Set up Continuous Learning Model using Watson Studio.

Goals:

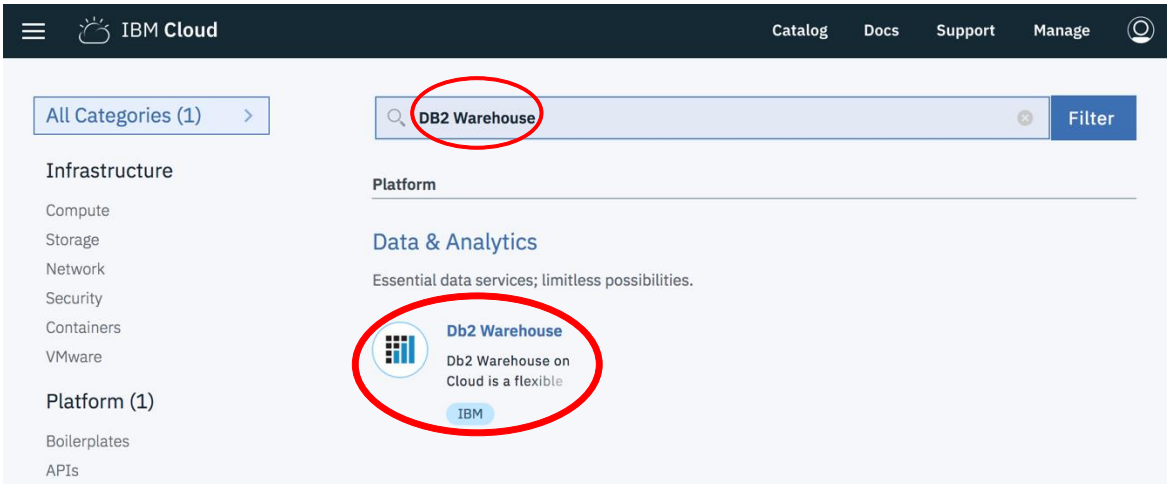
In this lab you will

- Requisition a DB2 Warehouse Service
- Upload data to your Watson Studio project.
- Train, compare, and select a machine learning model.
- Set up continuous learning capabilities.
- Deploy a machine learning model.

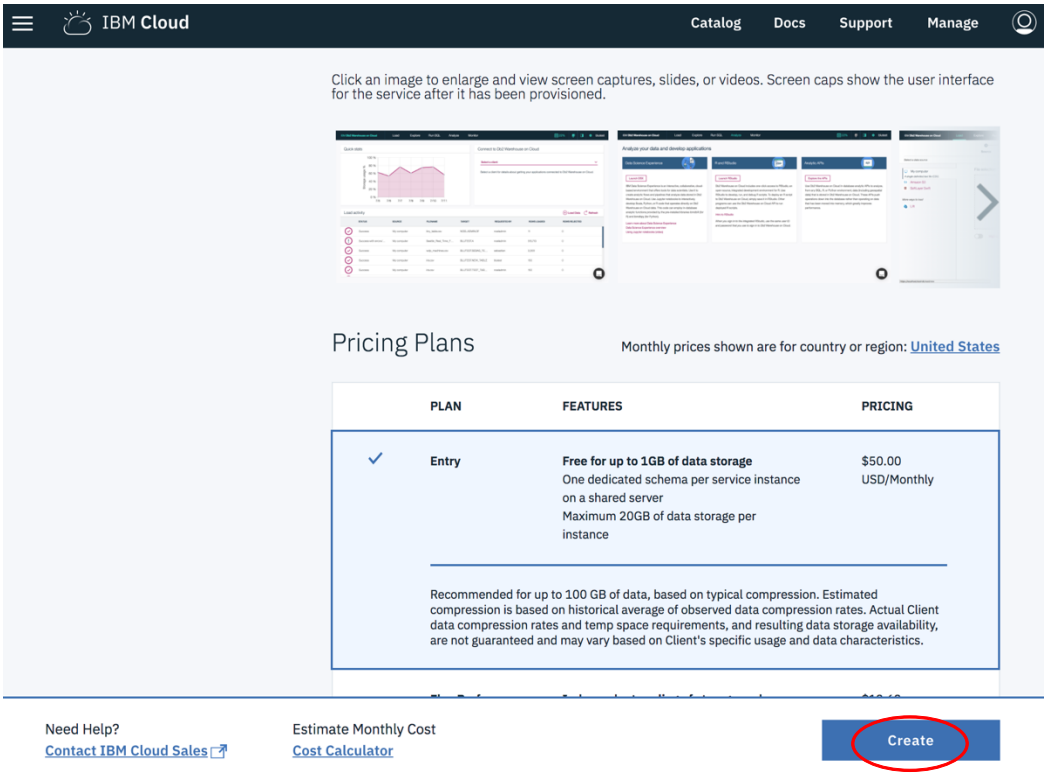
Exercise Instructions

Step 1.1 Requisition a DB2 Warehouse Service

1. Open https://console.bluemix.net/dashboard/apps/
2. Provide your username and password.
3. Access the catalog page. https://console.bluemix.net/catalog/
4. Type DB2 Warehouse into the search bar and click on the DB2 Warehouse icon.



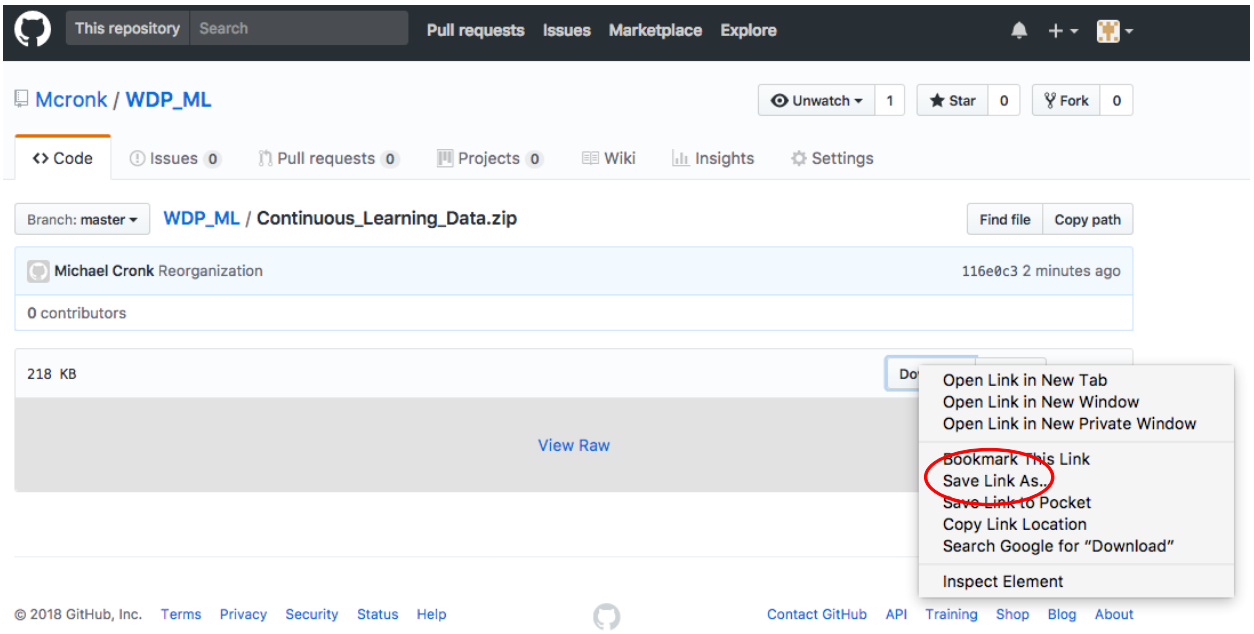
5. Scroll down, ensure that Entry is selected and click Create. Note that you will not be charged for this service for the lab.



Step 1.2 Upload a New Data Asset.

Before we build our models, we need to load data into our project.

1. Open https://github.com/bleonardb3/ML-POT/blob/master/Lab-2/data/Continuous_Learning_Data.zip
2. Right click Download and click Save Link As.

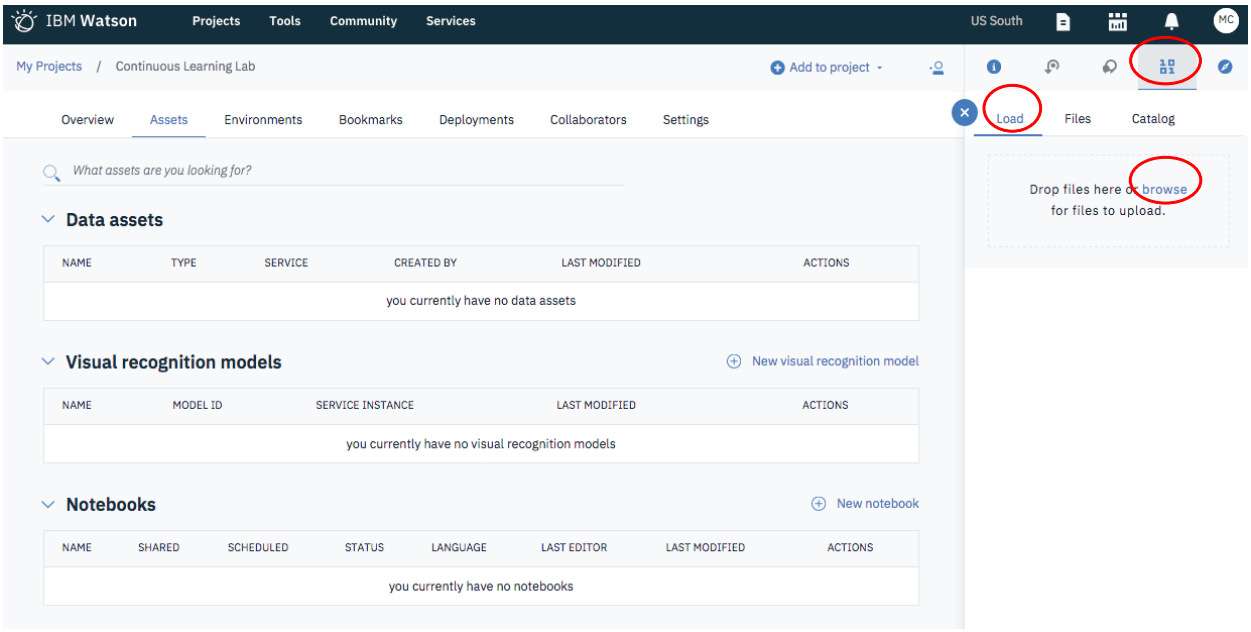


3. Save the zip file to your desktop and decompress the contents. This zip file contains data for building information for each month.

This zip file contains data for building information for each month. We will begin by building our model for the month of September and then add in data for later months to show how we can build our model to continuously adjust over time.

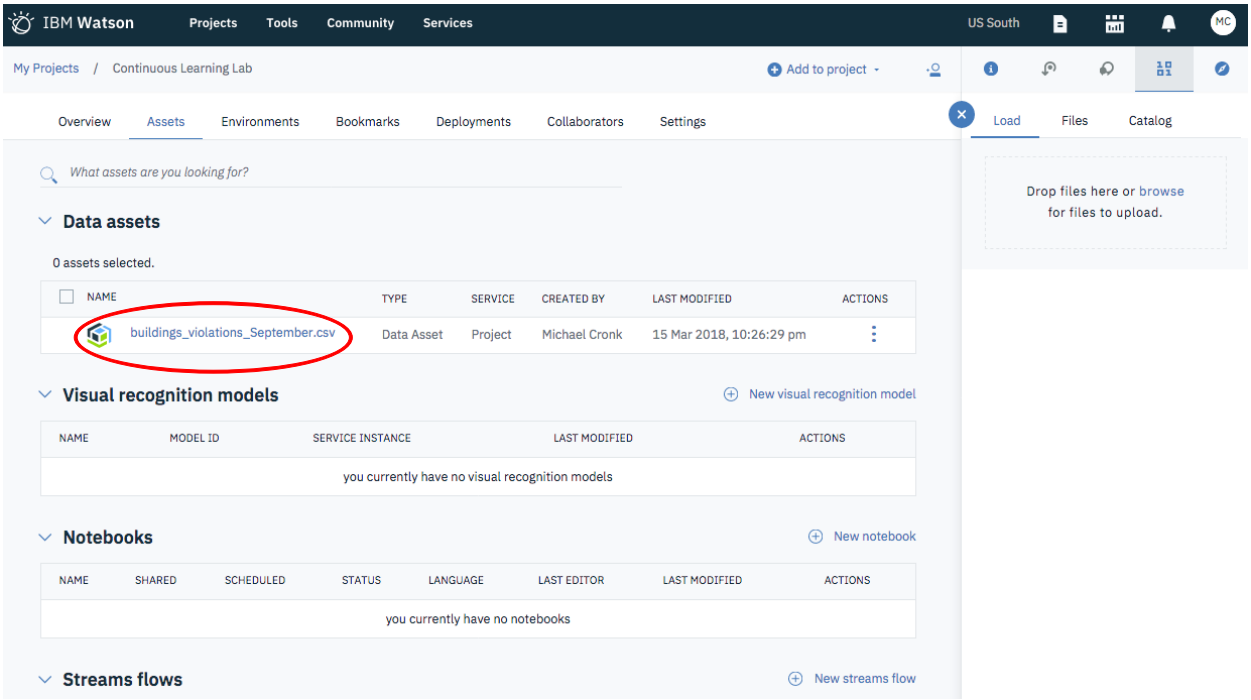
4. Return to your assets page.

5. In the top right of the screen, click the blue icon with 1's and 0's. This will open the data pallet.
6. Click on Load, browse, select the building_violations_September.csv file and click open.

