

Lab: Introduction to Watson OpenScale

Introduction

Businesses today are increasingly certain that AI will be a driving force in the evolution of their industries over the next few years. Yet for every successful AI project, there are many that fail to reach widespread adoption in the business and achieve their expected outcomes. This is partly because the mechanics of AI deployment can be complex, and there are still gaps in skills and tooling that can make it difficult for data science, IT operations, and business teams to work in lockstep. But beyond the operational challenges, there are also much more profound issues of trust and transparency that businesses need to address before they can truly turn AI into a business advantage.

Knowledge workers must be able to trust AI and explain the decisions it helps make before they will incorporate it in their business processes. If AI is a black box that simply takes in data and produces obscure, unexplainable outcomes, then there is no way for the business to judge whether these systems are producing fair, accurate outcomes, or have confidence in AI's ability to augment decision-making. Equally, the business will not be able to explain outcomes to customers, auditors, or compliance teams.

IBM Watson OpenScale is an open platform that helps remove barriers to enterprise-scale AI. Watson OpenScale enables the enterprise to:

- Measure performance of production AI and its impact on business goals
- Track actionable metrics in a single console
- Explain AI outcomes
- Detect and mitigate harmful bias to improve outcomes
- Accept feedback to compute accuracy measures
- Accelerate the integration of AI into existing business applications.

Objectives

The goal of this lab is to familiarize the user with the features of Watson OpenScale. After completing this lab, you will understand how to:

1. Import a machine learning model
2. Deploy the model
3. Provision Watson OpenScale
4. Configure the payload logging database and Machine Learning provider
5. Score Data
6. Prepare Deployed Model for Monitoring
7. Configure Payload Logging
8. Configure Quality
9. Configure Fairness
10. Configure Drift
11. Submit Feedback and View Quality Metrics

12. Score Data and View Fairness Metrics
13. Explain a Transaction.

Lab Use Case

Traditional lenders are under pressure to expand their digital portfolio of financial services to a larger and more diverse audience, which requires a new approach to credit risk modeling. Their data science teams currently rely on standard modeling techniques - like decision trees and logistic regression - which work well for moderate datasets and make recommendations that can be easily explained. This satisfies regulatory requirements that credit lending decisions must be transparent and explainable.

To provide credit access to a wider and riskier population, applicant credit histories must expand beyond traditional credit, like mortgages and car loans, to alternate credit sources like utility and mobile phone plan payment histories, plus education and job titles. These new data sources offer promise, but also introduce risk by increasing the likelihood of unexpected correlations which introduce bias based on an applicant's age, gender, or other personal traits.

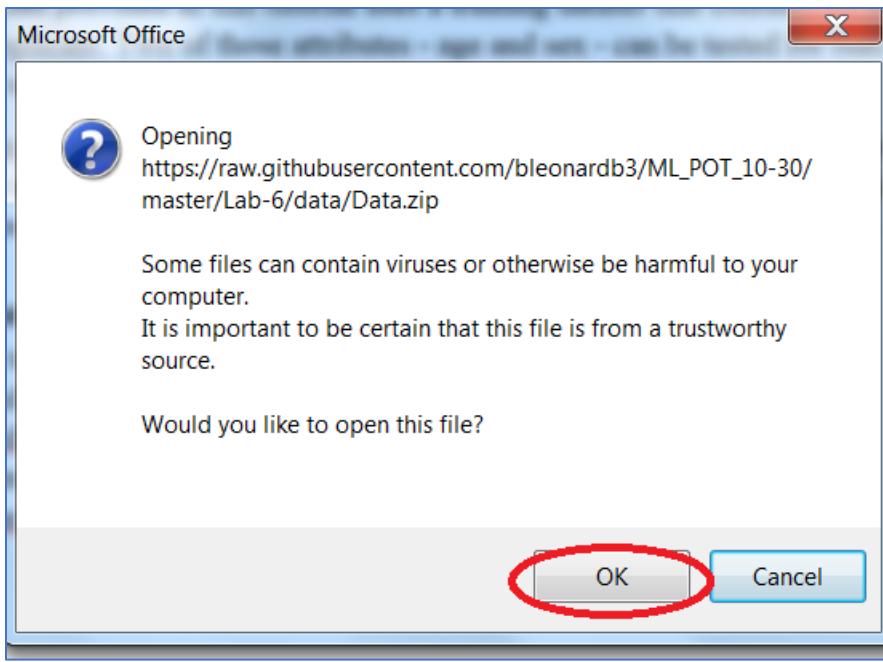
The data science techniques most suited to these diverse datasets, such as gradient boosted trees and neural networks, can generate highly accurate risk models, but at a cost. Such "black box" models generate opaque predictions that must somehow become transparent, to ensure regulatory approval such as Article 22 of the General Data Protection Regulation (GDPR), or the federal Fair Credit Reporting Act (FCRA) managed by the Consumer Financial Protection Bureau.

The credit risk model provided in this tutorial uses a training dataset that contains 20 attributes about each loan applicant. Two of those attributes - age and sex - can be tested for bias. For this tutorial, the focus will be on bias against sex and age.

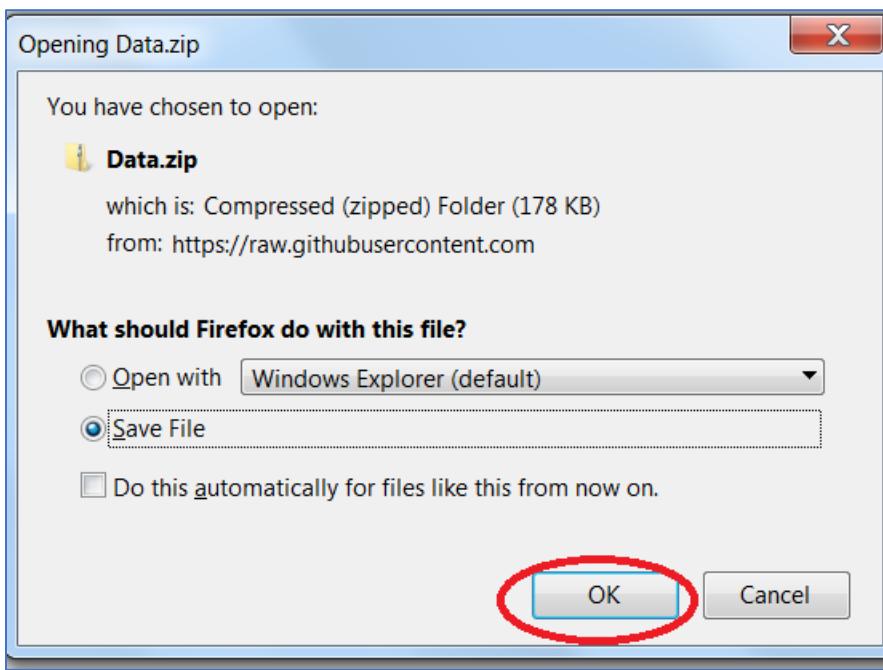
Watson OpenScale will monitor the deployed model's propensity for a favorable outcome ("No Risk") for one group (the Reference Group) over another (the Monitored Group). In this tutorial, the Monitored Group for sex is `female`, while the Monitored Group for age is `19 to 25`.

Download the Lab Files

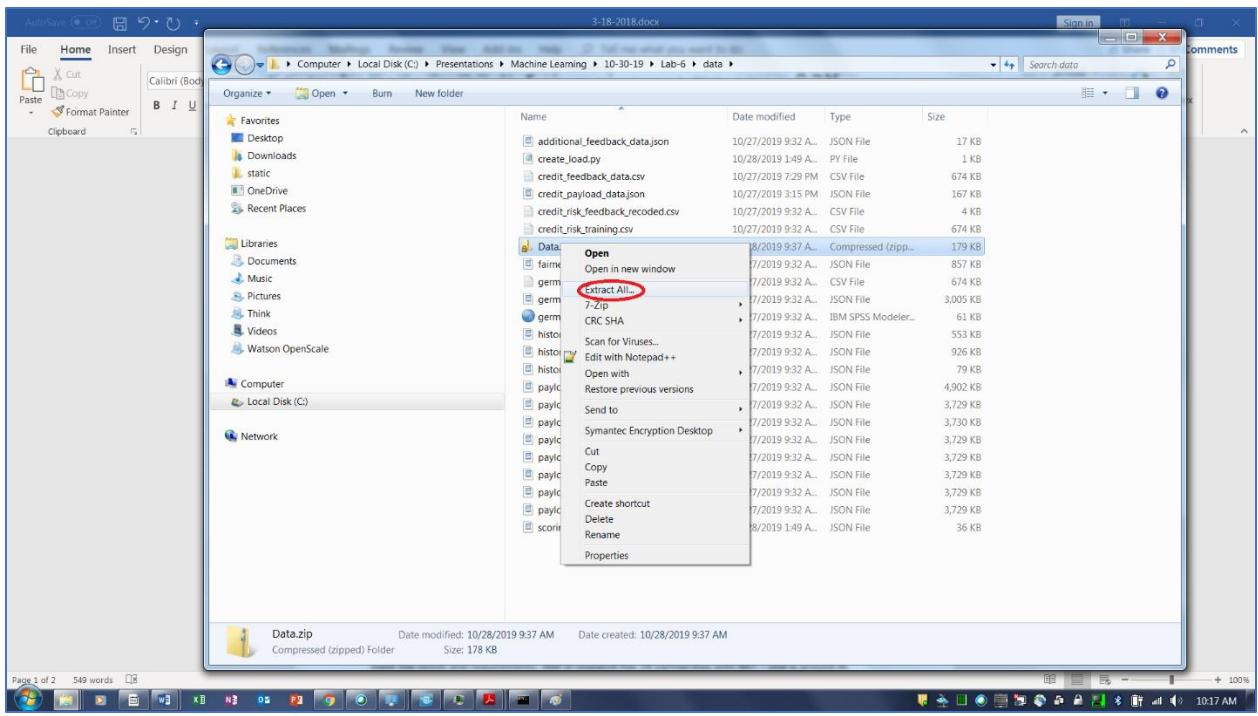
1. Click [here](#) to download the Data.zip file.
 1. `credit_feedback_data.csv`
 2. `credit_payload_data.json`
 3. `german_credit_data_biased_training.csv`
 4. `scoring.json`
2. Click **OK**. Note your browser may not prompt this message.



3. Click **OK**.



4. Navigate to the folder where the file is saved. Select the Data.zip file, right-click, and click **Extract All**.



Import the Credit Risk Model

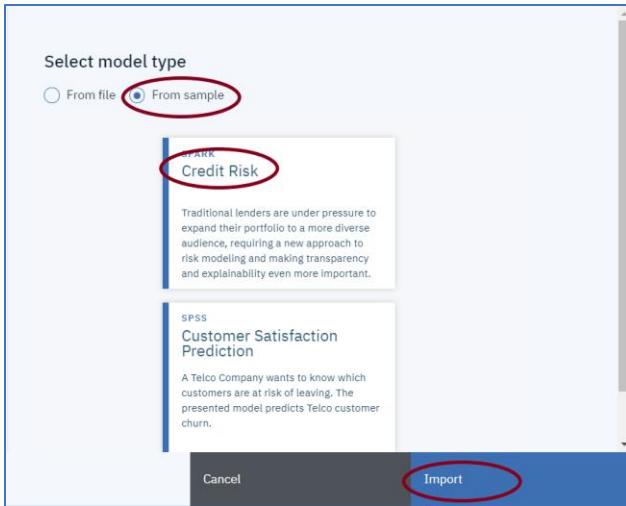
- From the Watson Studio project, click on **Add to project**.

The screenshot shows the IBM Watson Studio interface. The top navigation bar has 'IBM Watson Studio' and 'Upgrade' buttons. Below it, the 'My projects / Watson Studio Labs' section is visible. The main navigation bar includes 'Overview', 'Assets' (which is underlined and highlighted), 'Environments', 'Jobs', 'Deployments', 'Access Control', and 'Settings'. On the right side, there are icons for upload, user, launch IDE, and a red-circled 'Add to project' button.

- Click on **Watson Machine Learning**.

The screenshot shows a 'Choose asset type' dialog box. It lists various asset types: Data, Connection, Connected data, AutoAI experiment, Notebook, Dashboard, Visual Recognition ..., Natural Language Cl..., Watson Machine Learn..., Modeler flow, Data Refinery flow, Streams flow, and Decision Optimizatio... (with a 'NEW' badge). The 'Watson Machine Learn...' option is circled in red.

- From the **Import model** page, click on **From sample** for the **Select model type**, click on **Credit Risk**, and then click on **Import**.



Deploy the Credit Risk Model

1. From the **Model** page, click on the **Deployments** tab.

Machine learning service	WatsonMachineLearning
Model Type	mllib-2.3
Runtime environment	spark-2.3
Training date	27 Oct 2019, 1:11 PM
Label column	Risk
Latest version	fd202530-8cb7-48f1-a172-1575bee01ddd

2. Click on **Add Deployments**.

NAME	STATUS	DEPLOYMENT TYPE	ACTIONS
Your model is not deployed.			

3. From the **Create Deployment** page, type **credit-risk-deploy** for the **Name**, select **Web service** for the Deployment type, and click on **Save**.

Create Deployment

Define deployment details

Name: credit-risk-deploy (circled)

Description: Deployment description

Deployment type: Web service (circled)

Save (circled)

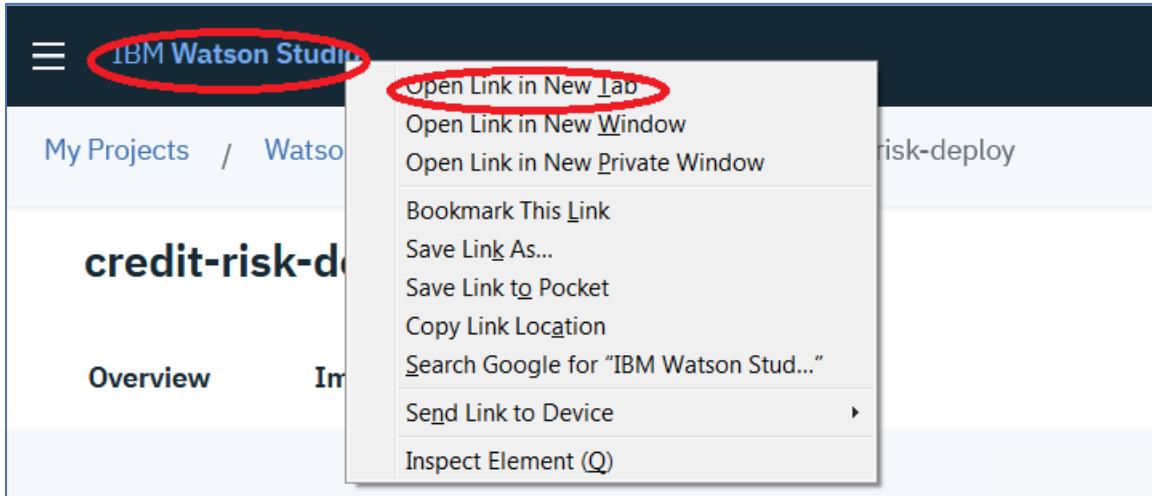
- The deployment status should go from **INITIALIZING** to **DEPLOY SUCCESS**. If the status doesn't change after a minute or so, refresh the browser.

