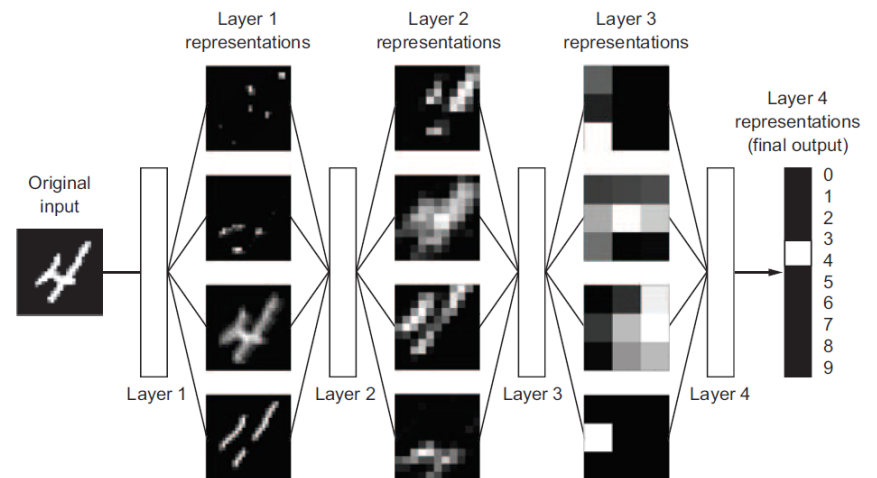


Suggested Project

Identifying people from images of their faces



Objective

The objective of the exercise is to build a model that allows the identification of different people from an image of their face. The model should correctly classify any new image from one of the selected people.

Winona Ryder



Jean Chretien



Carlos Menem



Ariel Sharon



Alvaro Uribe



Colin Powell



Recep Tayyip Erdogan



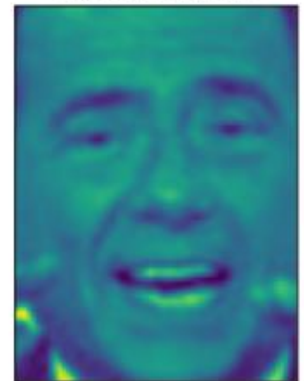
Gray Davis



George Robertson



Silvio Berlusconi



Planning

A four-step planning is proposed to develop a complete modelling project. Each phase will emphasize an important aspect of conducting AI consulting projects.

Proposed development phases for the project

Think strategies



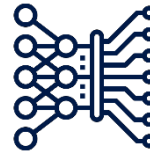
- Ensure the clear understanding of the objectives
- Design of the strategy for its resolution
- Identification of IT resources to be used
- Detail working plan for ensuring the delivery at the end of the week

Research options



- Downloading and exploratory analysis of the database
- Understanding the data
- Preprocessing of data
- Research of the best models and metrics to use. **At least one type of neural network must be selected.**

Find solutions



- Selection of the best approach
- Development of the models
- Validation and final adjustments
- Understanding the output

Communicate results



- Development of the presentation for the communication of the results
- Preparation of data and graphic support
- Thinking of possible improvements and next steps



Thinking the strategy

Organization of the workflow and workload is paramount to solve the problems in a timely manner. It is necessary to define a clear strategy to achieve success in the realization of the project.

Main questions to ask

What is the objective?

- Identify different people from an image of their face. That is, create clusters from the dataset that contain photos of the same person.

Where is the data?

- The data is the Labeled Faces in the Wild (LFW) people dataset.
- You can access it throughout the sklearn datasets:
from sklearn.datasets **import** fetch_lfw_people

How to solve it?

- Design a strategy to solve the problem. It should include:
 - Your initial approaches.
 - A schedule with deadlines for the data analysis, model selection, model development...
 - The resources you will need, such as libraries or more data.

Recommendations

- Create a new environment for the project.
- **Document your code.**
- Check the official documentation of functions and libraries.





Researching the options

The main objective is to create clusters of images and there are plenty of clustering algorithms. Try and compare the output of several algorithms to decide which is the most suitable.

Clustering algorithms

Tasks

- Which algorithms and goodness of fit metrics could be used?
- What conclusions could we obtain from this preliminary models?
- Preprocessing the data (for example, applying PCA) makes the model better?
- Are they useful for the objective?
- Compare the main characteristics of the models used





Finding the solutions

Choose the models that work best for you and work on them to improve the results. Try to modify the input hyperparameters of the model and to work more on the the preprocessing.

Selecting models

Tasks

- Explain why you choose these models and compare results.
- Explain the preprocessing you are going to do and why you think it is the best.
- Explain the benefits and drawbacks of applying PCA.
- What does PCA make to images?
- What number of clusters is the most adequate?
- What other models serve to solve the problem?





Communicating results

Submit a report explaining your results. Use images, data, text and anything else you think will be helpful to understand your results. No assumptions: everything you say must be based in facts. Attach to your delivery the code of the final models worked on.

Understanding the results

Tasks

- Explain the models you used to explore the data and the preprocessing you have done.
- Think about how you would explain your result to someone who doesn't know coding.
- If you have a model based on a linear regression with an accuracy of 0,80 and another model based on neural nets with an accuracy of 0,85, what model would you use to solve a major problem of the company?
- Explain what you would need to improve the accuracy of your model and how.
- Explain what is overfitting and underfitting and how you have avoided it in your model.