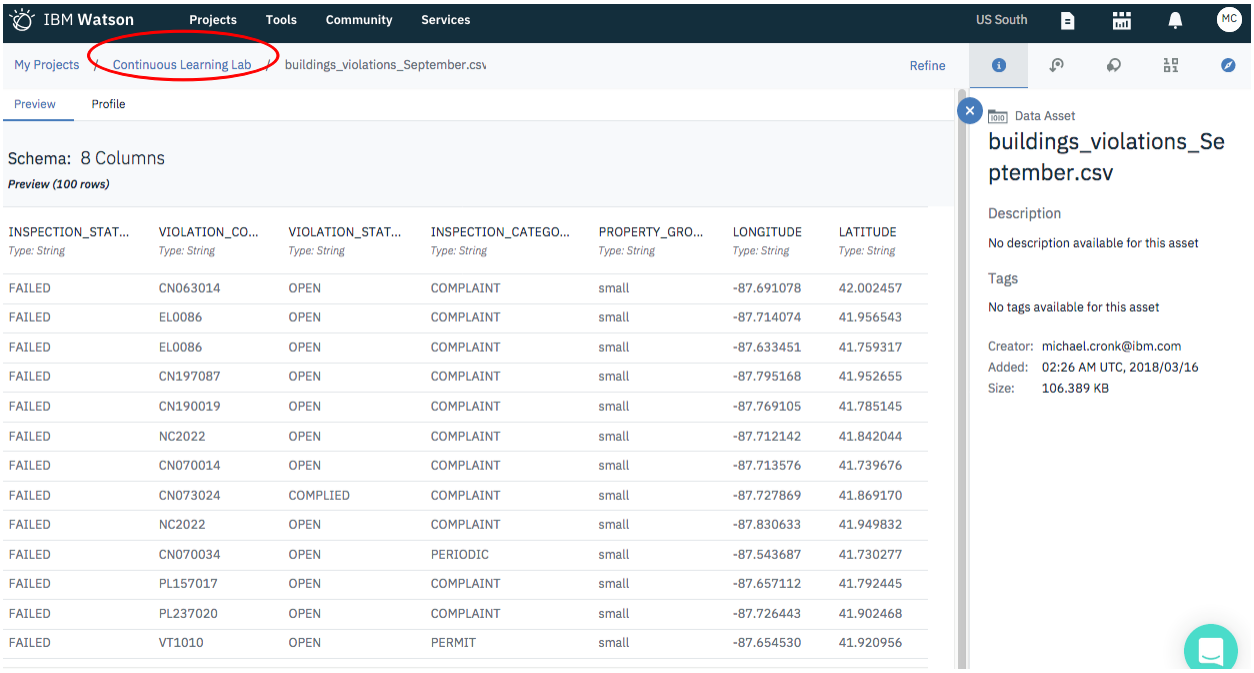


The CSV file should now be listed under "Data assets." It is now accessible by Watson Data Platform modeling tools and applications.

7. Click on building_violations_September.csv under Data assets.

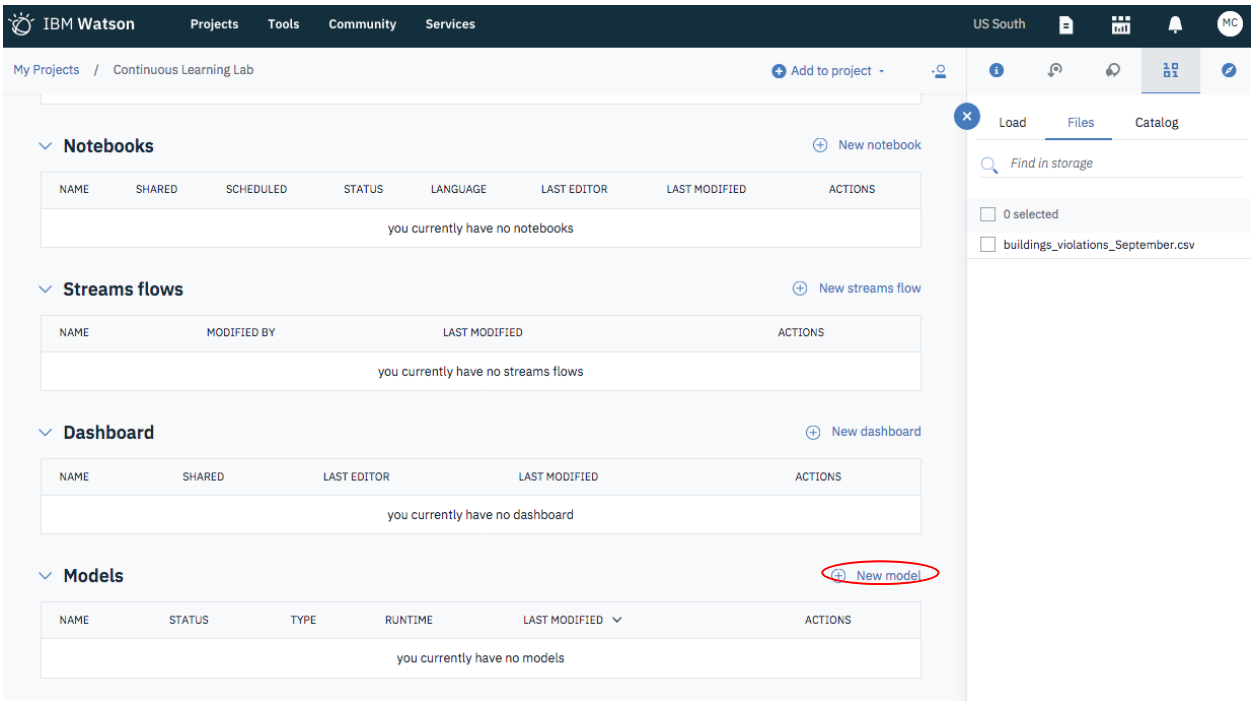


After clicking on building_violations_September.csv, you are brought to a data view. Once you are satisfied that this is the data you want, click on your project name to return to your project's assets.

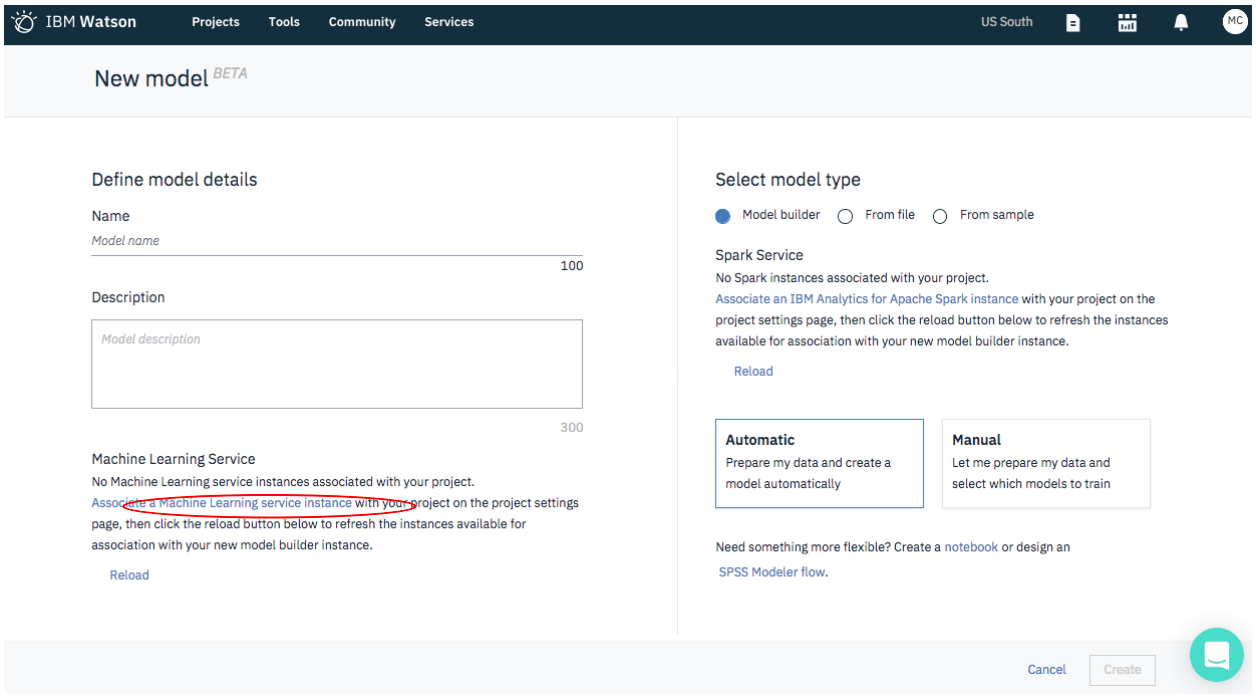


Step 1.3 Train, Compare, and Select a Machine Learning Model

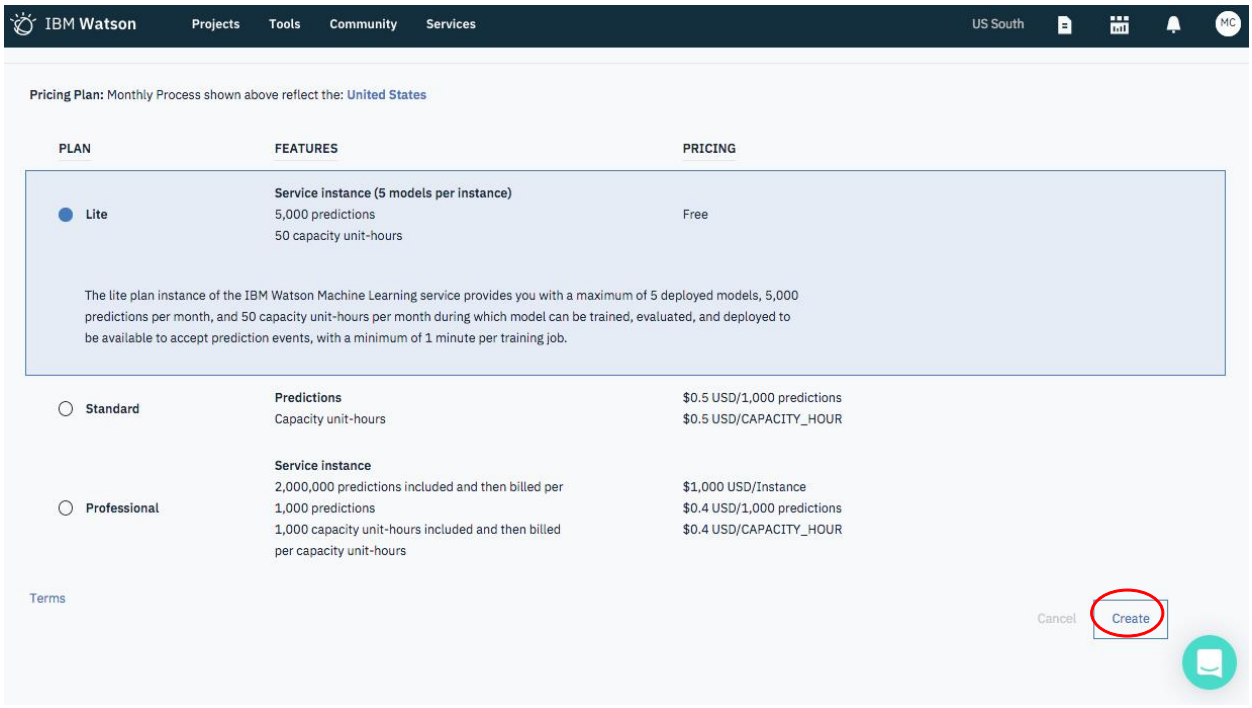
1. Scroll down and click New model.



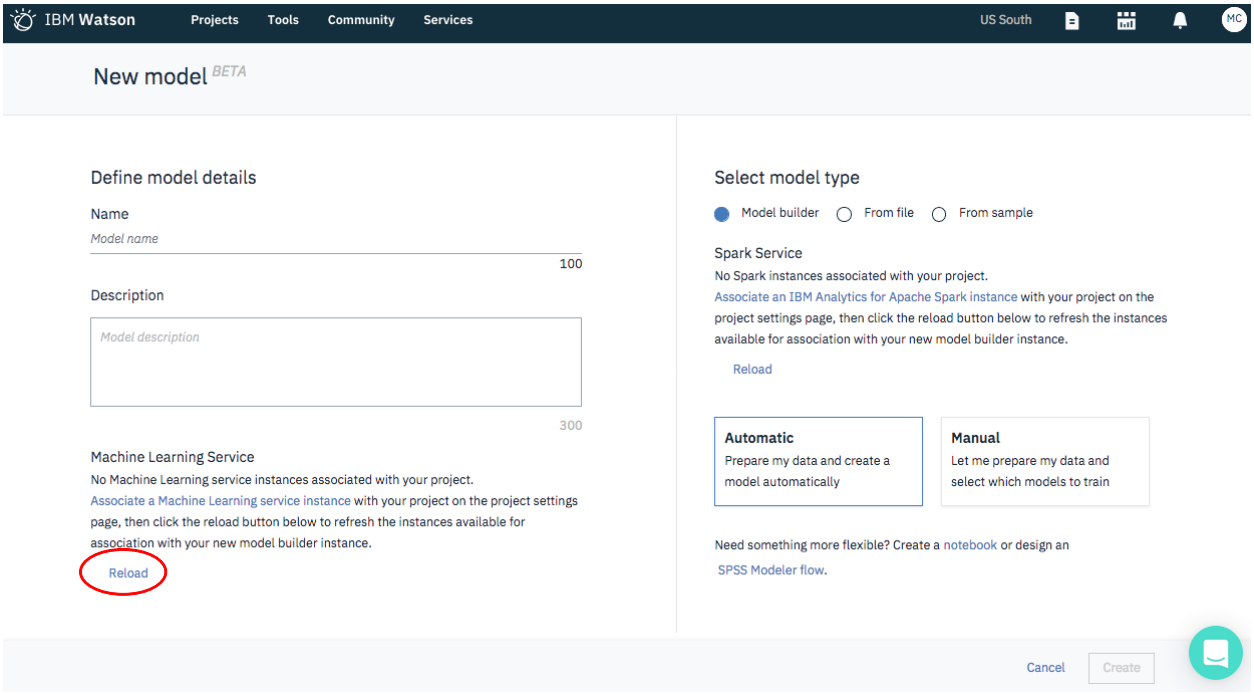
2. After being brought to the New Model page, we need to create a Machine Learning Service to build our models. Click on Associate Machine Learning service instance.



