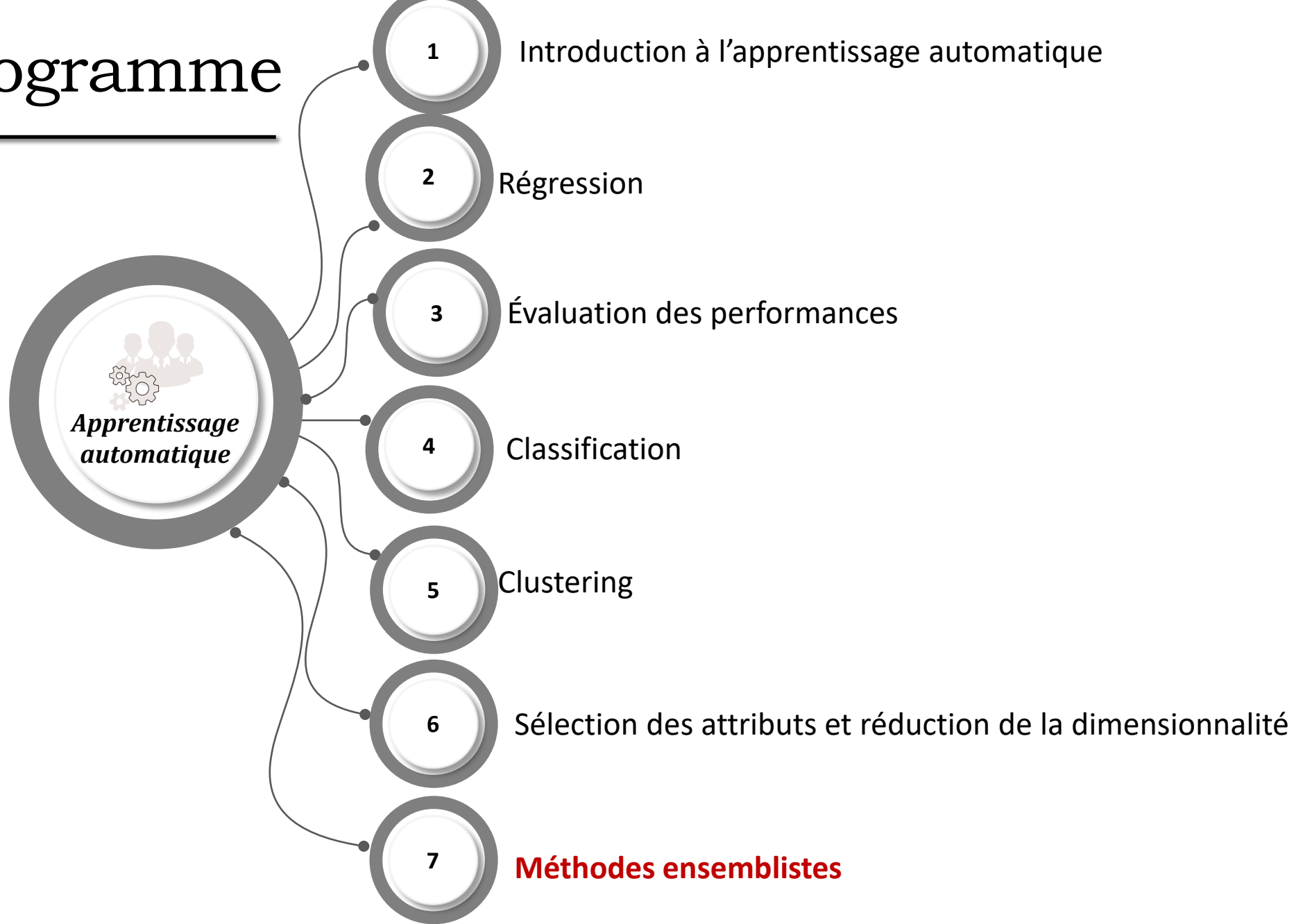


Apprentissage Automatique

