

gitops cicd continuous-delivery git

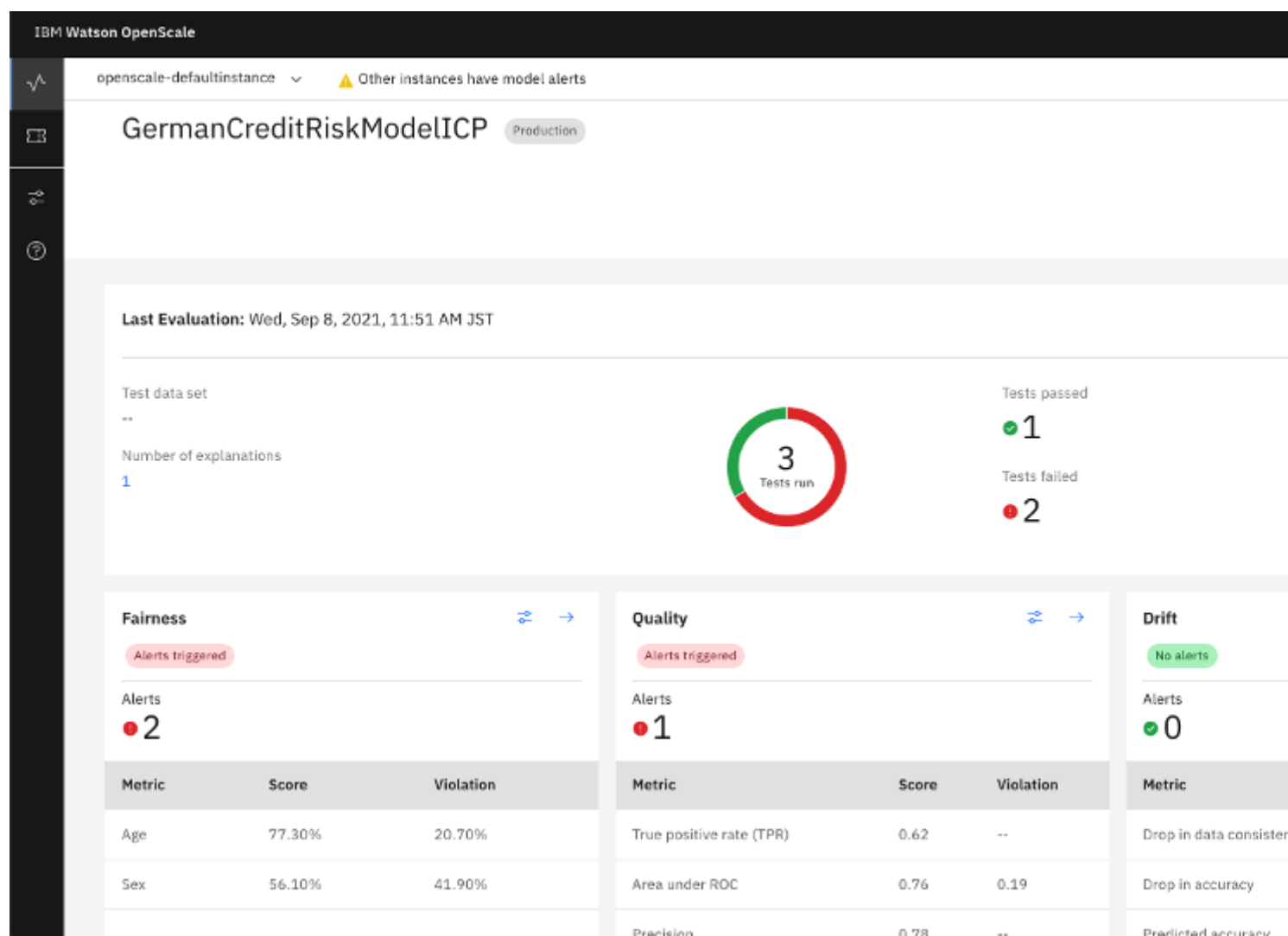
Watson OpenScale

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

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

- **Fairness:** Configure a monitor for fairness to check if your model produces similar results for different groups, like gender or race. Set thresholds to measure predictions for each group compared to a reference group.
- **Quality:** Configure a monitor for quality to assess your model's performance against new data. Set quality thresholds to track when a metric value falls outside an acceptable range.
- **Drift:** Configure a monitor for drift to ensure your deployments are up-to-date with new feature importance to determine the impact of feature drift on your model.
- **Explainability:** Configure explainability settings to understand which features impact the model's predictions. Different methods like SHAP and LIME are available to help you understand the model's predictions.

All of the evaluation results can be reviewed and monitored in a single dashboard. For example:



For more information check out the [official documentation](#) or the [example Notebooks](#)

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

- whether the model is performing at a constant high accuracy. If performance drops below the threshold we set, an alert is triggered.
- whether the test data produces output that is similar to the training data. If there is significant 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, we ensure that older customers are being treated fairly as compared to younger customers.

Results are automatically checked by Watson OpenScale whenever new data is processed. If one of the checks fails, an alert will be sent to the responsible person so that they can take action.

