

# Datasheets for Datasets

## Motivation

Fashion industry motivates more consumption to pursue the growth of business. As the result, the quality of garment is not going to be more durable, better or more resilience but to be cheaper and faster to make. Thanks to globalisation with cost pressure, the quality of fibre is very low and make it harder to recycle into new garments.

A recognition to go towards sustainable fashion has become more popular and it concerns more than just addressing fashion textiles or products. It addresses the entire manner in which clothing is produced, who produces it, and how long the life span of a product is before it reaches the landfill. This sustainable movement combats the large carbon footprint that the fashion industry and fast fashion have created by reducing greenhouse gas emissions. Reducing the environmental impact of fashion can combat air pollution, water pollution and overall climate change that could possibly prevent millions of premature deaths over the next century.

Today, there are more approach to reach the sustainability by using AI. The technology that is widely used is deep learning method with optical character recognition (OCR) for image classification. In Zelus Image Project, the goal is to use AI to detect the type of clothes with user friendly interface. The project has two important parts such as deep learning model and API. We are using Lightning Flash model to train the dataset and FastAPI for API process. In deep learning model, we classified the label from 491 categories to 10 categories to simplify the training process and to detect general type of clothes and accessories.

## Composition

The dataset represents 4000 pictures of clothes and it is representative of the larger set of our Vinted database which contain more than 100,000 pictures. Consider to have smaller size of categories, we minimised the data manually into 10 categories with 400 pictures each. Then, the dataset is divided into train and val folder where it has 1/5 of the train folder.

Labelling is a crucial part of image classification. In the project, the label is associated with each instance such as dresses/1367721161.jpeg or sweater/1367693746.jpeg.

Then, we trained the dataset using transfer learning with Lightning Flash. We created a model from the ImageClassificationData class, set the trainer with certain number of epoch, GPUs and the precision. Later on, we did finetuning the model and saved the model before we use it in API.

In the API process, it became more challenging to make friendly user interface. Some features added so users can upload image and take a picture and the model can recognise with high accuracy if the picture is t-shirt, jeans, glasses, shoes, or any other type of clothes.

Source of dataset : <https://github.com/pierre-warnier/vinted>

Source of model and API : <https://github.com/TimStrauven/Zelus-api>

## **Maintenance**

The dataset will not be update and it will not continue to be supported/hosted/maintained. If others want to build on/contribute to the dataset, please read the documentation and send the message to Tim Strauven

## **The Team**

The project is done by Biniam Berhe, Fortuné BT, Rosyidah Nadiah, and Tim Strauven on behalf of Bencode Brussels 2022. All Bencode Brussels students was involved in data collection process and preprocessing.