

Broad Functionality: **Data Drift**

Principle: **Model Monitoring and Bias**

Understanding Functionality: As the difference between data in the training and production changes over time, the performance of the model gets degraded, this concept of change in the distribution of data from training and production is called Data Drift. Since it is vital to keep track between similarities/differences between training data and production data, MLflow should log and profile the data in production to evaluate metrics like KL Divergence, Hellinger Index, and Population Stability Index.

Does MLflow supports the functionality *directly*: **No**

Does MLflow supports the functionality *indirectly*: **Yes**

- Explanation: The MLflow is highly flexible in integrating other open-source libraries. “Whylogs” is an open-source and lightweight library, that can be integrated with MLflow, to capture the data quality and monitoring of distribution in the features. Whylogs profile the data from train and production and keeps the track of changing quality. That enables Whylogs to measure the magnitude of data drifting in production.
- Extra remark: Profiling is basically the creation of a statistical summary of the dataset. Which can be used to compare different datasets, i.e., Training and Production. However, many of the metrics can be visualized using the Whylogs platform, through subscription-based usage of the library.

Screenshots for MLflow supports the functionality *directly or indirectly*: **NA**