

Watson Machine Learning Overview

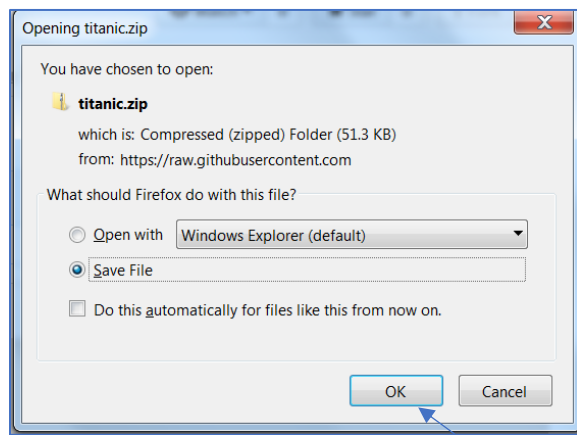
This lab will introduce the Watson Machine Learning capability using the Titanic dataset. The lab will consist of the following steps:


1. Adding a data asset to the Watson Studio Labs project
2. Creating a Model to predict whether a passenger would survive
3. Deploying and Testing the Model
4. Deploying a simple web front-end and connecting it to the Titanic deployed model.

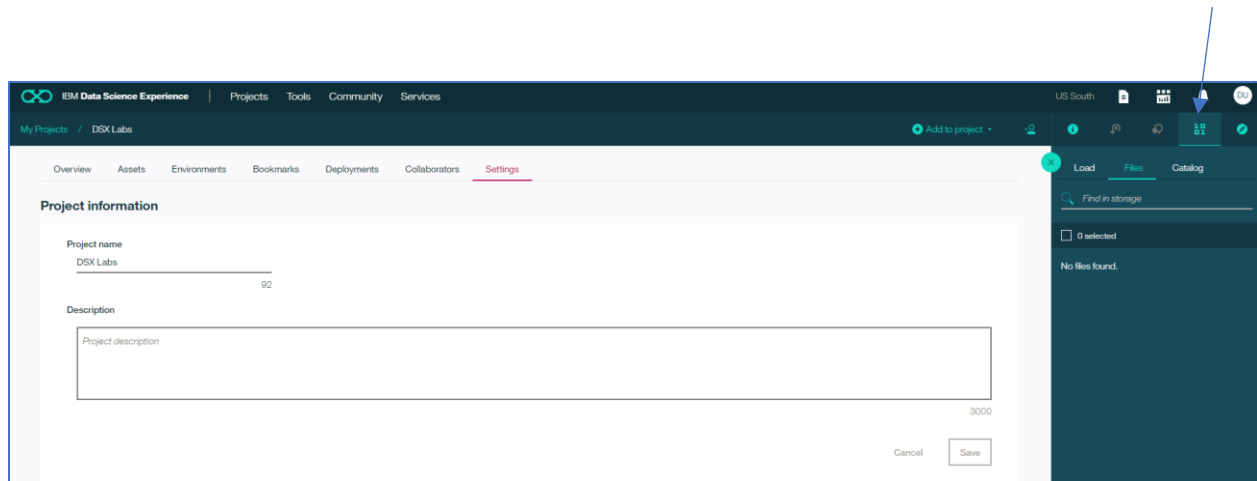
Step 1: Adding a Data Asset to the project

This step can be skipped if the [Titanic Data](#) was already downloaded in a previous lab.

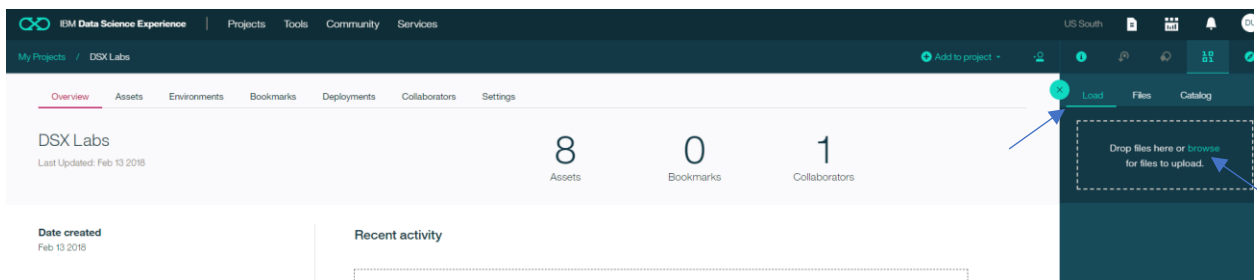
1. Download the Titanic data file from the following location by clicking on the link [Titanic Data](#).
2. Click on the **OK** button in the pop-up dialog.



3. Navigate to the directory where the file has been downloaded. Unzip the titanic.zip file. There should be two files (1) titanic_cleansed.csv and (2) titanic.csv. You will use the **titanic_cleansed.csv** for this lab.
4. Go back to your Watson Studio Labs project. Click on the  icon.

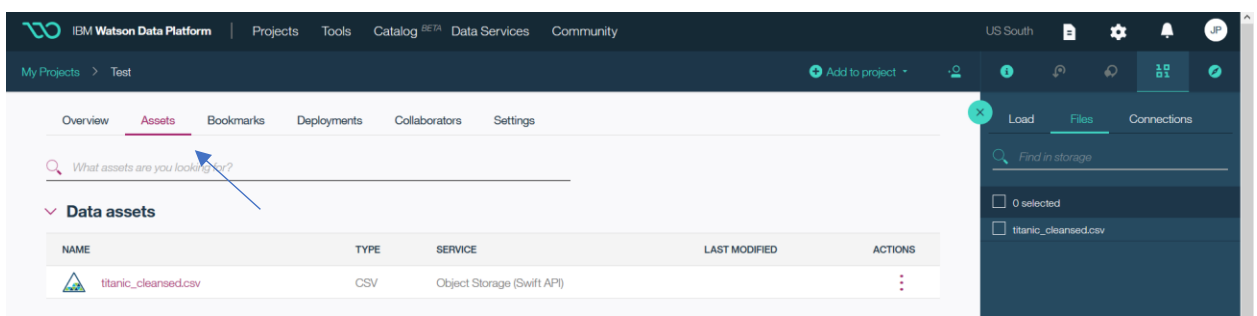


3. Click on **Load** and then **browse** and then go to the folder where the titanic_cleansed.csv is stored. Select titanic_cleansed.csv and then click Open.



Step 2: Create a Model to predict survival

1. Click on the **Assets** Tab



2. Click on **New Watson Machine Learning model**.

My Projects / Watson Studio Labs + Add to project

▼ **Models**

Natural Language Classifier models BETA + New Natural Language Classifier model

| NAME | MODEL ID | SERVICE INSTANCE | LAST MODIFIED | ACTIONS |
|--|----------|------------------|---------------|---------|
| You don't have any Natural Language Classifier models yet. | | | | |

Visual Recognition models + New Visual Recognition model

| NAME | MODEL ID | SERVICE INSTANCE | LAST MODIFIED | ACTIONS |
|---|----------|------------------|---------------|---------|
| You don't have any Visual Recognition models yet. | | | | |

Watson Machine Learning models + New Watson Machine Learning model

| NAME | STATUS | TYPE | RUNTIME | LAST MODIFIED | ACTIONS |
|--|--------|------|---------|---------------|---------|
| You don't have any Watson Machine Learning models yet. | | | | | |

- Enter a **Model Name** (eg Titanic), optionally a **Description**, leave the default for the **Machine Learning Service** (should be the one created in the pre-reqs), leave **Model Builder** selected, select the **Spark Scala Service**, select **Manual**, and click on **Create**.

Define model details

Name
Titanic-WML 89

Description

Model description

300

Machine Learning Service
Machine Learning

Select model type

☒ Model builder ☐ From file ☐ From sample

Select runtime

Only Spark environments supporting Scala kernels can be used for model builder creation.

Default Spark Scala 2.11

The selected runtime uses one driver with 1 vCPU and 4 GB RAM, and 2 executors each with 1 vCPU and 4 GB RAM. This runtime consumes 1.5 capacity units per hour.

⚠ Your Spark runtime will be automatically stopped when you save your model, or after 3 hours of inactivity. To avoid consuming extra capacity unit hours delete your model builder instance or stop your runtime when you are finished with it.

Learn more about capacity unit hours and Watson Studio pricing plans.

