

# QCM - Evaluation écrite - L3 Numérique

Répondez à l'ensemble du quiz. Une fois terminé, vous recevrez un courriel avec vos réponses.

*\* Indique une question obligatoire*

---

1. Adresse e-mail \*

---

2. Votre nom et votre prénom (NOM Prénom) \*

---

## Data Mining (Analyse exploratoire des données)

Questions portant sur le premier jour de la formation.

3. Pourquoi est-il important d'explorer les données avant de modéliser ?

*Plusieurs réponses possibles.*

- ☐ Pour éviter les erreurs causées par des données mal préparées.
- ☐ Pour comprendre la structure des données et leur distribution.
- ☐ Pour identifier les valeurs aberrantes et les données manquantes.
- ☐ Toutes les réponses ci-dessus.

4. Quelle méthode est la plus adaptée pour détecter des outliers dans une variable quantitative ?

*Une seule réponse possible.*

- ☐ Utiliser une heatmap.
- ☐ Calculer le Z-score ou l'IQR (Interquartile Range).
- ☐ Appliquer un test de corrélation de Spearman.
- ☐ Observer uniquement les moyennes et médianes.

5. Lequel des éléments suivants n'est pas une statistique descriptive ?

*Une seule réponse possible.*

- ☐ La moyenne.
- ☐ L'écart-type.
- ☐ Le scatterplot.
- ☐ Le mode.

6. Quelle est la principale différence entre la corrélation de Pearson et celle de Spearman ?

*Une seule réponse possible.*

- ☐ Pearson mesure les relations monotones, tandis que Spearman mesure les relations linéaires.
- ☐ Pearson utilise les rangs des observations, tandis que Spearman utilise les valeurs brutes.
- ☐ Pearson est sensible aux outliers, tandis que Spearman est plus robuste face aux outliers.
- ☐ Spearman nécessite une distribution normale des données, tandis que Pearson n'en a pas besoin.

7. Si deux variables ont une corrélation proche de zéro, cela signifie-t-il qu'elles ne sont pas liées ?

*Une seule réponse possible.*

- ☐ Oui, elles ne sont pas liées.
- ☐ Non, elles peuvent avoir une relation non linéaire.
- ☐ Oui, elles sont indépendantes.
- ☐ Non, elles sont monotones.

8. Que faire si un jeu de données contient des valeurs manquantes ?

*Plusieurs réponses possibles.*

- ☐ Toujours supprimer les lignes avec des données manquantes.
- ☐ Remplacer les données manquantes par la moyenne ou la médiane.
- ☐ Cela dépend de la quantité et de l'importance des données manquantes.
- ☐ Ignorer le problème.

9. Lors de l'analyse d'une matrice de corrélation, que pouvons-nous conclure si deux variables ont une corrélation proche de 1 ?

*Une seule réponse possible.*

- ☐ Les deux variables sont indépendantes.
- ☐ Les deux variables sont fortement liées et évoluent de manière similaire.
- ☐ Les deux variables sont peu liées, mais présentent une relation non linéaire.
- ☐ Les deux variables sont corrélées de manière monotone.

10. Quel graphique est le plus adapté pour visualiser la distribution d'une variable continue, en prenant en compte les valeurs aberrantes et la répartition des quartiles ?

*Une seule réponse possible.*

- ☐ Un histogramme.
- ☐ Un boxplot.
- ☐ Un scatterplot.
- ☐ Une matrice de corrélation avec heatmap.

## Machine Learning

Questions portant sur le second jour de la formation.

11. Qu'est-ce que le Machine Learning ?

*Une seule réponse possible.*

- ☐ Une méthode pour écrire des programmes sans erreurs
- ☐ Une technique où les ordinateurs apprennent à partir de données pour prendre des décisions
- ☐ Un langage de programmation
- ☐ Une base de données intelligente

12. Quel est le rôle de l'apprentissage supervisé ?

*Une seule réponse possible.*

- ☐ Identifier des groupes dans des données non étiquetées
- ☐ Apprendre à partir de données étiquetées pour faire des prédictions
- ☐ Améliorer les performances d'un modèle sans données historiques
- ☐ Choisir les meilleurs hyperparamètres automatiquement

13. Quelles phrases sont justes par rapport à la séparation d'un jeu de données en 2 parties (entraînement et test) ? (plusieurs réponses possibles) \*

*Plusieurs réponses possibles.*

- ☐ Cela permet d'évaluer correctement les performances d'un modèle.
- ☐ Dans un problème de classification, il faut s'assurer que les classes sont équitablement réparties.
- ☐ Cela consiste à diviser un jeu de données en 2 sous échantillons de dimension identique.
- ☐ Avec des séries temporelles nous pouvons diviser aléatoirement le jeu de données d'entrée.
- ☐ Cela permet d'éviter le sur-apprentissage.

14. Parmi les algorithmes suivants, lequel est utilisé pour la **classification** ? \*

*Plusieurs réponses possibles.*

- ☐ K-Means
- ☐ Régression linéaire
- ☐ Régression logistique

15. Quelles phrases sont justes au sujet du "sur-apprentissage" ? (plusieurs réponses possibles) \*

*Plusieurs réponses possibles.*

- ☐ Le sur-apprentissage est une méthode de standardisation des données.
- ☐ Une solution pour éviter le sur-apprentissage est d'ajouter des données d'entraînements.
- ☐ L'overfitting arrive quand le modèle est plus performant sur les données de test qu'à l'entraînement.
- ☐ Une des méthodes les plus efficaces pour éviter le sur-apprentissage est la cross-validation.
- ☐ Une solution pour éviter le sur-apprentissage est de sélectionner les variables explicatives non-corrélées.

16. Parmi les algorithmes suivants, lesquels sont utilisés pour des tâches d'apprentissage non supervisées ? \*

(plusieurs réponses possibles)

*Plusieurs réponses possibles.*

- ☐ Modèle de classification descendante hiérarchique ("Hierarchical clustering")
- ☐ K-means ("K-moyennes")
- ☐ Logistic regression ("Régression logistique")
- ☐ Multiple linear regression ("Régression linéaire multiple")

17. Quelle est la principale différence entre l'apprentissage supervisé et non supervisé ? \*

*Une seule réponse possible.*

- ☐ L'apprentissage supervisé utilise des données étiquetées, l'apprentissage non supervisé utilise des données non étiquetées
- ☐ L'apprentissage non supervisé est plus rapide
- ☐ L'apprentissage supervisé est utilisé uniquement pour la régression
- ☐ L'apprentissage non supervisé nécessite une supervision humaine constante

18. Quelle est la principale caractéristique d'un algorithme d'apprentissage non supervisé comme **K-Means** ?

*Une seule réponse possible.*

- ☐ Prédire des valeurs continues
- ☐ Découvrir des structures cachées dans des données non étiquetées
- ☐ Apprendre des règles prédéfinies
- ☐ Créer un modèle avec des données étiquetées