



# Techniques d'évaluation des performances et de sélection de modèles

**Présentée par :**

EDDRIOUCH AYA

**Supervisée par :**

MME. HARCHLI FIDAE

# Plan Du Travail :

- 01 Techniques d'évaluation des performances

---

- 02 Techniques de sélection de modèles

---

- 03 Défis et Considérations

---




# Techniques D'évaluation Des Performances

## MÉTRIQUES POUR LES MODÈLES DE CLASSIFICATION

- Matrice de confusion
- Exactitude (**Accuracy**)
- Précision
- Rappel
- F1-score
- Courbes ROC et AUC

## MÉTRIQUES POUR LES MODÈLES DE RÉGRESSION

- Erreur quadratique moyenne (**MSE**)
- Erreur absolue moyenne (**MAE**)
- Le coefficient de détermination (**R<sup>2</sup>**)



# Matrice De Confusion

Tableau qui montre les prédictions **correctes** et **incorrectes** pour chaque classe. Permet d'observer les faux positifs et les faux négatifs

		Classe réelle	
		1	0
Classe prédite	1	Vrais Positifs (TP)	Faux Negatives (FN)
	0	Faux Positifs (FP)	Vrais Négatifs (TN)

# Medical Data

Douleur thoracique	Bonne circulation sanguine	Artères bloquées	Poids	Maladie cardiaque
Non	Non	Non	125	Non
Oui	Oui	Oui	180	Oui
Oui	Oui	Non	210	Non
...	...	...	...	...

# Medical Data

Douleur thoracique	Bonne circulation sanguine	Artères bloquées	Poids	Maladie cardiaque
Non	Non	Non	125	Non
Oui	Oui	Oui	180	Oui
Oui	Oui	Non	210	Non
...	...	...	...	....

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
Yes	Yes	Yes	180	Yes
Yes	Yes	No	210	No

TRAINING DATA

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
Yes	Yes	Yes	180	Yes
Yes	Yes	No	210	No

TESTING DATA

		REEL	
PREDICTION		A une maladie cardiaque	N'a pas de maladie cardiaque
	A une maladie cardiaque	142	22
	N'a pas de maladie cardiaque	29	110

# Exactitude (Accuracy)

Proportion de bonnes prédictions sur l'ensemble des données

**NB:** Si les classes sont déséquilibrées, cette métrique peut être trompeuse.

$$\begin{aligned}\text{Accuracy} &= \frac{VP + VN}{VP + VN + FP + FN} \\ &= \frac{142 + 110}{142 + 110 + 22 + 29} = 83 \%\end{aligned}$$

Signifie que **83%** des prédictions du modèle sont correctes



# Précision

Indique parmi tous les exemples que le modèle a prédits comme positifs, quelle proportion est réellement positive.

$$\begin{aligned}\text{Précision} &= \frac{VP}{VP + FP} \\ &= \frac{142}{142 + 22} = 86 \%\end{aligned}$$

Signifie que **86%** des patients identifiés comme malades par l'algorithme sont réellement malades.

# Rappel

On appelle rappel ou sensibilité, le taux de vrais positifs, c'est-à-dire la proportion d'exemples positifs correctement identifiés

$$\begin{aligned}\text{Rappel} &= \frac{\text{VP}}{\text{VP} + \text{FN}} \\ &= \frac{142}{142 + 22} = 83 \%\end{aligned}$$

Signifie que **83%** des vrais malades ont été correctement détectés

# ✦ F1-score

On appelle F-mesure la moyenne harmonique de la précision et du rappel

$$\begin{aligned}\text{F1-score} &= 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} = \frac{2TP}{2TP + FP + FN} \\ &= \frac{2 \times 142}{2 \times 142 + 29 + 22} = 84 \%\end{aligned}$$

