




Data drift with Frouros

Ignacio Heredia (iheredia@ifca.unican.es)
on behalf of Jaime (cespedes@ifca.unican.es)



Data drift

For example:

- you train a fish classifier AI model
 - you deploy a webcam underwater
 - you start classifying the images, but the webcam gets progressively dirty
-  *when/how do you know you can not longer trust the predictions?*

Data drift can be defined as the process of trying to detect a significant change in the concept previously learned by a model (**concept drift**), or a change related to the feature/covariate distributions (**data drift**) that can end up producing a performance decay in model's performance.

Data drift

Typical workflow

- define a clean reference dataset (ie. the dataset you used to train your production model)
- train a drift detector on this dataset
- define a test dataset (ie. the dataset your production model is being evaluated on)
- evaluate the test dataset on this drift detector
- make a decision based on the returned p-value

Our solution - Frouros

- **Github**

<https://github.com/IFCA-Advanced-Computing/frouros>

- **Docs**

<https://frouros.readthedocs.io/en/latest/>

- **Tutorial:** Detecting image data drift with MNIST

https://frouros.readthedocs.io/en/latest/examples/data_drift/MMD_advance.html

Usecase exploration

Exploratory work has been carried out in the EMSO usecase in the iMagine project, where we want to detect when an underwater cam gets dirty and needs to be cleaned again.





Questions