# **Deploy**

## Lab overview

*Note: This lab requires that you have completed Lab 06 Analyze: AutoAI.*

In the previous lab, you created a model from both a notebook and from AutoAI. You will learn how to Deploy these models in this lab.

In our scenario, Trade Co. uses these steps to deploy their machine learning models into production.



## Persona represented in this lab

The Developer persona is the likely role to perform the various Deploy tasks in this lab. However, the Data Scientist persona could perform these tasks as well.

|  |  |
| --- | --- |
| **Persona (Role)** | **Capabilities** |
| A close up of sunglasses  Description generated with high confidence  Developer | Developers create and maintain the end-user applications that utilize the output from all the other personas on the CPD platform. |

## Logging into the CPD web client (if you have not already done so)

1. If you are starting this lab stand-alone (without going through previous labs) do the following:
2. Double-click the desktop icon: Cloud Pak for Data Web Client.

A picture containing object, clock

Description automatically generated

1. The CPD web client GUI displays as shown. Use cpduser and cpdaccess for the *Username* and *Password* and click Sign in.

A screenshot of a cell phone

Description automatically generated

## Reviewing the notebook deployment space

Think of model deployment as the equivalent of writing a self-service application that takes the model and makes it available through a REST API interface. Application developers access and consume the model through the same interface. While this is a manual process in most organizations, Cloud Pak for Data can automate deploying and maintaining models without writing a single line of code.

CPD eliminates the need for to do the following:

* + Writing code to perform the above capability and using a runtime to deploy it.
  + Creating a runtime on bare metal machines that require OS installation, network, storage, etc.
  + Creating a runtime on a virtual machine in VMware on Intel, or IBM POWER VM® on a POWER platform.
  + Creating a runtime in Docker or CRI-O requires someone to build the image and deploy it on one of the above platforms.

Each of the above requires manpower and machine resources. Using CPD, you can bypass this and quickly harvest insight from your data in a repetitive manner by integrating it with your CI/CD pipeline.

### **Associating a deployment space**

1. In the CPD web client, click the Navigation Menu a Projects.

A picture containing clock, meter

Description automatically generatedA picture containing dark, clock, light, meter

Description automatically generated

1. Select the project: CPD Workshop Analytics Project.

A screenshot of a cell phone

Description automatically generated

1. Click section Settings.

A screenshot of a cell phone

Description automatically generated

1. Scroll down to find Associated deployment space,

Notice it says this project is currently not associated with a deployment space.

Click Associate a deployment space,

A screenshot of a cell phone

Description automatically generated

Click section Existing,

Click Deployment Space deployment-space-analytics-project-workshop,

Click Associate,

A screenshot of a cell phone

Description automatically generatedA picture containing drawing

Description automatically generated

|  |  |
| --- | --- |
| A close up of sunglasses  Description generated with high confidence  Developer | Note: this deployment space was generated by running the notebook in the previous lab.      However, you could have created a deployment space through the CPD web client as well. |

1. The project should now show that it is associated with this deployment space. (You may need to scroll down to see this.)

A screenshot of a social media post

Description automatically generated

1. Scroll up and click Assets,

A screenshot of a cell phone

Description automatically generated

1. If the data asset add screen is open, close it.

A screenshot of a cell phone

Description automatically generated

1. Click + New data asset (This is how you open the data asset screen.)

A close up of a logo

Description automatically generated

1. Click Files, then select the two files: model\_batch\_score.csv and model\_eval.csv.

*(Note: these files were created by the notebook),* then click ellipses a Add as data asset and Apply.

