### **Titre du projet**

**"Analyse des avis des employés : évaluation de l'environnement de travail"**

### **✅ Objectif du projet**

Fournir un outil simple et visuel pour aider les candidats à mieux comprendre la culture et l’environnement de travail dans différentes entreprises, à travers une analyse automatisée des avis laissés par leurs employés (positifs et négatifs).

### **💡 Problématique**

Aujourd’hui, de nombreux candidats consultent des sites comme Glassdoor ou Indeed pour lire les avis d’employés avant de postuler. Cependant, ces avis sont souvent nombreux, non structurés, et difficiles à analyser rapidement.  
 **Comment peut-on extraire automatiquement les thèmes clés et le sentiment général autour d’une entreprise pour faciliter la prise de décision du candidat ?**

### **📦 Données utilisées**

Des avis anonymes sur les entreprises contenant :

* Un titre ("headline")
* Les points positifs ("pros")
* Les points négatifs ("cons")

Chaque avis peut aussi être lié à une entreprise, une date, un poste, etc.

### **🛠️ Solutions mises en place**

* **Nettoyage des données** : traitement des valeurs manquantes, fusion des champs texte.
* **Prétraitement NLP** : lower casing, lemmatisation, suppression des stopwords (pour le topic modeling).
* **Création d’embeddings** : via des modèles transformers pré-entraînés (Sentence Transformers).
* **Topic Modeling** : clustering des thèmes dominants dans les avis (positifs et négatifs).
* **Analyse de sentiment** : entraînement d’un modèle supervisé pour classifier automatiquement les avis comme positifs ou négatifs.
* **Dashboard final (Streamlit ou similaire)** : interface interactive pour explorer les entreprises.

### **👀 Démonstration du produit final (MVP)**

* Sélection d’une entreprise dans une liste
* Affichage des **principaux thèmes positifs et négatifs** (sous forme de nuage de mots ou de graphiques)
* Résumé du **sentiment global** de l’entreprise
* Liste des **3 avantages et 3 inconvénients les plus cités**
* (Optionnel) Comparaison entre entreprises du même secteur

### **🌍 Impact potentiel dans le monde réel**

* Pour les **candidats** : prendre des décisions plus éclairées avant de postuler.
* Pour les **équipes RH** : identifier les points d’amélioration fréquents dans les retours internes.
* Pour les **plateformes de recrutement** : proposer un résumé visuel instantané de la réputation d’une entreprise.
* Pour les **data analysts** : démonstration concrète d’un pipeline de NLP appliqué à un cas réel.

### **🚀 Axes d’amélioration / prochaines étapes**

* Prise en compte du **temps** : évolution du sentiment dans le temps.
* Comparaison par **poste ou localisation**.
* Passage à une **analyse multilingue**.
* Intégration de **nouveaux modèles fine-tunés** pour améliorer la précision du sentiment.