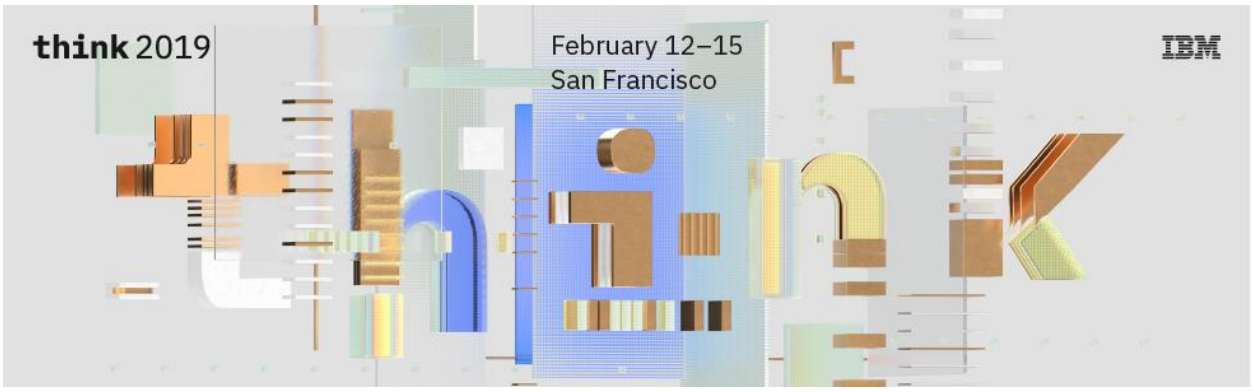


Lab Center – Hands-on Lab

Session #3259

Session Title: Introduction to Data Science using Watson Studio

Lab-3: 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

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.

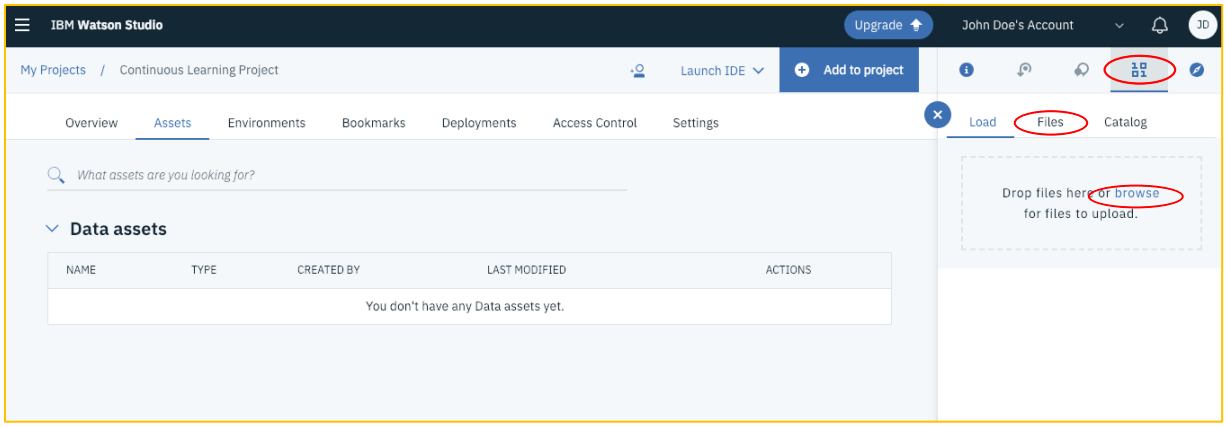
Introduction

In this lab, you will be using building a machine learning model to predict the appropriate drugs to treat patients with heart disease. This model will be deployed and made available online allowing any medical professional to see a list of drugs considered the most appropriate for each of their patients. That said, the doctors will still have the final decision. You are starting with a small dataset and will continue to be provided with data on a daily basis. Continuous learning will be set up to monitor the model performance and ensure the highest performing model is always available for doctors to use.

Step 1: Upload a New Data Asset.

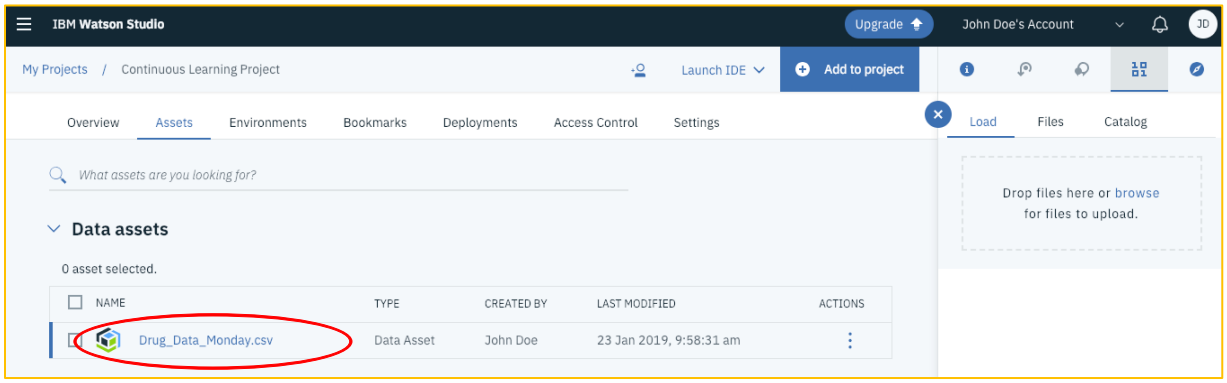
Before we build our models, we need to load data into our project.

1. Open Continuous Learning Data
2. Save the zip file to your desktop and unzip the contents.
3. Return to Watson Studio. In the top right of the screen, click the blue icon with 1's and 0's. This will open the data pallet.
4. Click on Load, browse, select the Drug_Data_Monday.csv file and click open.

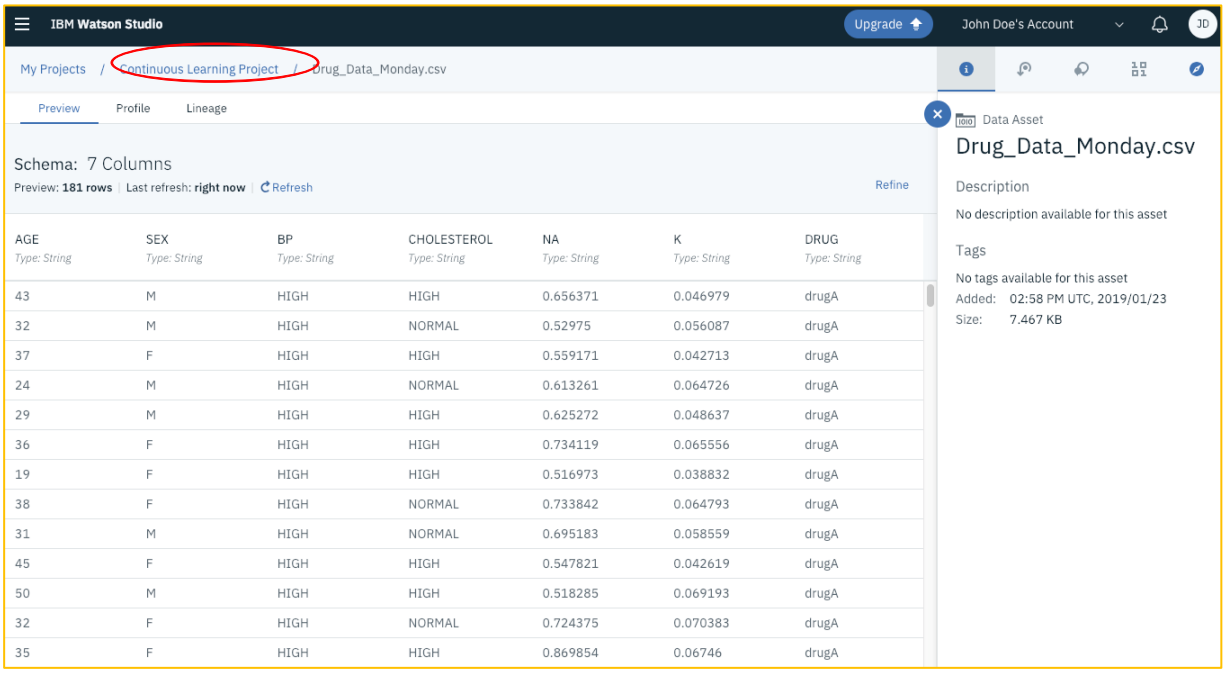


The CSV file should now be listed under "Data assets." It is now accessible by the Watson Studio modeling tools and applications.

5. Click on Drug_Data_Monday.csv under Data assets.

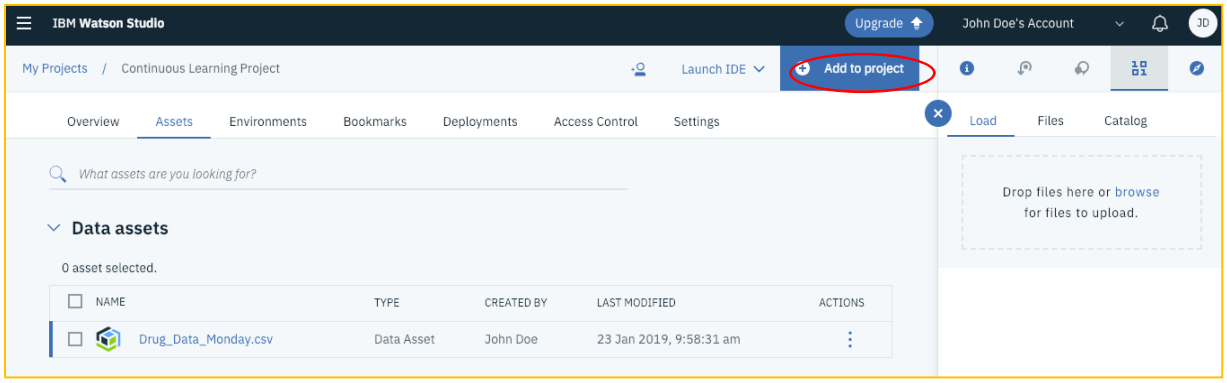


After clicking on Drug_Data_Monday.csv, you are brought to a data view. Once you are satisfied that this is the data you want, click on your project name to return to your project's assets.

