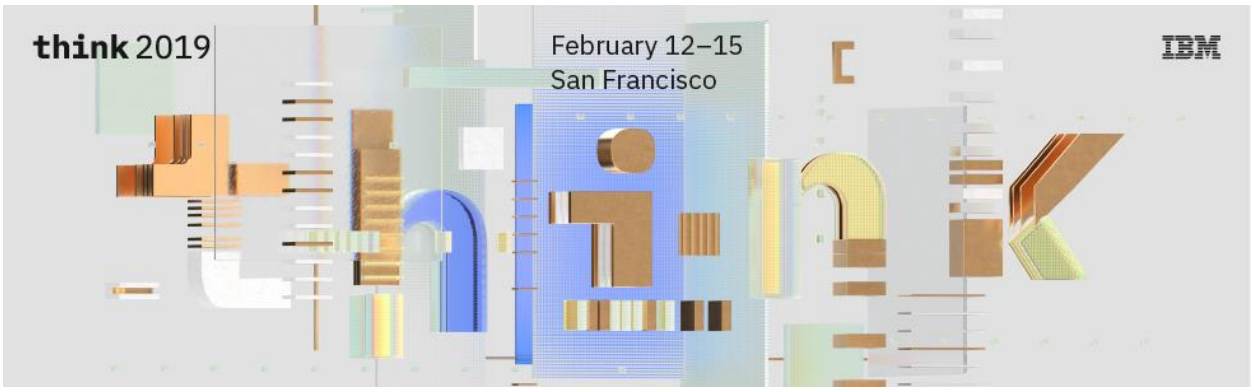


Lab Center – Hands-on Lab

Session #3259

Session Title: Introduction to Data Science using Watson Studio

Lab-2: Deploy a Continuous Machine Learning Model using Watson Studio

Bernard Beekman, IBM, beekmanb@us.ibm.com

Michael Cronk, IBM, michael.cronk@ibm.com

Table of Contents

Disclaimer 3
Introduction 5
Goals: 5
Step 1 Upload a New Data Asset 5
Step 2 Train, Compare, and Select a Machine Learning Model 8
Step 3 Set up Continuous Learning Model Capabilities 16
Step 4 Deploy a Machine Learning Model. 26
We Value Your Feedback! 33

Disclaimer

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results like those stated here.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts.
In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may

need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, ibm.com and [names of other referenced IBM products and services used in the presentation] are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.

U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.

Introduction

In this lab, you will be using 2017 Chicago building data to make Chicago a safer place by building a model to predict when buildings are likely to fail inspection. You can then use the model to find which buildings are most dangerous and attend to those first. Continuous learning will be set up to monitor the model performance and adjust if necessary.

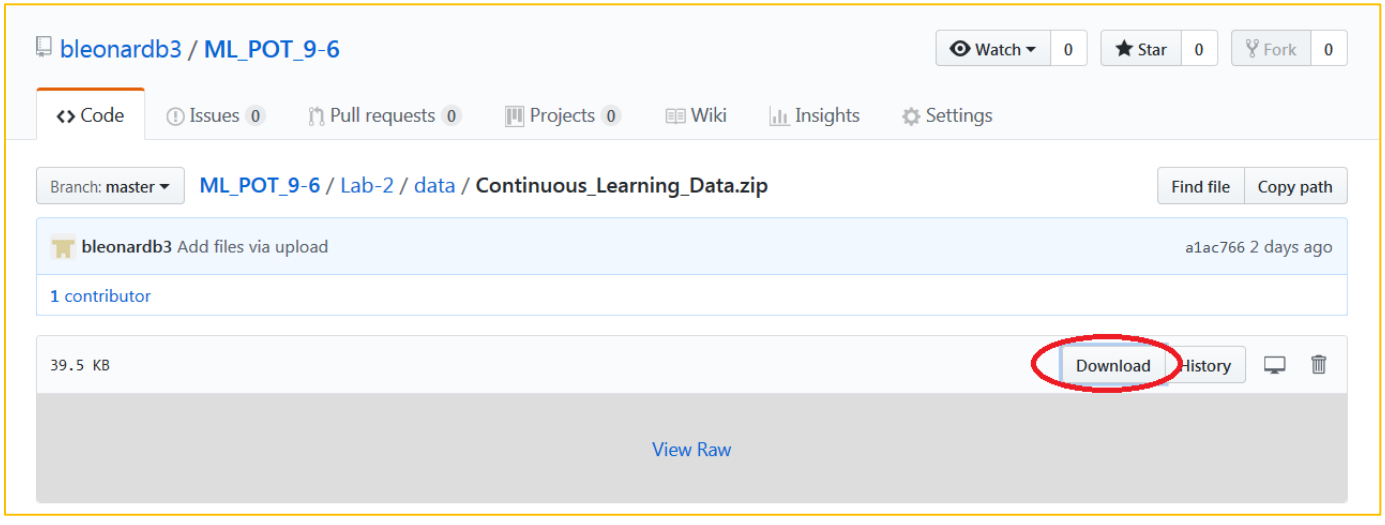
Goals:

- 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.

Step 1 Upload a New Data Asset.

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

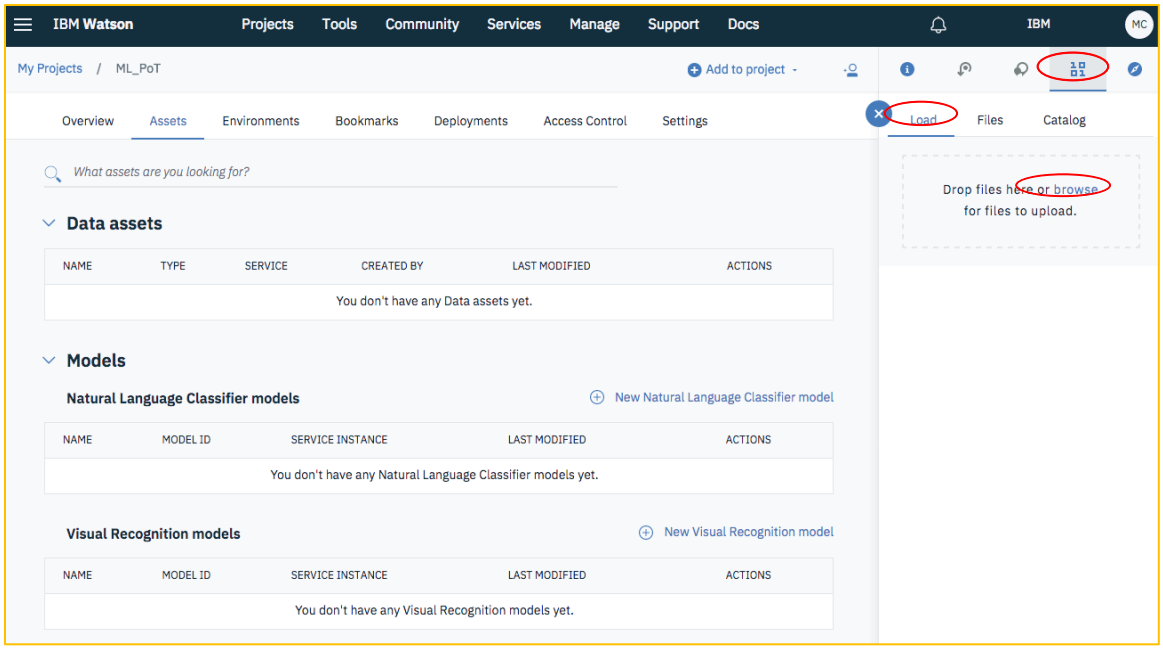
1. Click on Building Data
2. Click on Download.



3. Save the zip file to your desktop and unzip the contents. This zip file contains data for building inspections for each month. We will begin by training our model for the month of September and then add October and November data to show how we can continuously adjust over time.
4. Return to your assets page. From Watson Studio, you can click on the Projects

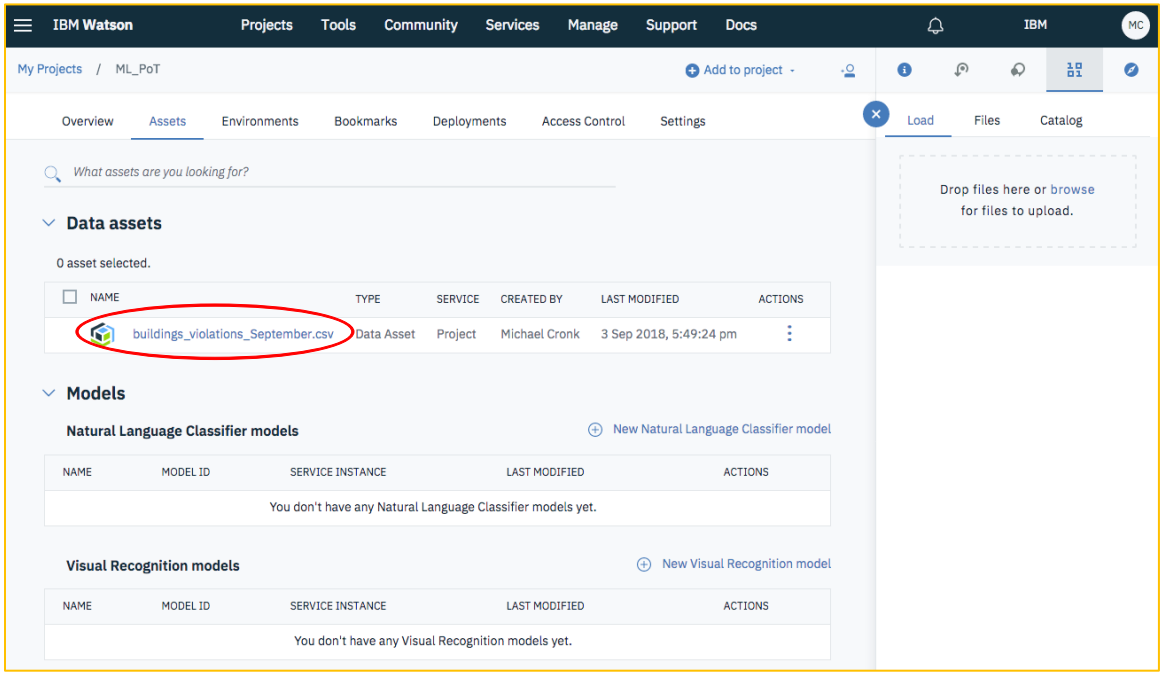
dropdown on the top navbar and then click on the name of your project.

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.

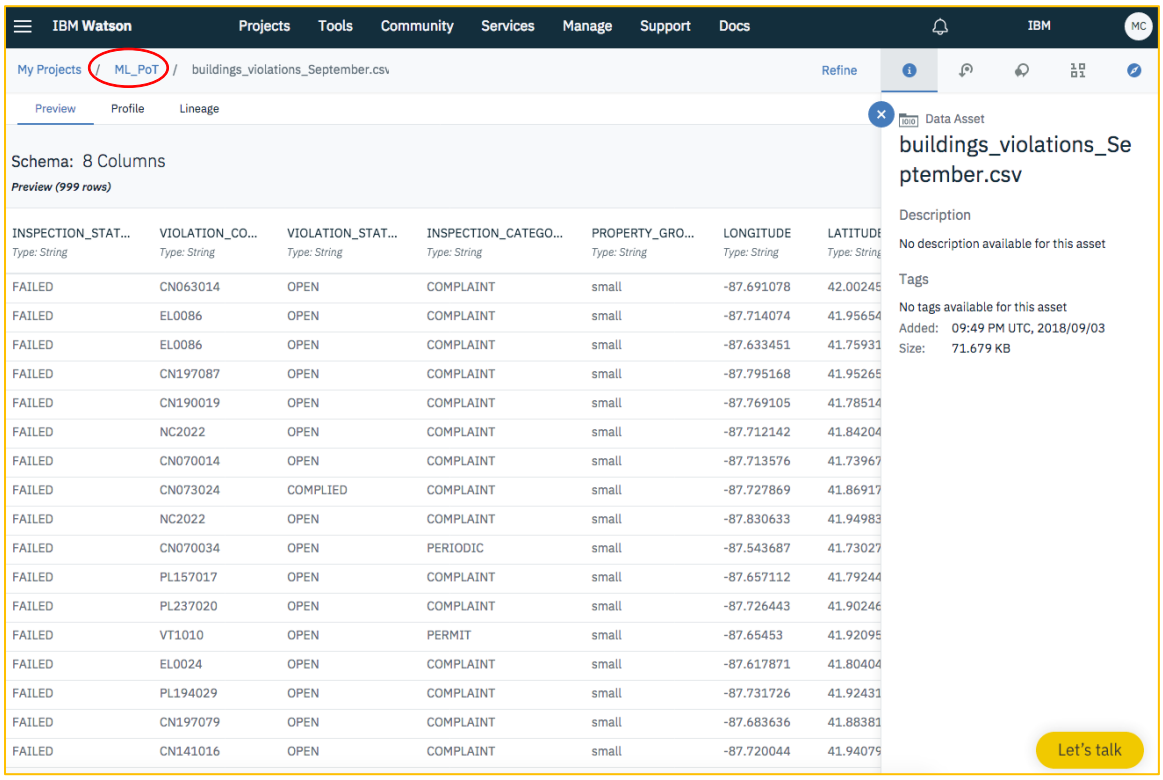


The CSV file should now be listed under "Data assets." It is now accessible by the Watson Studio modeling tools and applications.

