

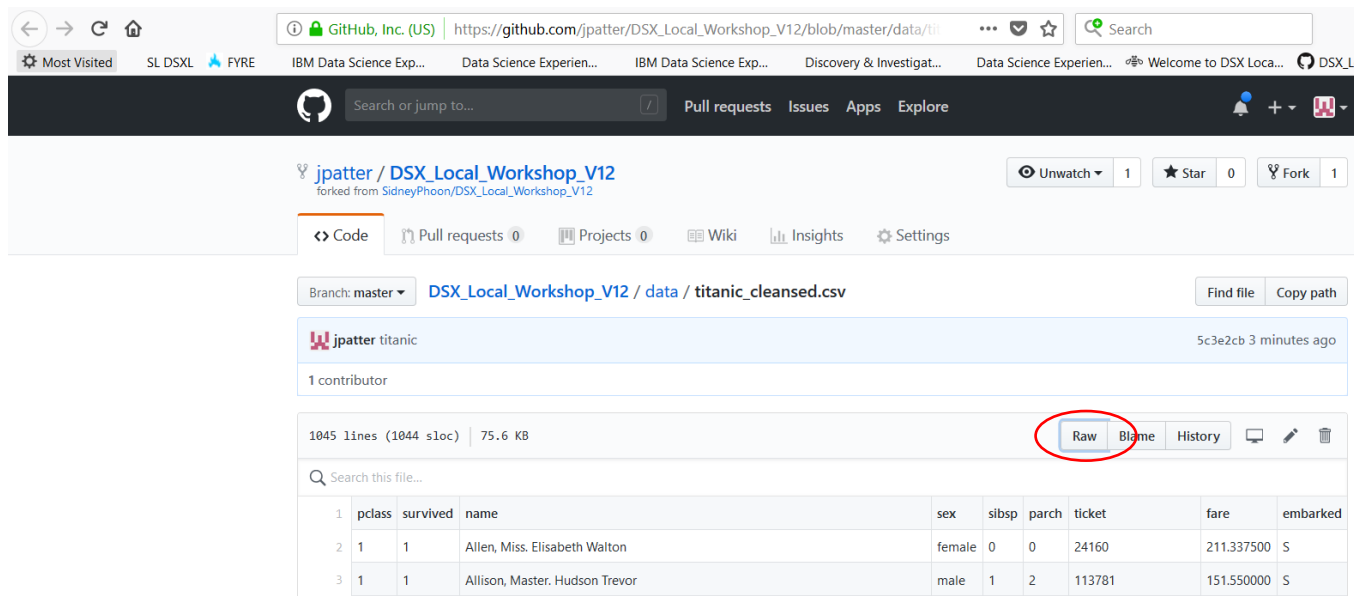
# Watson Machine Learning Overview

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...**



The screenshot shows a GitHub repository page for 'jpatter / DSX\_Local\_Workshop\_V12'. The file 'titanic\_cleansed.csv' is selected, showing 1045 lines and 75.6 KB. The 'Raw' button is circled in red. Below the file information, a table displays the first three rows of the dataset.

	pclass	survived	name	sex	sibsp	parch	ticket	fare	embarked
1	1	1	Allen, Miss. Elisabeth Walton	female	0	0	24160	211.337500	S
3	1	1	Allison, Master. Hudson Trevor	male	1	2	113781	151.550000	S

3. Save the file in your local filesystem
4. In your WSL project go to **Data Sets** and select **add data set**

IBM Watson Studio

Home > Projects > WSL\_Demos\_Joel

WSL\_Demos\_Joel

Assets 68Data Sources 0Jobs 0Environments 8Collaborators 1

Recent

Data sets32

> Notebooks11

> RStudio8

> Scripts10

> Models2

Model groups0

Cognos dashboards1

Data Refinerv flows0

Search by data set name

Data sets32

Name	Type	Size	Data Source	Last Modified
History_Transactions_v4.csv	CSV	5.97 MB	Local file	28 Oct 2019
Current_Transactions_v6.csv	CSV	4.84 MB	Local file	28 Oct 2019

5. Browse or drag the **titanic\_cleansed.csv** file

https://ec2-34-228-214-86.compute-1.amazonaws.com/#/projects/DSX\_Local\_V12\_Workshop

Most VisitedSL DSX FYRE IBM Data Science Exp... Data Science Experien... IBM Data Science Exp... Discovery & Investigat... Data Science Experien... Welcome to DSX Loca... DSX L