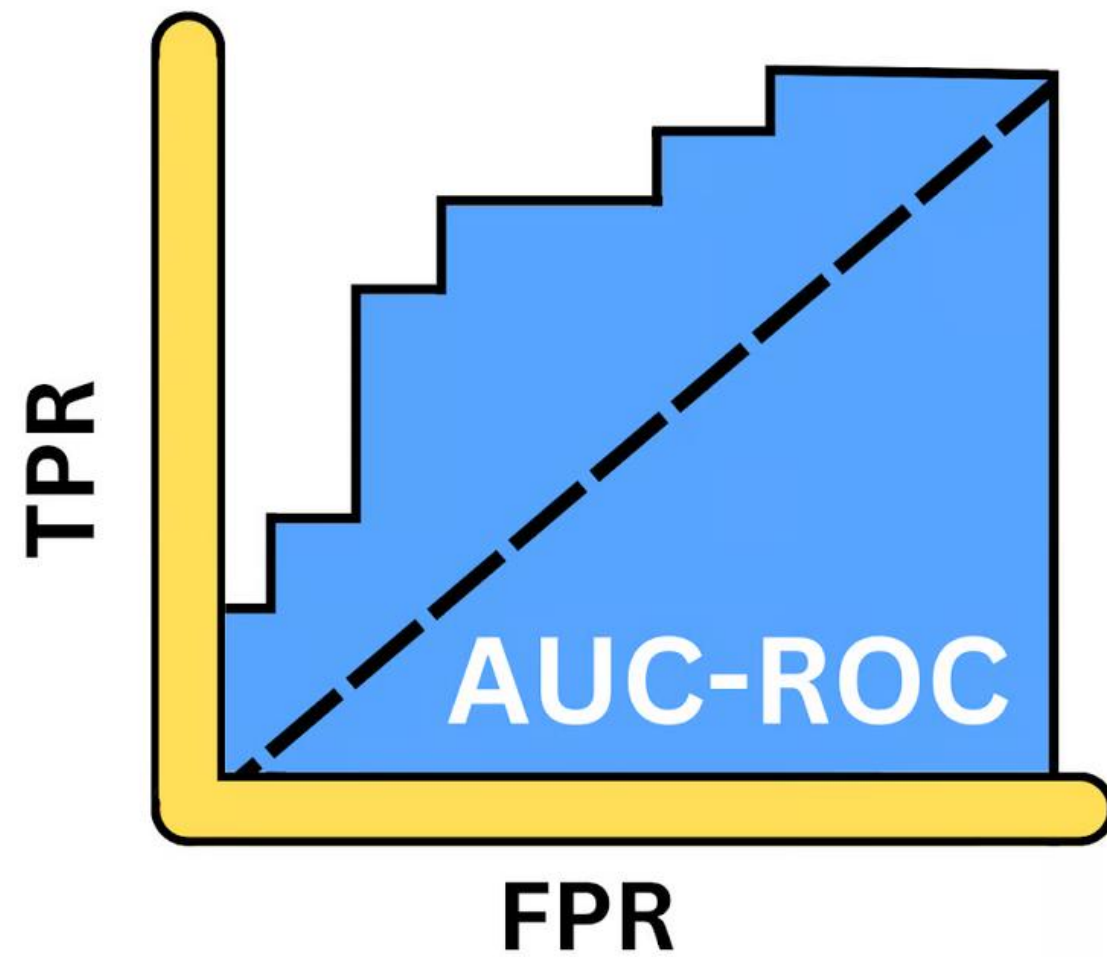
Indique que le modèle a **un bon compromis** entre éviter les faux positifs et éviter les faux négatifs.



# Courbes ROC et AUC

- ROC (**Receiver Operating Characteristic**) est un outil visuel qui montre la performance d'un modèle à différents seuils de classification. En traçant le taux de vrais positifs (rappel) contre le taux de faux positifs.
- AUC, ou **Area Under the Curve**, est une valeur scalaire unique allant de 0 à 1, qui donne un aperçu des performances du modèle. Vous ne calculez l'AUC qu'après avoir généré la courbe ROC.

