

Évaluation d'algorithmes

*Quels métriques
pouvons-nous utiliser pour
évaluer les algorithmes?*

Atelier #9





Nouveau jeu de données pratique!

Jeu de données : Force de compression du ciment

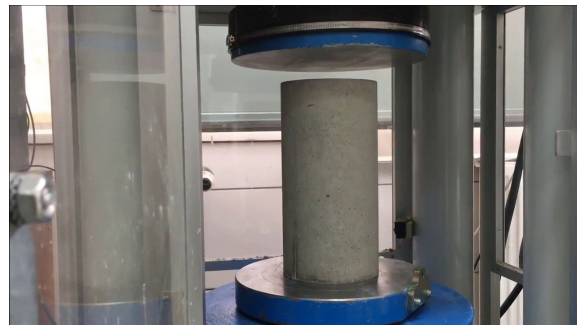
Lien : <https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>

Type de problématique : Régression

Nom du jeu de données : concrete.csv

TÂCHES

1. Télécharger le jeu de données.
2. Charger le jeu de données dans le notebook pratique.
3. Effectuer une exploration sommaire du jeu de données.



**Disponible dans les jeux de données importés sur Google Classroom.*



Nouveau jeu de données théorique!

Jeu de données : Efficacité énergétique de bâtiments

Lien : <https://archive.ics.uci.edu/ml/datasets/Energy+efficiency>

Type de problématique : Régression

Nom du jeu de donnée : energy_efficiency.csv

TÂCHES

1. Télécharger le jeu de données.
2. Charger le jeu de données dans le notebook théorique.



**Disponible dans les jeux de données importés sur Google Classroom.*



Importance du choix de métrique

À considérer avant de sélectionner un métrique :

- Très important de sélectionner le métrique approprié pour l'algorithme.
- Il faut faire la différence entre les métriques de classification et de régression.
- Le choix du métrique influence comment la performance d'un algorithme est évaluée.
- Influence l'importance et le choix des variables du jeu de données dans les résultats.





Métriques pour la CLASSIFICATION

Métrique	Terme anglophone
Précision de classification	<i>Classification accuracy</i>
Aire sous la courbe ROC	<i>Area under the ROC curve</i>
Matrice de confusion	<i>Confusion matrix</i>
Rapport de classification	<i>Classification report</i>



Précision de classification

$$\textit{Précision} = \frac{\textit{Bonnes prédictions}}{\textit{Prédictions totales}}$$

Définition : Nombre de bonnes prédictions par rapport à toutes les prédictions effectuées.

Utilisation :

- C'est le métrique le plus fréquemment utilisé (*aussi surutilisé).
- Son utilisation optimal requiert des classes avec un nombre égal d'observations, ce qui n'est généralement pas le cas.
- Équivalent à l'*accuracy* en anglais.



Matrice de confusion (**français*)

		Classes prédites		Somme
		Positif (1)	Négatif (0)	
Classes réelles	Positif (1)	Vrai positif TP = 100	Faux négatif FN = 10	110
	Négatif (0)	Faux positif FP = 5	Vrais négatif TN = 50	55
Somme		105	60	165

Vrai positif = Lorsque le patient a le diabète et que nous avons prédit qu'il aura le diabète.

Vrai négatif = Lorsque le patient n'a pas le diabète et que nous avons prédit qu'il n'aura pas le diabète.

Faux positif = Lorsque le patient n'a pas le diabète mais que nous avons prédit qu'il aura le diabète.

Faux négatif = Lorsque le patient a le diabète mais que nous avons prédit qu'il n'aura pas le diabète.

Question

Dans le contexte du jeu de données sur le diabète, quel est le métrique qu'il faut minimiser?



Matrice de confusion (*français)

Exemple fictif		Classes prédites		Somme
		Positif	Négatif	
Classes réelles	Positif	Vrai positif TP = 100	Faux négatif FN = 10 Erreur type 2	110
	Négatif	Faux positif FP = 5 Erreur type 1	Vrais négatifs TN = 50	55
Somme		105	60	165

Vrai positif = Lorsque le patient a le diabète et que nous avons prédit qu'il aura le diabète.

Vrai négatif = Lorsque le patient n'a pas le diabète et que nous avons prédit qu'il n'aura pas le diabète.

Faux positif = Lorsque le patient n'a pas le diabète mais que nous avons prédit qu'il aura le diabète.