Automatic

Prepare my data and create a model automatically

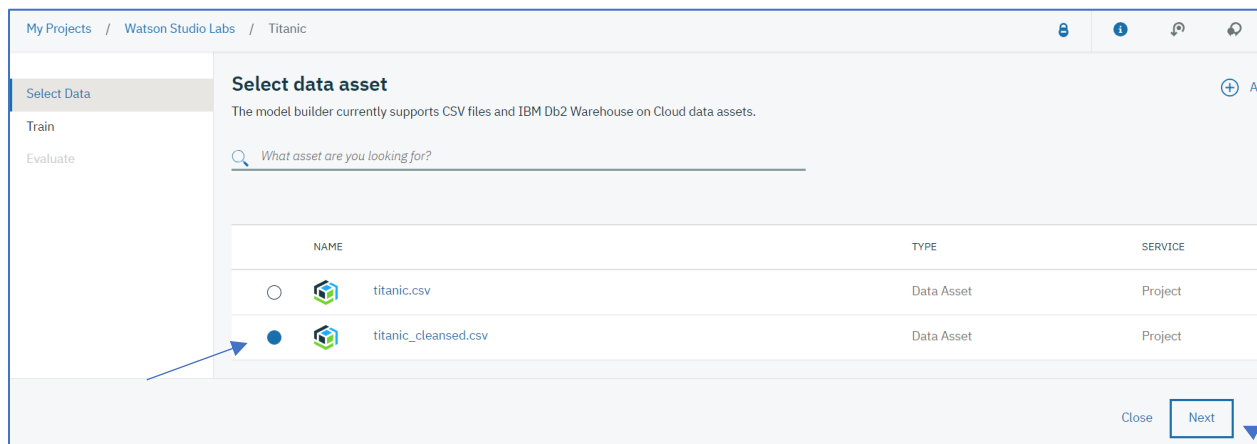
Manual

Let me prepare my data and select which models to train

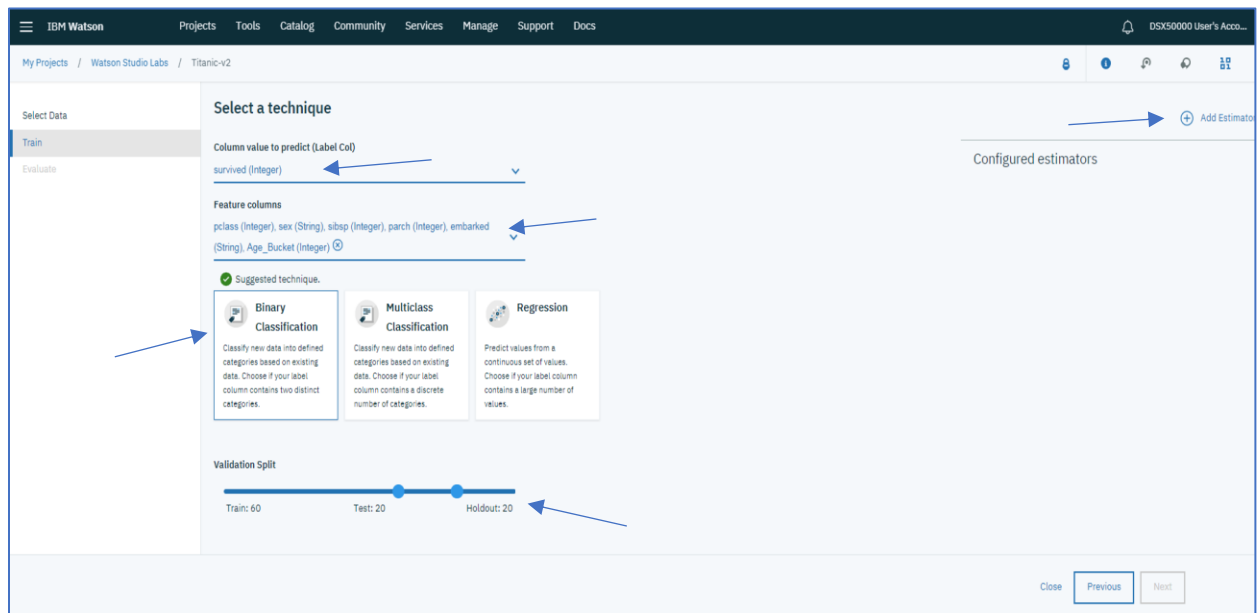
Need something more flexible? Create a notebook or design a Modeler flow

Cancel **Create**

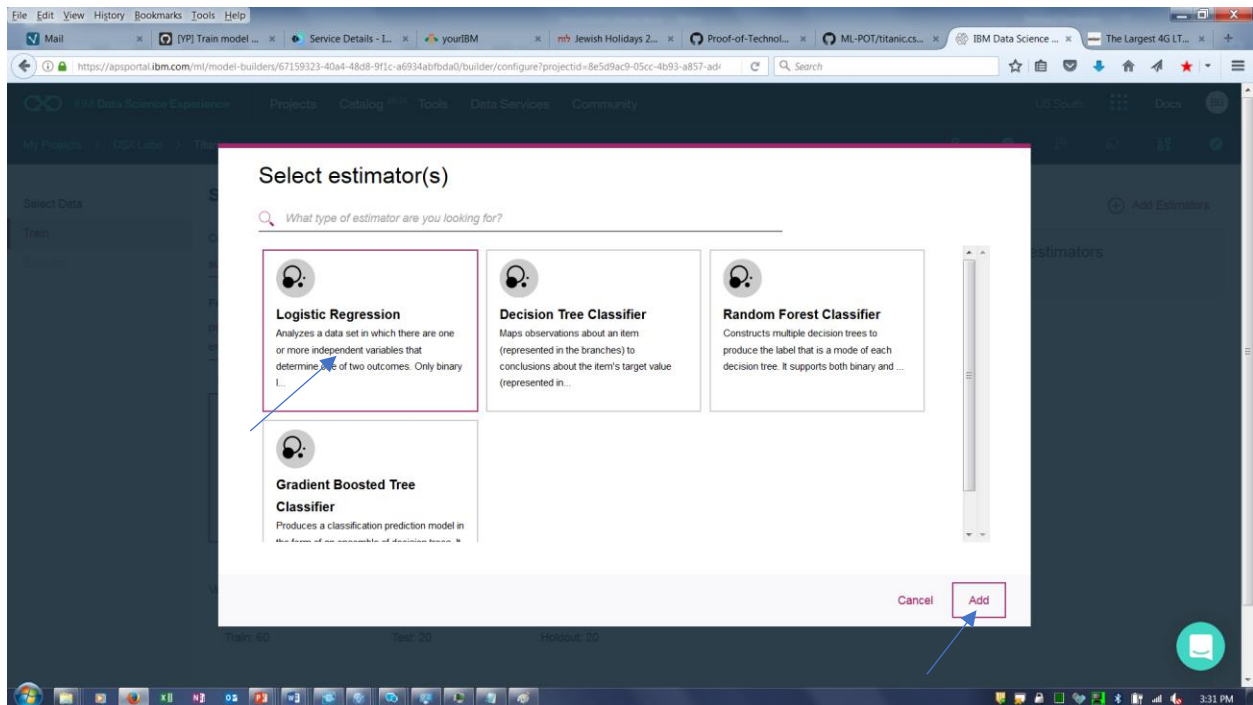
- Click on the `titanic_cleansed.csv` and click on **Next**



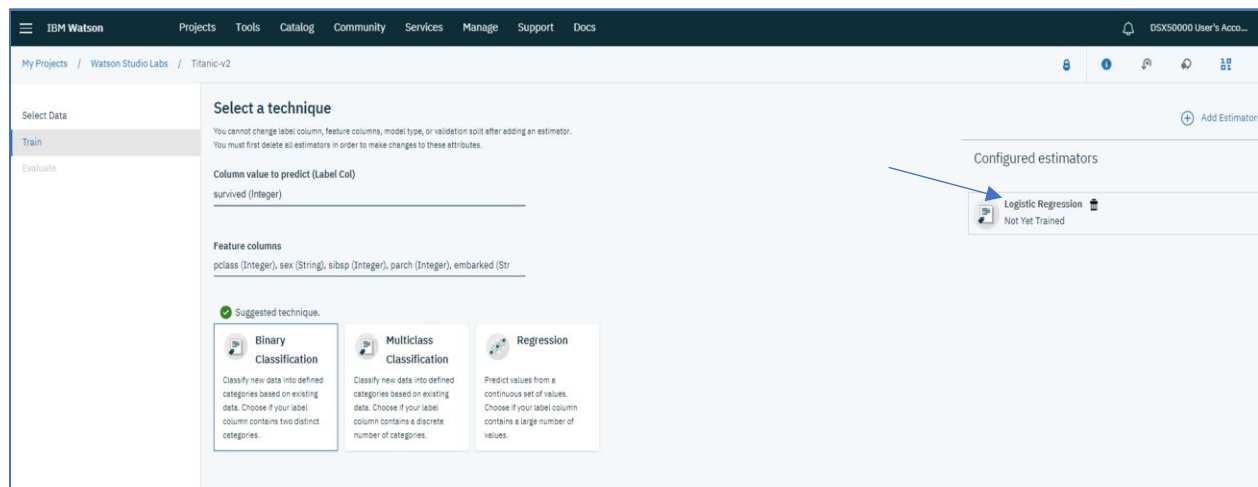
- For **Column value to predict (Label Col)** select **survived**. For **Feature columns** select the following features (**pclass, sex, sibsp, parch, fare, embarked, and Age_Bucket**) . Click on the **Binary Classification** Box (which is suggested by the service). Adjust the **Validation Split** as desired. Click on **Add Estimators** to add the specific models to use.