7. Click on building_violations_September.csv under Data assets.

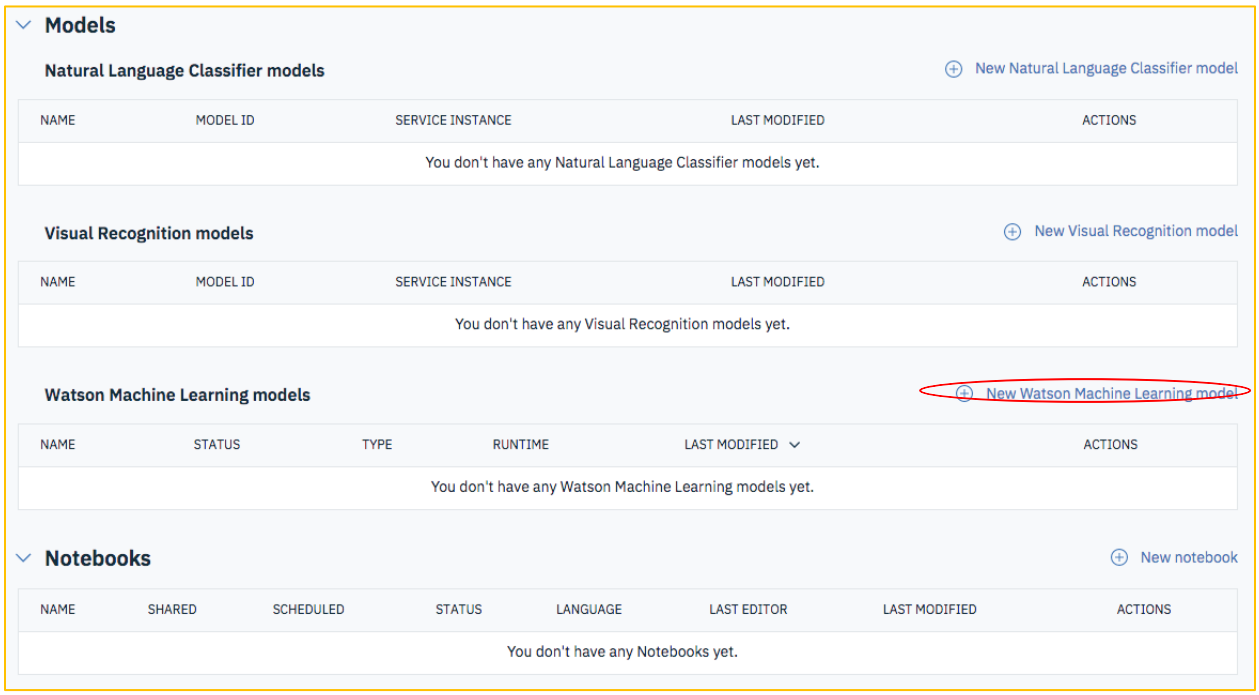


After clicking on building_violations_September.csv, you are brought to a data view. Once you are satisfied that this is the data you want, click on your project name to return to your project's assets.

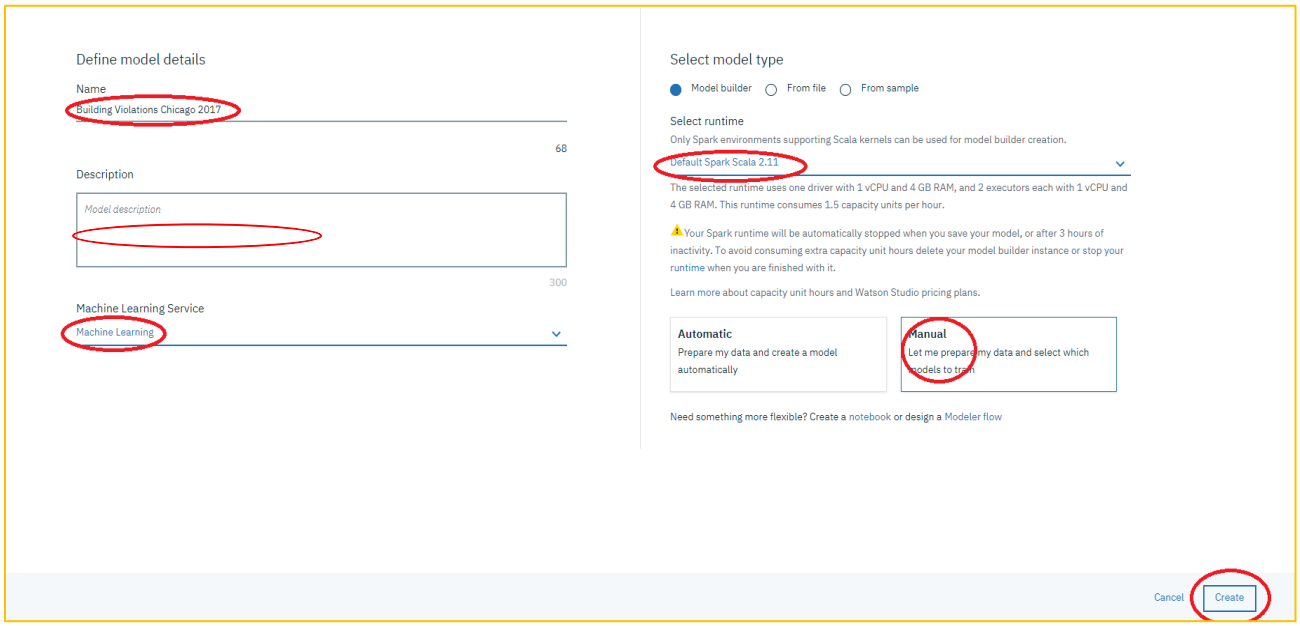


Step 2 Train, Compare, and Select a Machine Learning Model

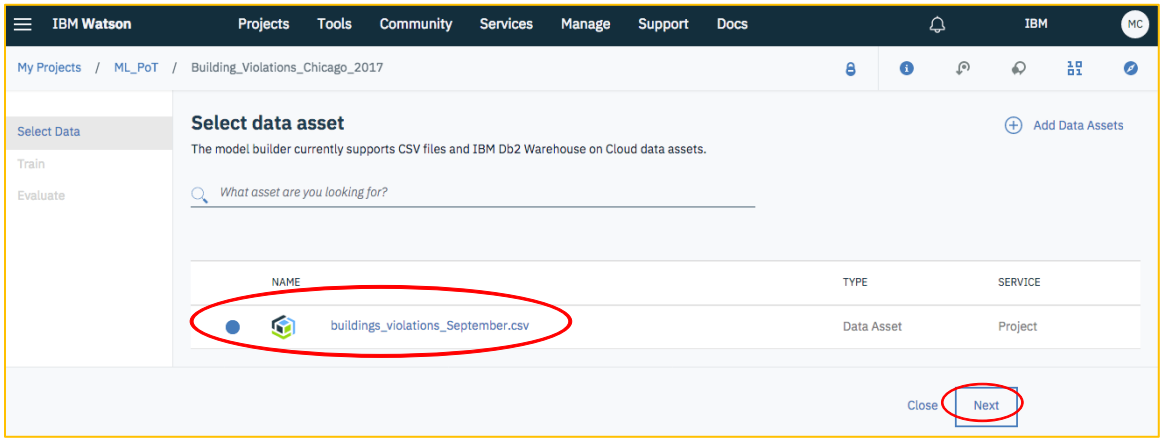
1. Scroll down and click New Watson Machine Learning model.



2. Enter a Name and a Description. Note that the Machine Learning Service should already be filled in from the prerequisites. Select the Default Spark Scala environment, 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.

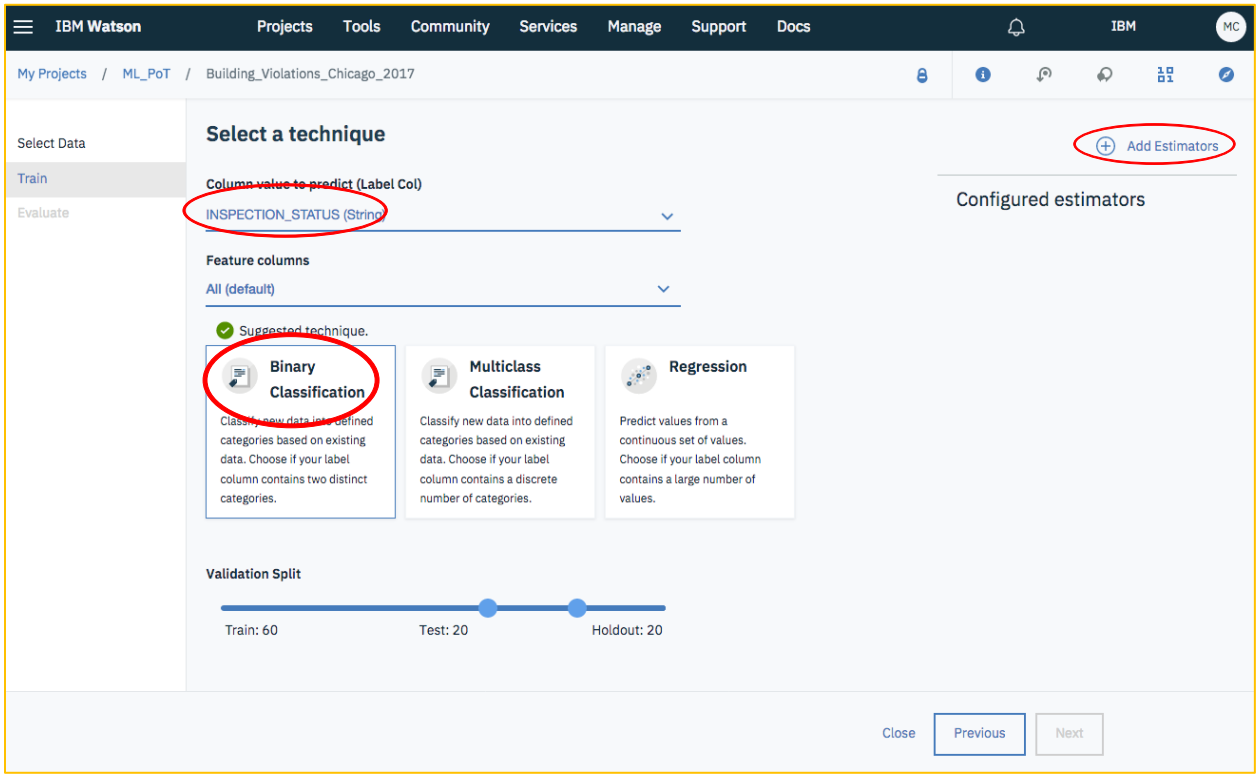


3. You will be brought to a "Select data asset" page. Select building_violations_September.csv as our data asset and click Next.



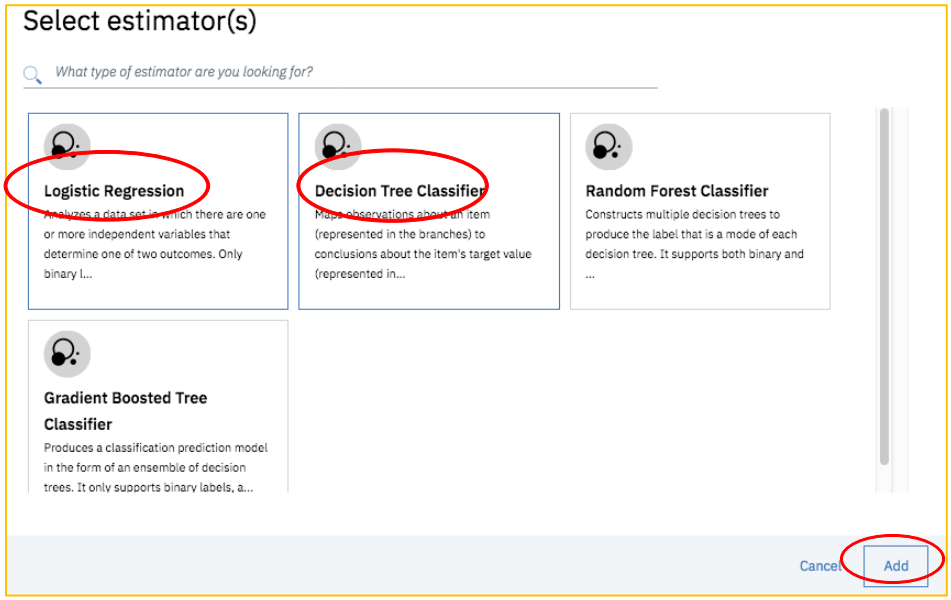
We will now choose which techniques may work well given our data and given our goal to predict which buildings are most likely to not pass inspection.

