

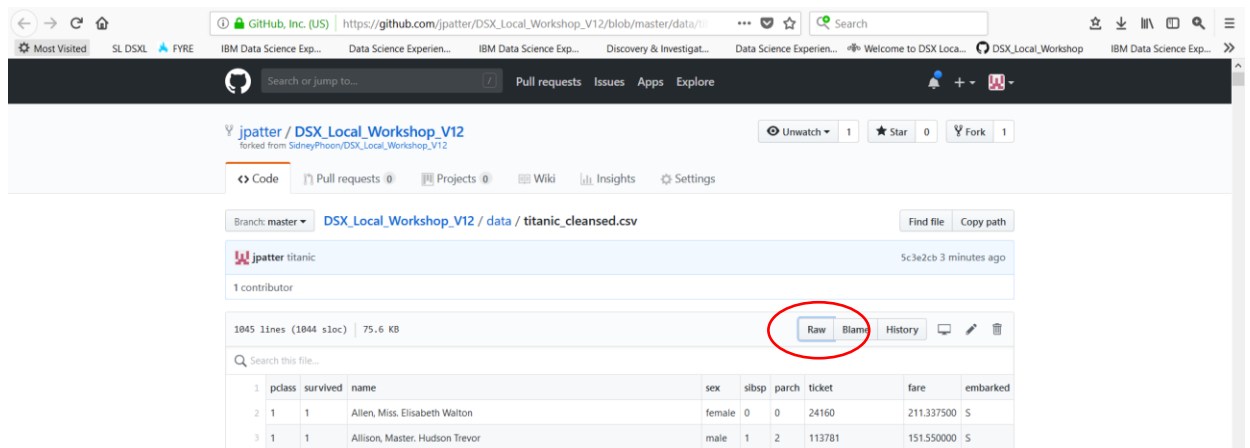
Watson Machine Learning Overview

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

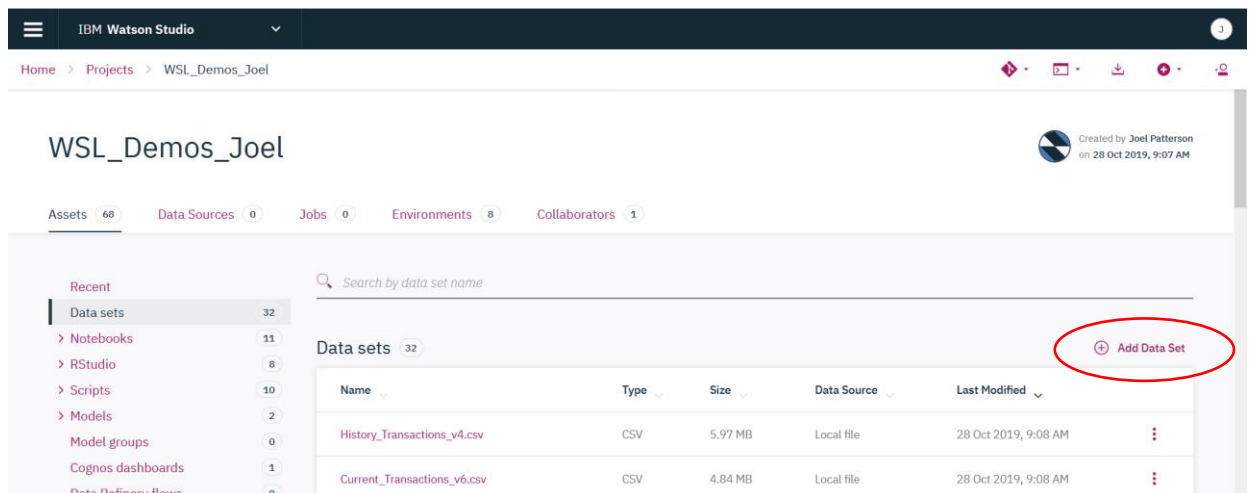
1. Adding a data asset to the WSL project
2. Creating a Model to predict whether a person would survive
3. Testing the Model

