

Watson Machine Learning Overview

This lab will introduce the Watson Machine Learning capability within IBM Data Science Experience. The lab will consist of the following steps:

1. Setting up the environment
2. Adding a data asset to the DSX Labs project
3. Creating a Machine Learning model
4. Deploying and Testing the Model

Step 1: Setting up your environment

To use IBM Watson Machine Learning you must have the following service instances in your Bluemix dashboard:

- Watson Machine Learning
- Object Storage
- Apache Spark

The Object Storage and Apache Spark service instances should already exist having been created when your DSX account was provisioned. We now need to provision a Machine Learning Service.

Step 1.1: Creating a Machine Learning Instance

To create a Machine Learning service instance, you must perform the following steps:

1. Go to the **Settings** tab in your project

Overview Assets Environments Bookmarks Deployments Collaborators **Settings**

Project information

Project name
 Lockheed Martin PoT 2/28 - 3/1
 70

Description
 Lockheed Martin PoT 2/28 - 3/1
 2970

Cancel Save

2. Scroll down to **Associated Services** select **add associated service** and select **Machine Learning**

Associated Services + add associated service ^

NAME	SERVICE TYPE	PLAN
DSX-Spark	Spark	Personal

- Amazon EMR Spark
- IBM Analytics Engine
- Machine Learning
- Spark

3. Select **New**. Select the **Lite (Free)** instance and **Create**

IBM Watson Data Platform | Projects Tools Catalog BETA Data Services Community US South [Settings] [Notifications] [User]

Machine Learning

Existing **New**

Machine Learning

IBM Watson Machine Learning is a full-service Bluemix offering that makes it easy for developers and data scientists to work together to integrate predictive capabilities with their applications. The Machine Learning service is a set of REST APIs that you can call from any programming language to develop applications that make smarter decisions, solve tough problems, and improve user outcomes.

Features

SPSS analytics platform features
 SPSS streams management and deployment with realtime scoring and batch processing options.

Spark and Python Machine Learning features
 Take advantage of Spark MLlib and scikit-learn machine learning models management and deployment - online, batch and streaming.

Integration with Data Science Experience
 Visit <http://datascience.ibm.com>. Create and train predictive analytics models with the best tools and the latest expertise in a social environment built by data scientists.

Pricing Plan: Monthly Process shown above reflect the: **United States**


Plan	Features	Pricing
● Lite	Service instance (5 models per instance) 5,000 predictions 5 compute hours	Free

4. Give a name to your new Machine Learning instance and **Confirm**


Confirm Creation

Organization: jpatter@us.ibm.com


Plan

Lite 

Space

dev 

Service Name

MyMachineLearning 

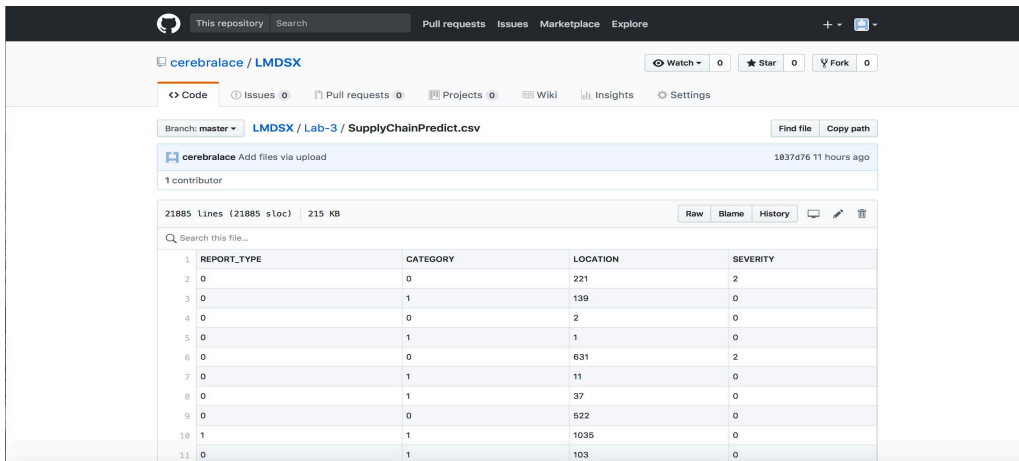
Cancel **Confirm**

Step 2: Adding a Data Asset to the project

1. Download the modified data file from [Modified Supply Chain Data Set](#)

The data in this file has already been prepared and it is ready to be input into the Modeling step. Select the link above.

