

EXAMEN 1 – Exploration données – Ricardo Vallejo

Partie I : Questions de compréhension (7 points)

1. Énumérez les familles caractéristiques de description des variables quantitatives.
Les caractéristiques des tendance centrale
Les caractéristiques de dispersion
Les caractéristiques de formes
2. Le coefficient d'asymétrie d'une distribution est égale à -2,3. Que peut-on dire de la distribution
La distribution est décalée vers la droite et la moyenne est inférieure à la médiane.
- 3.

3. Examinez le diagramme de la figure 1 et remplissez le tableau suivant correspondant à cette figure.

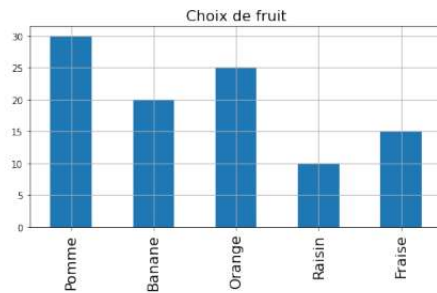


FIGURE 1 – Diagramme de choix de fruits

Response :

Modalite	nj	fj
Pomme	30	0.3
Banane	20	0.2
Orange	25	0.25
Raisin	10	0.1
Fraise	15	0.15
Total	100	1

4. Un collègue vous demande de calculer l'effectif cumulé à partir du tableau précédent. Quelle sera votre réaction/réponse à sa requête ?

Que si bien est calculable comme on montre dans le tableau, l'effectif cumule dans cette cas n'as pas de significat parce que l'agrupation de modalités n'as pas des sense dans le contexte réel.

Modalite	nj	fj	Effectif Cumule
Pomme	30	0.3	30
Banane	20	0.2	50
Orange	25	0.25	75
Raisin	10	0.1	85
Fraise	15	0.15	100
Total	100	1	

5. Analysez le diagramme en radar de la figure 2 et extrayez trois connaissances.

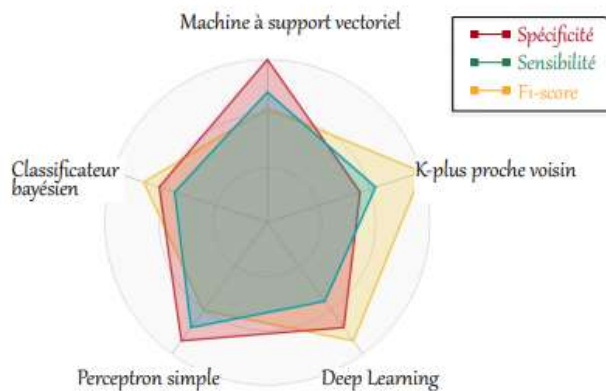


FIGURE 2 – Diagramme en radar

- Dans les applications de deep Learning, on utilise plus le Fi-Score que la spécificité pour évaluer le model, et le moins utilise c'est la sensibilité.
- Les paramètres de f1-score et spécificité, sont les plus utilisés dans les algorithmes de IA.
- Pour les machines de support vectoriel on utilise la spécificité plutôt que le F1-score comme critère d'évaluation.