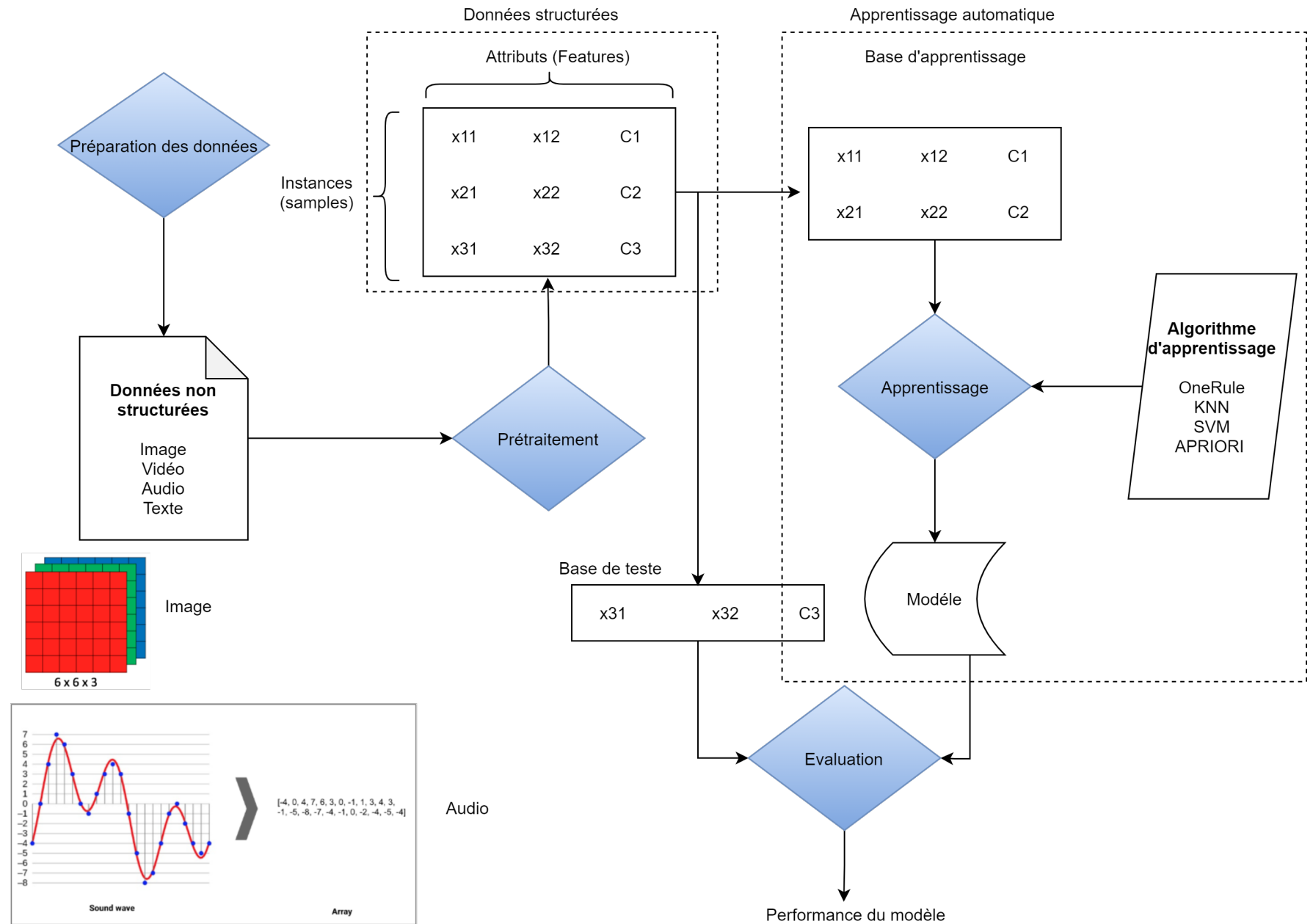
*Intelligence Artificielle et
Sciences de Données
(IASD)*