MODEL
credit-risk

NAME	STATUS	DEPLOYMENT TYPE	ACTIONS
credit-risk-deploy	DEPLOY_SUCCESS (circled)	Web Service	

Provision Watson OpenScale.

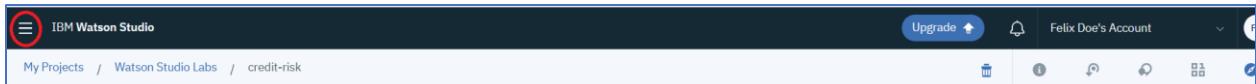
- Right-click on the **IBM Watson Studio** label and click on **Open link in New Tab**.



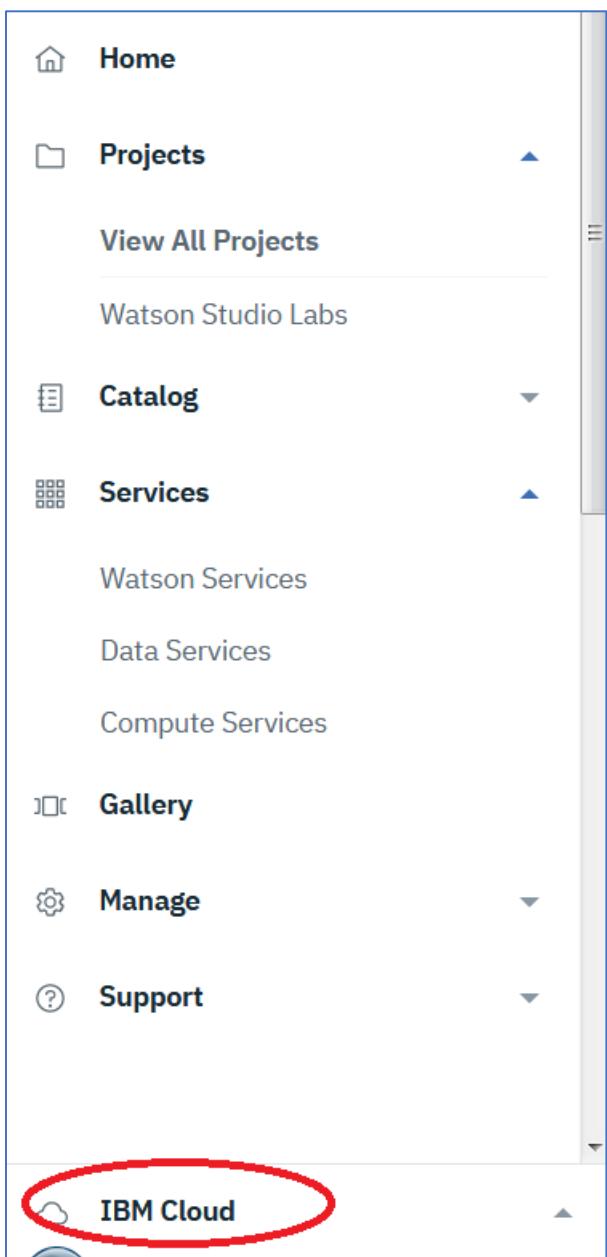
- Click on the new **Watson Studio** browser tab.



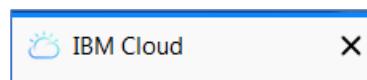
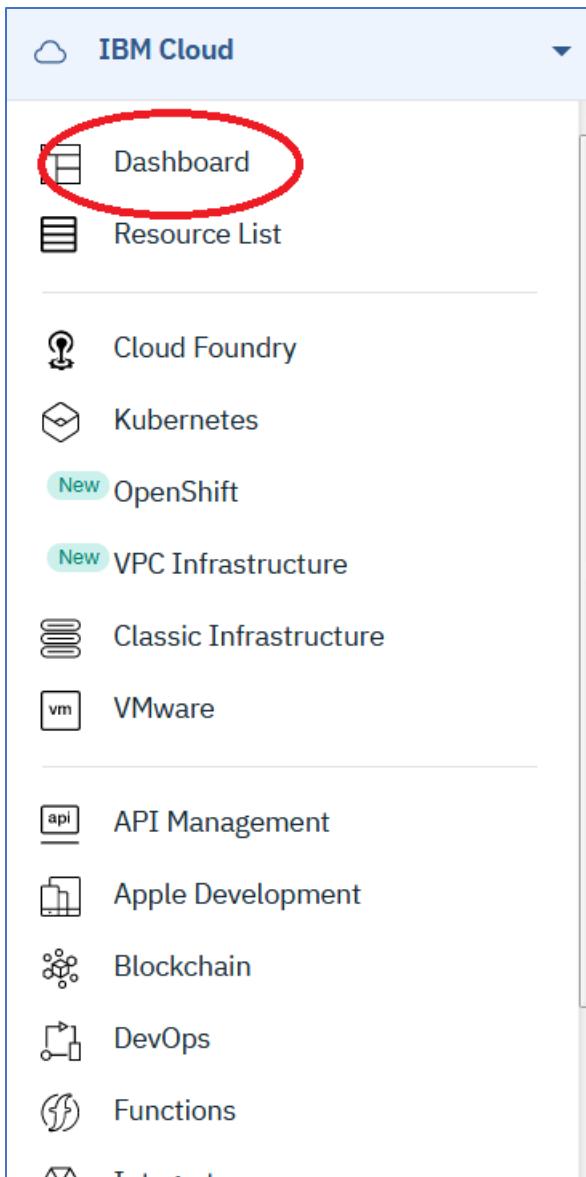
3. Click on the hamburger icon  in the top left corner.



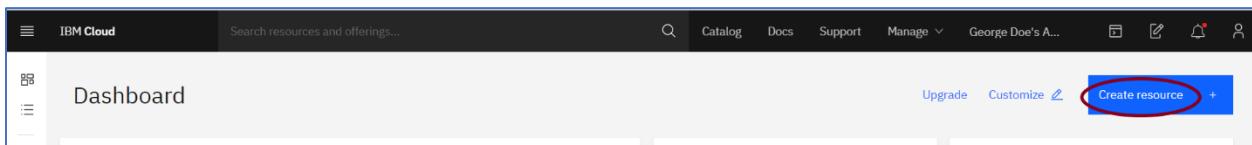
4. Click on **IBM Cloud**.



5. Click on **Dashboard**.



6. Note that the browser tab has been renamed to IBM Cloud.
7. Click on Catalog



8. Enter Watson OpenScale and hit the <Enter> key.

The screenshot shows the IBM Cloud Catalog interface. On the left, there's a sidebar with 'Catalog' and 'Featured' sections, which includes 'Services' and 'Software'. The main area is titled 'IBM Cloud products' and displays a message: 'Over 190+ products available for you to customize and build the solutions that you need for your business'. A search bar at the top has 'Watson OpenScale' typed into it, with the entire input field circled in red.

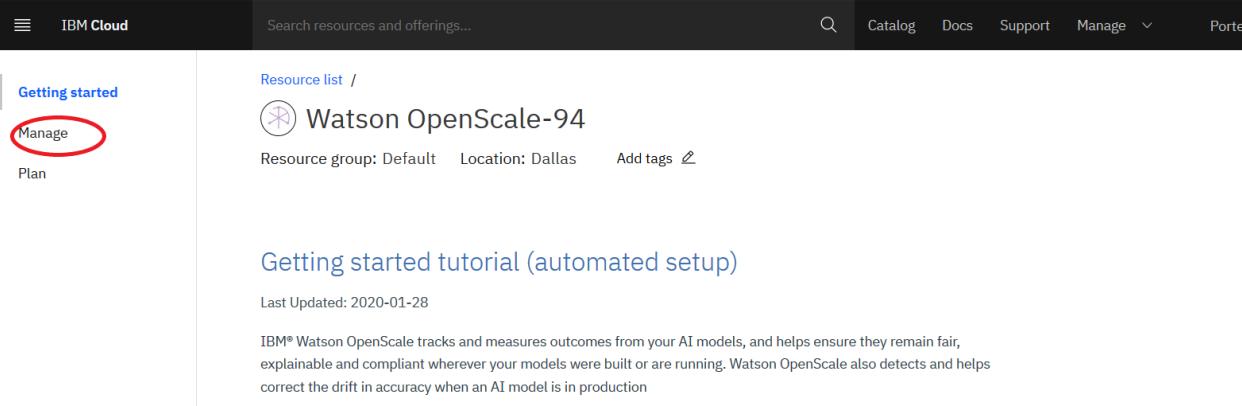
9. Click on Watson OpenScale.

This screenshot shows the search results for 'Watson OpenScale'. It displays one result: 'Watson OpenScale' with a red circle around its name. Below the title, it says 'IBM Watson OpenScale is an enterprise-grade environment for AI infused applications that provides enterprises...'. At the bottom, there are links for 'Services • Lite • Free • IAM-enabled'.

10. Click on Create.

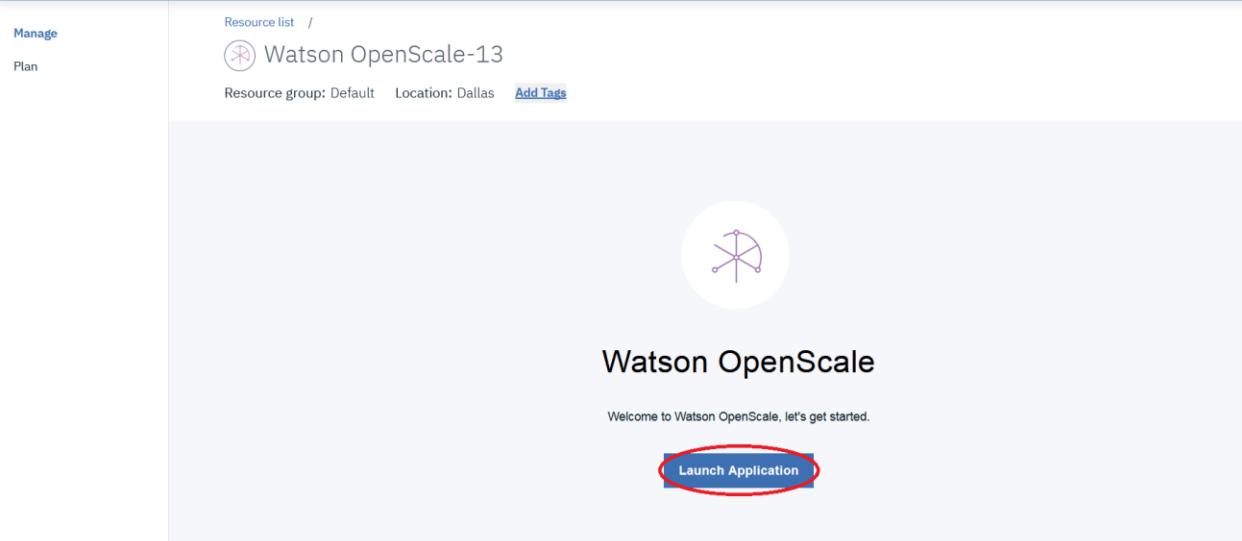
This screenshot shows the 'Watson OpenScale' service creation page. At the top, there are tabs for 'Create' (which is selected) and 'About'. Below that, there's a section to 'Select a region' with 'Dallas' chosen. Under 'Select a pricing plan', it says 'Lite' and 'Free'. To the right, there's a 'Summary' section with details like 'Region: Dallas', 'Plan: Lite', and 'Service name: Watson OpenScale-13'. The 'Create' button is highlighted with a red circle. At the bottom right, there's a 'View terms' link.

11. Click on **Manage**.



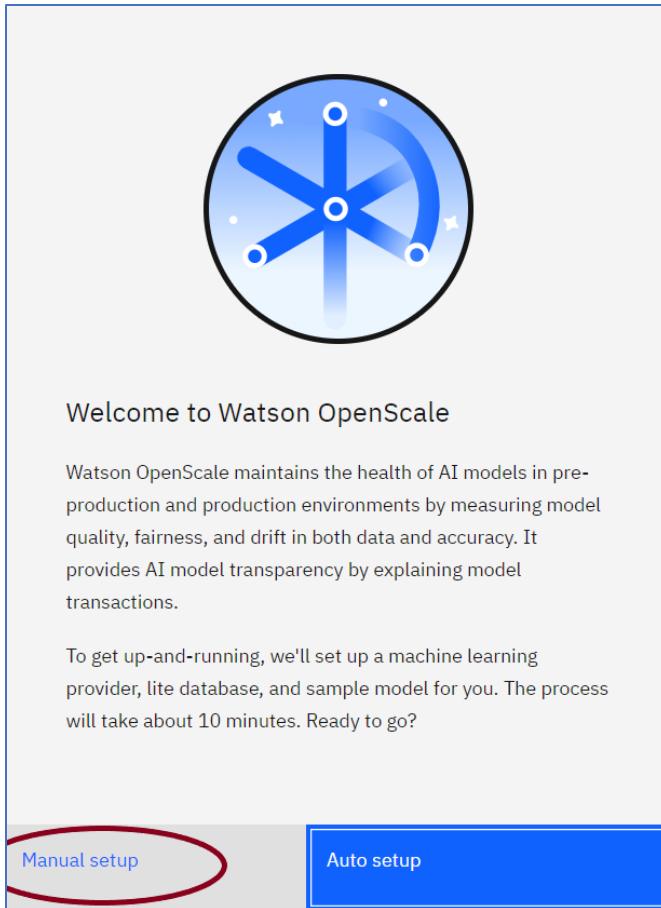
The screenshot shows the IBM Cloud interface. On the left, there's a sidebar with 'Getting started' and 'Manage' (which is circled in red). The main content area shows a resource named 'Watson OpenScale-94'. It includes details like 'Resource group: Default', 'Location: Dallas', and 'Add tags'. Below this, there's a 'Getting started tutorial (automated setup)' section with a last update date of '2020-01-28' and a brief description of what Watson OpenScale does.

12. Click on **Launch Application**.



The screenshot shows the Watson OpenScale application launch screen. On the left, there's a sidebar with 'Manage' (circled in red) and 'Plan'. The main content area shows a resource named 'Watson OpenScale-13'. It includes details like 'Resource group: Default', 'Location: Dallas', and 'Add Tags'. Below this, there's a large circular icon with a gear-like symbol. To its right, the text 'Watson OpenScale' is displayed. Underneath, it says 'Welcome to Watson OpenScale, let's get started.' At the bottom, there's a blue button labeled 'Launch Application' (which is circled in red).

13. Make sure to click on **Manual setup**.



Setup System

1. In the system setup, we need to set up a database to collect the payload logging data. We also need to specify which deployed model will be monitored.
2. Click on **Free Lite plan database** for **Database type** and click on **Save**.