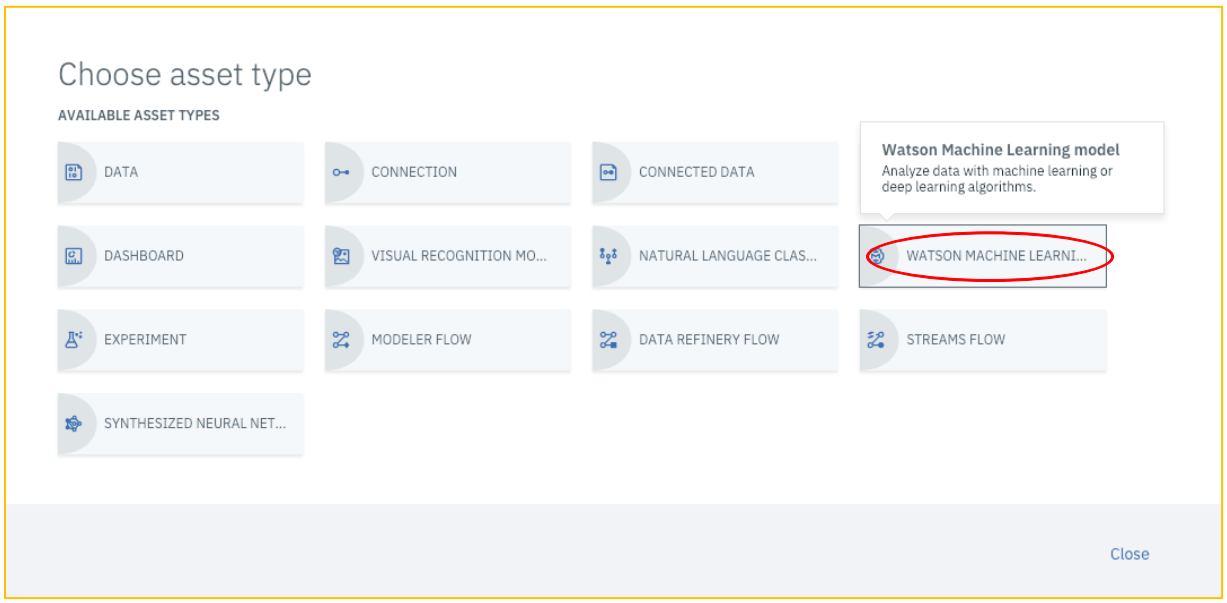


Step 2: Train, Compare, and Select a Machine Learning Model

1. Click on Add to project.

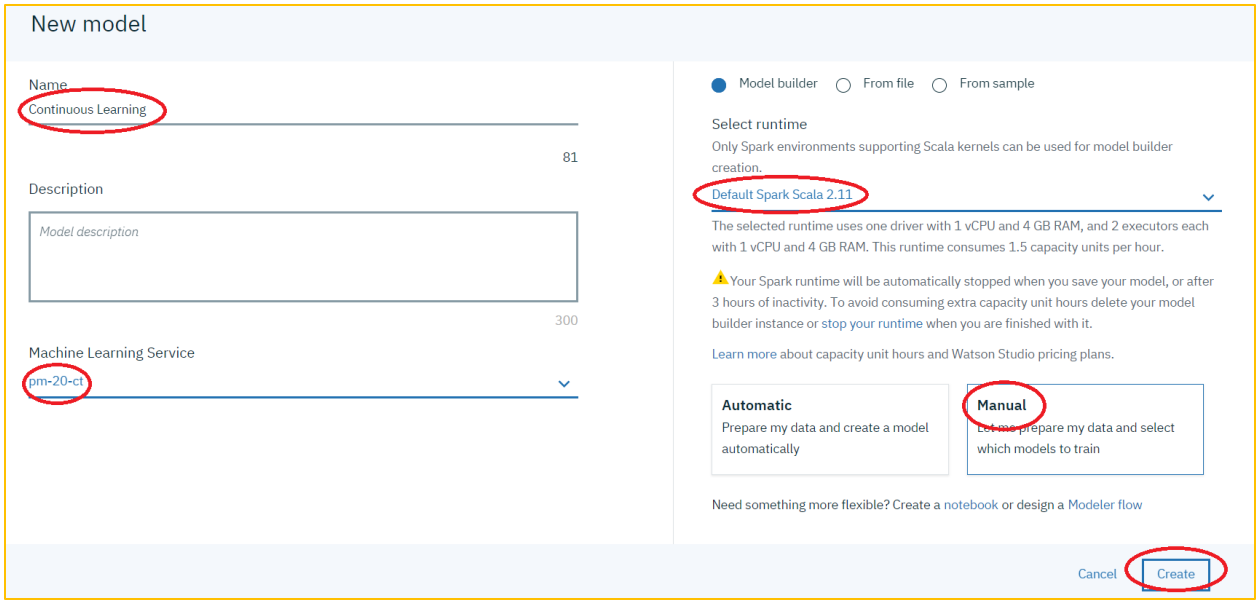


2. Click on Watson Machine Learning model.

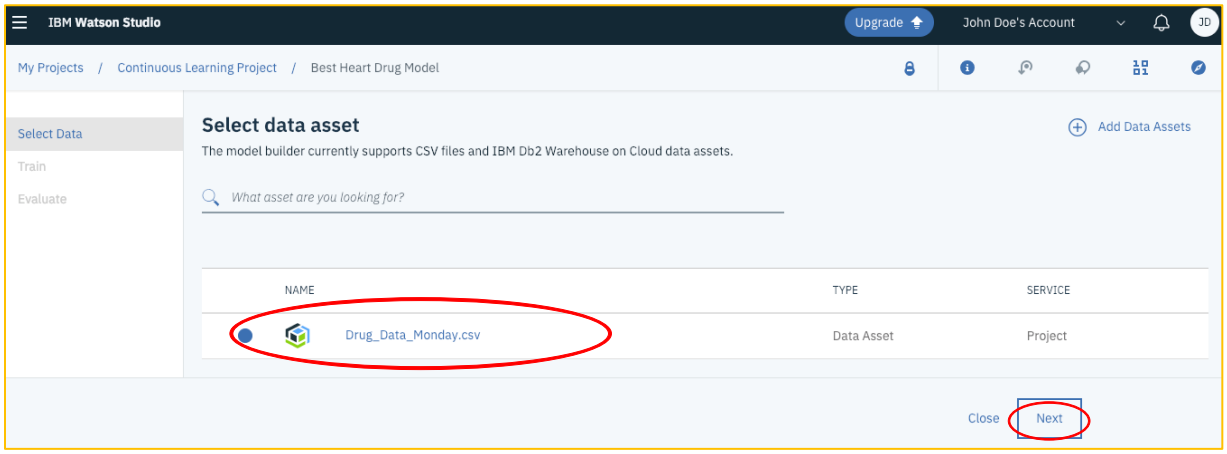


We are creating a model using a point-and-click UI called “Model builder” but can also create models through, APIs, notebooks, SPSS flows, Deep Learning flows, or the synthesized neural network tooling.

3. Enter a Name and a Description, select Default Spark Scala 2.11 under runtime, select Manual, and click Create.

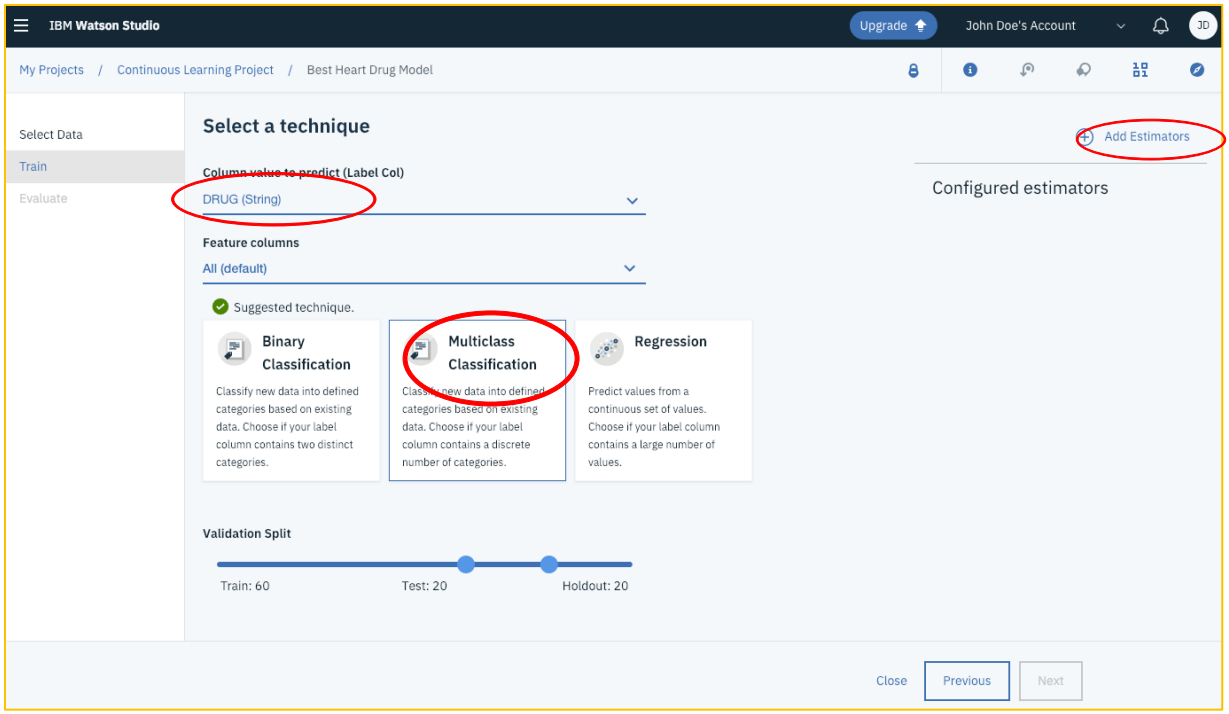


4. You will be brought to a “Select data asset” page. Select Drug_Data_Monday.csv as our data asset and click Next.



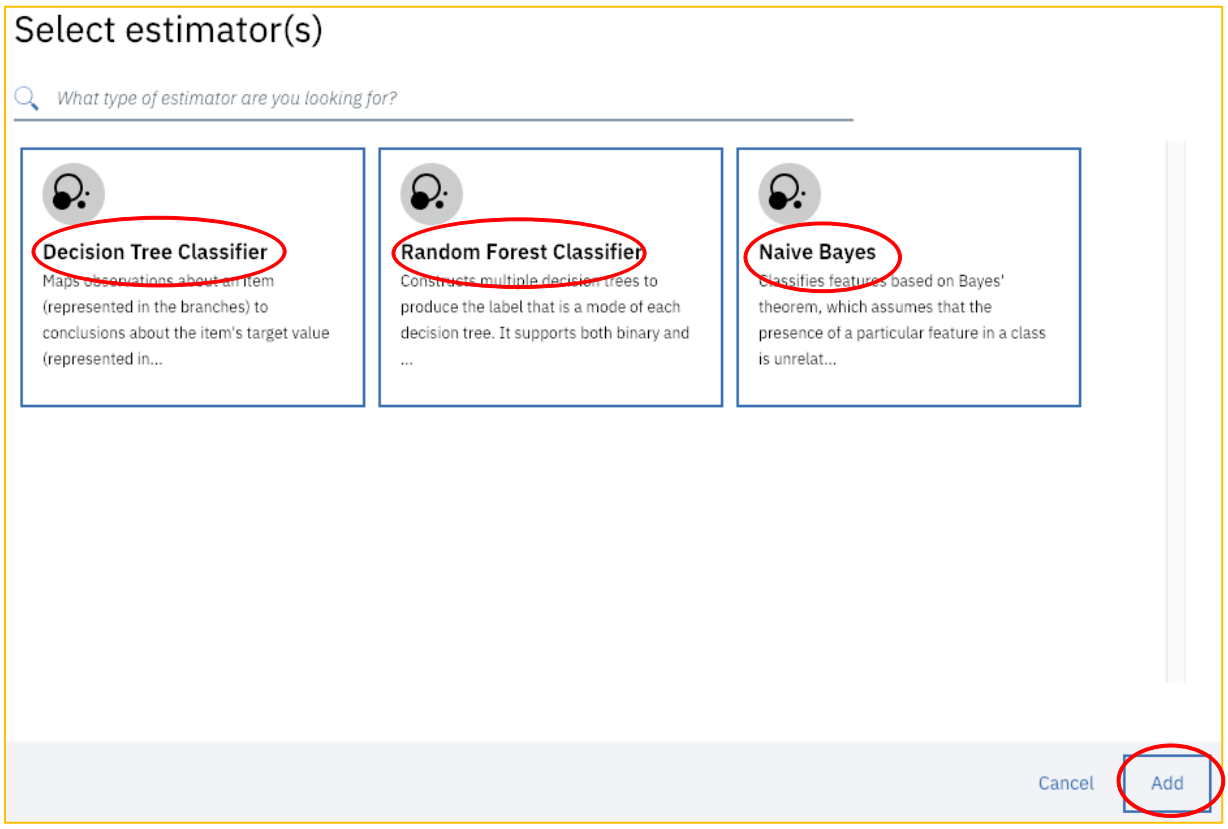
We will now choose which techniques may work well given our data and given our goal to predict which drug is the best suited for a given patient.