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



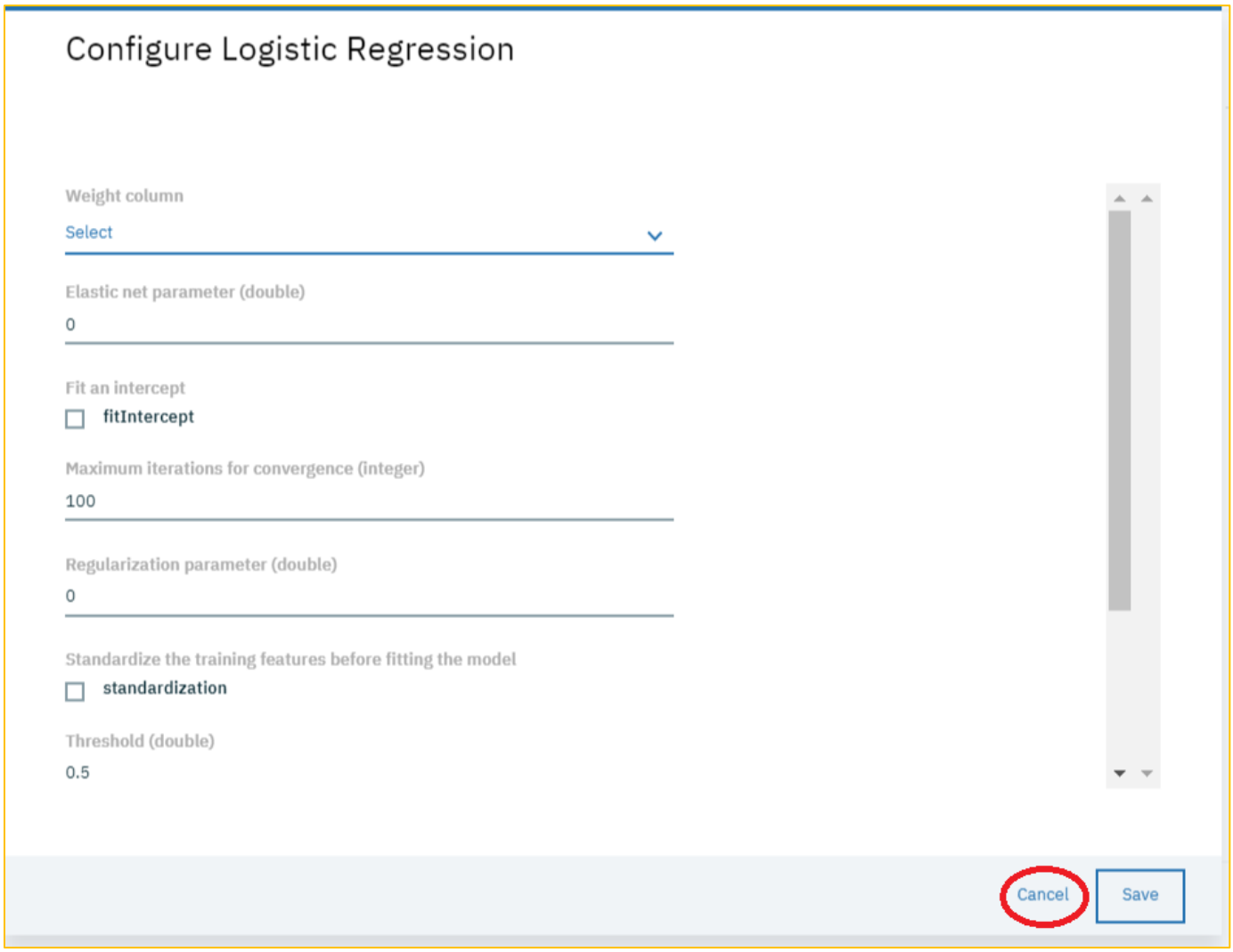
Watson Studio only presents the estimators that would work with our selections on the previous page.

7. Select Logistic Regression and Decision Tree Classifier.
8. Click Add.

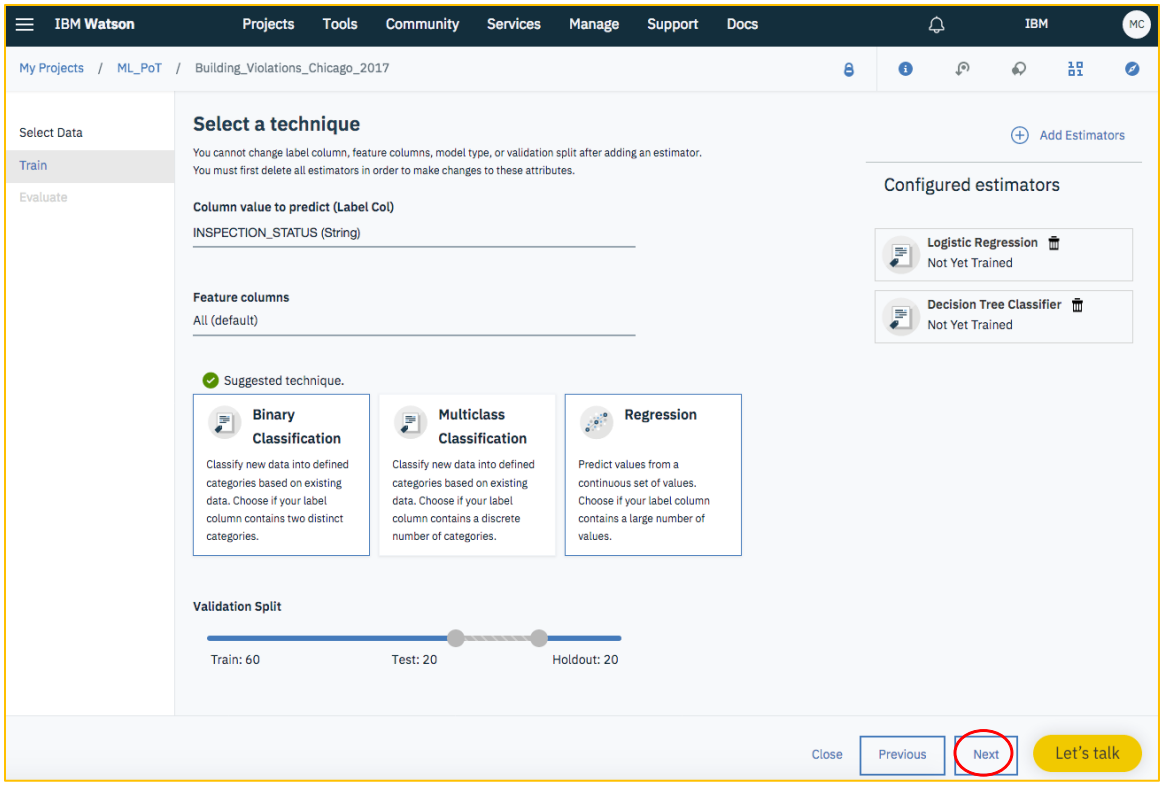


9. Note that you can click on the configured estimators to tune the hyperparameters associated with the algorithm. Click on Logistic Regression.

10. We will not change any hyperparameters, so click on Cancel.

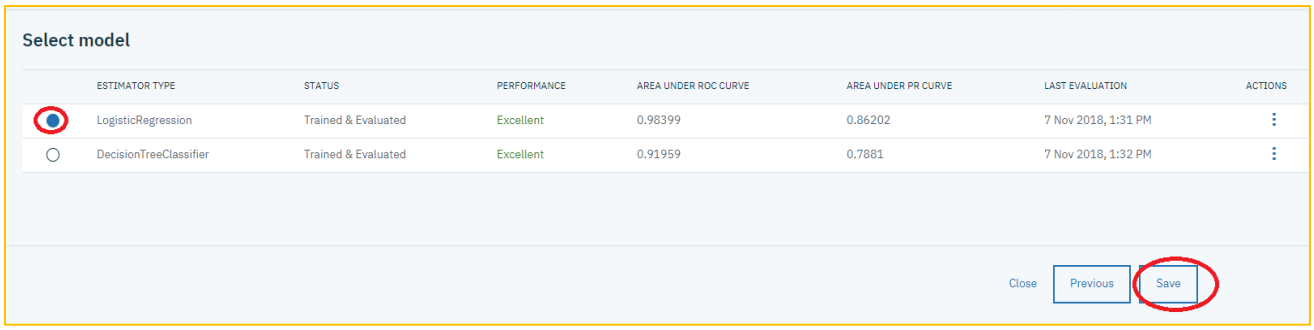


11. Back on the “Select a technique” page, click Next.

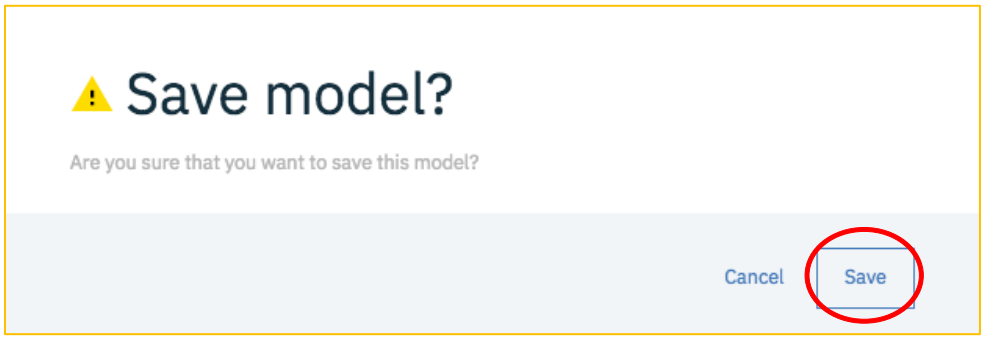


The Select model page will allow us to compare the results of different estimator types.

12. Both models did well. The Logistic Regression model performed slightly better. Select Logistic Regression.
13. Click Save.