IBM Data Science Experience Local

Projects > DSX\_Local\_V12\_Workshop\_IP

All Notebooks RStudio Models

Data Sets (17)

NAME
titanic_cleansed
new_customer_scores
TelcoModelEval
customer_churn
customer
customer_profile_data
historical_braze_events

Local FileRemote Data Set

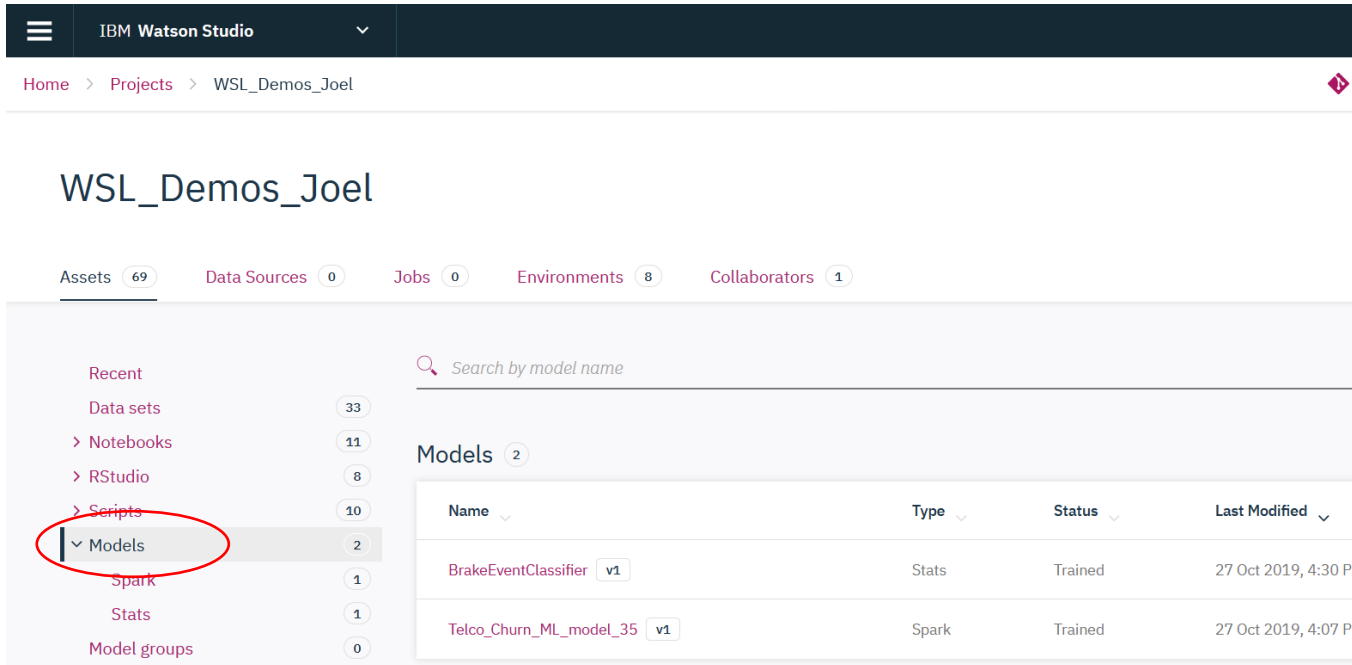
↑

Drag and drop your files here

Select from your local file system

## Step 2: Create a Model to predict survival

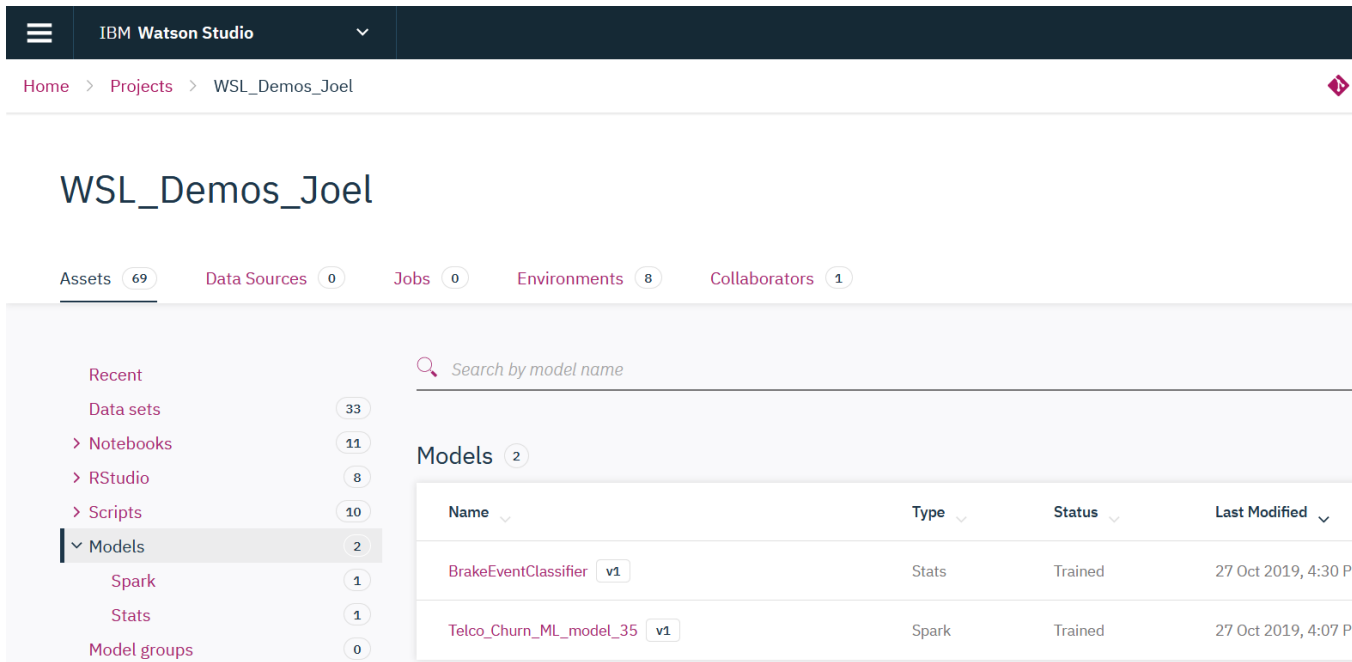
### 1. Select **Models**



The screenshot shows the IBM Watson Studio interface for a project named 'WSL\_Demos\_Joel'. The left sidebar contains a list of assets: Recent, Data sets (33), Notebooks (11), RStudio (8), Scripts (10), Models (2), Spark (1), Stats (1), and Model groups (0). The 'Models' item is highlighted with a red circle. The main content area displays a search bar 'Search by model name' and a table of models.

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

### 2. Select **Add Model**



This screenshot is identical to the one above, showing the IBM Watson Studio interface for the 'WSL\_Demos\_Joel' project. The 'Models' item in the left sidebar is highlighted with a red circle, and the main content area displays the same table of models.

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

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 ⓘ ☐ Decision Optimization ⓘ



**Method \***

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

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

300

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**.

 IBM Watson Studio 

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

Configured  
 Auto Data Preparation

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

IBM Watson Studio

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