A screenshot of the "System setup" dialog box. On the left, there's a sidebar with "Database" selected. The main area has a "Database" section with a "Required" label and a "Description" paragraph. A note at the bottom states: "Note: The free Lite plan database is not GDPR compliant. If your model processes personally identifiable information (PII), you must purchase a new database or use an existing database that does conform to GDPR rules. [Learn more](#)." To the right is a "Database type" dropdown menu with options: "Choose an option" (highlighted with a red oval), "Free lite plan database" (also highlighted with a red oval), "Compose for PostgreSQL", "Databases for PostgreSQL", and "Db2".

System setup	
<p>Database</p> <p>Machine learning providers</p>	<p>Required</p> <p>Database</p> <p>Description</p> <p>Select a database to store your model transactions and model evaluation results. A free database is available for Lite plan users to get started. Alternatively, you can use an existing PostgreSQL or Db2 database (lite Db2 plans not supported). You can also purchase a new database.</p> <p><i>Note: The free Lite plan database is not GDPR compliant. If your model processes personally identifiable information (PII), you must purchase a new database or use an existing database that does conform to GDPR rules. Learn more.</i></p>
	<p>Database type</p> <p>Choose an option</p> <p>Free lite plan database</p> <p>Compose for PostgreSQL</p> <p>Databases for PostgreSQL</p> <p>Db2</p>

3. Click **Save**.

System setup

<p>Database</p> <p>Machine learning providers</p>	<p>Required</p> <p>Database</p> <p>Description</p> <p>Select a database to store your model transactions and model evaluation results. A free database is available for Lite plan users to get started. Alternatively, you can use an existing PostgreSQL or Db2 database (lite Db2 plans not supported). You can also purchase a new database.</p> <p>Note: The free Lite plan database is not GDPR compliant. If your model processes personally identifiable information (PII), you must purchase a new database or use an existing database that does conform to GDPR rules. Learn more.</p>	<p>Database type</p> <p>Free lite plan database</p> <p>Purchase a database</p> <p>Cancel Save</p>
---	---	---

4. Click on **Machine learning providers**.

System setup

<p>Database</p> <p>Machine learning providers</p>	<p>Required</p> <p>Database</p> <p>Description</p> <p>The Watson OpenScale database stores your model transactions and model evaluation results.</p> <p>Database</p> <p>Database type</p> <p>Free lite plan database</p> <p>Database</p> <p>Internal database</p> <p>Schema</p> <p>public</p>
---	--

5. Click on **Add machine learning provider**.

System setup

Prepare Watson OpenScale for use by connecting it to a database, a machine learning provider, and integrated services.

<p>Database</p> <p>Machine learning providers</p> <p>Integrations (optional)</p>	<p>Required</p> <p>Machine learning providers</p> <p>Description</p> <p>Watson OpenScale connects to deployed models stored in a machine learning environment.</p> <p>Add machine learning provider +</p>
--	--

6. Click on pencil icon to edit the connection.

System setup

Prepare Watson OpenScale for use by connecting it to a database, a machine learning provider, and integrated services.

Database Machine learning providers Integrations (optional) [beta](#)

[Back to all providers](#)

New provider

Description [e](#)
Click edit to enter provider description.

Connection

Click edit to enter the connection information. [e](#)

7. Click on **Watson Machine Learning** for the **Service provider type**.

System setup

Prepare Watson OpenScale for use by connecting it to a database, a machine learning provider, and integrated services.

Database Machine learning providers Integrations (optional) [beta](#)

Machine learning providers

Connection

Description
Connect to the provider where your deployed models are stored and specify if the environment is a pre-production or production environment.

Pre-production environments
Test pre-production models by uploading test data sets (csv files) and running evaluations. When the model is ready, approve it for production.

Production environments
Monitor production models by logging model transactions and sending feedback (labeled test data) to Watson OpenScale for continuous evaluation.

Service provider type

Choose an option

Watson Machine Learning [e](#)

Custom Environment

Amazon SageMaker

Microsoft Azure ML Studio

Microsoft Azure ML Service

8. Click on **WatsonMachineLearning** for the **Watson Machine Learning service**. Click on **Production** for the **Environment type**. Click on **Save**.

System setup

Prepare Watson OpenScale for use by connecting it to a database, a machine learning provider, and integrated services.

Database Machine learning providers Integrations (optional) beta

Connection

Description
Connect to the provider where your deployed models are stored and specify if the environment is a pre-production or production environment.

Pre-production environments
Test pre-production models by uploading test data sets (csv files) and running evaluations. When the model is ready, approve it for production.

Production environments
Monitor production models by logging model transactions and sending feedback (labeled test data) to Watson OpenScale for continuous evaluation.

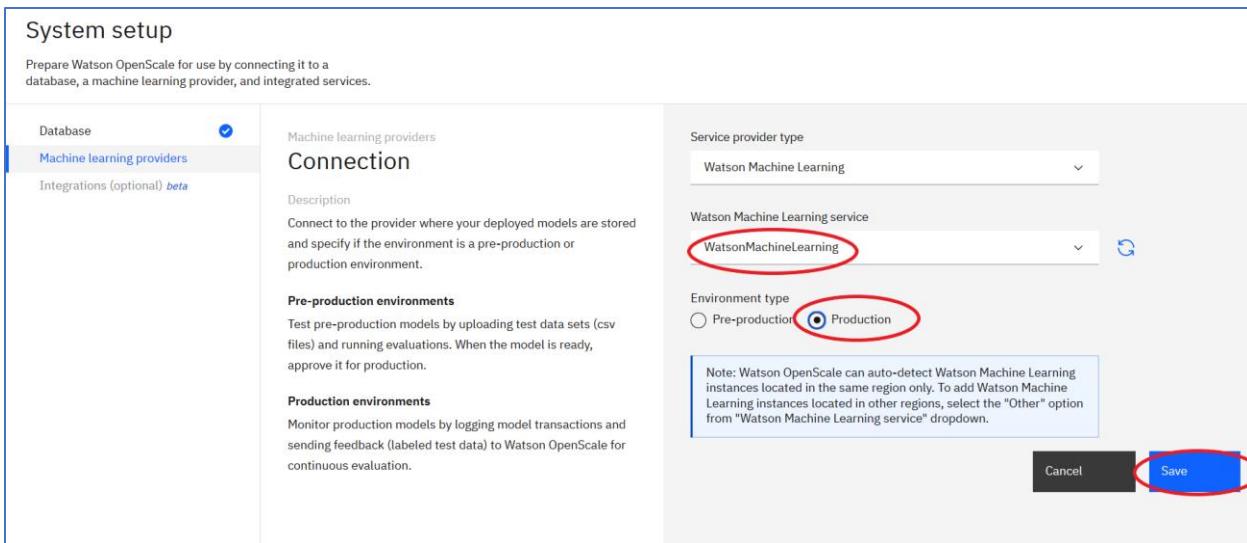
Service provider type
Watson Machine Learning

Watson Machine Learning service
WatsonMachineLearning

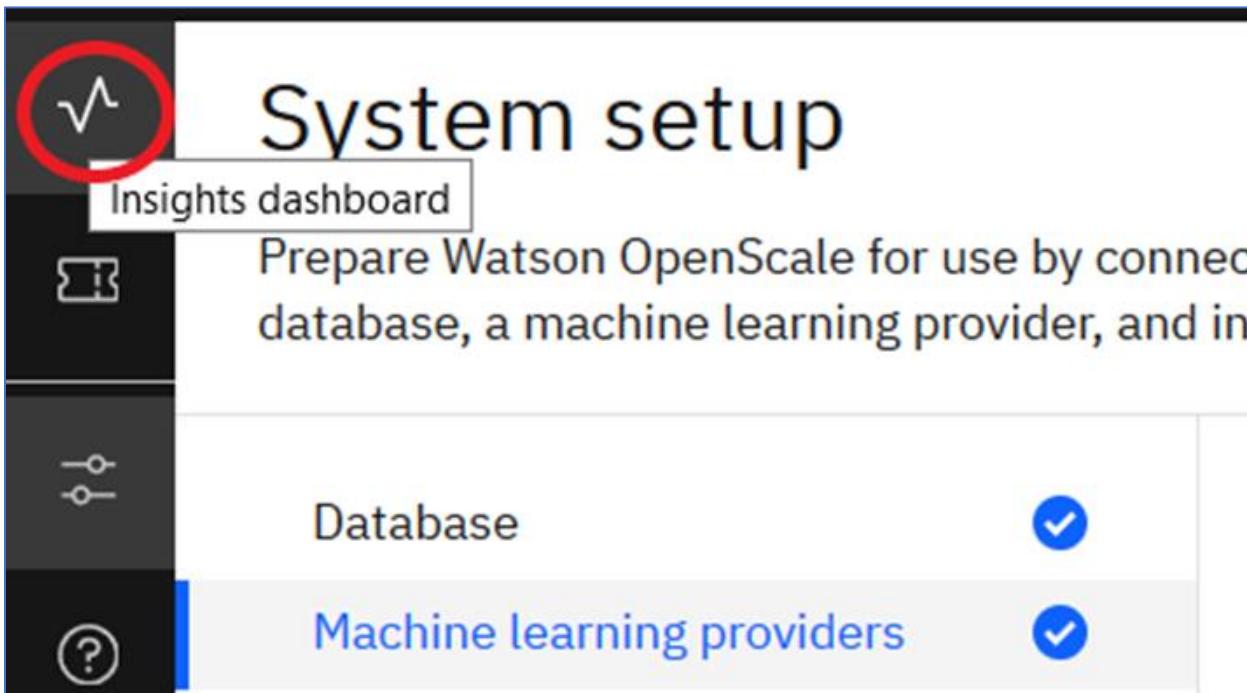
Environment type
 Pre-production Production

Note: Watson OpenScale can auto-detect Watson Machine Learning instances located in the same region only. To add Watson Machine Learning instances located in other regions, select the "Other" option from "Watson Machine Learning service" dropdown.

Save



9. Click on the  icon.



10. Click on **Add** to add a deployment to monitor.

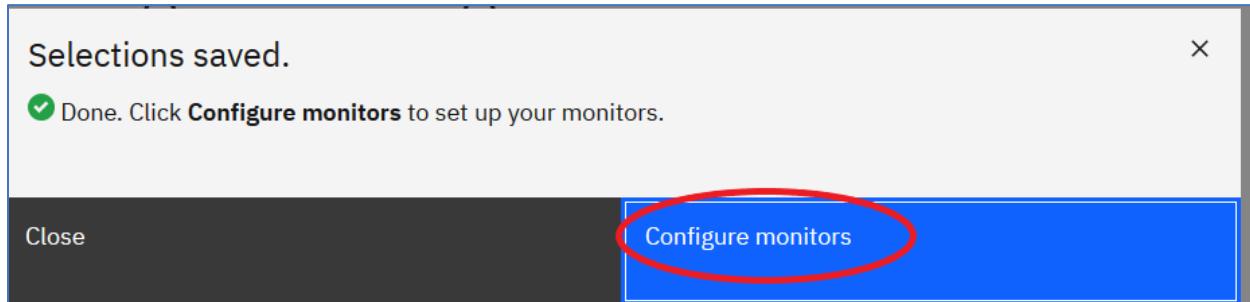
b

The screenshot shows the Insights Dashboard with a light gray background. At the top left is a sidebar with a blue header containing the text "Model Monitors". Below this is a section with four categories: "Deployments Monitored" (0), "Quality Alerts" (0), "Fairness Alerts" (0), and "Drift Alerts" (0). To the right of these is a large central area with the text "Add a deployed model to get started." and a blue "Add" button. A red oval highlights the "Add" button.

11. Click on **credit-risk-deploy** and then click on **Configure** to configure the model details, and the quality, fairness, and drift monitors.

The screenshot shows a modal dialog titled "Select a model deployment" with a close button "X" in the top right corner. The instruction "Select the deployment you want to monitor." is displayed. Under "Machine learning Provider", a dropdown menu shows "New provider (Production)". The main table lists one deployment: "credit-risk-deploy" (Description) created on "Tue, Mar 24, 2020, 11:52 AM EDT". A red oval highlights the "credit-risk-deploy" entry. At the bottom, there are "Cancel" and "Configure" buttons, with the "Configure" button circled in red.

12. Click on **Configure monitors**.



13. Setup is now complete. Do not proceed until you do the scoring step below.

Score Data

Before proceeding with the monitor configuration, we need to send scoring data to the deployed model in order to generate payload logging data that the monitors can consume. We will use the `credit_risk_data.json` file as sample data that Watson Studio will submit to the deployed model.

1. Click on the **Watson Studio** Browser tab
2. You should be back at the **credit-risk Model** page. Click on **credit-risk-deploy**.

A screenshot of the Watson Studio Model page for the "credit-risk" model. The top navigation bar shows "My Projects / Watson Studio Labs / credit-risk". Below the navigation is a "MODEL" section with the name "credit-risk". There are four tabs: "Overview", "Evaluation", "Deployments" (which is selected and highlighted in blue), and "Lineage". Under the "Deployments" tab, there is a table with columns: NAME, STATUS, DEPLOYMENT TYPE, and ACTIONS. A single row is present in the table, showing "credit-risk-deploy" in the NAME column, "DEPLOY_SUCCESS" in the STATUS column, "Web Service" in the DEPLOYMENT TYPE column, and an "Actions" button in the ACTIONS column. A red oval highlights the "credit-risk-deploy" entry in the table.

3. Click on the **Test** tab.

credit-risk-deploy	
Overview	Implementation
Deployment	
Name	credit-risk-deploy
Type	Web Service
Deployment ID	203b492f-f9ca-4166-95c9-a93e3018f091
Status	DEPLOY_SUCCESS
Asset type	model
Asset name	credit-risk
Machine learning service	WatsonMachineLearning
Created	27 Oct 2019 01:18pm
Last modified	27 Oct 2019 03:20pm

4. Click on the  icon to accept input as JSON.

credit-risk-deploy

- Overview
- Implementation
- Test

Enter input data



Provide input data as JSON

Paste the request payload here

Predict

5. Copy the file contents of **credit_payload_data.json** into the **input data** area and click on **Predict**.

credit-risk-deploy

Overview

Implementation

Test

Enter input data



```
{  
  "fields":  
    ["CheckingStatus", "LoanDuration", "CreditHistory", "Loa  
    nPurpose", "LoanAmount", "ExistingSavings", "Employme  
    ntDuration", "InstallmentPercent", "Sex", "OthersOnLoan  
    ", "CurrentResidenceDuration", "OwnsProperty", "Age", "I  
    nstallmentPlans", "Housing", "ExistingCreditsCount", "Jo  
    b", "Dependents", "Telephone", "ForeignWorker"],  
  "values":  
    []
```

Predict

6. The results should appear as below.

credit-risk-deploy

Overview

Implementation

Test

Enter input data



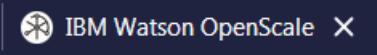
```
{  
  "fields":  
    ["CheckingStatus", "LoanDuration", "CreditHistory", "Loa  
    nPurpose", "LoanAmount", "ExistingSavings", "Employme  
    ntDuration", "InstallmentPercent", "Sex", "OthersOnLoan  
    ", "CurrentResidenceDuration", "OwnsProperty", "Age", "I  
    nstallmentPlans", "Housing", "ExistingCreditsCount", "Jo  
    b", "Dependents", "Telephone", "ForeignWorker"],  
  "values":  
    []
```

```
{  
  "fields": [  
    "CheckingStatus",  
    "LoanDuration",  
    "CreditHistory",  
    "LoanPurpose",  
    "LoanAmount",  
    "ExistingSavings",  
    "EmploymentDuration",  
    "InstallmentPercent",  
    "Sex",  
    "OthersOnLoan",  
    "CurrentResidenceDuration",  
    "OwnsProperty",  
    "Age",  
    "InstallmentPlans",  
    "Housing",  
    "Job",  
    "Dependents",  
    "Telephone",  
    "ForeignWorker"]  
}
```

Predict

Configure Model Details

1. Switch back to Watson OpenScale by clicking on the Watson OpenScale browser tab.



2. Click on pencil icon to edit the **Model input**.

The screenshot shows the "Model details" page. On the left, there's a "Model details" section with a "Description" field containing placeholder text about training data and model output. On the right, there's a "Model input" section with a descriptive text and a blue-bordered edit icon (pencil symbol) in the top right corner.