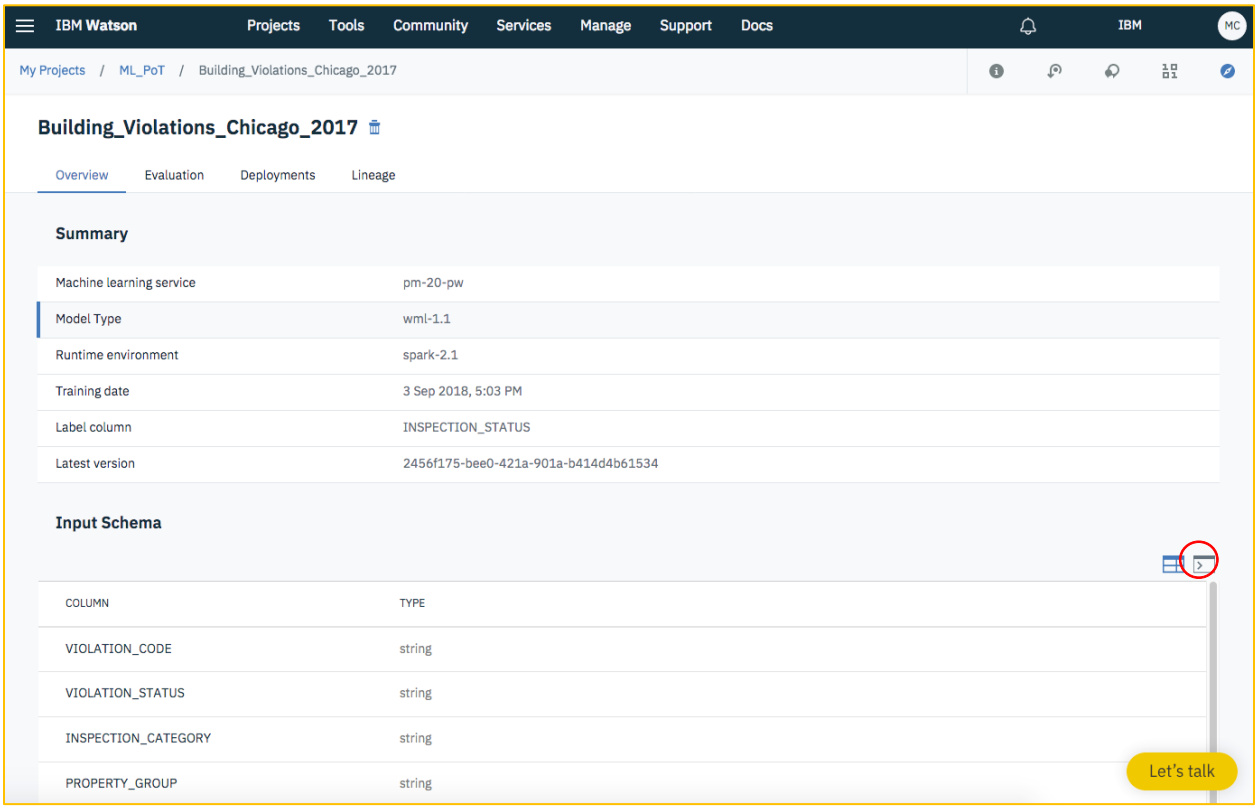


14. When the “Save model” window appears, click Save.

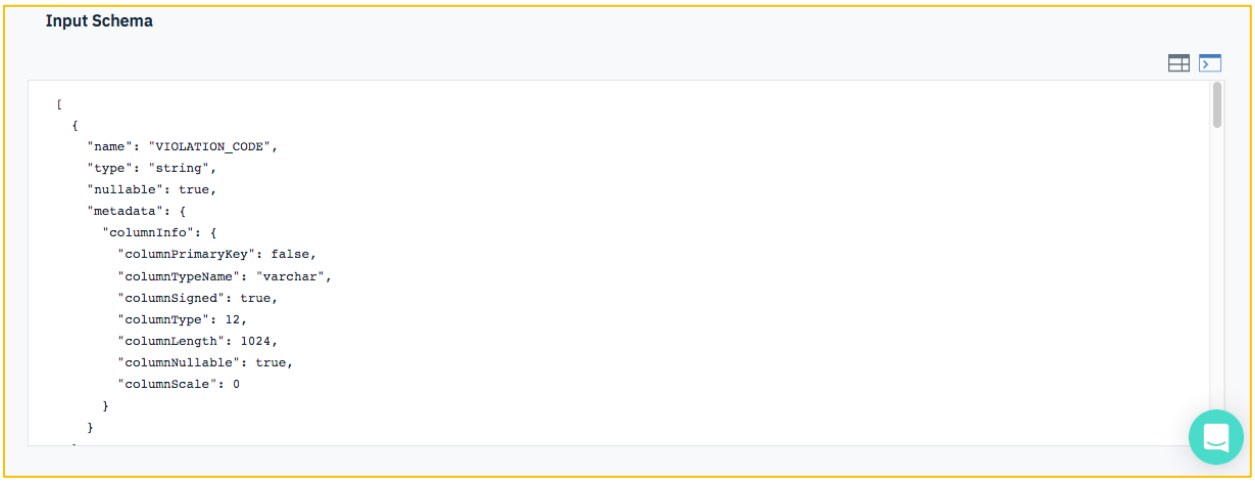


We have just saved the model in our Watson Machine Learning service and can now view information about the model specifications, details of construction, and input schema. The model can also be exposed as an API and used by data pipelines, applications, or other external systems.

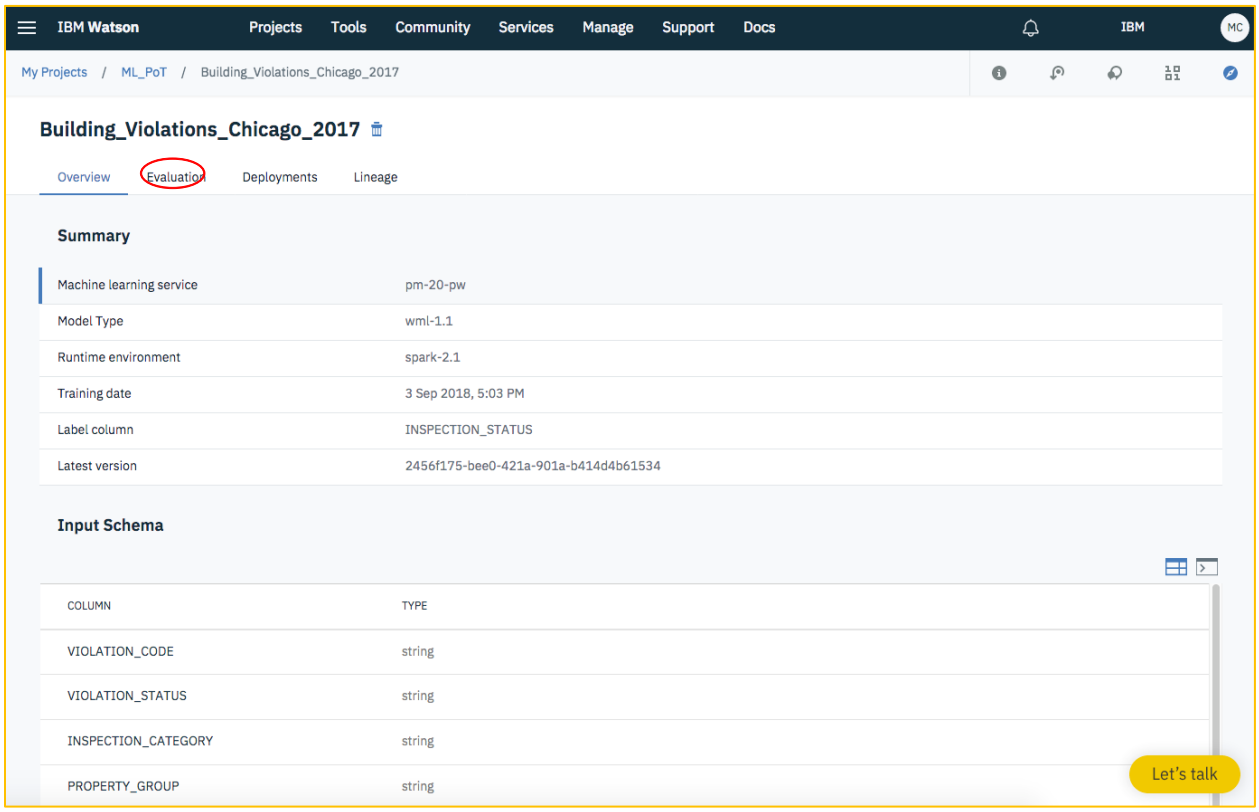
15. Click on the console button with an “angle bracket” to view the schema in JSON format.



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



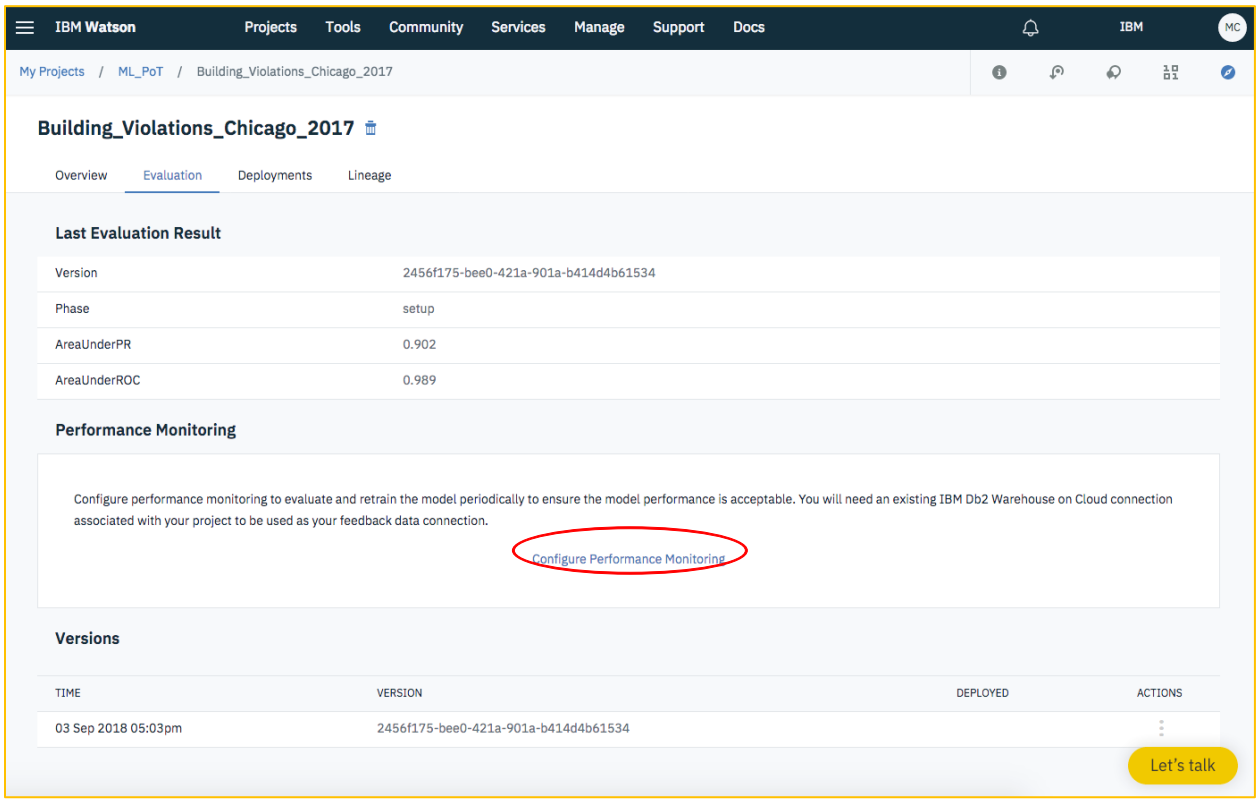
16. Scroll up and click on the Evaluation tab.



Step 3 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.

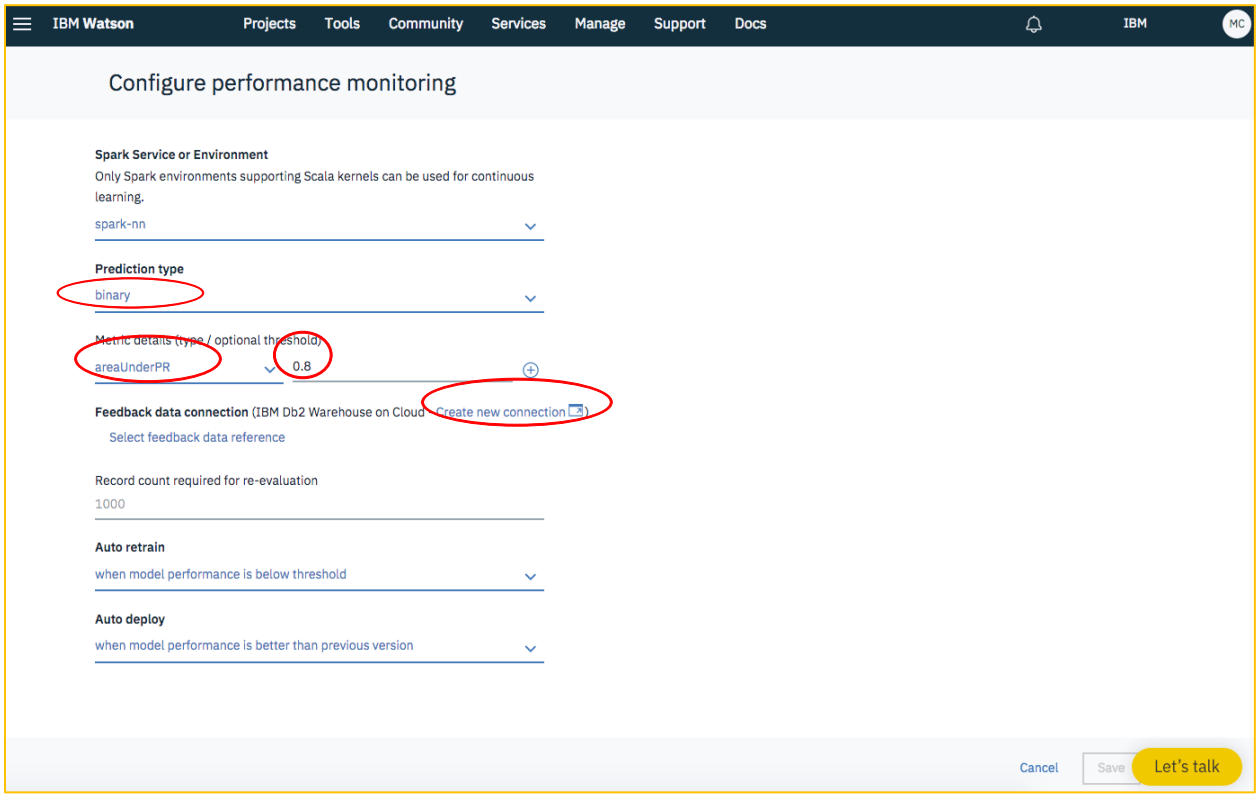
1. Click on Configure Performance Monitoring.



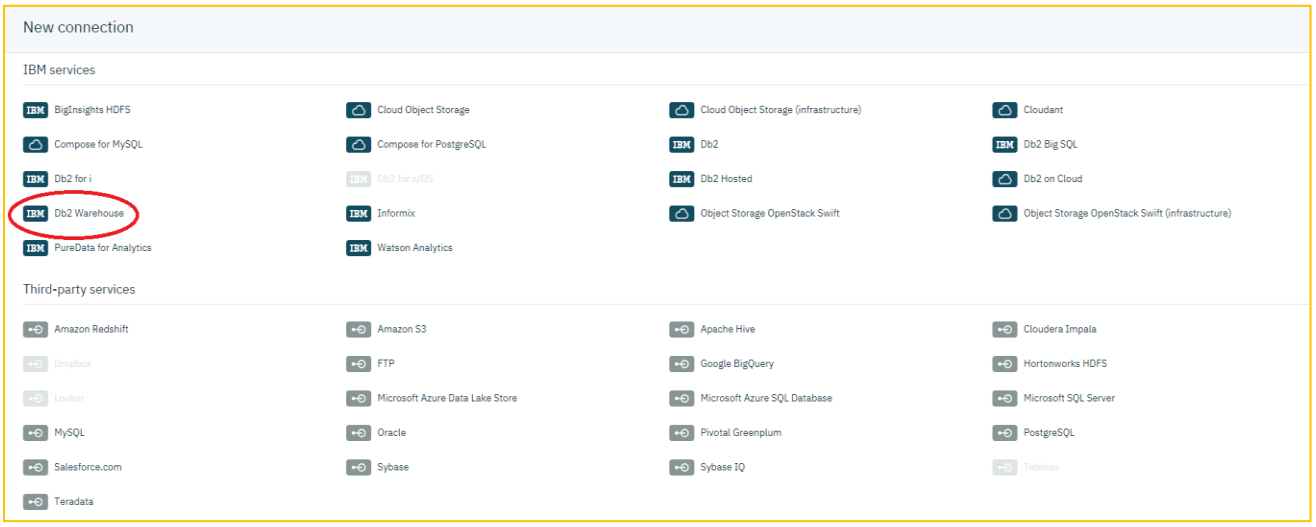
2. Under Prediction type, select binary.
3. Under Metric details, select areaUnderPR and enter 0.8.