Faux négatif = Lorsque le patient a le diabète mais que nous avons prédit qu'il n'aura pas le diabète.

Question

Dans le contexte du jeu de données sur le diabète, quel est le métrique qu'il faut minimiser?

Erreurs de type 1 vs type 2

Erreur de type 1 :

- Faux positif - Prédire une valeur positive alors qu'elle est négative.
- On minimise l'erreur de type 1 en optimisant la métrique d'exactitude (*precision*) d'un modèle.

Erreur de type 2 :

- Faux négatif - Prédire une valeur négative alors qu'elle est positive.
- On minimise l'erreur de type 2 en optimisant la métrique de sensibilité/rappel (*sensitivity/recall*) d'un modèle.

Type I Error



Type II Error



Matrice de confusion (*anglais)

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Rappel/
Recall

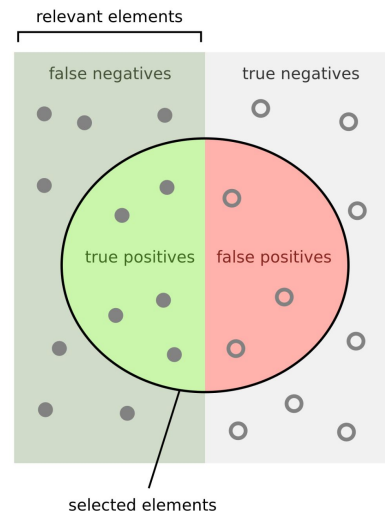


Rapport de classification

Définition : Le rapport de classification amalgame les principaux métriques qui permettent d'évaluer la fiabilité d'un modèle par la précision, le rappel (*recall*), le score f1 et le support.

Utilisation :

- Utilisé principalement pour obtenir le score f1.
- Le score f1 est utilisé dans les classification binaire.
- Le score f1 permet de minimiser les erreurs de types 1 et 2.



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$



Aire sous la courbe ROC

Définition : Représente la capacité d'un modèle à faire la différence entre une classe positive et négative.

Utilisation :

- Utilisé pour les classifications binaires.
- Une aire sous la courbe de 1 signifie que toutes les prédictions sont correctes.
- Une aire sous la courbe de 0.5 est équivalente à une valeur aléatoire (donc mauvaise).
- Peut être séparé entre la sensitivité (*sensitivity*) et la spécificité (*specificity*).





Sensitivité vs Spécificité

(sensitivity vs specificity)

Sensitivité

- Définition : Taux de vrais positifs.
- Synonymes : sensitivity, rappel, recall

$$\textit{Sensitivité/Retour} = \frac{\textit{Vrais positifs}}{\textit{Total de positifs}}$$

Spécificité

- Définition : Taux de vrais négatifs.
- Synonymes : specificity

$$\textit{Spécificité} = \frac{\textit{Vrais négatifs}}{\textit{Total de négatifs}}$$



Métriques pour la RÉGRESSION

Métrique	Terme anglophone
Erreur absolue moyenne	<i>Mean absolute error (MAE)</i>
Erreur quadratique moyenne	<i>Mean squared error (MSE / RMSE)</i>
R au carré (R^2)	<i>R squared</i>

Liste de tous les métriques disponibles sur Sklearn

https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics



Moyenne des erreurs absolues (MAE)

Définition : Moyenne de la somme des différences absolues entre les valeurs prédites et réelles.

Utilisation :

- Un MAE se situe entre 0 et l'infini.
- Un MAE de 0 signifie des prédictions parfaites. Plus on s'éloigne de 0, plus les prédictions sont mauvaises.
- Permet de répondre à la question suivante : "En moyenne, quelle est la distance entre la valeur prédite et la valeur réelle.
- Obtenir une idée du niveau auquel les valeurs prédites sont erronées.
- Interpréter comme un ordre de magnitude mais ne permet pas de savoir dans quelle direction les données sont mauvaises.
- Utiliser comme outil pour comparer des modèles entre eux et les optimiser.
- Métrique intuitif et populaire.



