

## 7. Un test de classification de bout en bout

Comme dernière étape, vous recevez un ensemble de données appelé `data/ccdefault.csv`. L'ensemble de données représente le défaut des clients des cartes de crédit. Il a 30 000 cas et 24 attributs différents. Plus de détails sur l'ensemble de données sont disponibles sur `data/ccdefault.txt`. Dans cette tâche, vous devez créer trois modèles, comparer leurs résultats et trouver la solution idéale. Voici les étapes suggérées :

1. Chargez les données.
2. Effectuez des analyses exploratoires (par exemple, comment les diverses caractéristiques et la variable cible sont distribuées).
3. Entraînez un modèle pour prédire la variable cible (risque de "par défaut").
  - Utiliser trois modèles différents (régression logistique, arbre de décision et forêt aléatoire).
  - Comparer les performances des modèles (par exemple, AUC).
  - Défendre votre choix du meilleur modèle (par exemple, quelles sont les forces et les faiblesses de chacun de ces modèles ?).
4. Que feriez-vous de plus avec ces données ? Quelque chose pour vous aider à trouver une meilleure solution ?
5. Mettre votre solution sur github et m'envoyez le lien.
6. Possibilit