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

4. Click on Create a new connection.



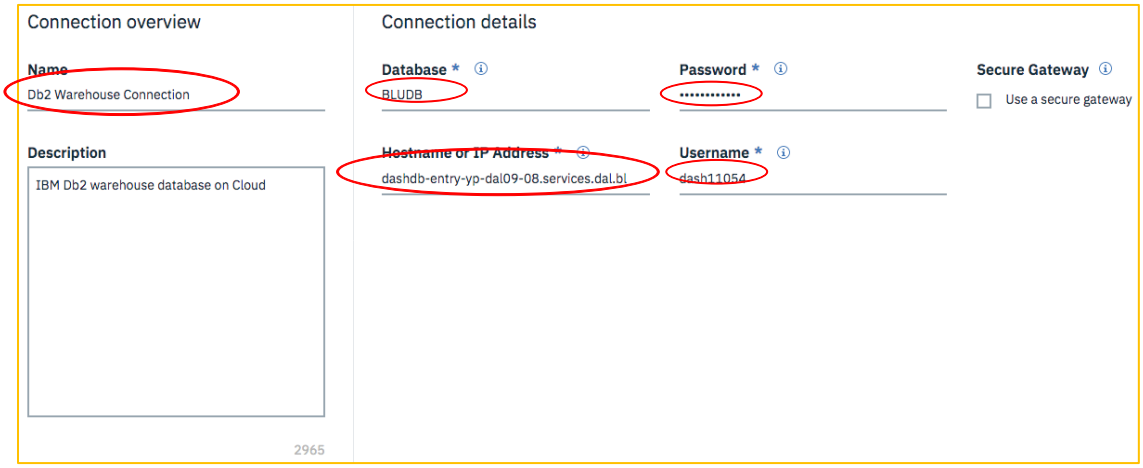
5. On the “New connection” page, click DB2 Warehouse.



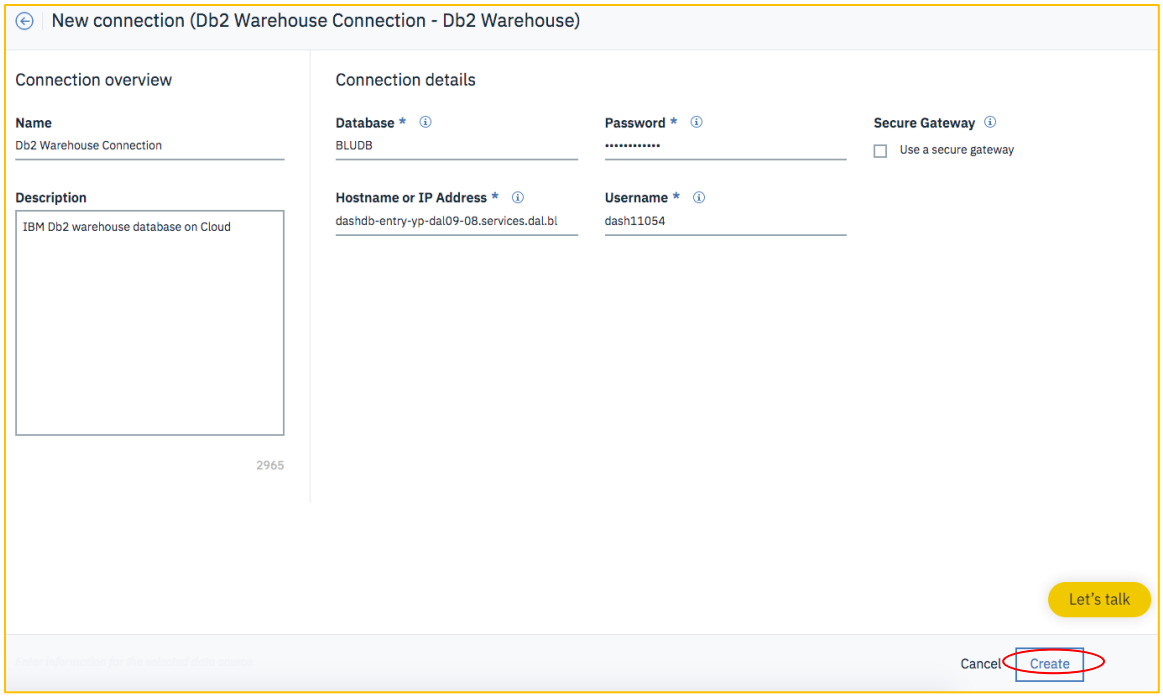
6. Access [DB2 Credentials]

Ask your instructor for your assigned database.

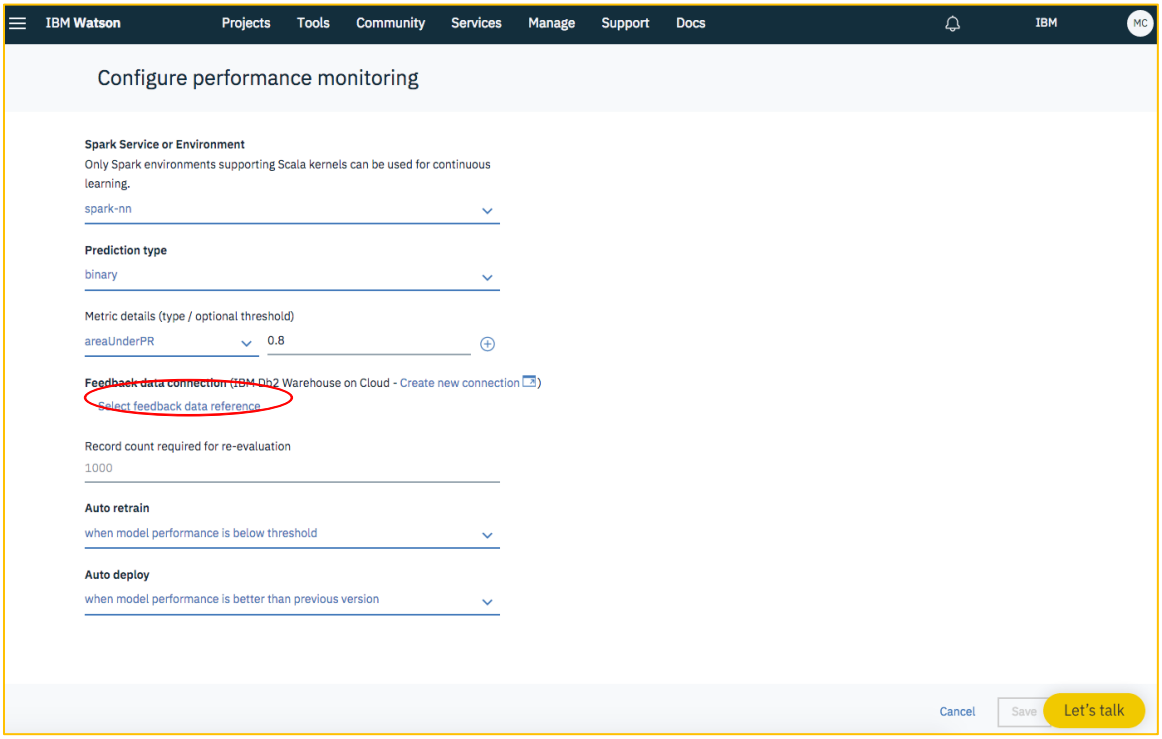
7. Enter a Name for your connection (e.g. DB2 Warehouse Connection).
8. Using the credentials cut and paste the fields under Hostname, Database, Username, and Password.



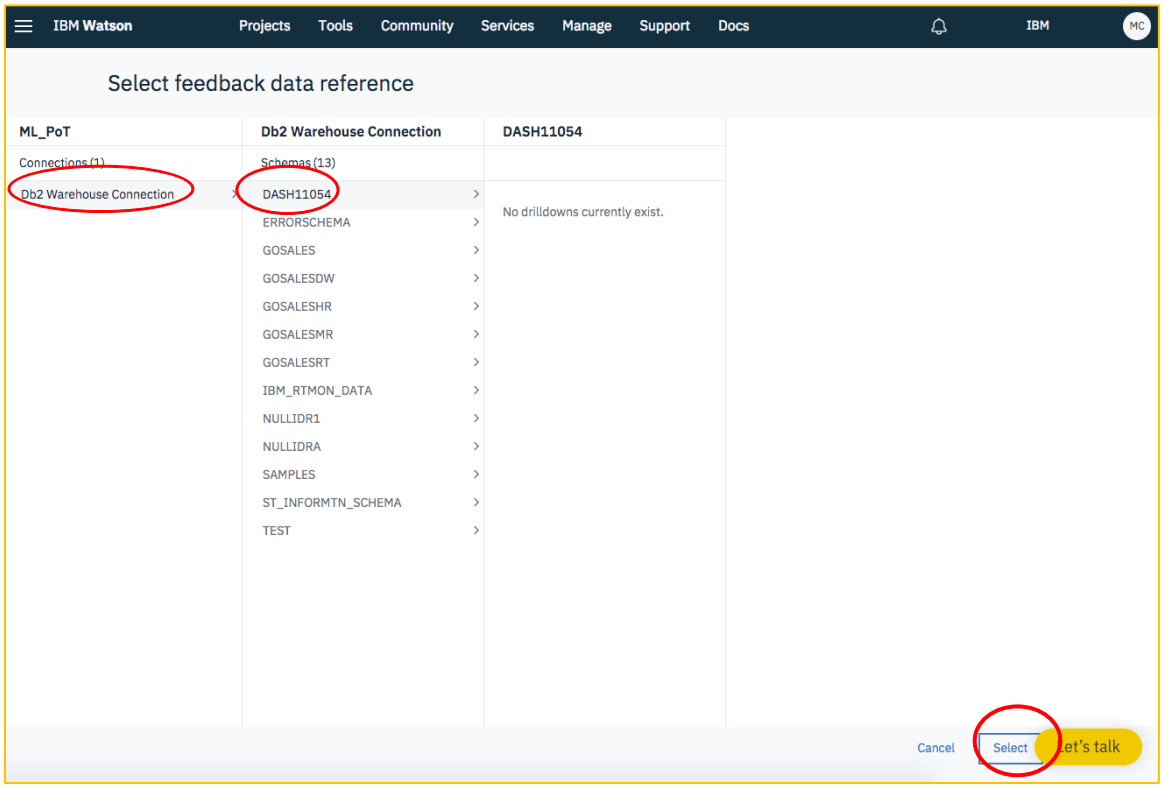
9. Click Create.


