

Projet Python for Data Analysis

- Iris dataset

• Théo Philippe

• DI4

I/ Dataset

- The dataset is composed by four columns which all describe flower dimensions, such as information about sepal and petal. Thus they are three flower which are considered here :

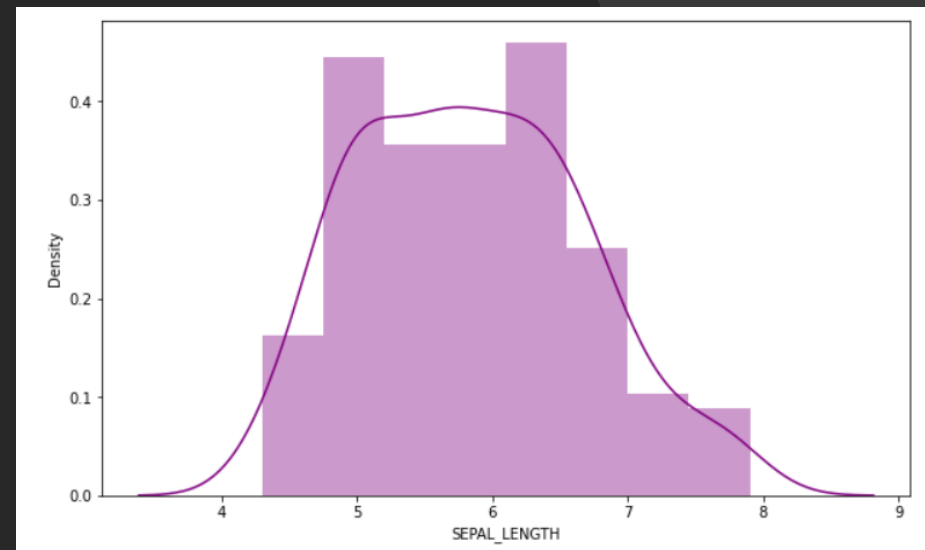
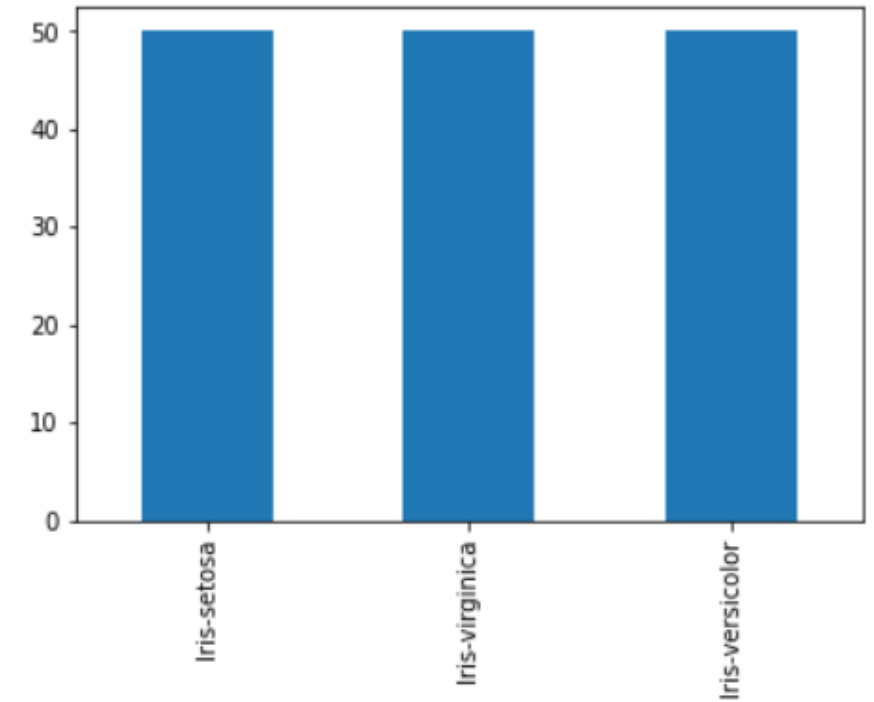
Iris Setosa – Iris Virginica – Iris Versicolor

- This dataset contains 5 columns and 150 lines. Furthermore, the fact that they are a few columns made us think that we need to consider them all.

