gitops   **cicd**   **continuous-delivery**   **git**
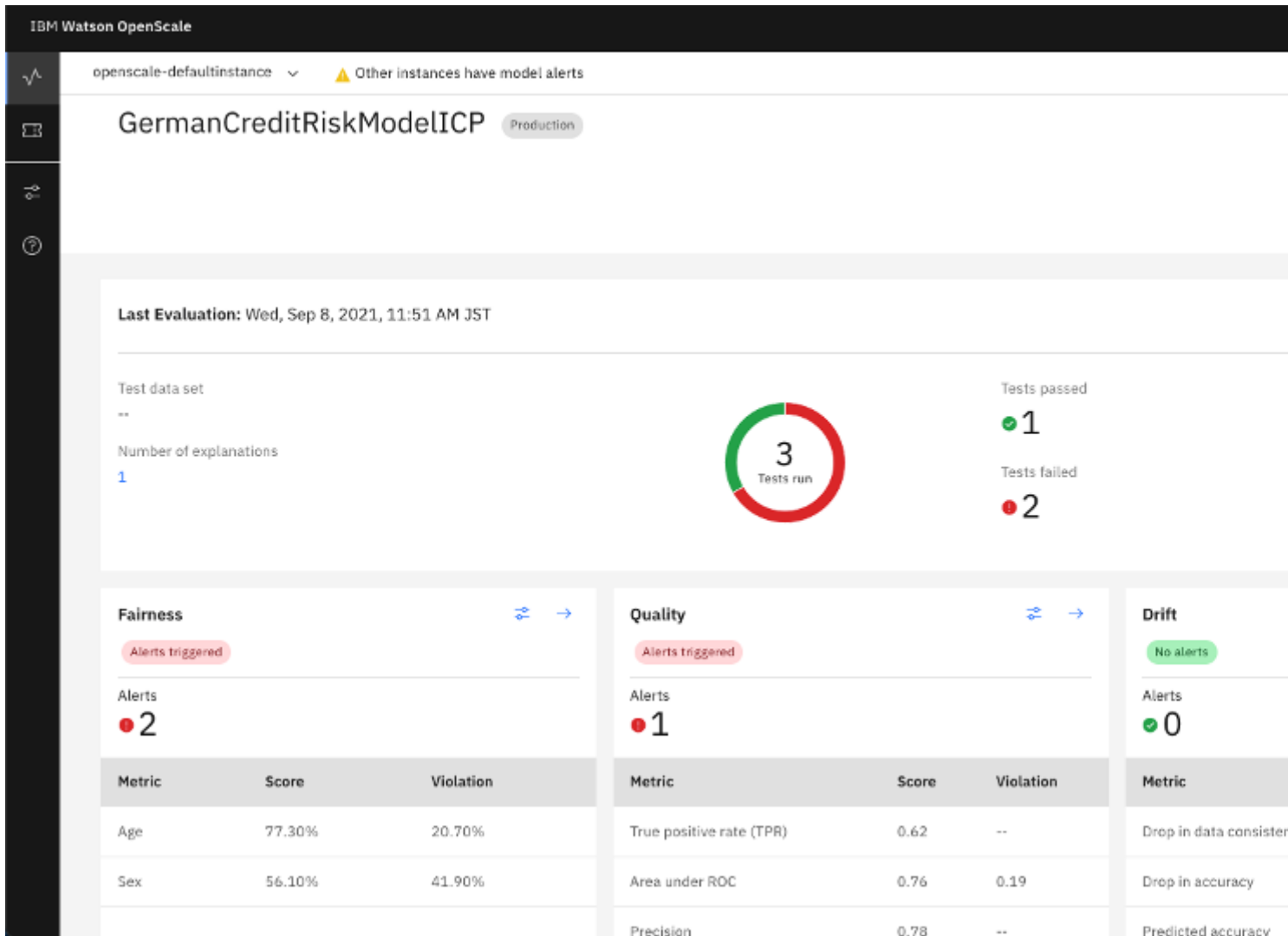
Watson OpenScale

