

T5 Boot camp

Data science

Project written description

Fashion classification model

mohammad alajmi
mohammad almalki

Instructor:
Mejdal Al-Qahtani

abstract:

In this project we are aiming to build an intelligent system that differentiate between the clothes, we are aiming to train the model on greyscale images because it could be represented as a series of data.

The data set should contain a different types of clothes such as tops, pants, sneakers etc.

The main goal of this project:

The main goal of this project is to build a classification model that classify the clothes by input the images of the clothes.

Design:

The used data set in this project is derived from different websites, such as red bubble and Namshi, the data collected using web scraping library called beautiful soap, it is about classifying the cloth based on 9 labels, the data has about 60,000 thousand rows before the cleaning and 785 columns.

Data:

Each row is a grayscale image, each image is 28 pixels in height and 28 pixels in width, for a total of 784 pixels in total.

The training and test data sets have 785 columns, The first column consists of the class labels, the rest of columns represent the values of each pixel, The pixel-values are integers between 0 and 255, Each value is the darkness of the pixel

The labels are:

- 0 T-shirt/top
- 1 Trouser
- 2 Pullover
- 3 Dress
- 4 Coat
- 5 Sandal
- 6 Shirt
- 7 Sneaker
- 8 Bag
- 9 Ankle boot

Algorithms:

in this project we used different models to achieve this goal mentioned as follows:

1. KNN
2. SVM
3. NN
4. CNN

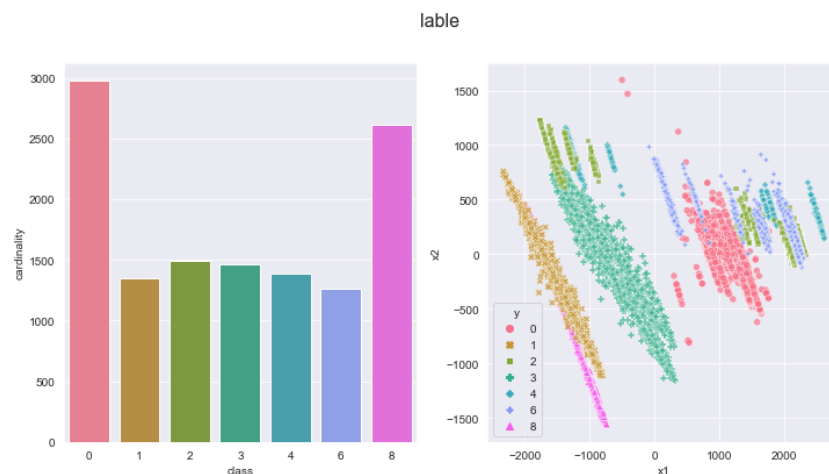
The used tools:

I have used different tools to achieve this goal such as:

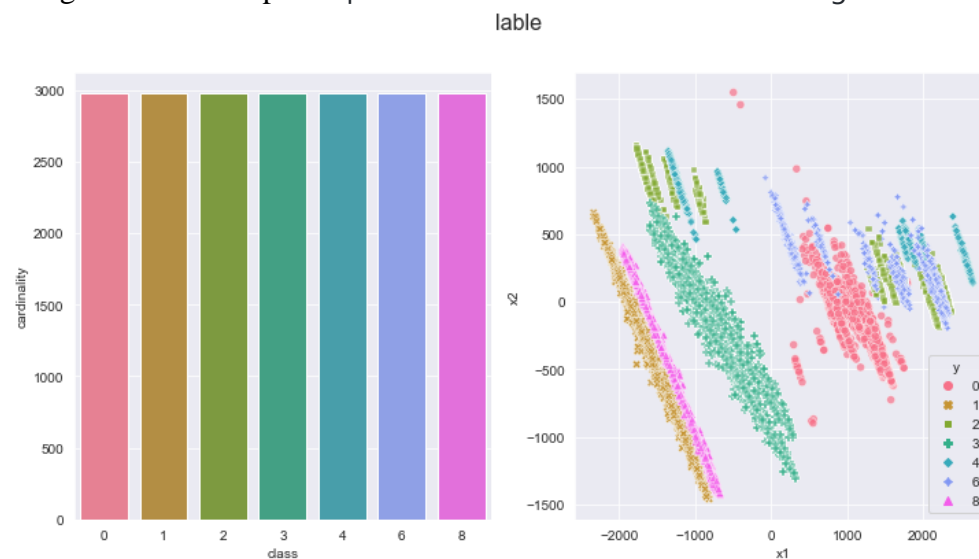
- 1- Python programming language
- 2- Numby library
- 3- Pandas' library
- 4- Matplotlib library
- 5- SeaBorn
- 6- Sklearn
- 7- BeautifulSoup
- 8- Request library
- 9- PIL
- 10- Glob

Communication:

In figure 1 the used plot represents the Dataset before of balancing it.



In figure 2 the used plot represents the Dataset after of balancing it.



In figure 3 the used plot represents the confusion of the dateset between the true labels and the predicted labels

