

Description du problème

Vous êtes engagé par le directeur des ressources humaines d'une entreprise pour enquêter sur les raisons pour lesquelles les employés démissionnent. Il vous a fourni une base de données sur les employées de l'année précédente. Il veut comprendre quels sont les indicateurs les plus pertinents pour savoir qui envisage de quitter.

Vous devez travailler avec l'ensemble d'apprentissage (train.csv) pour analyser les données et entraîner des algorithmes de classification. Une fois satisfait avec les algorithmes, vous allez les utiliser pour prédire les résultats (si la personne quittera l'entreprise ou non) dans l'ensemble de test (test.csv).

Description des variables

- Id - numéro d'identification unique
- satisfaction_level - niveau de satisfaction de l'employé
- last_evaluation - score dans la dernière évaluation
- number_project - le nombre de projet dans lesquels l'employée participe
- average_monthly_hours - la moyenne des heures mensuelles
- time_spend_company - temps avec l'entreprise
- Work_accident - accident du travail
- left - quitter l'entreprise - variable cible
- promotion_last_5years - une promotion au cours de 5 dernières années
- department - le département
- salary - niveau de salaire