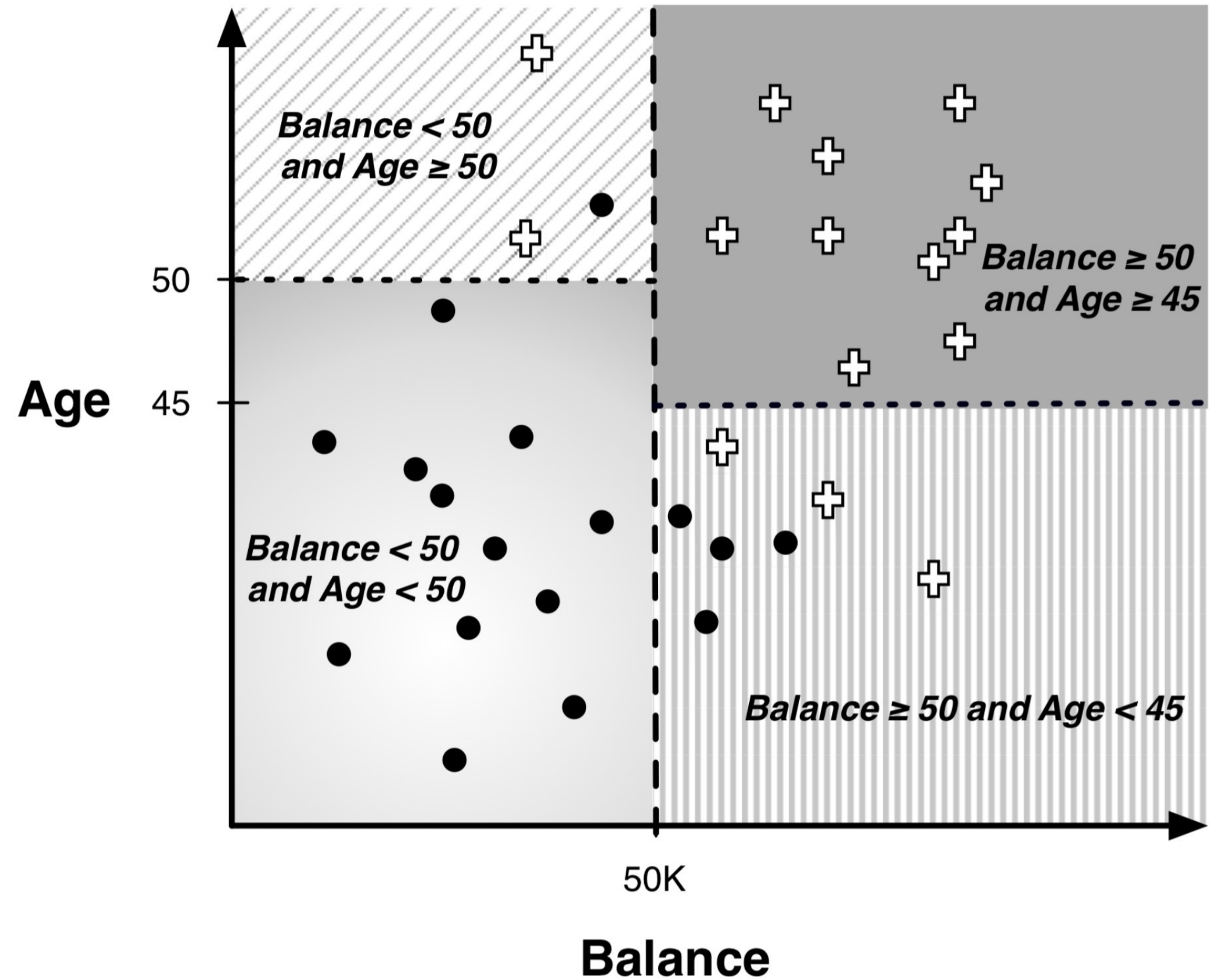