5. On the “Select a technique” page, click on the Select Label Col dropdown and select DRUG (String).
6. Select Multiclass Classification.
7. Click Add Estimators in the upper right corner of the page.

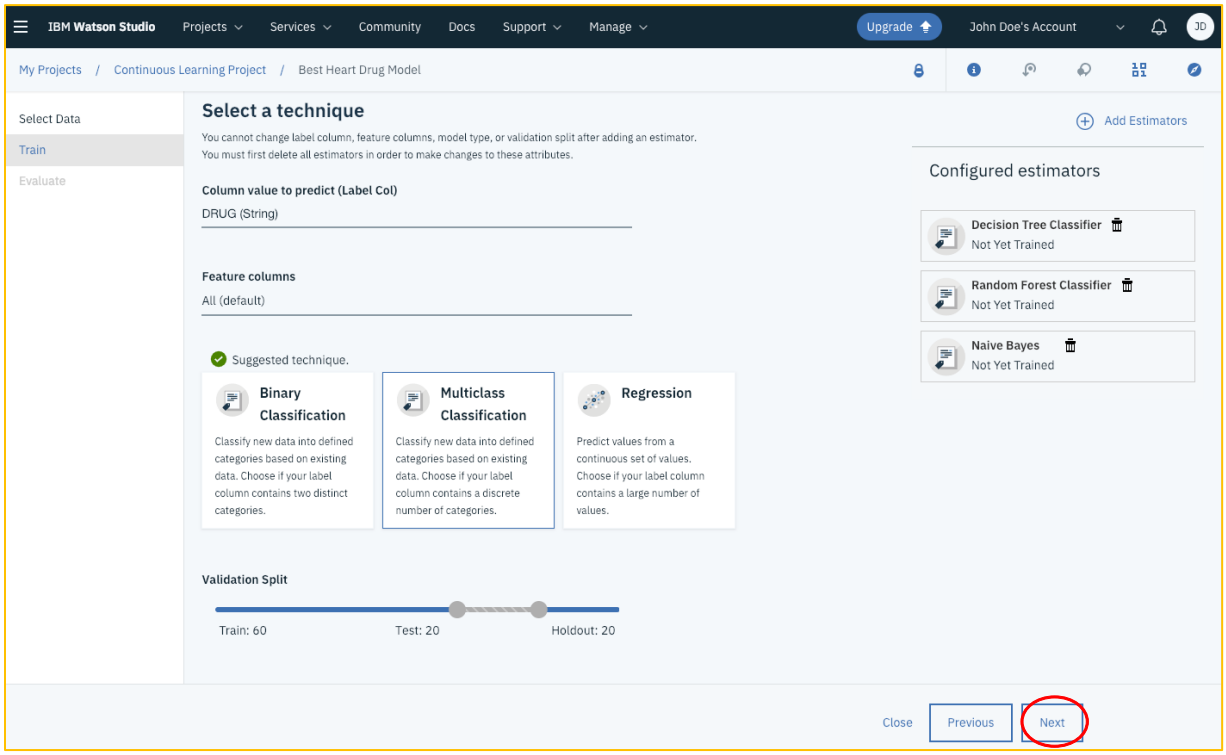


Watson Studio only presents the estimators that would work with our selections on the previous page.

8. Select all three estimator options. Click Add.

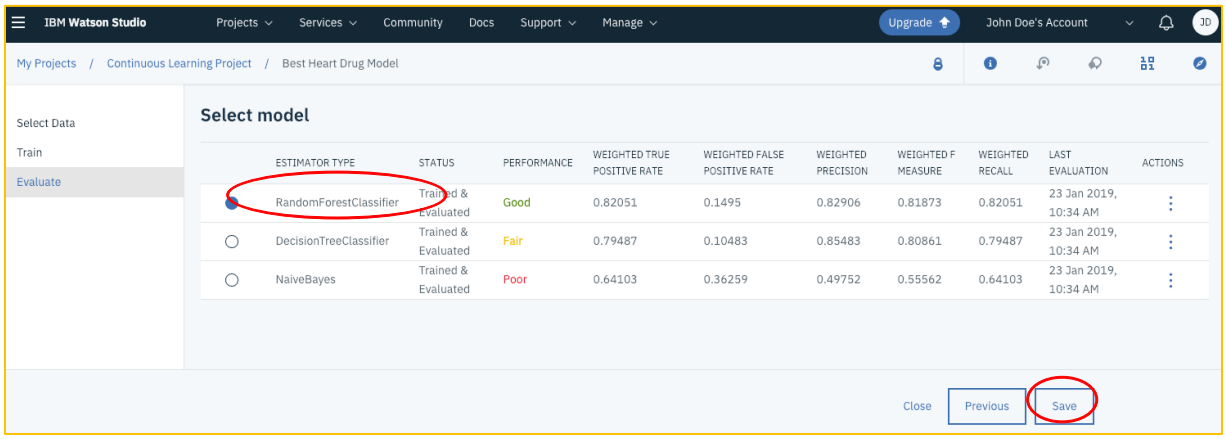


9. Back on the “Select a technique” page, click Next.

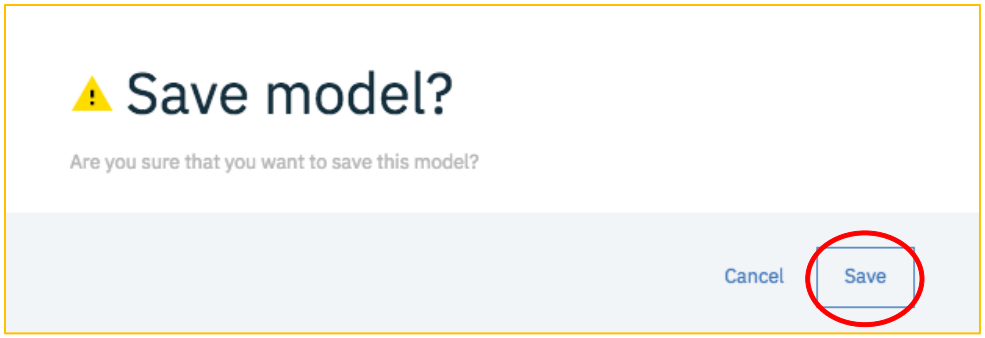


The Select model page will allow us to compare the results of different estimator types.

10. For this tutorial, select RandomForestClassifier.
11. Click Save.

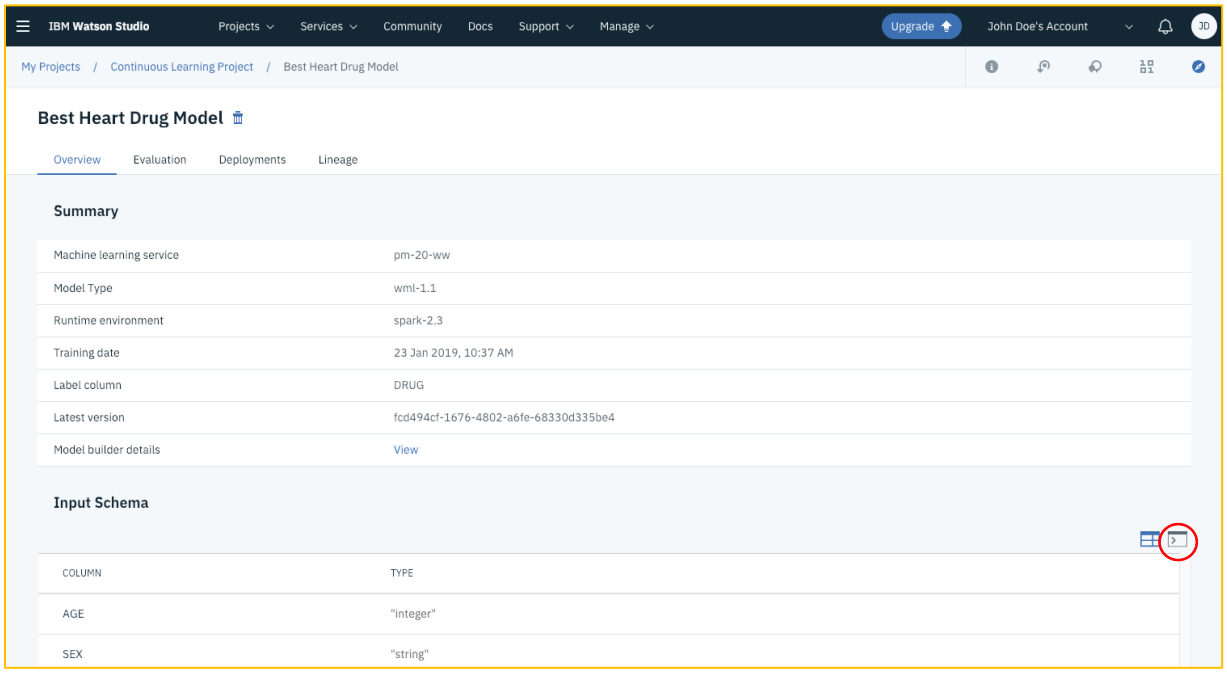


12. When the “Save model” window appears, click Save.

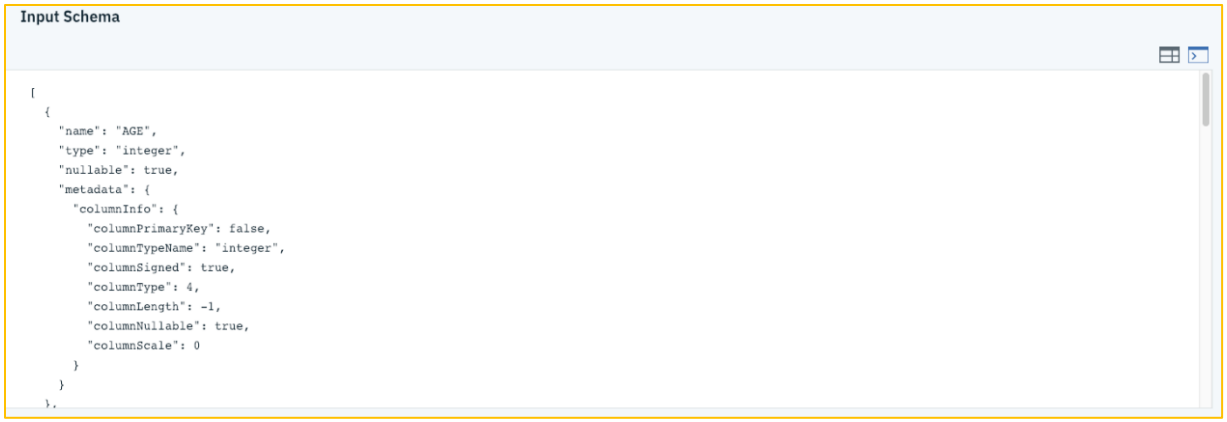


We have just saved the model in our Watson Machine Learning service and can now view information about the model specifications, details of construction, and input schema. The model can also be exposed as an API and used by data pipelines, applications, or other external systems.