	SEPAL_LENGTH	SEPAL_WIDTH	PETAL_LENGTH	PETAL_WIDTH	FLOWER
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa

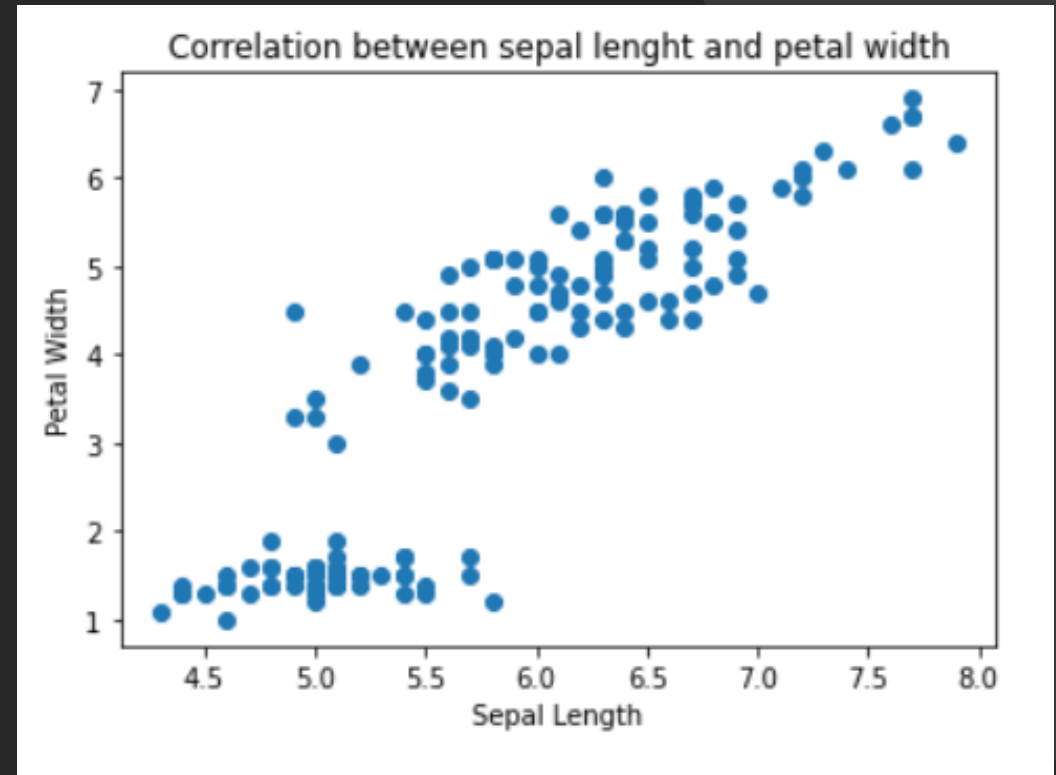
II/ Visualization

- First we consider the quality of our dataset, thus we can see how many train line in this dataset concern each flower. Hence, we notice that flowers are equally distributed.
- However, a quick visualization of a flower feature can be ressourceful for our analysis. Visualization that may us see that they are no direct cluster in this only variable that can yet show us the different flowers.



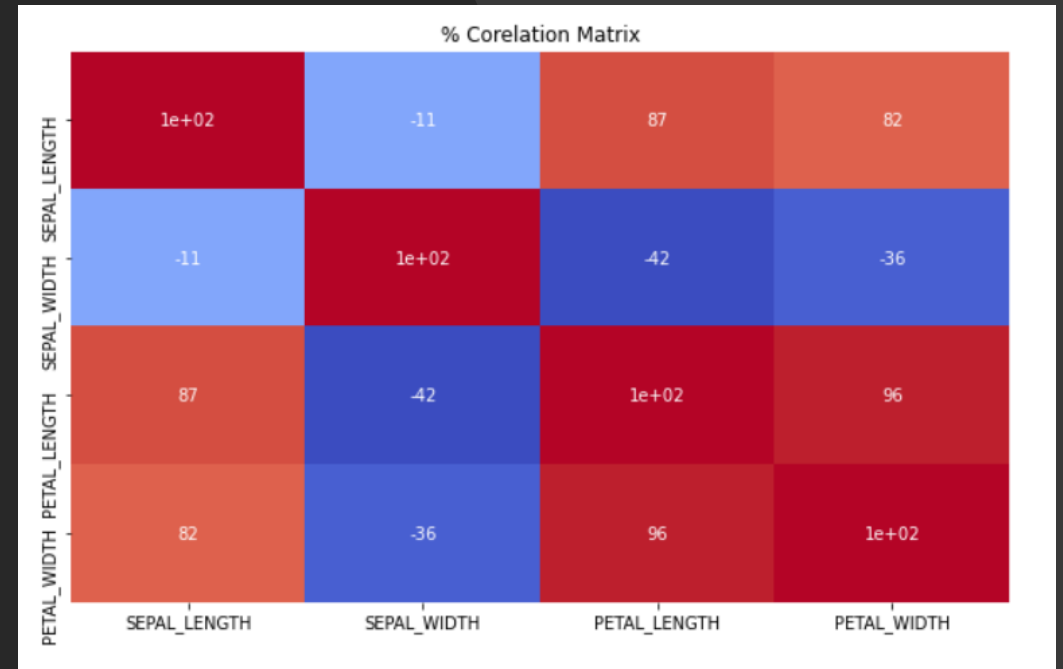
II/ Visualisation

- Besides, we show the correlation between petal and sepal dimensions, and notice a quite correlation between those two. However, there are two groups which seem to step up from this visualization.



III/ Modélisation

- Even if the small number of columns allows us to consider them all, the plot of the correlation matrix helps us to better our understanding from the dataset and the further modelization.