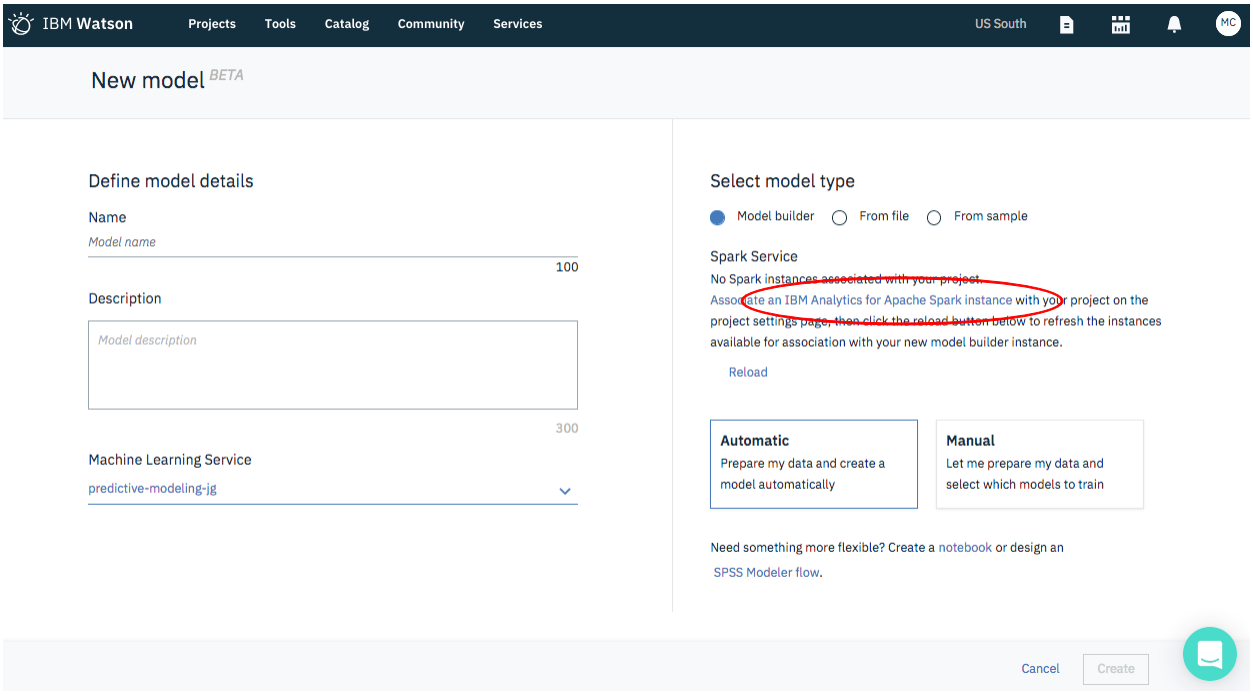
3. Scroll down, click on Lite and Create. This will allow us to train, evaluate, and deploy up to 5 models.



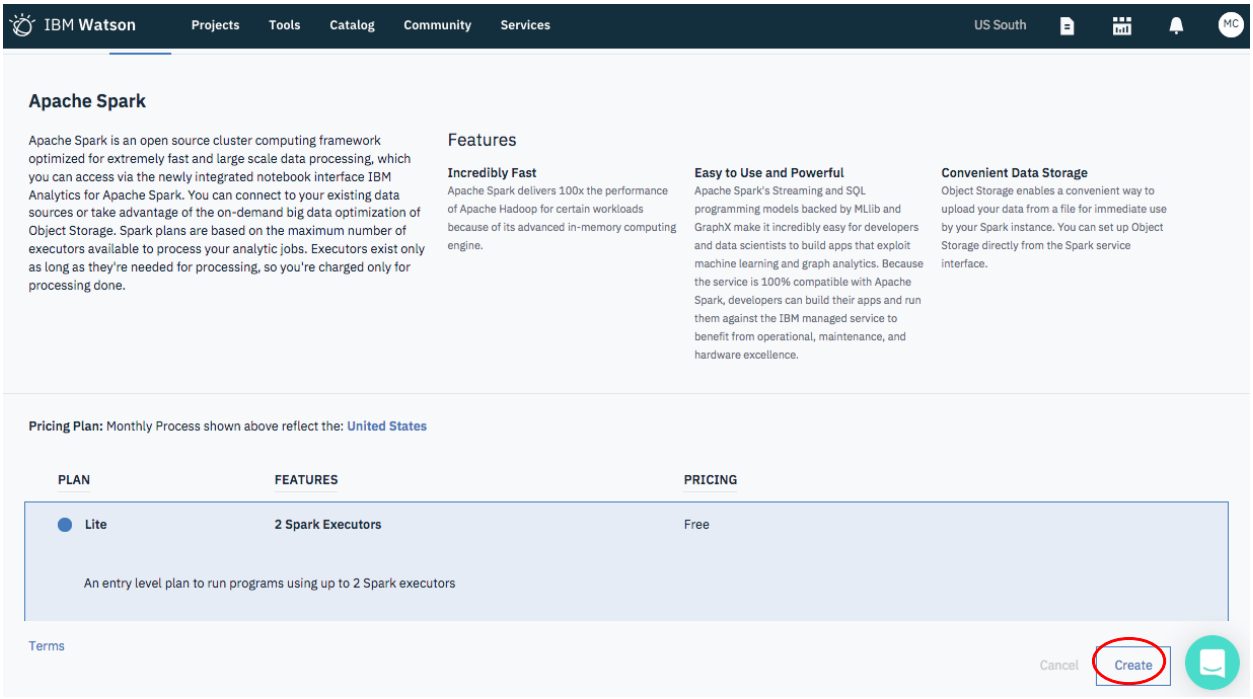
4. When the Confirm Creation window appears, keep the defaults and click Confirm.
5. Once you have returned to the New model page, click Reload under Machine Learning service.



6. Click on Associate IBM Analytics Apache Spark instance.

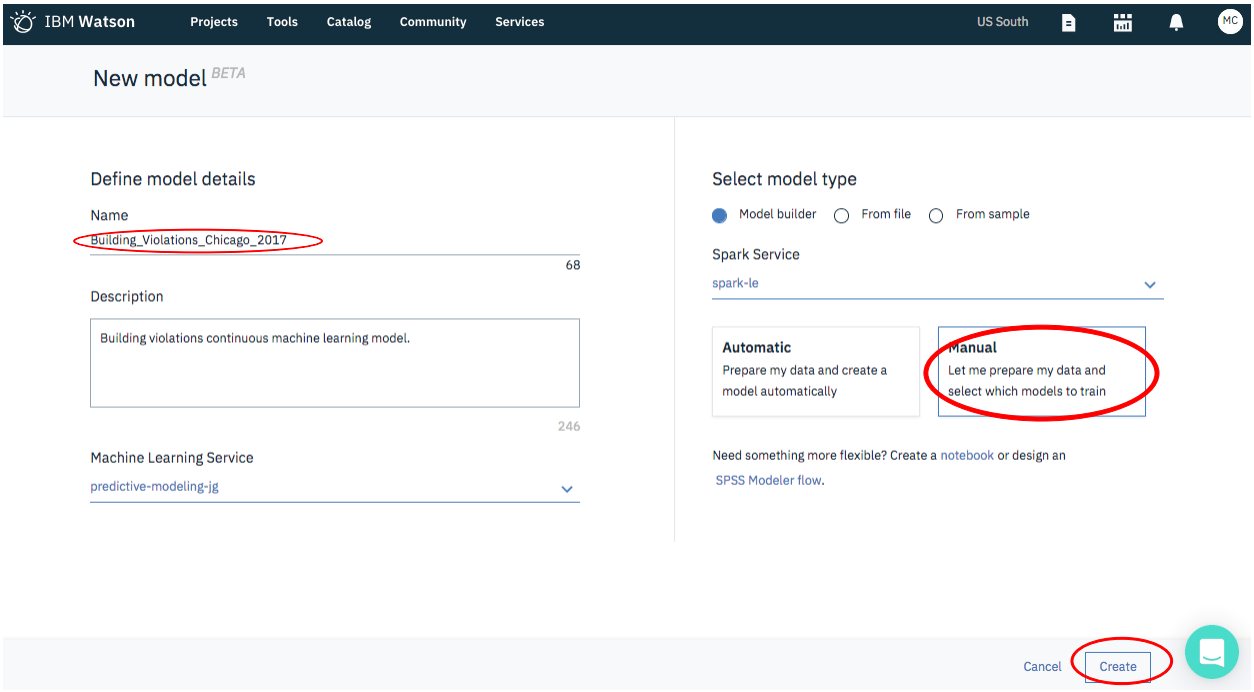


7. Scroll down, click Lite, and click Create. This will requisition two Spark executors.

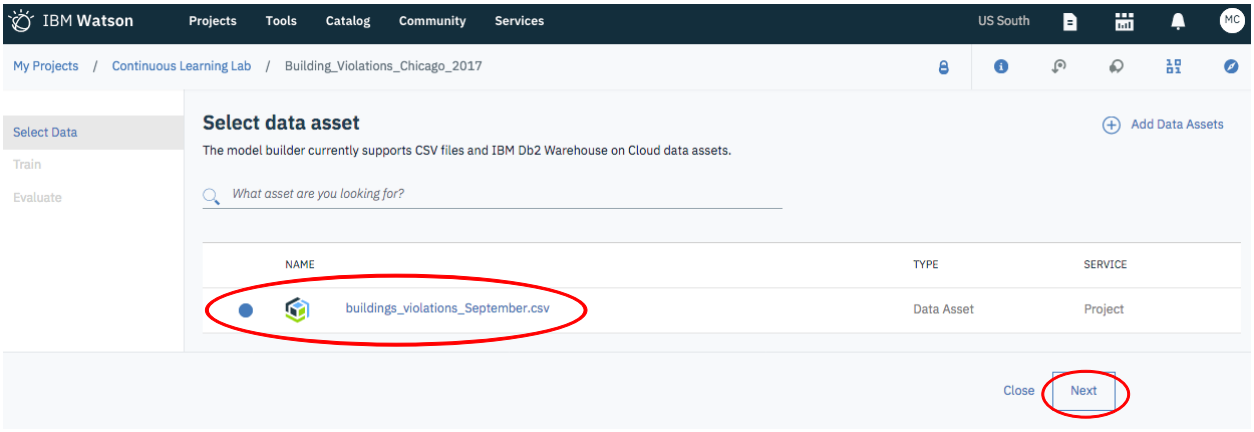


Now that we have the necessary resources, we can define our model.

8. Return to the New model page, enter a Name and a Description, select Manual, and click Create. We are creating a model using “Model builder” but can also create models in notebooks as well as with SPSS or Deep Learning flow modelers.

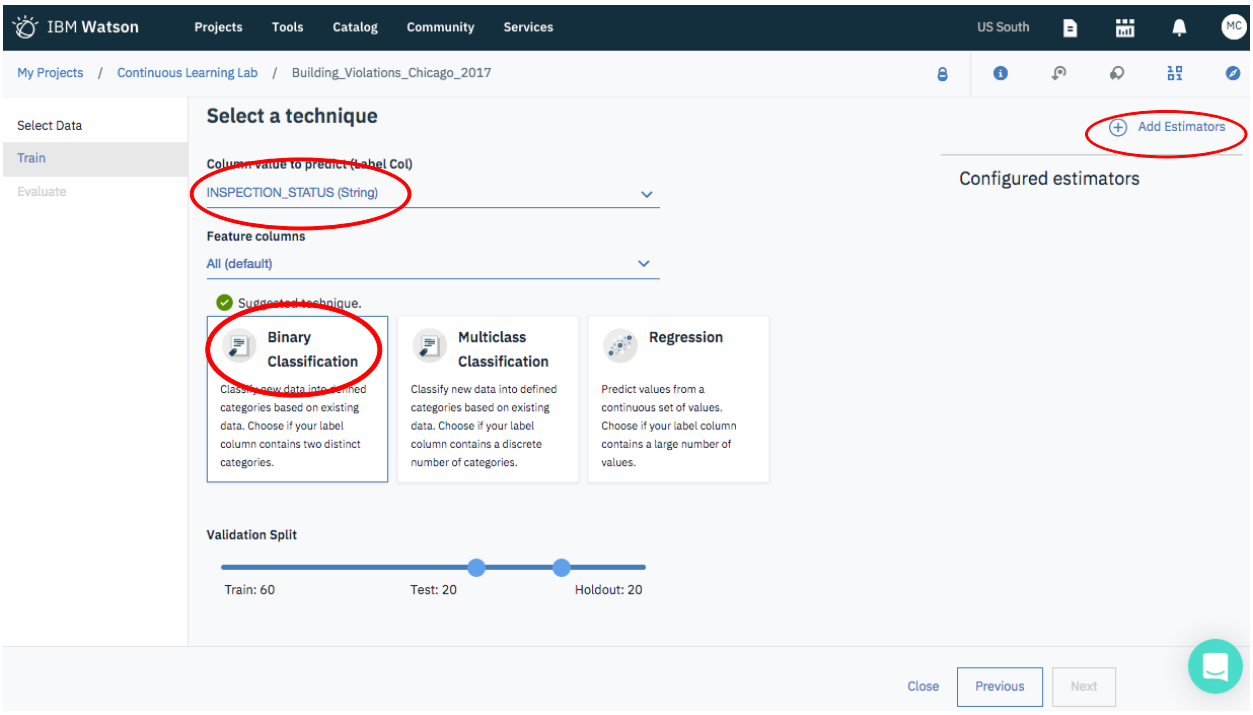


9. You will be brought to a “Select data asset” page. Select building_violations_September.csv as our data asset and click Next.



