

Service ID S00088

Location Remote



Assessment of dataset quality

Provider service

Laboratoire National de Meterologie et d'Essais (LNE)

Link to content

<https://agrifoodtef.eu/catalogue-of-services/assessment-dataset-quality>

Type of Sector

Arable farming, Food processing, Greenhouse, Horticulture, Livestock farming, Tree Crops, Viticulture

Accepted type of products

Data, Design / Documentation

Type of service

Conformity assessment, Data analysis, Desk assessment

Description

With the development of AI models in most sectors of society, the importance of the data used to train and test these models has been an important concern for the AI supplier. Understanding the quality of the datasets used, especially in sectors such as the agricultural and agrifood domains where data can be rare, is thus a prerequisite to ensure the correct development of AI models and build trust regarding these approaches for society. This service aims to answer this challenge by assessing the quality of datasets provided by the client. The features of the datasets that will be analyzed are based on agreed definition by the AI community: completeness, balance, diversity, accuracy... The dataset's quality assessment will begin by establishing the list of the system's influencing factors necessary for the constitution of a database. These factors will be identified both by using LNE expertise and outside sources, such as agrifood-relevant experts. Using these factors, LNE will analyse the distribution (or weight) of the various influencing factors in the databases and compare it to that under normal conditions of use in order to assess the representativity and balance of the dataset.

How can the service help you

The service helps the TEF client that wants to ensure the quality of a dataset and understand the features of the data, thus answering the requirements regarding data quality of the AI Regulation.

How the service will be delivered

Depending on the needs of the customer, it will be possible to focus the analysis on specific features of the data. The scope of the service will be defined with the customer. The dataset should be provided by the customer or at least available for open use. The dataset can also be provided through another service of the TEF.

Service customisation

A report will be provided to the customer at the end of the service, explaining the methodology and detailing all the analysis and conclusions of the study. The service generally takes around one month to be provided once the data has been delivered to LNE. Work is generally done on local servers in the LNE infrastructure located in Trappes, France, but in case of confidentiality requirements, it is possible to adapt the work.