DR N. DIF

Programme



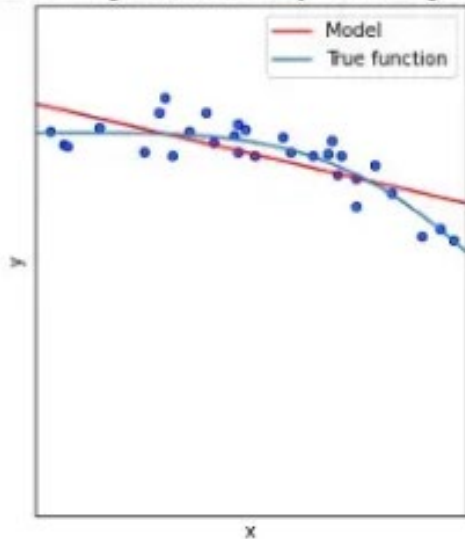
Rappel



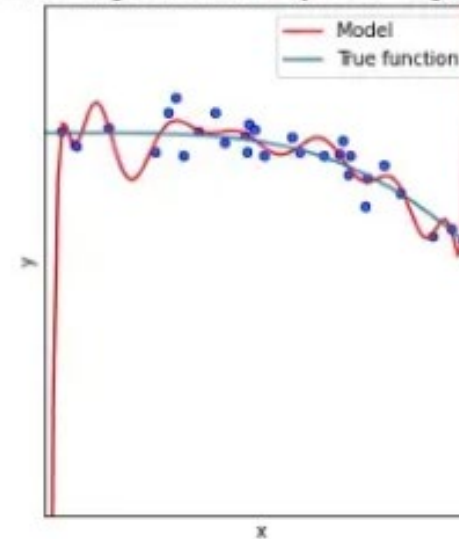
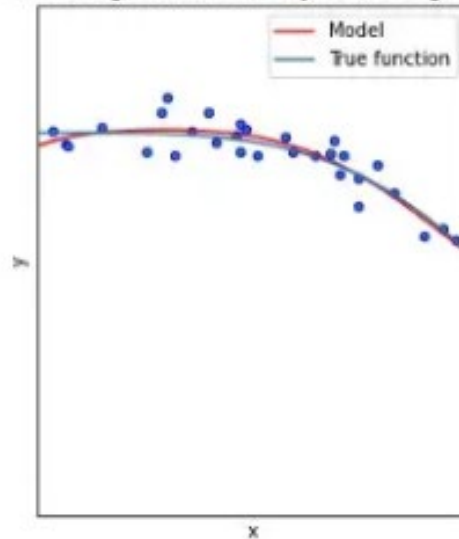
1. MÉTHODES ENSEMBLISTES

Bias et variance

Linear Regression with Polynomial Degree 1 Linear Regression with Polynomial Degree 5 Linear Regression with Polynomial Degree 15



Un modèle caractérisé par un biais élevé. Le modèle apprend facilement le lien entre l'entrée et la sortie (droite). Les modèles avec un biais élevé sont généralement caractérisés par un **underfitting**.



Un modèle caractérisé par une variance élevée. La variance du modèle mesure la performance du modèle sur différentes données d'entraînement. Les modèles avec une variance élevée sont généralement caractérisés par un **problème d'overfitting**. (Les arbres de décision ont généralement un biais faible et une variance élevée, car ils sont sensibles aux changements de données.)

1. MÉTHODES ENSEMBLISTES

Bias et variance

Faire un compromis entre la variance et les biais est un problème connu en machine learning. Un modèle performant est caractérisé par des variance faible et un biais faible. Réduire la variance et le biais en parallèle peut créer un conflit, car réduire trop la variance peut augmenter le biais et vice-versa. Généralement, les modèles complexes sont caractérisé par des biais faible et une variance élevé, et un modèle simple est caractérisé par une variance faible et un biais élevé.

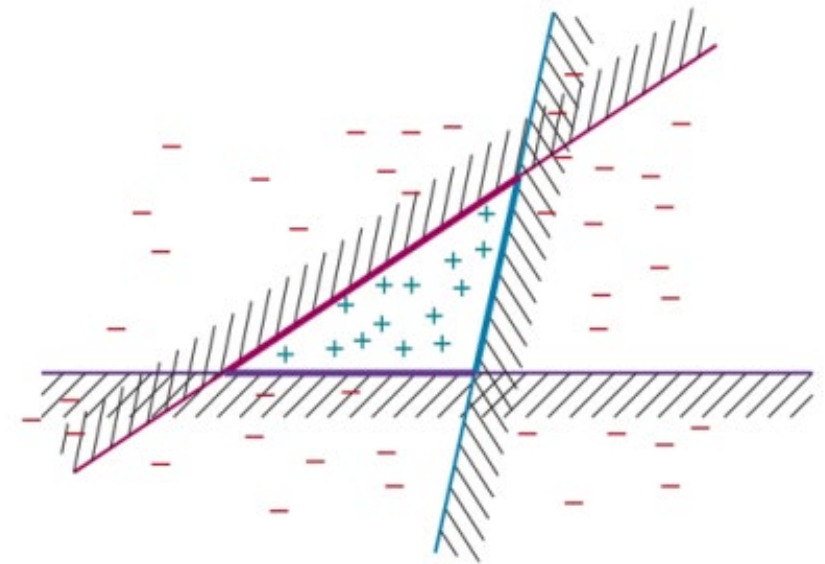


LET GROUP OUR
DECISIONS

1. MÉTHODES ENSEMBLISTES

1.1. Principe

- L'apprentissage ensembliste consiste à sélectionner une collection, ou un ensemble, d'hypothèses (h_1, h_2, \dots, h_n) et de combiner leurs prédictions par moyenne, en vote ou par une autre technique d'apprentissage automatique. Nous appelons les modèles de base des hypothèses individuelles et leur combinaison un modèle d'ensembliste.



