

Test Technique Crédit Agricole Brie Picardie

Contexte et objectif :

Vous avez à disposition l'historique de recrutement de l'entreprise CANDI-DATA spécialisée dans la production et livraison de produit Data-centric.

Dans le cadre d'un audit interne, le PDG souhaite s'assurer que ses recruteurs sont objectifs dans la sélection des candidats. Ainsi, il vous a demandé de mener l'enquête et de lui présenter vos conclusions.

Dans un second temps, le PDG souhaite aussi améliorer le processus de recrutement grâce à l'intelligence artificielle.

Livrables attendus :

- Un dossier zippé contenant votre Jupyter Notebook/ ou Google Colaboratory (écrit en python) avec tous les travaux que vous jugez nécessaires pour réaliser les 2 demandes du PDG. Vous pouvez éventuellement inclure un readme si nécessaire (veillez à garder les outputs de toutes vos cellules Jupyter).
- Une synthèse des différentes analyses réalisées dans le cadre de l'audit interne (incluse directement dans le Jupyter Notebook ou dans un fichier à part).

Pour vous guider :

- L'historique de recrutement est le fichier csv → Data.csv que vous trouverez joint à cet énoncé. La signification des colonnes :
 - Date : Date de candidature
 - Couleur des yeux : Couleur des yeux du candidat
 - Age : âge du candidat en année

- Exp : nombre d'années d'expérience professionnelle du candidat
- Salaire : le salaire demandé par le candidat
- Sexe : Sexe du candidat
- Diplôme : Diplôme du candidat
- Métier : Métier auquel le candidat à postuler
- Note : Note du test technique passé par le candidat (de 0 à 100)
- Dispo : « oui » si disponible immédiatement, sinon « non ».
- Embauche : si le candidat a été embauché ou non
- Ville d'origine : Ville de naissance du candidat :
 - 1 : Paris
 - 2 : Nancy
 - 3 : Lyon
 - 4 : Marseille
 - 5 : Barcelone
 - 6 : Séville
 - 7 : Madrid
 - 8 : Valence
 - 9 : New York
 - 10 : San Francisco
 - 11 : Miami
 - 12 : Chicago
 - 13 : Los Angeles
 - 14 : Washington
 - 15 : Las Vegas
- Quelques pistes de réflexion pour la partie 1 (audit interne) :
 - Est-ce que les données à disposition peuvent expliquer en partie ou non l'embauche d'un candidat ? (Vous pouvez présenter l'analyse sous forme de graphe, distribution, test statistique ou tout autre format que vous jugez pertinent)

- **(Questions bonus)** Y a-t-il une dépendance statistiquement significative entre :
 - Le sexe et le métier ?
 - Le salaire demandé et la couleur des yeux ?
 - Le nombre d'années d'expérience et la note du test technique ?

- Pour la partie 2 (usage de l'intelligence artificielle pour aider aux recrutements) :
 - Concevoir un modèle qui prédit si un candidat sera recruté ou non.
 - Quelles sont les variables les plus importantes de votre modèle ? Qu'en pensez-vous ?
 - Pensez-vous que votre modèle pourra être utilisé par l'équipe recrutement ?
 - Proposez des pistes d'améliorations de votre modèle.
 - Proposez d'autres méthodes statistiques/machine learning pour répondre aux enjeux du PDG.