3. Click on **Numeric/categorical** for the **Data type**. Click on **Binary classification** for the **Algorithm type**. Click on **Save and Continue**.

The screenshot shows a modal dialog titled "Select input type and algorithm". It has a "Description" field with placeholder text. Below it are three options: "Numeric/categorical" (selected), "Image", and "Unstructured text". On the right, there's a configuration panel with two dropdowns: "Data type" set to "Numeric/categorical" and "Algorithm type" set to "Binary classification". At the bottom right of the dialog is a "Save and continue" button, which is highlighted with a red circle.

4. Click on pencil icon to edit **Training data**.

Model details

Description

Provide information about the training data and deployed model output to prepare Watson OpenScale for monitoring and providing explanations for model transactions.

Model input

Data type
Numeric/categorical

Algorithm type
Binary classification

Training data

To connect to the training data, click the edit icon.

Training data label

5. Accept the defaults for the database configuration and click **Next**.

Dashboard / Configure

credit-risk-deploy

Model details

Quality

Fairness

Explainability

Drift

Endpoints

This is the preferred option. Use this option if the training data is stored in a Db2 database or cloud object storage.

Restricted (statistics only)

Use this option if you do not want to provide the training data. Analyze the training data in a data science notebook to generate a file containing training data statistics.

dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net

SSL port
50001

Database
BLUDB

Username
cmb91569

Password
.....

Schema
CMB91569

Table
CREDIT_RISK_TRAIN_DATA

Cancel **Next**

6. We've already set up the logging (**Logging is active**) so click **Next**.

Model details

Examining model output

Description

Watson OpenScale tracks and stores transactions processed by this model. To prepare Watson OpenScale, send a scoring request. You can send a scoring request in one of the following ways:

Send JSON payload

Enter payload data and send a scoring request directly from Watson OpenScale.

Use the payload logging API

Send a scoring request from an external application or data science notebook using code snippets.

Scoring method

Automatic logging

Logging is active Click Next

Cancel **Back** **Next**

7. Watson OpenScale has determined that Risk is the label (target) column. Click **Next**.

Dashboard / Configure
credit-risk-deploy

Model details

Select the label column

Description

From the selected training features, select a single feature as the label. The label represents the correct prediction (ground-truth) for each record.

Features (1)	Type
Risk	A

Cancel Back Next

8. All of the features are used to train the model. Watson OpenScale has determined the feature types. Click **Next**.

Dashboard / Configure
credit-risk-deploy

Model details

Select the training features

Description

From the training data, select the features used to train the model.

For each selected feature with a numeric data type, indicate if the numeric feature values represent codes or categories by checking the Categorical checkbox. This instructs OpenScale to process the feature as a set of classifications (ex. ZIP code) rather than continuous numeric values (ex. median household income).

Features (20)	Type	Categorical
Age	81	<input type="checkbox"/>
CheckingStatus	A	<input checked="" type="checkbox"/>
CreditHistory	A	<input checked="" type="checkbox"/>
CurrentResidenceDuration	81	<input type="checkbox"/>
Dependents	81	<input type="checkbox"/>
EmploymentDuration	A	<input checked="" type="checkbox"/>

Cancel Back Next

9. Watson OpenScale has determined the feature that contains the prediction generated by the AI deployment. Click **Save**.

Dashboard / Configure

credit-risk-deploy

Model details	<input type="radio"/>
Quality	<input type="radio"/>
Fairness	<input type="radio"/>
Explainability	<input checked="" type="radio"/>
Drift	<input type="radio"/>
Endpoints	<input checked="" type="radio"/>

Model details

Specify model output details

Description

Select the prediction column

From the model output data, select the column that contains the prediction generated by the deployed model.

Select the prediction probability column

The prediction probability column contains the model's confidence in the prediction it provides.

Probability column detected
Watson OpenScale automatically selected the probability column for this model.

Features (3)	Type	Prediction	Probability
prediction	<input type="checkbox"/>	<input type="checkbox"/>	
predictedLabel	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
probability	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

Cancel Back Save

10. Model details are now complete. Note that Explainability input is also complete. The next step is to configure Quality.

Configure Quality

The Quality Monitor evaluates how well your deployed model predicts accurate outcomes. It identifies when model quality declines so you can retrain appropriately.

1. Click **Quality** to configure the quality inputs.

Dashboard / Configure

credit-risk-deploy

Model details	<input checked="" type="radio"/>
Quality	<input checked="" type="radio"/>
Fairness	<input type="radio"/>
Explainability	<input checked="" type="radio"/>
Drift	<input type="radio"/>
Endpoints	<input checked="" type="radio"/>

Model details

Description

Provide information about the training data and deployed model output to prepare Watson OpenScale for monitoring and providing explanations for model transactions.

Model input

Data type: Numeric/categorical
Algorithm type: Binary classification

Training data

Storage type: Database or cloud storage
Location of training data: Db2
Hostname or IP address: dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net
SSL port: 50001
Database: BLUDB
Username: cmib91569
Password:

2. Click on the pencil icon to edit the **Quality threshold**.

Dashboard / Configure

credit-risk-deploy

Model details	<input checked="" type="radio"/>
Quality	<input type="radio"/>
Fairness	<input type="radio"/>
Explainability	<input checked="" type="radio"/>
Drift	<input type="radio"/>
Endpoints	<input checked="" type="radio"/>

Quality

Description

The Quality monitor evaluates how well your model predicts accurate outcomes. It identifies when model quality declines, so you can retrain your model appropriately.

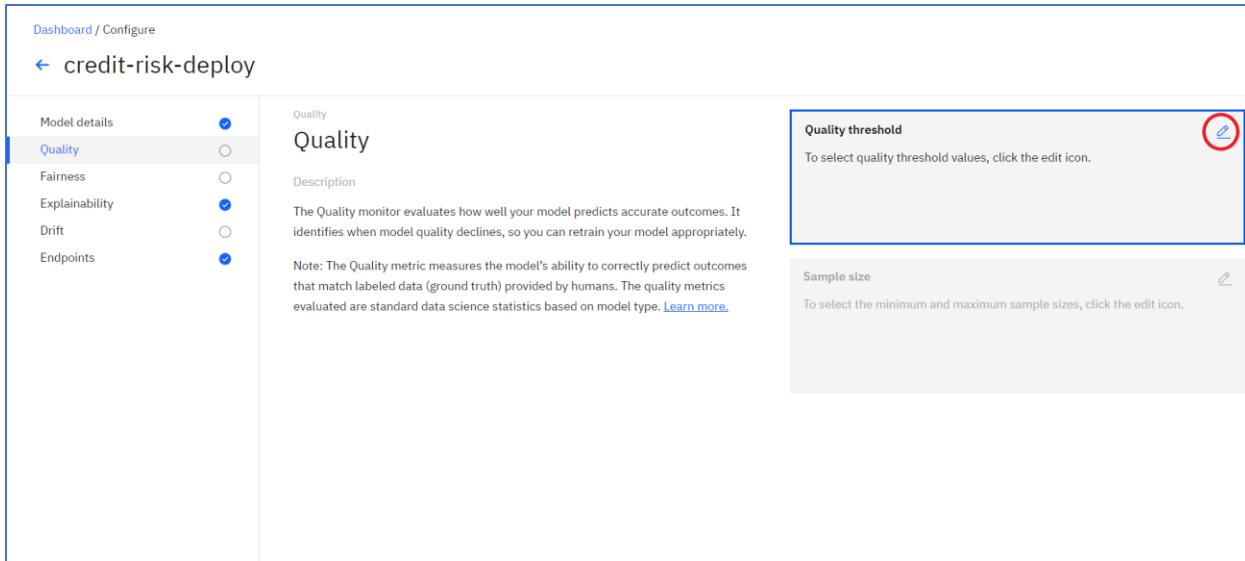
Note: The Quality metric measures the model's ability to correctly predict outcomes that match labeled data (ground truth) provided by humans. The quality metrics evaluated are standard data science statistics based on model type. [Learn more](#).

Quality threshold

To select quality threshold values, click the edit icon.

Sample size

To select the minimum and maximum sample sizes, click the edit icon.



3. Enter .9 for the Threshold value and click **Next**.

Dashboard / Configure

credit-risk-deploy

Model details	<input checked="" type="radio"/>
Quality	<input type="radio"/>
Fairness	<input type="radio"/>
Explainability	<input checked="" type="radio"/>
Drift	<input type="radio"/>
Endpoints	<input checked="" type="radio"/>

Import settings

[Go to model summary](#)

Quality threshold

Description

Area under ROC

For binary classification models, it will check the score which measures the model's ability to distinguish two classes. The higher the score, the better the model is at identifying class A as class A and class B as class B. The score of 1 is perfect, while a score of .50 is equal to a random guess.

Additional quality metrics captured for this model include:

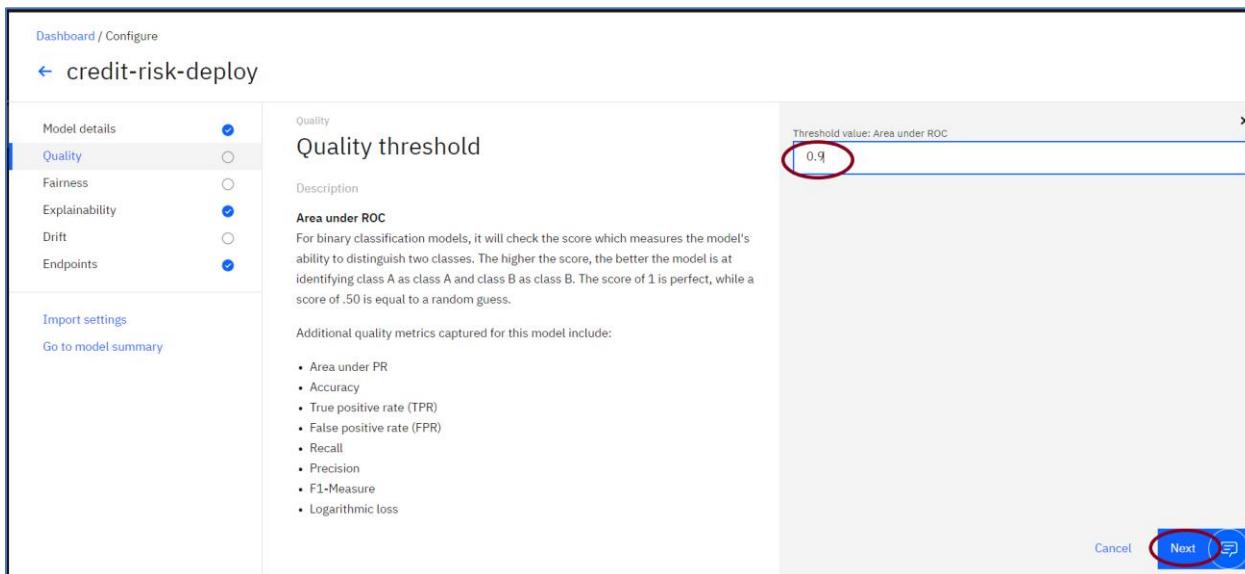
- Area under PR
- Accuracy
- True positive rate (TPR)
- False positive rate (FPR)
- Recall
- Precision
- F1-Measure
- Logarithmic loss

Threshold value: Area under ROC

0.9

Cancel

Next



4. Set minimum and maximum values for the sample size to be analyzed for quality. 100 is chosen for the minimum sample size, and 10000 is chosen for the maximum. Click **Save**.

Dashboard / Configure

credit-risk-deploy

Model details

- Quality
- Fairness
- Explainability
- Drift
- Endpoints

Quality

Sample size

Description

Ensure that your minimum sample size is large enough to be accurately represent the variety of requests the deployment receives. Set the maximum sample size to limit the volume of data analyzed.

Minimum sample size (number of transactions) (circled)

Maximum sample size (number of transactions) (circled)

Cancel Back Save (circled)

5. This completes the **Quality** configuration. The next step is to configure Fairness.

Configure Fairness

The Fairness monitor checks your deployment for biases. It tracks when the model shows a propensity to provide a particular outcome more often for one group over another.

1. Click on **Fairness** to configure the Fairness monitor.

Dashboard / Configure

credit-risk-deploy

Model details

- Quality
- Fairness (circled)
- Explainability
- Drift
- Endpoints

Quality

Description

The Quality monitor evaluates how well your model predicts accurate outcomes. It identifies when model quality declines, so you can retrain your model appropriately.

Note: The Quality metric measures the model's ability to correctly predict outcomes that match labeled data (ground truth) provided by humans. The quality metrics evaluated are standard data science statistics based on model type. [Learn more](#).

Quality threshold

Threshold value: Area under ROC
0.8

Sample size

Minimum sample size
100

Maximum sample size
10,000

2. Click on the pencil icon to edit the **Favorable outcomes**.

Dashboard / Configure

credit-risk-deploy

Model details

Quality

Fairness

Explainability

Drift

Endpoints

Fairness

Description

The Fairness monitor checks your deployments for biases. It tracks when the model shows a tendency to provide a favorable (preferable) outcome more often for one group over another. You will specify which values represent favorable outcomes, select the features to monitor for bias (for example, Age or Sex), and specify the groups to monitor for each selected feature.

Favorable outcomes

To select the favorable outcomes, click the edit icon.

Sample size

To select the minimum sample size, click the edit icon.

Features to evaluate (0)

Add feature +

7:50 PM

3. Check the Favorable check box for **No Risk** and check the Unfavorable check box for **Risk**. Then click **Next**.

Dashboard / Configure

credit-risk-deploy

Model details

Quality

Fairness

Explainability

Drift

Endpoints

Select the favorable outcomes

Description

For each group, Watson OpenScale will calculate the percentage of transactions that receive a favorable outcome.

Select the values that represent favorable (preferable) outcomes. You can also add a value manually if it is not included in the list.

Select the favorable outcomes

Enter a value

Add value

Values	Favorable	Unfavorable
No Risk	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Risk	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Cancel **Next**

7:50 PM

4. Change the **Minimum sample size** to 200 and click **Next**.

Dashboard / Configure

credit-risk-deploy

Model details

Quality

Fairness **Selected**

Explainability

Drift

Endpoints

Fairness

Sample size

Description

Ensure that your minimum sample size is large enough to be accurately represent the variety of requests the deployment receives.