- Select **Logistic Regression**. You can select more if you wish to see the results of multiple models. Select **Add**.




7. Note you can adjust the algorithm's hyperparameters by clicking on Logistic Regression.



8. For now we will leave the hyperparameters unchanged, so click **Cancel**.

Configure Logistic Regression

Weight column
[Select](#) 

Elastic net parameter (double)
0


Fit an intercept
☐ `fitIntercept`


Maximum iterations for convergence (integer)
100

Regularization parameter (double)
0

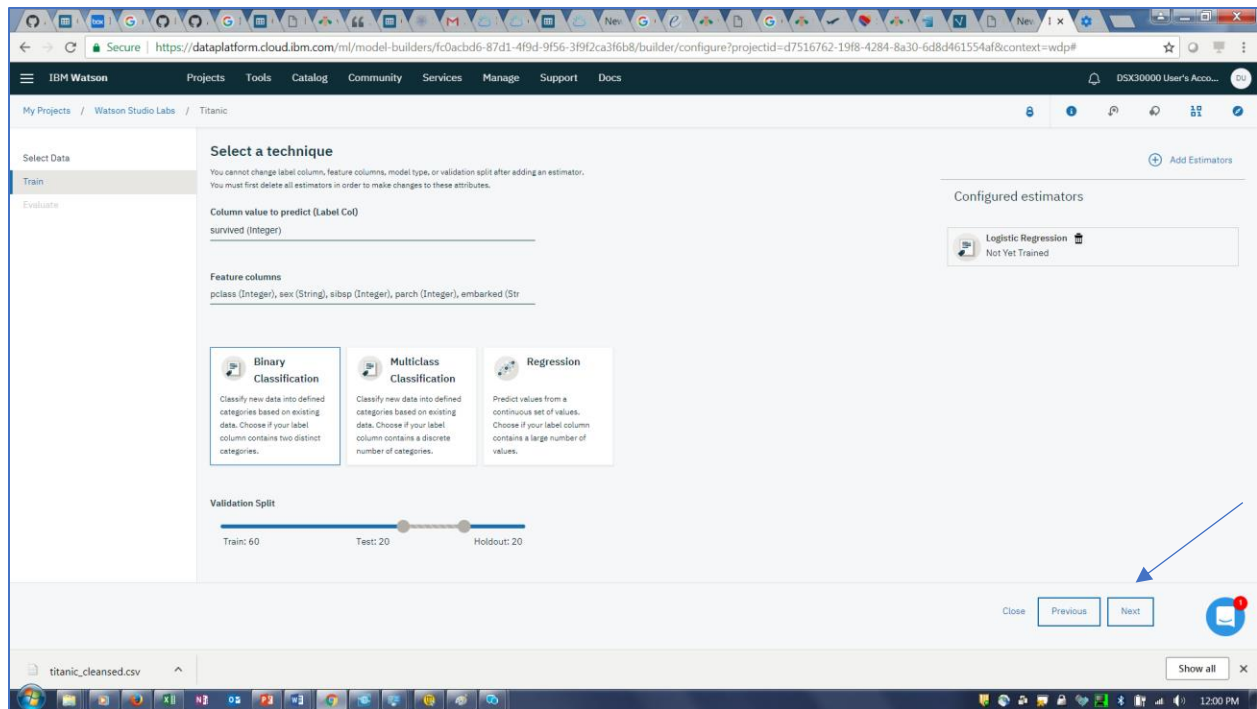
Standardize the training features before fitting the model
☐ `standardization`

Threshold (double)
0.5



 [Cancel](#) [Save](#)

9. Select the **Next** button.

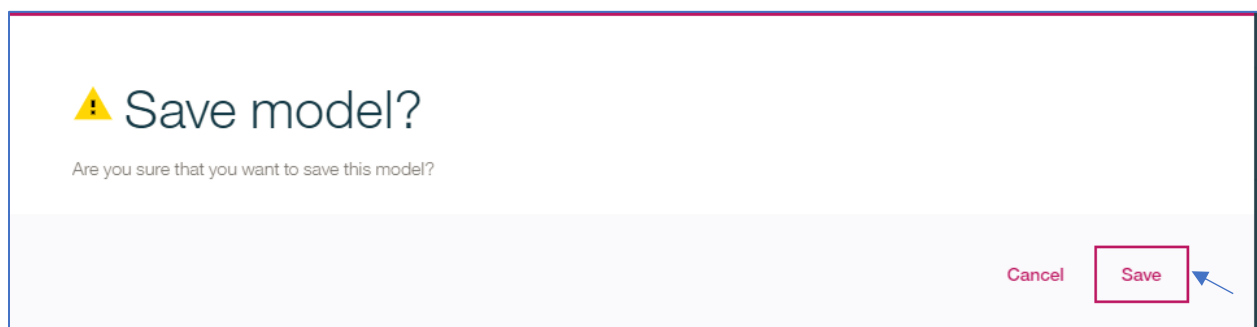


10. The system trains and evaluates each model. If more than one model was selected, the models would be listed in descending order of quality with the best result at the top. Note: if a model fails to run (rare, but happens), select Previous, delete the estimator, and re-add it. Then run again. Click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

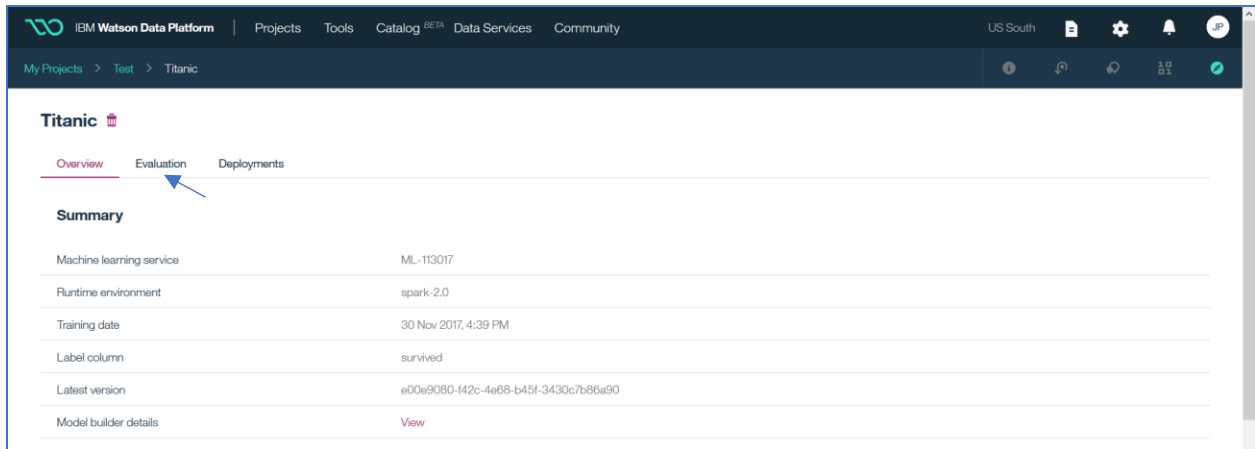
The screenshot shows the 'Select model' screen with a table of model performance metrics. A blue arrow points to the 'LogisticRegression' row, and another blue arrow points to the 'Save' button at the bottom right.

| ESTIMATOR TYPE | STATUS | PERFORMANCE | AREA UNDER ROC CURVE | AREA UNDER PR CURVE | LAST EVALUATION |
|--------------------|---------------------|-------------|----------------------|---------------------|-----------------------|
| LogisticRegression | Trained & Evaluated | Good | 0.88379 | 0.87048 | 22 Jul 2018, 12:08 PM |