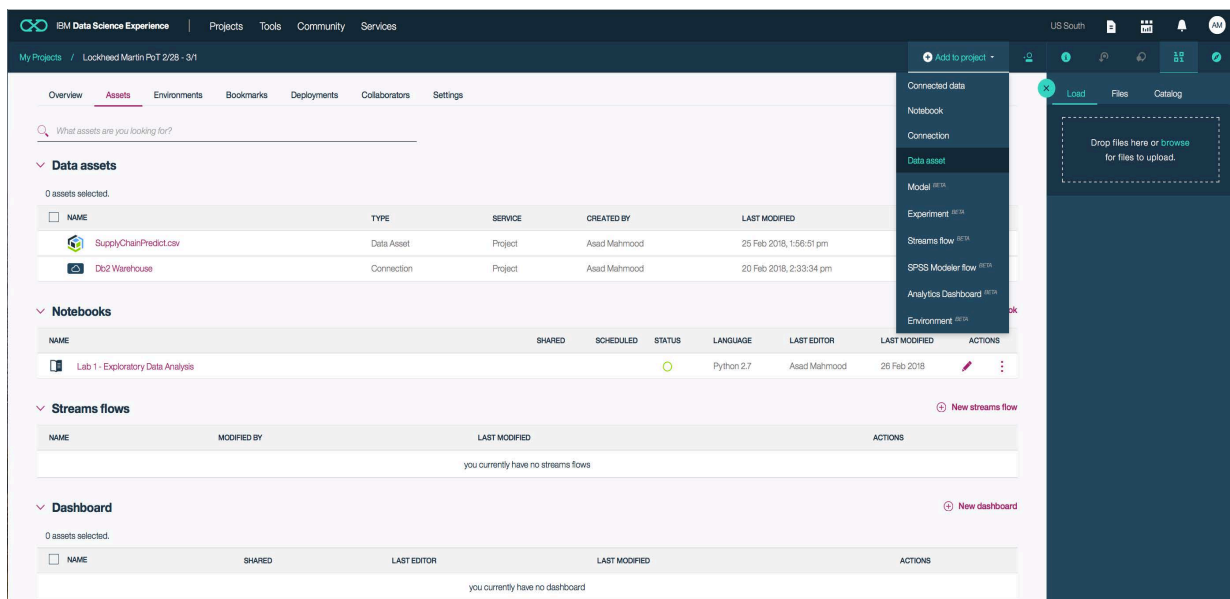
2. Right click on Raw, and click on Save link as



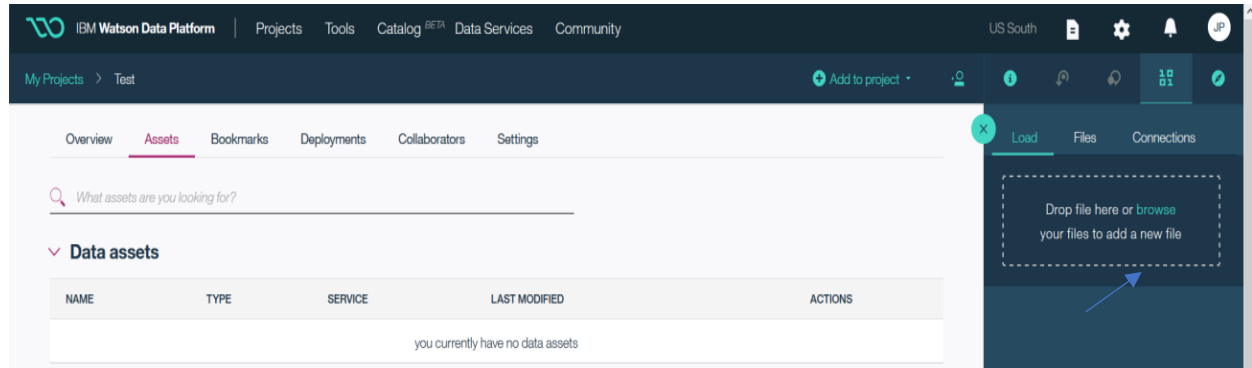
The screenshot shows the GitHub interface for the repository 'cerebrallace / LMDSX'. The file 'SupplyChainPredict.csv' is selected, showing its details: 21885 lines (21885 slots) and 215 KB. The 'Raw' button is highlighted. Below the file information, a table of data is visible.