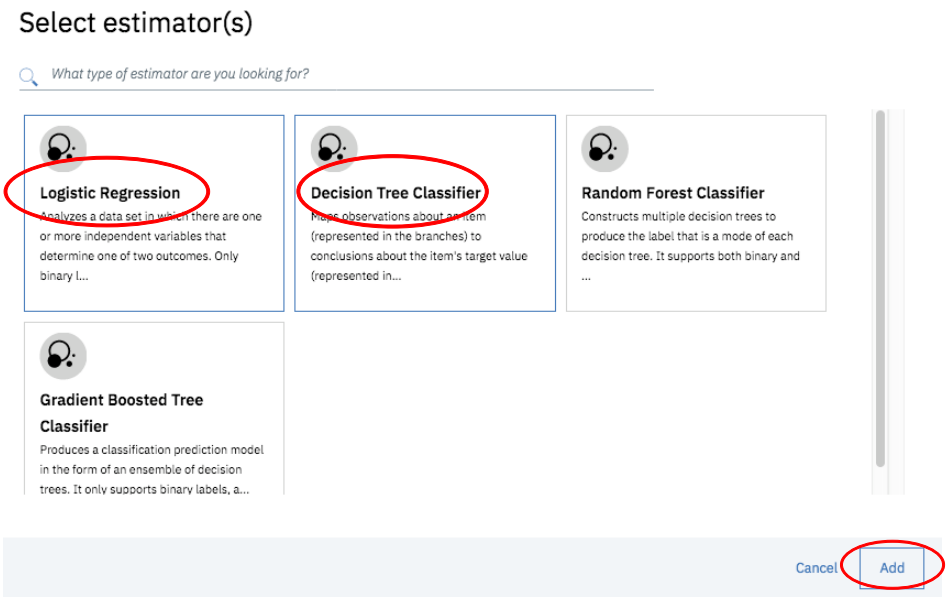
We will now choose which techniques may work well given our data and given our goal to predict which buildings are most likely to not pass inspection.

10. On the “Select a technique” page, click on the Select Label Col dropdown and select INSPECTION_STATUS(String).
11. Select Binary Classification. Note that this is often automatically suggested by Watson Studio.
12. Click Add Estimators in the upper right corner of the page.

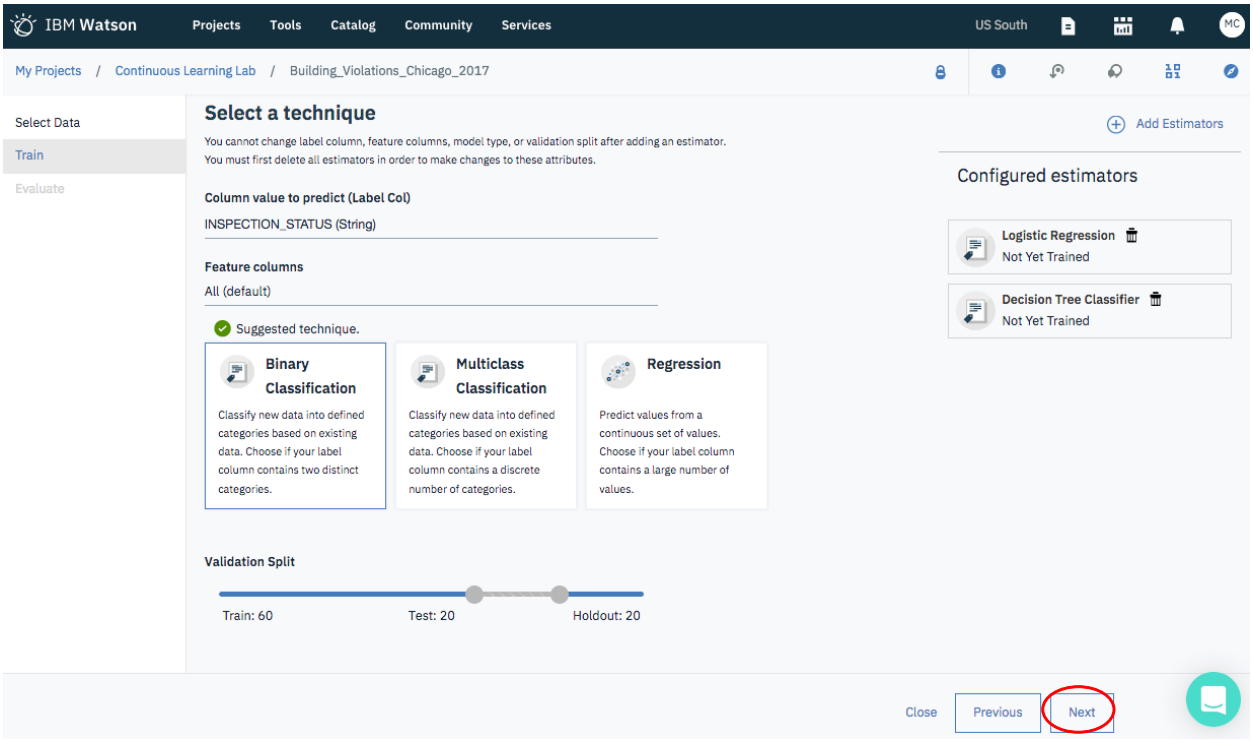


Watson Data Platform only presents the estimators that would work with our selections on the previous page.

13. Select Logistic Regression and Decision Tree Classifier.
14. Click Add.



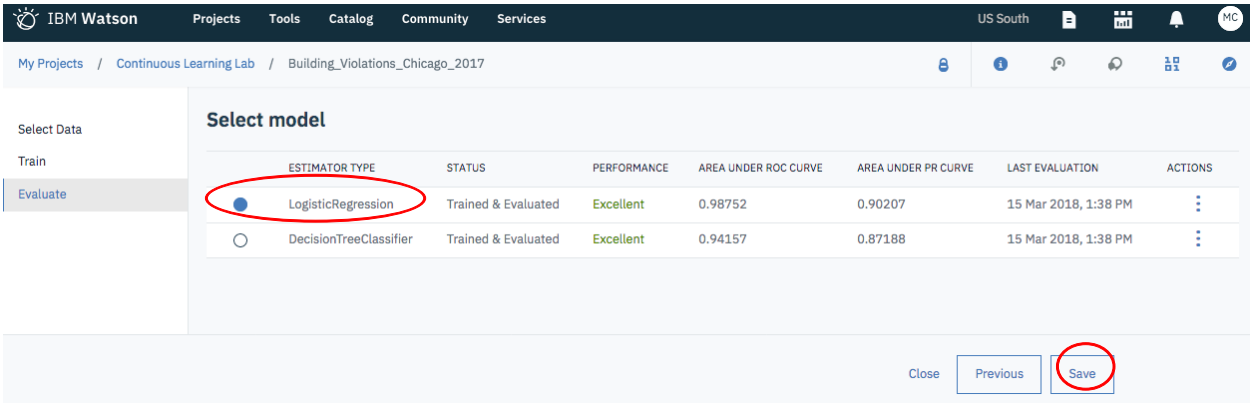
15. Back on the “Select a technique” page, click Next.



The Select model page will allow us to compare the results of different estimator types (An excellent feature for more technical users). This saves time for data scientists and business analysts by allowing them to focus on the problem at hand rather than dealing with the overhead associated with building out and tuning different models. We can simultaneously train and tune multiple algorithms and immediately compare the results.

16. Both models have performed well. For this tutorial, select Logistic Regression.

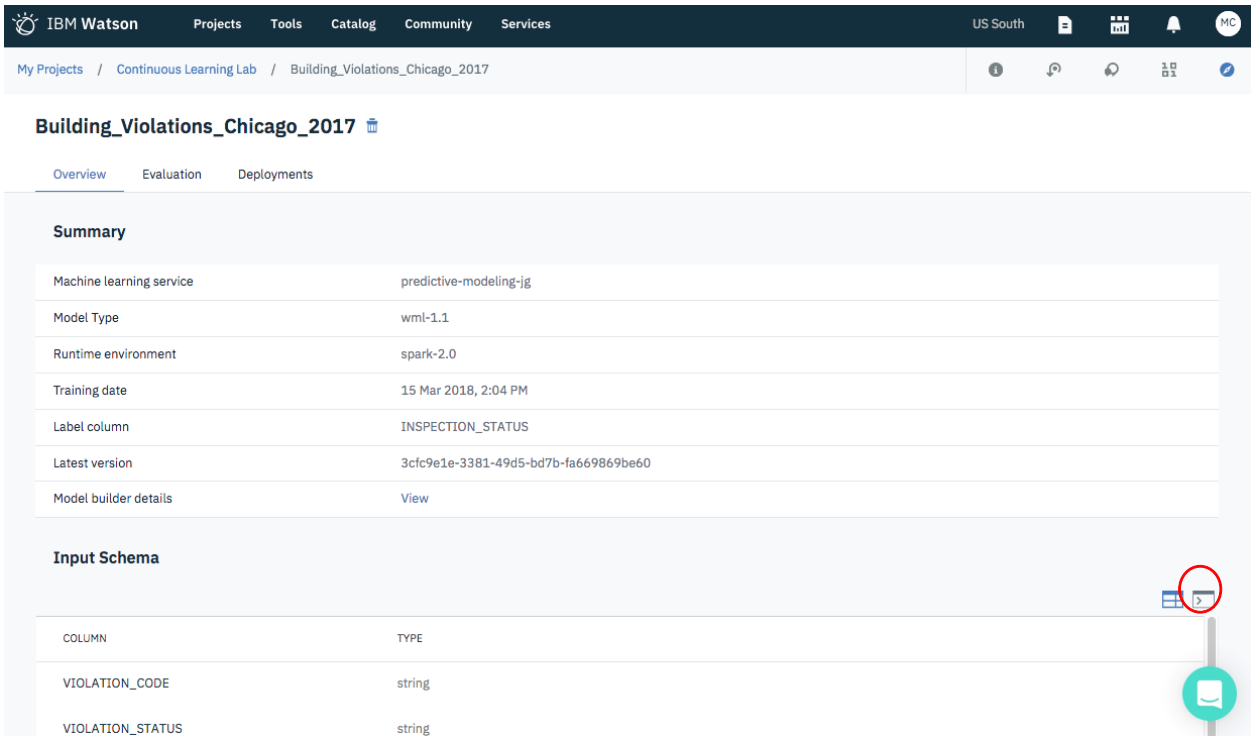
17. Click Save.



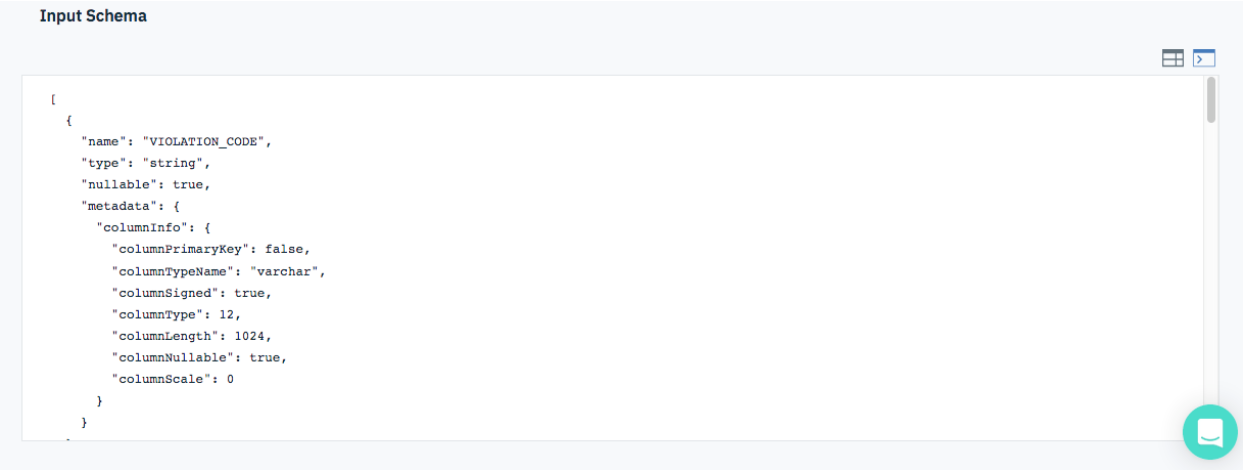
18. When the “Save model” window appears, click Save.

We have just saved the model in our Watson Machine Learning service and can now view information about the model specifications, details of construction, and input schema. The model can also be exposed as an API and called used by data pipelines, applications, or other external systems.

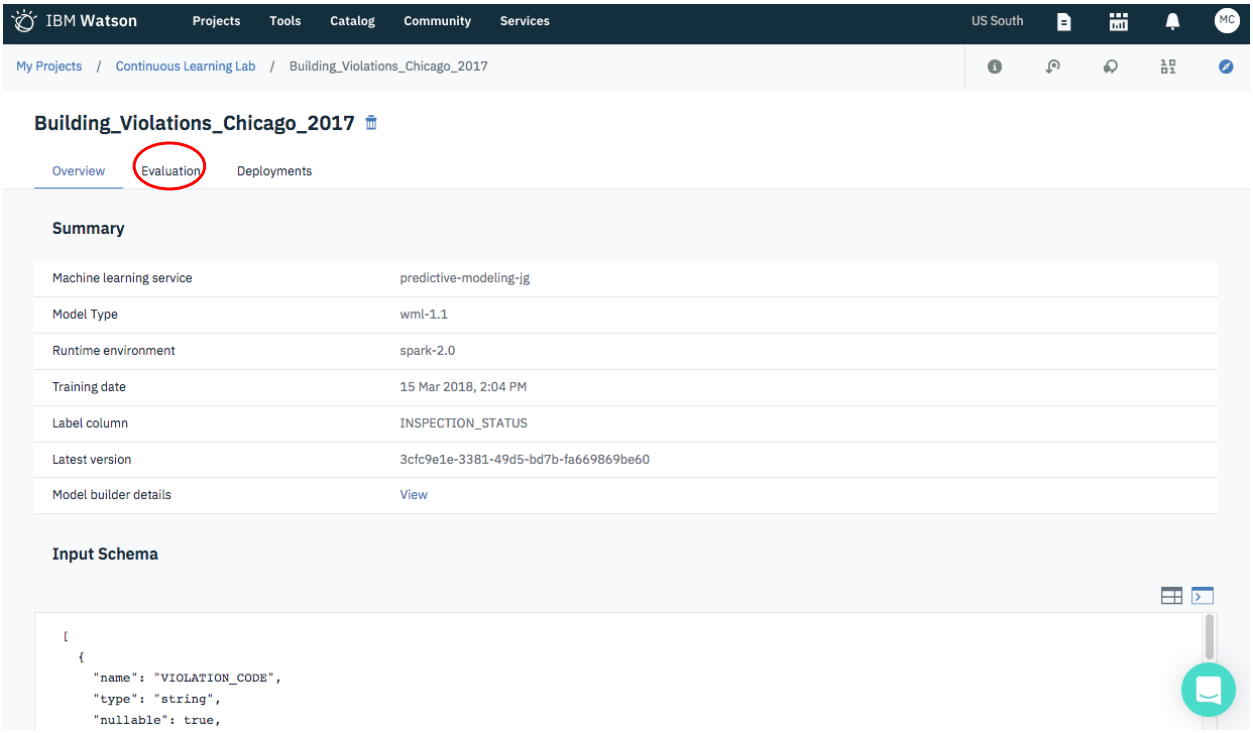
19. Click on the console button with an “angle bracket” to view the schema in JSON format.