After you deploy a machine learning model, the work doesn't stop. To guarante
functioning in production as expected, you must have a plan for monitoring the
updating it as needed. As part of your end-to-end MLOps process, consider IBM
evaluate model deployment to make sure they are fair, accurate, and performin

When OpenScale is installed or provisioned as part of your Cloud Pak suite, you
details for a deployment, then run scheduled evaluations that measure dimensi
thresholds you set. For example, if you want to test whether predicted outcome
various age groups, you can configure the Fairness monitor to evaluate the out
group, such as young adults, and compare the results to the age group most lik
results. If the results deviate more than a threshold you specify, you will get an
require attention. The dimensions you can test are:

- Fairness: Configure a monitor for fairness to check if your model produces l
  different groups, like gender or race. Set thresholds to measure predictions
  group compared to a reference group.

- Quality: Configure a monitor for quality to assess your model's performance
  data. Set quality thresholds to track when a metric value falls outside an ac

- Drift: Configure a monitor for drift to ensure your deployments are up-to-da
  feature importance to determine the impact of feature drift on your model.

- Explainability: Configure explainability settings to understand which featur
  model's predictions. Different methods like SHAP and LIME are available to

All of the evaluation results can be reviewed and monitored in a single dashboa
example:

IBM Watson OpenScale

openscale-defaultinstance ⌄   ⚠ Other instances have model alerts

GermanCreditRiskModelICP `Production`

Last Evaluation: Wed, Sep 8, 2021, 11:51 AM JST

Test data set
--

Number of explanations
1

3 Tests run

Tests passed
✓ 1

Tests failed
● 2

Fairness	⇄ →		Quality	⇄ →		Drift
Alerts triggered			Alerts triggered			No alerts

Alerts ● 2			Alerts ● 1			Alerts ✓ 0
Metric	Score	Violation	Metric	Score	Violation	Metric
Age	77.30%	20.70%	True positive rate (TPR)	0.62	--	Drop in data consister
Sex	56.10%	41.90%	Area under ROC	0.76	0.19	Drop in accuracy
			Precision	0.78	--	Predicted accuracy

For more information check out the official documentation or the example Note

In this use case we use Watson OpenScale to automatically verify:

- whether the model is performing at a constant high accuracy. If performano threshold we set, an alert is triggered.

- whether the test data produces output that is similar to the training data. If deviation, we are alerted that it might be time to retrain the model.

- whether the model is discriminating against a particular group. In this case for older customers to ensure they are being treated fairly as compared to t

Results are automatically checked by Watson OpenScale whenever new data is one of the checks fails, an alert will be send to the responsible person so that s

IBM

IBM Watson OpenScale

openscale-defaultinstance ⌄   ⚠ Other instances have model alerts

Dashboard / GermanCreditRiskModelICP / Evaluations

Evaluations

Model ID: b490ddfb-2c6a-4a16-a801-32e4587c600c   Created date: 9/8/2021   Configure monitors ⇄

Fairness

Age ⚠

Sex ⚠

Quality

Area under ROC ⚠

Area under PR

Accuracy

True positive rate (TPR)

False positive rate (FPR)

Recall

Precision

F1-Measure

Logarithmic loss

Drift

Drop in accuracy

The model's propensity to deliver favorable outcomes to one group over another. Learn more.

Time frame   Date range

Hourly ⌄   Past week ⌄

FAIRNESS SCORE

110%

90%

70%

Sep 2, 12:00 AM   Sep 3, 1:00 PM   Sep 5, 2:00 AM   Sep 6, 3:00 PM   Sep 8, 4:0

TIME [PAST WEEK]

update: October 27, 2023

Analytics

Predictions by Confidence

Chart Builder

ors: Dominik Kreuzberger, Julianne Forgo, Przemek Czuba, moritzscheele