Validation croisée (*cross-validation*)

1. **Concepts de base :**
    - Qu'est-ce que la validation croisée et pourquoi est-elle importante dans l'entraînement des modèles de machine learning ?
    - Quelle est la différence entre la validation simple (*train/test split*) et la validation croisée ?
2. **Types de validation croisée :**
    - Quelles sont les différences entre *k-fold cross-validation*, *leave-one-out cross-validation* (LOOCV) et *stratified k-fold cross-validation* ?
    - Dans quels cas utiliser *stratified k-fold cross-validation* plutôt qu'une validation croisée classique ?
3. **Applications et limites :**
    - Quels sont les avantages et les inconvénients de la validation croisée pour les ensembles de données déséquilibrés ?
    - Comment la validation croisée permet-elle d'éviter le surapprentissage (*overfitting*) ?
4. **Métriques et résultats :**
    - Que représente le score moyen obtenu lors d'une validation croisée ?
    - Comment interpréter la variance des scores sur les différents plis (*folds*) ?

Optimisation des hyperparamètres (*GridSearchCV* et *RandomizedSearchCV*)

1. **Concepts de base :**
    - Quelle est la différence entre les paramètres d'un modèle et ses hyperparamètres ?
    - Pourquoi les hyperparamètres nécessitent-ils une optimisation séparée ?
2. **Approches d'optimisation :**
    - Comment fonctionne *GridSearchCV* ? Quels en sont les avantages et inconvénients ?
    - Comment *RandomizedSearchCV* diffère-t-il de *GridSearchCV* et dans quels cas est-il préférable ?
    - Quels sont les facteurs influençant le choix de la méthode d'optimisation (taille des données, coût computationnel) ?
3. **Configuration et choix :**
    - Qu'est-ce que le paramètre `cv` dans *GridSearchCV* et pourquoi son choix est-il critique ?
    - Comment choisir les hyperparamètres et les plages de valeurs à tester ?
4. **Problèmes courants :**
    - Quels risques peuvent survenir si la validation croisée est mal configurée dans *GridSearchCV* ?

    - Que signifie le terme *data leakage* dans le contexte de l'optimisation des hyperparamètres, et comment l'éviter ?
5. **Métriques et performance :**
    - Comment évaluer les performances des modèles obtenus via *GridSearchCV* ou *RandomizedSearchCV* ?
    - Pourquoi privilégier une métrique spécifique (par exemple, *accuracy* vs *F1-score*) pour certains problèmes ?