This schema can be copied and used elsewhere to help existing systems easily interact with our model and the Watson Machine Learning service.



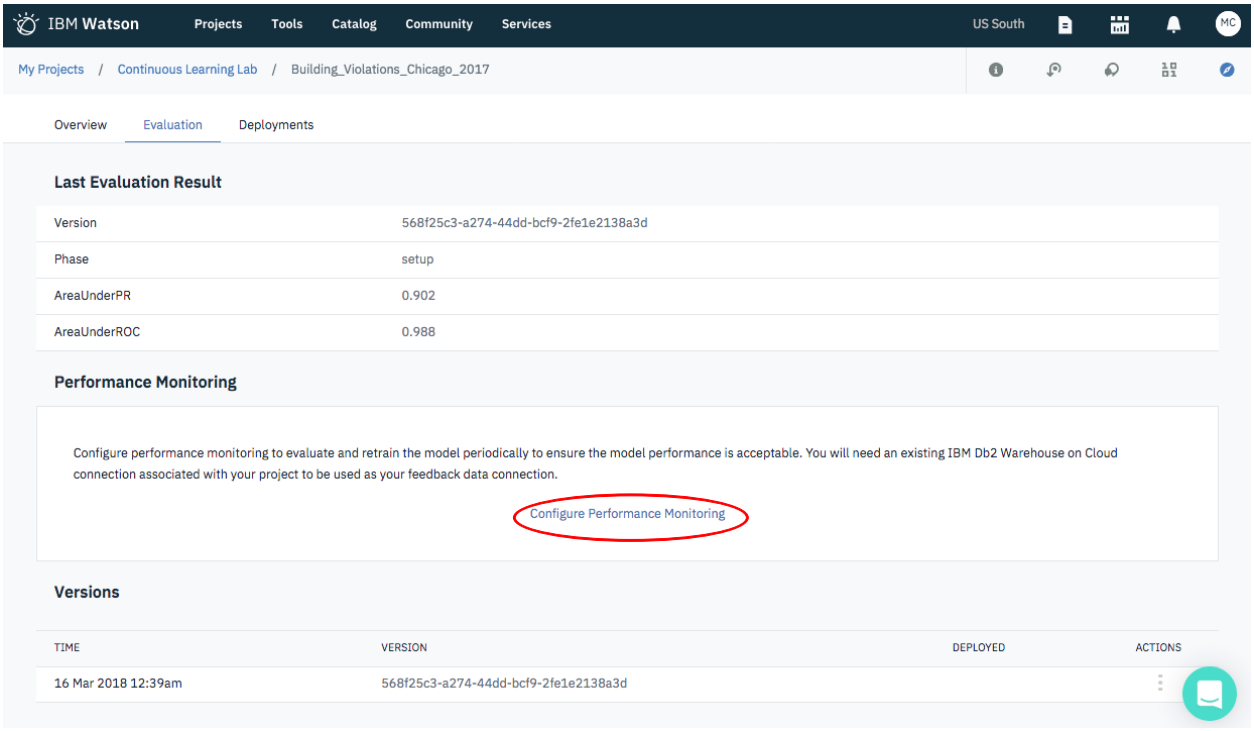
20. Scroll up and click on the Evaluation tab.



Step 1.4 Set up Continuous Learning Model Capabilities

Model performance is critical for solving data science problems. This page provides information and functionality to aid in continuous training, tuning, and redeployment. We will now set up parameters to automatically retrain our model when performance falls below a certain threshold.

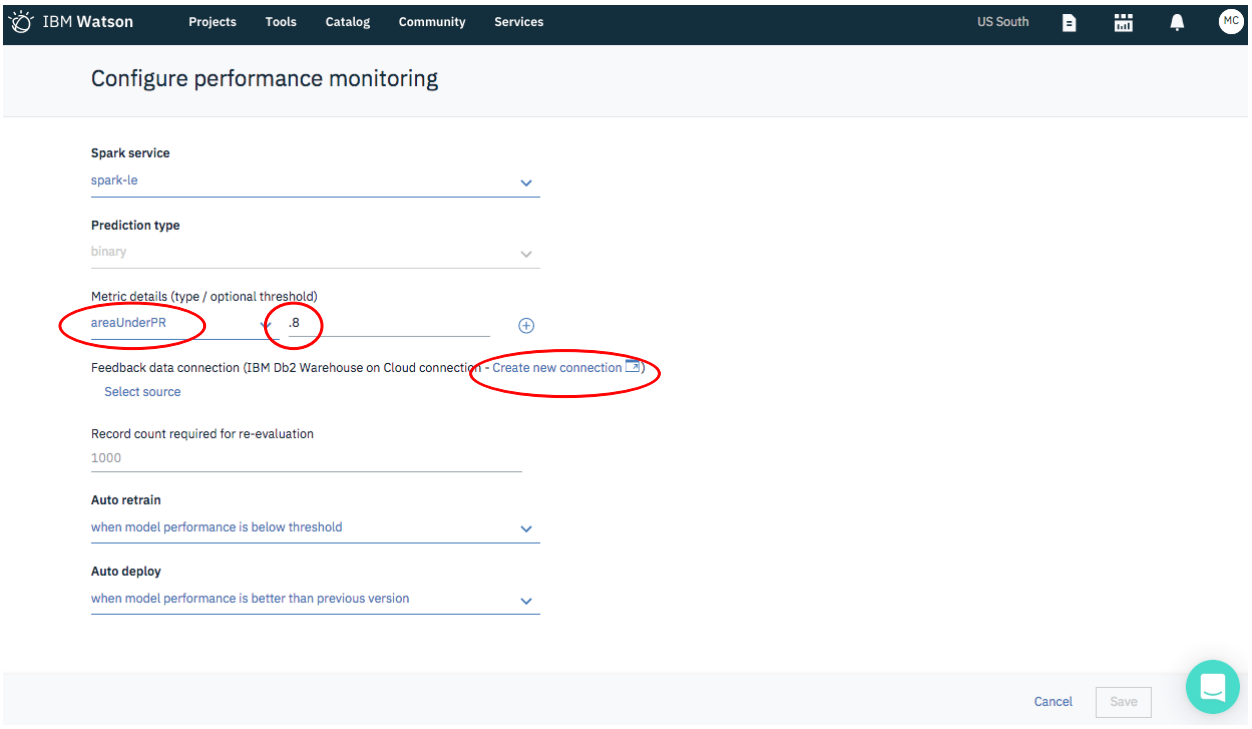
1. Click on Configure Performance Monitoring.



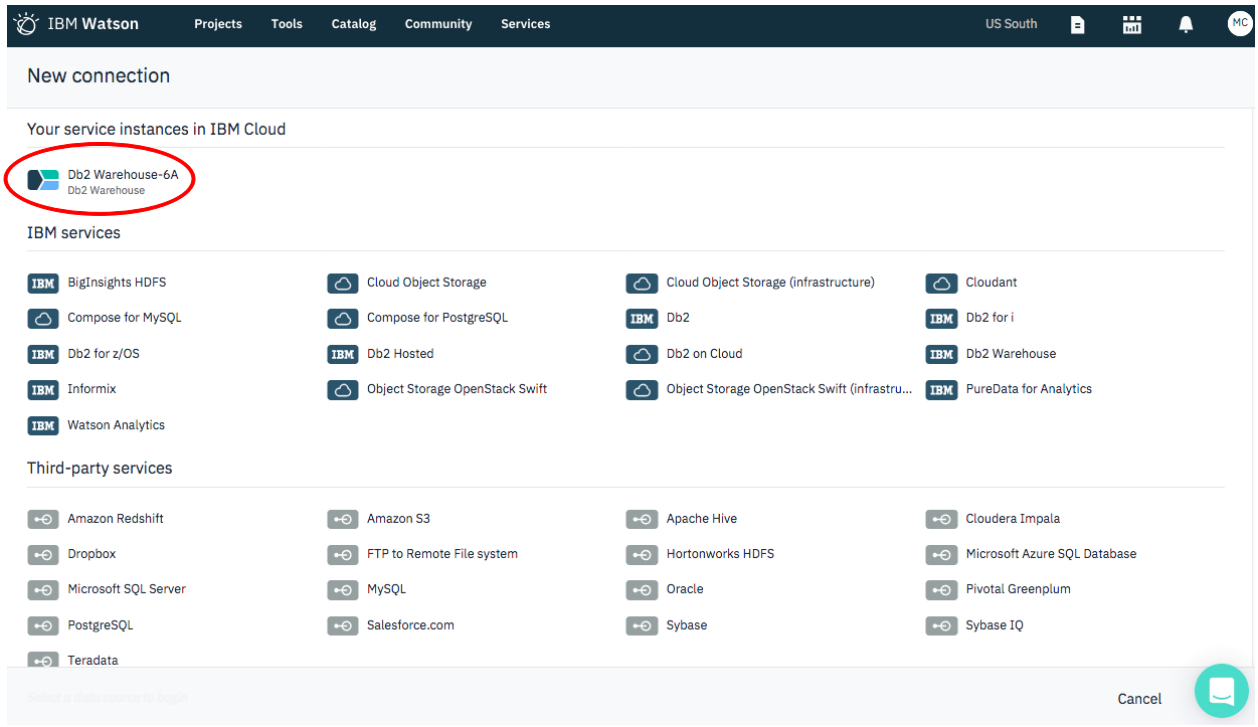
2. Under Metric details, select areaUnderPR and enter 0.8.

This relies on a database table of new data. We will use DB2 Warehouse on Cloud, a data store optimized for analytic data sets.

3. Click on Create a new connection.

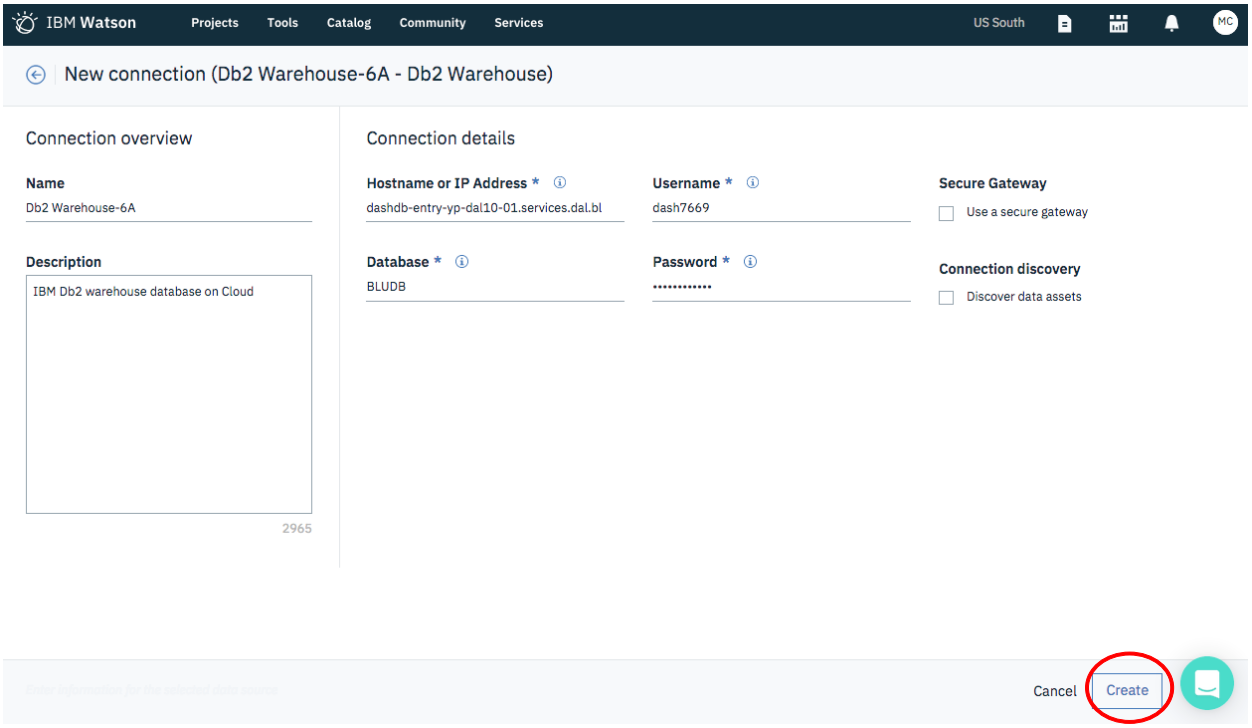


4. On the “New connection” page, click the DB2 Warehouse instance created during the lab prerequisites.

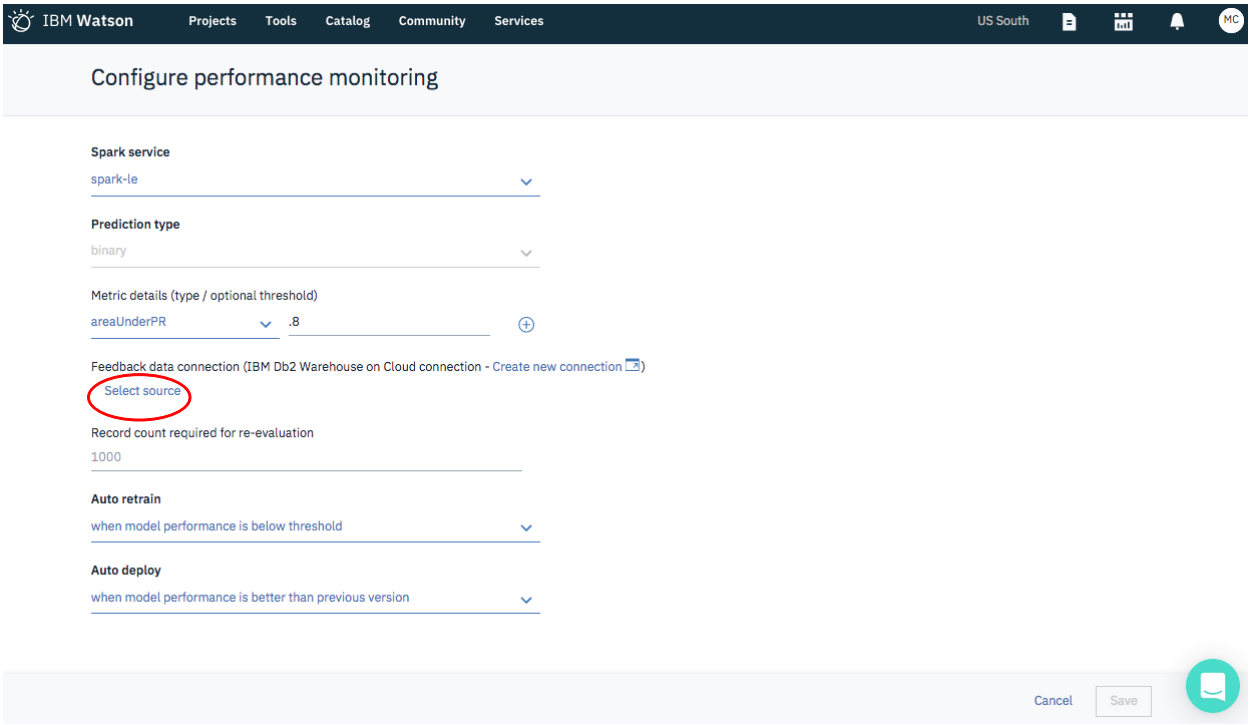


Notice the connection details have already been filled in.