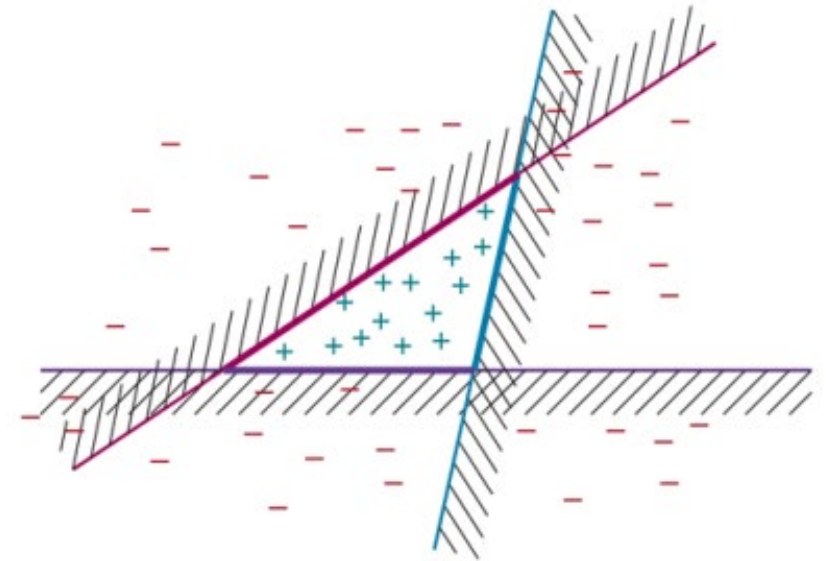
1. MÉTHODES ENSEMBLISTES

1.1. Principe

Pourquoi l'apprentissage ensembliste ?

- Résoudre le **dilemme biais-variance**, en réduisant la variance en par un ensemble de modèles sans les frais d'augmentation du biais.
- **Améliorer la performance** : il a été démontré que la performance d'un ensemble est généralement meilleure que celle des modèles individuels.

Quelle est l'accuracy de chaque classificateur linéaire et de l'ensemble?



Un ensemble de trois classificateurs linéaires peut représenter une région triangulaire qui ne peut pas être représentée par un seul classificateur linéaire

1. MÉTHODES ENSEMBLISTES

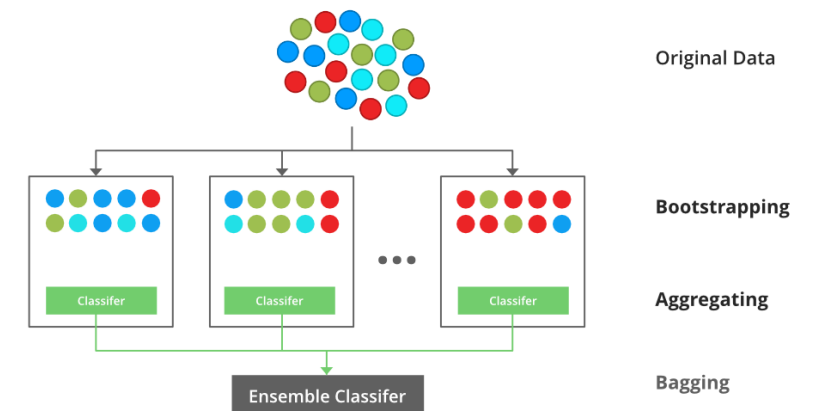
1.2. Bagging (bootstrap aggregating)

- Méthode ensembliste qui permet d'améliorer la performance des Weak learners. L'apprentissage est effectué en parallèle.
- Générer K datasets distincts par échantillonnage avec remplacement à partir de la dataset originale.
- Exécutez chaque algorithme sur chaque dataset pour générer K modèles différents.
- La prédiction finale dépend de la prédiction de chaque modèle : vote majoritaire en classification, moyenne non pondéré en régression.

$$h(x) = \frac{1}{K} \sum_{i=1}^K h_i(x)$$

Echantillonnage avec remplacement (Bootstrap):

Un échantillon aléatoire de données est sélectionné avec remplacement dans un jeu d'apprentissage, ce qui signifie que les points de données individuels peuvent être choisis plus d'une fois.



