**Lab 105** (**Run and interpret evaluation**)

In the previous lab, we configured OpenScale. In this lab, we will configure various model evaluation monitors, and then run an evaluation to calculate these metrics.

1. **Configure fairness monitors**
   1. From the list of Evaluations on the left, click on fairness.
   2. Note the description of fairness in the centre of the screen, which gives a good definition of what the monitor is evaluating. Click on the edit icon in the configuration tile.

A screenshot of a computer

Description automatically generated

* 1. You will manually configure the fairness files. Leave configure manually selected and click Next to proceed. To properly monitor a model for unfair bias, you must specify which model outcomes are favourable and which are unfavourable. For our use, favourable class is No Risk and non-favourable class is Risk. Click Next.

A screenshot of a computer

Description automatically generated

* 1. Set the minimum sample size to 50 and click next.

A screenshot of a computer

Description automatically generated

* 1. Leave the selected monitored metrics set to Disparate impact and click Next.

A screenshot of a computer

Description automatically generated

* 1. Leave the lower and upper thresholds for Disparate impact set to their defaults, and click Next.

A screenshot of a computer

Description automatically generated

* 1. In the next screen, select the fields that you need to monitor for fairness. For now, retain the default values suggested by OpenScale (after analyzing the training data). Click Next.

A screenshot of a computer

Description automatically generated

* 1. In the next screen, select the groups which are to be monitored against the reference group. Here, 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 screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* 1. Use the default alert threshold (80), and click Save to finish configuring the fairness monitor. It may take up to a minute for the configuration to save, at which point you will be returned to the model settings screen.

1. **Configure the quality monitor**
   1. From the list of evaluations on the left, click on Quality.
   2. Click the edit icon on the Quality thresholds tile.

A screenshot of a computer

Description automatically generated

* 1. Click the Edit icon on the Quality thresholds tile.
  2. Leave the default lower and upper threshold values as they are. You can click the Information icon to the right of each value for more information on how it is calculated. Click Next.

A screenshot of a computer

Description automatically generated

* 1. Set the Minimum sample size value to 50. Click Save to save the quality configuration.

1. **Configure the explainability service**
   1. In the Explainability section, click on General settings, and in the explanation method tile, click on the Edit icon.

A screenshot of a computer

Description automatically generated

* 1. Two different methods are available for explanations: Shapley Additive Explanations (SHAP) or Local Interpretable Model-agnostic Explanations (LIME). As described in hint that appears when you click the Information box, SHAP often provides more thorough explanations, but LIME is faster.
  2. Leave the LIME method selected and click Save.

1. **Run an evaluation**
   1. Return to the Insights Dashboard by clicking on the Dashboard link the upper left, and click on the tile for the model you configured (german-credit-risk-testing)
   2. Click on the Actions button to open the Actions menu, click on Evaluate now.

A screenshot of a computer

Description automatically generated

* 1. Import test data screen pops up, select import from CSV file and then browse the test file from your local system, and keep the option of Test data includes output as unchecked.

A screenshot of a computer

Description automatically generated

* 1. Click Upload and evaluate.

After some time, the results of the evaluation can be seen in the dashboard.

1. **Viewing and interpreting the results**
   1. Take a moment to review the results of the evaluation. Note that, based on the content of the random sample of the evaluation data, your results will vary each time you perform the evaluation.
   2. Review the different metrics in the quality table. In the screenshot below, notice that, if the measurement falls below the alert threshold set when you configured the quality monitor, the amount will be listed in the Violation column of the table.

A screenshot of a test results

Description automatically generated

* 1. Beyond meeting standards for quality and fairness, AI models in many cases are required to provide explanations into the decisions or predictions they make. This can be viewed in watsonx.governance by clicking on the arrow icon in the fairness tile, and then selecting view payload transactions.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* 1. From the table of transactions, click one of the Explain prediction links. The explainability service will use the LIME algorithm to generate a detailed explanation, which can take a few minutes to run.

A screenshot of a group

Description automatically generated

* 1. Once the explanation has been generated, scroll down to the graph, which shows the influence different features had in the model's outcome. Features in blue increased the final score, while those in red decreased it. For classification models, blue features contributed positively to the model's confidence in the prediction, while those in red decreased the confidence.

A screenshot of a computer

Description automatically generated

* 1. Click on the Inspect tab. On this tab, you can alter values associated with the record and re-submit it to the model to see how the final risk calculation changes. This can be useful for understanding how the model is working, or if a policyholder is looking for ways to decrease their risk assessment.

A screenshot of a computer

Description automatically generated