

# Projet — Données RH

David A. Ouedraogo   Senior Data Scientist / AI Engineer  
Formateur Data Scientist / IA Générative

ESIEA

17 novembre 2025

**Contexte** : L'entreprise souhaite mieux **recruter**, **fidéliser** et **développer ses talents**. Les données RH actuelles peuvent être exploitées pour :

- comprendre les profils d'employés existants
- prédire le type de poste ou de contrat adapté
- détecter les risques de départ et optimiser les politiques RH

## Objectifs Data Mining

- 1 **Classification** : prédire le type de contrat d'un employé
- 2 **Clustering** : trouver des groupes homogènes d'employés

**Valeur métier** : optimiser le recrutement et la gestion des talents.

# Description du dataset RH

**Fichier source :** dataset\_rh.csv **Nombre d'employés :** ~ 1200

**Domaine :** Ressources Humaines

| Colonne      | Description                                |
|--------------|--|
| EmployeID    | Identifiant unique de l'employé            |
| Nom          | Nom de famille de l'employé                |
| Prenom       | Prénom de l'employé                        |
| Age          | Âge de l'employé                           |
| Sexe         | Genre : Homme / Femme                      |
| Departement  | Service ou unité : IT, Marketing, RH, etc. |
| Poste        | Intitulé du poste occupé                   |
| Salaire      | Rémunération annuelle ou mensuelle         |
| NiveauEtude  | Niveau d'études (Bac, Licence, Master...)  |
| TypeContrat  | CDI, CDD, Stage, Alternance                |
| DateEmbauche | Date d'entrée dans l'entreprise            |

Variables disponibles dans le dataset RH

# Réflexion : Quel problème Data Mining en RH ?

Avec ces données RH, plusieurs besoins métier peuvent émerger :

- Recruter les **bons profils**
- Fidéliser les **talents clés**
- Optimiser les **coûts salariaux**
- Affecter les **bons profils aux bons postes**

**Question : Quel objectif analytique peut-on définir ?**

## Quels usages RH de nos données ?

- ① Prédire le **Type de Contrat** d'un employé
- ② Détecter les profils à **risque de départ**
- ③ Segmenter les employés selon leurs caractéristiques
- ④ Analyser l'équité **salariale**
- ⑤ Identifier les profils à **fort potentiel**

→ Plusieurs techniques possibles : classification, clustering, régression...

L'entreprise souhaite :

- Réduire le **turnover**
- Améliorer le **recrutement**
- Mieux connaître ses profils

**Conclusion** : Nous allons nous concentrer sur des analyses orientées **talent management**.

## Objectif : prédire le type de contrat d'un employé

- Sortie (cible) : CDI / CDD / Alternance / Stage
- Entrées : âge, diplôme, ancienneté, salaire, métier, etc.
- Algorithme choisi : **Arbre de décision**

**Valeur métier** : anticiper les besoins en emploi et fidélisation.

## **Objectif : regrouper les employés en profils homogènes**

Exemples d'interprétations :

- Junior / Senior
- Technique / Commercial / Support
- Haut salaire / Bas salaire

**Valeur métier** : segmentation RH pour recrutement, formation et rétention.



Faites votre choix entre les 2 objectifs suivants :

## **Objectif principal 1 : Classification**

- Construire un modèle prédictif du type de contrat
- Évaluer la précision sur les nouveaux employés

## **Objectif principal 2 : Clustering**

- Identifier naturellement des profils d'employés
- Interpréter les groupes pour recommandations RH