

Classification Supervisée

Le principe de la classification supervisée est de pouvoir mettre des étiquettes sur les différents échantillons d'un jeu de données. Pour pouvoir avoir des jeux de données de qualité pour que les algorithmes utilisant ces jeux de données aient un entraînement correct.

Les différents types d'algorithmes supervisés sont classés en fonction du type de donnée avec lesquels ils ont été nourris.

Les différents type d'algorithmes sont :

- **Répartition en k-means** : Pour cela, il établit une moyenne de référence parmi un jeu de données, ce qui permet alors de définir un profil type. Son avantage : sa précision. Et ce, même en traitant de grand volume de données rapidement.
- **Régression linéaire** :
La régression linéaire est utilisée pour prédire des valeurs en fonction de variable prédictive (prix d'une voiture en fonction de l'année, kilométrage, nombre de propriétaire par exemple).
- **Random Forest** :
Une Random Forest est plusieurs arbre de décision d'où le nom Forest. On va donc combiner plusieurs arbre pour avoir un résultats non pas plus précis mais plus robuste c'est-à-dire qu'il sera un peu plus fiable que l'arbre de décision. Son côté négative est qu'il peut prendre beaucoup de temps à calculer.

- **Arbre de décision** :

L'algorithme d'arbre de décision est utile pour résoudre les problème de classification et de régression. L'arbre est structuré en différent partie. Commenant par le nœud racine, celle-ci reprenne la première question de la condition. Puis on a les nœud internes chaque nœud interne représente une question ou un test sur une caractéristique. Puis on a les branches ce sont les différents réponse possible aux questions. Puis on a les feuilles ceux si sont les réponses finaux des questions. On peut se représenter l'arbre de décision comme le jeux de devinette Akinator dans laquelle il nous pose des questions et il essaye de deviner un personnage.