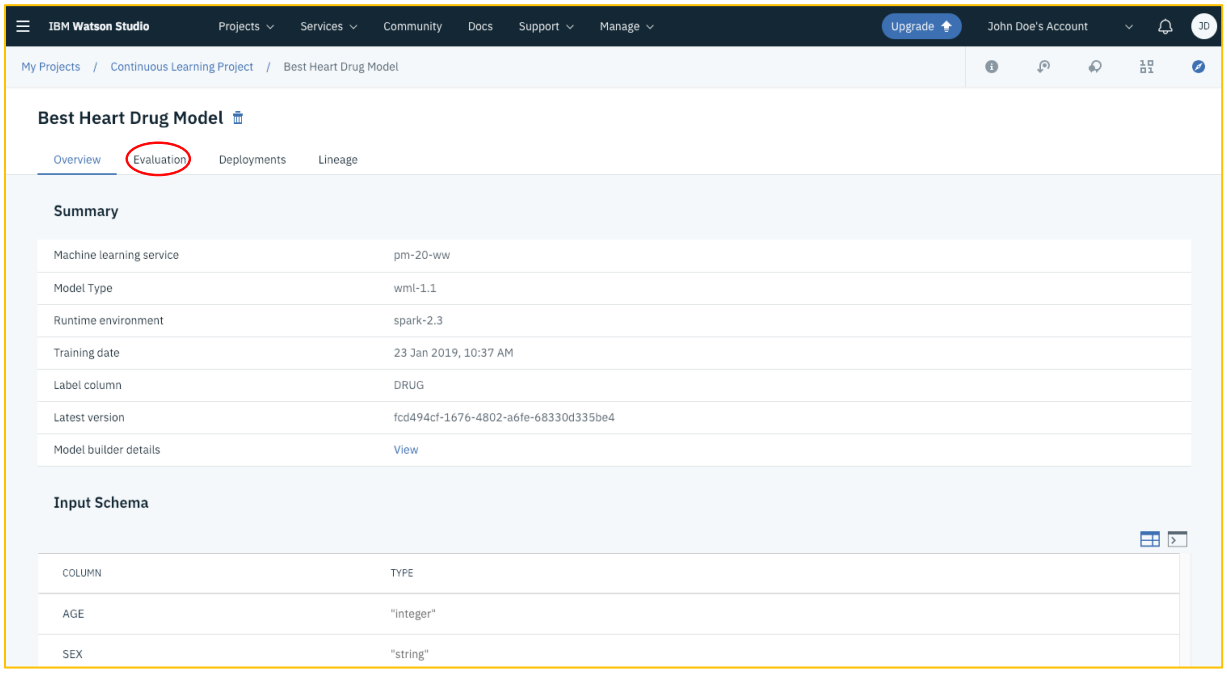
13. Click on the console button with an “angle bracket” to view the schema in JSON format.



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



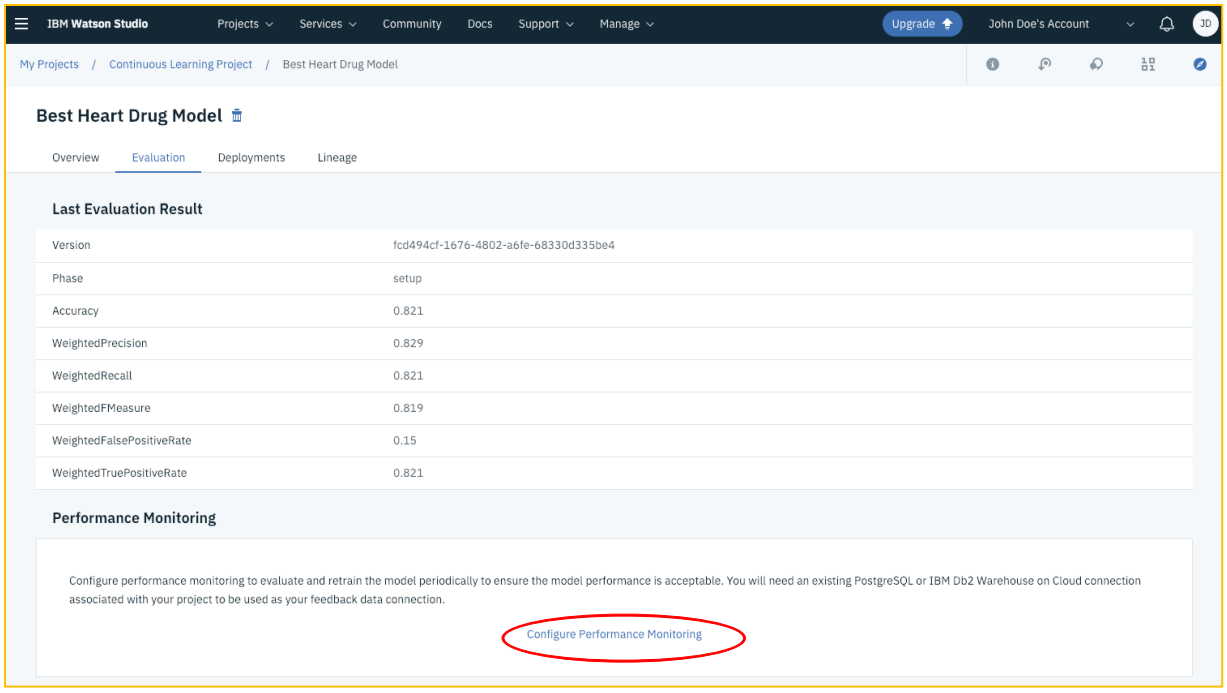
14. Scroll up and click on the Evaluation tab.



Step 3 Set up Continuous Learning Model Capabilities

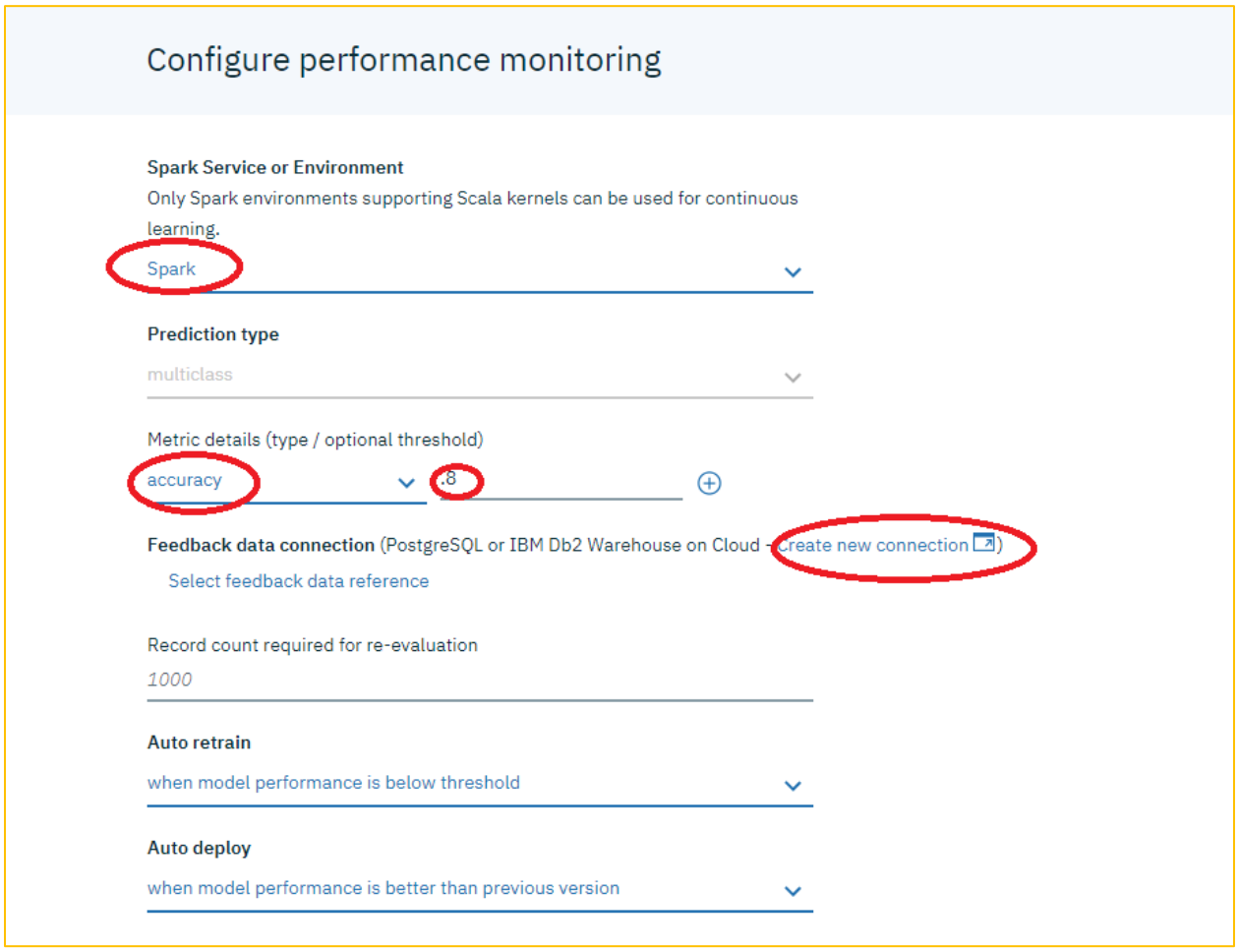
Model performance is critical for solving data science problems. This page provides information and functionality to aid in continuous training, tuning, and redeployment. We will now set up parameters to automatically retrain our model.

