

Test Technique

Stage Data Science

Algorithme de classification

Le but de cet exercice est de classer des offres d'emploi par type de métiers. Vous avez à disposition deux jeux de données, un jeu d'apprentissage pour développer votre algorithme de classification et un jeu de test, qui nous permettra d'évaluer les résultats de votre algorithme.

Nous nous limitons ici aux métiers de la banque-finance, il existe 26 labels différents :

Métier	Id
Assurance	24
Audit	30
Audit & Expertise Comptable	29
Banque de Détail	2
Clientèle Entreprise	15
Commercialisation	26
Conception et adaptation de produits	25
Conseil	32
Conseil en organisation et système d'information	34
Conseil en Stratégie et Management	33
Conseil et Ingénierie Financière	13
Conseil Opérationnel	35
Contrôle et conformité	432
Direction Financière	37
Expertise Comptable	31
Finance de Marchés	5
Finance d'Entreprise	11
Financement	12
Fonction centrales et risques	16
Fonctions Support	4
Front-Office des Banques	6
Gestion	27
Gestion d'actifs	19
Risk Management	17
Services Satellites	10
SSII	36

Vous êtes libre sur le choix de l'algorithme, le langage de programmation doit être exclusivement **R** ou **Python**.

MULTIPOSTING

Données

Les données sont composées de deux jeux comprenant :

- **Les données d'apprentissage** (data_train.csv) : 12 000 annonces labellisées (id, title, description et label)
- **Les données de test** (data_test.csv) : 300 annonces non labellisées (id, title et description)

id : id unique

title : titre de l'annonce

description : description du poste

label : label correspondant à un métier

Evaluation : les réponses sont évaluées sur la qualité du code, l'algorithme utilisé et les résultats.

Réponse : Veuillez joindre dans votre réponse **le code** utilisé ainsi que **les données de test classifiées** sous le format CSV suivant :

```
id, job
70828, 7
70829, 9
71193, 2
...
```