11. Click **Save** again on the next screen.



12. The system displays the model training summary. Click on Evaluation.



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

My Projects > Test > Titanic

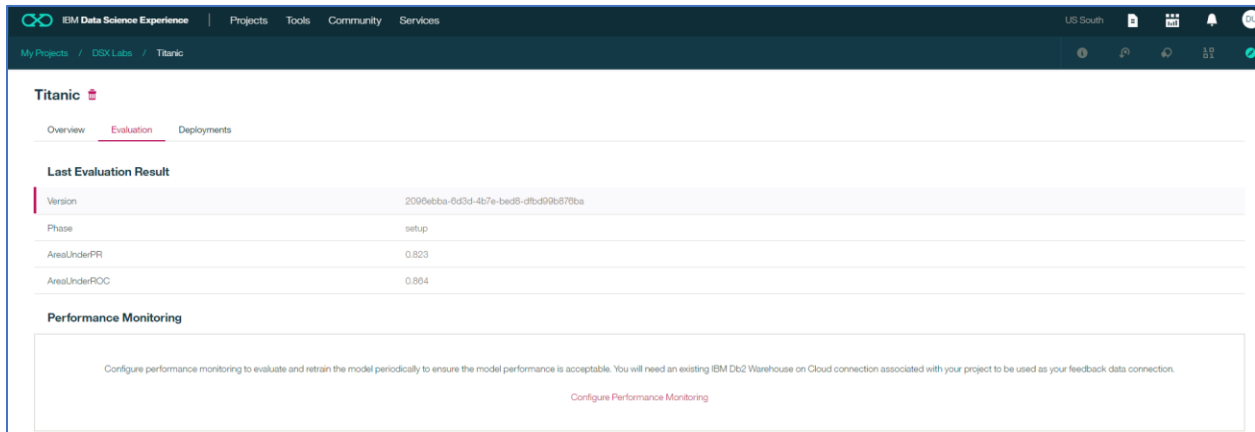
Titanic

Overview **Evaluation** Deployments

Summary

| | |
|--------------------------|--------------------------------------|
| Machine learning service | ML-113017 |
| Runtime environment | spark-2.0 |
| Training date | 30 Nov 2017, 4:39 PM |
| Label column | survived |
| Latest version | e00e9080-f42c-4e68-b45f-3430c7b86a90 |
| Model builder details | View |

13. The system displays the recorded evaluation statistics for the run. You can also set up Continuous Learning (Performance Monitoring) on this screen. We will not do this now.



IBM Data Science Experience | Projects Tools Community Services US South

My Projects / DSX Labs / Titanic

Titanic

Overview **Evaluation** Deployments

Last Evaluation Result

| | |
|--------------|--------------------------------------|
| Version | 2090ebba-6d3d-4b7e-bcd5-dfbd90b570ba |
| Phase | setup |
| AreaUnderPR | 0.823 |
| AreaUnderROC | 0.864 |

Performance Monitoring

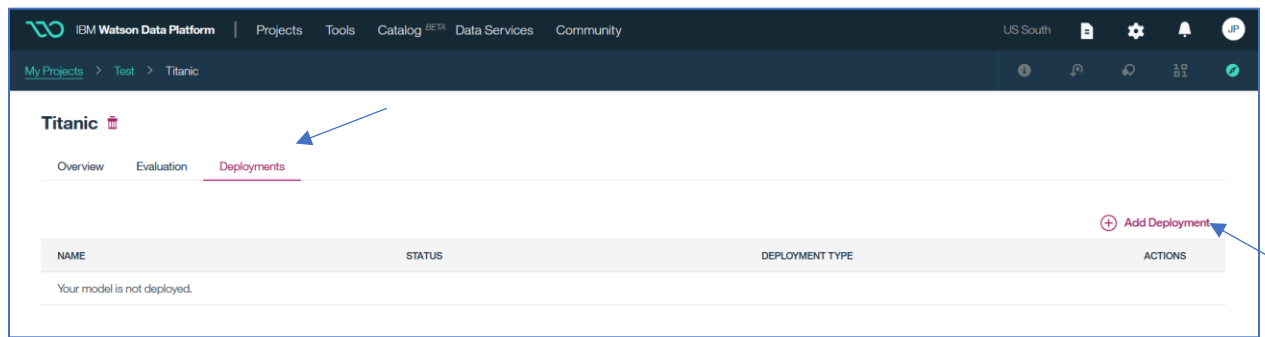
Configure performance monitoring to evaluate and retrain the model periodically to ensure the model performance is acceptable. You will need an existing IBM Db2 Warehouse on Cloud connection associated with your project to be used as your feedback data connection.

[Configure Performance Monitoring](#)

Step 3: Deploying a Model

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

1. Select the **Deployments** Tab
2. Click on **Add Deployment**



3. Enter Titanic_Deployment for **Name**, optionally a **Description**, select the **Web Service** radio button, and click on **Save**.

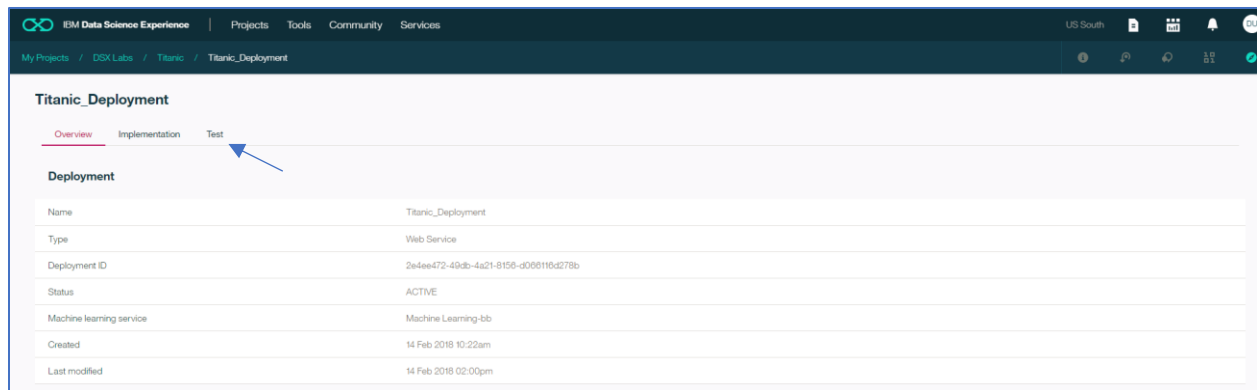
The screenshot shows the 'Define deployment details' form. The 'Name' field contains 'Titanic_Deployment'. The 'Description' field is empty. The 'Deployment type' section has two radio buttons: 'Web service' (selected) and 'Batch'. A blue arrow points to the 'Save' button at the bottom right of the form.

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

The screenshot shows the IBM Watson Data Platform interface. The top navigation bar includes 'My Projects', 'Test', and 'Titanic'. The 'Titanic' project page has tabs for 'Overview', 'Evaluation', 'Deployments', and 'Lineage'. The 'Deployments' tab is active, showing a table with columns: NAME, STATUS, DEPLOYMENT TYPE, and ACTIONS. A message states 'Your model is not deployed.' A blue arrow points to the 'Titanic_Deployment' entry in the table.

| NAME | STATUS | DEPLOYMENT TYPE | ACTIONS |
|--------------------|----------------|-----------------|---------|
| Titanic_Deployment | DEPLOY_SUCCESS | Web Service | |

5. The system displays information about the deployed service. Click on **Test** to test out the API.



6. Enter values for the following fields:

pclass - 1

sex – female

sibsp (number of siblings or spouses) – 0

Scroll down continue entering values for:

parch (number of parents or children) – 0

fare - 200

embarked – S

Age_Bucket - 1

and then click on **Predict**. Note that any values inputted for any of the fields not included in the model parameters (e.g. name) will not affect the prediction.