1. Click on Configure Performance Monitoring.

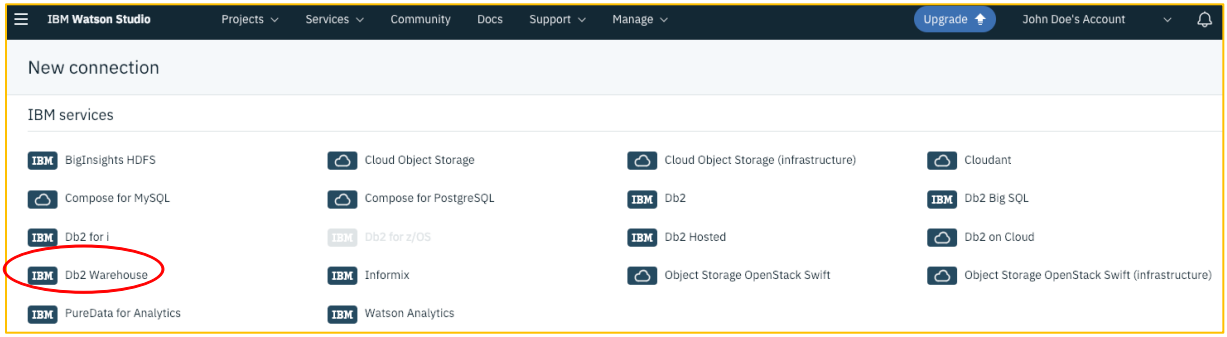


2. Under Spark Service or Environment, select the Spark service that was created in the prerequisites, not the Default Spark Scala 2.11 option.
3. Under Metric details, select Accuracy and put .8. This relies on a database table of new data. We will use DB2 Warehouse on Cloud, a data store optimized for analytic data sets.
4. Click on Create a new connection

.



5. On the “New connection” page, click Db2 Warehouse.

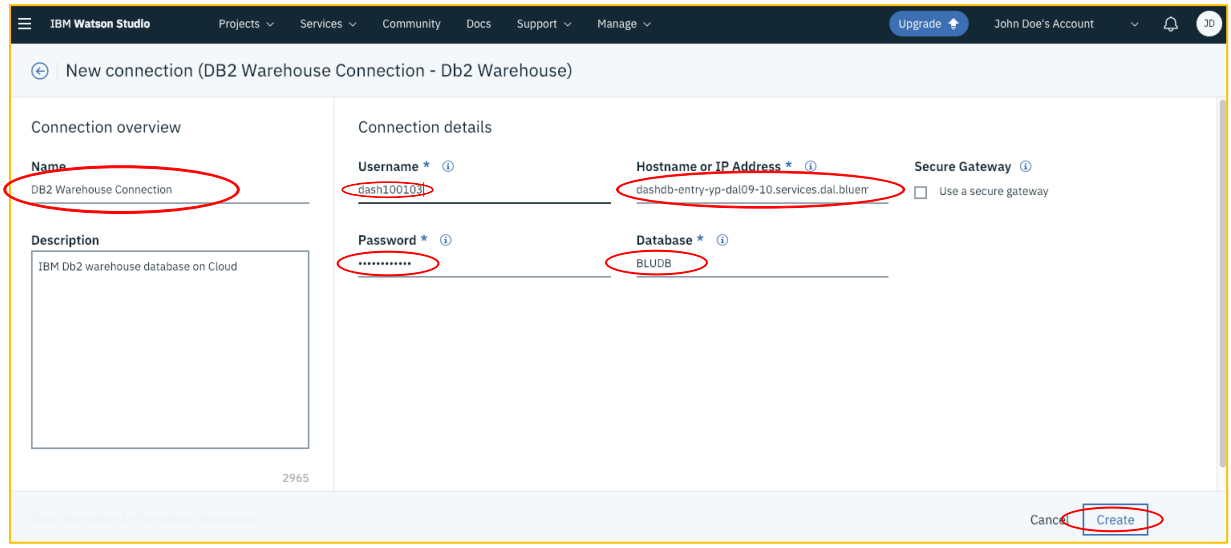


6. Click on Access DB2 Credentials

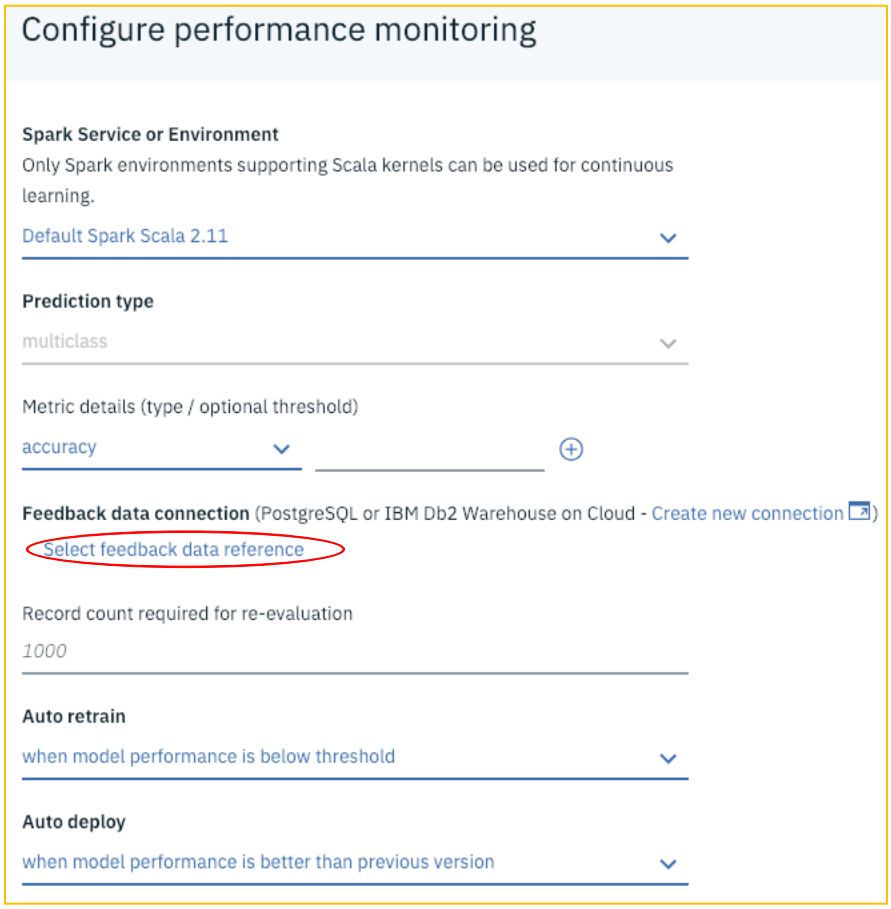
You should have an index card that provides your assigned database. If not, ask your instructor for your assigned database.

