

Projet de Fouille de Données

Projet de Fouille de Données : Analyse des Tendances du Marché de l'Emploi Français en Temps Réel

1. Contexte et Objectif du Projet

Le marché de l'emploi français évolue rapidement, influencé par les mutations technologiques, les politiques économiques et les transformations sociétales.

Ce projet vise à **analyser en temps réel les offres d'emploi publiées sur France Travail (ex Pôle Emploi)** afin d'identifier :

- Les tendances salariales selon les secteurs et les régions,
- Les profils types d'offres (types de contrats, exigences, qualifications),
- Les combinaisons de compétences les plus recherchées.

L'objectif global est de proposer une **analyse data-driven** du marché du travail français, appuyée sur des méthodes de fouille de données modernes.

2. Domaine d'Application

- **Thématique** : Marché du travail et recrutement en France
 - **Objectif opérationnel** : Comprendre la dynamique du marché (salaires, compétences, types de contrats, répartition géographique).
-

3. Source de Données

- **API France Travail (ex Pôle Emploi)**
 - Données ouvertes et structurées sur les offres d'emploi.
 - Variables clés : intitulé du poste, secteur, type de contrat, salaire, localisation, compétences, expérience requise, avantages.

- Données collectées en **temps réel** via requêtes API.
-

4. Méthodologie d'Analyse

L'analyse repose sur **trois approches de fouille de données** enseignées durant le cours : classification supervisée, classification non supervisée et recherche de règles d'association.

4.1 Classification Supervisée → Régression

Objectif : prédire le **salaire brut annuel** d'une offre à partir de ses caractéristiques.

Variables explicatives :

- Secteur ou domaine métier
- Localisation (région, département)
- Type de contrat (CDI, CDD, alternance)
- Expérience requise (junior, confirmé, senior)
- Compétences mentionnées

Méthodes envisagées :

- Régression linéaire multiple
- Arbre de régression
- Random Forest Regression

Résultats attendus :

- Identifier les **facteurs les plus influents sur le salaire**.
 - Visualiser la **variation des salaires moyens** par secteur et région.
-

4.2 Classification Non Supervisée → Clustering

Objectif : regrouper les offres selon leurs **profils caractéristiques**.

Exemples de clusters attendus :

- Offres à **fort salaire** : CDI, Île-de-France, domaine Tech.
- Offres à **salaire moyen** : Commerce, Santé, CDD, régions urbaines moyennes.

- Offres à **bas salaire** : Emplois saisonniers ou à temps partiel.

Méthodes envisagées :

- K-Means
- Classification ascendante hiérarchique (CAH)
- Visualisation avec ACP (Analyse en Composantes Principales)

Résultats attendus :

- Identifier les **profils types du marché de l'emploi**.
 - Aider à la **segmentation géographique et sectorielle** des offres.
-

4.3 Règles d'Association → Analyse des Co-Compétences

Objectif : découvrir **quelles compétences apparaissent souvent ensemble** dans les offres.

Exemples de règles attendues :

- {Python} → {SQL}
- {Linux, Docker} → {CI/CD}
- {Excel} → {Power BI}

Résultats attendus :

- Identifier les **compétences complémentaires** les plus fréquentes.
 - Construire un **graphe des relations entre compétences**, utile pour :
 - les candidats (plan de montée en compétences),
 - les formateurs (conception de parcours cohérents).
-

5. Valeur Ajoutée du Projet

Ce projet permettra de :

- ✓ Fournir une **vision en temps réel** du marché français de l'emploi.
- ✓ Identifier les **facteurs qui influencent les salaires**.
- ✓ Distinguer les **profils types d'offres** par région et secteur.
- ✓ Révéler les **synergies entre compétences** les plus valorisées.
- ✓ Offrir des **insights exploitables** aux chercheurs d'emploi, recruteurs, et acteurs de la formation.

6. Perspectives et Extension

À plus long terme, ce projet pourrait être enrichi par :

- L'intégration d'autres plateformes (Indeed, APEC).
- L'analyse de la **dynamique temporelle** des salaires (tendances mensuelles).