	REPORT_TYPE	CATEGORY	LOCATION	SEVERITY
1				
2	0	0	221	2
3	0	1	139	0
4	0	0	2	0
5	0	1	1	0
6	0	0	631	2
7	0	1	11	0
8	0	1	37	0
9	0	0	522	0
10	1	1	1035	0
11	0	1	103	0

3. Go back to your DSX project. Click on **add data assets** or the  icon.

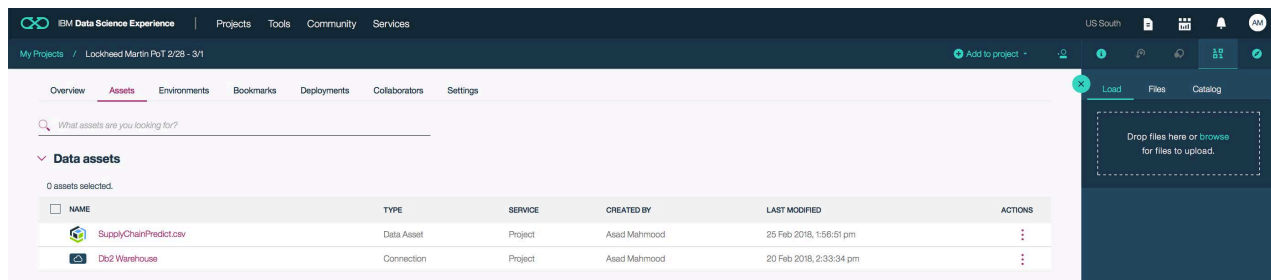


- Click on **Load** and then **browse** and then go to the folder where the SupplyChainPredict.csv is stored. Select SupplyChainPredict.csv and then click Open.

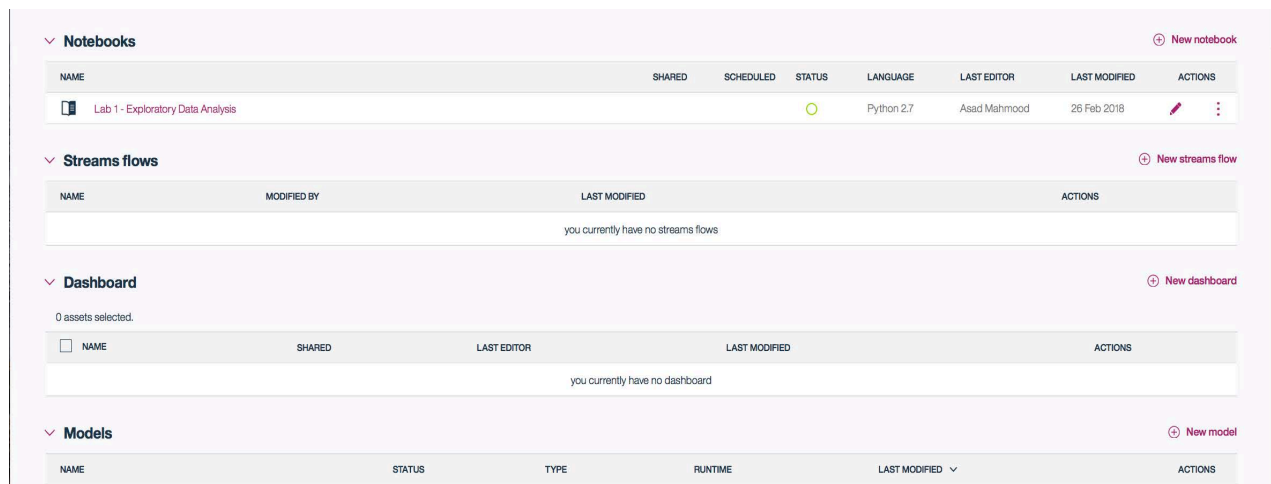


Step 3: Create a Model to predict survival

- Click on the **Assets** Tab



2. Click on **New model**.



3. Enter the Model **Name**, **Description**, Select **Manual**, and click on **Create**.

New model BETA

Define model details

Name
SupplyChainPredict

Description
Model description

Machine Learning Service
WML

Select model type

☒ Model builder ☐ From file ☐ From sample

Spark Service
Apache Spark-ID

Automatic
Prepare my data and create a model automatically

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

Need something more flexible? Create a [notebook](#) or design an [SPSS Modeler flow](#).