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

9. Click Create.

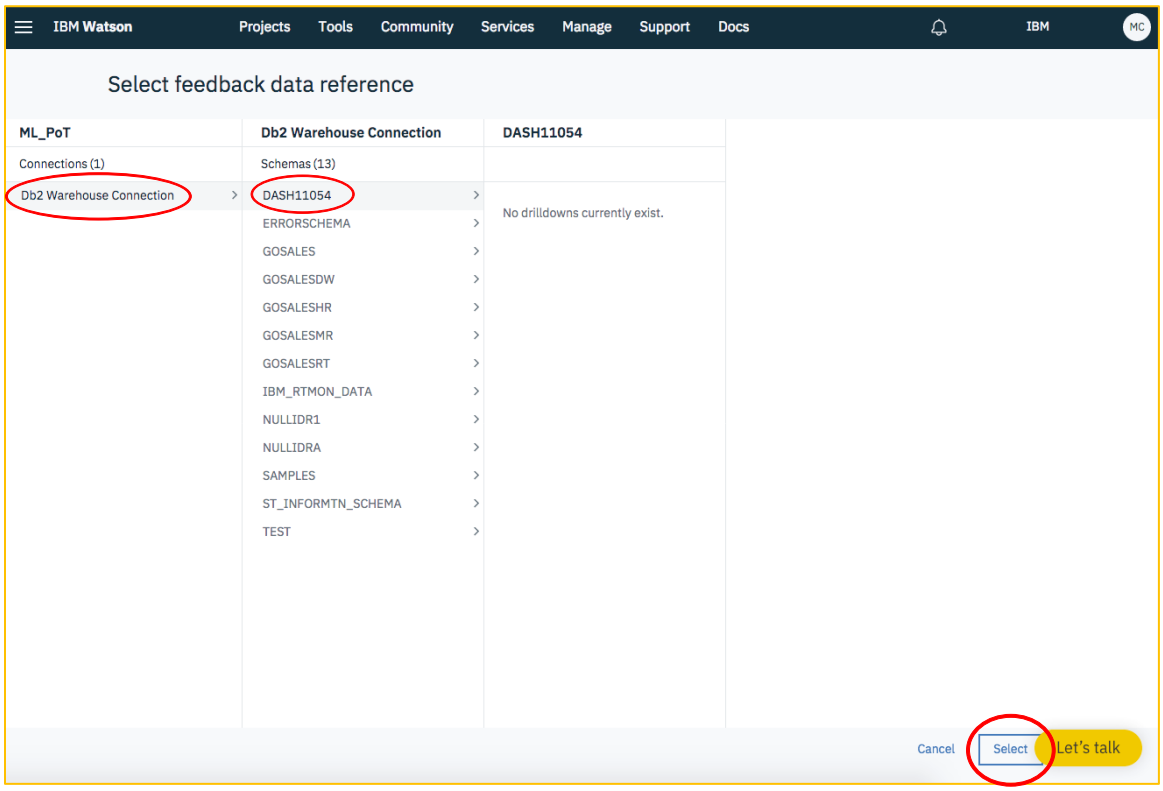


10. Return to the “Configure performance monitoring page” and click Select feedback data reference.

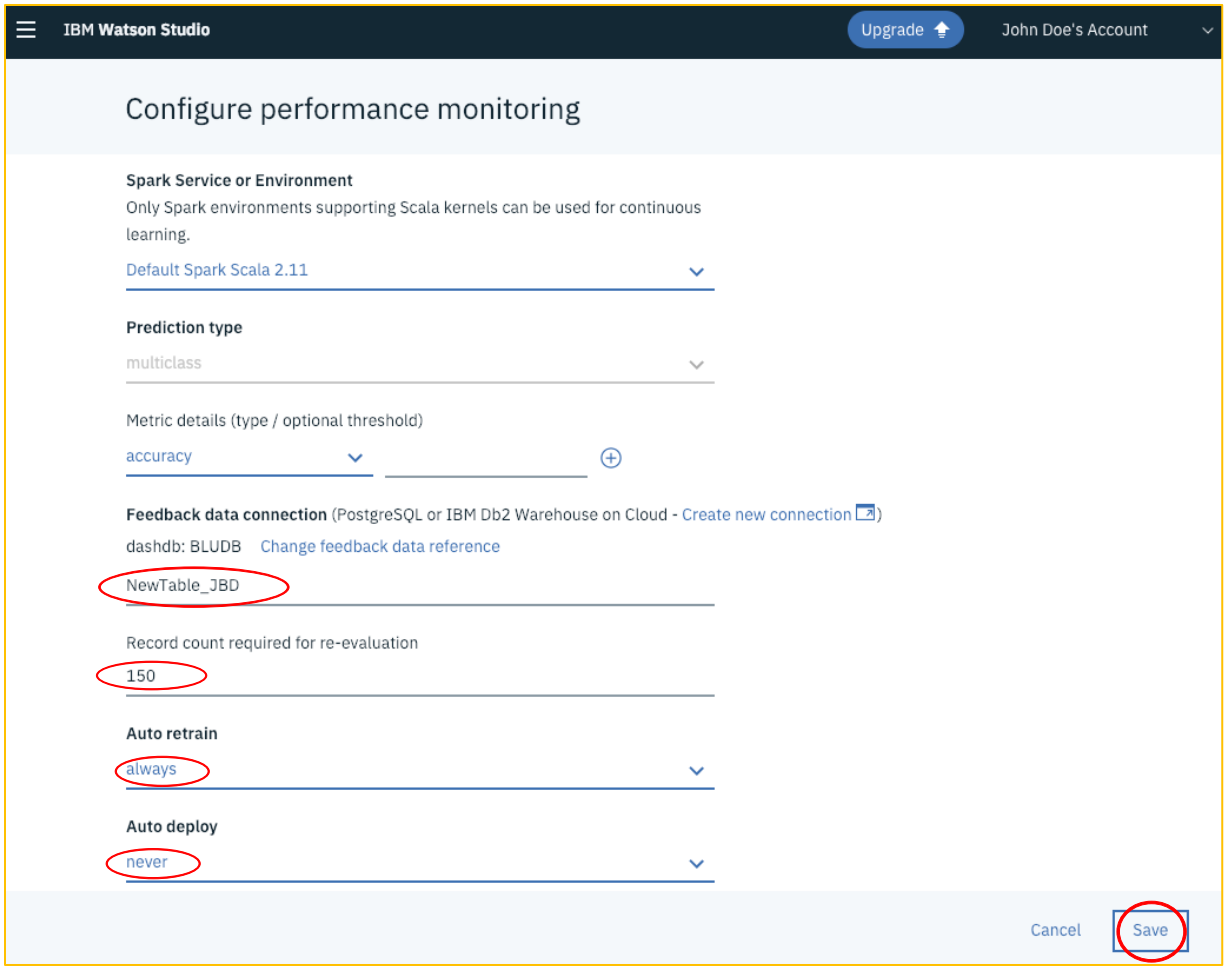


11. Click on your DB2 Warehouse-xx name.

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 150 as the Record count.
16. Under Auto retrain select always.
17. Under Auto deploy, Never.
18. Click Save.

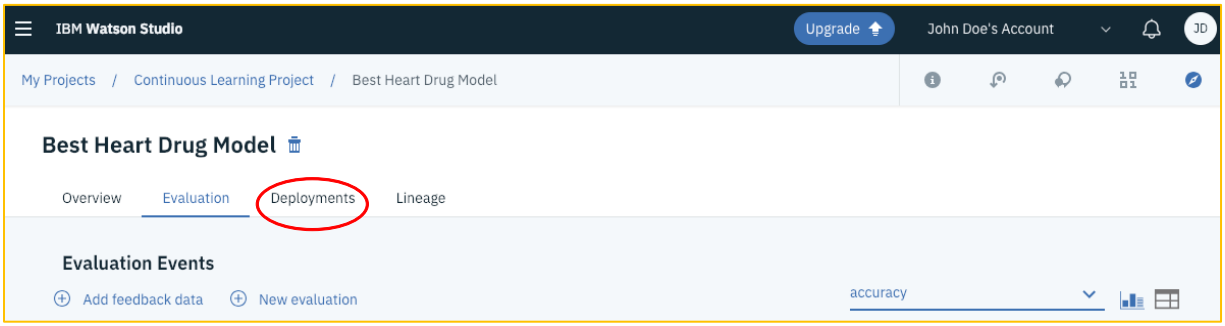


We have just set our model to retrain whenever new data is submitted. We set our model to never automatically redeploy as we would like to review each model before making it available to the medical staff who rely on it.

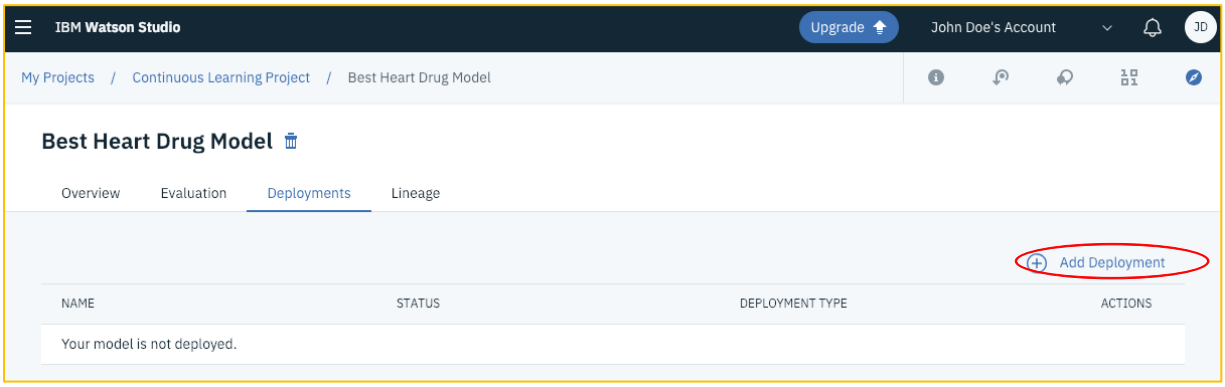
Step 4 Deploy a Machine Learning Model.

Now we will focus on deployment.

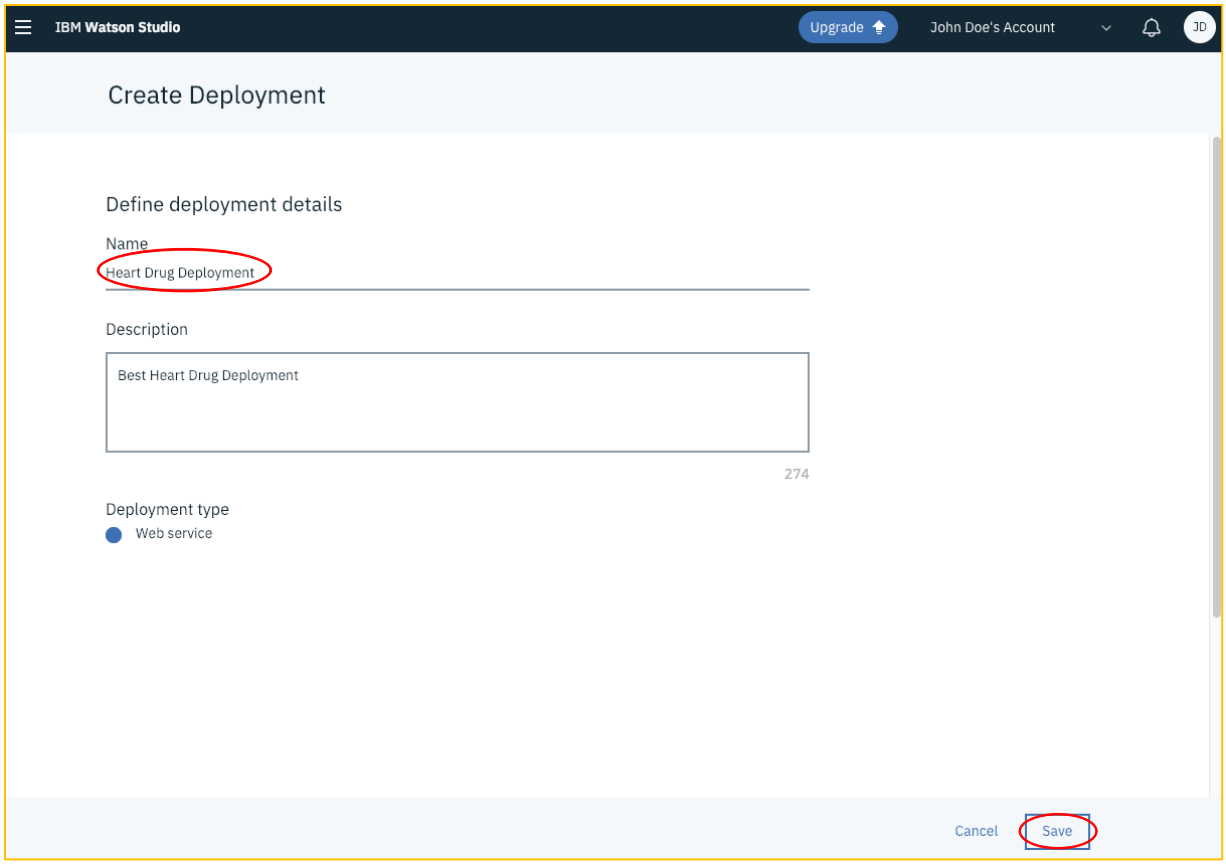
1. Click the Deployment tab.



2. Click Add Deployment.

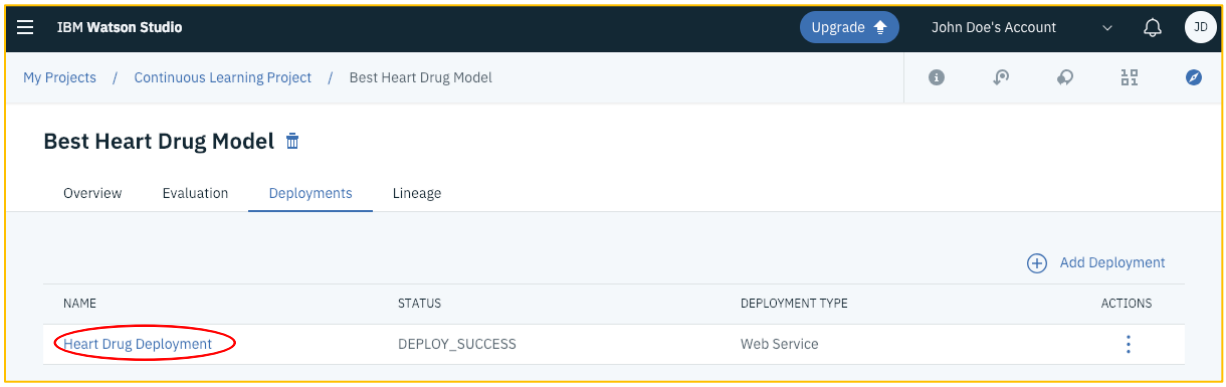


3. Enter a Name (e.g. Heart Drug Deployment).
4. Enter a Description.
5. Click Save.



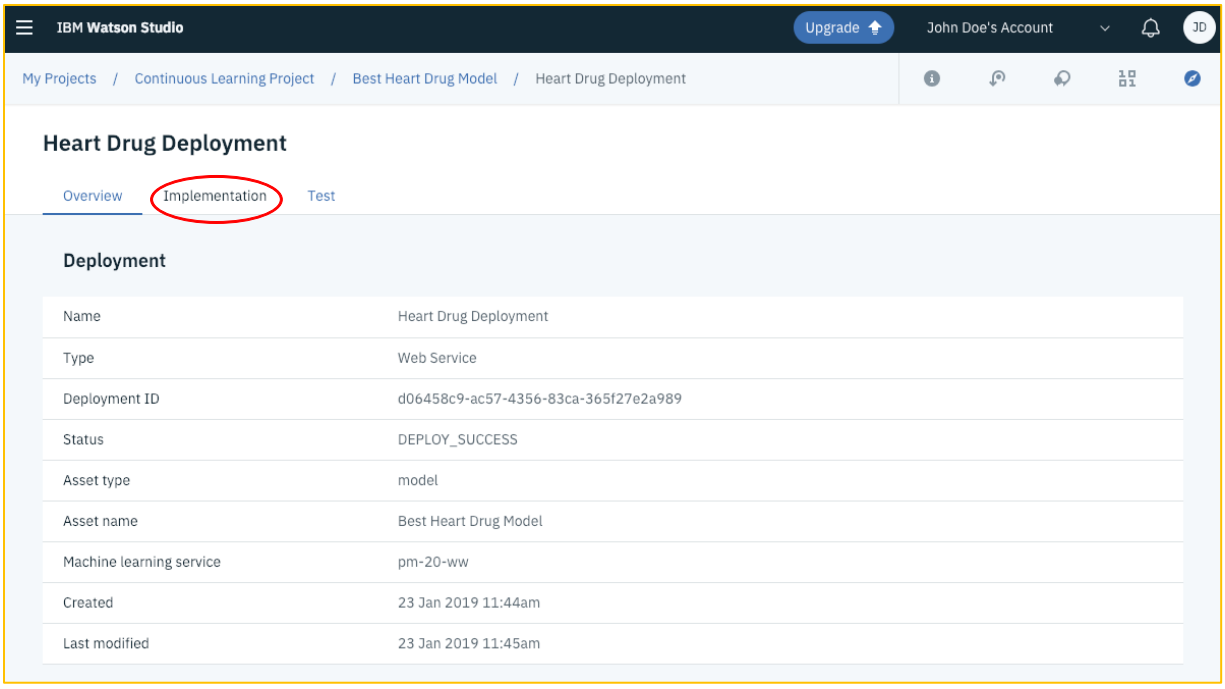
The model is now accessible by external systems.