Titanic_Deployment

Overview Implementation Test

Enter input data

☰

📄

pclass

1

name

sex

female

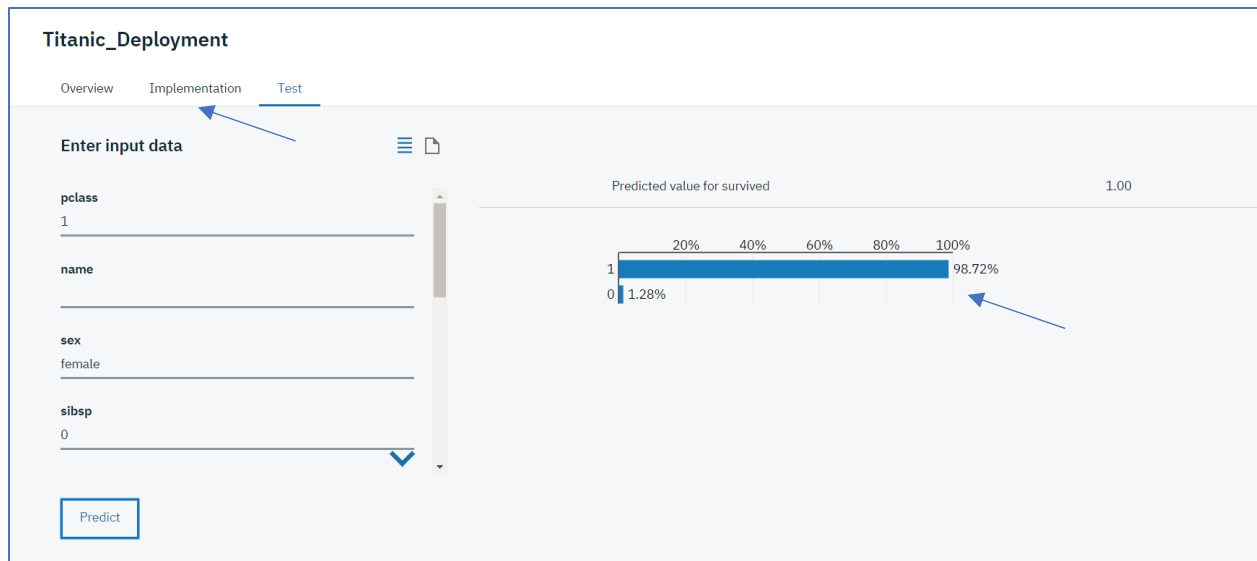
sibsp

0

✓

Predict

7. The predicted result should be returned which indicates that the model has been successfully deployed. Now click on the **Implementation** tab.



8. The Implementation panel provides information for the application developers to invoke the deployed model. It includes sample code in various programming languages and the scoring endpoint to be used when invoking the web service. Open Windows Notepad to copy and paste the scoring endpoint. We will be using this endpoint in Step 4.

The screenshot shows the 'Titanic_Deployment' interface with the 'Implementation' tab selected. It displays the 'Scoring End-point' as `https://ibm-watson-milmybluemix.net/v3/wml_instances/c4862b9a-5cb3-4d8e-9884-3255e55464e5/published_models/11465e1f-818a-4239-8db6-37312d8d32a7/deployments/2e4ee472-49db-4a21-8156-d0661b8d278b/online`. Below this, it shows 'Authorization: Bearer <token>' and 'Content-type: application/json'. A blue arrow points to the 'Authorization' field. Under 'Code Snippets', there are tabs for 'cURL', 'Java', 'JavaScript', 'Python', and 'Scala'. The 'cURL' tab is active, showing a cURL command to retrieve an authentication token and a TODO comment to manually define and pass values to be scored below.

```
# retrieve your $WML_SERVICE_CREDENTIALS_USERNAME, $WML_SERVICE_CREDENTIALS_PASSWORD, and $WML_SERVICE_CREDENTIALS_URL from the
# Service credentials associated with your IBM Cloud Watson Machine Learning Service Instance

curl --basic --user $WML_SERVICE_CREDENTIALS_USERNAME:$WML_SERVICE_CREDENTIALS_PASSWORD $WML_SERVICE_CREDENTIALS_URL/v3/identity/token

# the above CURL request will return an auth token that you will use as $WML_AUTH_TOKEN in the scoring request below
# TODO: manually define and pass values to be scored below
curl -X POST --header 'Content-Type: application/json' --header 'Accept: application/json' --header 'Authorization: Bearer $WML_AUTH_TOKEN' -d '{"fields": ["pclass", "name", "sex", "sibsp", "parch", "ticket", "fare", "embarked", "Age"]
```

Step 4: Deploy a simple web front-end to invoke the Watson Machine Learning service

This section will provide an example of a simple Python Flask web front-end application that invokes the Titanic scoring API demonstrating embedding machine learning in a web app. You

will click on a link below that will deploy the sample Python web application into your IBM Cloud account. A toolchain will be set up for continuous delivery of the application. The application code will be cloned from a public Git repository into a private Git repo in your account that will be set up as part of the toolchain. Each time you commit changes to the repo, the app will be built and deployed.

The toolchain uses tools that are part of the Continuous Delivery service. If an instance of that service isn't already in your account, when you click **Deploy**, it is automatically added with the free [Lite](#) plan selected.

You will need to configure the application to provide the credentials for your Watson Machine Learning service, and to provide the scoring endpoint.

1. Click on the **Deploy to IBM Cloud** link below to deploy a sample Python Flash web application into your IBM Cloud account. Note you may get this message – “*An IBM Cloud account is required. To get started, click Log In or Sign Up at the top of this page*”. If you get this message, click on **Log In**.

[Deploy to IBM Cloud](#)

2. Scroll down to the bottom. Click on the **Create+** button to create an IBM Cloud API key.

The Delivery Pipeline automates continuous deployment.

App name:

IBM Cloud API Key:

Region:

Organization:

Space:

Create+

*The value is required.

*The value is required.

*The value is required.

*The value is required.

*The value is required.

3. Click on the **Create** button.

×

Create API key

The following API key will be created to deploy your application from the Delivery Pipeline.

API Key for Titanic-20180722183628794

API keys can be managed from **Manage > Security > Platform API keys**.

Cancel

Create

- Please wait until the Region, Organization, and Space are filled in. **Note that these must match the corresponding Region, Organization, and Space where the Titanic model was deployed.** Switch the **Region** to US South, the Organization should display the userid, and the space should be filled in as well. If this is not the case, try a different **Region**. Click on **Deploy**.

The Delivery Pipeline automates continuous deployment.

App name:

Titanic-20180722183628794

i

IBM Cloud API Key:

.....