1. MÉTHODES ENSEMBLISTES

1.2. Bagging (bootstrap aggregating)

Quand utiliser le Bagging ?

- Pour réduire la variance si les données sont limitées ou le modèle a un problème de sur-apprentissage.
- Généralement utilisé sur les algorithmes sensibles (instable) aux changements de données (arbre de décision).

EXEMPLE : Random forest

Inconvénient : Le bagging des arbres de décision peut générer des arbres fortement corrélés

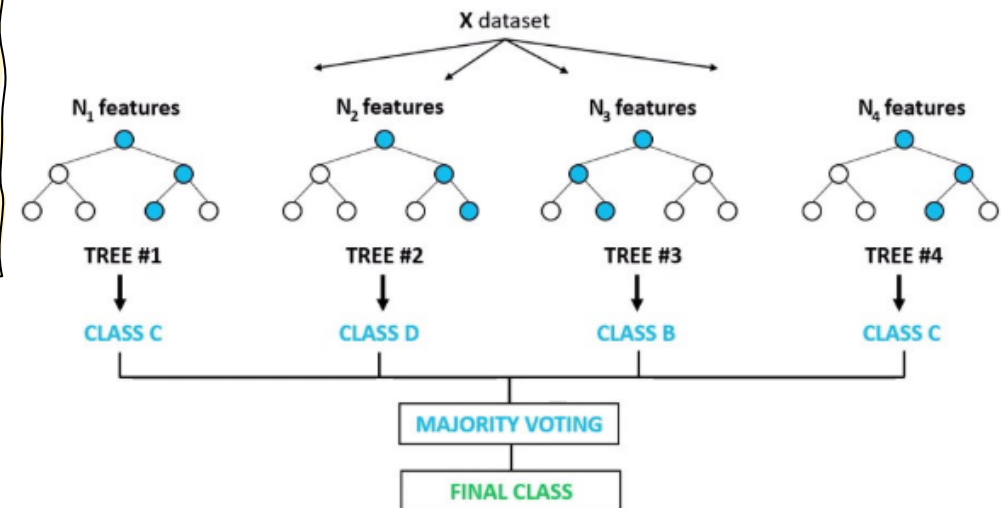
1. MÉTHODES ENSEMBLISTES

1.3. Random forest

- L'idée clé est de varier aléatoirement les choix d'attributs (plutôt que les instances d'apprentissage).
- A chaque itération, sélectionner aléatoirement K attributs. Généralement, dans chaque itération \sqrt{n} attributs est sélectionné en classification, et $\frac{n}{3}$ attributs en régression.

D'autres types d'arbres de décision : Extra-trees algorithm.