Step 1: Adding a Data Asset to the project

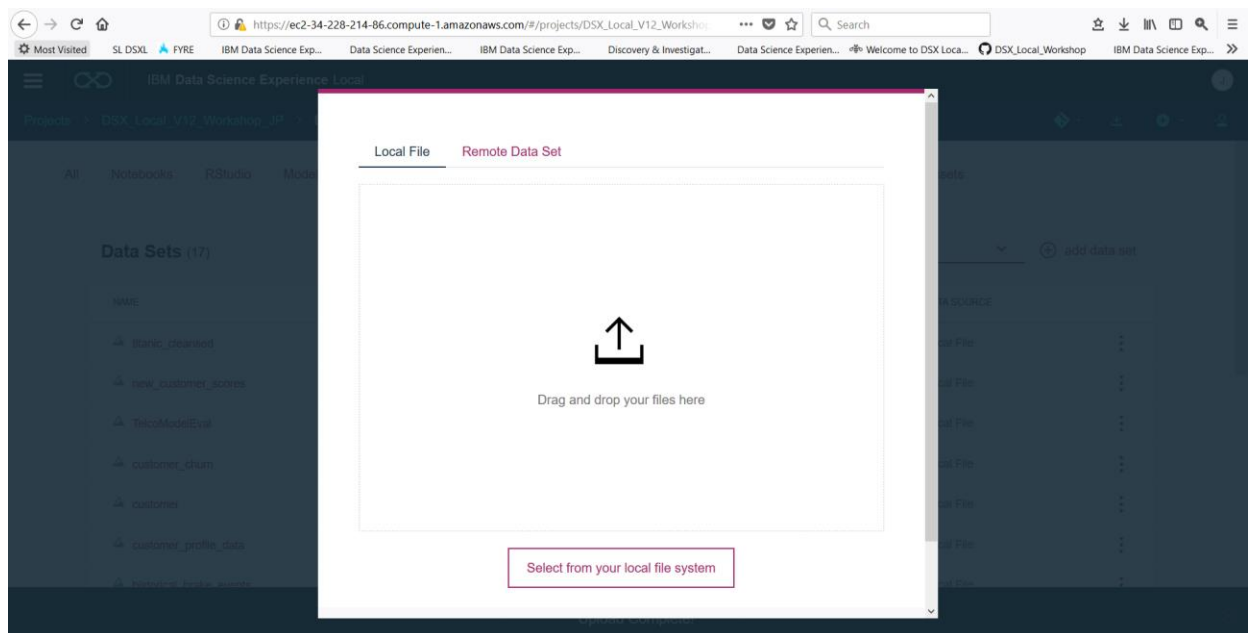
1. Download the Titanic data file from the following location by clicking on the link [Cleansed Titanic Data Set](#) and following the instructions below.
2. Right-click on **Raw** and select **Save link as...**



3. Save the file in your local filesystem
4. In your WSL 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 **Models**

IBM Watson Studio

Home > Projects > WSL_Demos_Joel

Created by Joel Patterson on 28 Oct 2019, 9:07 AM

Assets 69 Data Sources 0 Jobs 0 Environments 8 Collaborators 1

Recent

- Data sets 33
- > Notebooks 11
- > RStudio 8
- > Scripts 10
- Models 2**
- Spark 1
- Stats 1
- Model groups 0

Search by model name

Models 2 [Add Model](#)

Name	Type	Status	Last Modified
BrakeEventClassifier v1	Stats	Trained	27 Oct 2019, 4:30 PM
Telco_Churn_ML_model_35 v1	Spark	Trained	27 Oct 2019, 4:07 PM

2. Select **Add Model**

IBM Watson Studio

Home > Projects > WSL_Demos_Joel

Created by Joel Patterson on 28 Oct 2019, 9:07 AM

Assets 69 Data Sources 0 Jobs 0 Environments 8 Collaborators 1

Recent

- Data sets 33
- > Notebooks 11
- > RStudio 8
- > Scripts 10
- Models 2**
- Spark 1
- Stats 1
- Model groups 0

Search by model name

Models 2 [Add Model](#)

Name	Type	Status	Last Modified
BrakeEventClassifier v1	Stats	Trained	27 Oct 2019, 4:30 PM
Telco_Churn_ML_model_35 v1	Spark	Trained	27 Oct 2019, 4:07 PM

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

IBM Watson Studio

Projects > WSL_Demos_Joel > Add Model

Name *
TitanicModel

Description
Model description

Model type *
☒ Machine Learning ^① ☐ Decision Optimization ^①

Method *
Automatic
Prepare my data and create a model automatically.

Manual
Let me prepare my data and select which models to train.

Cancel Create

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

IBM Watson Studio

Projects > WSL_Demos_Joel > TitanicModel

Select Data

Prepare

Train

Evaluate

Select data asset

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

NAME
<input checked="" type="radio"/> titanic_cleansed.csv
<input type="radio"/> History_Transactions_v4.csv

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

Projects > WSL_Demos_Joel > TitanicModel

Select Data

Prepare

Train

Evaluate

Prepare data set

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	female	1	0	13502	77.9583	S	4

Close Previous **Next**

Configured transformers

+ Add a Transformer

Auto Data Preparation Automatic Transformer

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

Projects > WSL_Demos_Joel > TitanicModel

Select Data

Prepare

Train

Evaluate

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.

Configured estimators

+ Add Estimators

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

Projects > WSL_Demos_Joel > TitanicModel

Select Data

Prepare

Train

Evaluate

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.

Configured estimators

+ Add Estimators

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 I...

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...