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

1. Find the newly added data asset model\_batch\_score.csv, click on ellipses and Promote.

A screenshot of a computer screen

Description automatically generated

1. Click Promote to space.

A screenshot of a cell phone

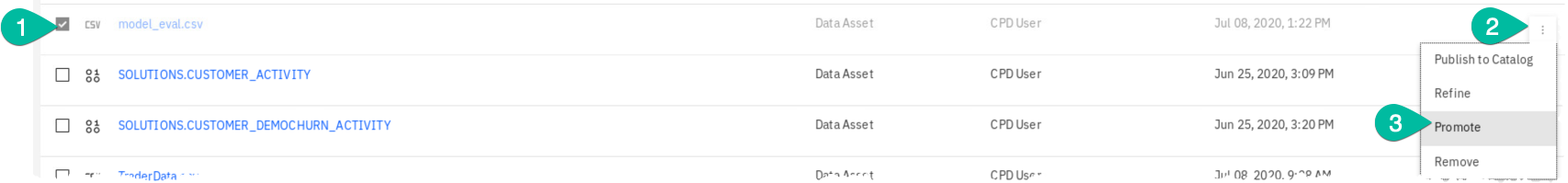
Description automatically generated

1. The data asset is now promoted to the associated Deployment Space.

A screenshot of a cell phone

Description automatically generated

1. Repeat for the data asset model\_eval.csv.



### **Working in the deployment space**

Now you can go to the deployment space to work in it:

1. Click Navigation menu a Analyze a Analytics deployments

A picture containing clock, meter

Description automatically generatedA screenshot of a cell phone

Description automatically generated

1. In this screen, note that you could create a new deployment space if you wanted to. You will not need to, however, because the one we need was created by the notebook.

Instead, click deployment-space-analytics-workshop.

A screenshot of a cell phone

Description automatically generated

1. In the Assets section, there is one Model: churn\_risk\_model (created by the notebook)

and there are two Data Assets (the ones you just promoted).

A screenshot of a social media post

Description automatically generated

1. In the Deployments section, there is one deployment: churn\_risk\_model-deployment

(which uses model churn\_risk\_model).

A screenshot of a cell phone screen with text

Description automatically generated

|  |  |
| --- | --- |
| A close up of sunglasses  Description generated with high confidence  Developer | The notebook created all of these CPD assets – part of the code that did it is shown here. |

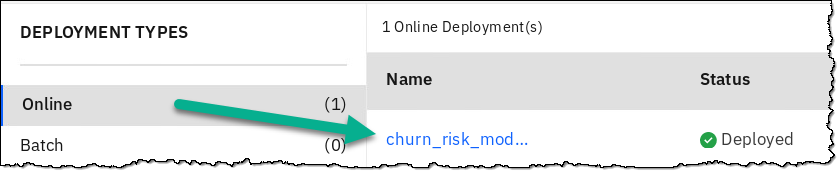
1. Return to section Assets (click on it).

Click on the model: churn\_risk\_model.

A screenshot of a cell phone

Description automatically generated

1. It shows one Online deployment (as seen previously in the Deployments section).



1. Click Batch a Create Deployment

A screenshot of a cell phone

Description automatically generated

1. Fill in name: churn-risk-model-deployment-batch.

A screenshot of a cell phone

Description automatically generated

1. Scroll down and find Select a Hardware definition. Select 1 standard CPU,4 GB RAM, and click Create.

A screenshot of a cell phone

Description automatically generated A picture containing drawing

Description automatically generated

1. You will be taken to the Jobs screen. Click Create job +.

A screenshot of a cell phone

Description automatically generated

1. Fill in the Job Name: churn-risk-model-deployment-batch-run.

A screenshot of a cell phone

Description automatically generated

1. Click on sections for INPUT – Specify input data.

A screenshot of a cell phone

Description automatically generated

1. Set the input value to model\_batch\_score.csv and Confirm.

A screenshot of a cell phone

Description automatically generated

1. Select the OUTPUT – Specify output data asset.

A screenshot of a cell phone

Description automatically generated

1. In this case, we will output to a new dataset. Enter churn-risk-model-deployment-batch-run for name and Confirm.

A screenshot of a cell phone

Description automatically generated

1. Select Create and Run.

A picture containing drawing

Description automatically generated

1. The job will queue, run, and complete.   
   *Note: You may need to refresh your browser to see the “Completed” result. You may have to wait around 5-10 minutes for the job to complete.*

A picture containing object, clock, meter

Description automatically generated

1. Click on the Start Time value when Completed.

A picture containing clock

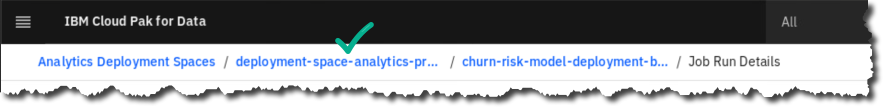
Description automatically generated

1. Review the job – select Show More at bottom right.

A screenshot of a social media post

Description automatically generated

1. After reviewing the job, click on the breadcrumb to return to the deployment space.



1. In the section Assets, find the new output file under Data Assets.

The file name is churn-risk-model-deployment-batch-run.csv.