Minimum sample size (number of transactions)

200

Cancel Back Next

5. Scroll if needed and click on the **Sex** check box and the **Age** check box to monitor those features for bias, and then click on **Next**.

Dashboard / Configure

credit-risk-deploy

Model details

Quality

Fairness **Selected**

Explainability

Drift

Endpoints

Fairness

Select the features to monitor

Description

For each feature you select, Watson OpenScale will monitor the deployed model's tendency to provide a favorable (preferred) outcome for one group over another.

Features are monitored individually, but the Watson OpenScale debiasing algorithm will correct bias issues for all monitored features together.

With the Lite plan, you can select up to 2 features to monitor. [View upgrade options.](#)

Select the features to monitor

Feature	Value
<input type="checkbox"/> Sex	81
<input type="checkbox"/> Age	81
<input type="checkbox"/> OthersOnLoan	81
<input type="checkbox"/> CurrentResidenceDuration	81
<input type="checkbox"/> OwnsProperty	81
<input type="checkbox"/> InstallmentPlans	81
<input type="checkbox"/> Housing	81

Cancel Back Next

6. Enter 19 for the **minimum value** and 25 for the maximum value and then click **Add value**. We are defining the age range 19-25 as the age monitoring group.

Dashboard / Configure

credit-risk-deploy

Model details	<input checked="" type="radio"/>
Quality	<input checked="" type="radio"/>
Fairness	<input type="radio"/>
Explainability	<input checked="" type="radio"/>
Drift	<input type="radio"/>
Endpoints	<input checked="" type="radio"/>

Fairness

Specify the monitored groups for [Age]

Description

Add value ranges and select the groups to monitor. Minimum and maximum values from the training data are presented for reference.

The percentage of favorable outcomes delivered to the monitored groups will be compared to the percentage of favorable outcomes delivered to the remaining groups (the reference groups) to check for potential bias.

A fairness score of 100% implies that the monitored group and reference group received an equal number of favorable values. Likewise, a fairness score of 50% implies that the monitored group received half as many favorable outcomes as the reference group.

Set the fairness alert threshold to track when the fairness value falls below an acceptable level.

Select the groups to monitor [Age]

Minimum value: 19 Maximum value: 74

19 25

7. Enter **26** as the **minimum value** and **74** as the **maximum value** and then click **Add value**. We are defining the age range 26-74 as the reference group.

Dashboard / Configure

credit-risk-deploy

Model details	<input checked="" type="radio"/>
Quality	<input checked="" type="radio"/>
Fairness	<input type="radio"/>
Explainability	<input checked="" type="radio"/>
Drift	<input type="radio"/>
Endpoints	<input checked="" type="radio"/>

Fairness

Specify the monitored groups for [Age]

Description

Add value ranges and select the groups to monitor. Minimum and maximum values from the training data are presented for reference.

The percentage of favorable outcomes delivered to the monitored groups will be compared to the percentage of favorable outcomes delivered to the remaining groups (the reference groups) to check for potential bias.

A fairness score of 100% implies that the monitored group and reference group received an equal number of favorable values. Likewise, a fairness score of 50% implies that the monitored group received half as many favorable outcomes as the reference group.

Set the fairness alert threshold to track when the fairness value falls below an acceptable level.

Select the groups to monitor [Age]

Minimum value: 19 Maximum value: 74

26 74

8. Check the **Monitored** check box for the 19-25 group and check the **Reference** check box for the 26-74 group. Set **95** for the **fairness alert threshold**. If OpenScale has already added “recommended” groups (pre-filled), uncheck those and ensure that the 19-25 group and 26-74 group you added are the only ones checked. Click **Next**.

Specify the monitored groups for [Age]

Values	Monitored	Reference	Recommended
19-25	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
26-74	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

Set fairness alert threshold [Age]
95

- Check the **Monitored** check box for female and check the **Reference** check box for male. Set **95** for the **fairness alert threshold**. Click on **Save**.

Specify the monitored groups for [Sex]

Values	Monitored	Reference	Recommended
female	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
male	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

Set fairness alert threshold [Sex]
95

- This completes the Fairness configuration. The next step is to configure Drift.

Configure Drift

The Drift monitor measures two types of changes. It measures the drop in accuracy of the deployed model during runtime. The model accuracy could drop if there is an increase in the number of transactions similar to ones that the model was not able to accurately evaluate in the training data.

It measures the drop in consistency of the data in runtime as compared to the characteristics of the data at training.

- Click on **Drift** to configure the Drift monitor.

The screenshot shows the Watson OpenScale dashboard under the 'credit-risk-deploy' model. On the left, a sidebar lists 'Model details' with several metrics checked: Quality (checked), Fairness (checked), Explainability (checked), Drift (unchecked, circled in red), and Endpoints (checked). The main content area is titled 'Quality' and describes the monitor's function of evaluating model quality. It includes a note about the Quality metric measuring prediction accuracy and a link to learn more. To the right, there are two sections: 'Quality threshold' (threshold value: Area under ROC 0.9) and 'Sample size' (minimum sample size 100, maximum sample size 10,000).

2. Click on the pencil icon to select a **drift model training option**.

The screenshot shows the Watson OpenScale dashboard under the 'credit-risk-deploy' model. The sidebar shows 'Model details' with Quality, Fairness, Explainability, and Endpoints checked, while Drift is unchecked. The main content area is titled 'Drift' and describes the drift test. It includes two sections: 'Drop in accuracy' (structured binary and multi-class classification models only) and 'Drop in data consistency'. To the right, there are three sections: 'Drift model' (with a red circle around the edit icon), 'Drift threshold' (with a red circle around the edit icon), and 'Sample size' (with a red circle around the edit icon).

3. Watson OpenScale will detect a drop in accuracy using a custom drift model generated from the training data. Similarly, it detects a drop in data consistency by analyzing your training data. Click on **Train in Watson OpenScale**. Click **Next**.

Dashboard / Configure

credit-risk-deploy

Model details	<input checked="" type="checkbox"/>
Quality	<input checked="" type="checkbox"/>
Fairness	<input checked="" type="checkbox"/>
Explainability	<input checked="" type="checkbox"/>
Drift	<input type="radio"/>
Endpoints	<input checked="" type="checkbox"/>

Train a drift model

Description

Watson OpenScale will detect a drop in accuracy using a custom drift model generated from your training data. Similarly, it detects a drop in data consistency by analyzing your training data. Watson OpenScale can analyze the data and train the model for you or you can do it yourself using a custom notebook.

Train in Watson OpenScale

If you connected your training data to Watson OpenScale and it is less than 500 MB, use this option.

Train in a data science notebook

If you did not connect your training data to Watson OpenScale or it exceeds 500 MB, use this option.

Training option

Train in Watson OpenScale

Train in a data science notebook

Cancel **Next**

4. Set the drift alert threshold. Leave **10%**. Click **Next**.

Dashboard / Configure

credit-risk-deploy

Model details	<input checked="" type="checkbox"/>
Quality	<input checked="" type="checkbox"/>
Fairness	<input checked="" type="checkbox"/>
Explainability	<input checked="" type="checkbox"/>
Drift	<input type="radio"/>
Endpoints	<input checked="" type="checkbox"/>

Drift threshold

Description

Watson OpenScale will track the degree of change in model accuracy when compared to the accuracy at training time.

For example, if the accuracy of the model tested against the training data is 90% and the drift threshold is set to 5%, the estimated accuracy of the model on incoming transactions must be 85% or greater to avoid a threshold violation.

Testing the model on a sample of the training data set determines the baseline for measuring change. The drift measure will be reported as an estimated value with a specified margin of error. The drift threshold must be 5% or greater.

Drift threshold

10 %

Cancel Back **Next**

5. Set the **Sample size** to compute Drift. Select **200**. Click **Save**.

Dashboard / Configure

credit-risk-deploy

Model details	<input checked="" type="checkbox"/>
Quality	<input checked="" type="checkbox"/>
Fairness	<input checked="" type="checkbox"/>
Explainability	<input checked="" type="checkbox"/>
Drift	<input type="radio"/>
Endpoints	<input checked="" type="checkbox"/>

Drift

Sample size

Description

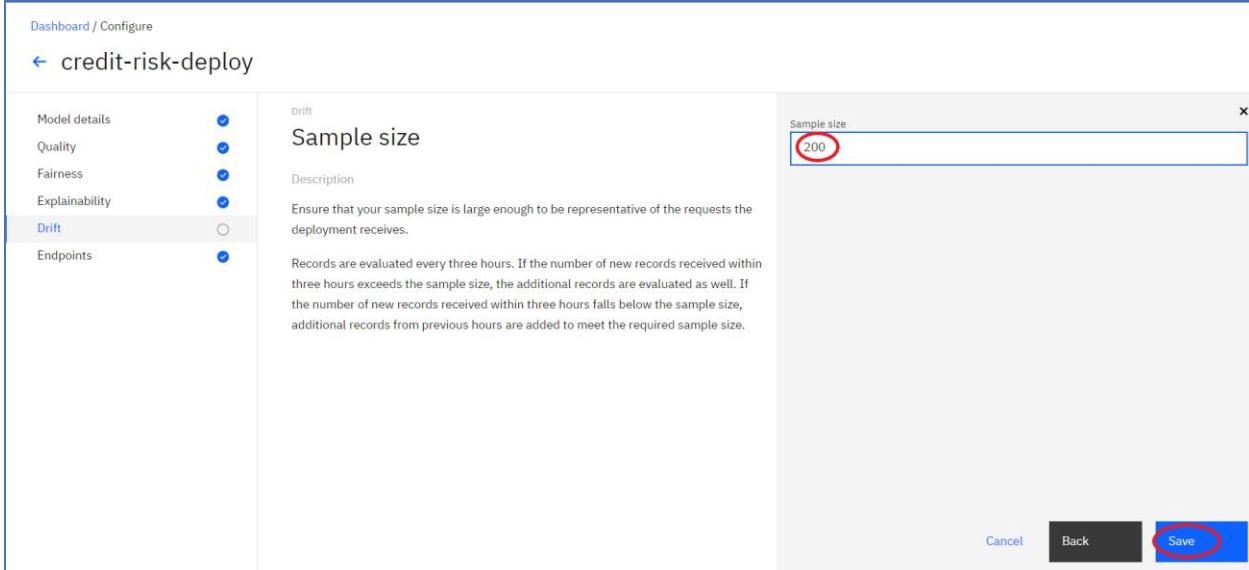
Ensure that your sample size is large enough to be representative of the requests the deployment receives.

Records are evaluated every three hours. If the number of new records received within three hours exceeds the sample size, the additional records are evaluated as well. If the number of new records received within three hours falls below the sample size, additional records from previous hours are added to meet the required sample size.

Sample size

200

Cancel Back Save



6. This completes the Drift configuration.

Submit Feedback and View Quality Metrics

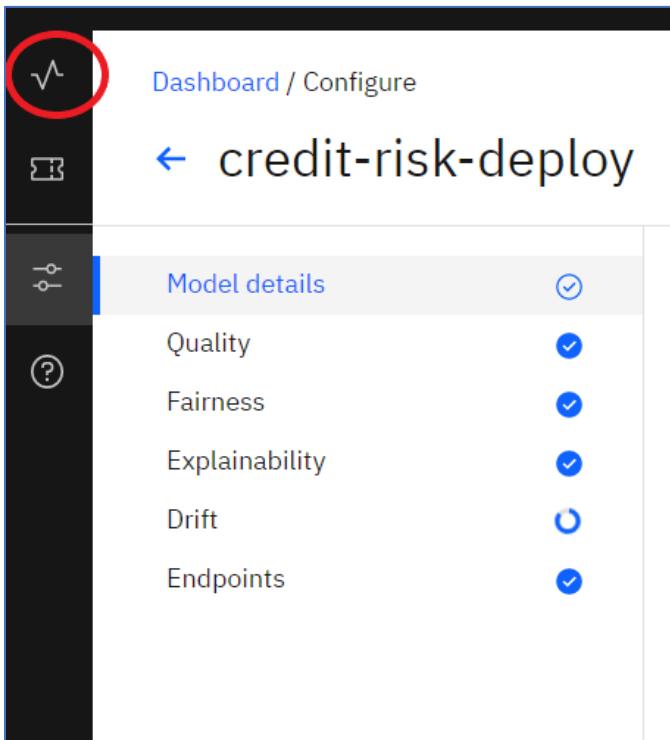
In order to measure quality, scored transactions including human labeled feedback must be provided.

1. Click on the  to display the **Insights Dashboard**.

Dashboard / Configure

credit-risk-deploy

	Model details	<input checked="" type="checkbox"/>
	Quality	<input checked="" type="checkbox"/>
	Fairness	<input checked="" type="checkbox"/>
	Explainability	<input checked="" type="checkbox"/>
	Drift	<input type="radio"/>
	Endpoints	<input checked="" type="checkbox"/>



2. Click on vertical ellipse  and then click on **View Details**.

Insights Dashboard

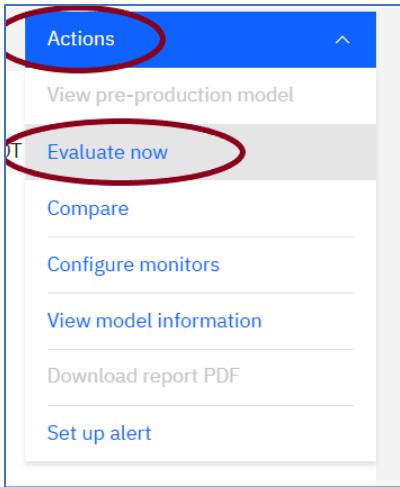
Model Monitors	Deployments	Quality Alerts	Fairness Alerts	Drift Alerts
1	Monitored 1	0	0	0

 Quality and Fairness metrics update every hour. Drift metrics update every 3 hours.