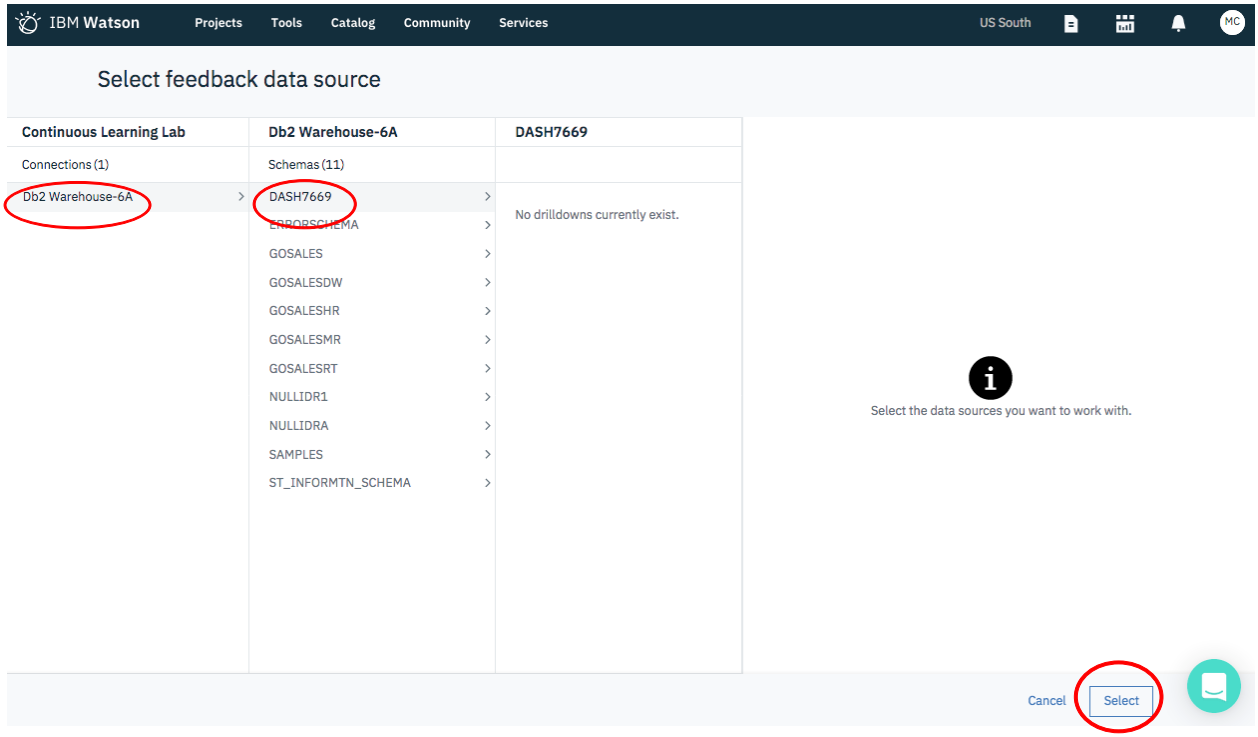
5. Click Create.



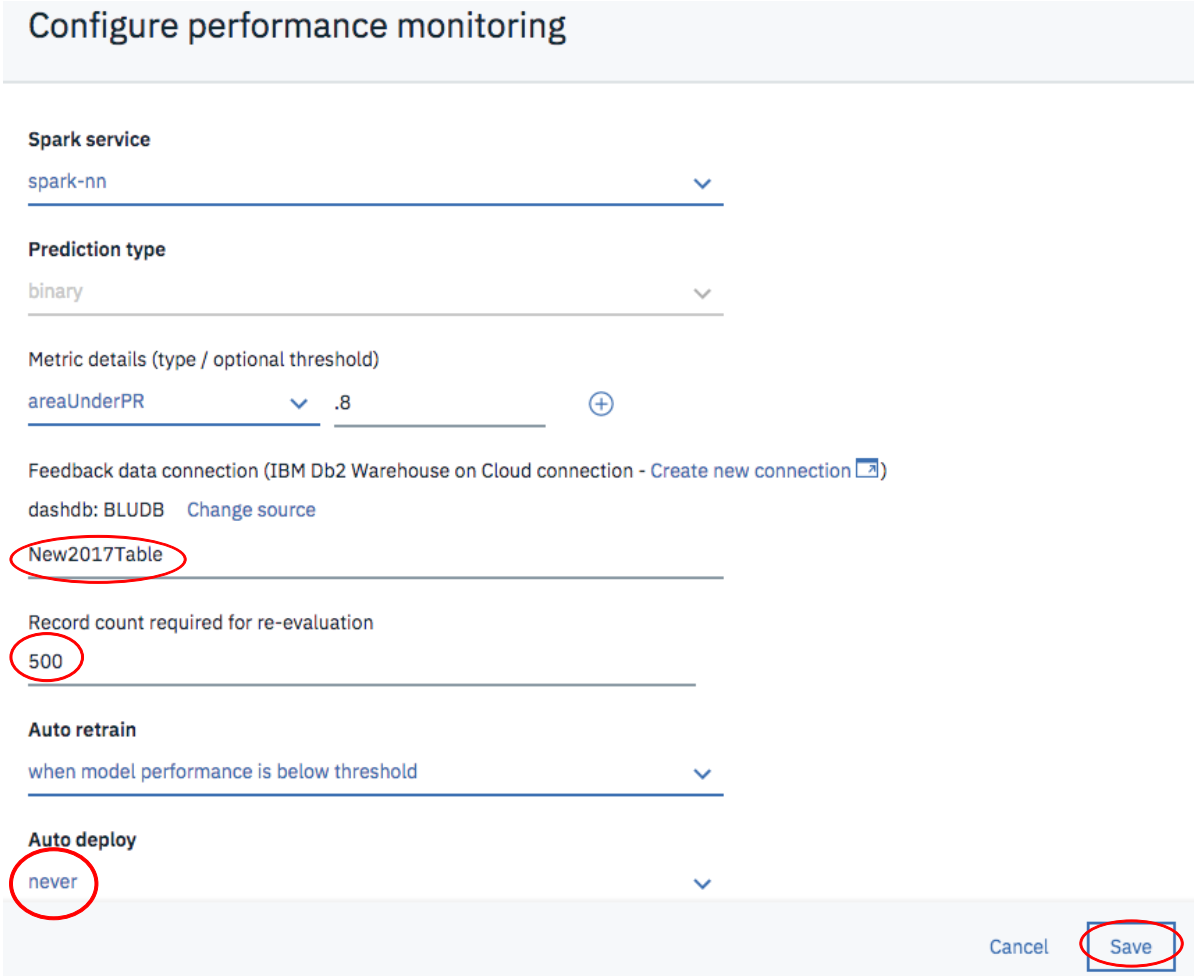
6. Return to the “Configure performance monitoring page” and click Select source.



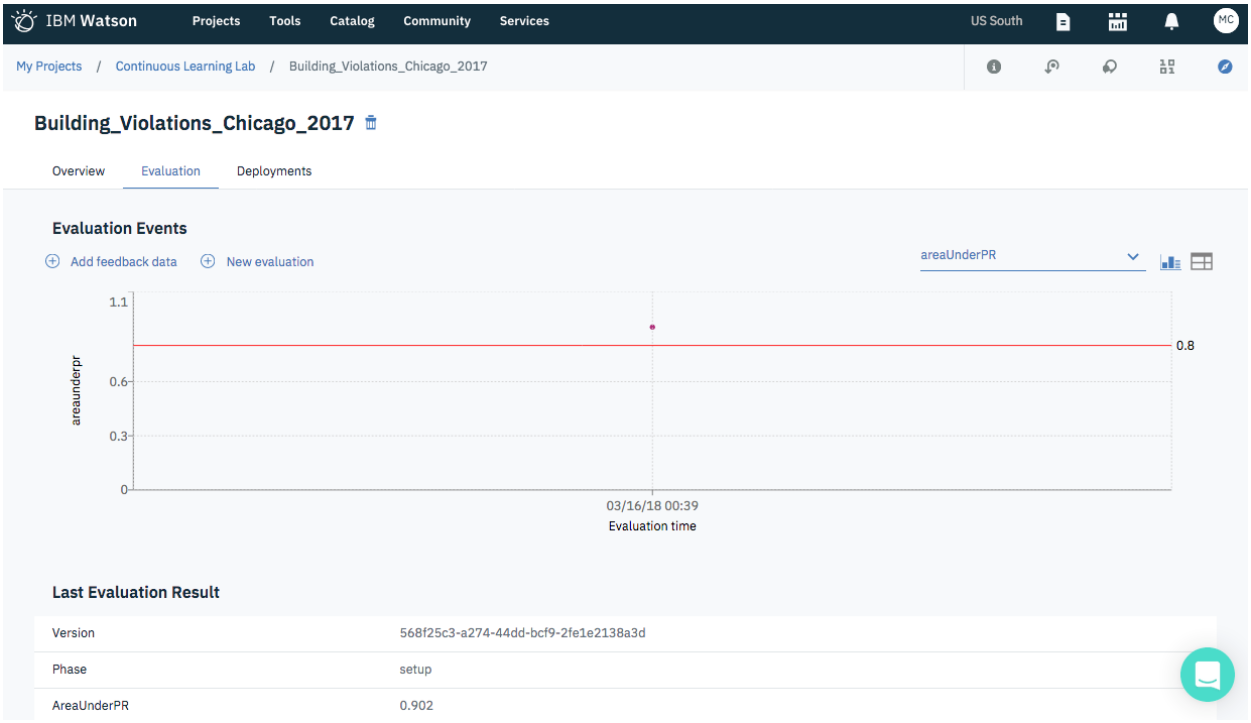
7. Click on your DB2 Warehouse-xx name.
8. Select the schema that matches DASH####.
9. Click select.



10. After returning to the “Configure performance monitoring” page, enter “New2017Table” as the table name.
11. Enter 500 as the Record count.
12. Under Auto deploy, select never.
13. Click Save.

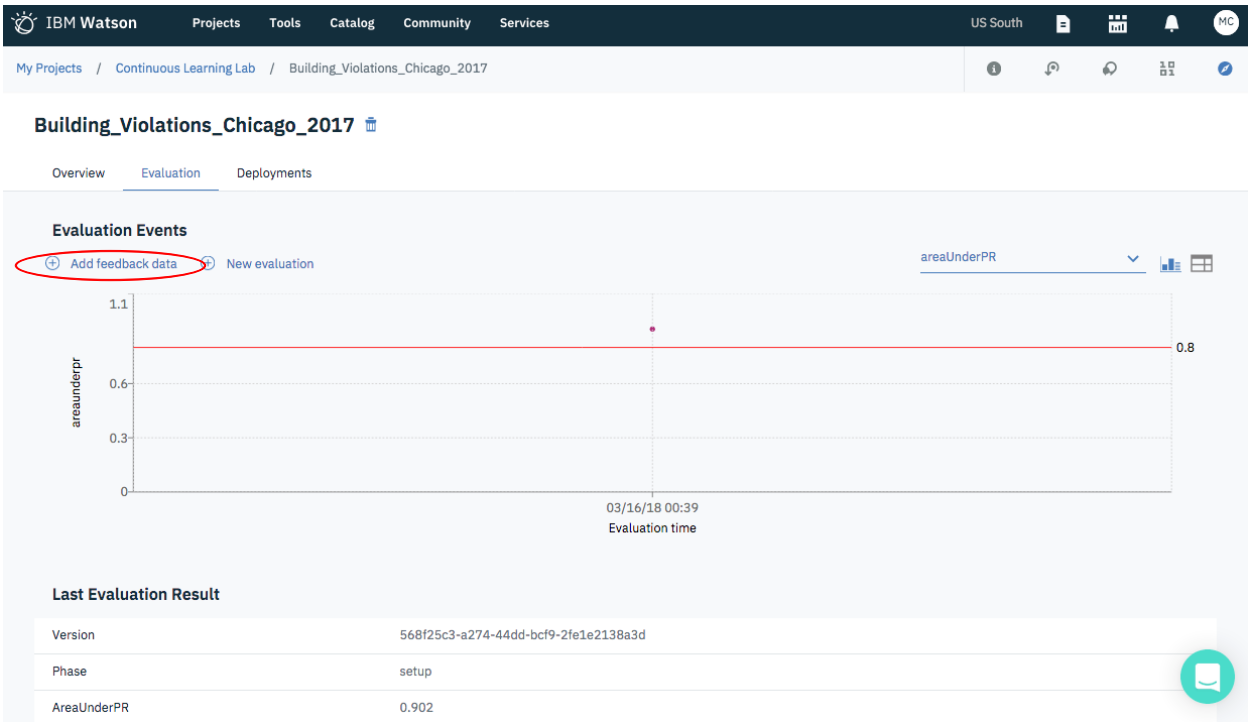


We have just set our model to retrain whenever its performance falls below 0.8. The evaluation screen has changed to reflect the decisions we have just made. Notice the graph lists our current areaunderPR and the threshold we have set. Clicking on Add feedback data will upload a new dataset. Clicking on new evaluation will then retrain the model and check to see how the model compares to our 0.8 threshold.