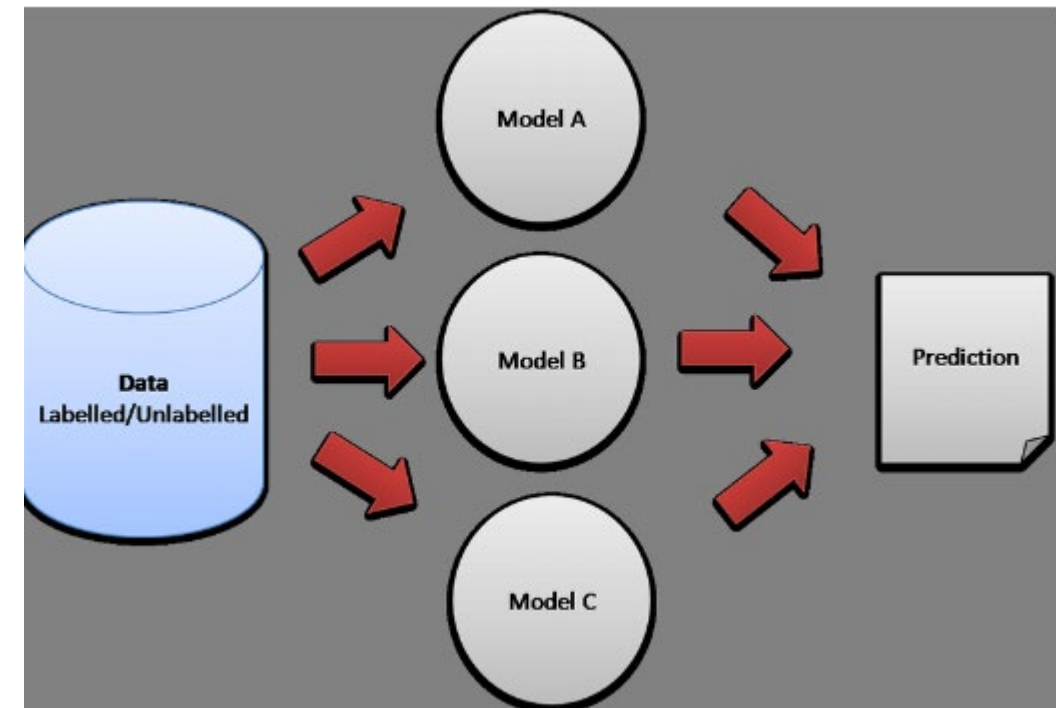
Random Forest Classifier



1. MÉTHODES ENSEMBLISTES

1.4. Model Averaging Ensemble

- L'idée est combiner plusieurs modèles de base de différentes classes sur les mêmes données.
- La combinaison des prédictions est effectuée à travers un vote majoritaire ou moyennage non pondéré



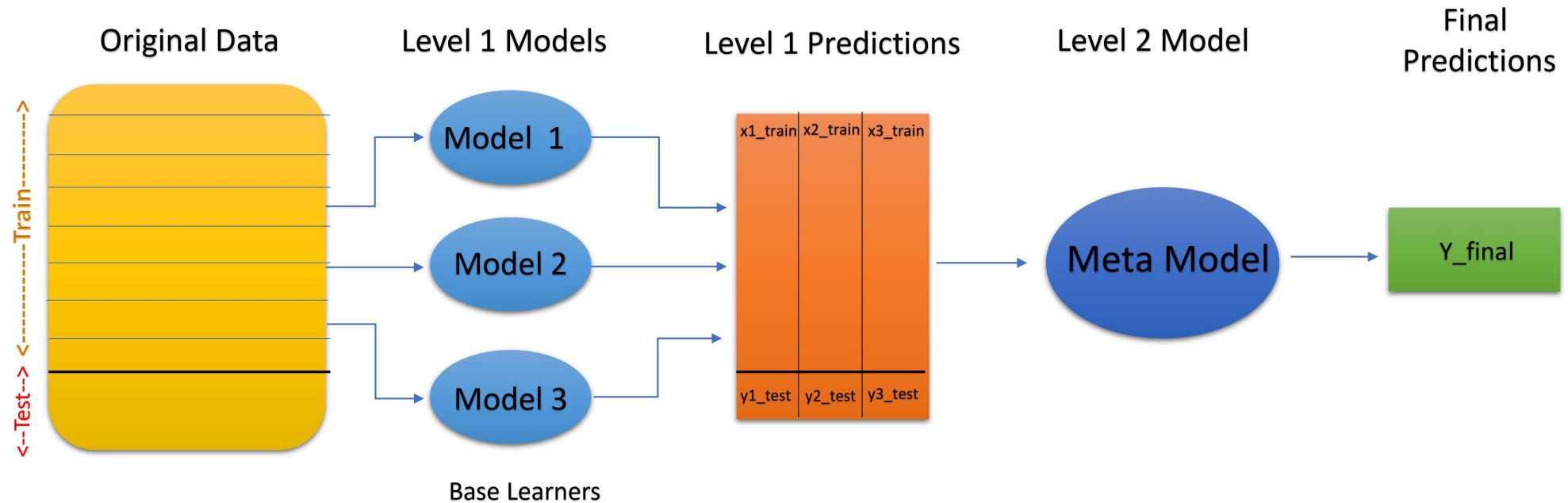
1. MÉTHODES ENSEMBLISTES

1.4. Stacking

- Une méthode ensembliste à base de la technique model ensembling average basée sur le meta-learning.
- L'idée est combiner plusieurs modèles de base de différentes classes (weak learners hétérogènes) sur les mêmes données.
- La combinaison des prédictions est effectué à travers un autre algorithme d'apprentissage automatique (Meta-learner).

