

Rahma ALBEKBASHY

Adham ELBAHRAWY

Projet IA 2022

1

1.

En utilisant la librairie pandas :

```
import pandas as pd
df = pd.read_csv("synthetic.csv")
cols = df.columns.values.tolist()
cols
```

On peut voir tous les attributs du jeu de données ainsi qu'une colonne "Class". On peut en déduire qu'il existe `len(cols) - 1` ie 14 attributs. ### 2. En exécutant la commande `len(df["Class"].unique())` on peut conclure qu'il existe 4 classes différentes dans le jeu de données. ### 3. En utilisant la commande `df["Class"].value_counts().sort_values(ignore_index=True)`, on obtient l'affichage suivant qui nous indique le nombre d'instances par classe.

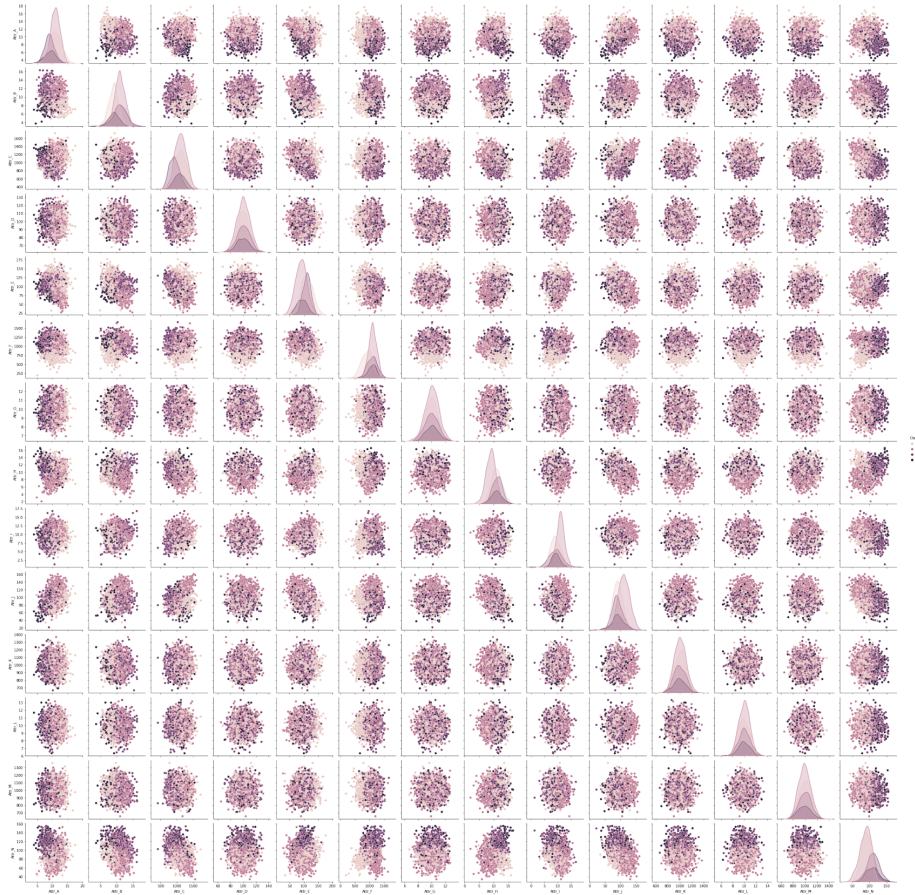
```
0    244
1    472
2    674
3    908
Name: Class, dtype: int64
```

4.

En exécutant le code suivant :

```
import seaborn as sns
sns.pairplot(df, hue="Class")
```

On observe un chevauchement important des classes sur les pairplots ainsi que sur les histogrammes, les données ne sont donc pas linéairement séparables.