Cancel Create

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

Select data asset

The model builder currently supports CSV files and IBM Dc2 Warehouse on Cloud data assets.

What asset are you looking for?

NAME	TYPE	SERVICE
SupplyChainPredict.csv	Data Asset	Project

Close Next

5. For **Column value to predict (Label Col)** select **Severity**. For **Feature columns** select the following features (**Report_Type, Category, Location**) . Click on the **Multiclass Classification** Box (which is suggested by the service). Adjust the **Validation Split** as desired. Click on **Add Estimators** to add the specific models to use.

IBM Data Science Experience

Projects

Tools

Community

Services

US South

My Projects

/

Lockheed Martin PoT 2/28 - 3/1

/

SupplyChainModel

Select Data

Train

Evaluate

Select a technique

Column value to predict (Label Col)

SEVERITY (Integer)

Feature columns

REPORT_TYPE (Integer), CATEGORY (Integer), LOCATION (Integer)

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

Close

Previous

Next

6. Select **Random Forest Classifier**. You can select more if you wish to see the results of multiple models. Select **Add**.

Select estimator(s)

What type of estimator are you looking for?

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 ...

Naive Bayes

Classifies features based on Bayes' theorem, which assumes that the presence of a particular feature in a class is unrelat...

Cancel

Add

7. Select the **Next** button.

Select a technique

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

Column value to predict (Label Col)
SEVERITY (Integer)

Feature columns
REPORT_TYPE (Integer), CATEGORY (Integer), LOCATION (Integer)

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

Random Forest Classifier
Not Yet Trained

Close Previous Next

- The system trains and evaluates each model. If more than one model was selected, the models would be listed in descending order of quality with the best result at the top.
Note: if a model fails to run (rare, but happens), select **Previous**, delete that model and re-add it. Then run again. Click on **Random Forest Classifier** (if it is the best) and then click **Save**.

- The system displays the model training summary.

SupplyChainPredict

Overview Evaluation Deployments

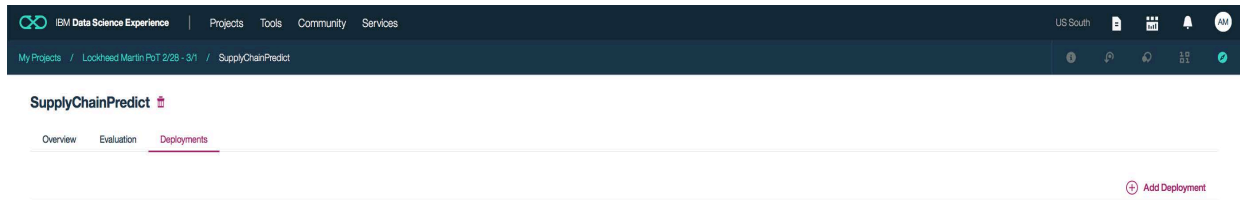
Summary

Machine learning service	WML
Model Type	wml-11
Runtime environment	spark-2.0
Training date	25 Feb 2018, 2:00 PM
Label column	SEVERITY
Latest version	6b46516c-d2b8-4ca3-963f-a2f30e33166d