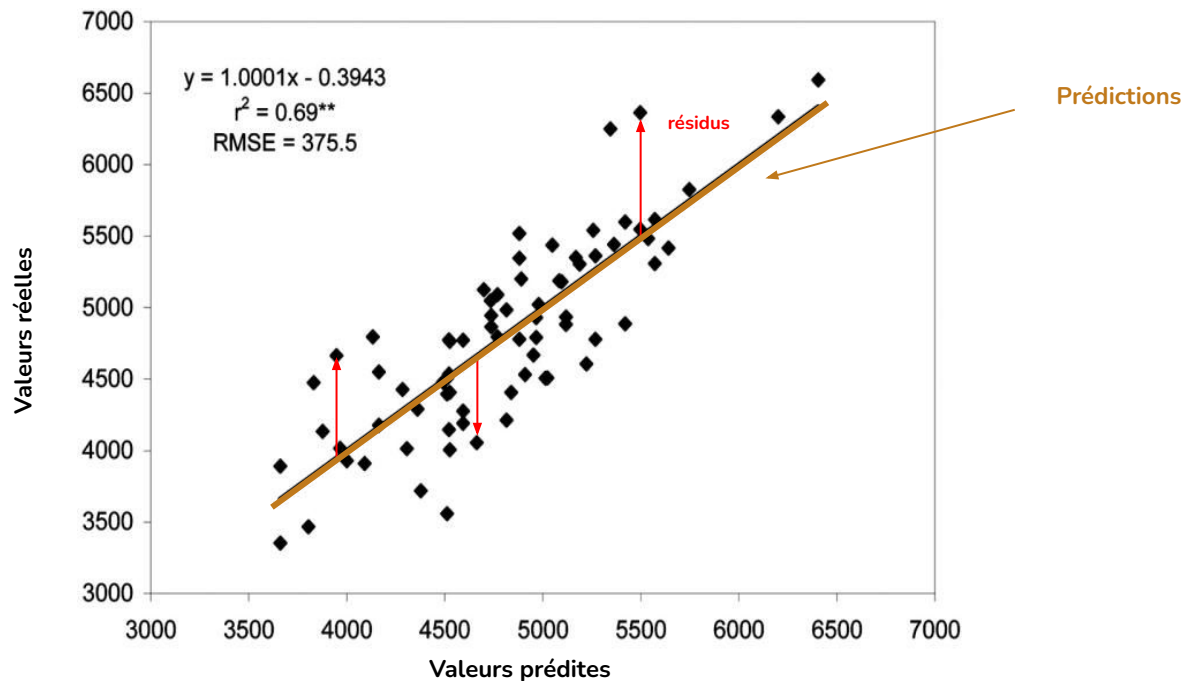
Erreur quadratique moyenne (MSE / RMSE)

Définition : Moyenne des résidus au carré (erreurs) entre les valeurs prédites et réelles.

Utilisation :

- Un MSE se situe entre 0 et l'infini.
- Un MSE de 0 signifie des prédictions parfaites. Plus on s'éloigne de 0, plus les prédictions sont mauvaises.
- Permet de considérer les grandes erreurs en pénalisant le métrique.
- Les résidus sont une mesure de la distance entre la droite de régression (valeurs prédites) et les points individuels (valeurs réelles).
- La racine carrée de l'erreur quadratique moyenne (RMSE) indiquent la dispersion des résidus.

Erreur quadratique moyenne (MSE / RMSE)





R au carré (R^2 / *R-squared*)

Définition : Proportion de variance expliquée par le modèle.

Utilisation :

- La valeur du R-carré se situe entre 0 et 1.
- Elle est considérée satisfaisante à partir de 0.7 selon le consensus général.
- Une valeur de 1 signifie que le modèle s'ajuste parfaitement aux données.
- Une valeur de 0 signifie que le modèle s'ajuste aucunement aux données.
- Attention : Le R-carré est dépendant du nombre d'observations et du nombre de colonnes dans un jeu de données. Il n'est donc pas idéal lorsqu'on a un grand/large jeu de données.



Tableau d'utilisation

Métrique	Description	Utilisation
Erreur absolue moyenne (MAE)	Moyenne des différences absolues entre les valeurs réelles et prédites.	Donner un poids égal à toutes les valeurs prédites par rapport aux valeurs réelles.
Erreur quadratique moyenne (MSE / RMSE)	Racine carrée de la moyenne des différences au carré entre les valeurs réelles et prédites.	Pénaliser les valeurs prédites qui sont extrêmes (i.e. très loin de la valeur réelle).
R au carré (R^2)	Pourcentage de la variance des variables dépendantes expliqué par la variable cible.	