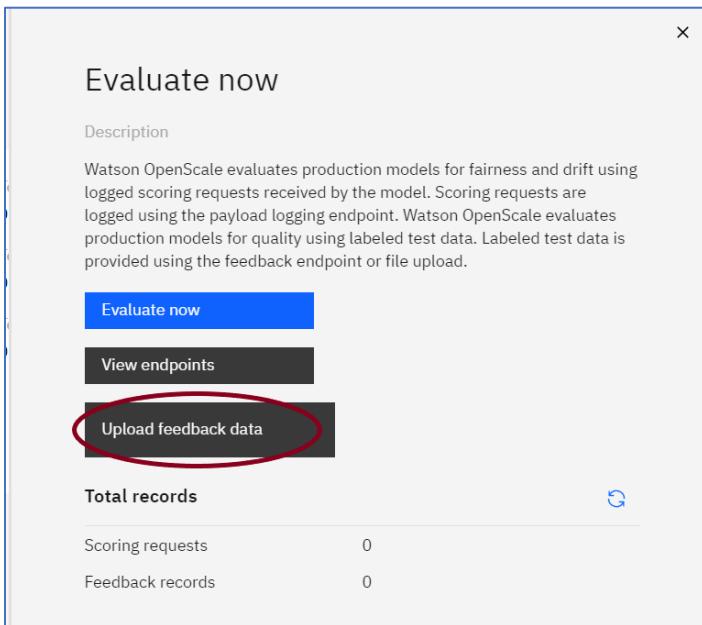
Watson Machine Learning	
credit-risk-deploy	
Issues 0	View details
Quality	Configure monitors Remove deployment
N/A	N/A


Evaluation pending

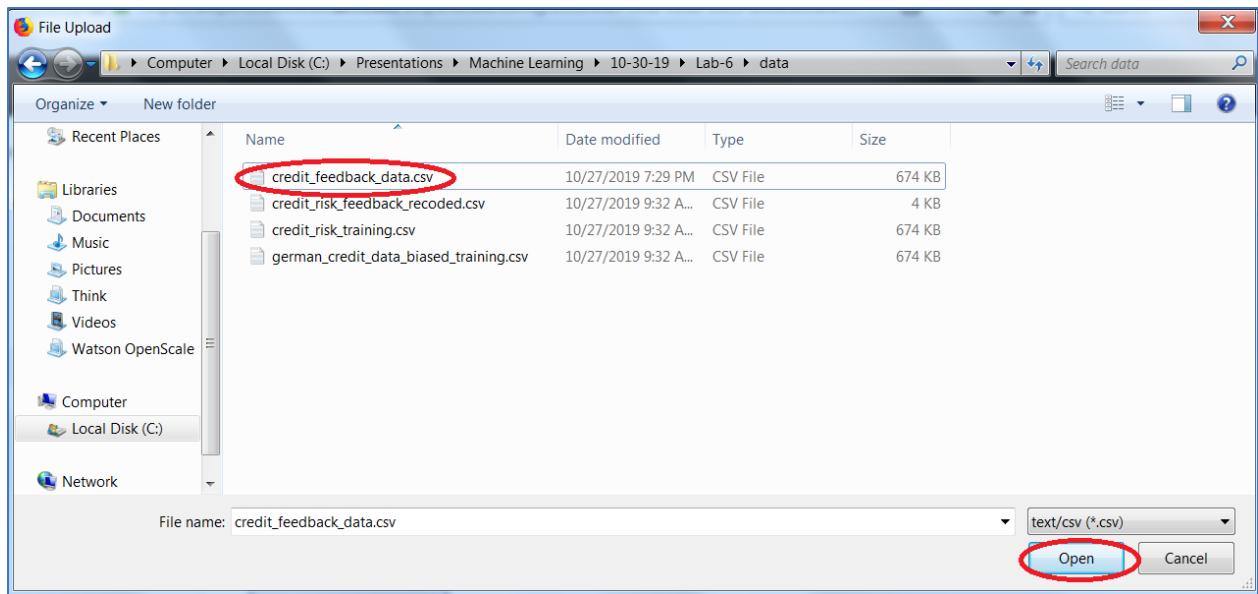
3. It will take a minute or so to bring up the **Evaluations** page. Click on **Actions** and **Evaluate now**.



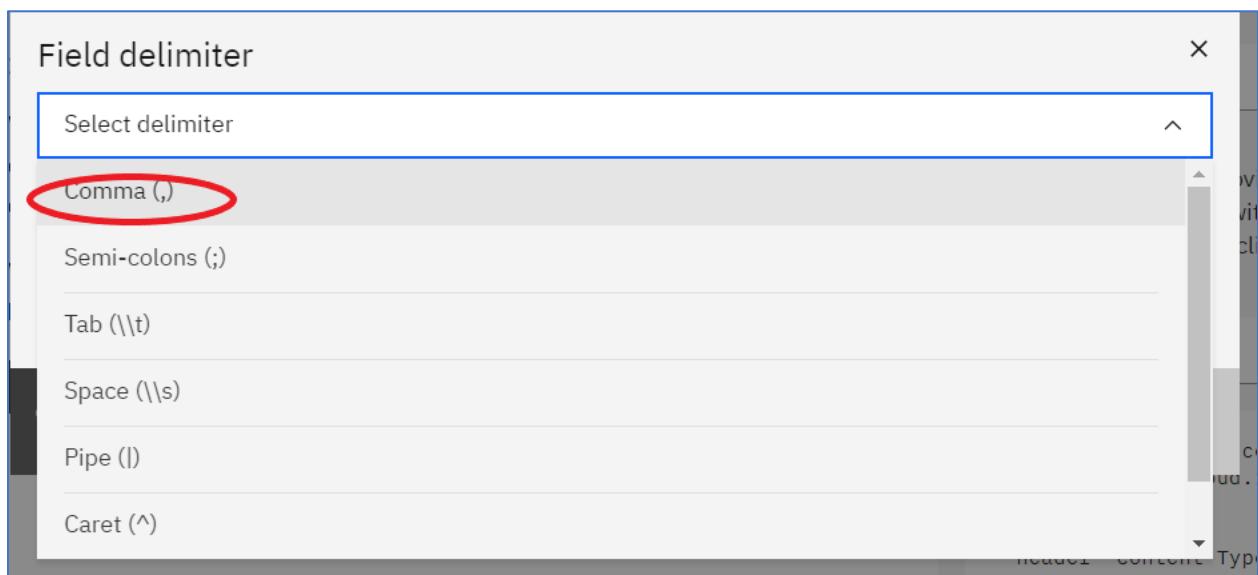
- Watson OpenScale evaluates models for quality using labeled test data. Labeled test data is provided using the feedback endpoint or by file upload. We will use the file upload options. Click on **Upload feedback data**.



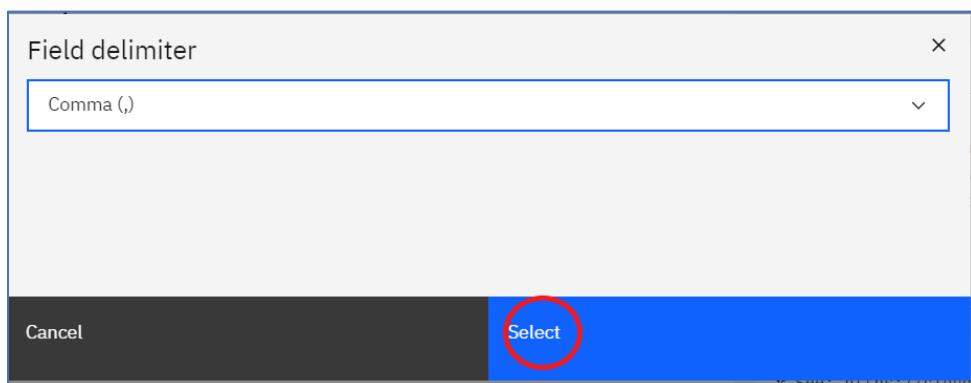
- Navigate to the feedback file **credit_feedback_data.csv**. Click on **Open**.



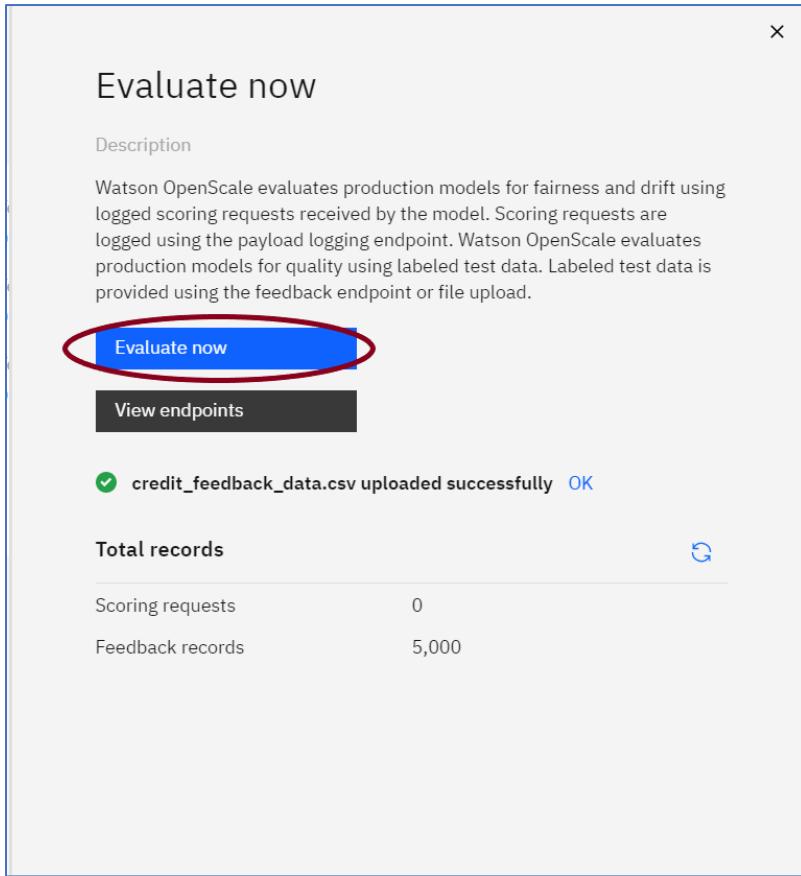
6. Select **Comma** as the **Delimiter**.



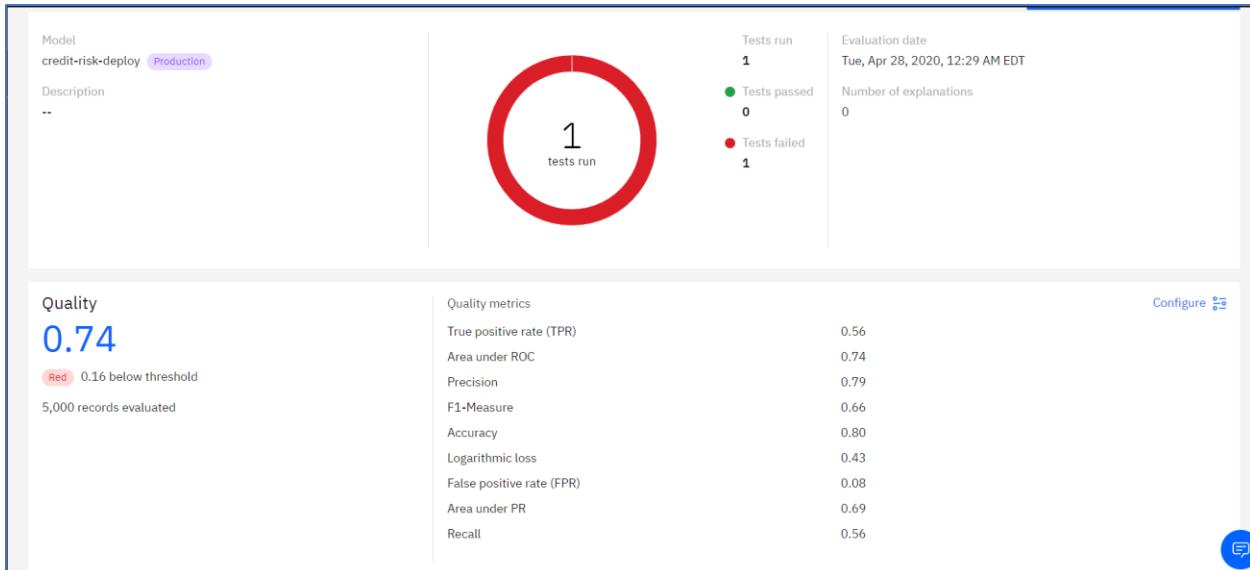
7. Click **Select**.



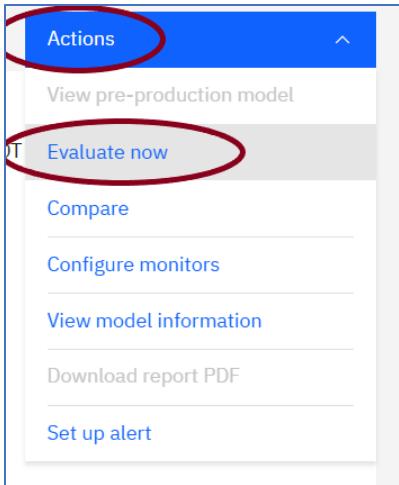
8. After the file is successfully uploaded, click on the **Evaluate now**.



9. Please wait about 3-5 minutes until the evaluation is done. The quality results are displayed. The quality test failed because the accuracy is below the quality threshold.



10. Click on **Actions** and click on **Evaluate Now**.



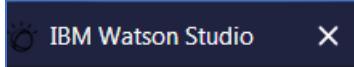
11. Leave Watson OpenScale at the Evaluation page while, you score transactions below.

A screenshot of the 'Evaluate now' page. It contains a descriptive text about how Watson OpenScale evaluates production models for fairness and drift using logged scoring requests. Below the text are three buttons: 'Evaluate now' (blue), 'View endpoints' (dark grey), and 'Upload feedback data' (dark grey).

