

Projet 2 : Prédiction de départs de salariés (Employee Attrition Prediction)

Objectif

Prédire si un employé risque de quitter l'entreprise à partir de données RH à l'aide de Spark MLlib en Scala.

Jeux de données

Dataset "HR Analytics" (par exemple sur Kaggle) contenant des colonnes comme :

- Age
- JobRole
- MonthlyIncome
- JobSatisfaction
- OverTime
- Attrition (cible : Oui / Non)

Étapes du projet

1. **Chargement des données**
2. **Prétraitement des données**
 - Supprimer les valeurs nulles
 - Encodage des colonnes catégorielles
3. **Sélection des colonnes et assemblage**
4. **Entraînement du modèle**
5. **Evaluation**
6. **Evaluation du modèle**
7. **Améliorations possibles :**
 - Cross-validation
 - Grid Search
 - Changement d'algorithme (RandomForest, GBTClassifier, etc.)