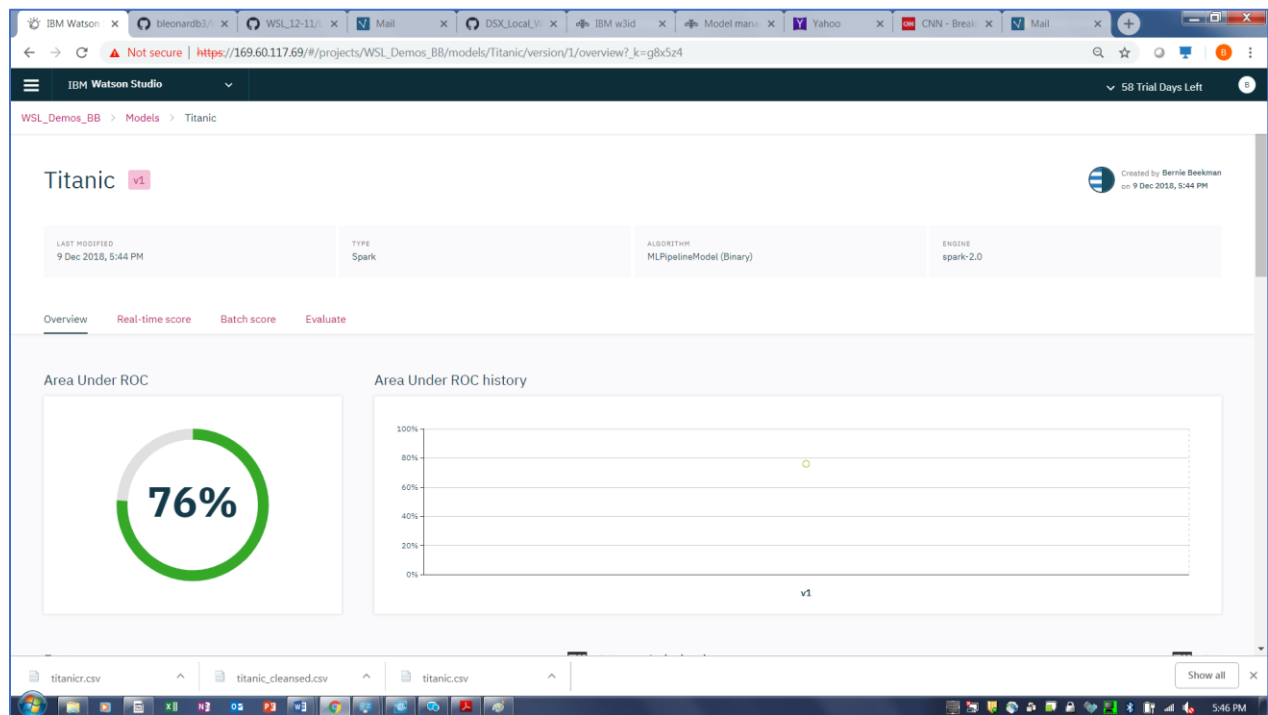
	ESTIMATOR TYPE	PERFORMANCE	AREA UNDER ROC CURVE	AREA UNDER PR CURVE	LAST VALIDATION	ACTIONS
<input checked="" type="radio"/>	Random Forest Classifier	Fair	0.7636	0.76519	9 Dec 2018, 5:40 PM	...
<input type="radio"/>	Logistic Regression	Fair	0.74916	0.75278	9 Dec 2018, 5:39 PM	...
<input type="radio"/>	Decision Tree Classifier	Fair	0.74597	0.74747	9 Dec 2018, 5:40 PM	...
<input type="radio"/>	Gradient Boosted Tree Classifier	Fair	0.72702	0.73131	9 Dec 2018, 5:41 PM	...

ClosePreviousSave

2. Confirm the save.



3. The model is saved, and a view of the model is displayed.



4. The Titanic model now appears in the list of models.

WSL_Demos_BB

Created by **Bernie Beekman**
on 8 Dec 2018, 4:26 PM

Assets 35 Data Sources 0 Jobs 3 Environments 10 Collaborators 1

- Recent
- Data sets 14
- Notebooks 6
- RStudio 8
- Scripts 3
- Models 3
 - SciKit Learn 1
 - Spark 2
- Model groups 0
- Cognos dashboards 0
- Data Refinery flows 0
- Model flows 0

🔍 Search by model name

Models 3

➕ Add Model

Name	Type	Status	Last Modified	
Classic - v1	Spark	Trained	9 Dec 2018, 5:44 PM	⋮
Telco_Churn_ML_model v1	Spark	Trained	8 Dec 2018, 4:47 PM	⋮
CreditCard_Default_Model v2	SciKit Learn	Trained	22 Aug 2018, 2:48 AM	⋮