5. On aurait pas besoin d'un encodage one-hot pour les arbres de décisions vue qu'elles ne prennent pas en compte les relations d'ordre des classes; par contre, il cela n'est pas le cas pour les réseaux de neurones, donc c'est plutôt intéressant d'encoder en one-hot les classes pour ces derniers. D'autre part, en regardant de nouveau les histogrammes, on peut observer que les courbes ont une forme assez Gaussienne, il n'y a donc pas d'outliers. On pourrait, cependant, standardiser les données mais cela n'a pas énormément d'intérêt. ### 6. Cela évite d'avoir un modèle qui apprend les données par coeur (overfitting) et le data snooping, ie cela nous permet de le tester sur des données qu'il n'a jamais vu. Par suite, on pourra avoir deux erreurs à tester, l'erreur d'entraînement et l'erreur de test dont on se fie pour la méthode d'early stopping.

2

2.1

Les 3 quartiles sont des valeurs strictement (dans notre contexte) supérieures à 25%, 50% et 75% respectivement à la population pour un attribut donné. Ils ont

calculés à l'aide de la fonction `quantile()` à laquelle on passe 0.25, 0.5 et 0.75.

2.2

a completer

3

3.1

Desicion Tree				
4				
classes	C0	C1	C2	C3
Accuracy	84.76%	86.98%	84.52%	90.06%
Precision	0.71	0.75	0.58	0.66
Recall	0.83	0.88	0.49	0.12
F1_Score	0.76	0.81	0.53	0.20

Modèle					
Desicion					
Tree 4					
True Label		0	1	2	3
	0	130	19	6	1
	1	15	174	2	2
	2	25	20	44	0
	3	12	8	23	6
Predicted label					

4

À la fin, tout dépend en ce qu'on cherche, si comme le cour on teste les cas des maladies (Diagnostiques médiceaux), on doit se voir dans les résultats les rappels (Recall) dans nos arbres la meilleur c'est L'arbre preds_arbre7.csv , et dans la prediction de la prof c'est y_pred_DT6.csv

Résultat Arbre 7

```

... Éxactitude de la classe 0: 83.23232323232324 %
Éxactitude de la classe 1: 86.47058823529412 %
Éxactitude de la classe 2: 87.4493927125506 %
Éxactitude de la classe 3: 91.45299145299145 %
Precision de la classe 0: 0.7552447552447552
Precision de la classe 1: 0.7461928934010152
Precision de la classe 2: 0.6421052631578947
Precision de la classe 3: 0.68
Recall de la classe 0: 0.6923076923076923
Recall de la classe 1: 0.8855421686746988
Recall de la classe 2: 0.6853932584269663
Recall de la classe 3: 0.3469387755102041
F1 Score de la classe 0: 0.7224080267558527
F1 Score de la classe 1: 0.8099173553719009
F1 Score de la classe 2: 0.6630434782608696
F1 Score de la classe 3: 0.4594594594594595
predicted    0    1    2    3
actual
0           108   30   14    4
1             11  147    6    2
2             14   12   61    2
3             10    8   14   17

```

arbre 7

Résultat Prédiction 6

```

Exactitude de la classe 0: 89.85507246376811 %
Exactitude de la classe 1: 90.3688524590164 %
Exactitude de la classe 2: 91.32231404958677 %
Exactitude de la classe 3: 92.78131634819533 %
Precision de la classe 0: 0.8496732026143791
Precision de la classe 1: 0.84
Precision de la classe 2: 0.7473684210526316
Precision de la classe 3: 0.7027027027027027
Recall de la classe 0: 0.8333333333333334
Recall de la classe 1: 0.8855421686746988
Recall de la classe 2: 0.797752808988764
Recall de la classe 3: 0.5306122448979592
F1 Score de la classe 0: 0.8414239482200647
F1 Score de la classe 1: 0.8621700879765396
F1 Score de la classe 2: 0.7717391304347826
F1 Score de la classe 3: 0.6046511627906977
predicted    0    1    2    3
actual
0           130   14   10   2
1             9  147    3   7
2             9    7   71   2
3             5    7   11  26

```

DT6