eye

Create +

i

Region

Organization

Space

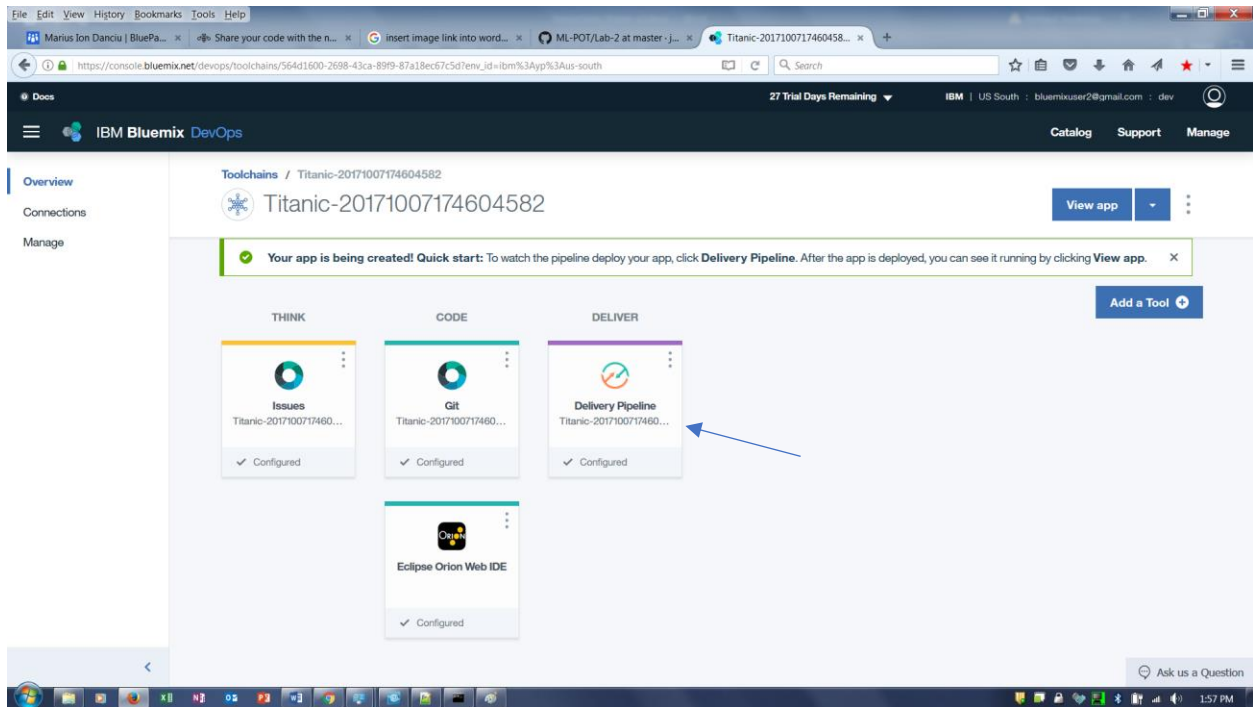
US South (Production) ▼

DSXUser30000@gm... ▼

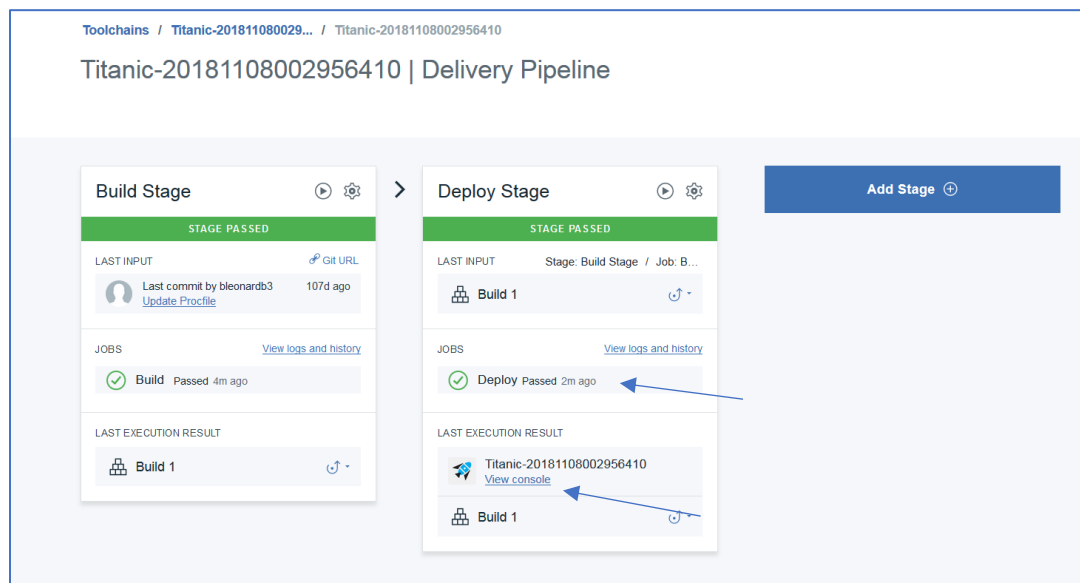
dev ▼

Deploy

5. Your app is being created! To watch the pipeline deploy your app, click **Delivery Pipeline**.



6. After the app is deployed successfully (should say Deploy Passed in the Deploy stage-may take about 2 minutes), view the running app by clicking on **View Console**



7. Click on **Visit App URL**

Cloud Foundry apps /

Titanic-20180925140946060 ● This app is awake. [Visit App URL](#)

Org: DSXUser52000@gmail.com Location: US South Space: dev

Runtime

| Metric | Value | Status |
|------------------------|--------|---|
| BUILDPACK | Python | |
| INSTANCES | 1 | All instances are running Health is 100% |
| MB MEMORY PER INSTANCE | 128 | |
| TOTAL MB ALLOCATION | 128 | 128 MB still available |

8. The web form collecting the Titanic passenger data should appear. Note that the application is not functional until we connect it to the Watson Machine Learning service so if you Submit you will get an error! Close the Titanic Prediction browser tab.

To determine the survival prediction, please enter the following:

Passenger Class:

☐ First

☐ Second

☐ Third

Name:

Gender:

☐ Male

☐ Female

Number of siblings/spouses:

Number of parents/children:

Ticket:

Fare:

Embark Location:

☐ South Hampton

☐ Cherbourg

☐ Queenstown

Age:

9. We are now going to connect the application to the Watson Machine Learning service that was created earlier. Scroll down until you see the Connections panel. Click on **Create Connection**.

The screenshot shows the IBM Cloud console interface for a Cloud Foundry application. The application is named 'Titanic-20171130220547025' and is in a 'Running' state. It has 1 instance and is using a Python buildpack. The connections panel is visible, showing 'No services are connected to this app' and a 'Create connection' button, which is highlighted by a blue arrow.

10. You should see at least 3 services listed, a Cloud Object Storage service, a Spark service, and a Watson Machine Learning service. Point the cursor on the **Machine Learning** service for your application, and then click on **Connect**.

The screenshot shows the 'Connect Existing Compatible Service' dialog in the IBM Cloud console. It displays a table of available services:

| SERVICES | RESOURCE GROUP | PLAN | SERVICE OFFERING |
|-------------------------|----------------|------|----------------------|
| cloud-object-storage-id | Default | Lite | Cloud Object Storage |
| Machine Learning | Default | Lite | Machine Learning |
| Spark | -- | Lite | Apache Spark |

The 'Machine Learning' service is highlighted, and a blue arrow points to the 'Connect' button.

11. A **Connect IAM-enabled service** pop up will appear. Just click **Connect**.

Connect IAM-Enabled Service

To connect, you can customize the ServiceID and access role used for this binding. Restaging your app is required to connect this service and may result in application downtime.

Access Role for Connection ⓘ

Writer

Service ID for Connection (Optional) ⓘ

Auto Generate

Cancel Connect

12. A **Restage app** pop up will appear. Click on **Restage**.


Restage app

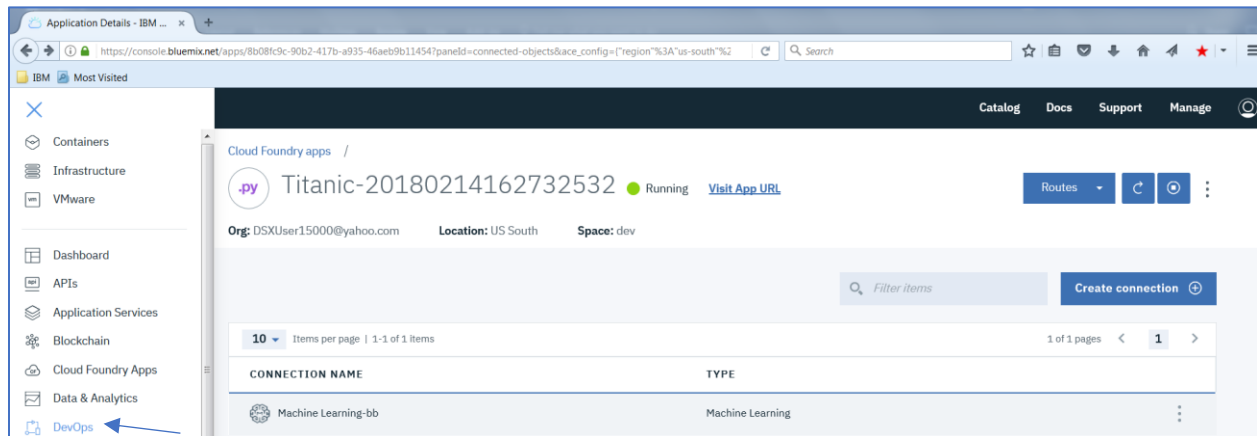
Your 'Titanic-20171007174604582' app must be restaged to use the new 'Watson Machine Learning' service. Restaging makes this service available for use. Do you want to restage it now?

Cancel Restage

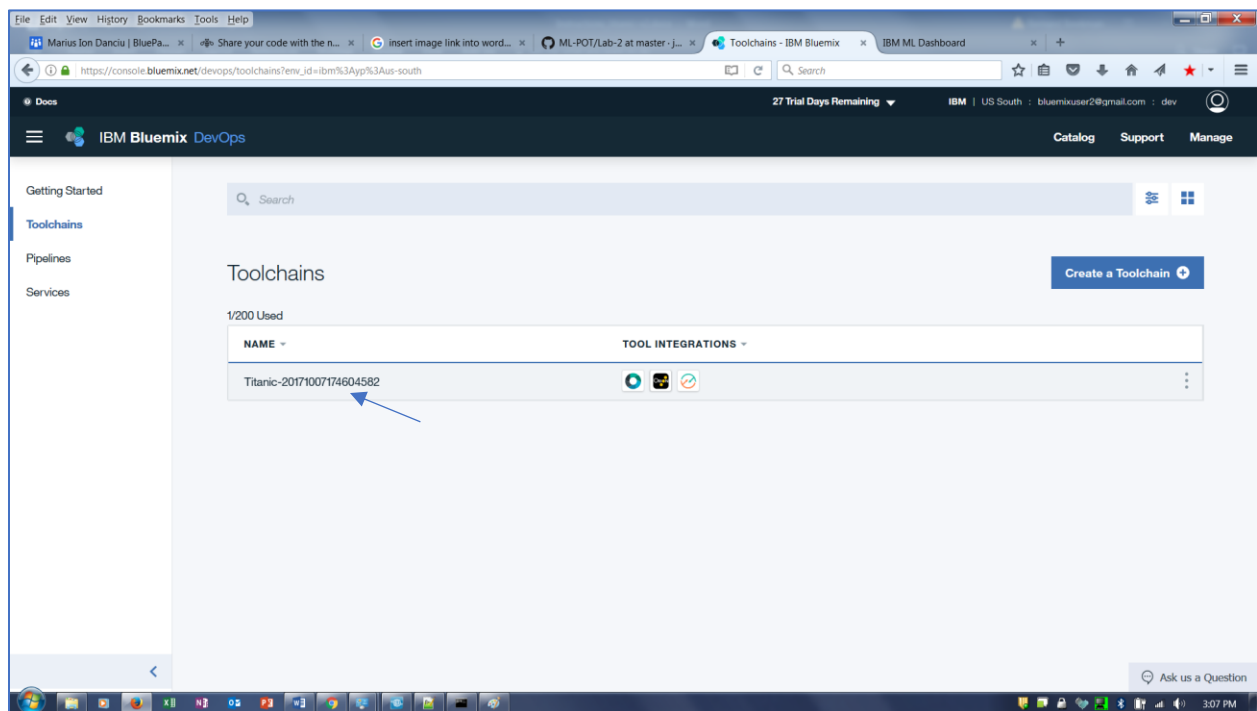
13. Wait for the application status to change to  This app is awake , or something similar.