New connection (Db2 Warehouse Connection - Db2 Warehouse)
Connection overview
Name
Db2 Warehouse Connection
Description
IBM Db2 warehouse database on Cloud
2965
Connection details
Database *
BLUDB
Password *
Secure Gateway
Use a secure gateway
Hostname or IP Address *
dashdb-entry-yp-dal09-08.services.dal.bl
Username *
dash11054
Let's talk
Cancel
Create

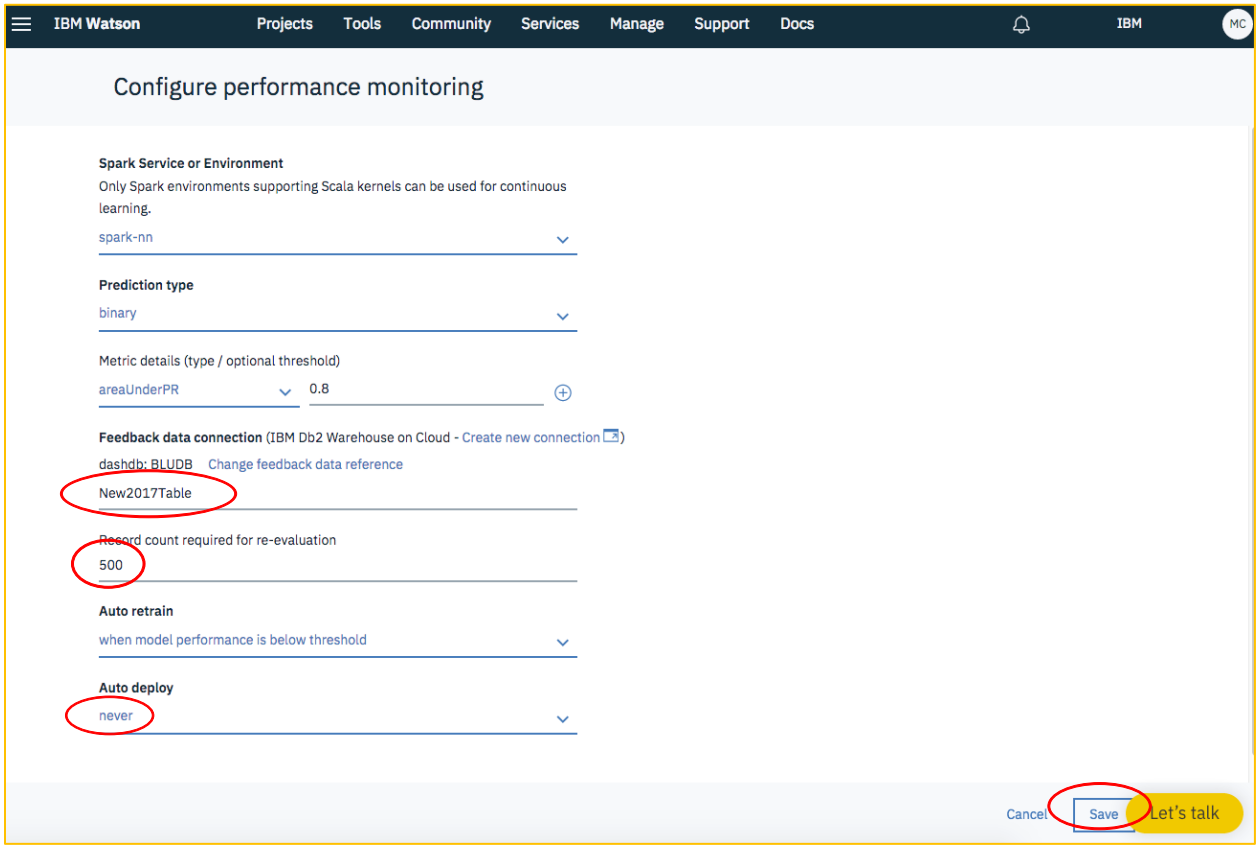
10. Return to the “Configure performance monitoring page” and click Select feedback data reference.



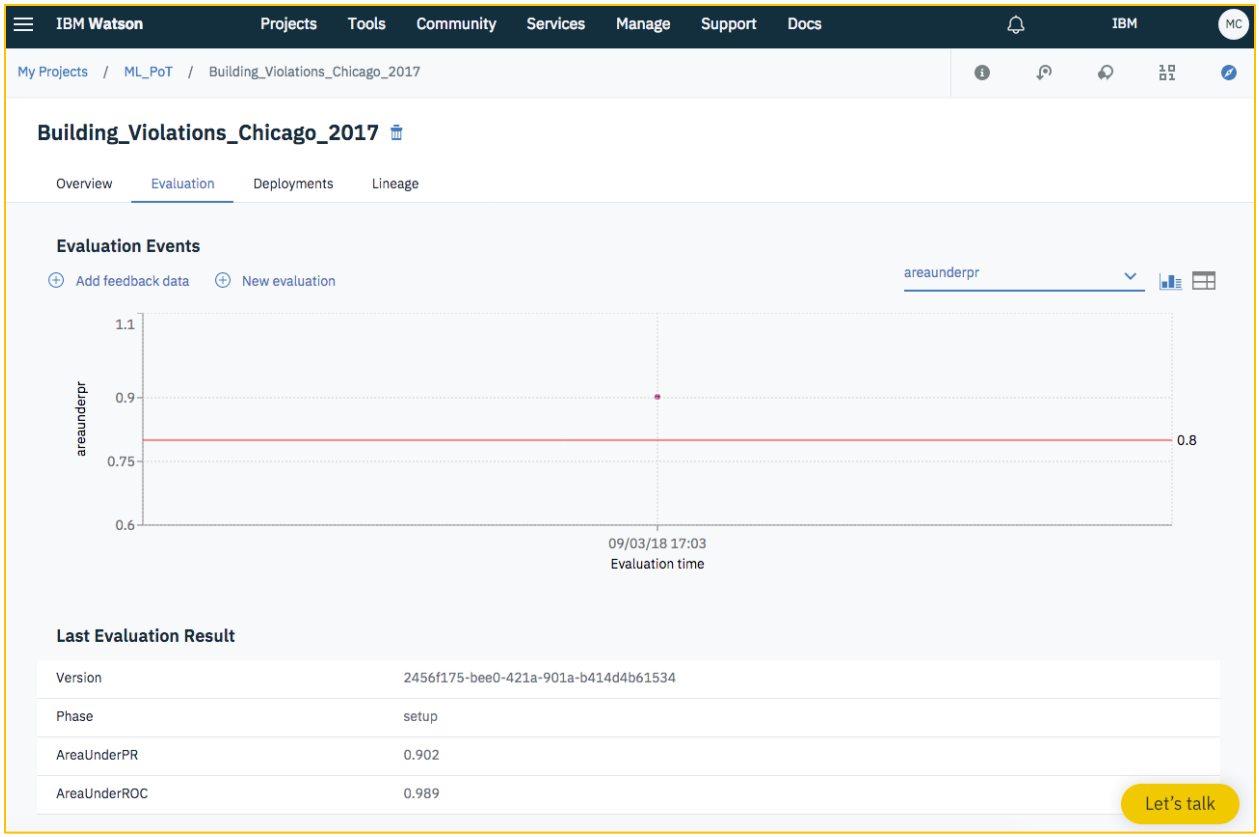
11. Click on your DB2 Warehouse-xx name.
12. Select the schema that matches DASH####.
13. Click select.



14. After returning to the “Configure performance monitoring” page, enter a unique table name (attendees are sharing the DB2 Warehouse service so make your name unique by appending your initials to the table name).
15. Enter 500 as the Record count.
16. Leave the default for Auto retrain. Under Auto deploy, select never.
17. Click Save.

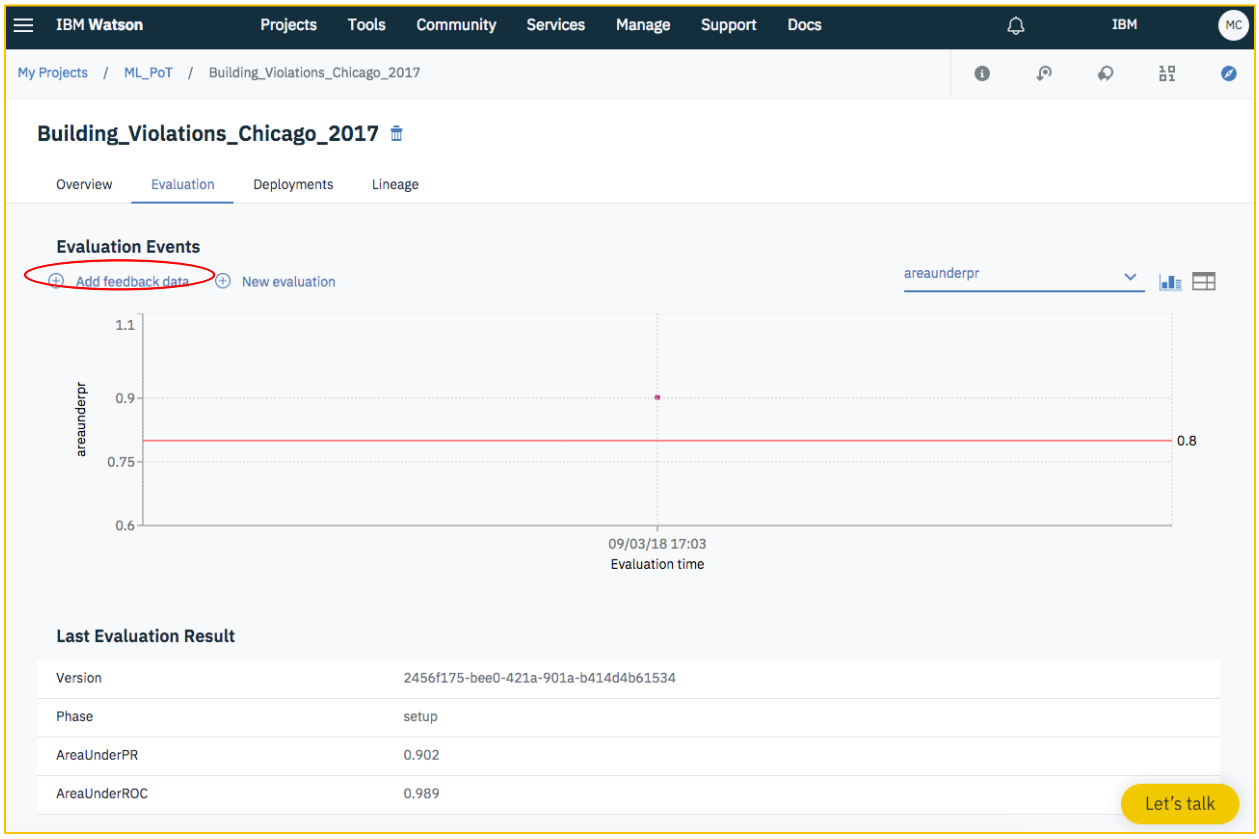


We have just set our model to retrain whenever its performance falls below 0.8. The evaluation screen has changed to reflect the decisions we have just made. Notice the graph lists our current areaunderPR and the threshold we have set. Clicking on Add feedback data will upload a new dataset. Clicking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



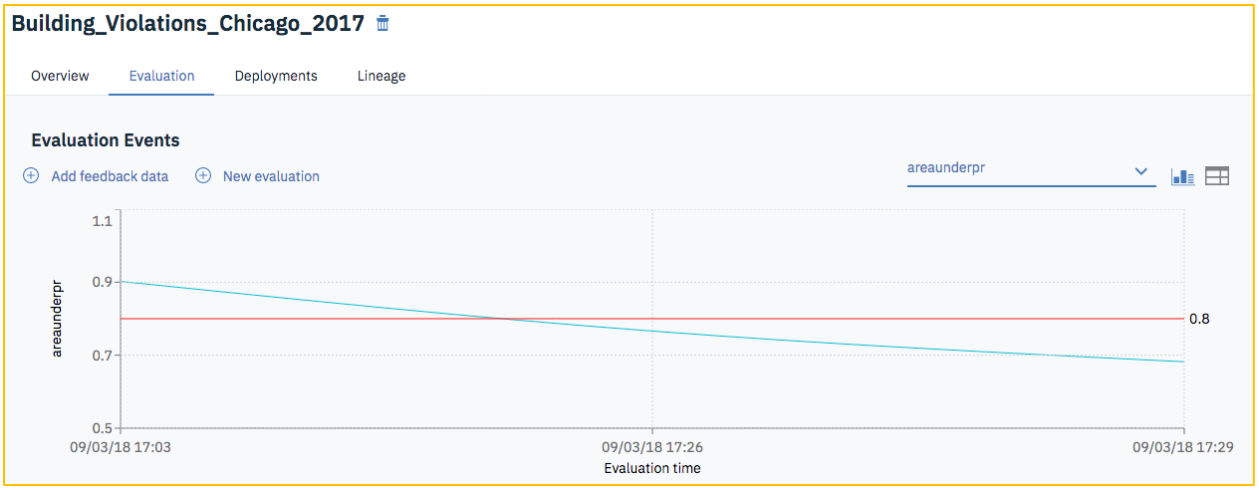
Suppose September has passed, and we now have October data. We can trigger a new evaluation by adding more feedback data.