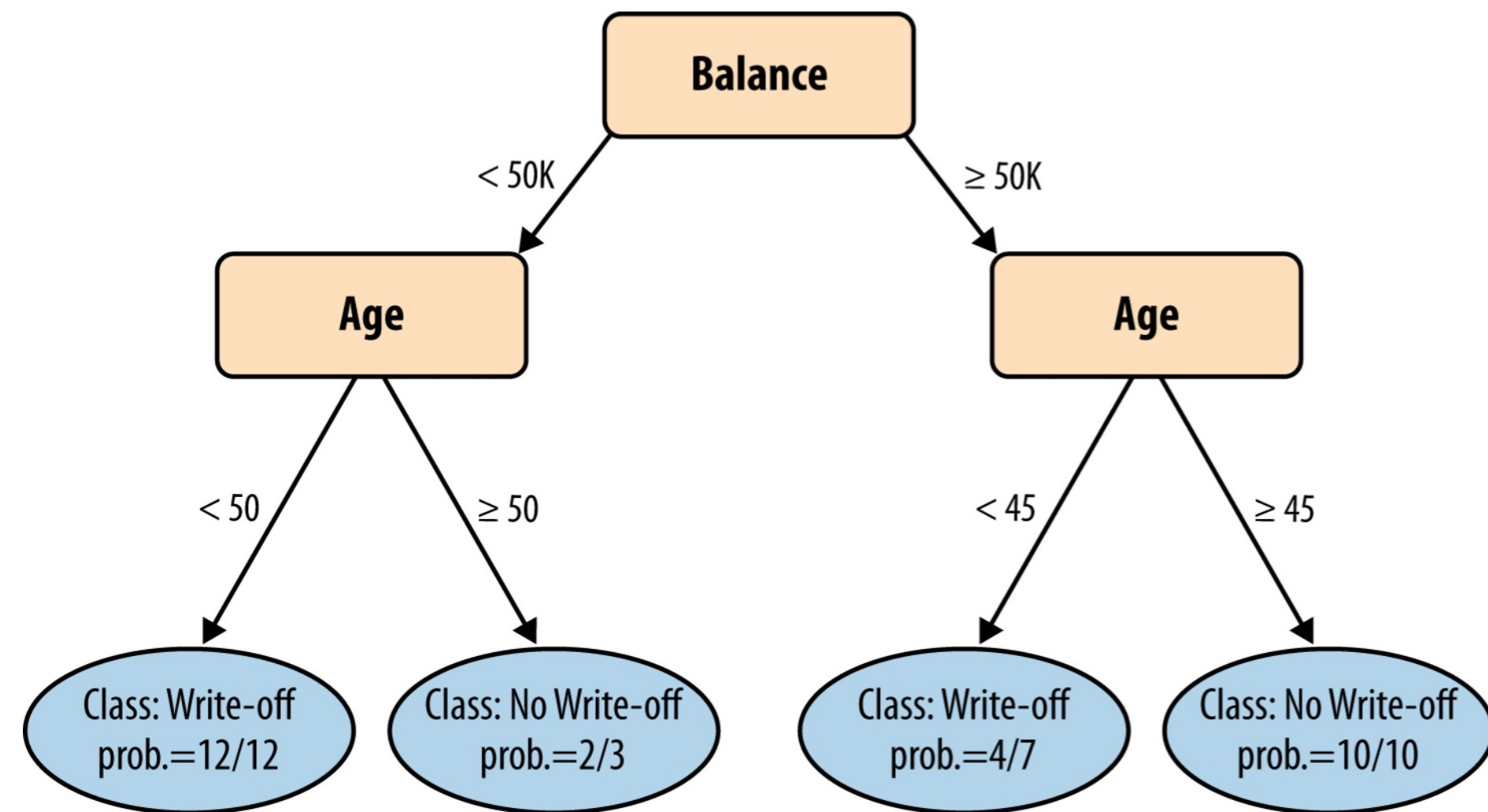


