Mercury

Mercury.robust

robust test

BBVA

Learn more at Robust Tutorial

mercury-robust is a framework for performing robust testing of models and/or datasets.

It provides a series of predefined and configurable tests cases to ensure the robustness or their ML pipelines. For example, you can quickly check if your model is discriminating a collective or check if the training of your model is reproducible

DataTest

Same Schema

This test ensures that the DataFrame has the same columns and feature types as the ones specified in the DataSchema.

Data Drift

Checks that the individual feature distributions have not changed significantly between a reference DataSchema and a pandas.DataFrame.

Linear Combinations

Checks that you have no redundant or unnecessary columns in your pandas.Dataframe.

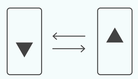
No Duplicates

Checks that you don't have repeated samples in your dataset, which can add bias on your performance metrics.

Noisy Labels

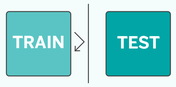
Checks that labels of a dataset are of a minimum quality. We consider low quality labels when we have a high number of wrongly labeled samples or when separation between labels is not evident.

Cohort Performance



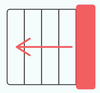
This test compares a particular metric (e.g. 'accuracy') between several groups specified by a categorical variable in your pandas.DataFrame ('group_col').

Sample Leaking



Checks that you don't have repeated samples in your dataset, which can add bias on your performance metrics.

Label Leaking



Checks that you don't have any feature leaking information about the target variable.

ModelTest

Model Simplicity

This test compares the performance of your model to a simpler baseline (by default a linear model, although you can specify your custom baseline).

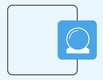
Model Reproducibility

Trains a model twice and checks that predictions (or a certain metric) of the two versions are not too different.

Drift Metric Resistance

This test adds artificial drift to a reference dataset and tests your model on it. If a chosen metric (e.g. 'accuracy') changes above the threshold, the test will fail, indicating that your model is weak against drift.

Drift Predictions Resistance



This test adds artificial drift to a reference dataset and tests your model on it. If lots of predictions change the test will fail, indicating that your model is weak against drift.

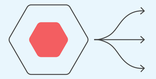
Feature Checker

This test estimates the feature importance of your features and retrains your model removing the least important ones one step at a time.

Tree Coverage

This test only works with tree-based models (mainly the scikit-learn ones). It checks that, once you have a model trained, given a test dataset, the samples "activate" a minimum amount of branches in your tree(s).

Classification Invariance



This test checks that your classifier has a minimum of robustness against data perturbations (defined by you). The test receives two versions of the same dataset: one without corruptions and other with a certain level of corruption.

TEST SUITE

Holds a set of tests so you only have to run them once