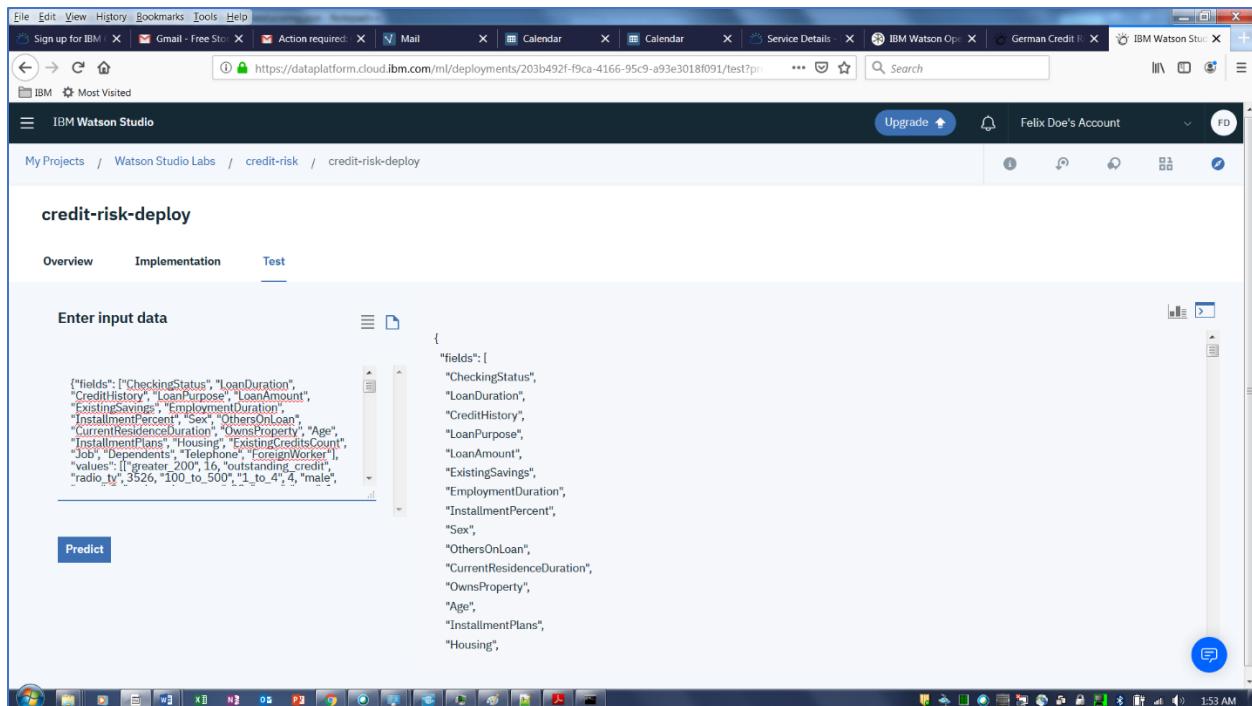
Score Transactions and View Fairness and Drift Metrics

In order to display Fairness metrics and Drift metrics, we need to direct transactions to the deployed model. We will use the scoring.json file as sample data that Watson Studio will submit to the deployed model

1. Return to Watson Studio by clicking on the **Watson Studio** browser tab.



2. You should be at the **Test** tab of the **credit-risk-deploy** page.



3. CLEAR out the contents of the **input data** area.
 4. Navigate to where the scoring.json file and cut and paste the contents of the file into the **input data** area.

```
{"fields": ["CheckingStatus", "LoanDuration", "CreditHistory", "LoanPurpose", "LoanAmount", "ExistingSavings", "EmploymentDuration", "InstallmentPercent", "Sex", "OthersOnLoan", "Cur
```

5. Click on Predict

credit-risk-deploy

Overview Implementation Test

Enter input data

☰

```
{"fields": ["CheckingStatus", "LoanDuration", "CreditHistory", "LoanPurpose", "LoanAmount", "ExistingSavings", "EmploymentDuration", "InstallmentPercent", "Sex", "OthersOnLoan", "CurrentResidenceDuration", "OwnsProperty", "Age", "InstallmentPlans", "Housing", "ExistingCreditsCount", "Job", "Dependents", "Telephone", "ForeignWorker"], "values": [{"radio_tv": 352, "100_to_500": 16, "1_to_4": 4, "male": 1}], "radio_tv": 352, "100_to_500": 16, "1_to_4": 4, "male": 1}
```

Predict

6. The result is displayed below.

The screenshot shows the IBM Watson Studio interface. The top navigation bar includes 'File', 'Edit', 'View', 'History', 'Bookmarks', 'Tools', and 'Help'. Below the bar, there are several tabs: 'Sign up for IBM', 'Gmail - Free Sto...', 'Action required...', 'Mail', 'Calendar', 'Service Details', 'IBM Watson Open...', 'German Credit ...', and 'IBM Watson Studio'. The main area is titled 'credit-risk-deploy' under 'My Projects / Watson Studio Labs / credit-risk / credit-risk-deploy'. The 'Test' tab is active. Below it, a section titled 'Enter input data' contains a JSON object:

```
{"fields": ["CheckingStatus", "LoanDuration", "CreditHistory", "LoanPurpose", "LoanAmount", "ExistingSavings", "EmploymentDuration", "InstallmentPercent", "Sex", "OthersOnLoan", "CurrentResidenceDuration", "OwnsProperty", "Age", "InstallmentPlans", "Housing", "ExistingCreditCount", "Job", "Dependents", "Telephone", "TotalWorkers", "values": [{"greater": 200, "le": 16, "outstanding": "credit", "radio_tv": 3526, "le_100": 500, "between": "1_to_4", "male": true}], "label": "Predict"}
```

A red box highlights the 'fields' key and its associated array of field names.

7. Return to Watson OpenScale by clicking on the **Watson OpenScale** browser tab.

and then click on the refresh icon, You should now see 200 Scoring requests.

The screenshot shows the Watson OpenScale interface. The top bar has a 'Logout' button. Below it, a modal window titled 'Evaluate now' is open. The window contains the following text:

Description
Watson OpenScale evaluates production models for fairness and drift using logged scoring requests received by the model. Scoring requests are logged using the payload logging endpoint. Watson OpenScale evaluates production models for quality using labeled test data. Labeled test data is provided using the feedback endpoint or file upload.

Below the text are three buttons: 'Evaluate now' (blue), 'View endpoints' (grey), and 'Upload feedback data' (grey). At the bottom, there is a section titled 'Total records' with two entries: 'Scoring requests' (200) and 'Feedback records' (5,000). To the right of the 'Total records' section is a blue circular refresh icon with a white arrow.

8. Click **Evaluate now**.

Evaluate now

Description

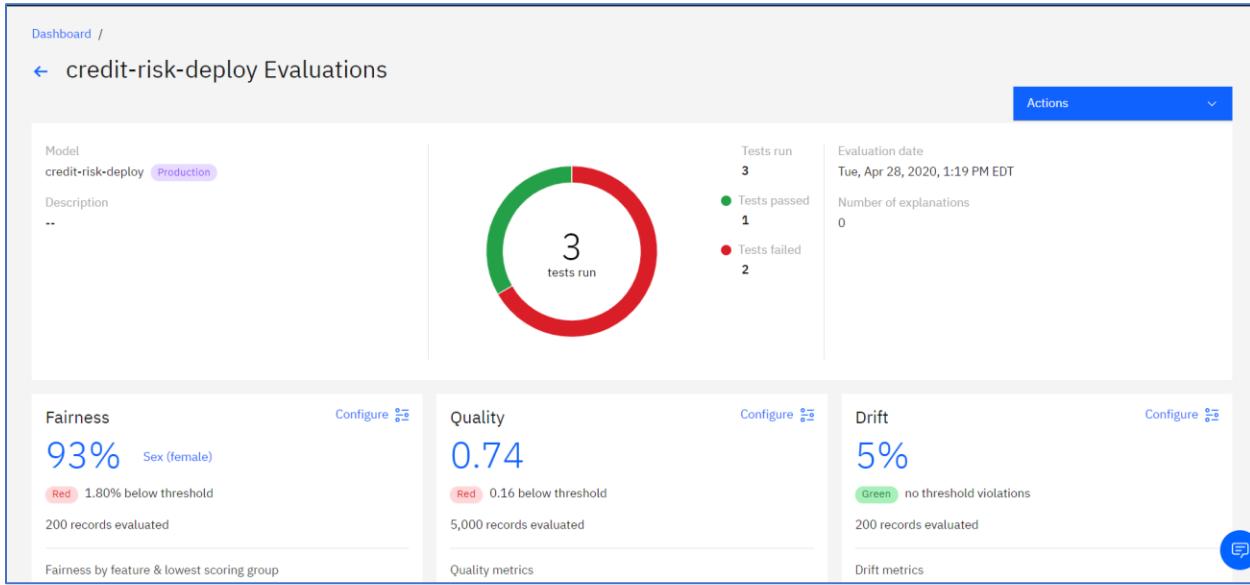
Watson OpenScale evaluates production models for fairness and drift using logged scoring requests received by the model. Scoring requests are logged using the payload logging endpoint. Watson OpenScale evaluates production models for quality using labeled test data. Labeled test data is provided using the feedback endpoint or file upload.

Evaluate now

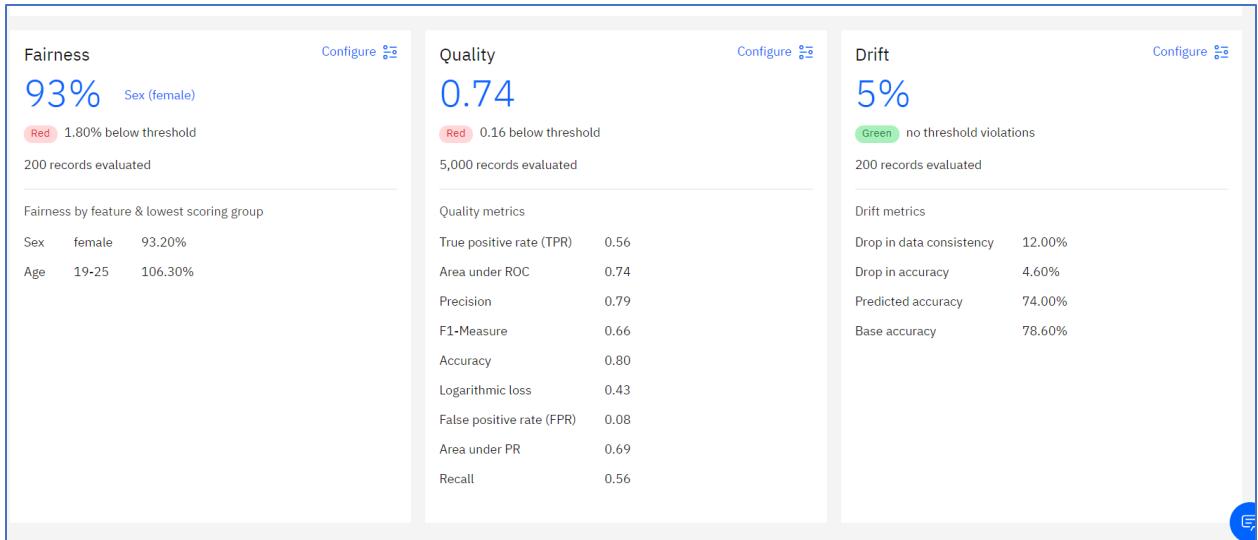
View endpoints

Upload feedback data

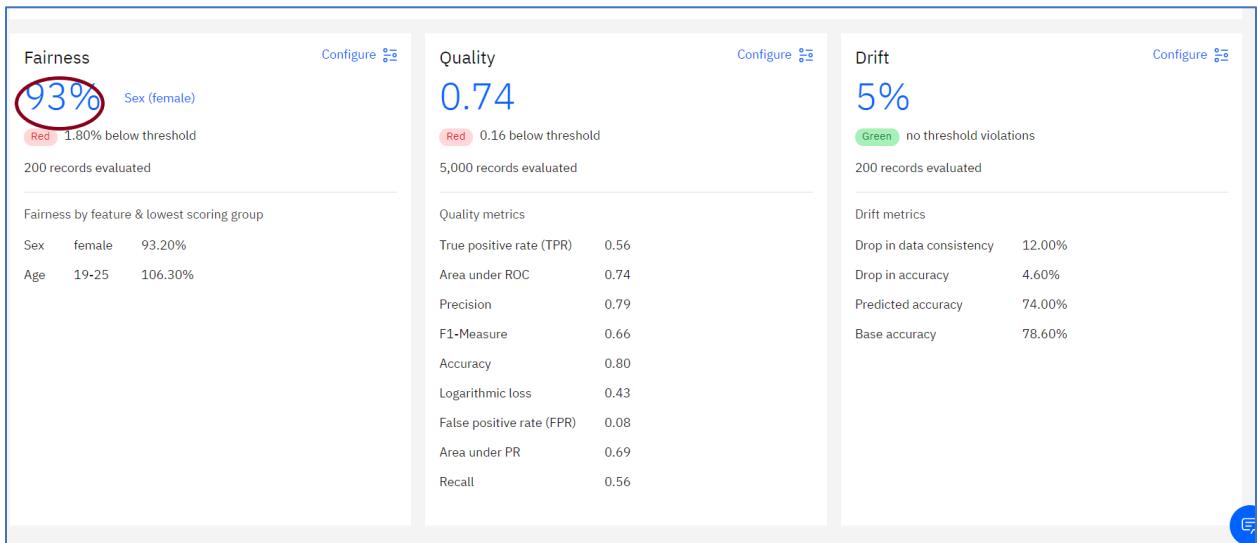
9. Please wait 3-5 minutes for the Evaluations page to display the metrics. Three tests have now been run. The quality test from before, and two new tests. One for fairness, and one for Drift. The fairness test has failed because 93% fairness is lower than our threshold of 95%. The Drift metric passed as the 5% drift is less than our drift threshold of 10%.
- Note, that lower drift is better, whereas lower fairness is obviously worse.



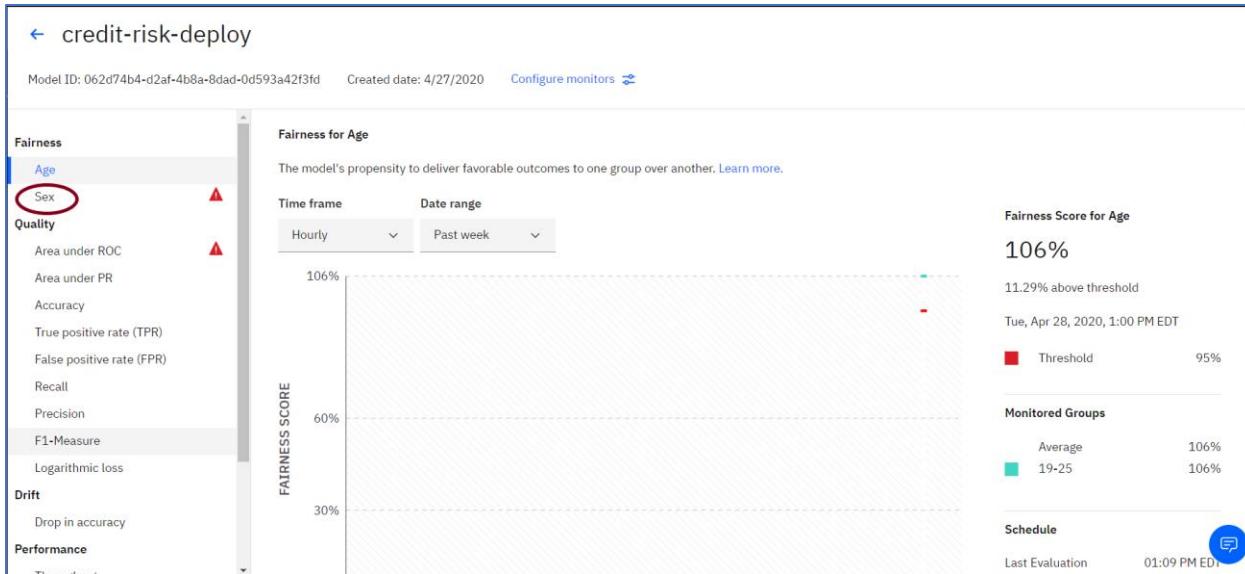
10. Scroll down to see the metric details.