CancelAdd

8. Select Next.

IBM Watson Studio

Projects > WSL_Demos_Joel > TitanicModel

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

Close

Previous

Next

Logistic Regression
Not yet trained

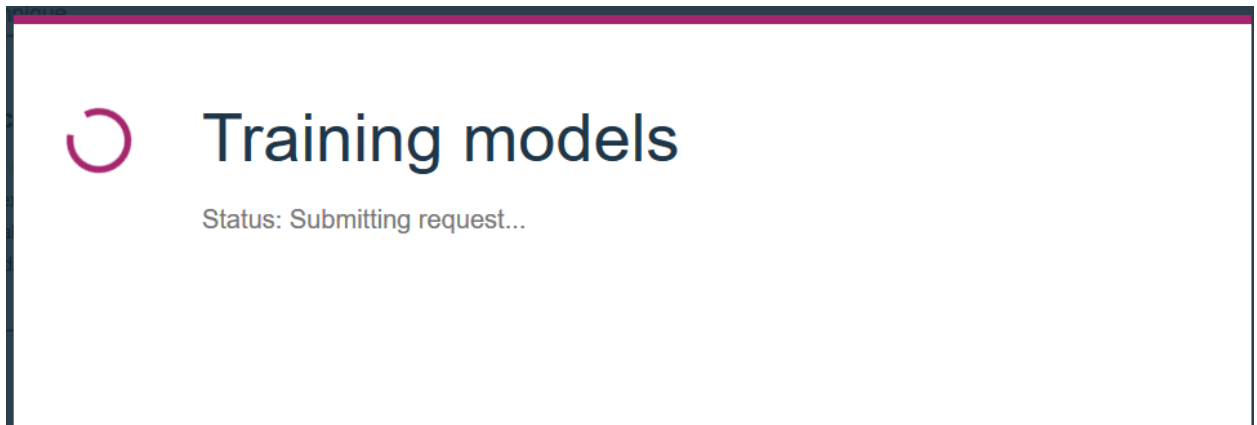
Decision Tree Classifier
Not yet trained

Random Forest Classifier
Not yet trained

Gradient Boosted Tree Classifier
Not yet trained

+ Add Estimators

9. Wait for all models to be trained



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

Select model						
	ESTIMATOR TYPE	PERFORMANCE	AREA UNDER ROC CURVE	AREA UNDER PR CURVE	LAST VALIDATION	ACTIONS
<input checked="" type="radio"/>	Logistic Regression	Fair	0.76348	0.792	28 Oct 2019, 11:37 AM	...
<input type="radio"/>	Gradient Boosted Tree Classifier	Fair	0.74848	0.74584	28 Oct 2019, 11:38 AM	...
<input type="radio"/>	Random Forest Classifier	Fair	0.74636	0.81421	28 Oct 2019, 11:37 AM	...
<input type="radio"/>	Decision Tree Classifier	Fair	0.74273	0.78815	28 Oct 2019, 11:37 AM	...

Close Previous Save

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

IBM Watson Studio

Projects > WSL_Demos_Joel > TitanicModel


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"/>	Logistic Regression	Fair	0.76348	0.792	28 Oct 2019, 11:37 AM	...
<input type="radio"/>	Gradient Boosted Tree Classifier	Fair	0.74848	0.74584	28 Oct 2019, 11:38 AM	...
<input type="radio"/>	Random Forest Classifier	Fair	0.74636	0.81421	28 Oct 2019, 11:37 AM	...
<input type="radio"/>	Decision Tree Classifier	Fair	0.74273	0.78815	28 Oct 2019, 11:37 AM	...

Close Previous Save

2. Confirm the save.

 **Save model?**

Are you sure that you want to save this model?

Cancel Save

3. The model now exists inside the **Models** tab of the project

IBM Watson Studio

WSL_Demos_Joel > Models > TitanicModel

TitanicModel v1

Created by Joel Patterson on 28 Oct 2019, 11:41 AM

LAST MODIFIED 28 Oct 2019, 11:41 AM	TYPE Spark	ALGORITHM MLPipelineModel (Binary)	ENGINE spark-2.0
--	---------------	---------------------------------------	---------------------

WSL_Demos_Joel

 Created by Joel Patterson
on 28 Oct 2019, 9:07 AM

Assets 70


Data Sources 0

Jobs 0

Environments 8







Collaborators 1

- Recent
- Data sets 33
- > Notebooks 11
- > RStudio 8
- > Scripts 10
- > Models 3
- Spark 2
- Stats 1
- Model groups 0
- Cognos dashboards 1
- Data Refinery flows 0

 Search by model name

Models 3

 Add Model

Name 	Type 	Status 	Last Modified 	
TitanicModel v1	Spark	Trained	28 Oct 2019, 11:41 AM	
BrakeEventClassifier v1	Stats	Trained	27 Oct 2019, 4:30 PM	
Telco_Churn_ML_model_35 v1	Spark	Trained	27 Oct 2019, 4:07 PM	