

— Problématique :

Après sa création, l'un arbre de décision est généralement exposé à deux problèmes, à savoir : sa complexité (en terme de taille), ainsi que le problème du sur-apprentissage. Faisant partie des algorithmes d'apprentissage supervisé, les arbres dépendent naturellement d'un ensemble d'apprentissage afin de construire leur modèle prédictif.

Le sur-apprentissage se caractérise par une très bonne performance de l'arbre sur cet ensemble de données d'apprentissage, ainsi qu'une mauvaise sur d'autre ensemble (ensemble de teste).

Afin de palier a ce problème, l'arbre de décision doit être élagué, cet élagage peut se faire de deux manières différentes :

1. Pre-Elagage

Comme son nom l'indique, cette technique permet d'élaguer l'arbre pendant sa création, et ce, en ajoutant des conditions d'arrêts, ce qui permettra l'arrêt du développement d'un sous ensemble de données avant d'atteindre l'homogénéité totale, si la condition d'arrêt est vérifiée a son niveau.

la condition pourrait être (entre autres) :

— Un seuil minimal d'instance

Qui, une fois atteint, le nœud est considéré comme étant une feuille.

— Longueur maximale de l'arbre

Ou seuls les nœuds ne dépassant pas cette longueur maximal seront développés.

2. Post-Elagage

Concernant cette technique, l'élagage de l'arbre se fait qu'après sa création,, et ce, en supprimant des sous arbre complet, en les remplaçant par les feuilles qui représentent la classe la plus fréquente au niveau de ce sous arbre.

Le parcours se fait de la racine a la feuille, pour chaque nœud (feuille ou interne), la précision de l'arbre est évaluée (calculée) avant et après sa suppression, si la différence est grande, le sous arbre sera élagué et remplacé donc par une feuille.

Éventuellement, d'autres nœuds seront perdus lors de cette suppression, cela ne va pas affecté la précision, qui a la base devait être augmentée?

Avec le calcul et la comparaison de la précision de l'arbre avant et après la suppression du sous arbre, Si la précision de l'arbre était meilleur (égale ou légèrement moins bonne) avec ces nœuds, le sous arbre n'aurait pas été supprimé, nous sommes donc sûr que la suppression du sous arbre ne ferai qu'améliorer la précision de l'arbre