# Courbes ROC et AUC



$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

# Métriques pour les modèles de régression

## MSE

### Mean Squared Error

- Mesure la différence quadratique moyenne entre les valeurs estimées et réelles.
- Lorsque les grandes erreurs sont plus critiques.

$$MSE = \frac{1}{n} \sum_i (y_{re'el} - y_{pre'dit})^2$$

## MAE

### Mean Absolute Error

- Cette métrique mesure la moyenne des erreurs absolues entre les prédictions et les valeurs réelles.
- Avec beaucoup de valeurs aberrantes ou des résidus non normaux.

$$MAE = \frac{1}{n} \sum_i |y_{re'el} - y_{pre'dit}|$$

## R<sup>2</sup>

### R-Squared

- Proportion de la variance des valeurs réelles expliquées par le modèle.
- Pour évaluer l'ajustement des modèles de régression linéaire.

$$R^2 = 1 - \frac{\sum_i (y_{re'el} - y_{pre'dit})^2}{\sum_i (y_{re'el} - \bar{y})^2}$$



# Techniques De Sélection De Modèles

## MÉTRIQUES DE PERFORMANCE

- Utilisation des métriques d'évaluation pour comparer et sélectionner le meilleur modèle

## RÉGULARISATION

- Régression Lasso (L1)
- Régression Ridge (L2)
- Régularisation Elastic Net (L1+L2)

# Métriques De Performance: Pour La Régression

- Il faut d'abord déterminer si les erreurs extrêmes doivent être pénalisées ou non. Si les grandes erreurs sont critiques, le **MSE** est préférable car il accentue l'impact des erreurs élevées. En revanche, si on veut une métrique plus robuste aux valeurs aberrantes, le **MAE** est plus utile, car il traite toutes les erreurs de manière égale. Enfin, pour comparer plusieurs modèles et évaluer leur capacité globale, le **R<sup>2</sup>** est un bon indicateur

# Métriques De Performance: Pour La Classification

- Si les classes sont équilibrées, **l'Accuracy** est suffisante. Cependant, en cas de déséquilibre, Accuracy devient trompeuse. Dans ce cas, si les faux positifs sont coûteux, on privilégiera la **Précision**. Si les faux négatifs sont critiques, le **Rappel** est plus important. Pour un bon compromis, notamment lorsque aucune des deux erreurs n'est prioritaire, le **F1-score** est une bonne option. Enfin, si le modèle doit produire des probabilités et être évalué globalement sur différents seuils de classification, **l'AUC-ROC** est recommandé

# Application Précédente (Medical Dataset)

## DEUX SITUATIONS POSSIBLES À SE PASSE:

- Si l'objectif est de ne manquer aucun malade → il vaut mieux maximiser le rappel = quitte à accepter plus de faux positifs.
- Si l'objectif est d'éviter les erreurs de diagnostic chez les non-malades → il faut maximiser la précision.

# La Régularisation

## L1

### Lasso

- Le modèle pénalise la valeur absolue des coefficients, à tel point que certains seront 0.
- Il est souvent appliquée aux données de grande dimension, ce qui la rend plus efficace en termes de calcul et de mémoire.
- Ou si on veut éliminer des variables inutiles.

## L2

### Ridge

- Le modèle pénalise le carré des coefficients. L'échelle et l'écart des coefficients sont réduits.
- On l'utilise si toutes les variables sont utiles et qu'on veut éviter des coefficients extrêmes.

## L1+L2

### Elastic Net

- Combiner L1 (Lasso) et L2 (Ridge) pour profiter des deux avantages.
- On l'utilise si on veut une sélection automatique de variables.

# PROBLÈMES ET DÉFIS

1

## L'optimisation des hyperparamètres

Choisir les bons hyperparamètres (ex : Le nombre de voisins dans un modèle KNN, taux d'apprentissage d'un réseau de neurones) est complexe et peut affecter la performance du modèle.

2

## Le déséquilibre des classes

Il se produit lorsqu'une classe est beaucoup plus représentée qu'une autre dans un jeu de données. Donc le modèle d'apprentissage automatique peut apprendre à ignorer la classe minoritaire et donner des résultats trompeurs.

3

## Le compromis biais-variance

Trouver le bon équilibre entre un modèle trop simple (biais élevé) et un modèle trop complexe (variance élevée).



# »» Conclusion

---

Pour conclure, les techniques d'évaluation des performances et de sélection de modèles jouent un rôle essentiel pour garantir la qualité, la fiabilité et la pertinence des modèles d'apprentissage automatique. Elles permettent de choisir des modèles à la fois performants et adaptés aux besoins du problème.

Dans cette continuité, l'optimisation des hyperparamètres, par des approches telles que la recherche en grille, aléatoire ou bayésienne, offre des leviers supplémentaires pour renforcer la performance et la robustesse des modèles.

---

**Merci Pour Votre  
Attention**