6. Click on your deployment (e.g. Heart Drug Deployment).



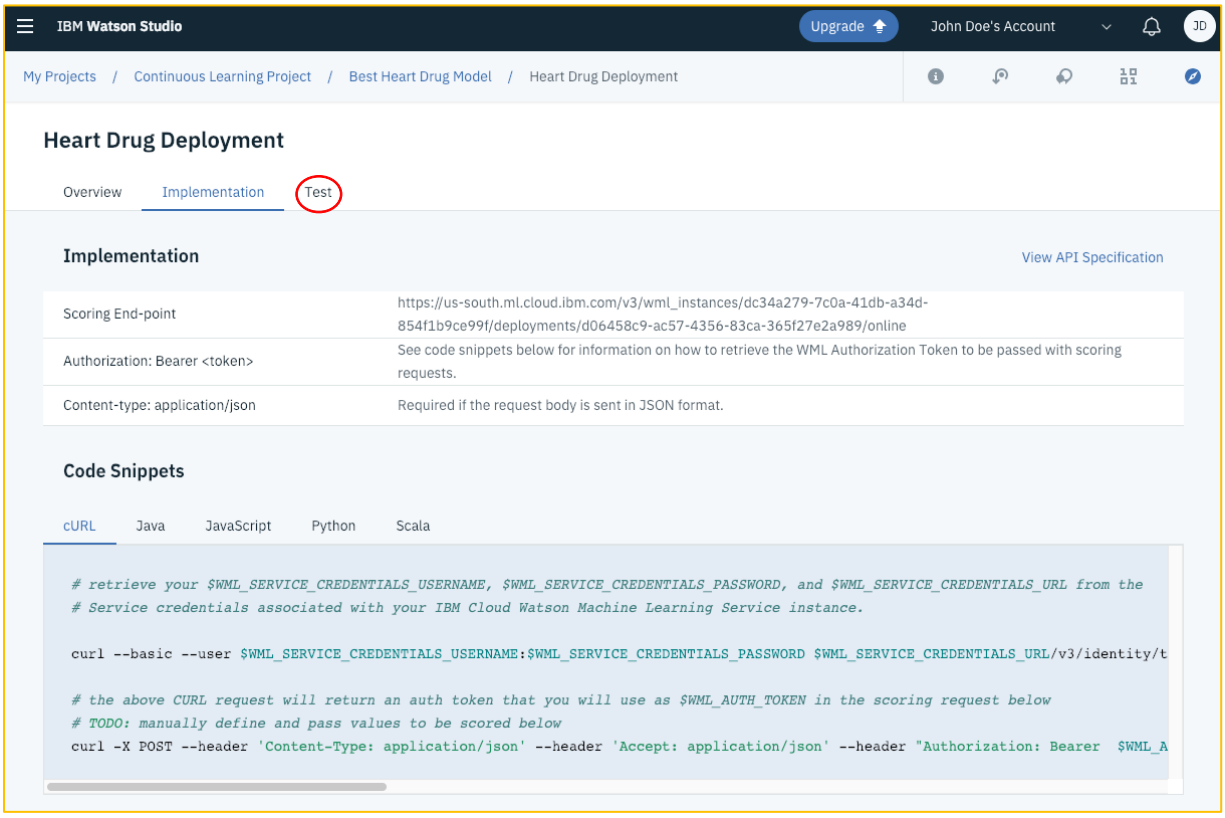
You are automatically brought to the Overview tab lists information related to the model deployment including services used and version number.

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:

   AGE: 62

   SEX: M

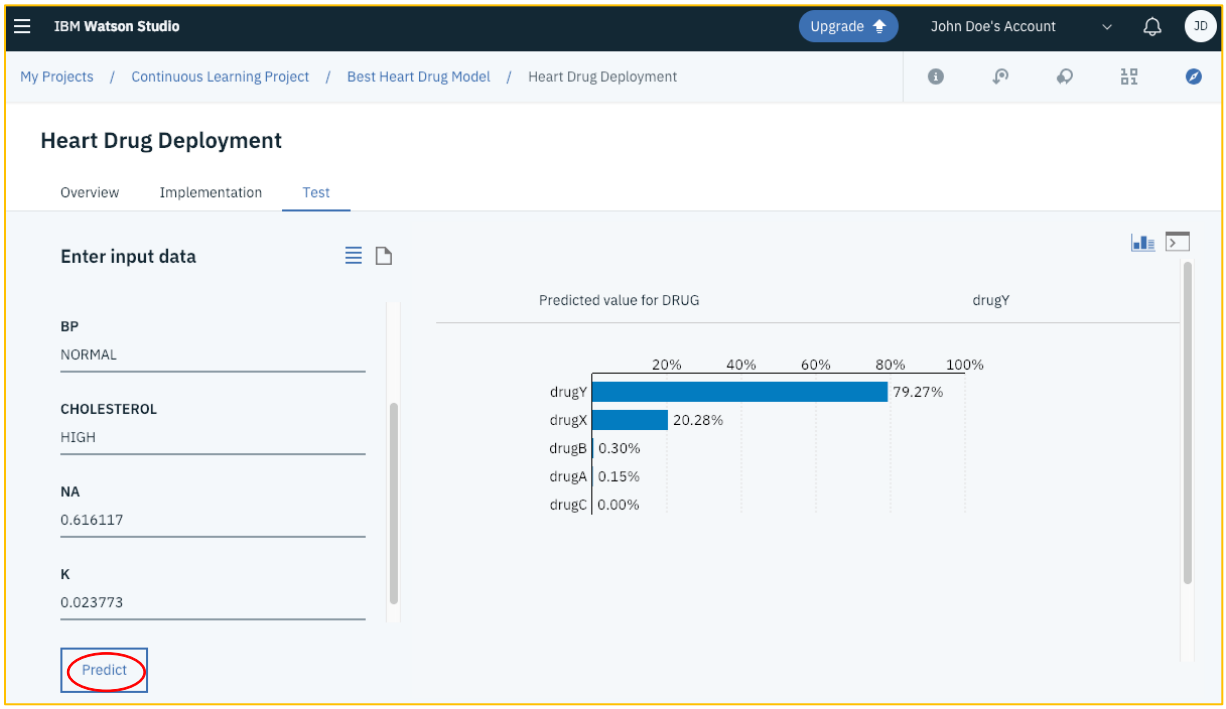
   BP: NORMAL

   CHOLESTEROL: HIGH

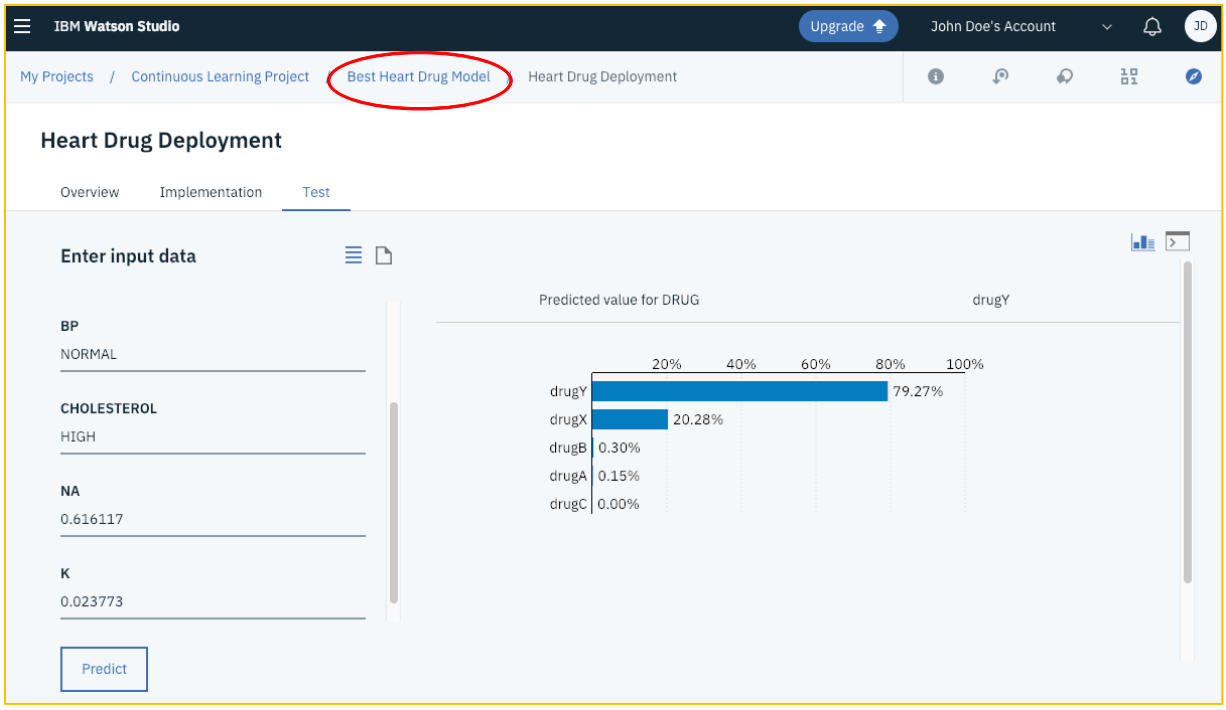
   NA: 0.616117

   K: 0.023773

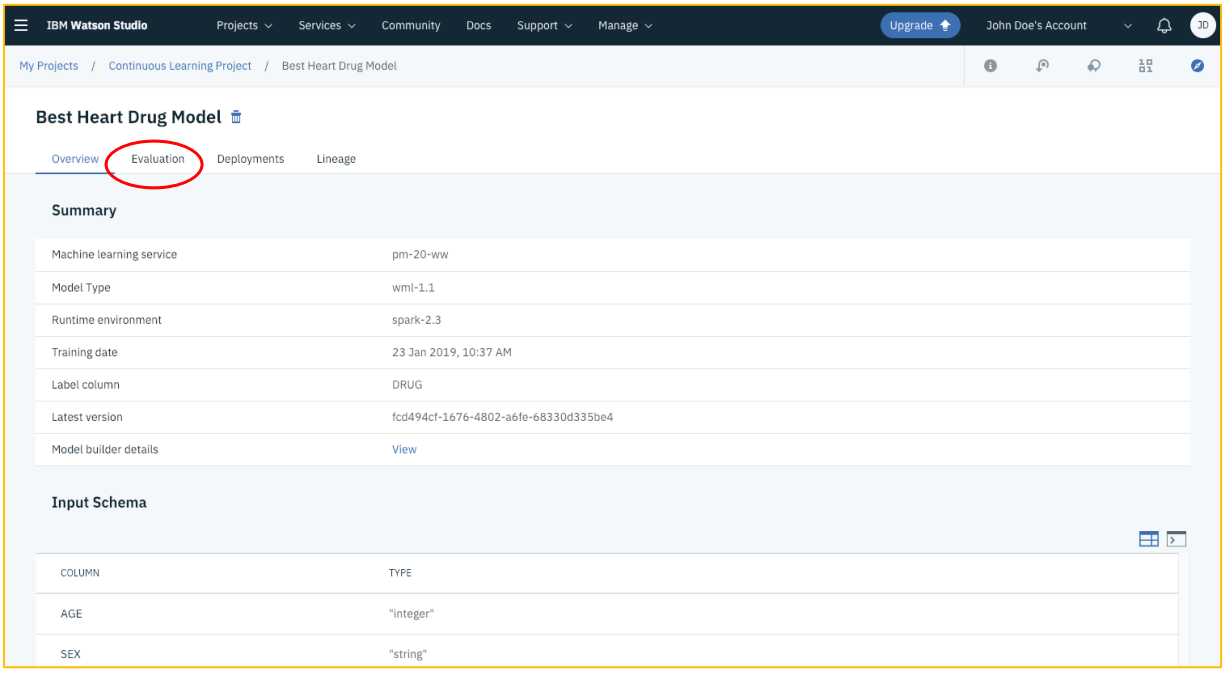
10. Click Predict.



11. Feel free to change the input data and run new predictions. When ready, click on your model name (e.g. Best Heart Drug Model)



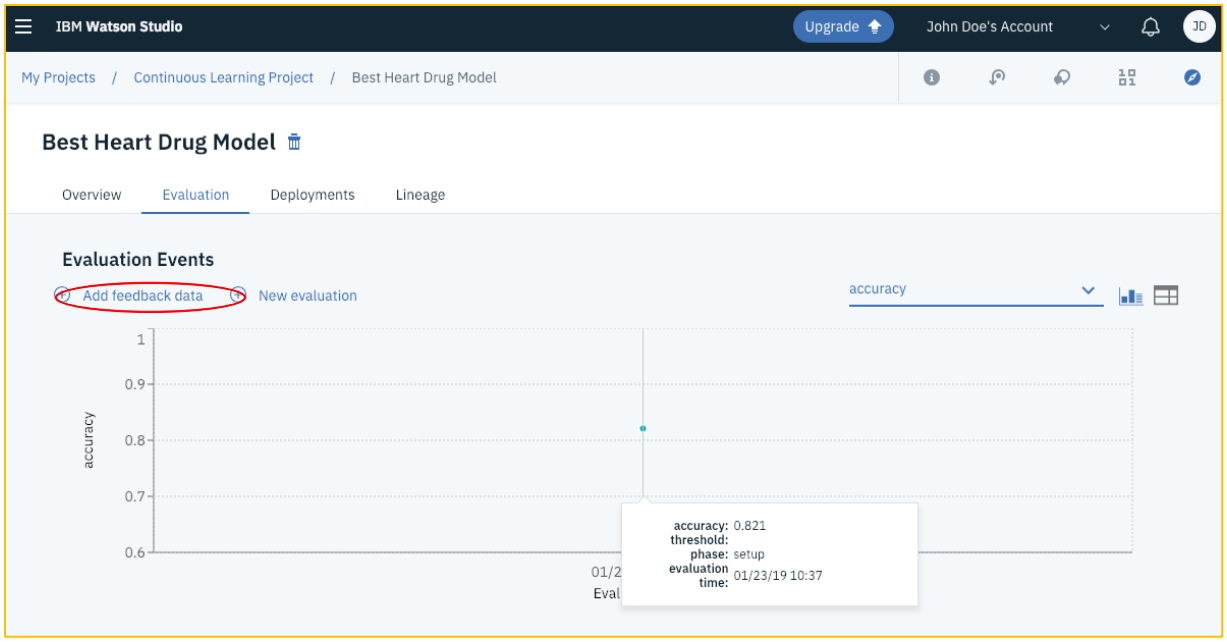
12. Click on the Evaluation tab.



13. If you scroll to the bottom you will be able to see which model version is deployed. Currently we only have one model.

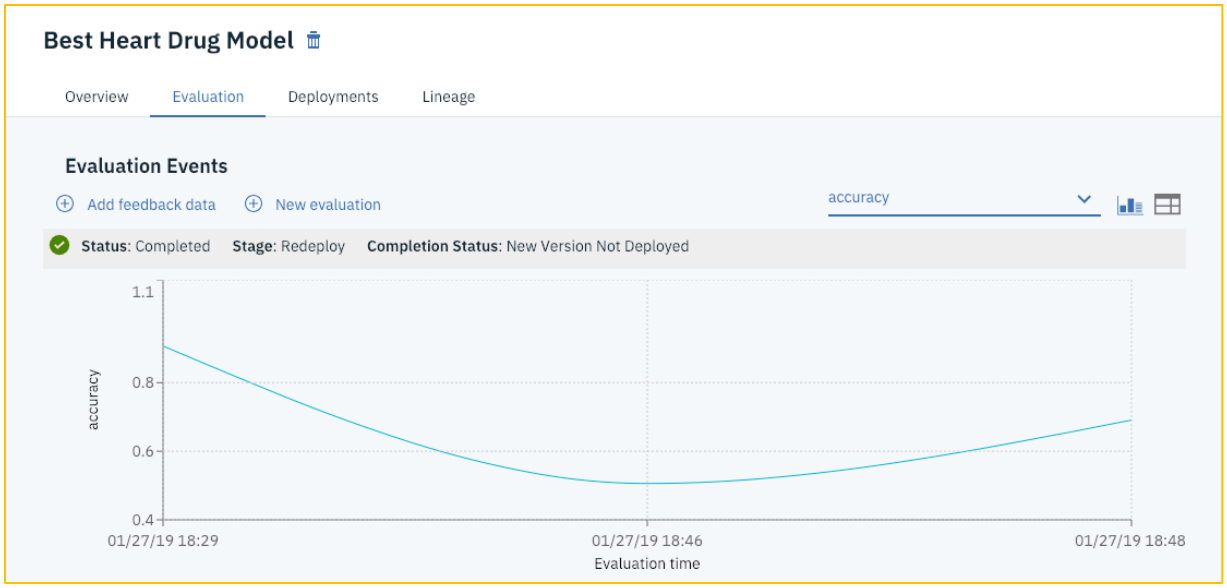
Suppose a day has passed and we have received more data to improve our model.

14. Scroll up and click on Add feedback data and select Drug_Data_Tuesday.csv.

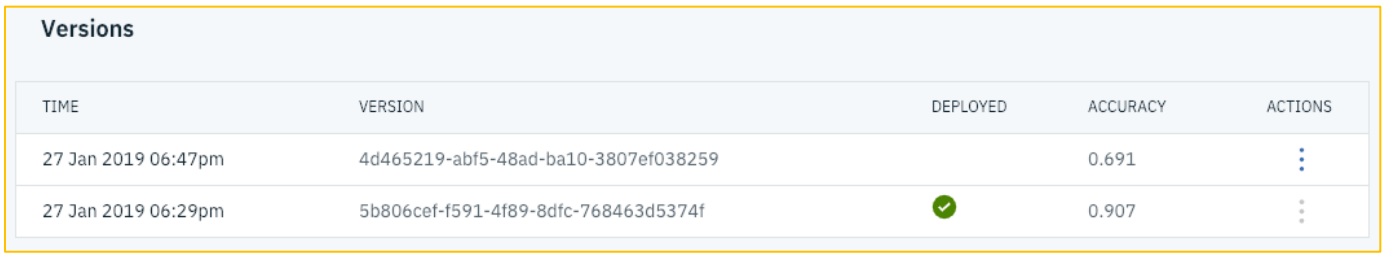


15. When the “New evaluation” window appears, click New Evaluation. Re-evaluating the model may take a few minutes. This is a great time to grab coffee.
16. Notice that additional data caused our original model’s performance to drop but retraining the model on the new unified dataset caused model performance to increase. Although we run these steps graphically, we can configure these steps to run easily through APIs.

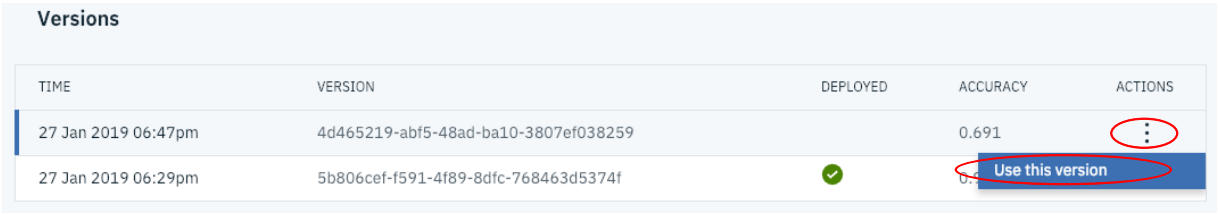
The continuous learning feature can be configured to automatically retrain or redeploy models under certain conditions (e.g. redeploy if accuracy falls below 0.9 or redeploy if the new model is better than the previous model). We currently have set our model to only allow manual redeployment. Our choice is now to either deploy this current model or re-run the model building process to build an entirely new model and choose the best algorithm to fit this new dataset.



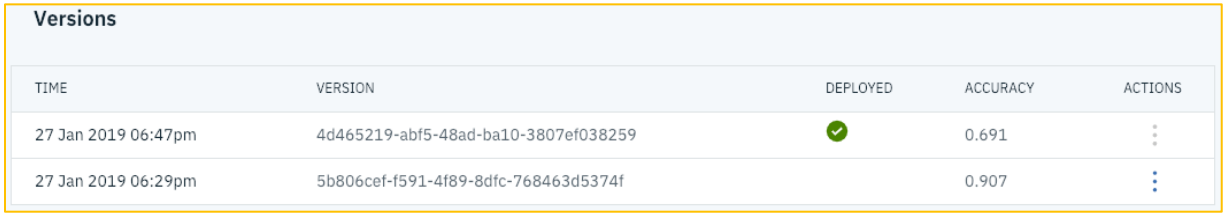
If you scroll down you will be able to see all model versions listed. We have two versions, but only our original model is deployed.



17. For the purpose of this lab, select the three dots under the Actions column and click Use this version to deploy the new model.



18. The new model version should now have a green checkmark under the Deployed column.



19. (Optional) scroll up and upload the final csv file, Drug_Data_Wednesday.csv, as feedback data.

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.

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.