Service ID   S00178

Location   Italy, Remote

Design of evaluation metrics for digital testing

Provider service

Politecnico di Milano (POLIMI), Università degli Studi di Milano (UMIL)

Link to content

https://agrifoodtef.eu/catalogue-of-services/design-evaluation-metrics-digital-testing

Type of Sector

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

Accepted type of products

Data, Design / Documentation

Type of service

Test design

Description

Any test activity involves three main components, i.e.: environment (where the tests take place), protocol (defining what activities are executed and how) and evaluation metrics (used to assess the results of the tests). This service concerns the last element; its goal is to design the best metrics to evaluate the performance of digital systems such as (for instance) AI models or Computer Vision software. The digital environment and the testing protocol metrics can be designed -if required- via services S00176 and S00177. Our team will identify and define with customers the most adequate set of quantitative metrics to assess the outcome of the digital testing activities. In order to ensure the relevance of the metrics with respect to the real-world use cases, the team will involve engineers and agronomists. The metrics will be adapted not only to the task that the digital system under test (e.g., a piece of software) is designed to perform, but also to the features of the data used for the tests. For instance, a customer that has developed a machine incorporating an AI model will be interested in testing the model on data generated by their own machine: the performance metrics will therefore need to be adapted to the specific features of those data.

How can the service help you

Building a digital system (e.g., an AI model) that solves a problem and designing the mathematical and data-processing operations needed to analyse data collected during digital experimentation with the system to evaluate its performance are two very different activities and involve very different competencies. Additionally, evaluation metrics are crucial to the identification of issues and ways to improve the performance of the system; therefore, their choice has a strong impact on product development.

This service supports customers who developed a digital solution, such as a piece of software, in designing the evaluation metrics necessary to process experimental data to evaluate system performance and suitability for the task. At the end of the service, customers are provided with a set of metrics tailored to their own system, data and necessities, which can be used to assess quantitatively the performance of their system. The metrics will be described in a report explaining clearly how and to what data to apply them, what aspects of system performance they are designed to capture, as well as any limitation in their applicability or significance.

If required, agrifoodTEF can support the customer in designing also the digital environment and testing protocol for the tests

How the service will be delivered

This service description is intentionally generic. Every instance of this service is, in fact, customised to adapt it to the needs and requirements of the specific customer. The following is an example of a service instance.

Example service: The customer is interested in measuring the capability of a computer vision software module to discriminate weeds from crops in RGB images. Consequently, a set of objective evaluation metrics gets designed to describe the quality of predictions, including classification accuracy, precision, and recall scores for all "crop" and "weed" instances observed. Metrics are designed considering the specific features of the data used for testing, provided by the customer; data have been collected using their own machine in relevant environments. The customer is also interested in measuring indicators such as "How often is Matricaria Chamomilla mistaken for bean crops?" Thus, quality of prediction scores are also computed not only at the macro level ("crop" vs. "weed") but also for individual plant species of interest to the customer ("bean" vs. "Matricaria").The customer is provided with the complete specifications of the designed performance metrics and with information about what data to use in order to optimise the significance of the experimental tests. Optionally, agrifoodTEF can support the customer in generating such data (e.g., through a dedicated experimental data collection campaign – via service S00113 – followed by a suitable data annotation – via service S00115 – or by augmenting existing data with additional examples – via service S00115).

Service customisation

The duration of this service is on average 1-3 weeks.The first phase involves one or more interviews, in person or remote, where the customer provides information about the features of the system(s) to be tested, the performance elements of interest and the type of data to be processed for performance evaluation.

Subsequently we design the evaluation metrics and check their compliance with the requirements by executing preliminary processing tests on data fragments (possibly provided by the customer, under NDA if needed).

During this phase we may provide the customer with feedback about data quality and suitability for the purpose.Should extensive data validation be needed by the customer, they may choose to leverage a separate agrifoodTEF service dedicated to such activities, i.e., S00115 (data validation and processing services).

S00115 also supports the customer in the remediation of data quality shortfalls and in devising strategies to improve data collection practices.At the end of the service, the customer receives a report with the design of the performance metrics and the outcomes of the preliminary processing tests.