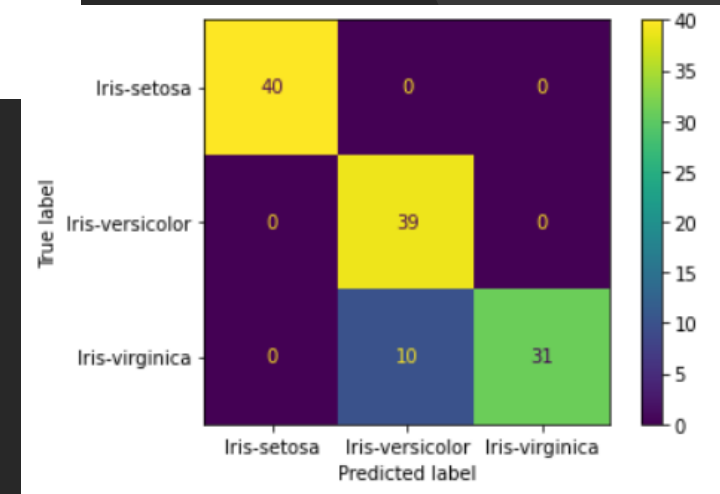
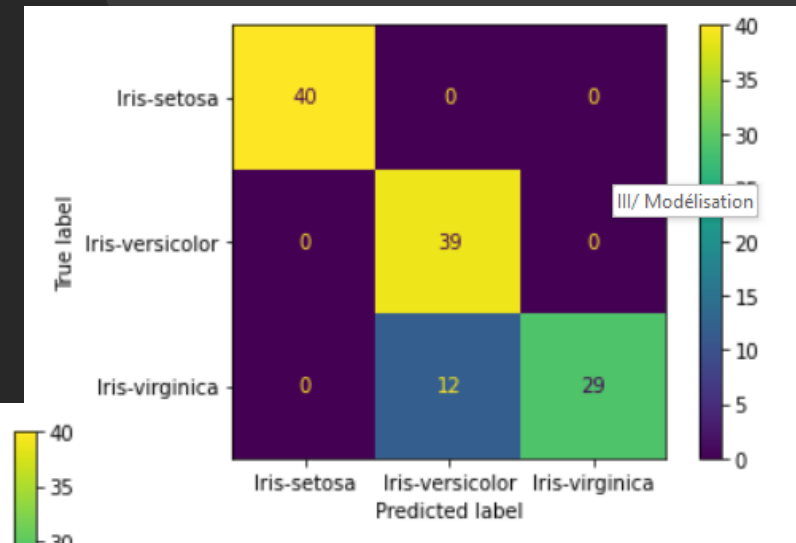
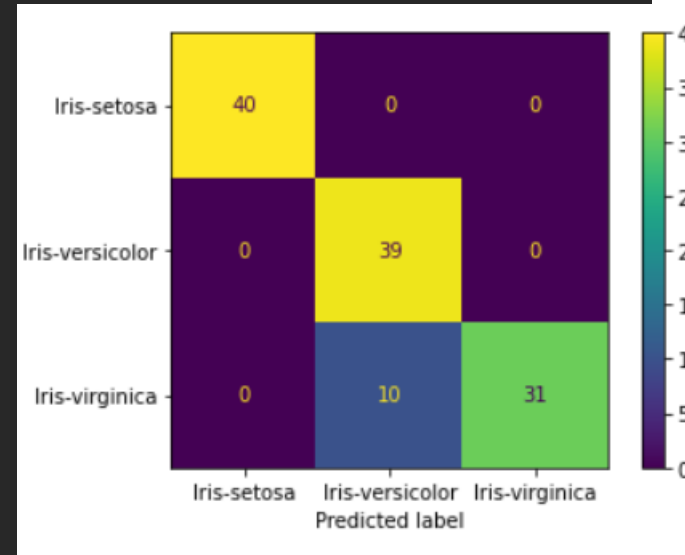
III/ Modélisation

- We consider three different models :

1. The Logistic Regression model
(Best accuracy 0,96)

2. The SVM model
(Best accuracy 0,93)

3. The Grid Search model
(Best accuracy 0,96)



III/ Web App

IRIS FLOWER DETECTION

First value :

Second value :

Third value :

Fourth value :



C:\Windows\System32\cmd.exe - app.py

Microsoft Windows [version 10.0.19041.928]
(c) Microsoft Corporation. Tous droits réservés.

C:\Users\theop\Desktop\Theo_Philippe_PythonData>app.py

C:\Python39\lib\site-packages\sklearn\base.py:310: UserWarning: Trying to unpickle estimator LogisticRegression from version 0.23.2 when using version 0.24.2. This might lead to breaking code or invalid results. Use at your own risk.

warnings.warn(

* Serving Flask app 'app' (lazy loading)

* Environment: production

WARNING: This is a development server. Do not use it in a production deployment.

Use a production WSGI server instead.

* Debug mode: on

* Restarting with stat

C:\Python39\lib\site-packages\sklearn\base.py:310: UserWarning: Trying to unpickle estimator LogisticRegression from version 0.23.2 when using version 0.24.2. This might lead to breaking code or invalid results. Use at your own risk.

warnings.warn(

* Debugger is active!

* Debugger PIN: 208-594-072

* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)

PREDICTION :

Iris-versicolor



[go back to home page](#)