3. ENSEMBLE MODELING

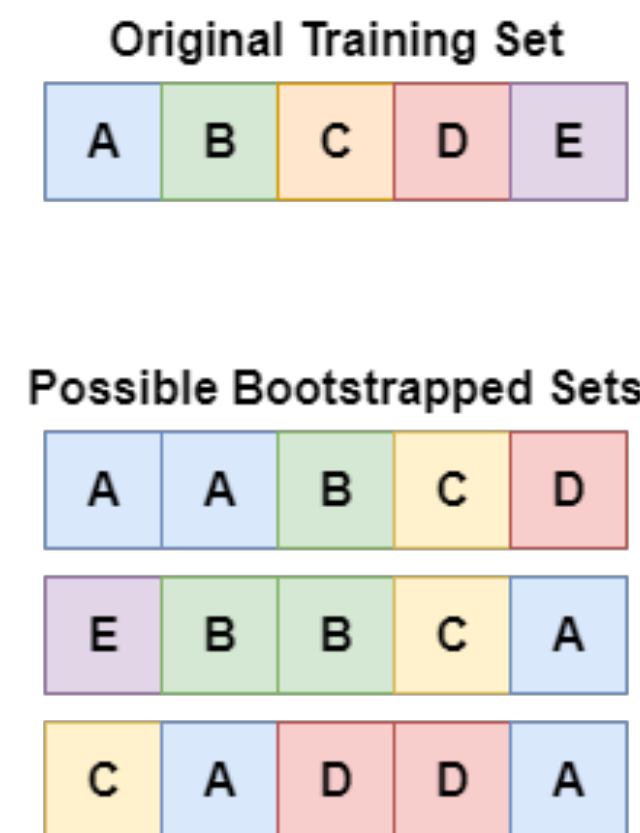
LEV KIWI

ARBRE DE DÉCISION



MÉTHODES ENSEMBLISTES

Les approches classiques



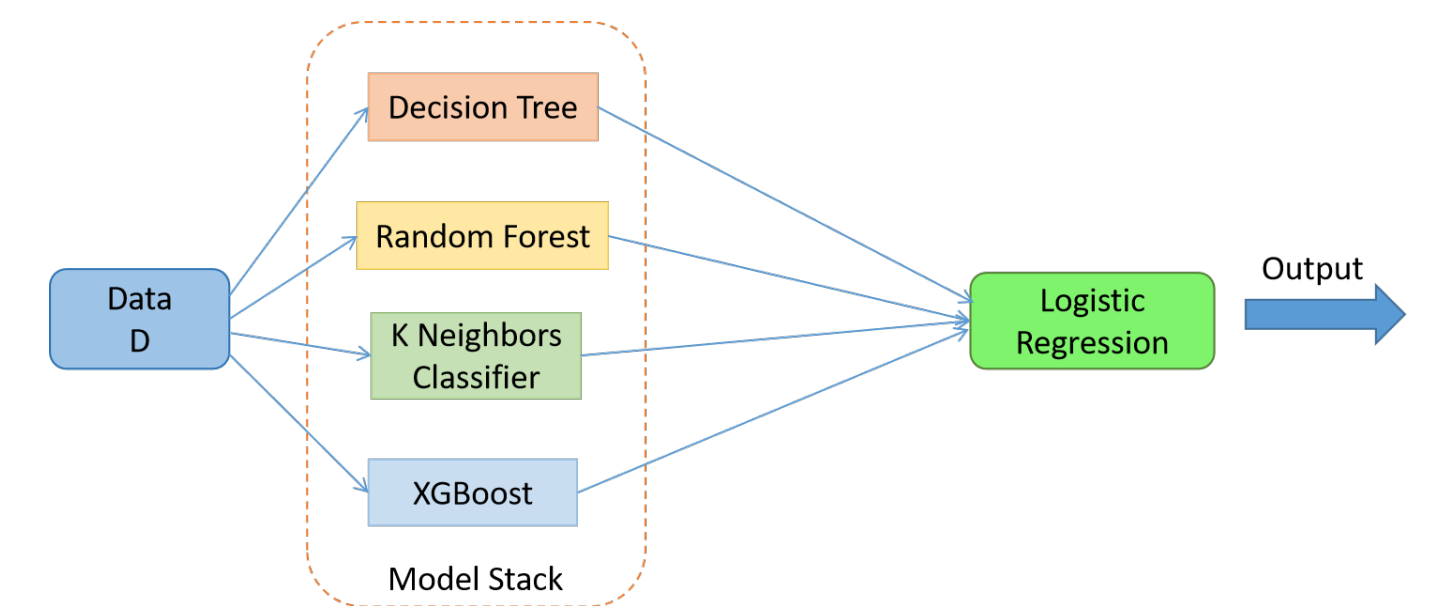
