

Python for data-science

Devoir à la maison

Hugo PIERRE

Présentation

- La dataset représente une liste de client d'une banque.
- Plusieurs informations sont disponibles sur ces derniers, comme la durée du dernier appel passé avec lui, quand, ce qu'il fait dans la vie, et si il a fait un placement après cette appel
- Les features exact :
 - travail, age, situation, étude, prêt conso, prêt immobilier, à découvert,
 - portable ou fixe, quand était ce dernier appel, durée du dernier appel
 - combien de contact avant cet appel pour cette campagne, nombre de jours depuis, combien de contact avant cet appel pour la campagne précédente, résultat de la campagne précédente (succès ou refus)
 - il y a-t-il eu un placement monétaire de fait

Quoi prédire

- Avec quels paramètres le client était le plus à même de faire un placement
- La récupération a été faite directement sur le site en téléchargeant un zip
- Il y a eu un parcours des données pour savoir exactement de quoi était composé la dataset
- Il y a eu un nettoyage de fait (on enlève les “unknown”, les Nan, les null etc..)
- En regardant différents graphiques, on peut voir qu’il y a des features qui ressortent pour pouvoir prédire le résultat (comme la durée de l’appel)
- J’ai effectué une classification par arbre
- Après plusieurs tests, il ressort que la profondeur idéale est 4

Résultat

- Score de test :
 - depth 1 : 0.86
 - depth 2 : 0.89
 - depth 3 : 0.9020
 - depth 4 : 0.9022
 - depth 6 : 0.9012