Suppose September has passed and we now have October data. We can trigger a new evaluation by adding more feedback data.

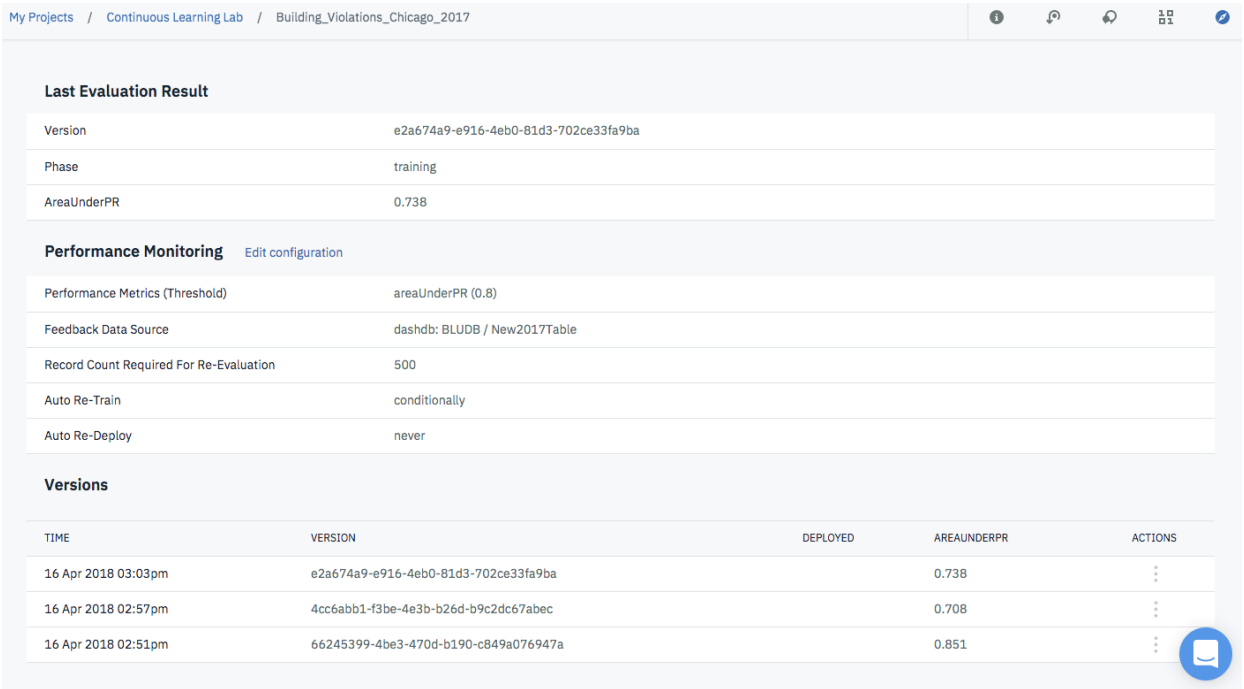
14. Click Add feedback data.
15. Select building_violations_October.csv.



16. When the “New evaluation” window appears, click New Evaluation. Re-evaluating the model may take a few minutes. This is a great time to grab coffee.

Notice that the October data has caused our model to fall below the 0.8 threshold and a new version has been trained. Although we run through these steps graphically, you can also configure a Watson Machine Learning continuous learning model entirely through APIs.

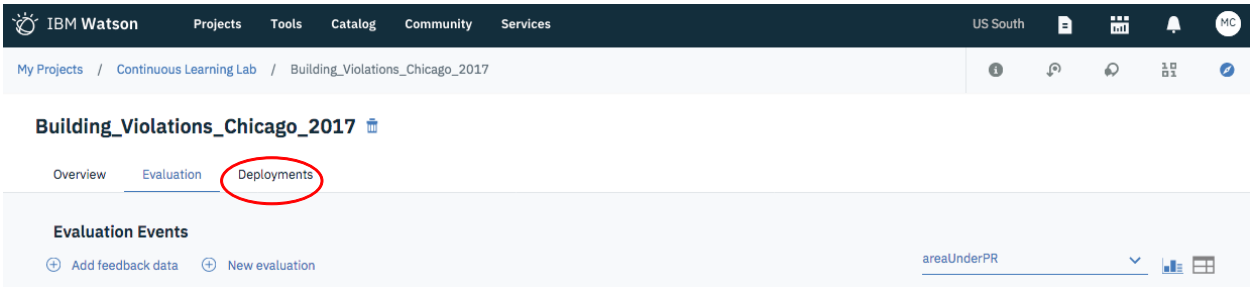
You can continue by adding the November data by the same process of hitting **Add feedback data** and adding more files. If you scroll down you will be able to see all model versions listed. We have three versions, but no deployments.



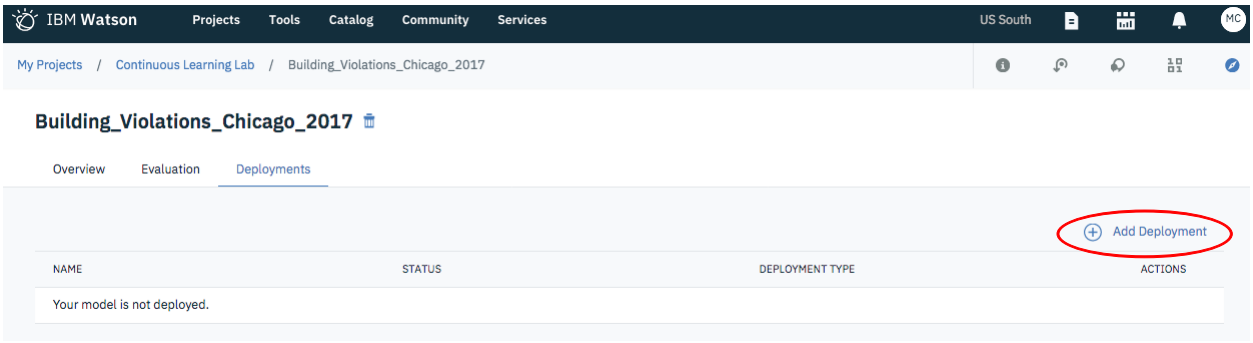
Lesson 1.5 Deploy a Machine Learning Model.

Now we will focus on deployment.

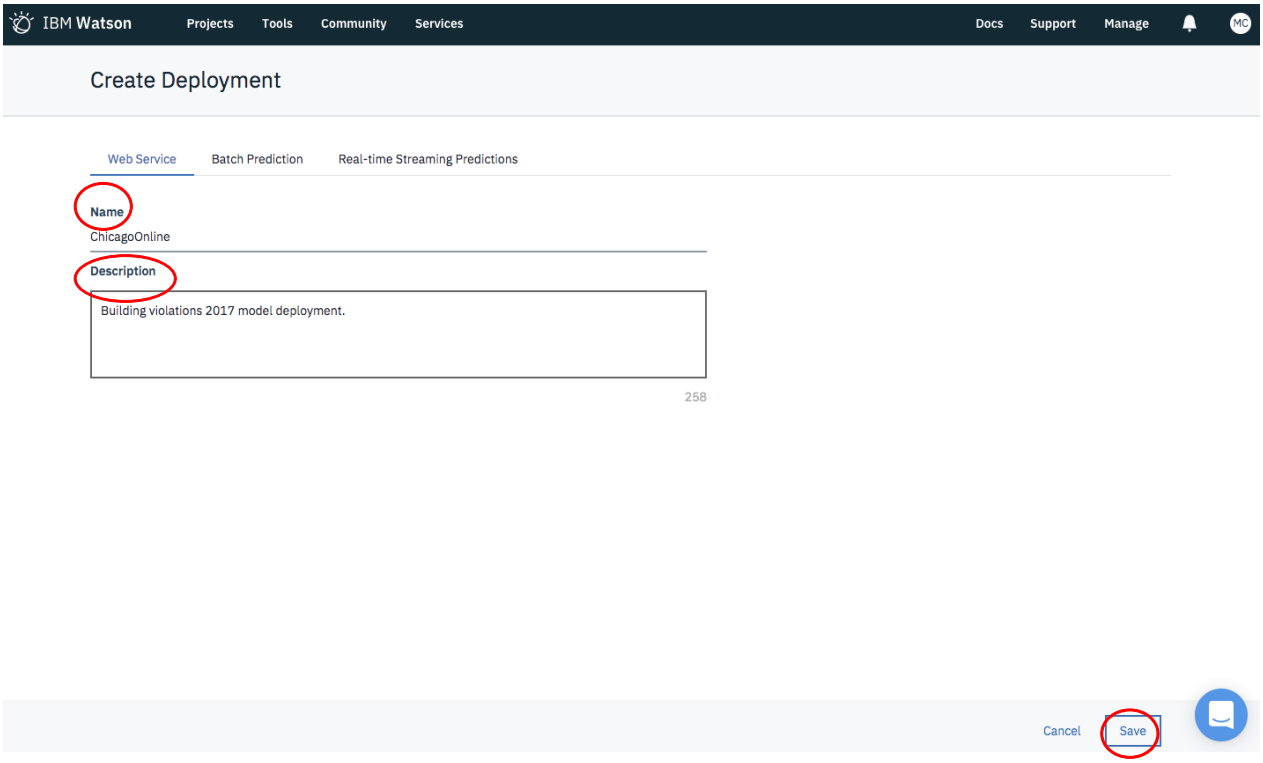
1. Click the Deployment tab.



2. Add Deployment.

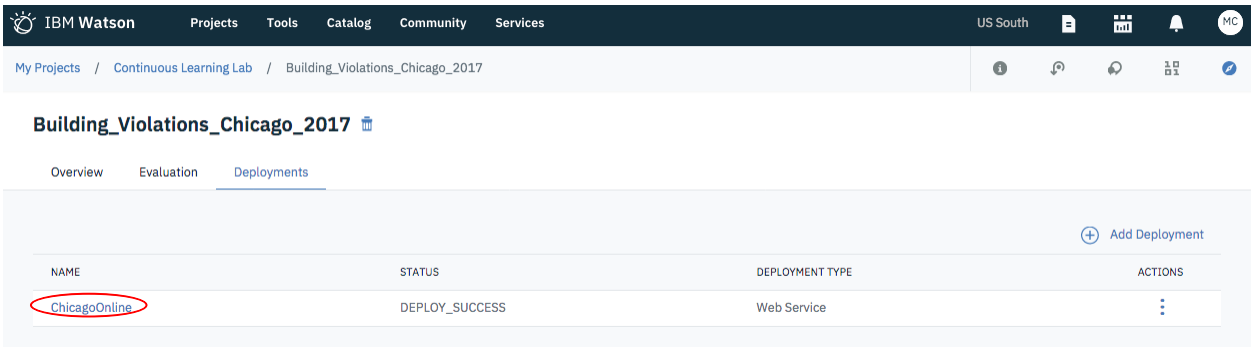


3. Enter a Name (e.g. ChicagoOnline).
4. Enter a Description.
5. Click Save.



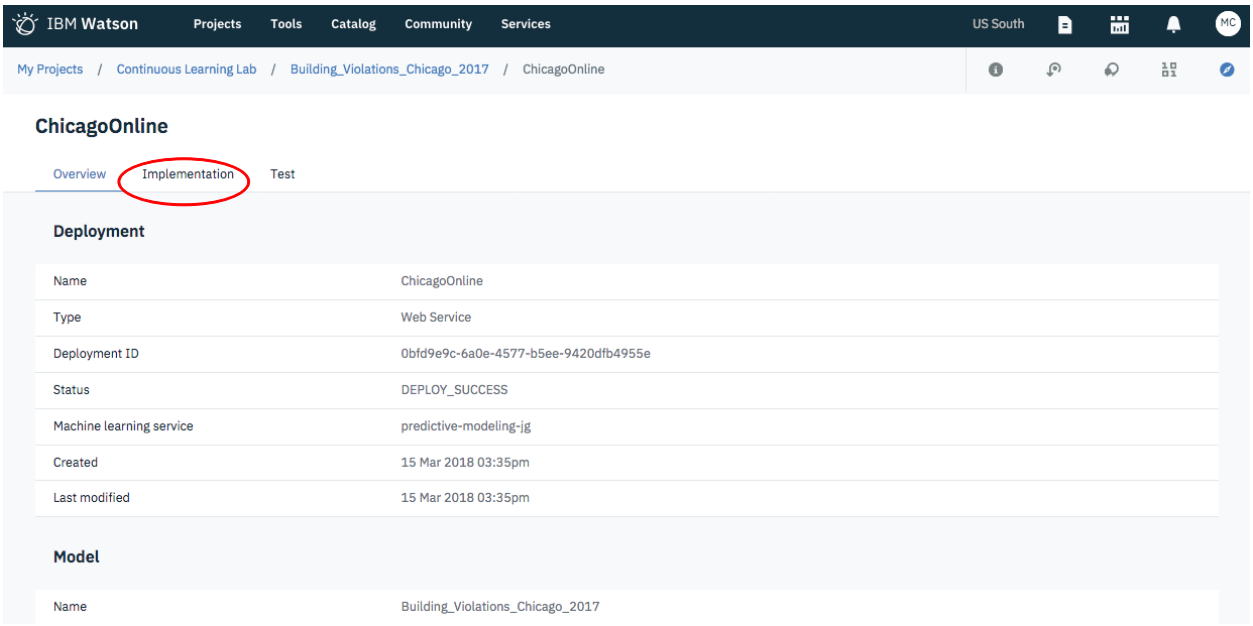
The model is now accessible by external systems.

