Service ID S00290



Location Italy, Remote

Data labelling

Provider service

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

Link to content

https://agrifoodtef.eu/catalogue-of-services/data-labelling

Type of Sector

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

Accepted type of products

Data

Type of service

Data augmentation

Description

This service concerns the curation of ground truth annotations for testing and experimentation data. More specifically, data labelling concerns the association of portions of data (e.g., image regions) to descriptions that provide information about the significance of those portions. In particular, labelling is required whenever the data are meant to be used for model training. We also provide labelling services for domain-specific data attributes, such as disease indicators and their confounding factors, by experts in agronomy. The labelled data will be accompanied by a report synthesising quality metrics and statistics like the number of data points annotated, percentage of the full set covered for the annotations, number of annotators and the inter-annotator agreement.

How can the service help you

Data labelling is a necessity whenever there is the need to train an AI model on the data. However, labelling can be (depending on data quantity) a very labour-intensive process and often requires specialised know-how about the use case that the data are associated with.

This service offers to the customer access to an expert team comprising engineers trained in the development and use of AI models and agronomists with extensive knowledge of real-world use cases.

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 promptly identifying the emergence of the Peronospora (downy mildew) disease in vineyards. Peronospora symptoms can be detected by inspecting changes on the leaf surface (appearance of small spots, gradual changes in the leaf colour). The customer has already implemented a computer vision algorithm to classify leaves as healthy or unhealthy from images. However, additional data are required to improve the performance and robustness of the solution. These data have been collected in the field via service S00113, but the customer requires that they be thoroughly annotated by expert agronomists who can recognise the presence of Peronospora symptoms at the level of individual leaves.

To fulfil the request from the customer, service S00290 is executed by first identifying segmentation masks as the most suitable annotation format.

Service customisation

The duration of this service heavily depends on the type of labelling that the customer requires, as well as on the amount of data to be processed. The service is delivered starting from the definition, together with the customer, of the most suitable labelling procedure based on the requirements of the customer's use case. This initial analysis will also provide an estimate of the time required to execute the service.

In this definition phase, we will select the most appropriate type of annotations to produce (e.g., pixel-level annotations of images vs. a single label for the whole image), as well as the level of granularity of labels (e.g., plant colour, variety, crop class). We will also agree with the customer on the expected annotation format (e.g., JSON, CSV) and the annotation tool to be used (e.g., in case the customers want to label additional data by themselves in the future).

Once the definition phase of the service is complete, agrifoodTEF will proceed with the setup and execution of the labelling. The final outcome of the service is the labelled datasets, accompanied by a report with evaluations of the data based on quality metrics.