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

IBM Watson Studio

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

## 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...

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

Configured

Logistic Not yet t

Decision Not yet t

Random Not yet t

Gradient Not yet t

Close

9. Wait for all models to be trained

Training models

Status: Submitting request...

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



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 VA
<input checked="" type="radio"/>	Logistic Regression	Fair	0.76348	0.792	28 Oct
<input type="radio"/>	Gradient Boosted Tree Classifier	Fair	0.74848	0.74584	28 Oct
<input type="radio"/>	Random Forest Classifier	Fair	0.74636	0.81421	28 Oct
<input type="radio"/>	Decision Tree Classifier	Fair	0.74273	0.78815	28 Oct

## 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 VA
<input checked="" type="radio"/>	Logistic Regression	Fair	0.76348	0.792	28 Oct
<input type="radio"/>	Gradient Boosted Tree Classifier	Fair	0.74848	0.74584	28 Oct
<input type="radio"/>	Random Forest Classifier	Fair	0.74636	0.81421	28 Oct
<input type="radio"/>	Decision Tree Classifier	Fair	0.74273	0.78815	28 Oct

- Confirm the save.



## Save model?

Are you sure that you want to save this model?

Cancel

Save

- The model now exists inside the **Models** tab of the project

IBM Watson Studio

WSL\_Demos\_Joel > Models > TitanicModel

## TitanicModel v1

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

# WSL\_Demos\_Joel

Assets70

Data Sources0


Jobs0

Environments8

Collaborators1

Recent

- Data sets33
- > Notebooks11
- > RStudio8
- > Scripts10
- ▼ Models3
- Spark2
- Stats1
- Model groups0
- Cognos dashboards1
- Data Refinery flows0

 Search by model name

## Models3

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