6. Click on your deployment (e.g. ChicagoOnline).



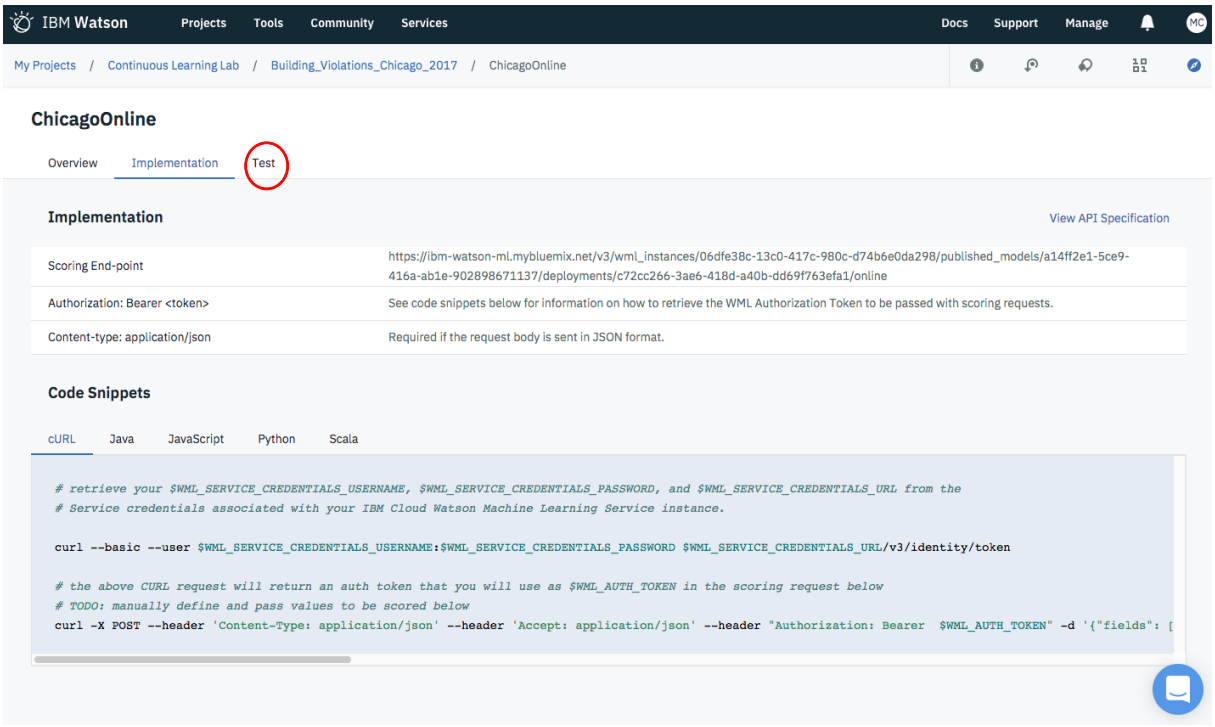
You are automatically brought to the Overview tab lists information related to the model deployment including services used and version number.

7. Click the Implementation tab.



The implementation tab provides developers information to help minimize the time it takes to develop models and place them in a production environment.

8. Click on the Test tab.



The Test tab allows manual testing of the deployed model and viewing of results.

9. Enter:

   VIOLATION_CODE: CN063014

   VIOLATION _STATUS: OPEN

   INSPECTION_CATEGORY: COMPLAINT

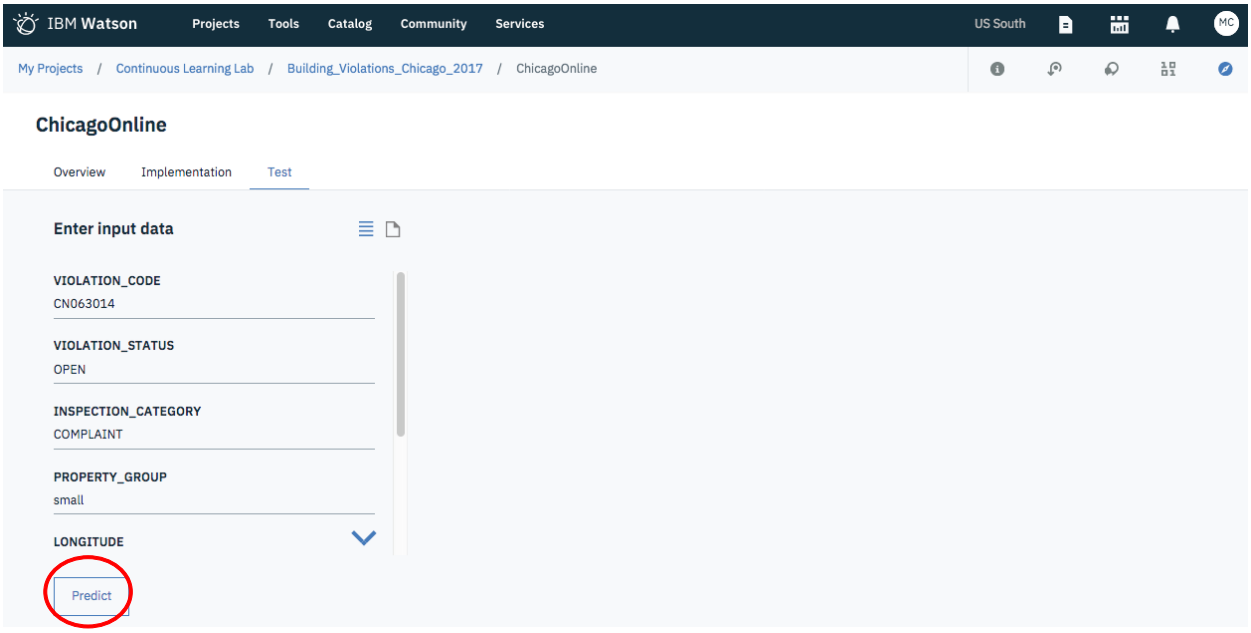
   PROPERTY_GROUP: small

   LONGITUDE: -87.691078

   LATITUDE: 42.002457

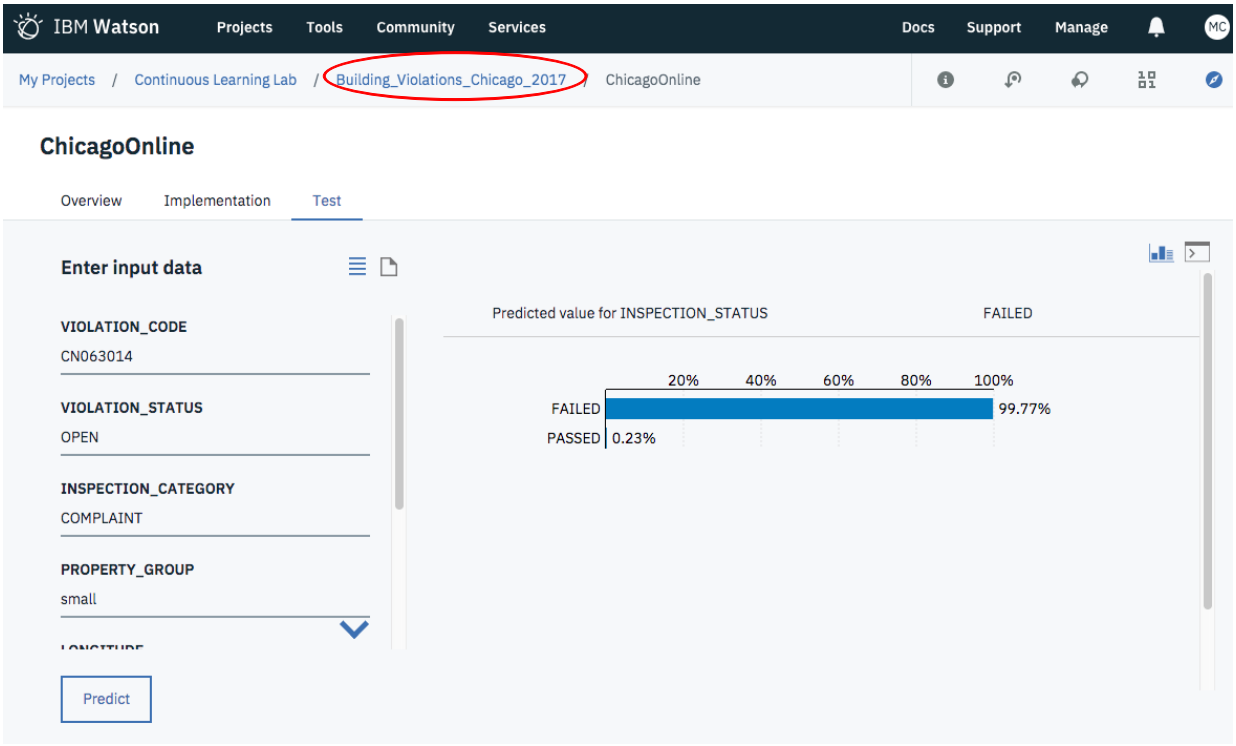
   DEPARTMENT_BUREAU: CONSERVATION

10. Click Predict.

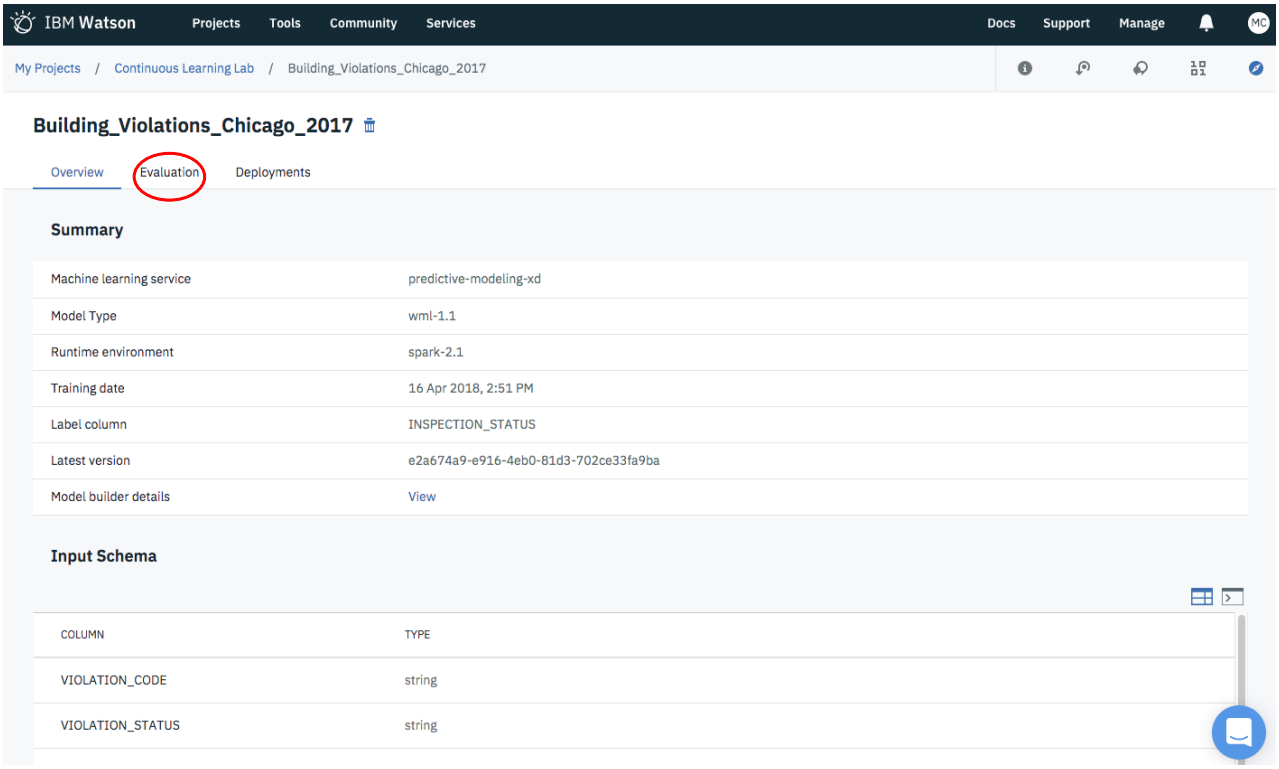


According to our model, a building with our input data has a 0.23% chance of passing inspection.

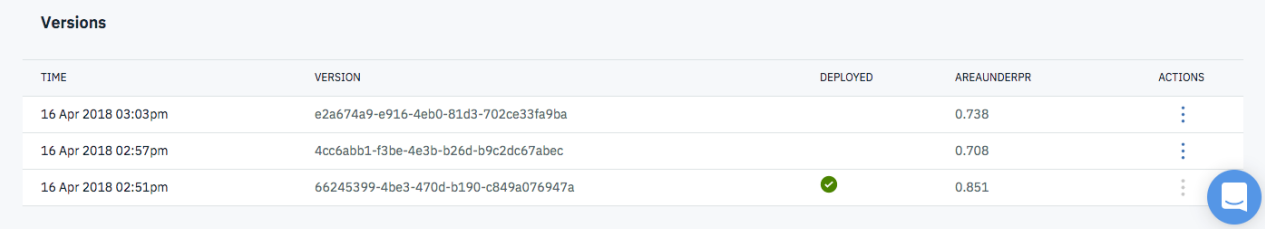
11. Feel free to change the input data and run new predictions. When ready, click on you model name (e.g. Building_Violations_Chicago_2017)



12. Click on the Evaluation tab.



If you scroll to the bottom you will be able to see the different model versions that have been created and which model is currently deployed.



During this lab we have very quickly compared various machine learning models and chosen the best one tuned to our dataset and objectives. We then created a continuous machine learning model that automatically monitors and retrains the most up-to-date machine learning model allowing Watson Machine Learning to keep applications, data pipelines, or external systems relying on the machine learning model as up to date as possible.

You have completed Lab 2