1. MÉTHODES ENSEMBLISTES

1.4. Stacking



1. MÉTHODES ENSEMBLISTES

1.4. Stacking

- 1- Les données de l'apprentissage sont divisées en n-folds.
- 2- Model 1 est entraîné sur les n-1 folds et les prédictions sont faites sur la fold restante.
- 3- Les prédictions sont ajoutées à la liste `X1_train`.
- 4- les étapes 2 et 3 sont répétées sur les autres folds pour obtenir un vecteur `X1_train` de taille n (n est le nombre d'instances dans la base d'apprentissage).
- 5- Les prédictions sont faites sur la base de test et les résultats sont ajoutées au vecteur `y1_test`.
- 6- De la même manière, on génère `x2_train`, `y2_test`, `x3_train` et `y3_test` en utilisant Model 2 et 3 pour générer la dataset level 1 predictions.
- 7- Entraîner le Meta learner sur Level 1 predictions.
- 8- Utiliser le meta-learner pour effectuer les prédictions sur la base de test

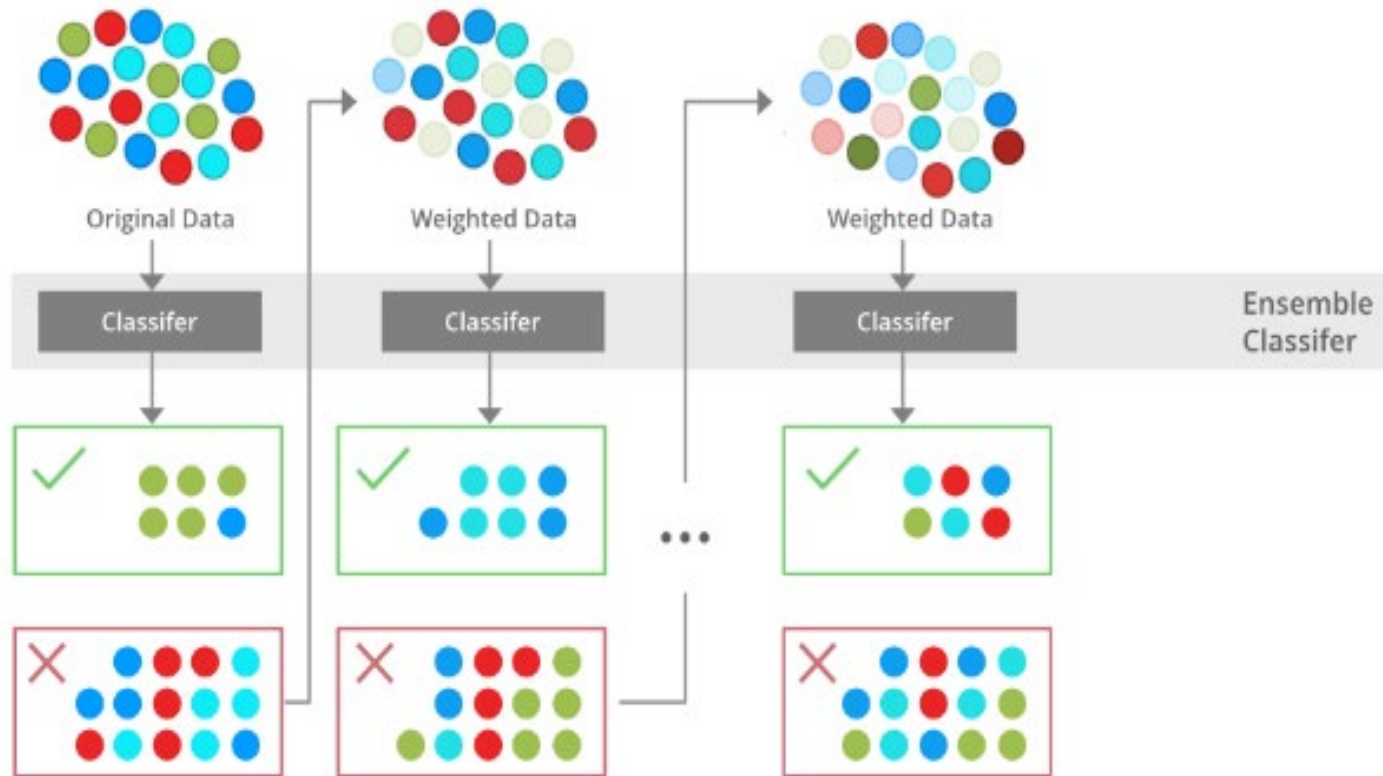
1. MÉTHODES ENSEMBLISTES

1.2. Boosting

- Les techniques de Boosting sont parmi les méthodes ensemblistes les plus populaires.
- Une méthode ensemblistes séquentielle qui utilise la notion d'une base d'apprentissage pondéré .
- **Une base d'apprentissage pondéré** associe un poids w_i à chaque instance. Un poids décrit le nombre d'exemples à compter durant l'apprentissage. Par exemple, si un exemple avait un poids de 3 et que les autres exemples avaient tous un poids de 1, cela équivaldrait à avoir 3 copies d'un exemple dans l'ensemble d'apprentissage.

