
Université Ibn Tofaïl

Faculté des Sciences, Kénitra

Mémoire de Projet de Fin d'Etudes

Master Intelligence Artificielle et Réalité Virtuelle

**Développement d'un Système Intelligent de Recrutement Assisté par
L'intelligence Artificielle**

Établissement d'accueil : La société MIDL

Elaboré par : Mlle. AOUATIF CHERDID

**Encadré par : Mr. TARIK BOUJIHA (ENSA Kénitra -UIT)
Mr. WALID BRIOUA (MIDL France, Paris)**

Soutenu le 26 Septembre 2024, devant le jury composé de :

- Mr Anas NOURI (FSK Université Ibn Tofaïl)
- Mme Raja TOUAHNI (FSK Université Ibn Tofaïl)
- Mr Idriss MOUMEN (FSK Université Ibn Tofaïl)
- Mme TARIK BOUJIHA (ENSAK Université Ibn Tofaïl)

Dédicace

Je souhaite exprimer ma reconnaissance en dédiant ce projet à plusieurs personnes qui ont joué un rôle essentiel dans sa réalisation.

À mes parents, dont le soutien inconditionnel et les encouragements constants ont été une source d'inspiration tout au long de mes études et de mes aspirations.

À mes professeurs et mentors, dont la sagesse et les conseils avisés m'ont permis de développer mes compétences et de progresser dans ce projet.

À mes amis et collègues, dont le soutien indéfectible et la motivation m'ont accompagné à chaque étape de ce projet, faisant de cette expérience une aventure collective enrichissante.

À toutes les personnes qui ont contribué, de près ou de loin, à ce projet. Leur collaboration et leur expertise ont été précieuses et ont grandement contribué à sa réussite. Je suis profondément reconnaissant envers chacune de ces personnes pour leur soutien et leur encouragement, qui ont été des éléments clés dans la concrétisation de ce projet.

Remerciement

Je tiens à exprimer ma profonde gratitude aux personnes qui ont contribué à la réalisation de ce projet et à mon parcours universitaire.

Tout d'abord, je remercie chaleureusement mes parents pour leur soutien indéfectible tout au long de mes études. Leur amour, leurs encouragements et leur confiance en moi ont été une source de motivation inestimable.

Je suis également reconnaissant envers **M. Vincent Azouni**, le CEO de la société Midl, qui m'a accueilli au sein de son entreprise et m'a offert l'opportunité de travailler dans un environnement stimulant et enrichissant. Son leadership et sa vision ont contribué à rendre cette expérience des plus formatrices.

En outre, je voudrais également remercier **M. Walid Brioua**, mon encadrant chez Midl, pour son appui constant, ses conseils avisés, et son accompagnement tout au long du projet. Son expertise a été déterminante pour la réussite de ce travail.

Je tiens aussi à exprimer ma gratitude envers **M. Tarik Boujiha**, mon encadrant académique, pour son soutien et ses conseils tout au long de mon cursus universitaire. Son expertise et sa passion pour l'enseignement m'ont inspiré et m'ont permis d'approfondir mes connaissances dans le domaine de l'informatique et de l'intelligence artificielle.

Enfin, je remercie toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce projet. Leurs encouragements et leurs suggestions ont été précieux pour moi.

Je suis reconnaissant de la chance d'avoir été entouré de personnes aussi bienveillantes et compétentes tout au long de mon parcours.

Abstract

The manual management of recruitment processes is often inefficient, leading to time loss, matching errors, and limited visibility of candidates' skills. To address these challenges, the **Midl-job** project has developed an innovative application that leverages advanced artificial intelligence technologies to optimize matching between candidates and recruiters.

Candidate and job data were collected via Firebase, then processed using Python libraries such as pandas and NLTK. Technical skills were vectorized using the TF-IDF method, while job titles were transformed into embeddings using the paraphrase-MiniLM-L6-v2 model. Similarity matrices were then calculated, and the K-Means algorithm was used to cluster job titles, thus improving the matching process between candidates and job offers.

Initial results show promising accuracy in matching. **Midljob** offers several versions : the Free version, which provides basic matching functionalities, and the Premium version, which adds personality comparison for a more comprehensive matching experience. Future developments will include the Pro version, which will integrate advanced features such as interview question generation and response analysis, using sophisticated language models to ensure data security.

This report summarizes the work completed as part of my final year project at **Midl**, highlighting our approach to improving recruitment processes through artificial intelligence, while optimizing productivity across various sectors. .

Keywords : Recruitment, Artificial Intelligence, Machine Learning, TF-IDF, Cosine Similarity, K-Means, Data Security, Midljob

Résumé

Le recrutement manuel est souvent un processus complexe et inefficace, confronté à des défis tels que la perte de temps, les erreurs de correspondance, et une visibilité limitée sur les compétences des candidats. Pour relever ces défis, le projet Midljob a développé une application innovante exploitant des technologies d'intelligence artificielle avancées pour optimiser le matching entre les candidats et les recruteurs.

Le développement de l'application a suivi une méthodologie rigoureuse. Les données des candidats et des offres ont été collectées depuis Firebase, puis préparées avec des bibliothèques Python comme pandas et NLTK. Les compétences techniques ont été vectorisées via TF-IDF, tandis que les intitulés de postes ont été transformés en embeddings avec le modèle paraphrase-MiniLM-L6-v2. Des matrices de similarité ont été calculées, et l'algorithme K-Means a permis de regrouper les intitulés de postes en clusters pour améliorer le matching entre candidats et offres.

Pour la version Free, les résultats montrent une performance élevée avec une précision qui permet d'améliorer significativement le processus de recrutement. Pour la version Premium, en plus des fonctionnalités de la version Free, nous avons ajouté la comparaison des personnalités entre les candidats et les recruteurs pour un matching encore plus complet et personnalisé. Pour la version Pro, nous envisageons d'intégrer des fonctionnalités avancées telles que la génération de questions d'entretien et l'analyse des réponses, en utilisant des modèles de langage sophistiqués pour garantir la sécurité des données.

Ce rapport résume les travaux réalisés dans le cadre de mon projet de fin d'études chez Midl, en offrant une vue d'ensemble de notre approche pour améliorer le processus de recrutement grâce à l'intelligence artificielle, tout en visant à rationaliser les processus et à accroître la productivité dans divers secteurs.

Introduction Générale

Le recrutement est un processus fondamental pour les entreprises, jouant un rôle crucial dans la constitution des équipes et la réalisation des objectifs organisationnels. Un recrutement efficace permet non seulement de trouver les talents les plus adaptés à un poste, mais aussi d'assurer une adéquation optimale entre les compétences des candidats et les besoins des entreprises. Cependant, le recrutement traditionnel, souvent manuel et subjectif, présente plusieurs défis majeurs qui limitent son efficacité. Dans un environnement économique en constante évolution, la capacité à attirer et à retenir les meilleurs talents devient un facteur clé de succès pour les organisations. Lorsqu'il est bien mené, le processus de recrutement contribue à réduire les coûts liés au turnover, à accroître la productivité, et à créer un environnement de travail dynamique et innovant. De plus, un bon recrutement est essentiel pour aligner les compétences des employés avec la stratégie à long terme de l'entreprise.

Malgré son importance, le recrutement traditionnel se heurte à plusieurs obstacles, tels que la perte de temps due au tri manuel des candidatures, l'organisation des entretiens et l'évaluation des compétences. Ces tâches chronophages ralentissent l'ensemble du processus. Les biais humains peuvent également entraîner des erreurs dans l'évaluation des candidats, aboutissant à des recrutements inappropriés ou sous-optimaux, et les recruteurs disposent souvent d'une visibilité limitée sur les compétences réelles des candidats, compliquant ainsi la sélection des profils les plus pertinents. Ces défis peuvent conduire à des coûts supplémentaires pour l'entreprise, des retards dans la prise de décision et des difficultés à identifier les candidats les plus qualifiés pour des postes spécifiques.

Face à ces défis, l'intégration de l'intelligence artificielle (IA) dans le processus de recrutement apparaît comme une solution innovante et efficace. L'IA permet d'automatiser les tâches répétitives et d'améliorer la précision des correspondances entre les candidats et les offres d'emploi. Grâce à des algorithmes de machine learning et de traitement du langage naturel (NLP), l'IA peut analyser les compétences des candidats et les exigences des postes afin de générer des correspondances plus précises. En outre, les modèles d'IA minimisent les biais humains en fournissant une évaluation objective des candidatures, tout en offrant aux recruteurs une meilleure visibilité sur les compétences, les expériences et les aspirations des postulants.

Le projet **Midljob** s'inscrit dans cette dynamique en proposant une application innovante qui exploite l'IA pour transformer le processus de recrutement. L'objectif principal de ce projet est de développer un modèle prédictif capable d'identifier les meilleures correspondances entre les profils des candidats et les exigences des postes, en tenant compte des compétences techniques, des expériences et des affinités culturelles.

La version Free de l'application repose sur des techniques telles que la vectorisation des compétences par TF-IDF, le calcul de similarité cosinus et le clustering des titres de postes pour améliorer la précision du matching. En outre, la version Premium enrichit l'expérience en ajoutant une comparaison des personnalités entre les candidats et les recruteurs, offrant ainsi un matching plus complet. Une version Pro est également prévue, intégrant des fonctionnalités avancées, qui seront détaillées dans la section des perspectives de ce rapport. Cette version Pro cherchera à répondre aux exigences les plus élevées des recruteurs en offrant un niveau de précision et d'efficacité inégalé.

Ce rapport est structuré en cinq chapitres pour offrir une vue d'ensemble complète et détaillée du projet Midljob. :

- **Chapitre 1** : Présentation de l'Organisme d'Accueil : Ce chapitre présente l'entreprise Midl, ses activités, et son positionnement stratégique dans le domaine du recrutement.
- **Chapitre 2** : Contexte du Projet : Ce chapitre expose le contexte général du projet, les défis rencontrés dans le recrutement traditionnel, et les objectifs poursuivis.
- **Chapitre 3** : État de l'Art : Ce chapitre passe en revue les avancées récentes dans le domaine du recrutement assisté par l'intelligence artificielle, ainsi que les technologies utilisées dans ce projet.
- **Chapitre 4** : Méthodologie : Ce chapitre décrit en détail les étapes méthodologiques suivies pour le développement du modèle et de l'application Midljob.
- **Chapitre 5** : Résultats : Ce chapitre présente les résultats obtenus, analyse leur pertinence pour le recrutement, et propose des améliorations pour les travaux futurs.

Table des matières

Dédicace	i
Remerciement	ii
Abstract	iii
Résumé	iv
Introduction	1
Table des matières	3
Table des figures	7
Liste des Abréviations	9
1 Présentation de l'organisme d'accueil	10
1.1 Introduction	10
1.2 Présentation de l'entreprise	10
1.3 Fiche d'identité	12
1.4 Contexte RH	13
1.5 L'environnement économique du stage	14
1.5.1 Le Secteur Économique	14
1.6 Conclusion	14
2 Contexte du Projet	15
2.1 Introduction	15
2.2 Contexte du Projet	15

2.3	Problématique	16
2.4	Objectif du Projet	16
2.5	Missions du projet.....	17
2.6	Risques du projet	17
2.7	Conduite du projet Avec Scrum.....	18
2.8	Les points positifs de Scrum.....	18
2.9	Diagramme de GANTT	20
2.10	Conclusion.....	21
3	État de l'Art	22
3.1	Introduction.....	22
3.2	Historique du Recrutement	22
3.3	L'Émergence de l'Intelligence Artificielle dans le Recrutement	23
3.4	Applications de l'Intelligence Artificielle dans le Recrutement.....	24
3.4.1	Le Machine Learning (Apprentissage Automatique)	24
3.4.2	Le Traitement du Langage Naturel (NLP).....	25
3.4.3	Les Systèmes de Recommandation.....	27
3.4.4	L'Analyse Prédictive	27
3.4.5	L'Évaluation Automatisée des Compétences.....	27
3.5	Solutions Existantes	28
3.5.1	Présentation de quelques outils innovants.....	28
3.5.2	Analyse des forces et faiblesses des solutions existantes.....	31
3.6	Les solutions proposées par le projet MidlJob.....	32
3.7	Conclusion.....	34
4	Méthodologie	35
4.1	Introduction.....	35
4.2	Structure de MidlJob.....	35
4.2.1	Architecture du Système	35
4.2.2	Version Gratuite.....	36
4.2.2.1	Collecte et préparation des Données	37
4.2.2.2	Structure et Source des Données	39

4.2.2.3	Attributs des Offres d'Emploi et des Profils de Candidats	39
4.2.2.4	Préparation des Données	42
4.2.3	Utilisation des Embeddings Sémantiques	43
4.2.3.1	Définition des Embeddings	43
4.2.3.2	Présentation du Modèle d'Embeddings "paraphrase-MiniLM-L6-v2"	44
4.2.3.3	Transformation des Intitulés en Embeddings et Clustering	45
4.2.4	Clustering des Intitulés de Poste avec K-Means	46
4.2.4.1	Présentation de l'Algorithme K-Means	46
4.2.4.2	Étapes du Clustering K-Means	46
4.2.4.3	Application des Embeddings dans K-Means	48
4.2.5	Vectorisation des Textes avec TF-IDF	48
4.2.5.1	Présentation de la Méthode TF-IDF	49
4.2.5.2	Calcul de la Méthode TF-IDF	49
4.2.5.3	Application de TF-IDF dans le Projet	50
4.2.6	Calcul des matrices de Similarité	51
4.2.6.1	Similarité Cosinus	51
4.2.6.2	Calcul de la Matrice de Similarité Globale	53
4.2.7	Sélection des Meilleures Correspondances	54
4.2.7.1	Tri des Correspondances par Scores de Similarité	54
4.2.7.2	Extraction des Meilleures Correspondances (Top N)	54
4.3	Version Premium	54
4.4	Choix d'API et Comparaison	55
4.4.1	Choix d'API	55
4.4.2	Comparaison des APIs	57
4.5	Conclusion	58

5 Résultats 59

5.1	Introduction.....	59
5.2	Résultats du Matching dans la Version Free.....	59
5.2.1	Prédiction des Meilleurs Candidats pour un Poste	60
5.2.2	Prédiction des Meilleurs Postes pour un Candidat	64
5.3	Résultats des Tests avec des Données Simulées	67

5.3.1	Prédiction des Meilleurs Candidats pour un Poste	67
5.3.2	Prédiction des Meilleurs Candidats pour un Poste	69
5.4	Résultats du Matching dans la Version Premium.....	72
5.5	Scénario de Présentation des Profils dans l'Application MidlJob.....	74
5.5.1	Visualisation des Profils des Candidats	74
5.5.2	Notification de Correspondance ("It's a Match !").....	76
5.5.3	Interface de Messagerie.....	76
5.6	Conclusion.....	77
Perspectives pour la Version Pro		78
Conclusion générale		80
Bibliographie		81

Table des figures

1.1	L'incubateur Station F	11
1.2	Logo Midl.....	12
1.3	La fiche d'identité de Midl.....	12
1.4	Organigramme de l'entreprise Midl	13
2.1	Architecture du SCRUM	18
2.2	Diagramme de Gantt.....	21
3.1	Interface LinkedIn Talent Solutions.....	29
3.2	l'interface HireVue utilise d'entretiens vidéo basés sur l'IA.....	30
3.3	l'interface de Textio montrant l'analyse d'une offre d'emploi.....	31
3.4	Processus global de Matching.....	33
4.1	Architecture du Système MidlJob	36
4.2	Interface de l'Application MidlJob	38
4.3	Interface de Saisie des Informations par les Utilisateurs.....	38
4.4	Extraction des Données et Création des Fichiers CSV	39
4.5	Représentation des relations sémantiques avec des embeddings	44
4.6	Relations Sémantiques entre Intitulés de Poste avec le Modèle "paraphrase-MiniLM-L6-v2"	45
4.7	Schéma du processus de clustering K-Means.....	47
4.8	Illustration du processus de vectorisation TF-IDF.....	51
4.9	Illustration de la similarité cosinus entre deux vecteurs	53
4.10	Architecture d'une application Web.....	56
4.11	Benchmarks de performance de FastAPI.....	58

5.1	Détails de l'Offre Data Scientist.....	60
5.2	Meilleures Correspondances - Candidats pour Data Scientist	61
5.3	Profil du Candidat Simo pour Data Scientist.....	61
5.4	Profil du Candidat Aouatif pour Data Analyst	62
5.5	Profil du Candidat Jad pour Data Engineer	63
5.6	Profil du Candidat Nihar pour Data Analyst	63
5.7	Recherche d'Offres Correspondant au Profil Développeur Web.....	64
5.8	Meilleures Correspondances pour le Profil Développeur Web.....	65
5.9	Détails du Poste de Développeur Web.....	65
5.10	Détails du Poste de Développeur Mobile.....	66
5.11	Meilleures Correspondances - Candidats pour le Poste de Web Developer.....	68
5.12	Meilleures Correspondances - Offres d'Emploi pour Anna Thompson (Infirmière).....	72
5.13	Présentation du Profil du Candidat Data Scientist.....	75
5.14	Liste des Profils Likés par le Recruteur dans l'Application MidlJob.....	75
5.15	Notification de Correspondance Réussie dans la Plateforme MidlJob.....	76
5.16	Interface de Messagerie pour le Matching des Candidats	77

Liste des Abréviations

AI	Intelligence Artificielle
NLP	Natural Language Processing (Traitement du Langage Naturel)
ML	Machine Learning (Apprentissage Automatique)
TF-IDF	Term Frequency-Inverse Document Frequency
K-Means	Algorithme de Clustering K-Moyennes
BERT	User Identifier (Identifiant Utilisateur)
GPT	Structured Query Language
UID	Curriculum Vitae
SQL	Contrat à Durée Indéterminée
CV	Contrat à Durée Déterminée
API	Application Programming Interface
ASGI	HyperText Markup Language
ATS	Cascading Style Sheets
MLP	User Experience
KPI	Key Performance Indicator (Indicateur Clé de Performance)

Chapitre 1

Présentation de l'organisme d'accueil

1.1 Introduction

Ce premier chapitre a pour objectif de fournir un cadre de référence essentiel pour comprendre l'environnement dans lequel j'ai effectué mon stage et les missions qui m'ont été confiées. En commençant par la présentation de l'entreprise Midl, son contexte économique, et son organisation interne, ce chapitre met en lumière les fondations sur lesquelles repose le projet MidlJob, une application innovante dédiée au recrutement. Il explore également le secteur économique dans lequel évolue Midl, ainsi que le cadre spécifique de mon stage, en décrivant les tâches et responsabilités qui m'ont été assignées au sein de cette start-up dynamique.

1.2 Présentation de l'entreprise

L'entreprise Midl, fondée en 2023 par M. Vincent Azouani, est une start-up innovante située dans l'incubateur Station F, dans le 13-ème arrondissement de Paris. Station F est un lieu dynamique qui offre un environnement propice à la croissance des jeunes entreprises, avec des espaces de travail partagés ou privés, ainsi que des conseils et un réseau précieux.



FIGURE 1.1 – L’incubateur Station F

Midl se spécialise dans le recrutement digital, avec l’ambition de moderniser et de simplifier le processus de recrutement en intégrant des technologies de pointe. L’entreprise cherche à relever les défis rencontrés par les candidats et les entreprises en réduisant les délais de réponse, en optimisant le temps consacré à la recherche d’emploi et au tri des candidatures, tout en améliorant l’expérience globale des utilisateurs. Leur objectif est de révolutionner le marché du recrutement en éliminant les obstacles traditionnels et en favorisant des interactions plus fluides et efficaces. L’entreprise développe MidlJob, une application web et mobile inspirée des applications de rencontre, visant à faciliter la connexion entre candidats et employeurs. En combinant des critères techniques et affinitaires, MidlJob aspire à créer des recrutements plus durables et pertinents, répondant ainsi aux besoins croissants des entreprises modernes.



FIGURE 1.2 – Logo Midl

La mission de Midl est de révolutionner le secteur du recrutement en offrant des solutions technologiques qui maximisent l'efficacité des processus d'embauche tout en minimisant les biais humains. La vision de l'entreprise est de devenir un leader mondial dans le domaine des technologies de recrutement, en aidant les entreprises à trouver les talents qui correspondent le mieux à leurs besoins culturels et techniques.

1.3 Fiche d'identité

Le tableau ci-dessous présente les informations essentielles concernant l'entreprise Midl :

Dénomination Sociale	Midl
Date de Création	Mai 2023
Dirigeant	M. Vincent Azouani
Coordonnées	Villes : Paris Adresse : 8 rue d'Artois 93000 Bobigny, France Tél. : +33 978 743 870 E-mail :: contact@midljob.com

FIGURE 1.3 – La fiche d'identité de Midl

1.4 Contexte RH

Au sein de Midl, le service où j'ai travaillé est constitué d'une équipe multidisciplinaire chargée de divers aspects du développement et de la gestion de l'application de job dating. L'équipe comprend des développeurs spécialisés dans Flutter pour la création de l'application mobile, des experts en intelligence artificielle (IA) pour l'élaboration des algorithmes de matching, ainsi que des spécialistes en gestion et maintenance des bases de données. La répartition du travail est bien définie : les développeurs se concentrent sur la programmation et l'intégration des fonctionnalités, les experts en IA travaillent sur les algorithmes et les modèles de recommandation, tandis que l'équipe de gestion des bases de données assure la structure, la sécurité et la performance des données.

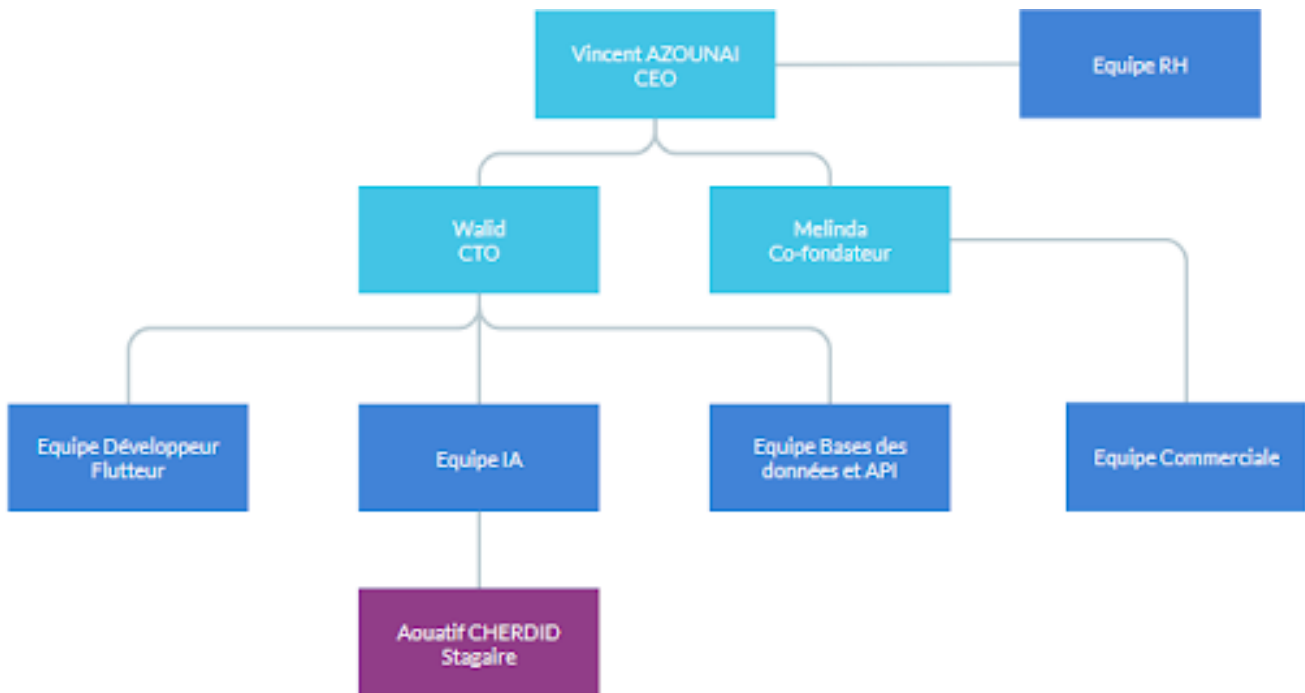


FIGURE 1.4 – Organigramme de l'entreprise Midl

Le manager/CTO, Walid, supervisait l'ensemble du projet avec mon maître de stage, Vincent Azouani, en veillant à ce que les différentes équipes collaborent efficacement pour atteindre les objectifs fixés. Pendant mon stage, j'ai également été en contact étroit avec mon maître de stage, qui m'a guidé tout au long de mes missions et assuré un suivi régulier de mon travail. Les contraintes particulières du service incluaient l'utilisation de technologies cloud comme Firebase pour la gestion

des données, ainsi que l'adhésion aux pratiques de développement Scrum avec le logiciel Jira, garantissant ainsi une flexibilité et une réactivité maximales dans le développement de l'application.

1.5 L'environnement économique du stage

1.5.1 Le Secteur Économique

Midl évolue dans le secteur des technologies de recrutement, un sous-secteur dynamique de l'industrie des technologies de l'information et de la communication (TIC). Ce secteur est en pleine croissance, porté par une demande accrue pour des solutions numériques capables d'optimiser les processus d'embauche. Les entreprises modernes cherchent à rationaliser leurs processus de recrutement pour gagner en efficacité, réduire les coûts, et attirer les meilleurs talents. Les plateformes de matching, les systèmes de gestion des candidatures (ATS), et les outils d'analyse prédictive sont désormais des éléments incontournables pour les ressources humaines.

En France, ainsi que dans le reste de l'Europe, le marché du recrutement numérique connaît une expansion rapide, stimulée par la transition vers des modèles de travail plus flexibles, l'augmentation du télétravail, et le besoin croissant de trouver rapidement des candidats qualifiés dans un marché du travail hautement compétitif.

1.6 Conclusion

En conclusion, ce chapitre a permis de contextualiser mon stage chez Midl en présentant l'entreprise, son secteur d'activité, et la structure organisationnelle qui m'entourait. Midl, en tant que start-up innovante, s'efforce de révolutionner le secteur du recrutement grâce à des solutions technologiques avancées. Ce contexte a grandement influencé les missions qui m'ont été confiées. Les tâches que j'ai réalisées, notamment dans le développement et l'optimisation de MidlJob, s'inscrivent dans cette dynamique de transformation. Ces éléments fournissent une base solide pour comprendre les aspects techniques et méthodologiques qui seront détaillés dans les chapitres suivants.

Chapitre 2

Contexte du Projet

2.1 Introduction

Ce deuxième chapitre est consacré à l’exploration de la problématique à laquelle le projet MidlJob cherche à répondre, ainsi qu’au contexte général qui a motivé le développement de ce projet. Nous aborderons les limitations des méthodes traditionnelles de recrutement, l’intégration de l’intelligence artificielle comme solution innovante, et les objectifs spécifiques du projet MidlJob.

2.2 Contexte du Projet

Le projet MidlJob s’inscrit dans la dynamique actuelle de transformation numérique du marché du recrutement en proposant une application innovante qui exploite l’intelligence artificielle (IA) pour révolutionner le processus de recrutement. Cette application est conçue pour identifier les meilleures correspondances entre les profils des candidats et les exigences des postes, en tenant compte des compétences techniques, des expériences et des affinités culturelles.

La version gratuite de l’application utilise des techniques avancées telles que la vectorisation des compétences via TF-IDF, le calcul de similarité cosinus et le clustering des titres de postes, visant à améliorer la précision du matching et à affiner la sélection des candidats. Pour enrichir l’expérience utilisateur, la version Premium introduit une comparaison des personnalités entre candidats et recruteurs, offrant un matching plus complet et pertinent. Une version Pro est également en développement, intégrant des fonctionnalités avancées pour répondre aux exigences les plus élevées des recruteurs, promettant ainsi un niveau de précision et d’efficacité inégalé.

MidlJob se positionne ainsi comme une solution stratégique pour les entreprises, cherchant à optimiser leurs processus de recrutement tout en améliorant la qualité des correspondances entre candidats et postes.

2.3 Problématique

Ce projet aborde une problématique majeure liée aux défis posés par les méthodes traditionnelles de recrutement, qui impactent à la fois les recruteurs et les candidats. Du côté des recruteurs, les processus manuels, souvent inefficaces, sont chronophages et sujets à des erreurs humaines, ce qui peut entraîner des décisions de recrutement sous-optimales et prolonger les délais pour pourvoir les postes vacants. De plus, les recruteurs disposent souvent d'une visibilité limitée sur les compétences réelles et l'expérience des candidats, ce qui complique la sélection des profils les mieux adaptés aux besoins spécifiques de l'entreprise.

Du côté des candidats, l'absence d'outils modernes pour faciliter leur mise en relation avec les recruteurs conduit à un processus de recrutement inefficace, où leurs compétences et expériences ne sont pas toujours valorisées. Cela peut les amener à être injustement écartés des postes pour lesquels ils sont qualifiés ou à être sélectionnés pour des emplois qui ne correspondent pas réellement à leurs aspirations ou compétences. Cette inefficacité globale dans la correspondance entre les candidats et les postes disponibles limite la précision et l'efficacité des recrutements, conduisant souvent à des résultats insatisfaisants pour les deux parties.

2.4 Objectif du Projet

L'objectif principal de ce projet est de fournir une solution complète et opérationnelle pour transformer le processus de recrutement en améliorant significativement la précision et l'efficacité du matching entre les candidats et les postes. En automatisant les étapes clés du recrutement, MidlJob vise à réduire les délais d'embauche, à minimiser les biais humains, et à offrir une meilleure visibilité sur les compétences et les expériences des candidats. Ce projet ambitionne de répondre aux besoins contemporains des entreprises en leur offrant un outil performant, capable de faciliter la sélection des talents les plus adaptés, tout en optimisant l'utilisation des ressources et en augmentant la qualité des recrutements.

2.5 Missions du projet

Les missions du projet MidlJob sont multiples et visent à transformer le processus de recrutement grâce à l'intelligence artificielle et aux technologies avancées de traitement des données. Les principales missions de ce projet sont les suivantes :

- **Automatisation du Matching Candidat-Offre :** Développer un système capable d'automatiser le processus de correspondance entre les candidats et les offres d'emploi, en se basant sur des critères précis tels que les compétences, les expériences et les aspirations des candidats.
- **Amélioration de la Précision des Recommandations :** Utiliser des algorithmes de machine learning et de traitement du langage naturel (NLP) pour affiner la précision des recommandations, en tenant compte des nuances sémantiques et contextuelles des profils et des descriptions de poste.
- **Optimisation du Temps de Recrutement :** Concevoir un système permettant de réduire les délais de recrutement en automatisant les étapes fastidieuses de présélection et en fournissant rapidement les meilleures correspondances aux recruteurs.

2.6 Risques du projet

Comme tout projet innovant, MidlJob présente plusieurs risques qui doivent être identifiés et gérés de manière proactive :

- **Risque Technique :** La complexité des technologies utilisées (IA, machine learning, NLP) peut entraîner des difficultés dans le développement et l'intégration des différentes composantes du système. Une gestion rigoureuse des phases de développement et des tests est essentielle pour atténuer ce risque.
- **Risque de Conformité :** L'utilisation de données personnelles dans le cadre du recrutement impose de respecter les réglementations sur la protection des données (comme le RGPD). Le non-respect de ces réglementations pourrait entraîner des sanctions et nuire à la réputation du projet.

2.7 Conduite du projet Avec Scrum

Inspirée du privé et de la gestion des projets informatiques, la méthode SCRUM est devenue de nos jours de plus en plus adoptée dans les équipes de développement. Cette méthode "agile" permet la réalisation de projets complexes en favorisant l'interaction avec les membres de l'équipe et les managers, la collaboration du client et la réactivité face aux changements. En tant qu'équipe de 6 membres, nous avons choisi d'adopter la méthodologie SCRUM. Actuellement, SCRUM est l'approche agile la plus largement adoptée par les équipes de développement, car elle embrasse les valeurs fondamentales du célèbre Manifeste Agile. Ces valeurs comprennent la collaboration étroite avec le client, l'acceptation du changement, l'interaction humaine dynamique et la focalisation sur la livraison de logiciels fonctionnels.

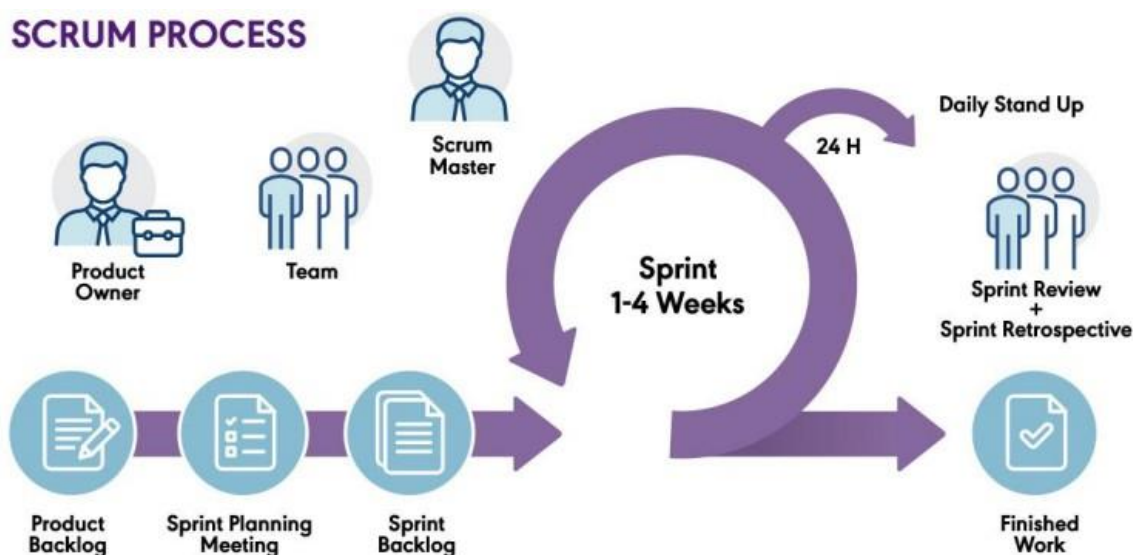


FIGURE 2.1 – Architecture du SCRUM

2.8 Les points positifs de Scrum

Scrum se distingue des autres méthodologies de développement agile à bien des égards. C'est actuellement le framework le plus utilisé et le plus fiable dans l'industrie du logiciel. Parmi les avantages bien connus de Scrum, on peut citer :

- **Adaptabilité facile :** Les processus Scrum sont itératifs et organisés en cycles de travail spécifiques, ce qui permet à l'équipe de se concentrer sur des fonctionnalités définies pour chaque période. Cela se traduit par des livrables mieux alignés sur les besoins des utilisateurs et permet une évolution ordonnée, transparente et fluide des modules en termes de fonctionnalités, de conception, de portée, et de caractéristiques.
- **Satisfaction des exigences :** Le client exprime ses attentes en évaluant la valeur apportée par chaque exigence du projet. L'équipe évalue ensuite ces attentes, et le Product Owner en établit les priorités en conséquence. Des vérifications régulières lors des démonstrations de sprint permettent de s'assurer que les exigences sont respectées, avec des retours fournis à l'équipe pour ajustement.
- **Capacité d'adaptation aux changements :** Scrum est conçu pour s'ajuster aux évolutions des exigences dictées par les besoins des clients ou les évolutions du marché, particulièrement fréquentes dans le cadre de projets complexes.
- **Minimisation des risques :** Grâce à sa capacité à anticiper et à réduire efficacement les risques, Scrum permet de commencer par la mise en œuvre des fonctionnalités les plus critiques, tout en maintenant une visibilité sur la progression de l'équipe dans le projet.
- **Amélioration de la qualité :** Le besoin d'obtenir une version fonctionnelle après chaque itération favorise la création de logiciels de qualité supérieure, en maintenant un standard élevé tout au long du développement.

La répartition des rôles au sein de SCRUM s'organise autour de trois entités principales : le Product Owner, l'équipe de développement et le Scrum Master. Ces équipes sont conçues pour être auto-organisées et transversales, leur permettant de choisir la meilleure approche pour accomplir leur travail sans dépendre de directives externes. Les sprints, également connus sous le nom d'itérations, sont des cycles de travail courts, généralement d'une durée de 1 à 4 semaines, utilisés pour créer des incréments de produits logiciels ou de services livrables. Au cours de notre stage chez Midl, la durée d'un sprint était d'une semaine.

Le Product Backlog joue un rôle central dans la gestion de toutes les exigences connues d'un projet Scrum. Il regroupe l'ensemble des exigences du client ainsi que les résultats de travail nécessaires pour mener à bien le projet. Les exigences, qu'elles soient fonctionnelles ou non fonctionnelles, ainsi que d'autres caractéristiques liées à l'expérience utilisateur et à la conception de l'interface utilisateur, sont toutes considérées comme des exigences. Les réunions de sprint, quant à elles, permettent de discuter et de planifier les activités du sprint à venir.

- **Sprint Planning :** L'objectif de la réunion de planification du sprint est de déterminer les activités à accomplir pendant le sprint en cours et la manière dont elles seront réalisées. Cette réunion a lieu au début de chaque sprint et détermine l'approche à adopter en fonction des étapes et des délais définis dans le Product Backlog.
- **Daily Scrum :** Chaque matin, l'équipe de développement se réunit pendant une période de 5 à 15 minutes pour évaluer la progression vers les objectifs du sprint. Pendant cette réunion, l'équipe discute du travail accompli la veille, des problèmes et des obstacles rencontrés, et demande de l'aide en cas de besoin.
- **Sprint Review :** Cette réunion se tient généralement à la fin du sprint et permet à l'équipe Scrum de présenter l'incrément de produit terminé aux parties prenantes. En plus de l'inspection des fonctionnalités produites pendant le sprint, le Product Owner fournit des commentaires et des observations constructives qui peuvent être intégrés dans le Product Backlog pour orienter le travail des futurs sprints.

2.9 Diagramme de GANTT

Le diagramme de Gantt est un instrument couramment employé en planification et en gestion de projets. Il offre une représentation temporelle des différentes tâches constituant un projet. Il se présente sous la forme d'un graphe connexe, valué et orienté, qui permet de visualiser de manière graphique la progression du projet. Ci-dessous, vous trouverez notre diagramme de Gantt, qui illustre les diverses activités que nous avons menées tout au long de la période de stage.

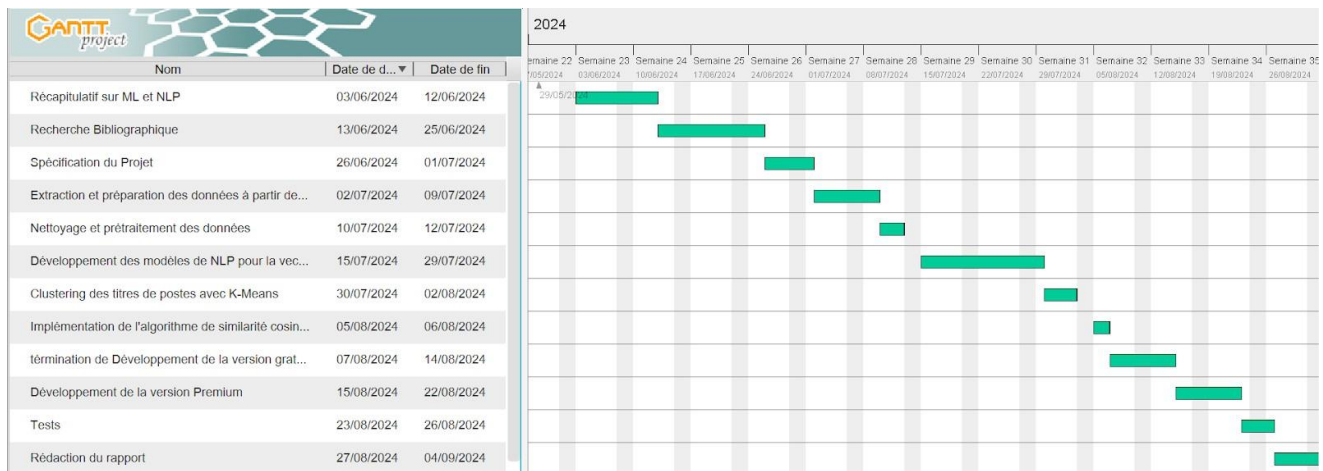


FIGURE 2.2 – Diagramme de Gantt

2.10 Conclusion

En conclusion de ce chapitre consacré à la problématique et au contexte de notre projet de réinvention du recrutement par l'intelligence artificielle, nous avons établi les bases nécessaires pour la compréhension et le développement de cette initiative innovante. Nous avons soigneusement défini le cadre du projet en présentant l'organisme d'accueil, en expliquant le contexte global, et en énonçant les objectifs spécifiques que nous visons à atteindre. De plus, une planification méthodique de la mise en œuvre de ce projet a été décrite en détail, garantissant une approche structurée et rigoureuse. Ce chapitre représente ainsi un fondement essentiel qui nous permettra de progresser avec assurance dans la réalisation de notre projet.

Chapitre 3

État de l'Art

3.1 Introduction

Le secteur du recrutement a subi d'importantes transformations au cours des dernières décennies, en grande partie sous l'impulsion des progrès technologiques. Les méthodes traditionnelles, souvent manuelles et basées sur des relations personnelles, sont devenues insuffisantes face aux exigences croissantes de rapidité, de précision et d'objectivité. L'intelligence artificielle (IA) s'est imposée comme une solution clé, non seulement pour automatiser les tâches répétitives, mais aussi pour améliorer la manière dont les talents sont identifiés et évalués. Ce chapitre se propose d'examiner l'évolution du recrutement et d'explorer comment les technologies telles que le machine learning et le traitement du langage naturel (NLP) redéfinissent ces pratiques pour répondre aux besoins contemporains des entreprises.

3.2 Historique du Recrutement

Au fil du temps, le recrutement a évolué pour s'adapter aux besoins changeants des entreprises et des candidats. Au début du XX siècle, il reposait principalement sur les relations personnelles, les recommandations et les affichages publics. Les entreprises recrutaient souvent au sein de leur cercle proche ou via des annonces dans les journaux locaux. Cette approche, bien qu'efficace à petite échelle, limitait l'accès à un vivier de talents diversifié et manquait de structure.

Avec l'industrialisation et la croissance économique après la Seconde Guerre mondiale, le besoin de main-d'œuvre qualifiée a augmenté. Les agences de recrutement sont apparues pour répondre

à cette demande, offrant des services spécialisés pour aider les entreprises à trouver les candidats adéquats. Elles utilisaient des méthodes plus structurées, comme les tests psychotechniques et les entretiens standardisés, pour évaluer les compétences des candidats de manière plus objective.

Les années 1990 ont marqué un tournant décisif avec l'arrivée d'Internet. Les sites d'emploi en ligne, tels que Monster et Indeed, ont révolutionné le recrutement en permettant aux entreprises de publier des offres d'emploi accessibles mondialement et aux candidats de postuler en quelques clics. Cette démocratisation du recrutement a élargi le vivier de candidats, mais a également conduit à une surcharge d'informations pour les recruteurs, rendant le tri des candidatures plus complexe et chronophage.

Malgré ces avancées, le recrutement traditionnel présente plusieurs limitations persistantes. Les processus restent souvent manuels et laborieux, impliquant le tri manuel des CV, l'organisation des entretiens et la vérification des références. Les décisions de recrutement peuvent être influencées par des biais cognitifs, conscients ou inconscients, affectant l'objectivité et l'équité du processus. De plus, il est souvent difficile d'évaluer précisément les compétences réelles et le potentiel des candidats, ce qui peut conduire à des correspondances sous-optimales entre les postes et les profils. Enfin, la gestion de volumes massifs de candidatures peut entraîner des délais prolongés, retardant le pourvoi des postes vacants et entraînant la perte de talents potentiels.

3.3 L'Émergence de l'Intelligence Artificielle dans le Recrutement

Face aux limites des méthodes classiques, l'intelligence artificielle s'est imposée comme une solution capable de transformer le recrutement. L'IA, en simulant des capacités cognitives humaines, offre des outils sophistiqués pour automatiser, optimiser et personnaliser plusieurs aspects du processus de recrutement. L'IA permet d'analyser de grandes quantités de données, d'identifier des schémas complexes et d'améliorer la précision du matching entre candidats et postes.

Avec l'explosion des données disponibles et l'avancement des technologies, l'IA est devenue plus accessible aux entreprises de toutes tailles. Les recruteurs disposent désormais d'outils sophistiqués pour analyser les profils des candidats, prédire leur performance future et optimiser le processus

de recrutement. Selon une étude de Deloitte (2020), plus de 50 % des entreprises utilisent déjà l'IA dans au moins une étape de leur processus de recrutement, et ce chiffre est en constante augmentation.

Année	Pourcentage d'adoption de l'IA
2015	10%
2016	15%
2017	22%
2018	30%
2019	40%
2020	55%
2021	65%
2022	75%
2023	85%

Tableau 3.1 : Croissance de l'adoption de l'IA dans le recrutement de 2015 à 2023 (Source : Deloitte, 2023)

3.4 Applications de l'Intelligence Artificielle dans le Recrutement

L'intégration de l'IA dans le recrutement a ouvert la voie à plusieurs applications innovantes qui optimisent les différentes étapes du processus, de l'analyse des candidatures à l'évaluation des compétences.

3.4.1 Le Machine Learning (Apprentissage Automatique)

Le machine learning est devenu une application fondamentale de l'IA dans le domaine du recrutement. Cette technologie utilise des algorithmes pour analyser de grands volumes de données, ce qui permet d'automatiser des tâches telles que le tri des CVs et la prédiction de la performance future des candidats. Voici les principaux outils et techniques de machine learning utilisés dans le recrutement :

- **Réseaux de Neurones** : Ces modèles sont capables de modéliser des relations complexes et non linéaires entre les données, ce qui améliore la précision des prédictions concernant les candidats. Ils sont particulièrement efficaces pour identifier des patterns dans des ensembles de données volumineux et variés.
- **Forêts Aléatoires** : Cette technique de classification utilise un ensemble d'arbres de décision pour classer les candidats. Elle réduit le risque de surapprentissage en combinant les prédictions de plusieurs arbres, ce qui permet d'améliorer la robustesse et la précision du modèle.
- **Machines à Vecteurs de Support (SVM)** : Les SVM sont utilisés pour classer les candidats en fonction des critères de sélection définis. Ils créent des hyperplans dans un espace multidimensionnel pour séparer les différentes classes de données, facilitant ainsi le tri des candidats en fonction de leurs compétences et qualifications.
- **K-Means** : Cet algorithme de clustering regroupe les candidats en fonction de caractéristiques communes. Il permet de segmenter les CV en clusters, facilitant l'identification de groupes de profils similaires. L'objectif est de mieux organiser et cibler les candidats lors du tri.
- **Similarité Cosinus** : Cet algorithme mesure la similitude entre deux vecteurs, par exemple entre un CV et une offre d'emploi. Il permet de comparer directement les compétences des candidats aux exigences des postes afin de prioriser ceux qui correspondent le mieux.

3.4.2 Le Traitement du Langage Naturel (NLP)

Le traitement du langage naturel (NLP) est une autre application clé de l'IA dans le recrutement. Il permet aux machines de comprendre et d'interpréter le langage humain dans des documents tels que les CVs et les lettres de motivation. Les principales technologies de NLP utilisées incluent :

- **TF-IDF** : Cette méthode de pondération est utilisée pour évaluer l'importance des mots dans les documents, facilitant ainsi le filtrage des CVs. Elle permet d'identifier les termes les plus significatifs pour un poste donné en comparant leur fréquence dans un document avec leur fréquence dans l'ensemble des documents.

- **Word2Vec** : Ce modèle de langage capture les relations sémantiques entre les mots en les représentant sous forme de vecteurs. Il est utilisé pour analyser les descriptions de poste et les CVs en capturant les similarités entre les termes et les concepts.
- **BERT** : permet une compréhension fine des nuances linguistiques en analysant les mots dans leur contexte bidirectionnel. Il est particulièrement efficace pour analyser les lettres de motivation et autres documents textuels en offrant une compréhension plus approfondie du langage.
- **GPT** : est utilisé pour la génération de texte basée sur des contextes prédéfinis. Il peut être employé pour rédiger automatiquement des réponses aux candidats ou pour générer du contenu textuel de manière fluide, bien qu'il puisse parfois produire des résultats imprévisibles.
- **paraphrase-MiniLM-L6-v2** : Modèle d'embedding pour obtenir des représentations vectorielles des phrases, utilisé pour améliorer la correspondance entre les descriptions de poste et les CVs.

Technologie	Avantages	Inconvénients	Exemple d'Utilisation
TF-IDF	Simple, rapide à implémenter	Ne capture pas le contexte	Filtrage initial des CVs
Word2Vec	Capture les relations sémantiques	Nécessite de grandes données pour l'entraînement	Analyse des descriptions de poste
BERT	Compréhension fine du contexte	Exigeant en ressources	Analyse des lettres de motivation
GPT	Génération de texte naturelle	Peut produire des résultats imprévisibles	Rédaction automatique de réponses aux candidats

Technologie	Avantages	Inconvénients	Exemple d'Utilisation
paraphrase- MiniLM-L6-v2	Représentations vectorielles de phrases efficaces	Peut nécessiter un ajustement pour des tâches spécifiques	Amélioration de la correspondance entre postes et CVs

Tableau 3.2 : Comparaison des technologies de traitement du langage naturel (NLP) dans le recrutement

3.4.3 Les Systèmes de Recommandation

Les systèmes de recommandation utilisent des algorithmes pour proposer des suggestions personnalisées aux recruteurs et aux candidats. Pour les recruteurs, ces systèmes permettent de découvrir des candidats pertinents en fonction des besoins du poste, même si ces derniers n'ont pas postulé. Pour les candidats, ils facilitent la recherche d'offres adaptées à leurs compétences et aspirations. Ces algorithmes analysent les comportements passés des utilisateurs pour fournir des recommandations précises, améliorant ainsi l'efficacité du matching.

3.4.4 L'Analyse Prédictive

L'analyse prédictive utilise des algorithmes pour anticiper des comportements futurs à partir de données passées. Dans le recrutement, elle permet aux entreprises de prévoir leurs besoins en personnel en fonction des tendances du marché et de la croissance de l'entreprise. Elle peut aussi identifier les risques de turnover, en évaluant la probabilité qu'un employé quitte l'entreprise.

3.4.5 L'Évaluation Automatisée des Compétences

L'IA permet de développer des outils d'évaluation automatisée pour tester les compétences techniques et comportementales des candidats. Ces outils incluent des tests en ligne adaptatifs et des simulations, qui offrent une évaluation précise du niveau de compétence des candidats. Les entretiens vidéo analysés par l'IA permettent aussi d'évaluer des aspects plus subtils, tels que les expressions faciales et le langage corporel.

3.5 Solutions Existantes

L'adoption de l'IA dans le recrutement a favorisé l'émergence de plusieurs solutions technologiques sur le marché.

3.5.1 Présentation de quelques outils innovants

LinkedIn Talent Solutions

LinkedIn Talent Solutions est une suite d'outils proposée par LinkedIn pour les professionnels du recrutement. Cette plateforme exploite l'intelligence artificielle pour optimiser le processus de recrutement et offre plusieurs fonctionnalités innovantes. Elle propose des recommandations de candidats basées sur les critères spécifiques du poste, aidant les recruteurs à identifier rapidement les profils les plus pertinents parmi plus de 900 millions de membres sur LinkedIn. De plus, elle fournit des insights approfondis sur le marché du travail, offrant des données et des analyses sur les tendances actuelles, les compétences en demande et la disponibilité des talents dans différents secteurs et régions. Ces informations permettent aux entreprises d'ajuster leurs stratégies de recrutement et de rester compétitives. Enfin, LinkedIn Talent Solutions permet de créer des publicités ciblées pour atteindre efficacement les candidats les plus adaptés. En diffusant les offres d'emploi auprès d'une audience qualifiée et en utilisant des critères de ciblage précis tels que les compétences, l'expérience ou la localisation, les recruteurs peuvent maximiser leur portée et attirer les meilleurs talents.

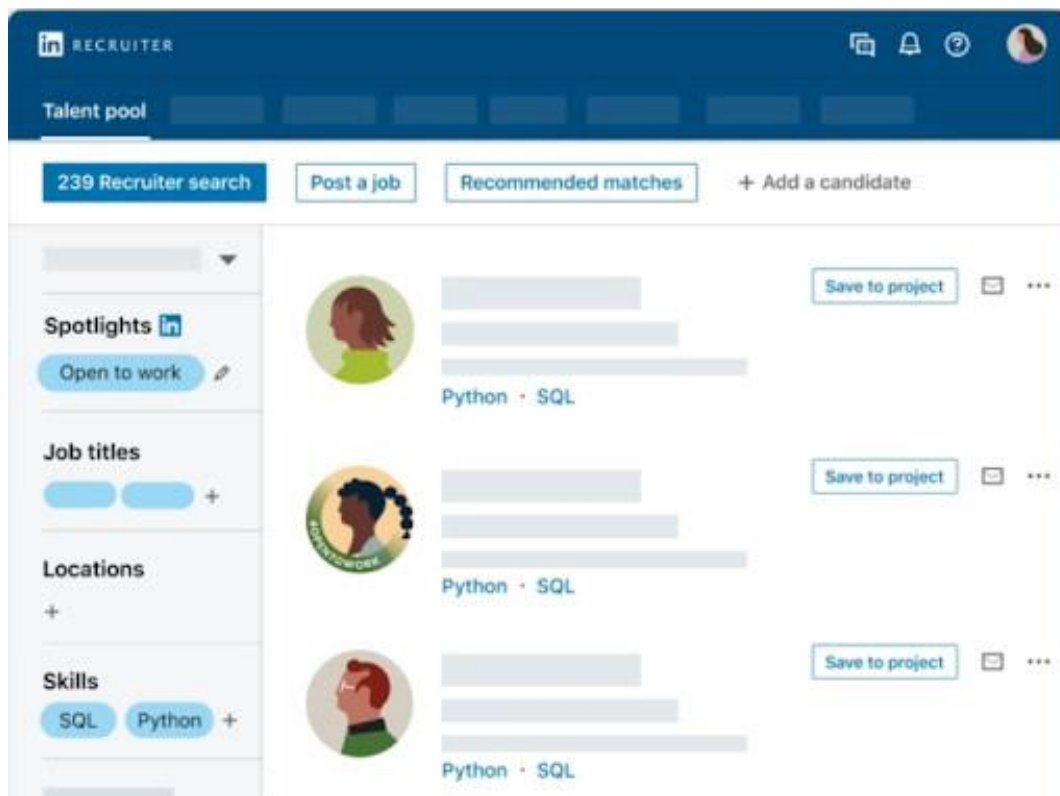


FIGURE 3.1 – Interface LinkedIn Talent Solutions

HireVue

HireVue est une plateforme d'entretien vidéo et d'évaluation basée sur l'intelligence artificielle, conçue pour aider les entreprises à optimiser leur processus de recrutement. Elle propose des entretiens vidéo automatisés, permettant aux candidats de répondre aux questions à leur convenance, ce qui accélère la phase de présélection. Grâce à l'analyse vocale, HireVue évalue les réponses des candidats en se concentrant sur le contenu et les compétences verbales, aidant ainsi à identifier les traits de personnalité et les compétences clés pertinentes pour le poste.

En outre, HireVue offre des tests d'évaluation des compétences techniques et cognitives, permettant aux recruteurs de mesurer de manière objective les aptitudes des candidats dans des domaines spécifiques. Ces évaluations peuvent inclure des tests de codage, des résolutions de problèmes ou des simulations de scénarios professionnels.

L'automatisation apportée par HireVue réduit considérablement la charge administrative liée au recrutement. En éliminant la nécessité de programmer des entretiens en personne et en fournissant des analyses approfondies des candidats, la plateforme permet aux recruteurs de gagner du temps

et de se concentrer sur les candidats les plus prometteurs. Cette efficacité accrue conduit à des cycles de recrutement plus rapides et à une meilleure expérience pour les candidats.



FIGURE 3.2 – l’interface HireVue utilise d’entretiens vidéo basés sur l’IA.

Textio

Textio est un outil d’écriture assistée par l’intelligence artificielle qui aide les recruteurs à rédiger des offres d’emploi plus efficaces, inclusives et attrayantes. En utilisant le traitement du langage naturel, Textio analyse le langage des annonces pour détecter les biais potentiels, tels que les termes genrés ou le jargon technique excessif, qui pourraient décourager certains candidats de postuler. L’outil propose des suggestions en temps réel pour optimiser le ton et le style de l’annonce, rendant le langage plus engageant et accessible à un public diversifié. De plus, Textio fournit des prédictions sur les performances des offres d’emploi en termes d’engagement des candidats, aidant les recruteurs à ajuster leur rédaction pour maximiser l’impact et attirer un vivier de talents plus large et diversifié.

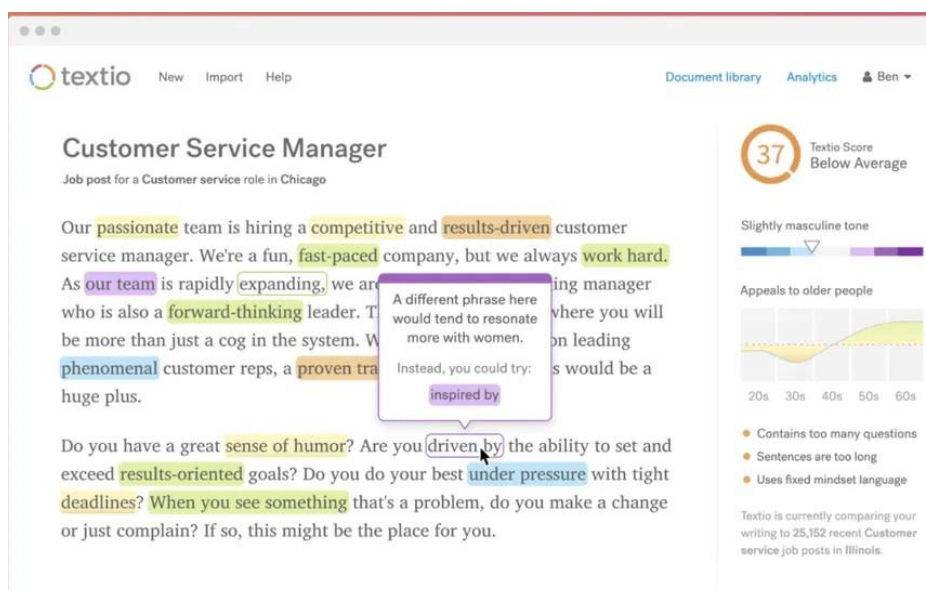


FIGURE 3.3 – l’interface de Textio montrant l’analyse d’une offre d’emploi.

Ideal

Ideal est un assistant virtuel de recrutement alimenté par l’intelligence artificielle, conçu pour optimiser le processus de recrutement des entreprises. Il propose un matching intelligent entre les candidats et les postes en analysant les compétences, les expériences et les qualifications pour identifier les profils les plus adaptés. Ideal améliore l’engagement des candidats en envoyant des messages personnalisés tout au long du processus, renforçant leur intérêt pour le poste et l’entreprise. De plus, il fournit une analyse approfondie des données sur des métriques clés comme le temps de recrutement et les taux de conversion, aidant les recruteurs à prendre des décisions éclairées et à améliorer continuellement leur processus de recrutement.

3.5.2 Analyse des forces et faiblesses des solutions existantes

Forces :

- **Efficacité accrue** : Automatisation des tâches répétitives, réduisant le temps de recrutement.
- **Amélioration de la qualité des recrutements** : Meilleure correspondance entre les candidats et les postes.

- **Expérience candidat améliorée** : Communication rapide et personnalisée.
- **Insights basés sur les données** : Aide à la prise de décision éclairée.

Faiblesses :

- **Biais algorithmiques** : Risque de reproduire ou d'amplifier les biais présents dans les données d'entraînement.
- **Manque de transparence** : Difficulté à expliquer les décisions prises par l'IA.
- **Dépendance à la technologie** : Nécessité de combiner l'IA avec le jugement humain.
- **Problèmes de confidentialité** : Collecte et traitement des données personnelles sensibles.

3.6 Les solutions proposées par le projet MidlJob

Les solutions de MidlJob sont organisées en trois versions principales, chacune adaptée à différents besoins en matière de recrutement. Ces versions permettent un processus de matching optimisé entre les recruteurs et les candidats, conçu pour évoluer avec les exigences du marché du travail.

1. Version Free :

Cette version gratuite est idéale pour les utilisateurs ayant des besoins de base. Elle propose un matching simple en se basant sur des attributs essentiels, tels que :

- Nom du poste
- Compétences techniques
- Expérience
- Type de contrat
- Langues maîtrisées

L'algorithme de matching utilise une approche basée sur la similarité cosinus pour comparer les compétences des candidats aux exigences des offres d'emploi.

Le schéma ci-dessous illustre le processus de matching global entre candidats et recruteurs dans MidlJob, applicable dès la version Free :

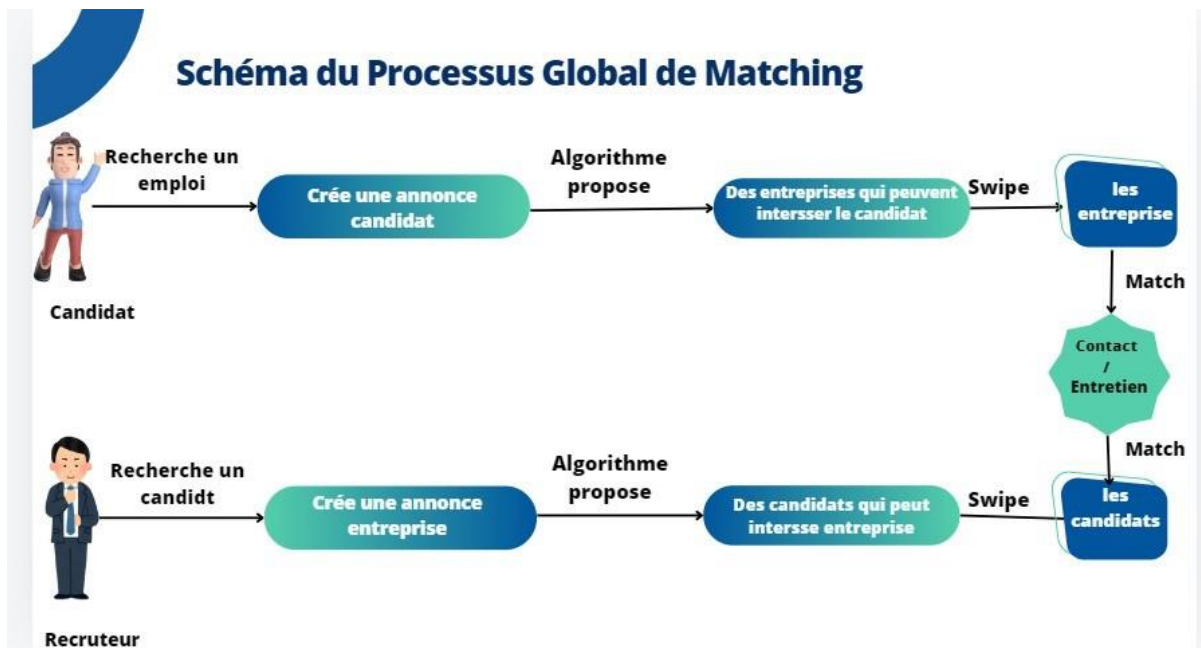


FIGURE 3.4 – Processus global de Matching

Dans ce schéma :

- Les candidats créent une annonce et accèdent à des entreprises correspondant à leurs critères.
- Les recruteurs créent une annonce d'emploi et accèdent à une liste de candidats pertinents.
- Le swipe permet aux deux parties de montrer leur intérêt.
- Un match se produit lorsque les deux parties expriment un intérêt mutuel, et elles sont mises en contact direct.

2. Version Premium :

Cette version avancée ajoute des fonctionnalités supplémentaires pour un matching plus précis et personnalisé. En plus des critères techniques de la version Free, elle introduit des éléments tels que :

- Soft skills (compétences non techniques)
- Compatibilité des personnalités entre candidats et recruteurs

Cette version prend en compte non seulement les compétences professionnelles, mais aussi les traits comportementaux et culturels pour un recrutement adapté à long terme. Le processus reste similaire au swipe, mais le matching est enrichi avec des données supplémentaires pour optimiser la compatibilité.

3. **Version Pro (en cours de développement) :**

La version Pro est en cours de développement et vise à offrir une gestion complète du processus de recrutement. Elle comprendra des fonctionnalités avancées telles que :

- Génération automatique de questions d’entretien
- Analyse des réponses des candidats pendant les entretiens.

Cette version s’appuie sur des modèles sophistiqués de traitement du langage naturel (NLP) pour améliorer la précision des résultats tout en garantissant la sécurité des données personnelles.

3.7 Conclusion

L’intelligence artificielle a introduit une nouvelle ère dans le domaine du recrutement, en automatisant les tâches courantes tout en améliorant l’efficacité et la précision des processus. Le machine learning, le NLP et les systèmes de recommandation ont permis aux entreprises de mieux faire correspondre les compétences des candidats aux postes disponibles, tout en réduisant les biais humains. Cependant, ces innovations soulèvent également des défis, notamment en matière de transparence, de biais algorithmiques et de protection des données personnelles. Pour garantir une adoption réussie, il est crucial de mettre en place des mesures éthiques et réglementaires solides. Le projet MidlJob illustre cette approche en combinant innovation technologique et responsabilité sociale, tout en offrant une solution de recrutement capable de répondre aux besoins actuels tout en anticipant les défis futurs.

Chapitre 4

Méthodologie

4.1 Introduction

Ce chapitre explore la méthodologie adoptée pour optimiser le processus de sélection des meilleures correspondances entre candidats et offres d'emploi dans le cadre du développement de MidlJob. En triant les correspondances selon un score de similarité, nous assurons un matching précis et pertinent, facilitant la décision des recruteurs et candidats. De plus, la version Premium de MidlJob offre un enrichissement supplémentaire en intégrant l'évaluation des traits de personnalité pour améliorer la compatibilité à long terme.

4.2 Structure de MidlJob

4.2.1 Architecture du Système

L'architecture de MidlJob a été conçue pour assurer un flux de données fluide et un traitement efficace. Elle est composée de plusieurs modules interconnectés, chacun étant responsable d'un aspect spécifique du traitement des données, depuis l'extraction initiale jusqu'à la correspondance (matching) entre les candidats et les offres d'emploi.

La Figure 4.1 illustre l'architecture complète du système. Les données sont d'abord extraites de Firebase, puis passent par des phases de prétraitement à l'aide d'outils tels que Pandas et NLTK. Par la suite, des embeddings sémantiques sont générés via le modèle paraphrase-MiniLM- L6-v2 pour les intitulés de postes. Le clustering de ces intitulés est réalisé à l'aide de l'algorithme

K-Means, tandis que pour les autres caractéristiques textuelles, telles que les compétences ou les descriptions de poste etc . . . , la méthode TF-IDF est appliquée.

Enfin, la similarité cosinus est utilisée pour effectuer un matching précis entre les candidats et les offres d'emploi. Cette architecture montre comment les différents modules interagissent de manière cohérente pour fournir des correspondances optimisées, répondant ainsi aux besoins variés des entreprises modernes.

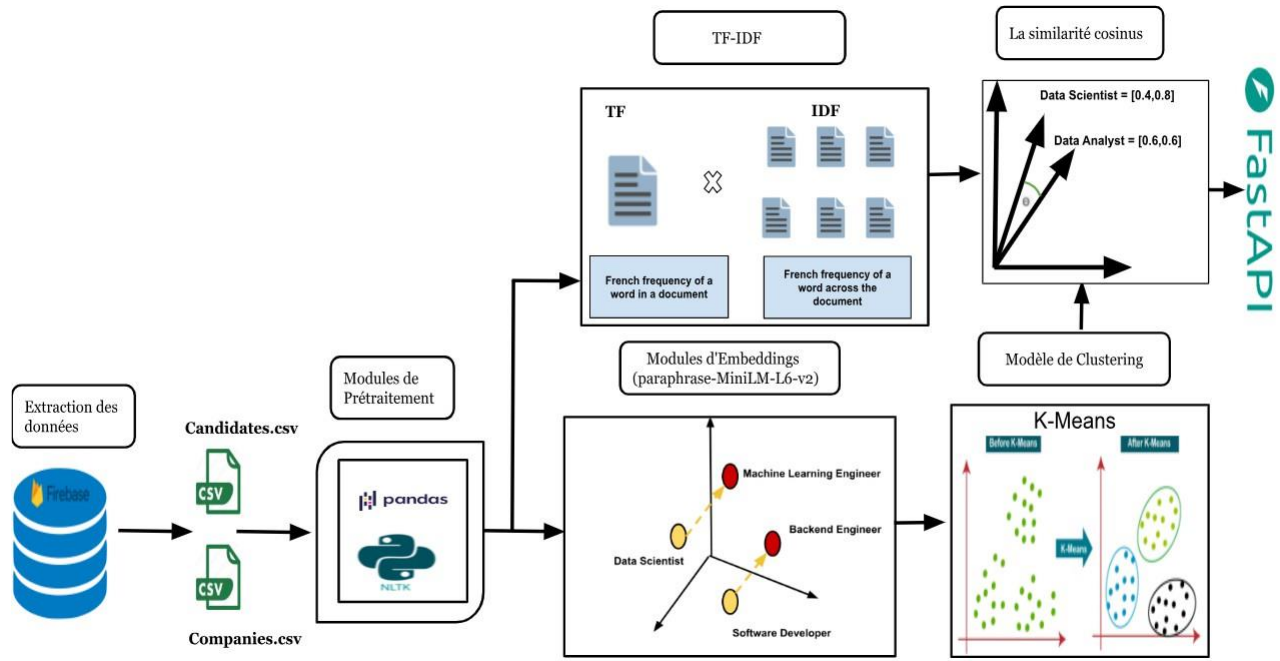


FIGURE 4.1 – Architecture du Système MidlJob

4.2.2 Version Gratuite

La version gratuite de MidlJob se concentre sur le matching basé sur des informations essentielles telles que :

- Le nom de poste recherché par le candidat
- Le nom de poste pour une offre d'emploi
- La description du poste

- La description du profil
- Les compétences techniques
- L'expérience
- Le type de contrat
- Les langues parlées
- Le niveau d'études, etc.

Entrées : Attributs de base des candidats et des offres d'emploi.

Sorties : Liste de correspondances optimales basées sur la similarité cosinus.

4.2.2.1 Collecte et préparation des Données

La collecte et la préparation des données sont des étapes cruciales pour garantir un matching de qualité. Les informations fournies par les utilisateurs (candidats et recruteurs) via l'application MidlJob sont stockées dans une base de données Firebase, un système NoSQL en temps réel. Ce choix permet une gestion flexible et instantanée des données, assurant qu'elles sont toujours à jour pour le processus de matching. Une fois collectées, les données sont exportées sous forme de fichiers CSV, ce qui facilite leur manipulation et leur analyse pour les étapes suivantes du processus. La Figure 4.2 montre l'interface d'accueil de l'application MidlJob, qui permet aux candidats et recruteurs de naviguer entre différentes options, en facilitant l'enregistrement de leurs données.



FIGURE 4.2 – Interface de l'Application MidlJob

L'interface de saisie des informations par les utilisateurs, illustrée dans la Figure 4.3, permet aux candidats de renseigner des informations essentielles telles que le niveau d'études, les compétences clés, les attentes salariales, ainsi que d'autres informations. Ces données sont ensuite utilisées dans le processus de matching avec les offres d'emploi.

Niveau d'études* :

☐ Autodidacte
 ☐ Diplôme national du brevet
 ☐ CAP
 ☐ BEP
 ☐ Bac
 ☐ DUT
 ☐ BTS
 ☐ Licence
 ☐ Master
 ☐ Doctorat

Salaire brut annuel souhaité* :

☐ Jusqu'à 20 000 €
 ☐ De 20.000 à 30.000 €
 ☐ De 30.000 à 40.000 €
 ☐ De 40.000 à 55.000 €
 ☐ De 55.000 à 70.000 €
 ☐ + de 70.000 €
 ☐ Saisir le salaire annuel

Compétences clés :

Avantages :

FIGURE 4.3 – Interface de Saisie des Informations par les Utilisateurs

4.2.2.2 Structure et Source des Données

Les informations fournies par les utilisateurs (candidats et recruteurs) sont extraites et transformées en attributs spécifiques dans des fichiers CSV, comme illustré dans la Figure 4.4. Ce processus permet une organisation claire et structurée des données pour le traitement ultérieur. Les fichiers `companies_job.csv` et `candidates_job.csv` contiennent respectivement les données des offres d'emploi et des profils de candidats.

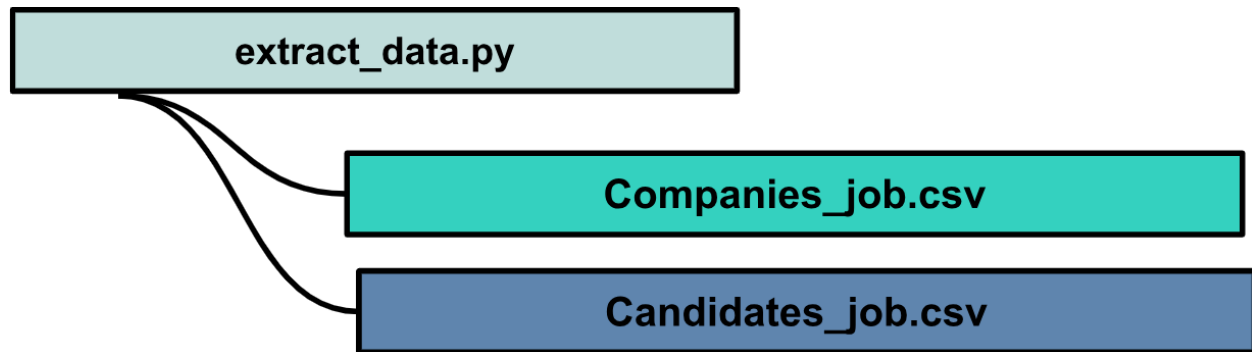


FIGURE 4.4 – Extraction des Données et Création des Fichiers CSV

4.2.2.3 Attributs des Offres d'Emploi et des Profils de Candidats

Les attributs extraits des offres d'emploi et des profils de candidats sont essentiels au processus de matching, facilitant la comparaison entre les exigences des entreprises et les qualifications, compétences et préférences des candidats. Ils sont soigneusement organisés dans des fichiers CSV pour un traitement optimisé.

Attributs des Offres d'Emploi

Les principaux attributs des offres d'emploi incluent :

Attribut	Description
Created At (Date de création)	La date à laquelle le profil du candidat a été créé.
created_by	Un dictionnaire contenant l'identifiant unique (uid) de l'utilisateur qui a créé l'offre d'emploi, ainsi qu'une liste de valeurs représentant les thèmes ou valeurs associées à cet utilisateur, comme l'innovation ou l'égalité des sexes.
Job Name (Titre du poste)	Le titre du poste proposé par l'entreprise.
Job Description (Description du poste)	Une description détaillée des responsabilités et des exigences du poste.
Technical Skills (Compétences techniques)	Les compétences techniques requises pour le poste.
Soft Skills (Compétences générales)	Les compétences non techniques ou générales nécessaires pour le poste.
Languages (Langues requises)	Les langues que le candidat doit maîtriser pour le poste.
Study Level (Niveau d'étude)	Le niveau d'étude requis pour le poste.
Contract Type (Type de contrat)	Le type de contrat proposé (CDI, CDD, freelance, etc.).
Experience (Expérience professionnelle)	Les années d'expérience ou les détails de l'expérience professionnelle requise pour le poste.
Availability (Disponibilité)	La disponibilité demandée pour le poste (immédiate, à partir de telle date, etc.).
Motivation (Motivation)	La motivation attendue du candidat pour le poste.
Job Time (Temps de travail)	Le type de temps de travail recherché par le candidat (temps plein, temps partiel, etc.).
Salary	Le salaire proposé pour le poste.

Tableau 4.1 : Détaille les attributs extraits pour les offres d'emploi (Companies_job.csv)

Attributs des Profils de Candidats

Les attributs extraits des profils de candidats comprennent :

Attribut	Description
Created At (Date de création)	La date à laquelle le profil du candidat a été créé.
created_by	Un dictionnaire contenant l'identifiant unique (uid) du candidat qui a créé le profil, ainsi qu'une liste de valeurs représentant les thèmes ou valeurs associés à ce candidat.
Job Name (Nom du poste recherché)	Le titre du poste que le candidat recherche ou pour lequel il postule.
Profil Description (Description du profil)	Une description détaillée du profil du candidat, incluant ses compétences, expériences, et objectifs.
Creator Name (Nom du créateur du profil)	Le nom du candidat.
Technical Skills (Compétences techniques)	Les compétences techniques que le candidat possède.
Soft Skills (Compétences générales)	Les compétences non techniques ou générales du candidat.
Study Level (Niveau d'étude)	Le niveau d'études atteint par le candidat.
Languages (Langues parlées)	Les langues que le candidat parle couramment.
Contract Types (Types de contrat souhaités)	Les types de contrat que le candidat préfère (CDI, CDD, freelance, etc.).
Job Time (Temps de travail)	Le type de temps de travail recherché par le candidat (temps plein, temps partiel, etc.).
Experience (Expérience professionnelle)	Les années d'expérience ou les détails de l'expérience professionnelle du candidat.
Availability (Disponibilité)	La disponibilité du candidat (immédiate, à partir de telle date, etc.).
Motivation (Motivation)	La motivation du candidat pour le poste ou le domaine.
Salary (Salaire attendu)	Le salaire que le candidat souhaite ou attend.

Tableau 4.2 : Présente les attributs pour les profils de candidats. (Candidates_job.csv)

4.2.2.4 Préparation des Données

Après la collecte des données, un processus de nettoyage est réalisé pour garantir une qualité optimale des informations. Ce processus comprend la gestion des valeurs manquantes, l'élimination des doublons et la normalisation des attributs, afin de préparer les données pour le matching.

Le traitement des fichiers CSV débute par une phase de nettoyage minutieuse. Cette étape consiste à compléter les valeurs manquantes et à convertir les informations dans un format compatible avec les algorithmes de traitement du langage naturel (NLP) utilisés ultérieurement. Voici les étapes essentielles de ce processus :

1. **Remplissage des valeurs manquantes** : Les champs vides sont complétés avec des chaînes vides (") pour éviter toute erreur lors de la vectorisation et garantir le bon fonctionnement des algorithmes de machine learning.
2. **Conversion des types de données** : Les colonnes pertinentes sont converties en chaînes de caractères afin d'assurer leur compatibilité avec les algorithmes d'analyse textuelle.
3. **Nettoyage des textes** : Cette étape, essentielle pour l'analyse NLP dans le projet Midl-Job, est réalisée à l'aide de la bibliothèque NLTK en Python et inclut plusieurs sous-étapes :
 - **Suppression des éléments non pertinents** : Les URL, chiffres, caractères spéciaux, caractères non-ASCII, et mentions d'utilisateurs (par exemple, "@utilisateur") sont éliminés, car ils n'apportent pas de valeur significative à l'analyse.
 - **Suppression de la ponctuation** : Les signes de ponctuation sont supprimés pour éviter d'ajouter du bruit à l'analyse textuelle.
 - **Conversion en minuscules** : Tous les textes sont convertis en minuscules pour assurer la cohérence de l'analyse, évitant que des mots identiques soient interprétés différemment en raison de la casse.
 - **Suppression des mots vides (stopwords)** : Les mots courants sans signification sémantique forte, tels que "est", "un", "le", etc., sont retirés pour se concentrer sur les mots-clés importants.

Enfin, les identifiants uniques (UID) sont extraits de la colonne `created_by` pour chaque enregistrement, ce qui permet d'associer de manière unique chaque profil de candidat ou offre d'emploi à une entrée unique. Cette étape est cruciale pour assurer un suivi précis et garantir la correspondance entre les offres et les candidats.

4.2.3 Utilisation des Embeddings Sémantiques

4.2.3.1 Définition des Embeddings

Les embeddings sont des vecteurs de dimension réduite qui représentent des textes, tels que des mots ou des phrases, dans un espace mathématique. Contrairement aux approches traditionnelles qui considèrent chaque mot individuellement, les embeddings tiennent compte du contexte global, ce qui permet de mieux saisir la signification et les relations entre les termes.

L'image ci-dessous, **Figure 4.5**, illustre les relations sémantiques capturées par les embeddings dans différents contextes. Par exemple :

- **Male-Female** : La relation sémantique entre "king" et "queen" est similaire à celle entre "man" et "woman". Cela montre que les embeddings peuvent comprendre les relations de genre.
- **Verb tense** : Les formes verbales comme "walked" et "walking" ou "swam" et "swimming" montrent comment les embeddings capturent les variations temporelles des verbes.
- **Country-Capital** : La relation entre des pays et leurs capitales, comme "Spain" et "Madrid", "Japan" et "Tokyo", est également bien représentée, illustrant la capacité des embeddings à saisir des connexions géographiques.

Ces vecteurs d'embeddings permettent ainsi de rapprocher des termes similaires dans un espace vectoriel, rendant visibles les relations sémantiques entre eux. Cela facilite des tâches complexes telles que la **classification**, le **clustering**, ou encore la **recherche de similarité** dans des textes.

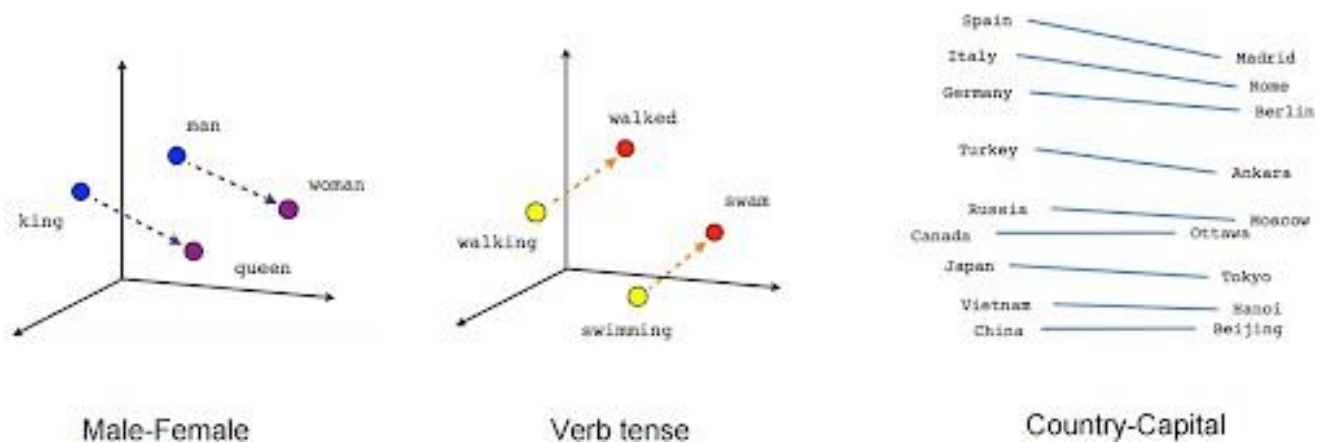


FIGURE 4.5 – Représentation des relations sémantiques avec des embeddings

4.2.3.2 Présentation du Modèle d'Embeddings "paraphrase-MiniLM-L6-v2"

Le modèle **paraphrase-MiniLM-L6-v2**, issu de la famille des sentence-transformers, a été choisi pour générer les embeddings des intitulés de poste. Ce modèle se distingue par sa capacité à produire des embeddings rapidement tout en capturant des nuances sémantiques importantes. Il est à la fois précis et léger, ce qui le rend adapté aux environnements de production ayant des contraintes de ressources.

L'image ci-dessus, **Figure 4.6**, illustre la représentation des relations sémantiques entre plusieurs intitulés de poste tels que Machine Learning Engineer, Backend Engineer, Data Scientist, et Software Developer. Le modèle est capable de capturer les similitudes entre ces titres, montrant par exemple comment "Machine Learning Engineer" est plus proche de "Data Scientist" qu'il ne l'est de "Backend Engineer", grâce à leur proximité sémantique.

Les avantages du modèle incluent :

- **Précision sémantique** : Il capture les relations sémantiques entre les phrases, facilitant ainsi la comparaison des textes selon leur signification globale.
- **Efficacité** : Il génère des embeddings rapidement, minimisant la latence.
- **Légereté** : Ce modèle est peu gourmand en ressources, ce qui est crucial dans les systèmes en temps réel.

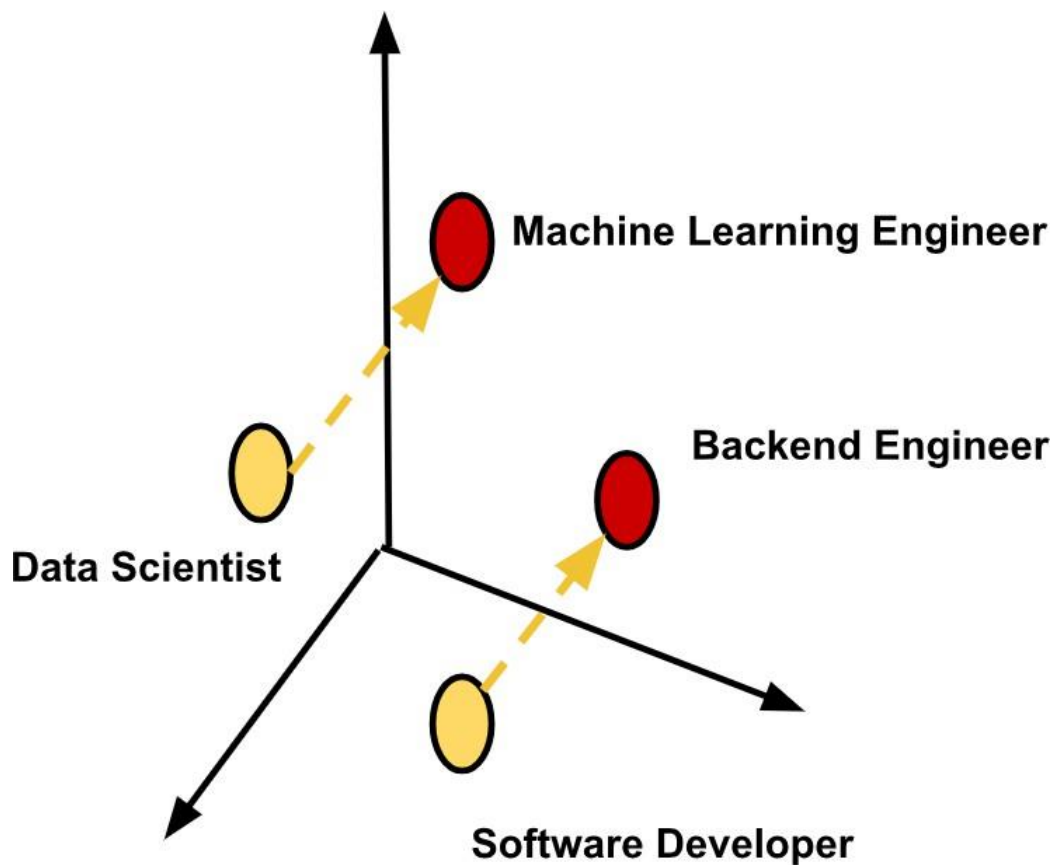


FIGURE 4.6 – Relations Sémantiques entre Intitulés de Poste avec le Modèle "paraphrase-MiniLM-L6-v2"

4.2.3.3 Transformation des Intitulés en Embeddings et Clustering

Dans ce projet, les intitulés de poste sont convertis en vecteurs d'embeddings via le modèle paraphrase-MiniLM-L6-v2. Chaque titre est ainsi transformé en vecteur numérique, capturant sa signification sémantique. Cela permet une comparaison plus précise des intitulés, en se basant sur leur contexte et leur relation sémantique plutôt que sur une simple correspondance de mots.

Les embeddings générés sont utilisés non seulement pour comparer les intitulés, mais aussi pour les regrouper en clusters via l'algorithme K-Means. Cela permet de structurer les offres d'emploi de manière cohérente, facilitant ainsi le matching entre les candidats et les postes. Ce processus réduit également l'impact des variations linguistiques en se concentrant sur la signification profonde des intitulés.

Le tableau ci-dessous présente quelques exemples de vecteurs d’embeddings générés pour différents intitulés de poste, illustrant comment ces représentations capturent les relations sémantiques entre les postes.

Nom de Poste	Vecteur d’Embeddings
Data Scientist	[0.23, -0.15, 0.67, ..., 0.45]
Ingénieur Machine Learning	[0.21, -0.17, 0.65, ..., 0.47]
Ingénieur Machine Learning	[0.19, -0.12, 0.69, ..., 0.42]
Développeur Web	[0.31, -0.08, 0.50, ..., 0.55]
Chef de Projet Informatique	[0.25, -0.14, 0.61, ..., 0.49]
Responsable Marketing	[0.18, -0.10, 0.62, ..., 0.44]
Technicien Réseau	[0.27, -0.09, 0.57, ..., 0.50]

Tableau 4.3 : Embeddings des Noms de Poste

4.2.4 Clustering des Intitulés de Poste avec K-Means

4.2.4.1 Présentation de l’Algorithme K-Means

K-Means est un algorithme non supervisé conçu pour diviser un ensemble de données en plusieurs clusters en fonction de leur similarité. Il fonctionne en regroupant les données autour de **K** centroïdes, où **K** est un paramètre défini à l’avance. Les données sont assignées au cluster dont le centroïde est le plus proche, ce qui permet de minimiser la variance intra-cluster et de maximiser la différence entre les clusters. Cet algorithme est largement utilisé dans diverses applications telles que la segmentation de marché, la reconnaissance de formes ou encore le traitement de grandes quantités de données.

4.2.4.2 Étapes du Clustering K-Means

Le schéma de la **Figure 4.7** illustre les principales étapes de l’algorithme K-Means :

— Initialisation :

Le processus débute par la sélection de K centroïdes aléatoires dans l’espace de données. Dans le schéma, les centroïdes sont représentés par des carrés verts et bleus, tandis que

les données non assignées sont les points roses (premier graphique, "Initialising $K = 2$ and Centroids").

— **Affectation :**

Chaque point de données est ensuite assigné au centroïde le plus proche, ce qui forme les premiers clusters. Dans le schéma, les points de données sont regroupés en clusters verts et bleus selon leur proximité avec les centroïdes (deuxième graphique, "Data Assignment").

— **Recalcul des centroïdes :**

Après cette affectation, les centroïdes sont mis à jour en recalculant la moyenne des points dans chaque cluster, ce qui déplace les centroïdes vers le centre des clusters nouvellement formés. Le schéma montre ce déplacement avec les carrés verts et bleus qui changent légèrement de position (troisième graphique, "Centroid Update").

— **Itérations :**

Ce processus d'affectation et de recalcul se répète plusieurs fois jusqu'à ce que les positions des centroïdes et des points ne changent plus. L'itération continue jusqu'à la stabilisation des clusters, comme illustré dans le dernier graphique, où les points sont définitivement répartis en clusters distincts (quatrième graphique, "Clustered Data").

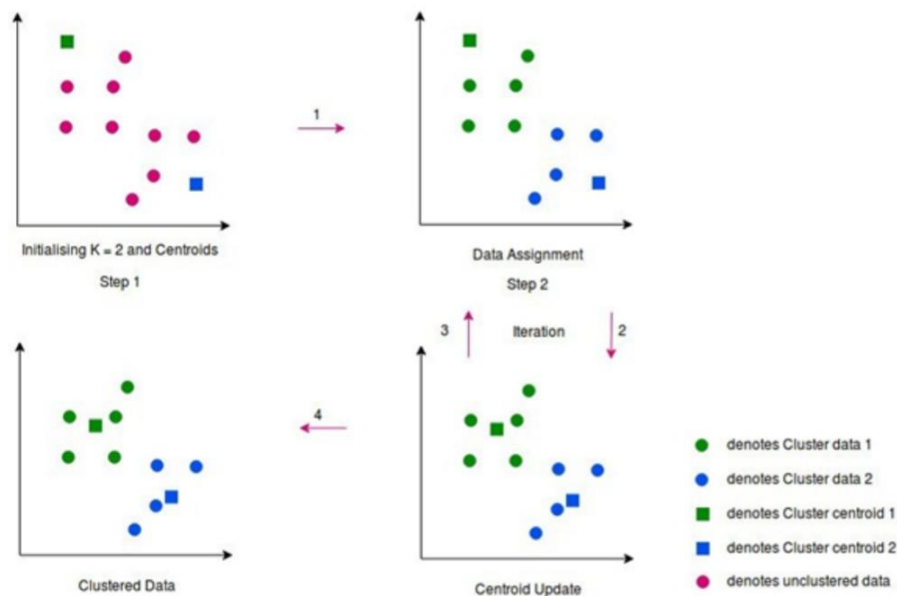


FIGURE 4.7 – Schéma du processus de clustering K-Means

4.2.4.3 Application des Embeddings dans K-Means

Dans ce projet, les vecteurs d'embeddings générés pour les intitulés de poste servent de base pour le clustering à l'aide de l'algorithme K-Means. Chaque intitulé, converti en vecteur, est regroupé avec d'autres postes similaires en fonction de leur proximité sémantique. Cette approche simplifie l'analyse et le rapprochement entre les profils des candidats et les offres d'emploi, facilitant ainsi l'identification des postes les plus pertinents dans le processus de recrutement.

Grâce à cette méthode, les intitulés sont organisés en groupes cohérents, ce qui améliore le traitement des autres caractéristiques textuelles et affine le matching.

Le tableau ci-dessous illustre comment ces vecteurs d'embeddings sont utilisés pour former des clusters avec K-Means. Chaque poste est affecté à un groupe en fonction de la similarité de son vecteur avec les autres. Par exemple, "Data Scientist" et "Ingénieur Machine Learning" sont regroupés dans le même cluster, indiquant qu'ils partagent des similarités sémantiques importantes.

Nom de Poste	Vecteur d'Embeddings	Cluster Assigné
Data Scientist	[0.23, -0.15, 0.67, ..., 0.45]	Cluster 1
Ingénieur Machine Learning	[0.21, -0.17, 0.65, ..., 0.47]	Cluster 1
Analyste de Données	[0.19, -0.12, 0.69, ..., 0.42]	Cluster 1
Développeur Web	[0.31, -0.08, 0.50, ..., 0.55]	Cluster 2
Chef de Projet Informatique	[0.25, -0.14, 0.61, ..., 0.49]	Cluster 3
Responsable Marketing	[0.18, -0.10, 0.62, ..., 0.44]	Cluster 4
Technicien Réseau	[0.27, -0.09, 0.57, ..., 0.50]	Cluster 2

Tableau 4.4 : Clustering des Noms de Poste avec K-Means

4.2.5 Vectorisation des Textes avec TF-IDF

La méthode TF-IDF est utilisée pour convertir les caractéristiques textuelles pertinentes en vecteurs numériques, ce qui permet d'affiner le processus de matching entre les offres d'emploi et les profils de candidats.

4.2.5.1 Présentation de la Méthode TF-IDF

Le TF-IDF (Term Frequency-Inverse Document Frequency) est une technique largement utilisée en traitement automatique du langage naturel (NLP) pour évaluer l'importance relative des mots dans un document, et ainsi les représenter sous forme de vecteurs numériques. La méthode repose sur deux composantes :

- **Fréquence des Termes (TF)** : Représente le nombre d'occurrences d'un terme dans un document, pondéré par la longueur de celui-ci.
- **Fréquence Inverse des Documents (IDF)** : Mesure la rareté d'un terme dans un corpus, afin de diminuer l'impact des mots fréquemment utilisés dans plusieurs documents.

4.2.5.2 Calcul de la Méthode TF-IDF

Le TF-IDF est calculé en deux étapes :

1. **Term Frequency (TF)** : La fréquence d'un terme est obtenue en divisant le nombre de fois où le terme apparaît dans un document par le nombre total de termes du document.

$$TF(t, d) = \frac{f_{t,d}}{N_d}$$

Où :

- $f_{t,d}$: Nombre de fois que le terme t apparaît dans le document d .
 - N_d : Nombre total de termes dans le document d .
2. **Inverse Document Frequency (IDF)** : L'IDF évalue la rareté d'un terme dans le corpus. Si un terme apparaît dans peu de documents, son IDF sera élevé, indiquant qu'il est un bon indicateur.

$$IDF(t) = \log \left(\frac{N}{n_t} \right)$$

Où :

- **N** : Nombre total de documents dans le corpus.
- **n_t** : Nombre de documents contenant le terme t .

La valeur TF-IDF pour un mot dans un document est obtenue en multipliant la fréquence du terme (TF) par son poids d'importance (IDF). Ce produit permet de pondérer les mots courants moins lourdement que les termes plus spécifiques et rares qui sont plus représentatifs du contenu du document.

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t)$$

4.2.5.3 Application de TF-IDF dans le Projet

Dans ce projet, la méthode TF-IDF a été utilisée pour vectoriser les principales caractéristiques textuelles des offres d'emploi et des profils de candidats. Les éléments textuels suivants ont été sélectionnés pour la vectorisation :

- Descriptions de poste et de profil
- Compétences techniques
- Langues parlées et requises
- Type et durée du contrat
- Disponibilité
- Expérience professionnelle
- Temps de travail

Le tableau ci-dessous illustre l'application du TF-IDF sur deux textes distincts, **Text1** et **Text2**. Les termes spécifiques à chaque texte, comme "basic" et "linux" pour **Text1**, ou "essential" et "dvc" pour **Text2**, obtiennent des valeurs TF-IDF plus élevées car ils apparaissent fréquemment dans leur texte respectif tout en étant absents dans l'autre. En revanche, les termes communs aux deux textes, tels que "data" et "science", obtiennent des valeurs TF-IDF plus faibles, car leur occurrence est répartie entre les deux textes, les rendant moins discriminants.

Text1 : Basic Linux Commands for Data

Text2 : Essential DVC Commands for Data Science

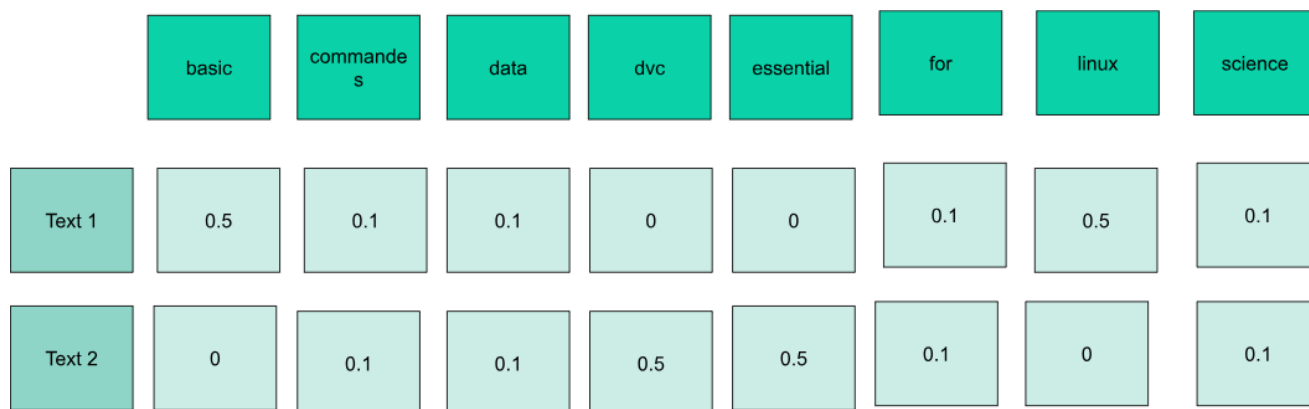


FIGURE 4.8 – Illustration du processus de vectorisation TF-IDF

4.2.6 Calcul des matrices de Similarité

Après avoir vectorisé les principales caractéristiques textuelles des offres d’emploi et des profils des candidats en utilisant la méthode TF-IDF, la prochaine étape consiste à calculer les matrices de similarité. Ces matrices permettent de quantifier la ressemblance entre les profils des candidats et les offres d’emploi en prenant en compte des critères tels que le titre de poste, les compétences techniques, l’expérience professionnelle et d’autres caractéristiques.

4.2.6.1 Similarité Cosinus

La similarité cosinus est une mesure mathématique utilisée pour évaluer la similarité entre deux vecteurs en calculant le cosinus de l’angle formé entre eux. Cette mesure est couramment utilisée dans le traitement du langage naturel (NLP) pour comparer des documents textuels transformés en vecteurs numériques, comme ceux générés par TF-IDF. La similarité cosinus permet de quantifier de manière précise le degré de similitude entre ces documents.

Fonctionnement :

La similarité cosinus évalue l'orientation relative de deux vecteurs, indépendamment de leur longueur. Cette approche est particulièrement utile pour comparer des documents textuels, car elle met l'accent sur l'importance relative des termes plutôt que sur leur quantité absolue.

- **Produit Scalaire :** Le calcul commence par déterminer le produit scalaire des deux vecteurs, représentant la somme des produits des composantes correspondantes des vecteurs.
- **Normes des Vecteurs :** Les normes (ou longueurs) des vecteurs sont ensuite calculées, normalisant ainsi le produit scalaire pour garantir que la comparaison soit basée uniquement sur l'orientation relative.
- **Détermination du Cosinus de l'Angle :** La similarité cosinus est obtenue en divisant le produit scalaire par le produit des normes des vecteurs. Le résultat est une valeur comprise entre -1 et 1 :
 - **1 :** Les vecteurs sont parfaitement alignés (similaires).
 - **0 :** Les vecteurs sont orthogonaux (aucune similarité).
 - **-1 :** Les vecteurs sont opposés (dissimilarité complète).

Formule de la similarité cosinus :

$$\text{Similarité Cosinus}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|}$$

Où :

- **A.B** est le produit scalaire des vecteurs **A** et **B**
- **||A||** et **||B||** sont les normes des vecteurs **A** et **B** respectivement

La Figure ci-dessous montre un exemple de la similarité cosinus entre deux vecteurs.

Vector Similarity Measures : Cosine

$$\text{Similarité Cosinus}(A, B) = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}} \quad \text{Cos(Data Scientist, Data Analyst)} = \frac{(0.4 \times 0.6) + (0.8 \times 0.6)}{\sqrt{(0.4^2 + 0.8^2)} \times \sqrt{(0.6^2 + 0.6^2)}}$$

$$= \frac{0.24 + 0.48}{\sqrt{0.16 + 0.64} \times \sqrt{0.36 + 0.36}}$$

$$= \frac{0.72}{\sqrt{0.8} \times \sqrt{0.72}}$$

$$= \frac{0.72}{0.894 \times 0.848}$$

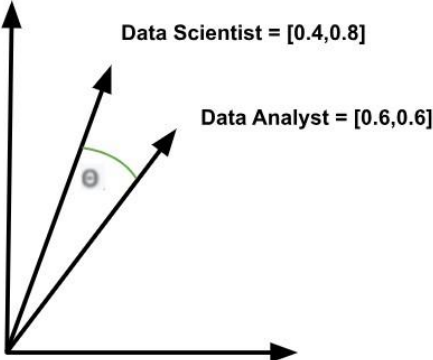
$$\approx 0.95$$


FIGURE 4.9 – Illustration de la similarité cosinus entre deux vecteurs

4.2.6.2 Calcul de la Matrice de Similarité Globale

Une fois que les matrices de similarité ont été calculées pour chaque caractéristique textuelle (par exemple, titres de poste, compétences, langues parlées), elles sont combinées pour former une matrice de similarité globale. Cette matrice permet d'obtenir une vue d'ensemble des correspondances entre les candidats et les offres en tenant compte de tous les aspects textuels vectorisés.

La matrice globale \mathbf{M}_{global} est obtenue en prenant la moyenne des différentes matrices de similarité individuelles :

$$M_{global} = \frac{1}{n} \sum_{i=1}^n M_i$$

Où n représente le nombre de matrices de similarité associées aux différentes caractéristiques textuelles.

4.2.7 Sélection des Meilleures Correspondances

4.2.7.1 Tri des Correspondances par Scores de Similarité

Après avoir calculé la **matrice de similarité globale**, les correspondances entre les candidats et les offres d'emploi sont triées en fonction de leur **score de similarité**. Ce score, noté S_{ij} , représente la proximité entre un candidat i et une offre d'emploi j . Plus le score S_{ij} est élevé, plus la correspondance entre le candidat et l'offre est pertinente. La formule suivante illustre le calcul du score de similarité pour une paire (candidat, offre) :

$$S_{ij} = M_{global}(i, j)$$

Ce score est extrait directement de la **matrice de similarité globale** M_{global} qui combine les similarités issues de différentes caractéristiques textuelles (compétences, descriptions de poste, langues, etc.). Ce processus permet de trier les paires candidat-offre du plus pertinent au moins pertinent.

4.2.7.2 Extraction des Meilleures Correspondances (Top N)

Une fois les correspondances triées par scores de similarité, le système extrait les **N meilleures correspondances** pour chaque candidat ou offre d'emploi. Ces correspondances, ayant les scores S_{ij} les plus élevés, sont jugées les plus pertinentes.

La formule suivante est utilisée pour sélectionner ces **Top N** correspondances :

$$\text{Top N} = \text{Top } N(S_{ij})$$

Ces correspondances constituent les recommandations finales, facilitant ainsi la prise de décision pour les recruteurs et les candidats.

4.3 Version Premium

La version Premium de MidlJob va au-delà des fonctionnalités offertes par la version gratuite en intégrant un critère déterminant pour le matching : la comparaison des personnalités entre les

candidats et les recruteurs. En plus des critères habituels tels que le titre du poste, la description du poste, les compétences techniques, l'expérience professionnelle, le type de contrat, le niveau d'étude et les langues maîtrisées, cette version évalue également la compatibilité de la personnalité des candidats avec les attentes culturelles et comportementales spécifiées par les recruteurs.

Le code et le modèle sous-jacents restent les mêmes que ceux utilisés dans la version gratuite. Cependant, la version Premium intègre des éléments supplémentaires pour enrichir le processus de matching, notamment les soft skills, la motivation et les avantages recherchés par les candidats.

Le système utilise des techniques avancées de traitement du langage naturel (NLP) pour analyser en profondeur les descriptions fournies par les candidats et les recruteurs. Ces descriptions sont transformées en traits de personnalité clés, qui sont ensuite comparés pour évaluer la correspondance. Par exemple, si un recruteur recherche un chef de projet proactif, capable de prendre des initiatives, de gérer le stress et de travailler efficacement en équipe, et qu'un candidat se décrit comme étant orienté vers l'action, avec une forte capacité de gestion du temps et une excellente résilience face au stress, le système détectera et valorise cette compatibilité.

Ce processus de matching enrichi ne se limite donc pas à l'alignement sur les compétences techniques et l'expérience, mais intègre également la compatibilité des traits de personnalité, les soft skills, la motivation et les avantages. Cette approche holistique du recrutement permet d'augmenter considérablement les chances de succès à long terme en alignant les candidats non seulement sur les qualifications professionnelles mais aussi sur la culture et les valeurs de l'entreprise. Ainsi, la satisfaction mutuelle et la rétention des talents sont renforcées, assurant une intégration harmonieuse du candidat au sein de l'équipe et de l'entreprise.

4.4 Choix d'API et Comparaison

4.4.1 Choix d'API

Une API (Interface de Programmation d'Applications) est un intermédiaire logiciel qui permet à deux applications de communiquer entre elles. Lorsque vous utilisez une application sur votre téléphone, celle-ci se connecte à Internet et envoie des données à un serveur. Le serveur traite ensuite les données et les renvoie à votre téléphone. L'application sur votre téléphone interprète alors les données et vous les présente de manière lisible.

Les API sont l'épine dorsale de l'architecture moderne car elles permettent aux applications d'être modulaires et découplées. Cela signifie que vous pouvez construire des applications rapidement et facilement, ce qui vous permet de les maintenir et de les mettre à jour facilement.

Une API est comme un serveur dans un restaurant. Le serveur prend votre commande et l'envoie à la cuisine. La cuisine prépare ensuite la nourriture et la renvoie au serveur. Le serveur vous apporte ensuite la nourriture.

Dans notre cas, un utilisateur envoie une requête qui contient des informations spécifiques, comme l'ID d'un candidat. Cette requête est traitée par l'API, qui extrait les informations nécessaires du serveur (par exemple, les offres d'emploi correspondantes) et les renvoie à l'utilisateur sous une forme compréhensible, telle qu'une liste de recommandations.

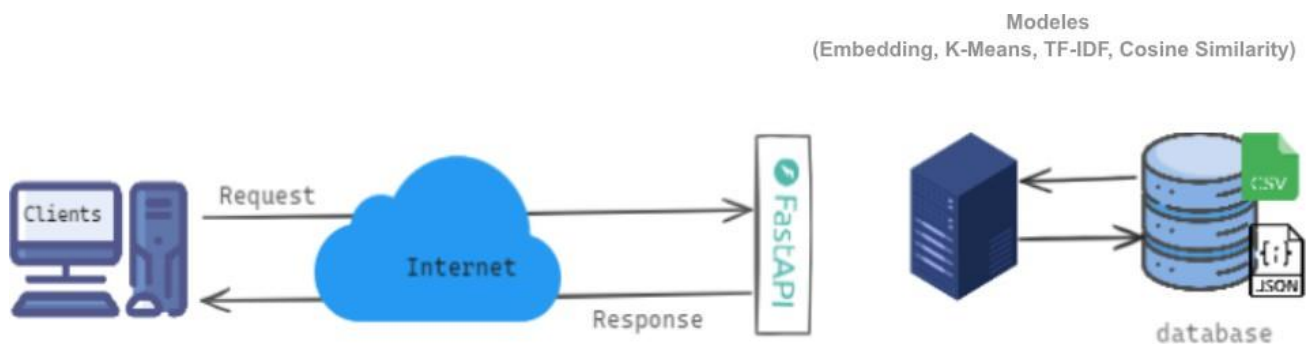


FIGURE 4.10 – Architecture d'une application Web

Les API sont également très importantes en apprentissage automatique car elles permettent à différentes applications de partager des données et de travailler ensemble, ce qui permet de gagner du temps et des efforts. Il existe de nombreux frameworks différents pour construire des API en Python. Certains des frameworks les plus populaires pour créer des API en Python sont Django, Flask et FastAPI. Ce tutoriel est une plongée profonde dans l'un des frameworks appelé FastAPI.

4.4.2 Comparaison des APIs

Comparaison de FastAPI avec Django et Flask

Les trois de ces frameworks sont des frameworks web Python que vous pouvez utiliser pour développer des applications web. Ils ont chacun leurs propres forces et faiblesse :

Django est un framework complet qui inclut tout ce dont vous avez besoin pour commencer, y compris un ORM intégré et un panneau d'administration. Il peut être un peu écrasant pour les débutants, mais sa documentation exhaustive le rend facile à apprendre.

Flask est un microframework léger et facile à démarrer. Il n'inclut pas autant de fonctionnalités que Django, mais il est parfait pour des projets simples.

FastAPI est un nouveau framework conçu pour être rapide et facile à utiliser. Il inclut des fonctionnalités telles que la validation automatique des données et la documentation.

	Django	Flask	FastAPI
Performance	Django est massif. Ce n'est pas le meilleur en terme de performance	Flask est un micro Framework web. Il se comporte mieux que Django	FastAPI est l'un des frameworks web les plus rapides avec un support asynchrone natif qui améliore l'efficacité du
Support asynchrone	Oui, avec une latence limitéeframework	Non. Nécessité	FastAPI fournit un support asynchrone natif
Facilité d'utilisation	Django est massif et donc un peu compliqué à apprendre	Flask est facile à apprendre et assez simple à utiliser	FastAPI est le plus simple des trois
Documentation interactive	Non interactive	Non	Oui (OpenAPI, Redoc)
Vérification des données	Non	Non	Oui

Tableau 4.5 : Comparaison de FastAPI avec Django et Flask

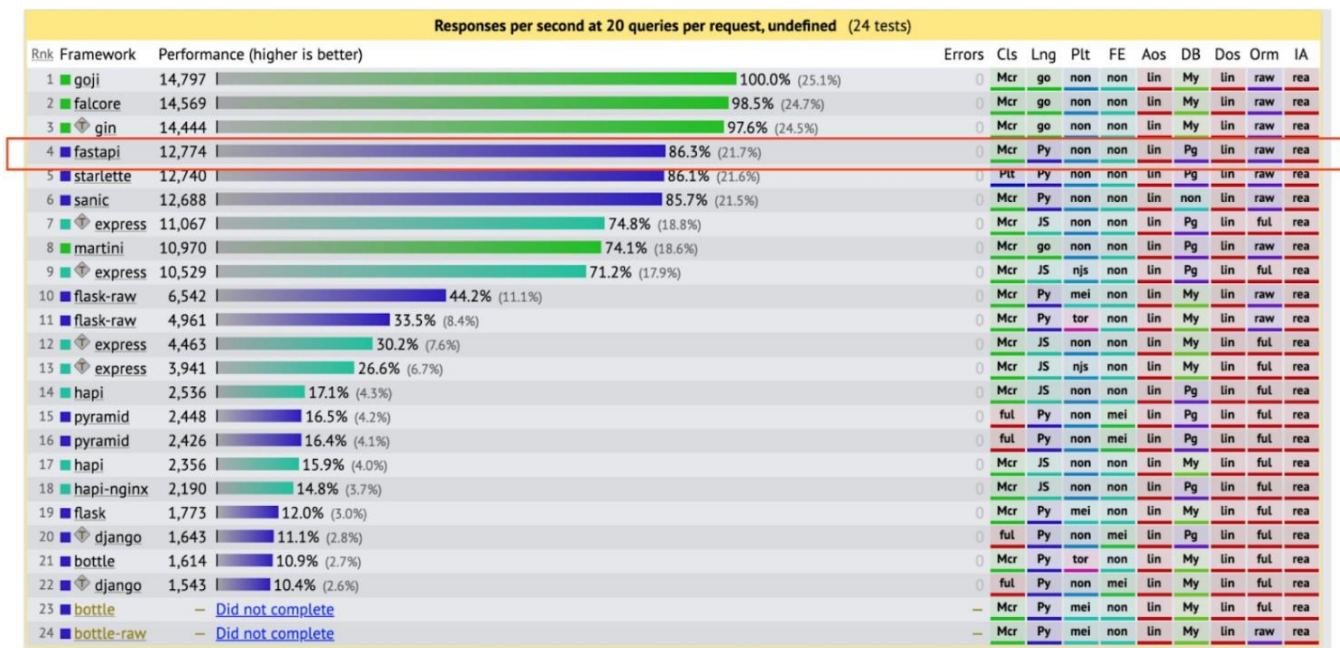


FIGURE 4.11 – Benchmarks de performance de FastAPI

4.5 Conclusion

En résumé, la méthodologie décrite dans ce chapitre démontre l'efficacité du système MidlJob dans la sélection des correspondances les plus pertinentes. En allant au-delà des simples compétences techniques, la version Premium enrichit encore le processus en considérant les aspects personnels et comportementaux, augmentant ainsi la probabilité d'un matching réussi à long terme pour les recruteurs et les candidats.

Chapitre 5

Résultats

5.1 Introduction

Ce chapitre vise à exposer les résultats obtenus à la suite de l’implémentation du projet Midl-Job, une solution innovante de matching entre les offres d’emploi et les candidats. En s’appuyant sur des algorithmes avancés de machine learning et de traitement du langage naturel (NLP), le système a été évalué à travers deux versions distinctes : la version gratuite et la version premium. Chacune de ces versions se différencie par les critères d’évaluation employés pour établir des correspondances optimales entre les candidats et les offres. Tandis que la version gratuite se concentre principalement sur les compétences techniques, les intitulés de poste et les critères de base (type de contrat, description du poste, etc.), la version premium apporte une dimension supplémentaire en prenant en compte des aspects plus subjectifs tels que les soft skills, les motivations personnelles et les valeurs culturelles du candidat. À travers cette analyse, nous examinerons la précision, les forces et les axes d’amélioration des deux versions, tout en explorant les perspectives futures pour l’évolution du système.

5.2 Résultats du Matching dans la Version Free

Dans le cadre de notre projet de matching entre les offres d’emploi et les profils de candidats, nous avons développé un système capable de prédire les meilleures correspondances pour un poste donné ainsi que les meilleures offres pour un candidat spécifique. Ces résultats sont obtenus grâce à l’utilisation d’algorithmes de machine learning, tels que la vectorisation TF-IDF, l’application

de l'algorithme K-Means pour le clustering, et l'algorithme de Cosine Similarity pour mesurer la similarité entre les profils et les offres.

5.2.1 Prédiction des Meilleurs Candidats pour un Poste

Pour illustrer les capacités de ce système, nous avons sélectionné une offre d'emploi pour un poste de Data Scientist. Cette offre se distingue par des compétences techniques spécifiques telles que Python, R, SQL, pandas, et scikit-learn, qui sont des technologies essentielles dans le domaine de la science des données. L'offre inclut également des critères supplémentaires tels qu'une disponibilité immédiate, un contrat à durée déterminée (CDD), ainsi qu'une préférence pour des candidats sans expérience, reflétant une certaine flexibilité dans les qualifications recherchées. Voici une présentation détaillée.

```
created_at: 2024-07-25 22:15:45.307000+00:00
remote_status: Full Remote
job_name: Data scientist
job_video_description: False
ignored_jobs:
salary: 💰 Jusqu'à 20 000 €
technical_skills: ['Python', 'SQL', 'R', 'pandas', 'NumPy', 'scikit-learn']
job_time: ⌚ Temps partiel
distance:
creator_name: facebook
contract_time:
experience: 🌱 Sans expérience
motivationLetter:
job_type: entreprise
image_url: https://firebasestorage.googleapis.com/v0/b/midlv1.appspot.com/o/default-images%2F1.png?alt=media&token=e92ecd66-0d60-45ef-b369-2c2f5d1215e6
availability: 🕒 Immédiate
advantages: ['🕒 Possibilité de travailler à temps partiel', '🏠 Possibilité de travailler à domicile', '📅 Accès à des événements professionnels']
created_by: {'uid': 'cMsOpntnRfZ1yrTkKb7Ajz3qFnk1', 'values': ['👉 Égalité des chances', '👉 Droits de l'homme', '👉 d'Égalité des sexes']}
activityDomains: ['🖥️ Informatique']
contract_types: ['📄 CDD']
study_level: 🎓 Master
job_audio_description:
likes_jobs:
profile_video_description: False
updated_at:
job_description: Nous recherchons un Data Scientist talentueux et motivé pour rejoindre notre équipe dynamique. En tant que Data Scientist, vous serez re
latLong: <google.cloud.firestore_v1_helpers.GeoPoint object at 0x0000021EB42AD6A0>
profile_audio_description:
soft_skills: ['Communication 🗣️', 'Intelligence émotionnelle ❤️', 'Confidentialité 🛡️', 'Leadership 👑', 'Analyse 🧠']
profile_description: Nous recherchons un Data Scientist en Recherche doté de solides compétences techniques et d'une passion pour l'innovation. Le candid
languages: ['Anglais GB', 'Français FR', 'Amazighe MA', 'Arabe SA', 'Chinois Cantonais HK']
```

FIGURE 5.1 – Détails de l'Offre Data Scientist

Lors du matching pour cette offre de Data Scientist, plusieurs candidats ont été identifiés comme étant des correspondances optimales, chacun ayant un score de similarité attribué selon leurs compétences, expériences et autres caractéristiques pertinentes.

```

/usr/local/lib/python3.10/dist-packages/transformers/tokenization_utils_base.py:1601: FutureWarning: 'clean_up_tokenization_spaces' was not set. It will be set to 'True' by default in future versions.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/tokenization_utils_base.py:1601: FutureWarning: 'clean_up_tokenization_spaces' was not set. It will be set to 'True' by default in future versions.
warnings.warn(
Meilleures correspondances pour le poste avec UID 'cMsOpntnR7ZlyrTkbb7Aj3qFnk1':
{'Best Match Candidate': 'Simo', 'Best Match Candidate UID': 'IeegLwplwvfm3aAM32aIV7lkEW52', 'Similarity Score': 0.65}
{'Best Match Candidate': 'AOUAII', 'Best Match Candidate UID': 'OTZ6adVqhZruV9gW377ynUnkcZ22', 'Similarity Score': 0.40}
{'Best Match Candidate': 'Jad', 'Best Match Candidate UID': 'LOF9tFXMDVal59MYNbLoJvG52Bm1', 'Similarity Score': 0.32}
{'Best Match Candidate': 'Nihan', 'Best Match Candidate UID': 'yRrTRP6RK5ZuKoHlMk6H2jkzEB3', 'Similarity Score': 0.20}

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of 'n_init' will change from 10 to 'auto' in 1.4. Set the value of 'n_init' explicitly to
super()._check_params_vs_input(X, default_n_init=10)

```

FIGURE 5.2 – Meilleures Correspondances - Candidats pour Data Scientist

1. Candidat : Simo

- **UID du Candidat** : CpfGRJ9nfJg0dgeGvLk4hvsSKRF3
- **Score de Similarité** : 0.65
- **Description** : Simo présente une excellente correspondance avec les exigences du poste de Data Scientist grâce à ses compétences en Python, pandas, SQL, NumPy, et scikit-learn. Ce poste à temps plein, de type CDD, correspond parfaitement à son profil, et sa disponibilité immédiate constitue un atout pour l'employeur.

Bien qu'il soit sans expérience professionnelle, son Master en informatique et sa maîtrise de plusieurs langues (dont l'anglais) renforcent son aptitude à occuper ce poste. La description du poste reflète ses compétences en data science, ce qui fait de lui un candidat pertinent pour cette opportunité.

```

created_at: 2024-07-25 22:36:44.073000+00:00
remote_status: Oui
job_name: Data scientist
job_video_description: false
ignored_jobs: [google.cloud.firestore_v1.document.DocumentReference object at 0x0000021EB3E7EBA0]
salary: 🍷 Jusqu'à 20 000 €
technical_skills: ['Python', 'R', 'SQL', 'NumPy', 'pandas', 'scikit-learn']
job_time: 🕒 Temps plein
distance: 30.0
creator_name: simo
contract_time:
experience: 🍷 Sans expérience
motivationLetter: Je vous écris pour exprimer ma profonde motivation à rejoindre votre équipe en tant que Data Scientist. Mon parcours académique et prof
job_type: candidat
image_url: https://firebasestorage.googleapis.com/v0/b/midlv1.appspot.com/o/default-images%2F2.png?alt=media&token=bf473604-6034-469f-99ad-775552a1ed3f
availability: 🕒 Immédiate
advantages: ['🍷 Accès à des événements professionnels', '🏠 Possibilité de travailler à domicile', '🕒 Possibilité de travailler à temps partiel']
created_by: {'uid': 'IeegLwplwvfm3aAM32aIV7lkEW52', 'values': ['🍷 Égalité des chances', '🍷 Droits de l'homme', '🍷 Égalité des sexes']}
activityDomains: ['🍷 Informatique']
contract_types: ['🍷 CDD']
study_level: 🍷 Master

```

FIGURE 5.3 – Profil du Candidat Simo pour Data Scientist

2. Candidat : Aouatif

- **UID du Candidat** : OTz6adYq2hZruV9qW377ynUnkcZ2
- **Score de Similarité** : 0.40
- **Description** : Aouatif est un candidat qualifié pour le poste de Data Analyst, grâce à ses compétences en Python, SQL, Big Data, Scikit-learn, et Pandas. Le poste, à temps plein en CDD, correspond à son profil. Avec 3 à 5 ans d'expérience et un Master en informatique, il possède l'expertise nécessaire pour ce rôle. Sa disponibilité immédiate le rend encore plus attractif pour les employeurs.

```
created_at: 2024-07-25 10:47:51.590000+00:00
remote_status: Oui
job_name: Data Analyst
job_video_description: false
ignored_jobs: [<google.cloud.firestore_v1.document.DocumentReference object at 0x0000021EB4283230>]
salary: 💰 Jusqu'à 20 000 €
technical_skills: ['Python', 'SQL', 'R', 'Big Data', 'Scikit-Learn', 'Pandas']
job_time: ⌚ Temps plein
distance: 45.0
creator_name: AOUATIF
contract_time:
experience: 🌱 3 à 5 ans d'expérience
motivationLetter: je suis un personne motiver et j'aime beacoup de travaille avec le goupe
job_type: candidat
image_url: https://firebasestorage.googleapis.com/v0/b/midlvl1.appspot.com/o/default-images%2F2.png?alt=media&token=bf473604-6034-469f-99ad-775552a1ed3f
availability: 📅 Immédiate
advantages: ["💎 Prime d'intéressement", "🏠 Plan d'épargne entreprise - PEE", "💰 Remboursement des frais de transports en commun"]
created_by: {'uid': 'OTZ6adYq2hZruV9qW377ynUnkcZ2', 'values': ["👤 Droits de l'homme ", "🌱 Protection de l'environnement ", "👦 Lutte contre la pauvreté"]}
activityDomains: ['💻 Informatique']
contract_types: ['📄 CDD']
study_level: 🎓 Master
```

FIGURE 5.4 – Profil du Candidat Aouatif pour Data Analyst

3. Candidat : Jad

- **UID du Candidat** : LO9FfxDYbL59MNYxlO3VG52Bm1
- **Score de Similarité** : 0.32
- **Description** : Jad présente une correspondance modérée pour le poste de Data Engineer, avec des compétences en Python, SQL, NumPy, et Google Cloud. Le poste à temps plein et sous contrat CDD correspond à son profil. Malgré un manque d'expérience professionnelle, son Master en informatique lui apporte une base solide. Sa disponibilité immédiate est un avantage pour les employeurs. Bien que son profil ne soit pas totalement aligné avec les attentes globales, son expertise en technologies cloud pourrait être précieuse pour certains projets.

```

created_at: 2024-07-25 23:07:16.750000+00:00
remote_status: Oui
job_name: Data engineer
job_video_description: false
ignored_jobs:
salary: 🏠 Jusqu'à 20 000 €
technical_skills: ['Python', 'SQL', 'Java', 'numpy', 'Google Cloud']
job_time: ⌚ Temps plein
distance: 30.0
creator_name: jad
contract_time:
experience: 🌱 Sans expérience
motivationLetter: Je suis passionné(e) par l'ingénierie des données et la manière dont les données peuvent transformer les entreprises. Avec [X] années
Mon expérience m'a également appris l'importance de la collaboration et de la communication efficace avec des équipes pluridisciplinaires. Je suis motiv
job_type: candidat
image_url: https://firebasestorage.googleapis.com/v0/b/midlv1.appspot.com/o/default-images%2F2.png?alt=media&token=bf473604-6034-469f-99ad-775552a1ed3f
availability: 🕒 Immédiate
advantages: ['🏠 Possibilité de travailler à domicile', '⌚ Possibilité de travailler à temps partiel']
created_by: {'uid': 'LOF9FtXMDYalS9MYNxL0JvG52Bm1', 'values': ['👉 Égalité des chances ', '🌱 Diversité et inclusion ', "👉 Droits de l'homme "]}
activityDomains: ['🖥 Informatique']
contract_types: ['📄 CDD']
studv_level: 🎓 Master

```

FIGURE 5.5 – Profil du Candidat Jad pour Data Engineer

4. Candidat : Nihar

- **UID du Candidat** : yRTrF6RK5ZuKolLmMK6JzKzEB3
- **Score de Similarité** : 0.20
- **Description** : Ce candidat Nihar présente des compétences en **Power BI**, **Python**, et **Hadoop**, mais ces compétences ne correspondent que partiellement aux exigences du poste de **Data Scientist**. Il a été sélectionné en raison du nombre limité d'annonces disponibles dans l'application.

```

created_at: 2024-07-04 07:39:14.558000+00:00
remote_status: Indifférent
job_name: Data Analyst
job_video_description: false
ignored_jobs:
salary: 🏠 Jusqu'à 20 000 €
technical_skills: ['Power BI', 'Python', 'Hadoop']
job_time: ⌚ Temps plein
distance: 30.0
creator_name: Nihar
contract_time:
experience: 🌱 1 à 3 ans d'expérience
motivationLetter: Motivation
job_type: candidat
image_url: https://firebasestorage.googleapis.com/v0/b/midlv1.appspot.com/o/default-images%2F2.png?alt=media&token=bf473604-6034-469f-99ad-775552a1ed3f
availability: 🕒 Immédiate
advantages: ['🚌 Remboursement des frais de transports en commun', '🍽 Remboursement des frais de repas', '❤ Assurance vie']
created_by: {'uid': 'yRTrF6RK5ZuKolLmMK6JzKzEB3', 'values': ['🚫 NoRacism ', '👉 Égalité des chances ', '🌱 Diversité et inclusion ']}
activityDomains: ['🖥 Informatique', '🏢 Management et Conseil aux Entreprises', '🏭 Industrie Manufacturière', '🌱 Énergie et Exploitation des Ressources']
contract_types: ['📄 Alternance']
study_level: 🎓 Master

```

FIGURE 5.6 – Profil du Candidat Nihar pour Data Analyst

Les résultats montrent une diversité dans les correspondances, avec Simo se démarquant comme le candidat le plus pertinent. Bien que des candidats comme Jad et Aouatif aient des scores plus faibles, leurs compétences pourraient répondre à des besoins complémentaires au sein de l'équipe.

5.2.2 Prédiction des Meilleurs Postes pour un Candidat

Le système de matching permet de prédire les meilleures offres d'emploi pour un candidat donné en comparant ses compétences techniques, ses valeurs et ses préférences avec les offres disponibles dans la base de données. L'objectif est de proposer une sélection d'offres qui correspond de manière optimale au profil du candidat, facilitant ainsi le processus de recrutement pour les employeurs et les candidats.

— Candidat : Dati

— **UID du Candidat** : CpfGRJ9nfJg0dgeGvLk4hvsSKRF3

— **Poste Recherché** : Développeur Web

— **Compétences Techniques** : Le candidat possède des compétences solides en HTML, CSS, SQL, JavaScript, Git, Django, et Python, et recherche un poste de Développeur Web. Ces compétences ont joué un rôle essentiel dans le score de correspondance obtenu.

```
created_at: 2024-07-25 23:22:02.579000+00:00
remote_status: Oui
job_name: Développeur.euse web
job_video_description: false
ignored_jobs:
salary: 🏆 Jusqu'à 20 000 €
technical_skills: ['HTML/CSS', 'SQL', 'JavaScript', 'Git', 'Django', 'python']
job_time: ⌚ Temps plein
distance: 30.0
creator_name: dati
contract_time:
experience: 🧑 Sans expérience
motivationLetter: Je suis passionné(e) par le développement web et les technologies qui transforment notre façon d'interagir avec le monde numérique. Av
job_type: candidat
image_url: https://firebasestorage.googleapis.com/v0/b/midlvl1.appspot.com/o/default-images%2F2.png?alt=media&token=bf473604-6034-469f-99ad-775552a1ed3f
availability: 🕒 Immédiate
advantages: ['⌚ Possibilité de travailler à temps partiel', '🏠 Possibilité de travailler à domicile']
created_by: {'uid': CpfGRJ9nfJg0dgeGvLk4hvsSKRF3, 'values': ['👤 Égalité des chances ', '👤 Droits de l'homme ', '👤 Égalité des sexes ']}
activityDomains: ['💻 Informatique']
contract_types: ['📄 CDD']
study_level: 🎓 Master
```

FIGURE 5.7 – Recherche d'Offres Correspondant au Profil Développeur Web

Notre système a identifié deux offres principales pour le candidat Dati. Lors du processus de matching, ces deux offres ont été sélectionnées comme étant les correspondances optimales. Chaque offre a obtenu un score de similarité basé sur les compétences, l'expérience, et d'autres caractéristiques pertinentes du candidat.

```
/usr/local/lib/python3.10/dist-packages/transformers/tokenization_utils_base.py:1601: FutureWarning: `clean_up_tokenization_spaces` was
warnings.warn(
Meilleures correspondances pour le candidat avec UID 'CpFGRJ9nfgDgdeGvL4hvsKSRF3':
{'Best Match Offer': 'Développeur-euse web', 'Best Match Offer ID': 'QBAlRkJLDobuJQuEyAGriu64Sgr2', 'Similarity Score': 0.6867}
{'Best Match Offer': 'Développeur-euse mobile', 'Best Match Offer ID': 'tRSEDpxgIxTqFQ1sv9gcdGPed0v2', 'Similarity Score': 0.1315}
```

FIGURE 5.8 – Meilleures Correspondances pour le Profil Développeur Web

1. Offre d'Emploi : Développeur-euse Web

- **UID du Offres d'Emploi** : QBAlRkJLDobuJQuEyAGriu64Sgr2
- **Score de Similarité** : 0.68
- **Détails du poste** : Cette offre requiert des compétences techniques en HTML/CSS, SQL, JavaScript, Git, Django, et Python, qui correspondent bien aux compétences du candidat. Le poste est à temps plein sous contrat CDD, avec des options de télétravail . La correspondance élevée entre les exigences du poste et les compétences du candidat est reflétée par un score de similarité de 0.68.

```
created_at: 2024-07-28 00:14:56.243000+00:00
remote_status: Oui
job_name: Développeur-euse web
job_video_description: False
ignored_jobs:
salary: 💰 Jusqu'à 20 000 €
technical_skills: ['HTML/CSS', 'SQL', 'JavaScript', 'Git', 'Django', 'Python']
job_time: ⌚ Temps plein
distance:
creator_name: hondas
contract_time:
experience: 🌟 Sans expérience
motivationLetter:
job_type: entreprise
image_url: https://firebasestorage.googleapis.com/v0/b/midlvl1.appspot.com/o/default-images%2F1.png?alt=media&token=e92ecd66-0d60-45ef-b369-2c2f5d1215e6
availability: 🟢 Immédiate
advantages: ['⌚ Possibilité de travailler à temps partiel', '🏠 Possibilité de travailler à domicile']
created_by: {'uid': 'QBAlRkJLDobuJQuEyAGriu64Sgr2', 'values': ['🗳️ Égalité des chances ', '👤 Droits de l'homme ', '♂️ Égalité des sexes ']}
activityDomains: ['💻 Informatique']
contract_types: ['📄 CDD']
study_level: 🎓 Master
```

FIGURE 5.9 – Détails du Poste de Développeur Web

2. Offre d'Emploi : Développeur-euse Mobile

- **UID du Offres d'Emploi** : tRSEDpxgIXqTqFQ1sv9gcGdPed0v2
- **Score de Similarité** : 0.131
- **Détails du poste** : Bien que le candidat ait des compétences techniques en Python, l'offre de Développeur Mobile requiert des compétences supplémentaires en Flutter et Flutterflow. Cette offre a été retenue en raison du nombre limité d'annonces disponibles et pourrait constituer une opportunité intéressante pour un développement professionnel futur.

Le rôle est à temps plein sous CDD. Bien que le candidat maîtrise Python, les autres technologies demandées ne sont pas totalement alignées avec son profil, ce qui explique un score de similarité de 0.131.

```
created_at: 2024-07-03 11:10:34.589000+00:00
remote_status: Indifférent
job_name: Développeur-euse mobile
job_video_description: False
ignored_jobs: [<google.cloud.firestore_v1.document.DocumentReference object at 0x0000021EB42F2DE0>, <google.cloud.firestore_v1.document.DocumentReference
salary: 💰 + de 70 000 €
technical_skills: ['flutter', 'python', 'flutterflow']
job_time: ⌚ Temps plein
distance:
creator_name: takwa
contract_time:
experience: 🌱 Sans expérience
motivationLetter:
job_type: entreprise
image_url: https://firebasestorage.googleapis.com/v0/b/midlvi.appspot.com/o/default-images%2F1.png?alt=media&token=e92ecd66-0d60-45ef-b369-2c2f5d1215e6
availability: 📅 Immédiate
advantages: ['❤️ Assurance vie', "🚫 Assurance perte d'emploi", "📈 Abondement à l'achat d'actions de l'entreprise"]
created_by: {'uid': 'tRSEDpxgIXqTqFQ1sv9gcGdPed0v2'} 'values': ['🚫 No Racism ', '🌱 Développement de compétences ', '💚 Égalité des chances ']]
activityDomains: ['🖥️ Informatique']
contract_types: ['📄 CDD']
```

FIGURE 5.10 – Détails du Poste de Développeur Mobile

Les résultats du matching ont montré que l'offre de Développeur-euse Web est la meilleure correspondance pour ce candidat, avec un score de similarité de 0.68, ce qui en fait une option prioritaire. Cette adéquation est principalement due à l'alignement des compétences techniques du candidat avec celles requises par l'offre. En revanche, bien que le score de similarité pour l'offre de Développeur-euse Mobile soit plus faible, cette offre peut néanmoins constituer une option intéressante pour un développement professionnel dans un autre domaine du développement logiciel.

5.3 Résultats des Tests avec des Données Simulées

En raison du lancement récent de la startup MidlJob et de la quantité limitée d'annonces disponibles dans la base de données, il a été nécessaire de recourir à des jeux de données externes, tels que ceux fournis par Kaggle, pour mener à bien les tests du système de matching. Cette approche a permis de simuler et d'évaluer l'efficacité du système dans des conditions variées, tout en compensant le manque de données réelles.

5.3.1 Prédiction des Meilleurs Candidats pour un Poste

Pour ce test, une offre de Développeur Web a été sélectionnée, présentant les caractéristiques suivantes :

ID de l'Offre : 398454096642776

Intitulé du Poste : Développeur Web

Description du Poste : Les développeurs web frontend conçoivent et mettent en œuvre des interfaces utilisateur pour les sites web, en veillant à ce qu'elles soient visuellement attrayantes et faciles à utiliser. Ils collaborent avec les designers et les développeurs backend pour offrir des expériences web fluides aux utilisateurs.

Compétences : HTML, CSS, JavaScript, frameworks frontend (par exemple, React, Angular), expérience utilisateur (UX)

Les résultats de la prédiction des meilleurs candidats pour ce poste sont les suivants :


```

{'Best Match Candidate': 'Jessica Moore', 'Best Match Candidate UID':
'69e4aac8-969a-4fc4-ac3d-99cce9e5fad5', 'Similarity Score':
0.3542061893041192}
{'Best Match Candidate': 'Wendy Thompson', 'Best Match Candidate UID':
'1c3becd7-ca7c-4f5f-8ad9-881868526db0', 'Similarity Score':
0.2921778797113233}
{'Best Match Candidate': 'Nathan Scott', 'Best Match Candidate UID':
'b45574b7-5493-4fb0-b423-dd2e1086fff6', 'Similarity Score':
0.19505516900277692}
{'Best Match Candidate': 'Mark Harris', 'Best Match Candidate UID':
'4c772bbb-bfa9-40e7-8545-8a3f265ac0ef', 'Similarity Score':
0.10713510291909226}
{'Best Match Candidate': 'William Johnson', 'Best Match Candidate UID':
'edd95e75-87a7-4ade-b365-flea959c81d4', 'Similarity Score':
0.10270563188196542}
{'Best Match Candidate': 'Sophia Martinez', 'Best Match Candidate UID':
'297a4788-df88-412a-8aac-085528b3cc78', 'Similarity Score':
0.10000708172634232}

```

FIGURE 5.11 – Meilleures Correspondances - Candidats pour le Poste de Web Developer

Quelques détails sur les trois meilleurs candidats sélectionnés pour ce poste :

Candidat : Jessica Moore

— **UID du Candidat :** 69e4aac8-969a-4fc4-ac3d-99cce9e5fad5

— **Score de Similarité :** 0.354

— **Description :** Jessica présente une bonne correspondance avec les compétences techniques requises pour le poste, notamment sa maîtrise de HTML, CSS, et JavaScript. Elle est identifiée comme la meilleure correspondante pour ce poste.

Candidat : Wendy Thompson

— **UID du Candidat :** 1c3becd7-ca7c-4f5f-8ad9-881868526db0

— **Score de Similarité :** 0.292

— **Description :** Wendy est une développeuse UI expérimentée avec des compétences en UI design et développement frontend. Bien que son expertise soit principalement orientée vers le design d'interface utilisateur, ses compétences en HTML, CSS, et JavaScript font d'elle une correspondance adéquate pour ce poste.

- **Salaire** : \$75,000
- **Langues** : Anglais, Chinois, Français
- **Niveau d'Études** : Bachelor's Degree
- **Type de Contrat** : CDI
- **Disponibilité** : 1 mois
- **Statut** : Remote

Candidat : Nathan Scott

- **UID du Candidat** : b45574b7-5493-4fb0-b423-dd2e1086fff6
- **Score de Similarité** : 0.195
- **Description** : Nathan possède des compétences de base en JavaScript et HTML, mais il manque d'expérience avec des frameworks frontend comme React ou Angular, ce qui se reflète dans un score de correspondance plus faible.

Parmi les candidats évalués pour le poste de Développeur Web, Jessica Moore se distingue comme la meilleure correspondante avec un score de 0.354, suivie de Wendy Thompson avec un score de 0.292. Jessica répond mieux aux exigences techniques du poste, tandis que Wendy, bien que spécialisée en design UI, reste une candidate intéressante en raison de ses compétences en développement frontend.

5.3.2 Prédiction des Meilleurs Candidats pour un Poste

Pour cette section, nous avons utilisé le profil d'Anna Thompson, une infirmière expérimentée, pour prédire les meilleures offres d'emploi correspondant à ses compétences et à son expérience. Dans la version premium, en plus de ses compétences techniques, le système évalue également ses soft skills, sa motivation et ses valeurs culturelles. Voici comment ces éléments influencent le matching dans la version premium.

Détails du Candidat :

- **UID du Candidat** : f7b47279-7767-4e93-8fa1-8a869f66a9c9
- **Intitulé du Poste** : Nurse (Infirmière)
- **Nom du Candidat** : Anna Thompson
- **Compétences Techniques** : Soins aux patients, Administration de médicaments, Thérapie intraveineuse (IV)
- **Compétences Interpersonnelles (Soft Skills)** : Compassion, Attention aux détails, Capacité à gérer des situations de stress
- **Description du Profil** : Infirmière expérimentée avec une solide expertise en soins aux patients et administration de médicaments.
- **Motivation** : Anna est motivée par la possibilité de travailler dans un environnement où elle peut continuer à améliorer ses compétences en soins infirmiers tout en apportant une contribution directe au bien-être des patients.
- **Valeurs Culturelles** : Anna adhère aux principes d'égalité d'accès aux soins et de respect des patients, des valeurs qui sont essentielles dans les environnements médicaux.
- **Salaire** : \$60,000
- **Langues** : Anglais, Français, Espagnol
- **Niveau d'Études** : Bachelor's Degree
- **Types de Contrats** : Temps plein
- **Statut** : Sur site
- **Expérience** : 4 ans
- **Durée du Contrat** : CDI
- **Disponibilité** : Immédiate

Les meilleurs postes correspondant aux compétences et au profil d'Anna Thompson sont les suivants :

1. Nurse Practitioner

- **ID de l'Offre :** 2006777111523978
- **Score de Similarité :** 0.344
- **Description :** Ce poste exige des compétences avancées en soins aux patients et en administration médicale, ce qui correspond bien aux compétences d'Anna. Le score de similarité montre une forte correspondance avec les exigences de l'offre.

2. Registered Nurse

- **ID de l'Offre :** 830586177954485
- **Score de Similarité :** 0.335
- **Description :** Ce poste de Registered Nurse requiert des compétences en soins infirmiers généralistes. Bien qu'il soit légèrement en dessous des exigences du poste de Nurse Practitioner, il reste une bonne option pour Anna.

3. Nurse Manager

- **ID de l'Offre :** 1053345514869033
- **Score de Similarité :** 0.316
- **Description :** Ce poste de gestionnaire requiert des compétences en supervision et gestion d'équipe. Bien que cela dépasse l'expérience actuelle d'Anna, ce poste pourrait être une opportunité de progression de carrière.

Les résultats de la prédiction montrent qu'Anna Thompson correspond particulièrement bien aux postes de Nurse Practitioner et de Registered Nurse, avec des scores de similarité respectifs de 0.344 et 0.335. Le poste de Nurse Manager, bien qu'un peu moins aligné, pourrait constituer une option intéressante pour son développement professionnel futur.

```

{'Best Match Offer': 'Nurse Practitioner', 'Best Match Offer ID': '2006777111523978', 'Similarity Score':
0.3444206571571074}
{'Best Match Offer': 'Nurse Practitioner', 'Best Match Offer ID': '2382673250573351', 'Similarity Score':
0.3444206571571074}
{'Best Match Offer': 'Nurse Practitioner', 'Best Match Offer ID': '90206306057788', 'Similarity Score':
0.3444206571571074}
{'Best Match Offer': 'Registered Nurse', 'Best Match Offer ID': '830586177954485', 'Similarity Score':
0.33509968821682334}
{'Best Match Offer': 'Registered Nurse', 'Best Match Offer ID': '2466750721870912', 'Similarity Score':
0.33509968821682334}
{'Best Match Offer': 'Registered Nurse', 'Best Match Offer ID': '2828247805840851', 'Similarity Score':
0.33509968821682334}
{'Best Match Offer': 'Registered Nurse', 'Best Match Offer ID': '1277230346836850', 'Similarity Score':
0.33509968821682334}
{'Best Match Offer': 'Registered Nurse', 'Best Match Offer ID': '2051493722156139', 'Similarity Score':
0.33509968821682334}
{'Best Match Offer': 'Registered Nurse', 'Best Match Offer ID': '1385010604943968', 'Similarity Score':
0.31833345920905187}
{'Best Match Offer': 'Registered Nurse', 'Best Match Offer ID': '2307153516206719', 'Similarity Score':
0.31833345920905187}
{'Best Match Offer': 'Nurse Manager', 'Best Match Offer ID': '1053345514869033', 'Similarity Score':
0.3165063160500517}
{'Best Match Offer': 'Nurse Manager', 'Best Match Offer ID': '2073220652563360', 'Similarity Score':
0.28628328974326467}
{'Best Match Offer': 'Nurse Manager', 'Best Match Offer ID': '3014747272177137', 'Similarity Score':
0.28628328974326467}
{'Best Match Offer': 'Nurse Practitioner', 'Best Match Offer ID': '2040957541238639', 'Similarity Score':
0.28101332297473813}
{'Best Match Offer': 'Nurse Practitioner', 'Best Match Offer ID': '441473044734741', 'Similarity Score':
0.28101332297473813}
{'Best Match Offer': 'Nurse Practitioner', 'Best Match Offer ID': '1252829260379029', 'Similarity Score':
0.28101332297473813}
{'Best Match Offer': 'Nurse Practitioner', 'Best Match Offer ID': '832490435008360', 'Similarity Score':
0.28101332297473813}
{'Best Match Offer': 'Nurse Practitioner', 'Best Match Offer ID': '1524311163188117', 'Similarity Score':
0.28101332297473813}
{'Best Match Offer': 'Family Nurse Practitioner', 'Best Match Offer ID': '3038257032555306', 'Similarity Score':
0.23507419084231707}
{'Best Match Offer': 'Nurse Practitioner', 'Best Match Offer ID': '1156025606824277', 'Similarity Score':
0.2114107874725221}
{'Best Match Offer': 'Nurse Practitioner', 'Best Match Offer ID': '2073782399665928', 'Similarity Score':
0.2114107874725221}
{'Best Match Offer': 'Registered Nurse', 'Best Match Offer ID': '1864996850784532', 'Similarity Score':
0.1892693614520577}
{'Best Match Offer': 'Registered Nurse', 'Best Match Offer ID': '2576700225981224', 'Similarity Score':
0.1892693614520577}
{'Best Match Offer': 'Nurse Manager', 'Best Match Offer ID': '1783038709502119', 'Similarity Score':
0.188066935809168}

```

FIGURE 5.12 – Meilleures Correspondances - Offres d'Emploi pour Anna Thompson (Infirmière)

5.4 Résultats du Matching dans la Version Premium

Dans cette version, en intégrant les soft skills, la motivation ainsi que les valeurs culturelles d'Anna dans le processus de matching, les résultats révèlent une différence significative par rapport à la version gratuite. Voici les correspondances les plus pertinentes pour Anna, obtenues grâce au système premium :

1. Nurse Practitioner

— **Score de Similarité** : 0.65

— **Description** : Ce poste exige des compétences avancées en soins aux patients, en ligne avec l'expérience d'Anna. En plus de cela, l'environnement hospitalier dans lequel elle

travaillerait valorise l'humanité dans les soins, ce qui correspond à ses motivations personnelles. L'accent est également mis sur l'égalité d'accès aux soins, une valeur importante pour Anna. De plus, la gestion du stress et la compassion sont des compétences clés pour ce rôle, des qualités qu'Anna possède, ce qui explique un score de similarité plus élevé par rapport à la version gratuite.

2. Registered Nurse

— **Score de Similarité : 0.67**

— **Description :** Ce poste correspond bien à l'expérience technique d'Anna. Ses compétences sont en alignement avec les exigences techniques du poste. Cependant, la culture d'entreprise met moins l'accent sur des valeurs telles que le respect et l'égalité d'accès aux soins, ce qui contribue à un score de similarité légèrement inférieur. De plus, bien que la compassion et l'attention aux détails soient valorisées, l'importance accordée à la gestion du stress est moins prononcée dans ce rôle, ce qui impacte également le score de correspondance.

3. Nurse Manager

— **Score de Similarité : 0.56**

— **Description :** Ce poste offre une opportunité de progression pour Anna, mais exige des compétences en gestion d'équipe, qu'elle n'a pas encore entièrement développées. Toutefois, le rôle est bien aligné avec son profil en termes de valeurs professionnelles, et l'approche de leadership au sein de l'organisation correspond à ses attentes. Malgré cela, le manque d'expérience en gestion justifie un score de similarité plus bas.

La version premium de MidlJob permet aux recruteurs de trouver non seulement des candidats techniquement qualifiés, mais également des candidats qui sont alignés avec la culture et les valeurs de l'entreprise. Cela favorise des relations de travail plus harmonieuses et une intégration réussie des nouveaux employés dans l'équipe.

5.5 Scénario de Présentation des Profils dans l'Application MidlJob

Après la conception et le développement des API de matching du système MidlJob, ces fonctionnalités ont été intégrées avec succès dans l'application en collaboration avec l'équipe de développement Flutter. Cette intégration offre désormais une présentation visuelle claire, intuitive et conviviale des profils de candidats et des offres d'emploi, optimisant ainsi le processus de recrutement pour les utilisateurs.

Grâce à cette interface améliorée, les recruteurs et les candidats peuvent accéder rapidement aux informations clés telles que les compétences techniques, les expériences, les valeurs et les soft skills. De plus, le système de matching met en avant les résultats les plus pertinents, facilitant ainsi la prise de décision pour les recruteurs et assurant une meilleure correspondance entre les profils et les besoins des entreprises.

5.5.1 Visualisation des Profils des Candidats

Les profils des candidats sont présentés de manière claire et concise, comme illustré dans Fig. 5.13. Ces interfaces permettent aux recruteurs de voir immédiatement les informations essentielles du candidat, telles que :

- **Compétences techniques**
- **Langues parlées**
- **Type de contrat souhaité**
- **Disponibilité**
- **Valeurs , soft skills et plus**

Le recruteur a également la possibilité de liker un profil ou de l'ajouter aux favoris pour un examen ultérieur. Cette fonctionnalité optimise le processus de sélection en mettant en avant les correspondances les plus pertinentes.

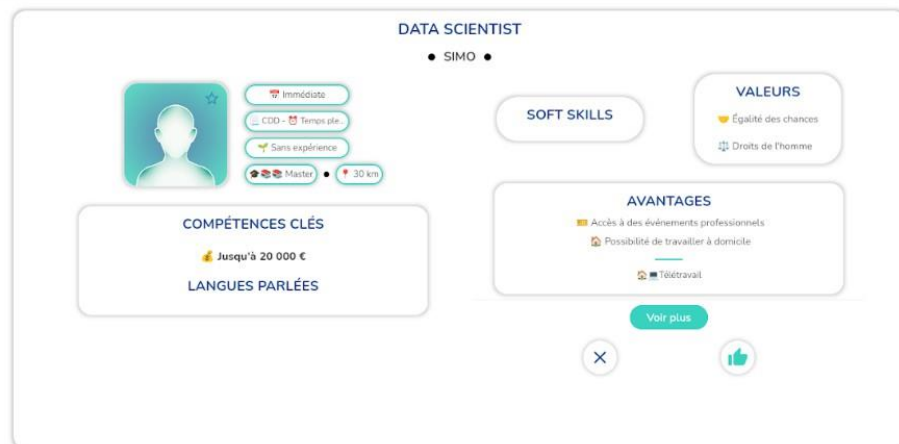


FIGURE 5.13 – Présentation du Profil du Candidat Data Scientist

Dans Fig. 5.14, on peut observer les profils que le recruteur a déjà likés, indiquant ainsi les candidats les plus pertinents pour le poste. Cette visualisation permet une meilleure gestion et organisation des profils, facilitant le suivi des candidatures les plus intéressantes pour le recruteur.

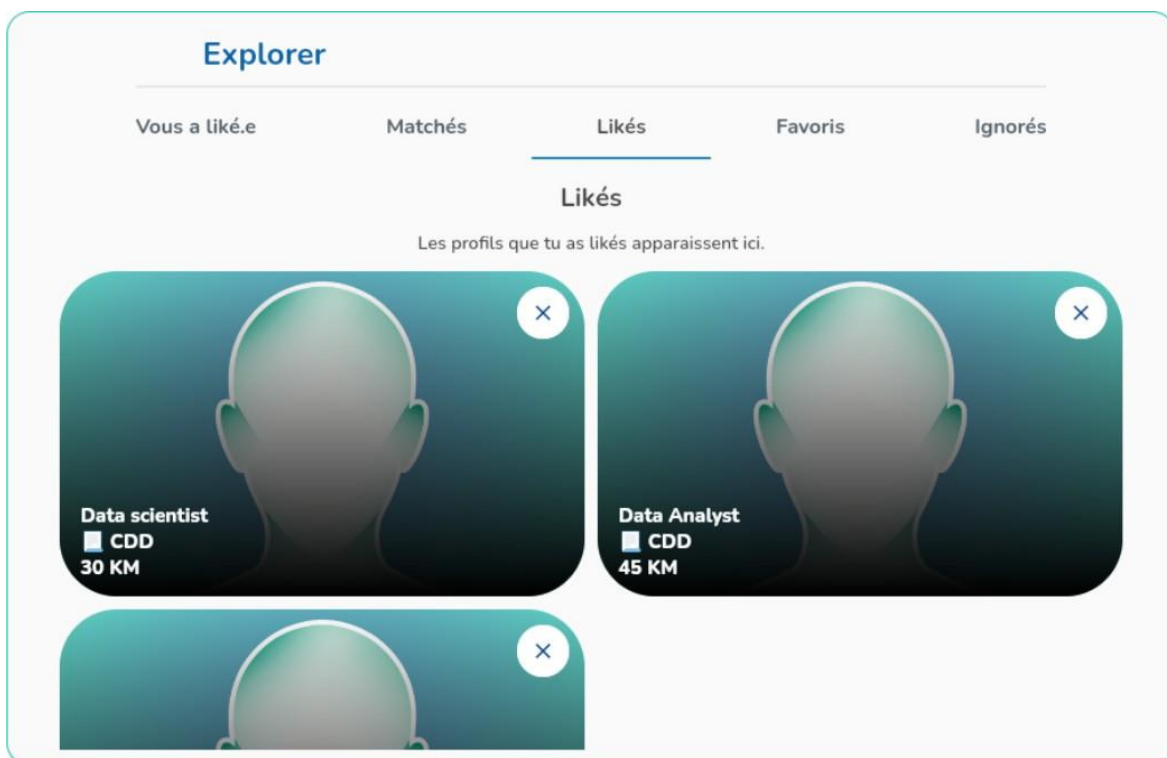


FIGURE 5.14 – Liste des Profils Likés par le Recruteur dans l'Application MidlJob

5.5.2 Notification de Correspondance ("It's a Match !")

Lorsque le recruteur et le candidat expriment un intérêt mutuel pour un poste, une notification "It's a Match !" apparaît, comme le montre Fig. 5.15. Ce système de correspondance permet d'accélérer le processus de mise en relation en créant un lien direct entre l'employeur et le candidat.

Une fois cette correspondance effectuée, les recruteurs peuvent immédiatement commencer à échanger avec le candidat via une interface de messagerie intégrée.

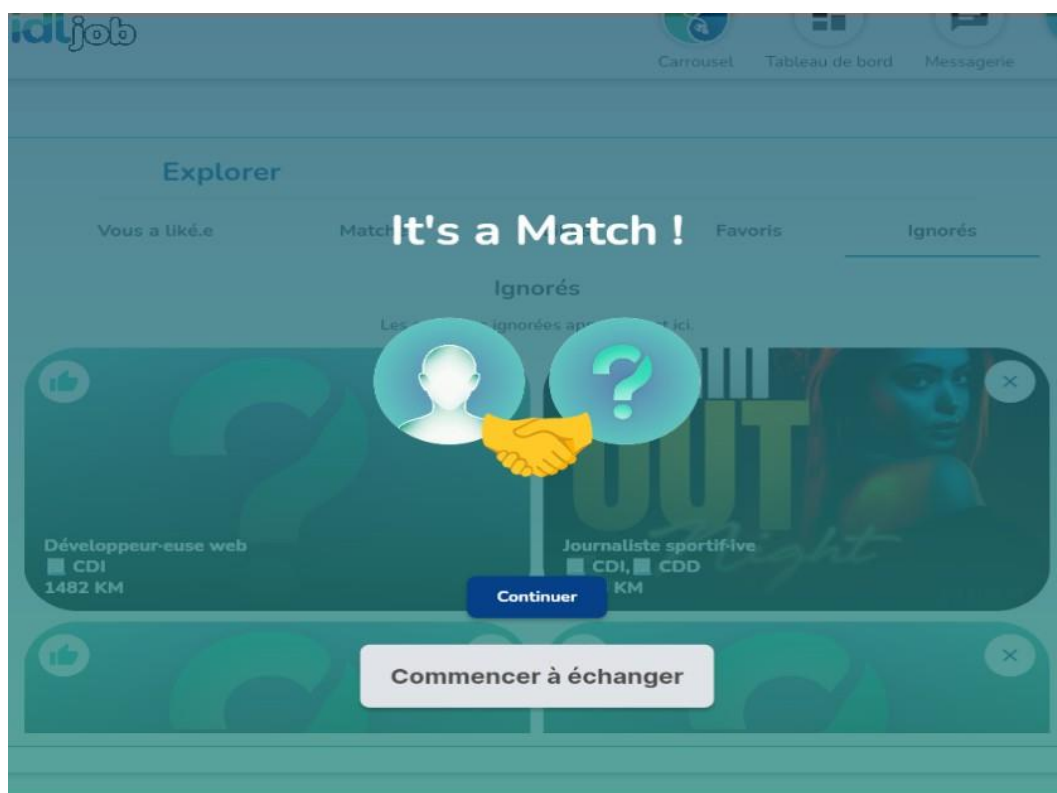


FIGURE 5.15 – Notification de Correspondance Réussie dans la Plateforme MidlJob

5.5.3 Interface de Messagerie

L'application propose également une interface de messagerie simplifiée (Fig. 5.16), permettant aux recruteurs de prendre contact directement avec les candidats une fois la correspondance établie. Cette messagerie instantanée rend les échanges plus fluides et favorise une interaction rapide, ce qui est essentiel pour les processus de recrutement modernes.

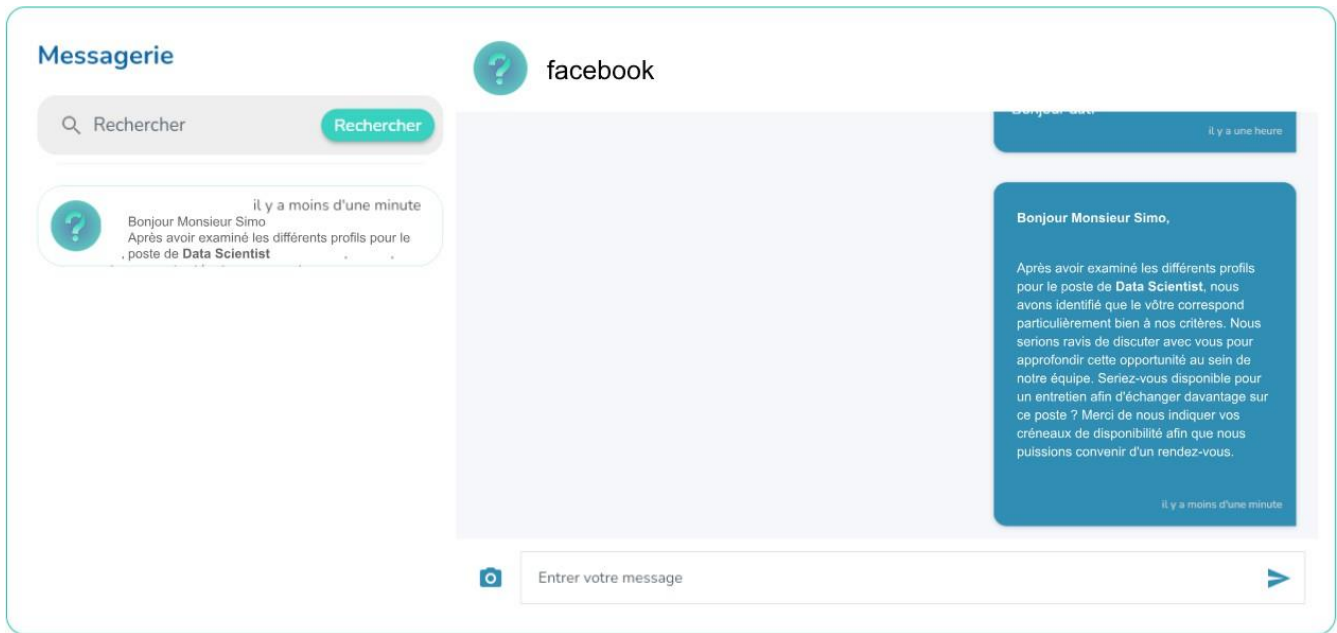


FIGURE 5.16 – Interface de Messagerie pour le Matching des Candidats

5.6 Conclusion

Les résultats obtenus au terme des tests menés dans le cadre du projet MidlJob démontrent la robustesse du système de matching dans les deux versions proposées. En effet, l'utilisation d'algorithmes tels que la vectorisation TF-IDF et le clustering K-Means a permis de fournir des recommandations pertinentes pour les recruteurs et les candidats. Cependant, des limitations subsistent, notamment en ce qui concerne le nombre d'annonces et de profils disponibles dans la base de données initiale, ainsi que la prise en compte de critères plus subjectifs dans la version gratuite.

L'intégration des soft skills et des valeurs culturelles dans la version premium offre une vision plus approfondie des correspondances, ouvrant la voie à des possibilités d'amélioration continue dans le domaine du recrutement digital. Dans l'ensemble, MidlJob se révèle être un outil prometteur pour faciliter les processus de recrutement, mais des ajustements seront nécessaires pour optimiser ses performances et répondre pleinement aux attentes des utilisateurs.

Perspectives pour la Version Pro

L'évolution de **MidlJob** vers une **version Pro** s'inscrit dans la volonté d'offrir une solution plus complète et performante, particulièrement adaptée aux grandes entreprises et aux organisations disposant de processus de recrutement complexes. Parmi les améliorations prévues pour cette version, on peut citer les fonctionnalités suivantes :

Génération de questions d'entretien personnalisées : Cette fonctionnalité repose sur l'utilisation de modèles de traitement du langage naturel (NLP) pour générer automatiquement des questions adaptées aux profils des candidats, en fonction des critères spécifiques des offres d'emploi. Cela permet aux recruteurs de bénéficier d'un outil d'évaluation plus complet et sur mesure.

Transcription et analyse des réponses vidéo : Les réponses vidéo des candidats pourront être automatiquement transcrites en texte pour une analyse plus précise. Cette fonctionnalité permettra d'évaluer non seulement les compétences techniques mais aussi la capacité des candidats à communiquer efficacement.

Développement d'un chatbot intégré : Un chatbot sera mis en place pour faciliter les échanges en temps réel entre recruteurs et candidats, permettant ainsi de fluidifier le processus de recrutement tout en assurant une réactivité immédiate.

État Actuel du Développement

Actuellement, je travaille sur la première tâche de la version Pro : la génération de questions d'entretien personnalisées. Bien que j'aie envisagé l'utilisation d'API telles que ChatGPT, des préoccupations concernant la confidentialité des données des candidats et des recruteurs m'ont

conduit à explorer des alternatives. Je prévois d'utiliser Llama 3, un grand modèle de langage (LLM), via Ollama, une plateforme permettant l'exécution locale de modèles de langage, que j'installerai sur les serveurs de l'entreprise pour garantir la sécurité des données tout en maintenant un haut niveau de performance analytique.

Ces perspectives montrent clairement que MidlJob, dans sa version Pro, s'inscrit dans une démarche de perfectionnement continu, en phase avec les besoins croissants des recruteurs modernes. En combinant innovation technologique et respect des impératifs de sécurité, MidlJob vise à devenir un acteur incontournable du recrutement intelligent, capable de s'adapter à des environnements toujours plus exigeants et dynamiques.

Conclusion générale

Le projet MidlJob met en lumière le potentiel transformateur de l'intelligence artificielle (IA) dans le domaine du recrutement. En automatisant les tâches répétitives et en optimisant le processus de correspondance entre candidats et recruteurs, cette solution propose une approche innovante pour surmonter les défis majeurs du recrutement, tels que la difficulté à identifier rapidement les meilleurs profils et à garantir une adéquation précise avec les besoins des entreprises.

L'intégration de technologies avancées comme la vectorisation TF-IDF, l'algorithme K-Means et la mesure de similarité cosinus permet à MidlJob de générer des correspondances précises. Cette solution ne se limite pas aux compétences techniques mais prend également en compte les affinités culturelles et comportementales, particulièrement dans sa version Premium. Les résultats des tests montrent des améliorations significatives en termes de gain de temps et de précision, profitant à la fois aux candidats et aux recruteurs.

Cependant, il existe encore des axes d'amélioration, notamment en ce qui concerne l'enrichissement des données disponibles et l'ajout de nouvelles fonctionnalités pour la version Pro. L'intégration de fonctionnalités telles que la génération automatique de questions d'entretien ou l'analyse des réponses vidéo renforcera davantage l'efficacité et l'adaptabilité de MidlJob, répondant ainsi aux exigences toujours plus complexes des entreprises contemporaines. En définitive, ce projet illustre comment l'IA peut profondément améliorer les processus de recrutement, les rendant plus objectifs, rapides et personnalisés.

Bibliographie

Faliagka, E., Tsakalidis, A., Tzimas, G. (2012). An Integrated E-Recruitment System for Automated Personality Mining and Applicant Ranking. [en ligne]. Disponible sur : <https://doi.org/10.1108/10662241211271545>.

Berkelaar, A., Pisaruk, M. (2018). The Use of Artificial Intelligence in Candidate Screening and Job Matching. [en ligne]. Disponible sur : <https://doi.org/10.1108/HRMID-04-2018-0051>.

Jovanovic, J., Joblon, D. (2020). How AI and Machine Learning Are Changing the Recruitment Process. [en ligne]. Disponible sur : <https://www.forbes.com/sites/forbestechcouncil/2020/09/02/how-ai-and-machine-learning-are-changing-the-recruitment-process/>.

Manning, C. D., Raghavan, P., Schütze, H. (2008). Introduction to Information Retrieval. [en ligne]. Disponible sur : <https://nlp.stanford.edu/IR-book/>.

Gupta, N., Gupta, M. (2019). AI in HR : Human Resource Analytics and Predictive Talent Acquisition. [en ligne]. Disponible sur : <https://www.ijitee.org/wp-content/uploads/papers/v8i6/F3454048619.pdf>.

Ramírez, S. (2021). FastAPI : Modern, Fast (High-Performance), Web Framework for Building APIs with Python 3.6+ Based on Standard Python Type Hints. [en ligne]. Disponible sur : <https://fastapi.tiangolo.com>.

Hugging Face (2021). paraphrase-MiniLM-L6-v2 Model Card. [en ligne]. Disponible sur : <https://huggingface.co/sentence-transformers/paraphrase-MiniLM-L6-v2>.

Résumé

Ce rapport présente la conception et le développement d'un **Système de Recrutement par l'Intelligence Artificielle**. Le système utilise des techniques avancées de **machine learning** et de **traitement du langage naturel (NLP)**, telles que la vectorisation **TF-IDF**, les **embeddings sémantiques**, et les algorithmes de **similarité cosinus** et **K-Means**, pour améliorer le matching entre les candidats et les recruteurs. La version **Free** prend en compte les **caractéristiques de base** des candidats, comme les compétences techniques, le titre du poste et l'expérience. En revanche, la version **Premium** enrichit l'expérience en ajoutant la **comparaison des personnalités** et des **compétences comportementales** (soft skills), offrant ainsi un recrutement plus précis et personnalisé. Ces fonctionnalités permettent d'aligner non seulement les compétences techniques mais aussi les affinités culturelles et comportementales entre les candidats et les recruteurs.

Abstract

This report outlines the design and development of an **Artificial Intelligence-Based Recruitment System**. The system leverages advanced **machine learning** and **natural language processing (NLP)** techniques, including **TF-IDF vectorization**, **semantic embeddings**, and **cosine similarity** and **K-Means algorithms**, to enhance matching between candidates and recruiters. The **Free version** considers **basic characteristics** such as technical skills, job title, and experience, while the **Premium version** adds **personality comparison** and **behavioral skills (soft skills)**, making the recruitment process more accurate and personalized. These features enable not only a match based on technical skills but also on **cultural and behavioral fit** between candidates and recruiters.