BAGGING

Le dataset est modifié afin d'entraîner un algorithme de manière différente. Le modèle final est assemblé en prenant la tendance centrale de tous les modèles



BOOSTING

Le même modèle est réentraîné de manière séquentielle en corrigeant l'erreur faite précédemment.



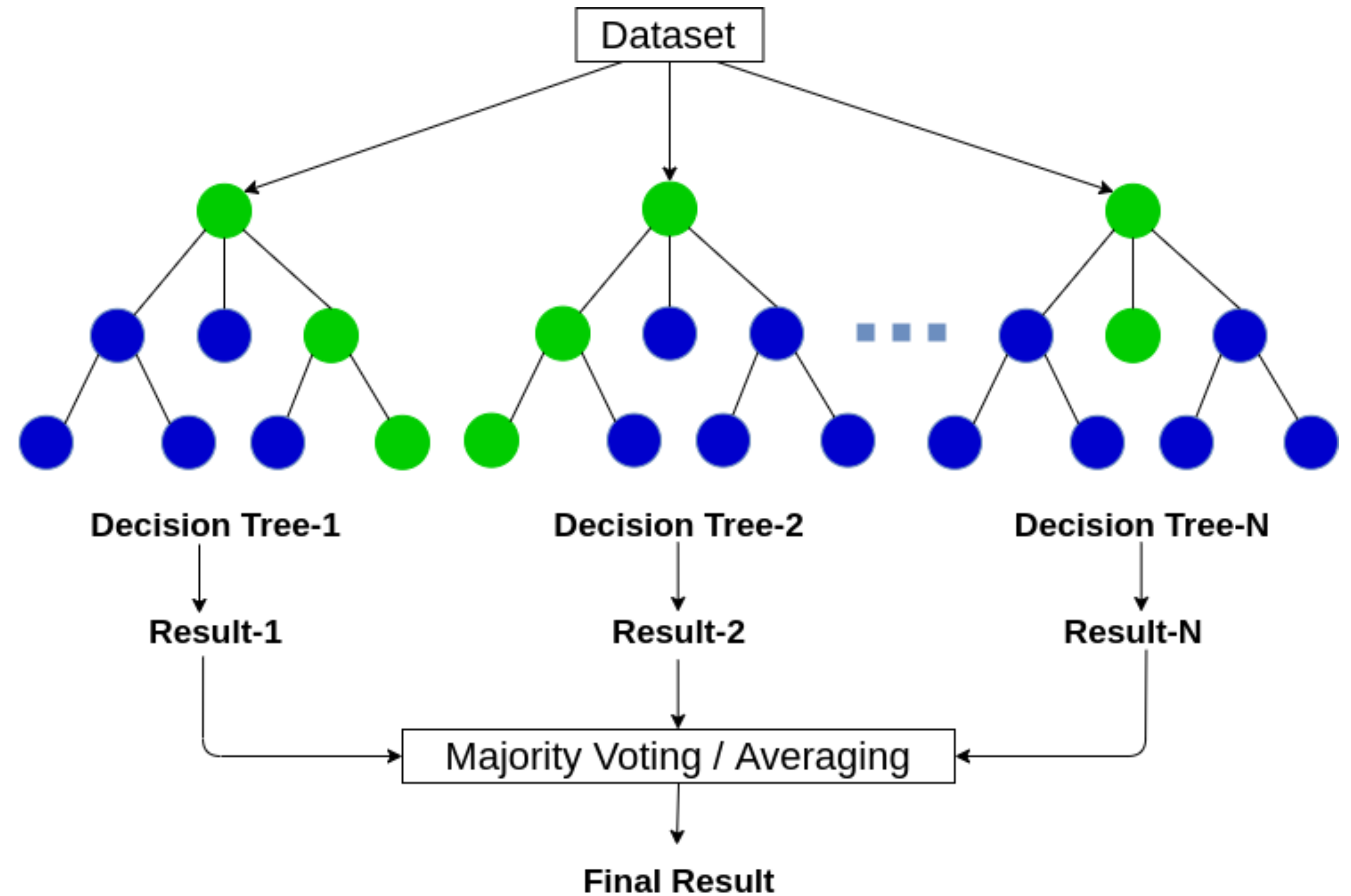
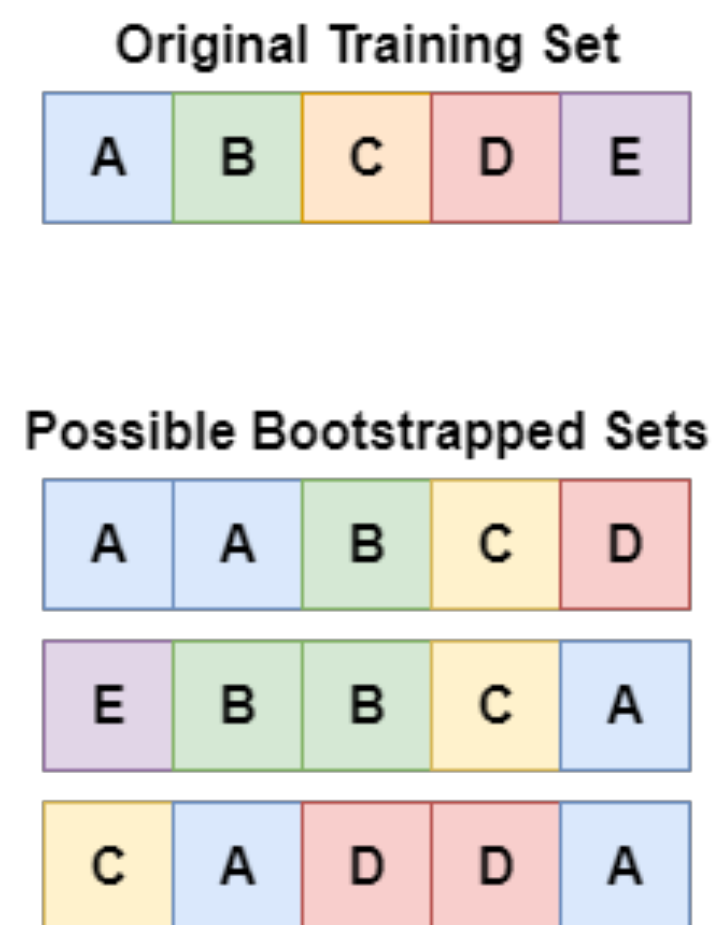
STACKING