1. MÉTHODES ENSEMBLISTES

1.2. Boosting



1. MÉTHODES ENSEMBLISTES

1.2. Boosting

1. Initialisez la dataset et attribuez des poids égaux à chaque instance.
2. Attribuez la dataset normalisé au modèle et identifiez les instances mal classifié.
3. Augmentez le poids des Instances mal classés et diminuez le poids des points de données correctement classés.
4. Si les résultats sont satisfaisantes
 Allez à l'étape 5
Sinon
 Allez à l'étape 2
5. Fin

1. MÉTHODES ENSEMBLISTES

1.2. Boosting

Des algorithmes de boosting

AdaBoost

GBM

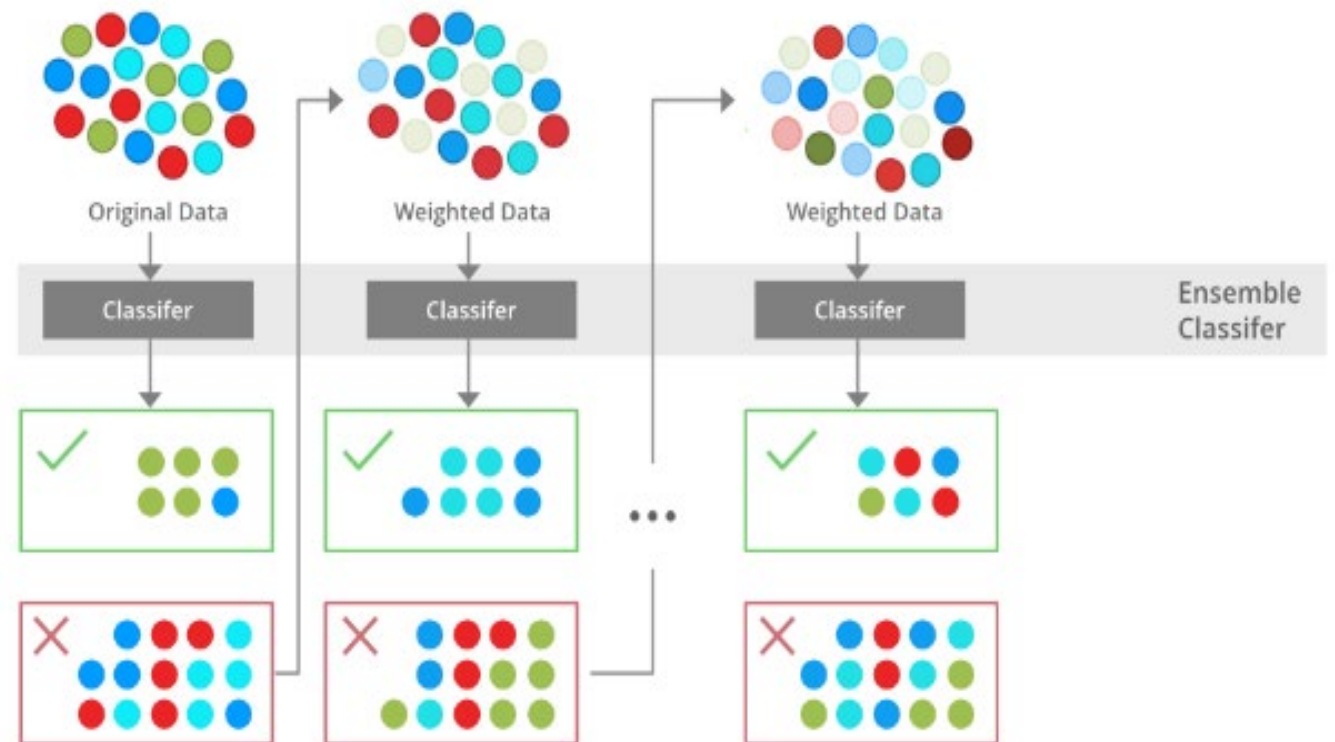
XGBM

GBM léger

CatBoost



**Combien de modèles
utilisent les algorithmes de
boosting?**



RÉFÉRENCES

Russell, S. J. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc..