

Proposition de Projet : Système Intelligent de Tri de CV

1. Idée du Projet

L'objectif est de développer un système intelligent de tri de CV basé sur l'intelligence artificielle pour automatiser et optimiser le processus de sélection des candidats dans le recrutement. Le problème ciblé est la perte de temps et les biais humains dans l'évaluation manuelle des CV, qui ralentissent le recrutement et peuvent exclure des candidats qualifiés. Ce système analysera les CV pour identifier les compétences, expériences et qualifications pertinentes, et les classera selon leur adéquation aux exigences d'un poste spécifique.

2. Pertinence par rapport aux ODD

Ce projet s'aligne sur l'ODD 8 : Travail décent et croissance économique, en favorisant une inclusion équitable dans le processus de recrutement. En réduisant les biais liés à l'évaluation humaine (genre, origine, etc.), le système promeut l'égalité des chances et accélère l'accès à l'emploi, contribuant ainsi à la réduction des inégalités (ODD 10). De plus, en optimisant le recrutement, il soutient la productivité des entreprises, favorisant la croissance économique.

3. Exemples de Littérature

- **Gugnani et al. (2020), "Resume Screening using NLP and Machine Learning"** : Cet article décrit un système utilisant le traitement du langage naturel (NLP) pour extraire des informations clés des CV et les classer via des modèles d'apprentissage machine. Il inspire notre approche pour l'extraction des compétences et l'évaluation automatisée.
- **Zhang et al. (2022), "Deep Learning for Resume Parsing and Matching"** : Ce travail utilise des réseaux neuronaux profonds pour analyser des CV et les associer à des descriptions de poste. Il met en évidence l'efficacité des modèles d'apprentissage profond pour traiter des données textuelles complexes, ce qui est pertinent pour notre projet.

4. Description des Données

Les données proviendront de bases de CV anonymisées disponibles publiquement (par exemple, Kaggle ou bases de données de recrutement open-source) et de descriptions de postes fournies par des entreprises partenaires. Les CV seront au format texte (PDF convertis en texte ou fichiers .docx), avec une taille estimée de 10 000 à 50 000 CV pour l'entraînement. Les étapes de prétraitement incluront la tokenisation, la suppression des mots vides, la normalisation du texte (minuscules, suppression des caractères spéciaux) et l'extraction d'entités nommées (compétences, expériences, formations).

5. Approche

Une approche d'apprentissage profond sera adoptée, en utilisant un modèle de traitement du langage naturel basé sur des transformers (par exemple, BERT). Cette méthode est justifiée par la complexité des données textuelles des CV, qui nécessitent une compréhension contextuelle fine pour extraire des informations pertinentes et les comparer aux descriptions de poste. Les modèles traditionnels d'apprentissage machine (par exemple, SVM ou arbres de décision) seraient moins performants pour capturer les nuances sémantiques des textes non structurés.