14. We now have tied the web application to the Watson Machine Learning service. Note that the Watson Machine Learning service could have more than one deployed model available to select and then embed in the web application. In our case, we have only one deployed model. We now need to copy the scoring endpoint of that deployed model

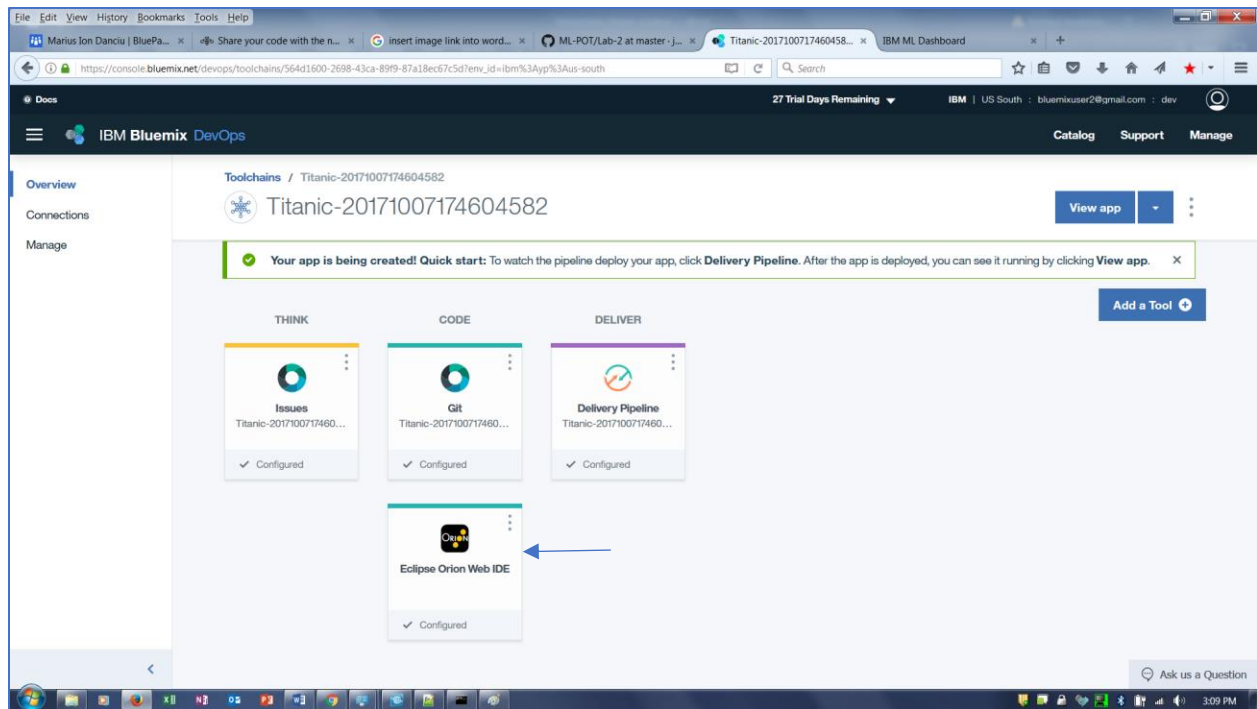
(previously copy and pasted into Notepad) and paste it in the web application code.
Click on the  icon and click on DevOps in the pulldown to navigate to the Toolchain.



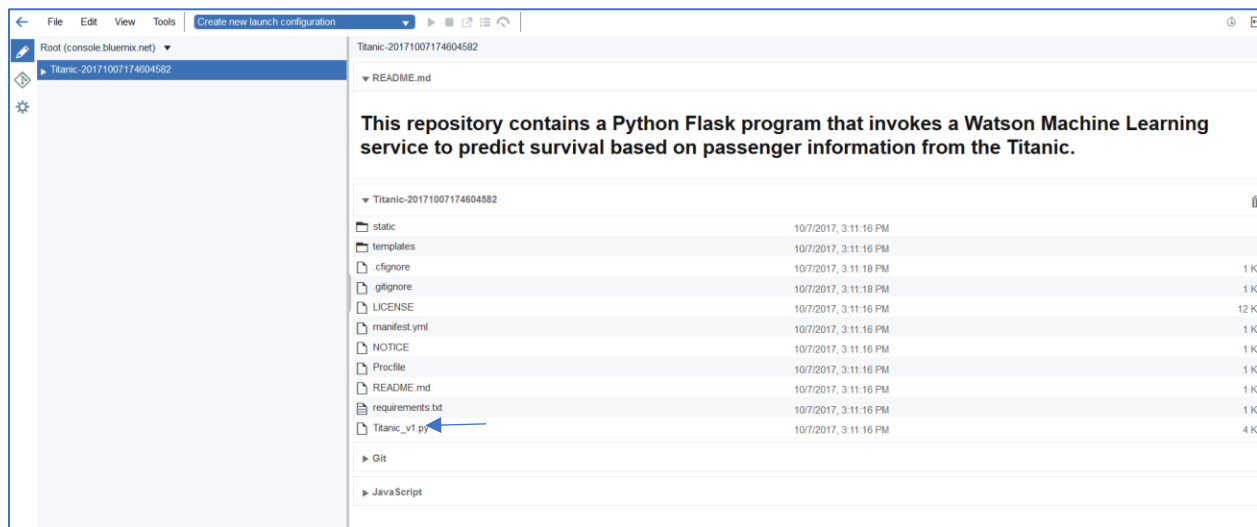
15. We are now going to paste the scoring endpoint into the application code. Click on the Toolchain (Titanic-2018xxxxx below).




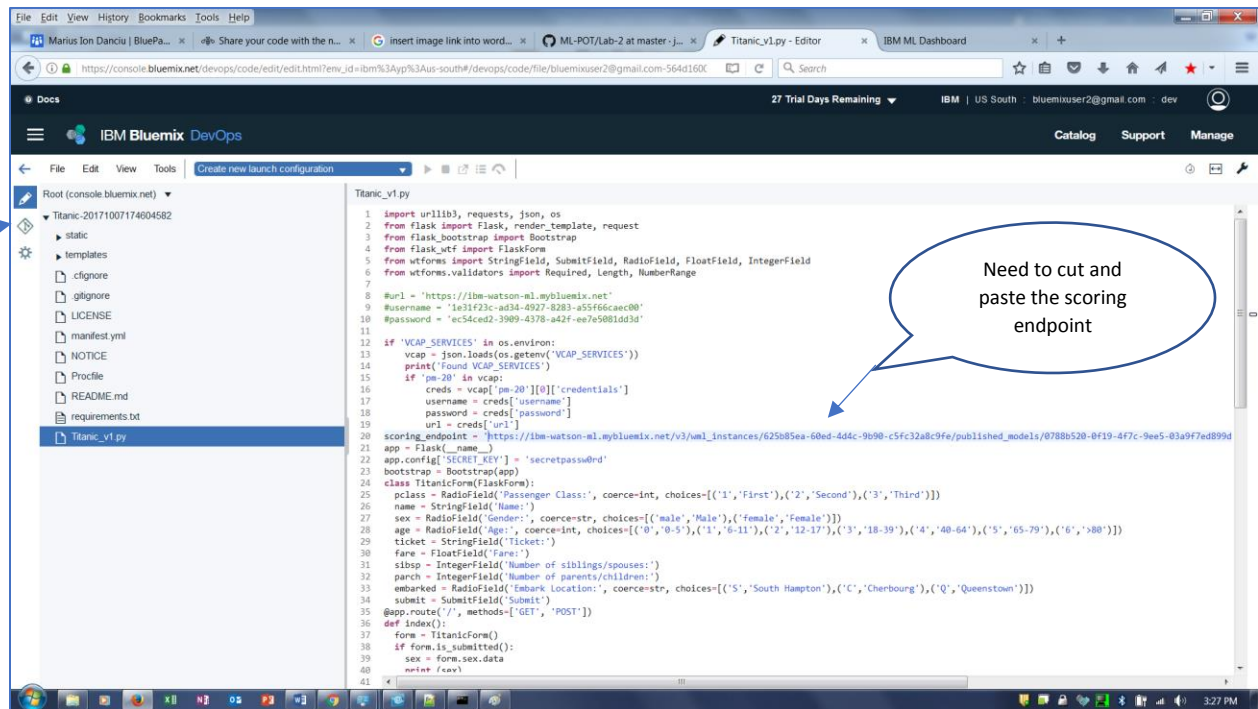
16. Click on the Eclipse Orion Web IDE.



