

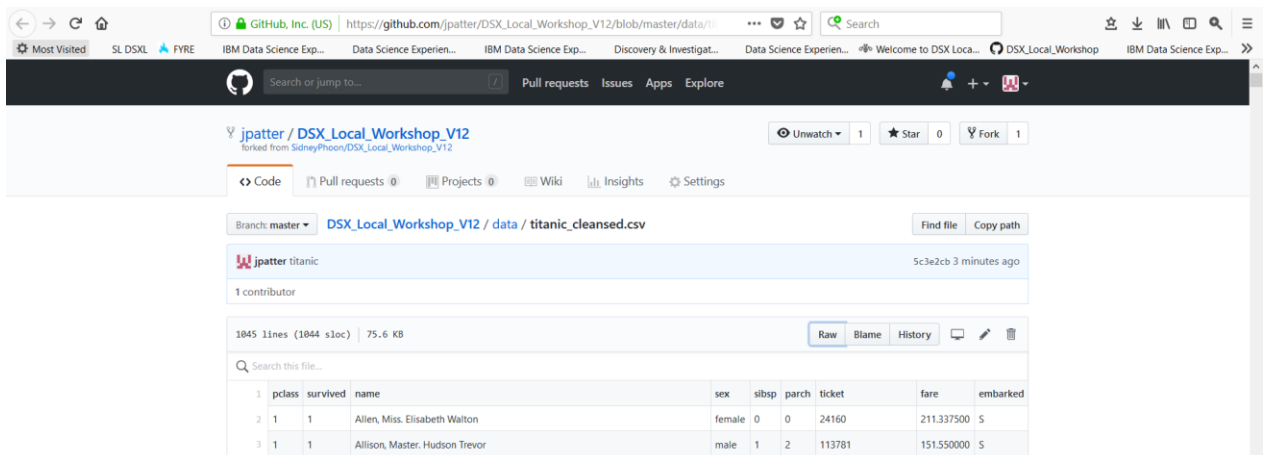
Watson Machine Learning Overview

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

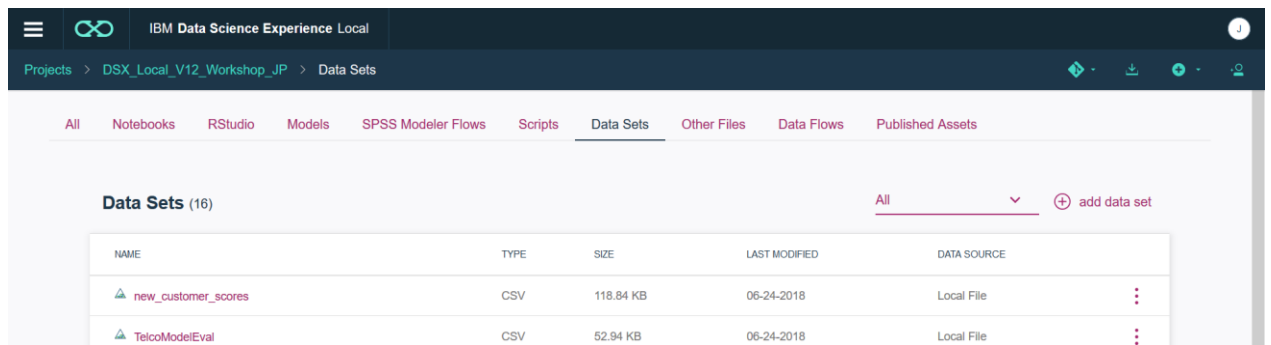
1. Download the Titanic data file from the following location by clicking on the link [Cleansed Titanic Data Set](#) and following the instructions below.
2. Right-click on **Raw** and select **Save link as...**



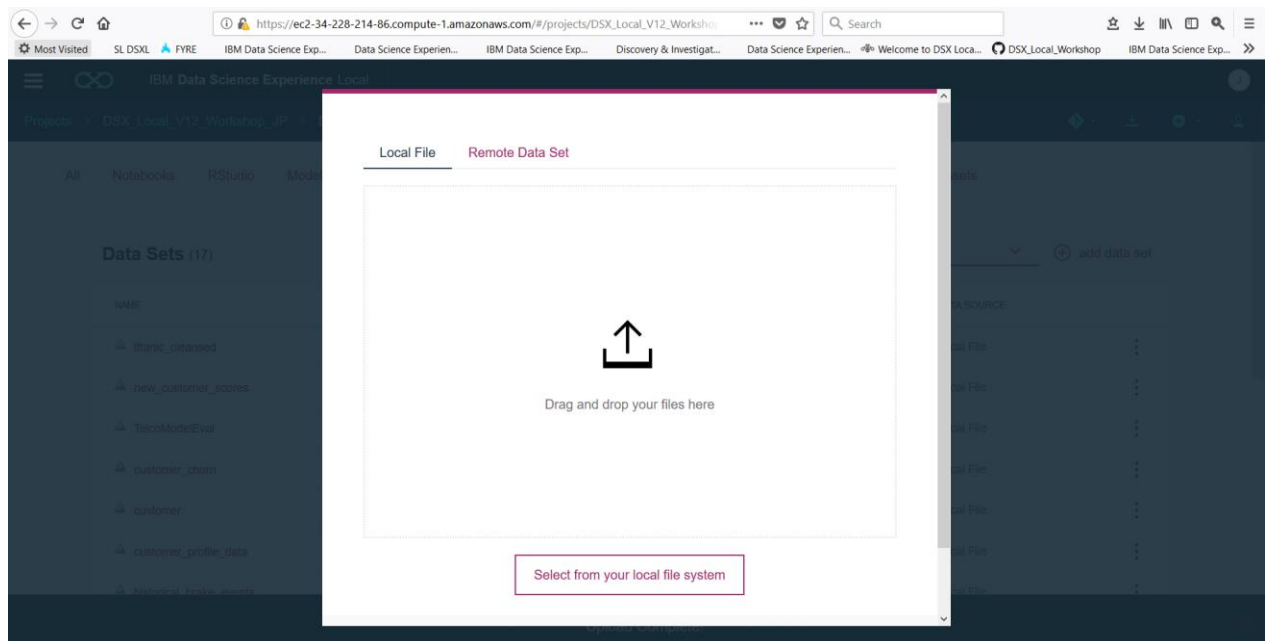
The screenshot shows the GitHub repository page for `jpattern / DSX_Local_Workshop_V12`. The file path `DSX_Local_Workshop_V12 / data / titanic_cleansed.csv` is highlighted. Below the file path, there is a table preview of the CSV data.

	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 DSXL project go to **Data Sets** and select **add data set**

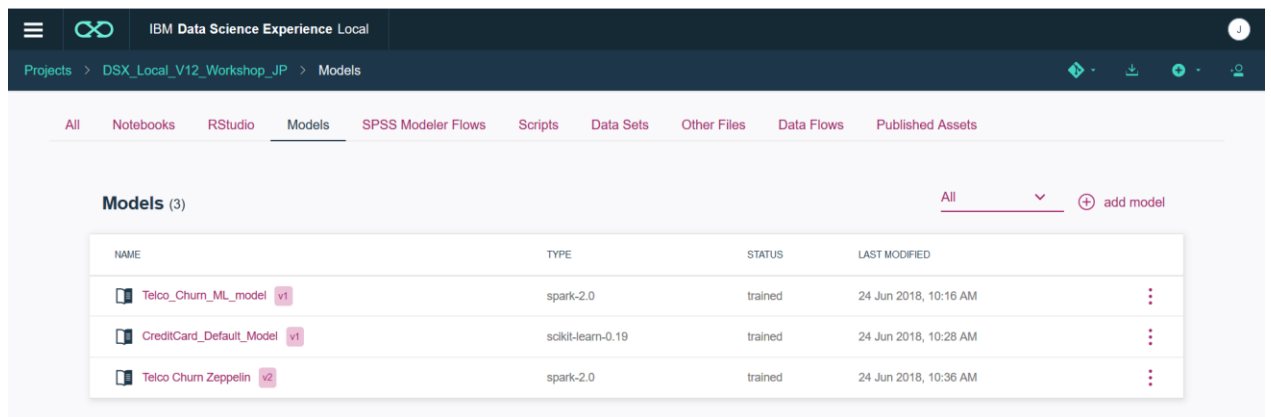


5. Browse or drag the **titanic_cleansed.csv** file



Step 2: Create a Model to predict survival

1. Select the **Models** tab



2. Select **add model**

The screenshot shows the 'Add Model' form in the IBM Data Science Experience Local interface. The form has two tabs: 'Blank' and 'From File'. The 'From File' tab is selected. The form includes the following fields and options:

- Name ***: A text input field with a character count of 100.
- Description**: A text area with a character count of 300.
- Select model type ***: Two radio buttons. 'Machine Learning' is selected, and 'Decision Optimization' is unselected.
- Method ***: Two options. 'Automatic' is selected and highlighted with a pink border. 'Manual' is unselected.

At the bottom right of the form, there are 'Cancel' and 'Create' buttons.

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 > DSX_Local_V12_Workshop_JP > Add Model

Blank From File

Name *
titanic-model 87

Description
Model description 300

Select 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.

Cancel Create

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

IBM Data Science Experience Local

Projects > DSX_Local_V12_Workshop_JP > titanic-model

Select Data

Prepare

Train

Evaluate

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"/>	customer_profile_data.csv	CSV	File System
<input type="radio"/>	historical_brake_events.csv	CSV	File System

Close Next

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

IBM Data Science Experience Local

Projects > DSX_Local_V12_Workshop_JP > titanic-model

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

Configure transformers

Auto Data Preparation Automatic Transformer

Close Previous Next

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

IBM Data Science Experience Local

Projects > DSX_Local_V12_Workshop_JP > titanic-model

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.

Validation Split

Train: 60 Test: 20 Holdout: 20

Configure 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 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...

Cancel Add

8. Select Next.

IBM Data Science Experience Local

Projects > DSX_Local_V12_Workshop_JP > titanic-model

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

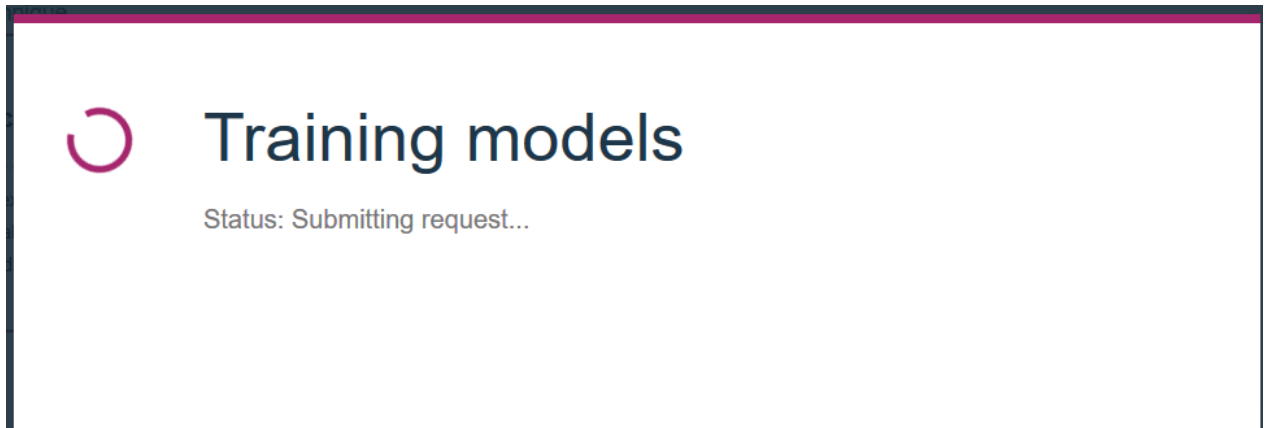
+ Add Estimators

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.

IBM Data Science Experience Local

Projects > DSX_Local_V12_Workshop_JP > titanic-model

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"/>	Decision Tree Classifier	Fair	0.77702	0.84566	24 Jun 2018, 4:23 PM	...
<input type="radio"/>	Logistic Regression	Fair	0.77659	0.81774	24 Jun 2018, 4:23 PM	...
<input type="radio"/>	Random Forest Classifier	Fair	0.77274	0.83924	24 Jun 2018, 4:23 PM	...
<input type="radio"/>	Gradient Boosted Tree Classifier	Fair	0.76758	0.8182	24 Jun 2018, 4:24 PM	...

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 Data Science Experience Local

Projects > DSX_Local_V12_Workshop_JP > titanic-model

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"/>	Decision Tree Classifier	Fair	0.77702	0.84566	24 Jun 2018, 4:23 PM	...
<input type="radio"/>	Logistic Regression	Fair	0.77659	0.81774	24 Jun 2018, 4:23 PM	...
<input type="radio"/>	Random Forest Classifier	Fair	0.77274	0.83924	24 Jun 2018, 4:23 PM	...
<input type="radio"/>	Gradient Boosted Tree Classifier	Fair	0.76758	0.8182	24 Jun 2018, 4:24 PM	...

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

AllNotebooksRStudioModelsSPSS Modeler FlowsScriptsData SetsOther FilesData FlowsPublished Assets

Models (4)

All▼+ add model

NAME	TYPE	STATUS	LAST MODIFIED	
 CreditCard_Default_Model v1	scikit-learn-0.19	trained	24 Jun 2018, 10:28 AM	⋮
 Telco Churn Zeppelin v2	spark-2.0	trained	24 Jun 2018, 10:36 AM	⋮
 Telco_Churn_ML_model v1	spark-2.0	trained	24 Jun 2018, 10:16 AM	⋮
 titanic-model v1	wml-sparkml-model-1.1	trained	24 Jun 2018, 4:30 PM	⋮

- Generate Script
- Test
- Batch Score
- Evaluate
- Publish to IBM Cloud
- Export
- Delete