1. On that file line, click Hover to uncover the download (down arrow) icon and Click it.

A screenshot of a cell phone screen with text

Description automatically generated

1. Save your File to the Documents folder and Save.

A screenshot of a social media post

Description automatically generated

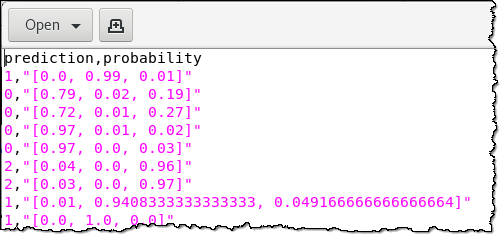
1. Review the downloaded file by clicking on the download (down arrow) icon in your Firefox browser, then select the file.

A screenshot of a cell phone

Description automatically generated

*Note: this file is located on the operating system under* ***/home/ibmuser/Downloads.***

1. The batch run output is available for review.



1. Close the file edit windows after reviewing.
2. Click Deployments.

Click Online deployment: churn\_risk\_model\_deployment.

A screenshot of a cell phone

Description automatically generated

1. Under section API reference, review the Endpoint for this model. This is what your applications can use to call this model.

A screenshot of a social media post

Description automatically generated

1. Review the code snippets for each language that can aid a Developer to easily embed the model into an application.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generated

## Deploying and testing the AutoAI model

### **Deploying the AutoAI model**

1. In the CPD web client, click the Navigation Menu a Projects.

A picture containing clock, meter

Description automatically generatedA picture containing meter

Description automatically generated

1. Select the project: CPD Workshop Analytics Project.

A screenshot of a cell phone

Description automatically generated

1. Under Assets, scroll down to Models.

On the model created from the AutoAI experiment (ChurnRisk AutoAI experiment – P4 RandomForestClassifierEstimator), click on ellipses then Promote and Promote to space.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. This promotes the model to the associated Deployment Space

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. Click Navigation menu a Analyze a Analytics deployments.

A picture containing clock, meter

Description automatically generatedA screenshot of a cell phone

Description automatically generated

1. Click deployment-space-analytics-project-workshop.

A screenshot of a cell phone

Description automatically generated

1. Under section Assets, find the model you just promoted from your project. (*Don’t click on it.)*

A screenshot of a cell phone

Description automatically generated

1. At the end of the row for this model, click Hover and click on the Deploy (rocket) icon.

A screenshot of a cell phone

Description automatically generated

1. Choose Online.

A screenshot of a cell phone

Description automatically generated

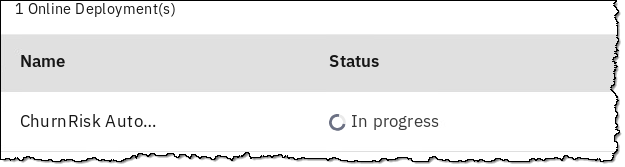
1. Name: ChurnRisk AutoAI Experiment deployment a Create.

A screenshot of a social media post

Description automatically generatedA picture containing drawing

Description automatically generated

1. You will see the online model being deployed.



1. While you are waiting for it to deploy, review the model Schema.

A screenshot of a cell phone

Description automatically generated

1. Select Deployments. The Online model is now deployed in short time with zero coding.

A screenshot of a cell phone

Description automatically generated

1. Click on the deployment.

A screenshot of a cell phone

Description automatically generated

### **Testing the AutoAI model**

1. Click on tab Test.

Fill in the first three input data fields as shown below.

A screenshot of a social media post

Description automatically generated

1. Scroll down and click on Predict.

(Note: filling in more fields would make the prediction more accurate, but we only did three fields to show you the process.)



1. Your machine learning prediction and probability is returned.

A screenshot of a cell phone

Description automatically generated

## Lab conclusion

Today, organizations need better, faster and more integrated methods for model deployment that can include online, batch and virtual testing. The Cloud Pak for Data deploy capabilities help serve this need so they can more easily infuse AI into their applications.

In our scenario, the Trade Co. developer used the platform to easily deploy and test models.



**\*\* End of Lab 07 - Deploy**

Lab by Burt Vialpando and Kent Rubin, IBM