Évaluation et discussion:

Comparaison des modèles :

Dans ce projet, nous devons prédire si un passager a **survécu** au naufrage du Titanic. L’objectif est de minimiser les **faux négatifs** (ne pas rater des survivants).

* **Précision (Precision)** mesure la proportion de prédictions positives correctes, mais elle ne prend pas en compte les faux négatifs.
* **L’Accuracy** peut être trompeuse si la classe majoritaire domine (ex. beaucoup plus de non-survivants que de survivants).
* **L’AUC-ROC** donne une vue d’ensemble, mais dans un contexte critique comme la survie, on préfère maximiser le rappel.
* **Le rappel (Recall)** est crucial ici, car il indique la capacité du modèle à détecter tous les survivants. Un rappel élevé signifie que peu de survivants sont classés à tort comme "décédés".

Classement des 7 algorithmes selon les métriques.

Lalgorithme avec le meilleur rappel ici est le Naive Bayes, suivi de la Regression Logistique, du Foret Aleatoire, de lArbre de decision, K-Nearest Neighbors, Support Vector Machine et Réseaux de neurones.

○ Comparaison entre les datasets (Titanic vs Bank Marketing).

A completer

Comparaison des méthodes de validation :

○ Observations sur les scores en split 70/30 vs cross-validation 5/7/10 folds.

* Les scores obtenus avec **CV (5 folds), CV (7 folds) et CV (10 folds)** sont très proches pour tous les modèles, ce qui montre que la **performance est stable** et ne dépend pas fortement de la manière dont les données sont divisées.
* Exemple : Régression Logistique : 0.797703 **(5 folds),** 0.796148 **(7 folds),** 0.792908 **(10 folds)**
* SVM : 0.831419 **(5 folds),** 0.834671 **(7 folds),** 0.834665 **(10 folds)**
* Cela indique que les modèles **généralisent bien** et ne sont pas excessivement sensibles aux variations des jeux d’entraînement et de test.
* En général, l’**accuracy du split 70/30 est légèrement supérieure** à celle obtenue en validation croisée.
* Exemple :
* **Régression Logistique** : CV (10 folds) = 0.792908, Split 70/30 = 0.809701
* **Forêt Aléatoire** : CV (10 folds) = 0.804378, Split 70/30 = 0.783582
* Cette différence peut s’expliquer par le fait que le split 70/30 teste le modèle sur **un seul ensemble de test**, tandis que la validation croisée répartit les données de manière plus équilibrée, limitant ainsi la dépendance à un jeu de test particulier.

○ Discussion sur la variance des scores et la robustesse de l’estimation.

La validation croisée démontre une **faible variance**, témoignant de la **fiabilité et de la cohérence** des estimations de performance des modèles. Parmi eux, la **Régression Logistique ,le SVM** et le **K-Nearest Neighbors** se distinguent par leur **robustesse**, affichant des scores remarquablement **stables** quelle que soit la méthode de validation employée. À l’inverse, l’**arbre de décision** révèle une **sensibilité accrue** aux variations de l’ensemble de test, avec des fluctuations plus marquées entre le split 70/30 et la validation croisée, suggérant une possible **tendance au surapprentissage**.

3. Overfitting :

○ Analyser la différence entre la performance en entraînement et en test/validation.

○ Observer si certains modèles ont une variance importante entre les folds ou

entre train/test.