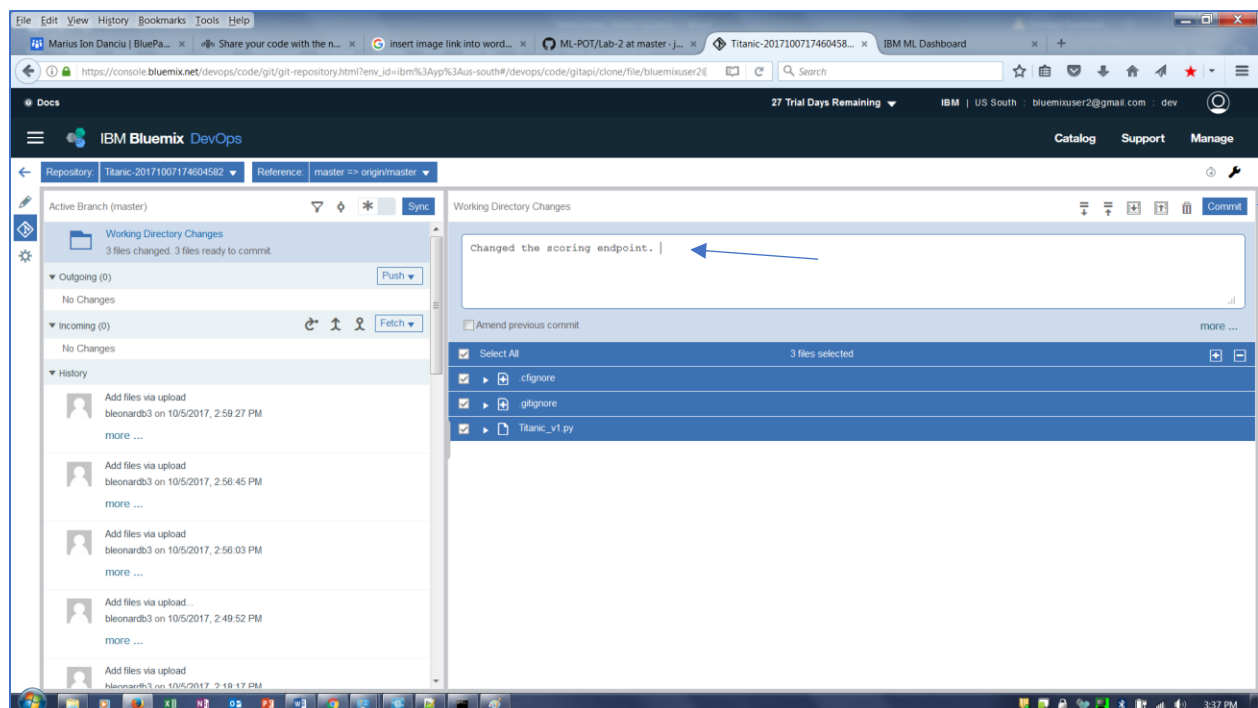
17. Click on the `Titanic_v1.py` file. This is a python source file.



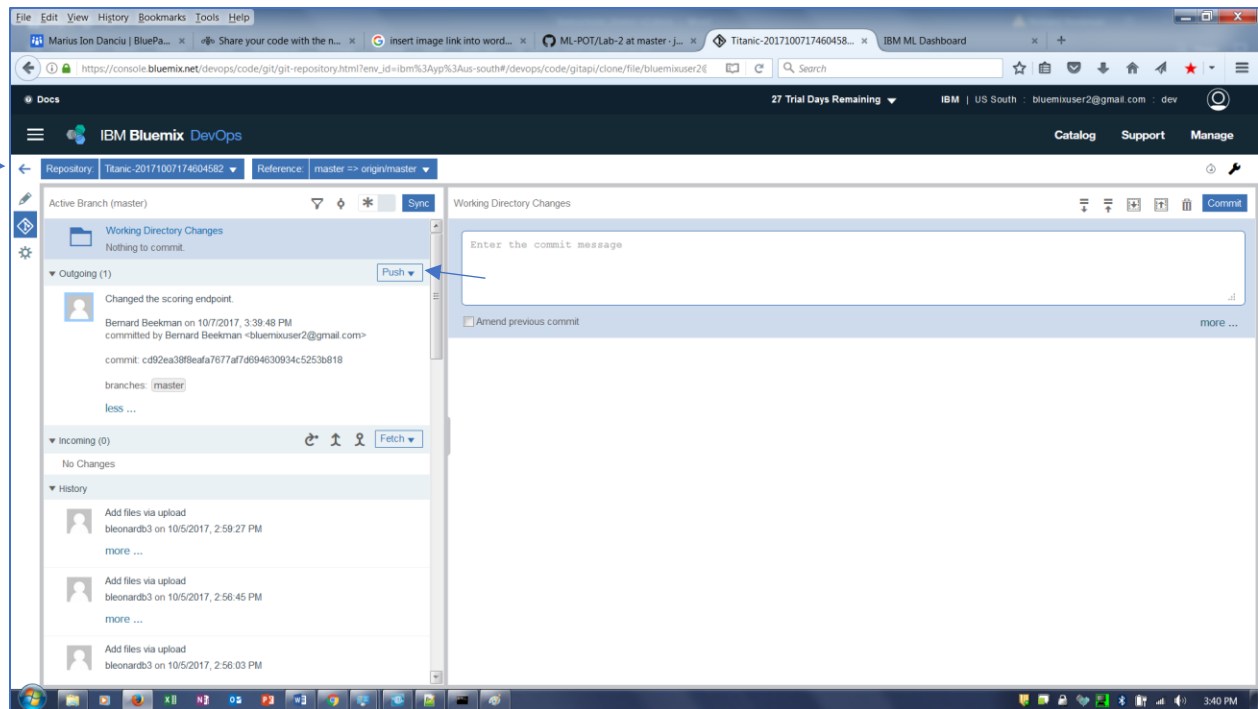
18. Go back to the Notepad file, and copy the scoring endpoint to the clipboard. Look around line 20 in the `Titanic_v1.py` file for the “scoring endpoint =”. Select the scoring endpoint value in line 20 (starting with `https://` may want to use Shift-End to get to the end of the line, and then back up one space to not select the endpoint quote – if you do just make sure to put it back in). Enter Ctrl-V to paste the new scoring endpoint from your Notepad file. Enter Ctrl-S or File > Save to save the file. Then click on the  icon on the top left.



19. The next step is to commit the change to the git repository. Enter “Changed the Scoring Endpoint” in the Enter Commit Message field, and then click on **Commit**.



20. Then click on **Push** to push the changes to the central Git repo which will start the build and deploy of the application. Click on the left arrow to return to the Toolchain.



21. Click on the **Delivery Pipeline** to view status of the deployment as before. Once the Deployment status shows **Deploy passed now** it shouldn't take longer than 2 minutes so reload the browser in case the UI didn't update). You can see the running app by **View Console**, and then **Visit App URL** (refer to steps 6,7 above).

22. The web form should appear. Enter data in all the fields and click on the **Submit** button. (the submit button is located at the bottom of the web form – you may need to scroll).

To determine the survival prediction, please enter the following:

Passenger Class:

☒ First

☐ Second

☐ Third

Name:

Gender:

☒ Male

☐ Female

Number of siblings/spouses:

Number of parents/children:

Ticket:

Fare:

Embark Location:

☒ South Hampton

☐ Cherbourg

☐ Queenstown

Age:

☒ 0-5

☐ 6-11

☐ 12-17

23. You should see something similar to the following depending on the values of the input fields that you entered. Click on the **Try Again!**, if you want to experiment with different inputs.

Titanic Prediction

prediction:survived
probability: 0.827966430684

[Try Again!](#)