Des modèles différents sont entraînés sur le même dataset, puis assemblés avec un méta-modèle

RANDOM FOREST

Procédure

1. Le dataset est décomposé en autant d'arbre qu'on souhaite avoir dans notre forêt.
2. Chaque arbre est entraîné sur une dataset bootstrap.
3. Le résultat final est obtenu en prenant la moyenne de tous les arbres.



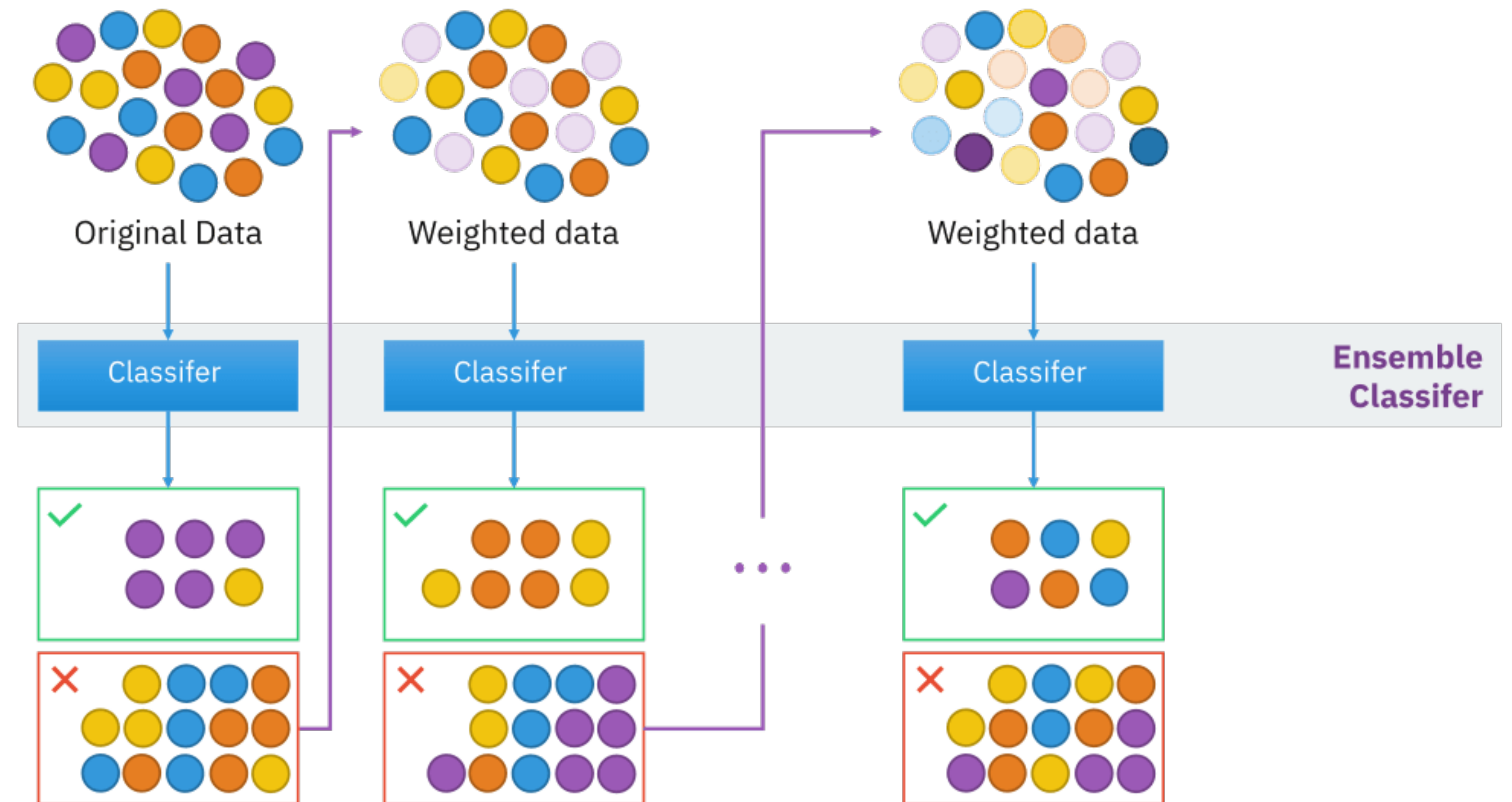
ADABOOST

Procédure

0. Initialisation des poids ω_i pour tous les points du dataset

Pour $k = 1, \dots, n$

1. Entraîner un modèle faible M_k
2. Calculer l'erreur de chaque modèle M_k et leur assigner une importance relative α_k
3. Mettre à jour les poids ω_i relativement à l'erreur globale des modèles



GRADIENT BOOSTING

Procédure

Similaire à AdaBoost sauf que les erreurs sont propagées dans les itérations suivantes.

Remarque. Le gradient boosting est généralement plus robuste qu'AdaBoost et moins sensible aux outliers.

Variants

XGBoost. Ajoute de la régularisation afin de mieux contrôler l'overfitting (cf. cours 5)

LightGBM. Modifie la stratégie d'entraînement afin d'être plus efficace avec de grands datasets