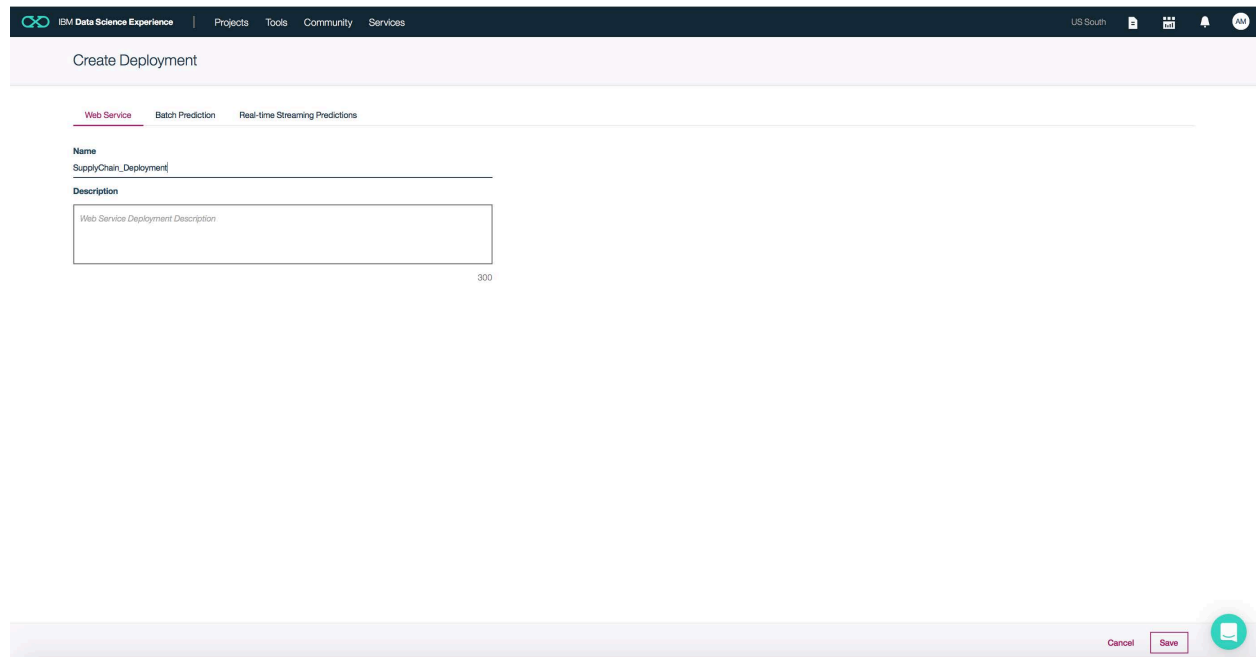
Step 4: Deploying a Model

We can deploy the model to enable applications to invoke it via an API call.

1. Select the **Deployments** Tab
2. Scroll down to the **Add Deployments** option. Click on **Add Deployments**



3. Select **Web Service** tab, enter SupplyChain_Deployment for **Name**, and click on **Save**.



4. The system responds with an acknowledgement that the model was successfully deployed. Click on **SupplyChain_Deployment** to test the deployed API.

The screenshot shows the 'Deployments' tab in the IBM Data Science Experience interface. The page has a dark blue header with navigation links: Overview, Assets, Environments, Bookmarks, Deployments (selected), Collaborators, and Settings. Below the header, there is a table with the following data:

NAME	TYPE	MODEL NAME	STATUS	ACTIONS
SupplyChain_Deployment	Web Service	SupplyChainPredict	DEPLOY_SUCCESS	

5. The system displays information about the deployed service including the endpoint to invoke by an application. Click on **Test** to test out the API.

The screenshot shows the 'SupplyChain_Deployment' page in the IBM Data Science Experience interface. The page has a dark blue header with navigation links: Overview, Implementation, and Test (selected). Below the header, there is a table with the following data:

Deployment	
Name	SupplyChain_Deployment
Type	Web Service
Deployment ID	e4cc1d3f-4225-4921-97b5-1886238aa4a2
Status	DEPLOY_SUCCESS
Machine learning service	WML
Created	25 Feb 2018 02:01pm
Last modified	26 Feb 2018 01:55pm

Model	
Name	SupplyChainPredict
Model ID	630ae616-7ef6-4245-b49c-c4b0d23d331b
Version ID	6b46616c-d2b8-4ca3-963f-a2930e33166d
URL	https://bm-watson-ml.mybluemix.net/v3/ml/assets/models/630ae616-7ef6-4245-b49c-c4b0d23d331b/versions/6b46616c-d2b8-4ca3-963f-a2930e33166d

6. Enter values for the input fields and then click on **Predict**.

IBM Watson Data Platform | Projects Tools Catalog ^{BETA} Data Services Community US South

My Projects > Test > Titanic > Titanic_Deployment

Titanic_Deployment

Overview Implementation **Test**

Input data

pclass	1
name	Allen, Miss. Elisabeth Walton
sex	female
sibsp	0

Predict

7. The predicted result is returned.