18. Click Add feedback data.
19. Select building_violations_October.csv.

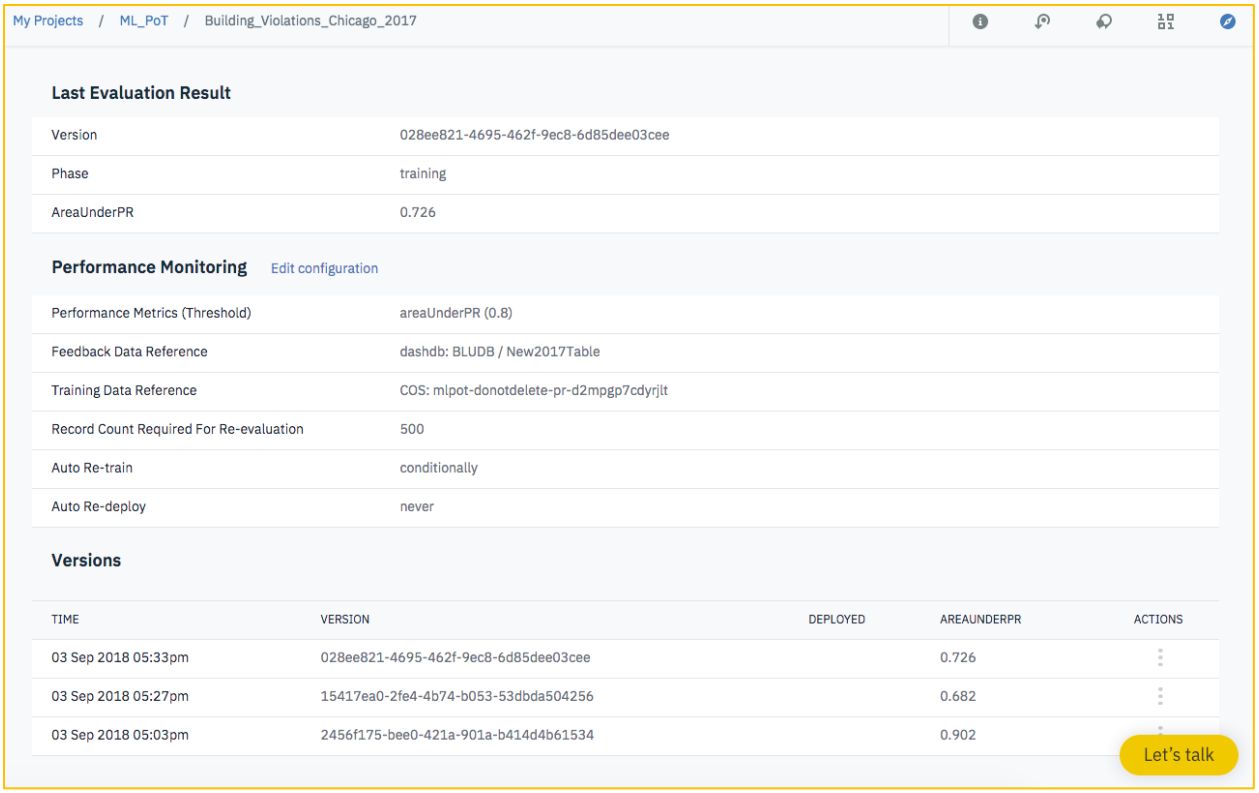


20. When the “New evaluation” window appears, click New Evaluation. Re-evaluating the model may take a few minutes. This is a great time to grab coffee.

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.



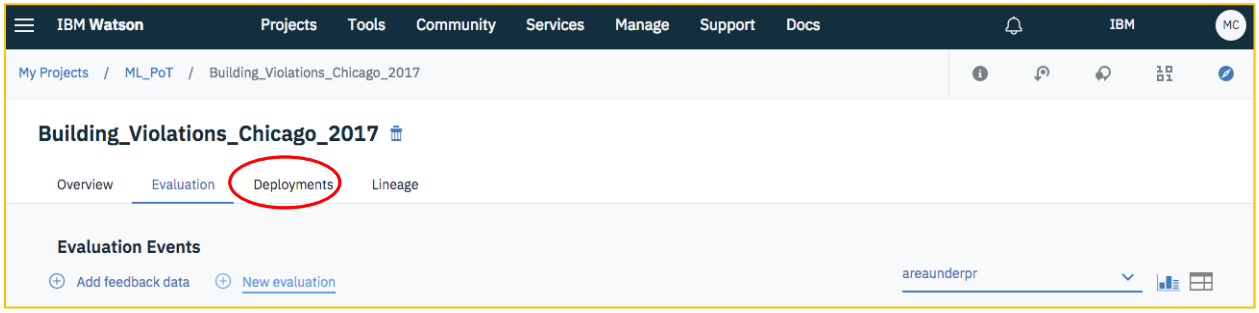
You can continue by adding the November data by the same process of hitting **Add feedback data** and adding the November file. If you scroll down you will be able to see all model versions listed. We have three versions, but no deployments.



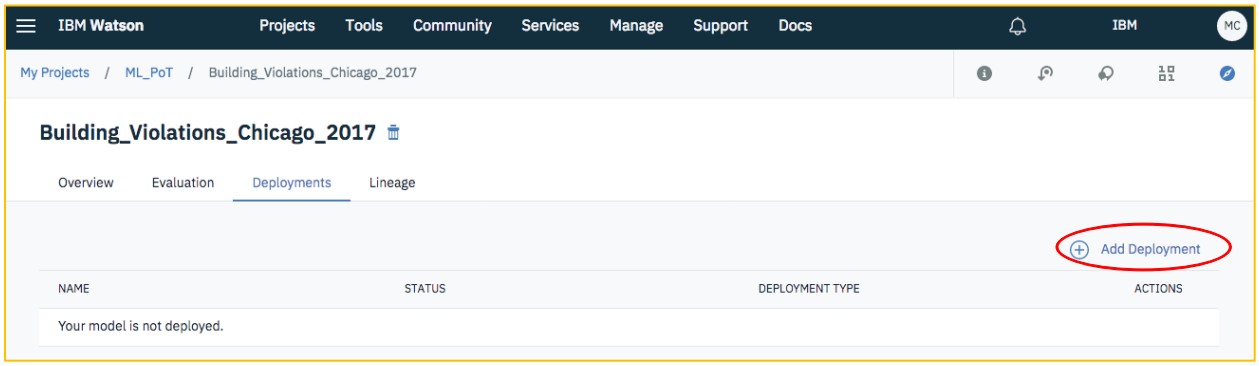
Step 4 Deploy a Machine Learning Model.

Now we will focus on deployment.

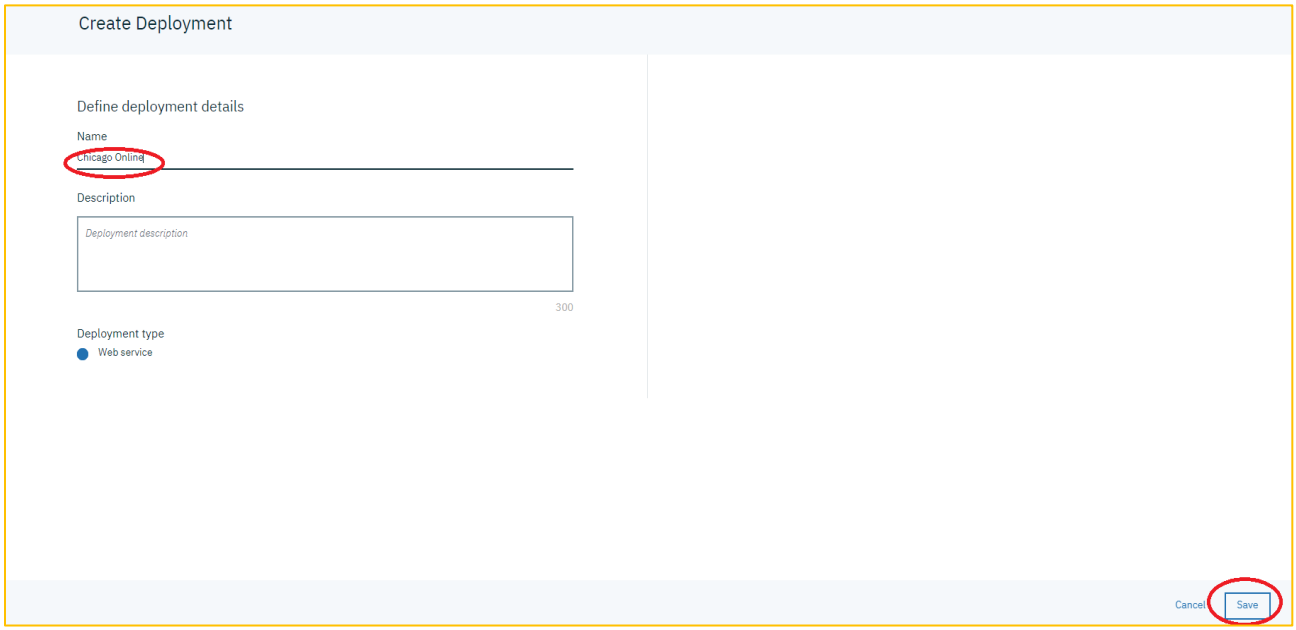
1. Click the Deployment tab.



2. Add Deployment.

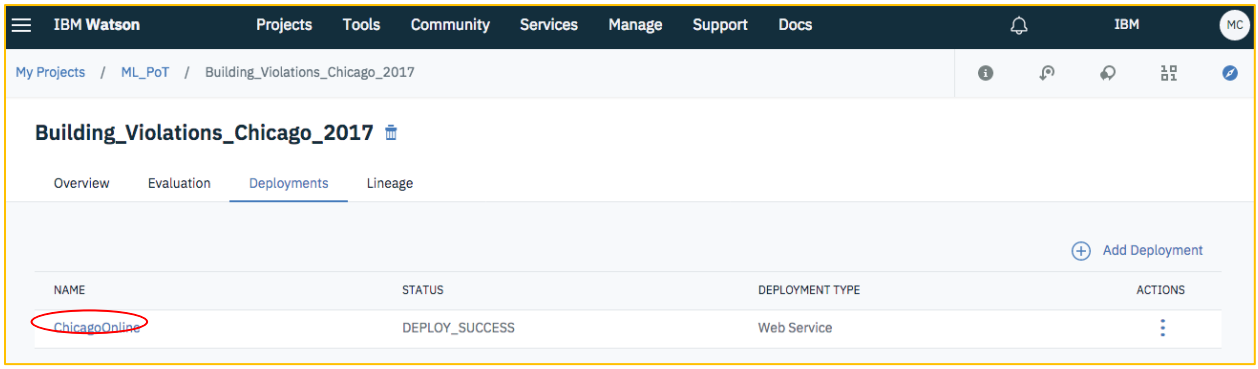


3. Enter a Name (e.g. ChicagoOnline).
4. Enter a Description.
5. Click Save.



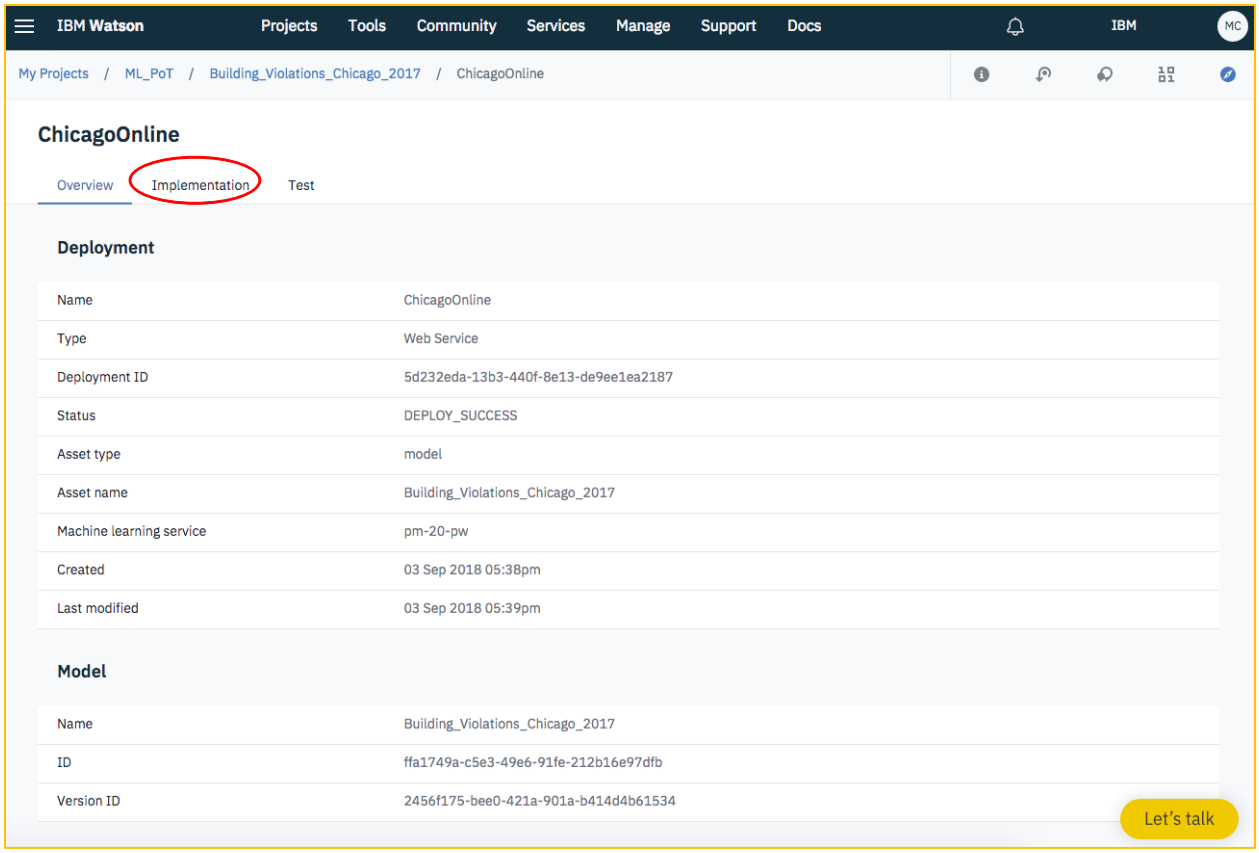
The model is now accessible by external systems.