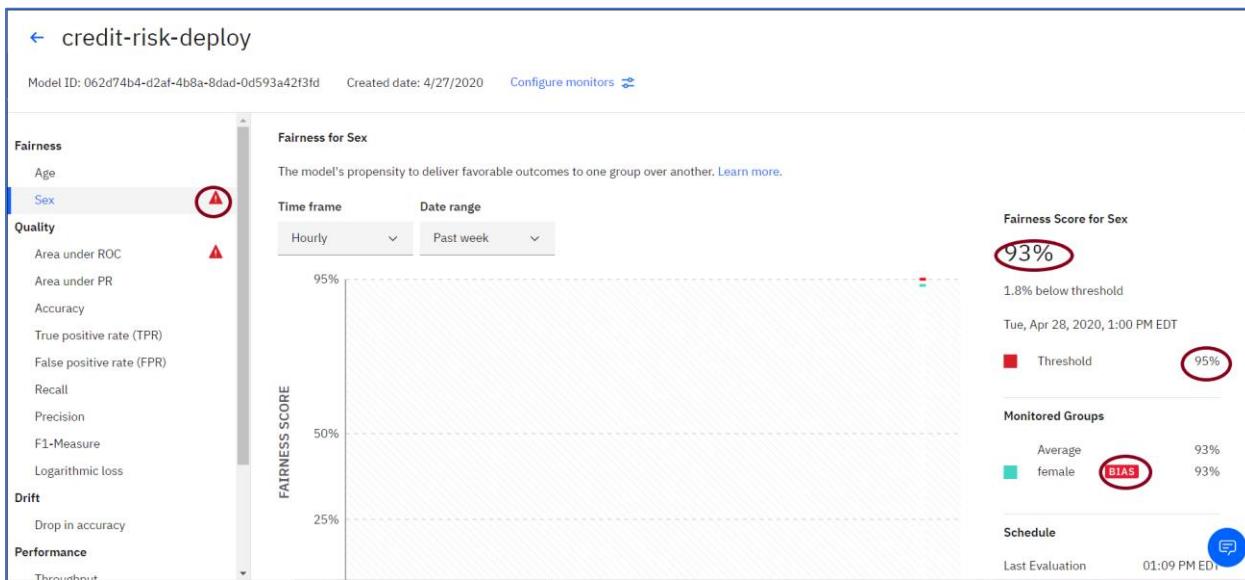
11. Click on the **Fairness** metric.



12. The time-series view is displayed. Click on **Sex**.



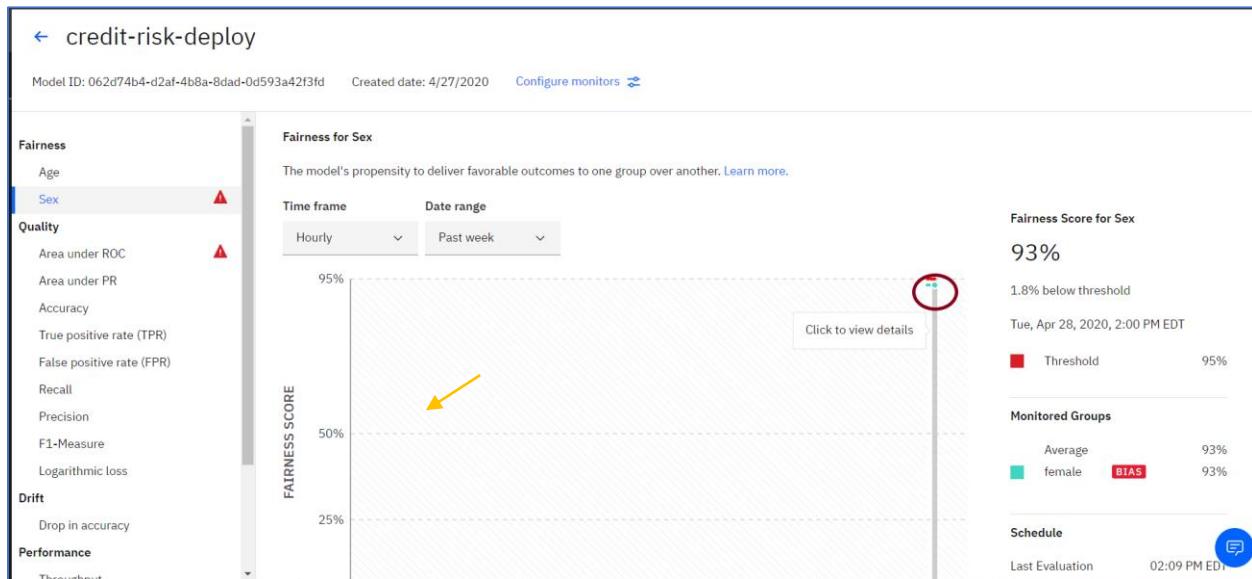
13. A time-series display for the Sex metric is shown. Currently, we have only run one evaluation. We can see the red triangle indicating an alert next to Sex. The Fairness score is listed at 93% which is less than the threshold of 95%.



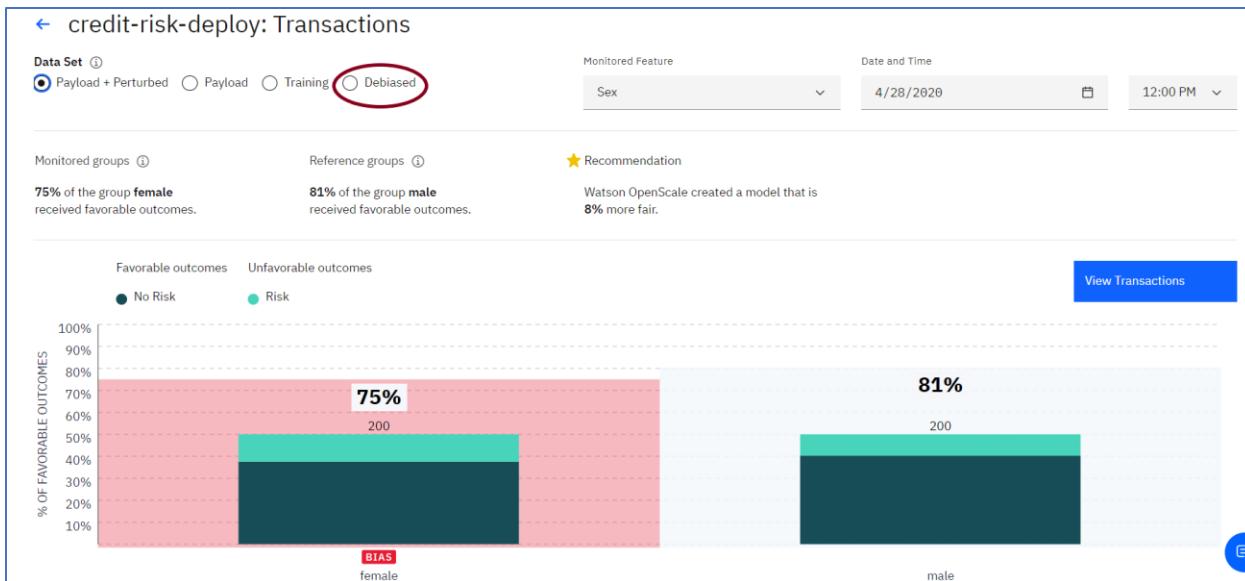
14. Scroll down and click on **Check fairness now**.



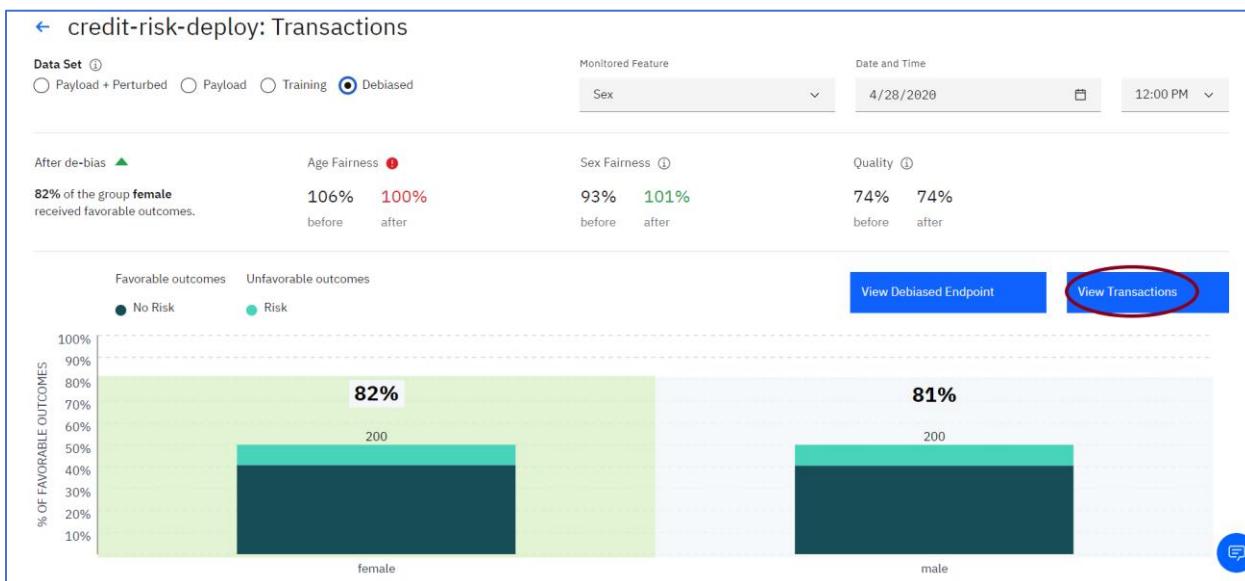
15. Click on the timeline to view details.



16. The discrepancy between Male and Female results are sufficient to trigger a bias alert.
Click on **Debiased**.



17. The results show that applying a debiasing scheme will reduce the bias to close to zero. The endpoint for invoking the debiasing algorithm can be obtained by clicking on View Debiased Endpoint. For now, click on **View Transaction** to display a list of transactions.

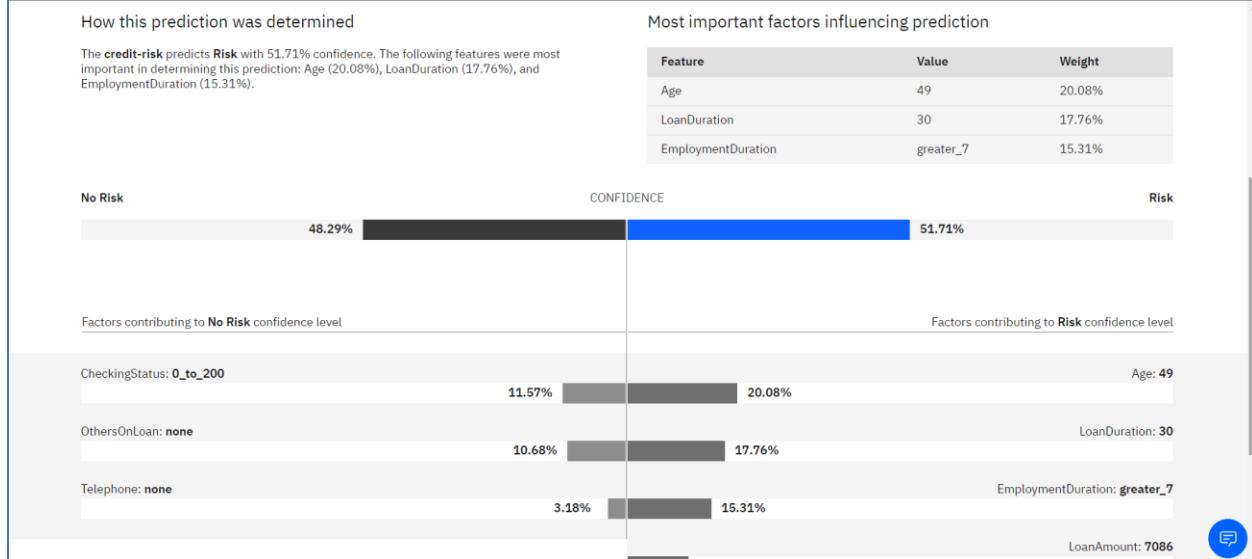


18. Click **Explain** next to a transaction to get an explanation of the factors that caused the deployed model to make the prediction it did for that transaction. I picked the first transaction that resulted in a Risk prediction. Note that it may take a few minutes for the transactions to show.

Transaction ID	Sex	Bias Detected	Outcome	Action
61a034fa736dabc57626b399cb139d38-1	male		No Risk	Explain
61a034fa736dabc57626b399cb139d38-10	female		No Risk	Explain
61a034fa736dabc57626b399cb139d38-100	male		No Risk	Explain
61a034fa736dabc57626b399cb139d38-101	male		No Risk	Explain
61a034fa736dabc57626b399cb139d38-102	male		No Risk	Explain
61a034fa736dabc57626b399cb139d38-103	male		No Risk	Explain
61a034fa736dabc57626b399cb139d38-104	male		No Risk	Explain
61a034fa736dabc57626b399cb139d38-105	female	✓	Risk	Explain
61a034fa736dabc57626b399cb139d38-106	female		No Risk	Explain
61a034fa736dabc57626b399cb139d38-107	male		No Risk	Explain

19. The results show two ways of explaining the prediction. One is using a LIME approach that provides factors “for” and “against” the decision. The second method is called **Contrastive Explanation**. It provides the minimum changes in features that would result in a different decision. It also provides the maximum changes in features that would leave the result the same.
 20. The Lime results are shown below.

20. The Lime results are shown below.



21. The Contrastive Explanation results are shown below.

← Explain a transaction	
<input type="text"/>	
 61a034fa736dabc57... x	
Details ⓘ	Minimum changes for No Risk outcome ⓘ
Transaction Deployment Model Name Type	61a034fa736dabc57626b399cb139d38-105 credit-risk-deploy credit-risk Original
	ExistingCreditsCount LoanAmount Age
	1.0 3238.5 36.0
	Maximum changes allowed for the same outcome ⓘ
	CheckingStatus LoanDuration CreditHistory
	0_to_200 30.0 credits_paid_to_date

22. Click on the Dashboard icon .

IBM Watson OpenScale

Need help? ⓘ

↶ Explain a transaction

🔍

61a034fa736dabc57...

?

Details ⓘ	Minimum changes for No Risk outcome ⓘ	Maximum changes allowed for the same outcome ⓘ
Transaction 61a034fa736dabc57626b399cb139d38-105	ExistingCreditsCount 1.0 LoanAmount 3238.5 Age 36.0	CheckingStatus 0_to_200 LoanDuration 30.0 CreditHistory credits_paid_to_date
Deployment credit-risk-deploy		
Model Name credit-risk		
Type Original		

23. The Dashboard view provides a snapshot view of how the deployed models are performing. Since we are only monitoring one deployment, there is only 1 monitoring tile displayed. You can drill down from this view to get the details on each of the metrics for a given deployment, as we've shown in this lab.

Insights Dashboard

Model Monitors

1

Add to dashboard +

Deployments Monitored	Quality Alerts	Fairness Alerts	Drift Alerts
1	1	1	0

ⓘ Quality and Fairness metrics update every hour. Drift metrics update every 3 hours.

Watson Machine Learning

credit-risk-deploy

Issues

2

QUALITY BIAS

Quality	Fairness	Drift
0.74	93%	5%
1 alert	1 alert	

Production

Evaluated 48 seconds ago

Congratulations! You have completed the Lab!!!

- ✓ Imported a machine learning model

- ✓ Deployed the model
- ✓ Provisioned Watson OpenScale
- ✓ Configured the payload logging database and Machine Learning provider
- ✓ Scored Data
- ✓ Prepared the Deployed Model for Monitoring
- ✓ Configured Payload Logging
- ✓ Configured Quality Monitoring
- ✓ Configured Fairness Monitoring
- ✓ Configured Drift Monitoring
- ✓ Submitted Feedback and Viewed Quality Metrics
- ✓ Scored Data and Viewed Fairness and Drift Metrics
- ✓ Explained a Transaction.