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



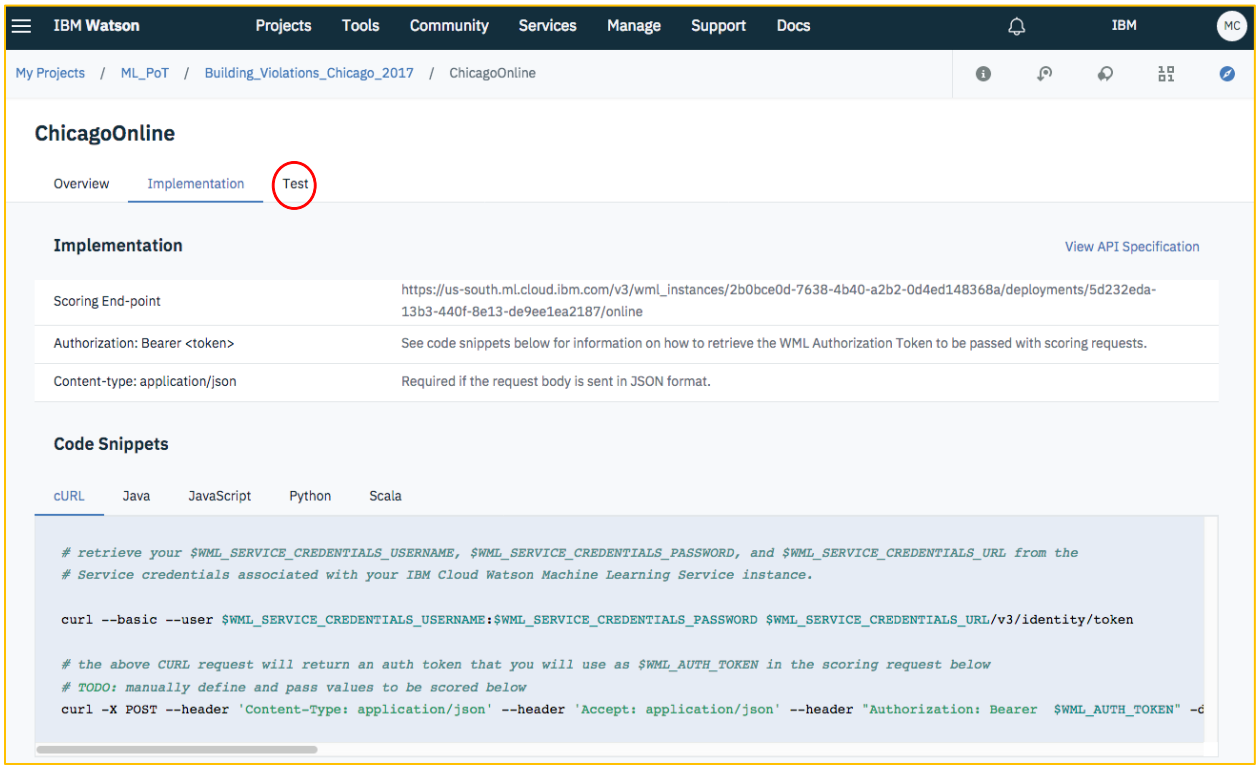
You are automatically brought to the Overview tab lists information related to the model deployment including services used and version number.

7. Click the Implementation tab.


IBM Watson
Projects
Tools
Community
Services
Manage
Support
Docs
IBM
MC
My Projects / ML_PoT / Building_Violations_Chicago_2017 / ChicagoOnline
ChicagoOnline
Overview
Implementation
Test
Deployment
Name ChicagoOnline
Type Web Service
Deployment ID 5d232eda-13b3-440f-8e13-de9ee1ea2187
Status DEPLOY_SUCCESS
Asset type model
Asset name Building_Violations_Chicago_2017
Machine learning service pm-20-pw
Created 03 Sep 2018 05:38pm
Last modified 03 Sep 2018 05:39pm
Model
Name Building_Violations_Chicago_2017
ID ffa1749a-c5e3-49e6-91fe-212b16e97dfb
Version ID 2456f175-bee0-421a-901a-b414d4b61534
Let's talk

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

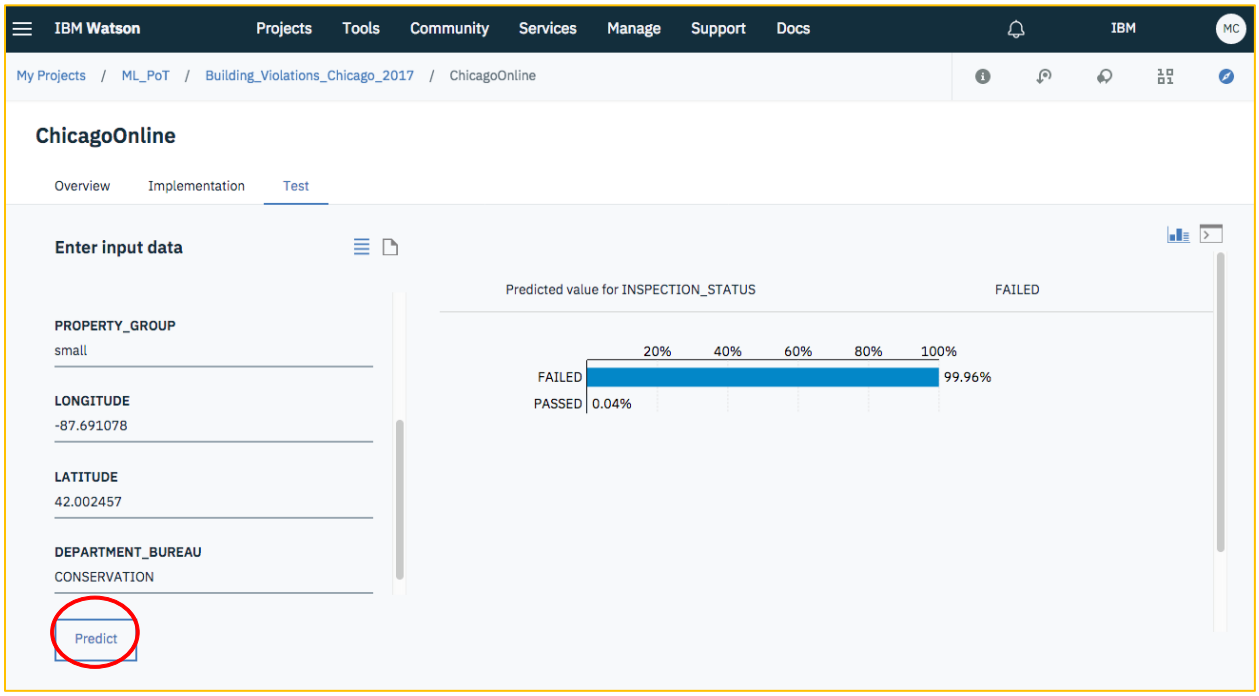
   PROPERTY_GROUP: small

   LONGITUDE: -87.691078

   LATITUDE: 42.002457

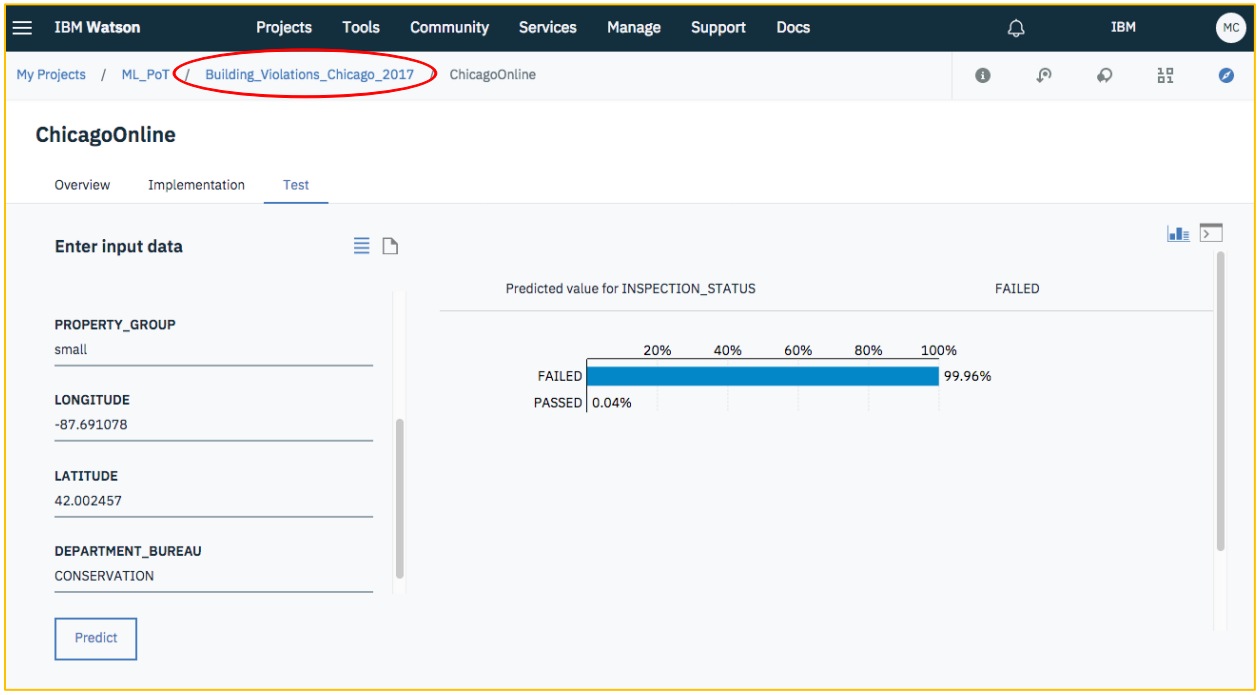
   DEPARTMENT_BUREAU: CONSERVATION

10. Click Predict.

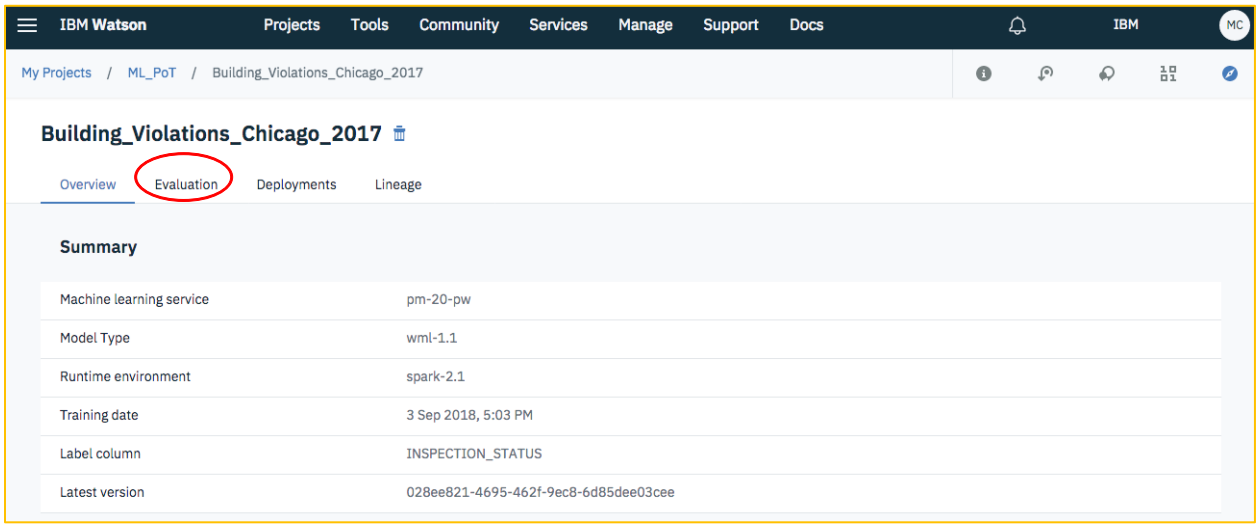


According to our model, a building with our input data has a 0.04% chance of passing inspection (your numbers may differ slightly)

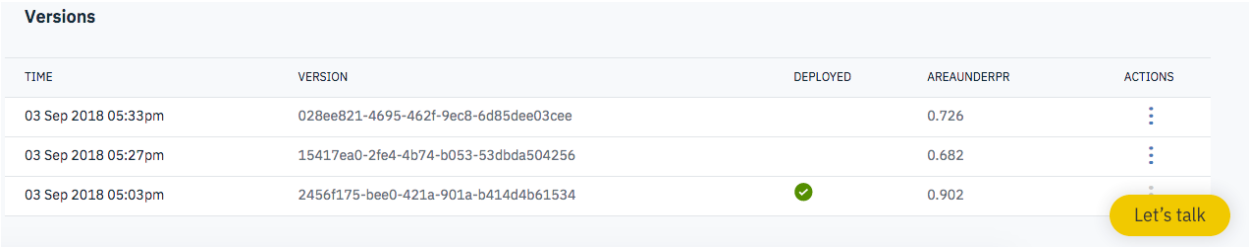
11. Feel free to change the input data and run new predictions. When ready, click on you model name (e.g. Building_Violations_Chicago_2017)



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. You can select the vertical ellipse under ACTIONS to change the deployed version.



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 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 02

We Value Your Feedback!

- Don’t forget to submit your Think 2019 session and speaker feedback! Your feedback is very important to us – we use it to continually improve the conference.
- Access the Think 2019 agenda tool to quickly submit your surveys from your smartphone, laptop or conference kiosk.