# Les Offres d’Emploi en Ligne, Nouvelle Source de Données sur le Marché du Travail : Illustration sur l’Année 2024

**Doumbia Abdoulaye, Bakayoko Abdoulaye,**

## Résumé

Avec l'évolution rapide d'Internet comme principal canal de diffusion des offres d'emploi, la collecte de données en ligne s'est imposée comme une méthode novatrice pour analyser le marché du travail. En 2024, notre étude a utilisé des techniques de web scraping pour récolter des données sur les offres d'emploi en Côte d'Ivoire, sous l'égide de l'Institut National de la Statistique (INS). L'objectif était de créer une base de données complète et exploitable. Cette initiative permet d'accéder à des informations dynamiques et actuelles, cruciales pour comprendre les tendances du marché et identifier les compétences recherchées. Grâce à des outils de traitement du langage naturel, nous avons analysé ces données, révélant des disparités dans la représentation des métiers en ligne. Cette base de données offre un aperçu précieux pour les chercheurs d'emploi, les recruteurs et les décideurs politiques, en complément des sources traditionnelles de données sur le marché du travail.

**Mots-clés :** offres d’emploi en ligne, web scraping, Big Data, NLP, Côte d'Ivoire, marché du travail, analyse de données

Table des matières

[Les Offres d’Emploi en Ligne, Nouvelle Source de Données sur le Marché du Travail : Illustration sur l’Année 2024 1](#_Toc170821545)

[Résumé 1](#_Toc170821546)

[Introduction 3](#_Toc170821547)

[I. Émergence des Offres d’Emploi en Ligne comme Source de Données 4](#_Toc170821548)

[1.1 L'Essor des Offres d’Emploi en Ligne en Côte d'Ivoire 4](#_Toc170821549)

[1.2 Comparaison avec les Sources Traditionnelles 4](#_Toc170821550)

[1.3 Internet comme Nouvelle Source de Données 6](#_Toc170821551)

[1.4 Champs d’Étude et Représentativité 6](#_Toc170821552)

[II. Production de Données Statistiques depuis les Offres en Ligne 7](#_Toc170821553)

[2.1 Collecte des Données : Approche par Web Scraping 7](#_Toc170821554)

[2.2 Transformation des Données Brutes en Données Statistiques 9](#_Toc170821555)

[3.2 Couverture et Représentativité des Données 10](#_Toc170821556)

[III. Caractéristiques de la Base d’Offres en Ligne 2024 11](#_Toc170821557)

[3.1 Composition et Temporalité des Données 11](#_Toc170821558)

[3.2 Couverture et Représentativité des Données 11](#_Toc170821559)

[3.3 Utilisations de la Base de Données 11](#_Toc170821560)

[Conclusion 12](#_Toc170821561)

[Bibliographie 12](#_Toc170821562)

[Annexes 12](#_Toc170821563)

[Remerciements 12](#_Toc170821564)

## Introduction

L'Internet est devenu une plateforme incontournable pour la diffusion des offres d'emploi. En Côte d'Ivoire, l'utilisation croissante des annonces en ligne a transformé le paysage du recrutement et offre une nouvelle opportunité pour l'analyse du marché du travail. Ce document explore la collecte et l'analyse des offres d'emploi en ligne publiées sur plusieurs sites ivoiriens en 2024, effectuées par l'Institut National de la Statistique (INS). L'objectif est d'examiner comment ces données peuvent compléter les sources traditionnelles et fournir des informations précieuses sur les tendances du marché.

## I. Émergence des Offres d’Emploi en Ligne comme Source de Données

### 1.1 L'Essor des Offres d’Emploi en Ligne en Côte d'Ivoire

Le nombre d'offres d'emploi publiées en ligne a connu une croissance exponentielle en Côte d'Ivoire. Cette tendance reflète un changement dans les pratiques de recrutement, passant des canaux traditionnels aux plateformes numériques. La collecte d'offres d'emploi via le web scraping, réalisée par l'INS, est devenue une méthode clé pour obtenir des données actualisées et représentatives des opportunités professionnelles.

### 1.2 Comparaison avec les Sources Traditionnelles

Les sources traditionnelles de données sur le marché du travail incluent les enquêtes menées par des institutions comme l'INS et les annonces publiées par des agences de recrutement. Cependant, ces méthodes présentent souvent des limites en termes de fréquence et de couverture. En comparaison, les offres d'emploi en ligne offrent une vue d'ensemble plus dynamique et immédiate du marché.

#### **Étapes de la Comparaison avec les Sources Traditionnelles :**

1. **Définir les Objectifs de l'Étude :**
   * **Objectif Principal :** Évaluer la complémentarité et les divergences entre les offres d'emploi en ligne et les données des sources traditionnelles.
   * **Objectifs Spécifiques :** Couverture géographique, diversité sectorielle, actualité et réactivité, qualité des données.
2. **Unité Internet :**
   * Les sites d’annonces d’emploi tels que :



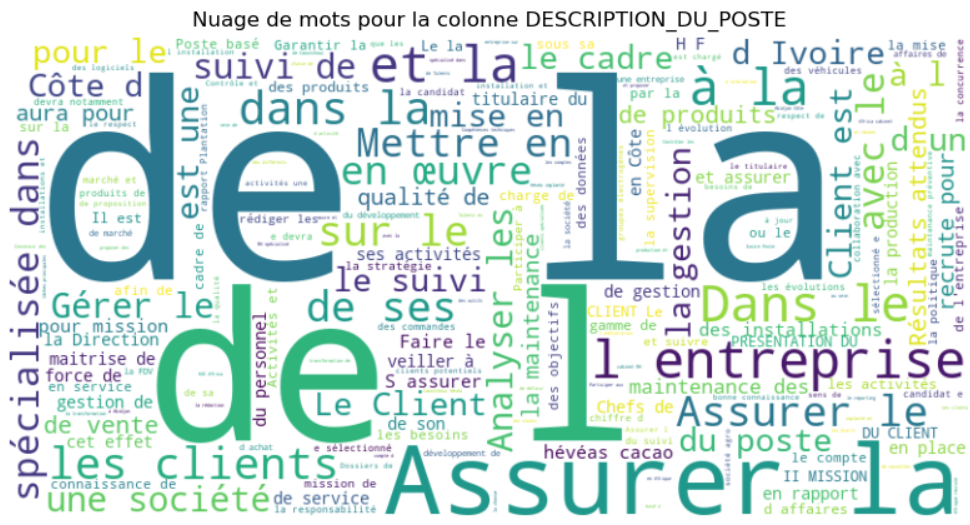
1. **Questions / Problématiques :**
   * **Problématique Générale :** Comment les offres d’emploi en ligne se complètent-elles ou se distinguent-elles des données des sources traditionnelles ?
   * **Questions Spécifiques :** Différences géographiques, représentation sectorielle, actualité des données, représentativité, précision des données.
2. **Variables et Indicateurs :**
   * **Couverture Géographique :** Région, ville.



* + **Diversité Sectorielle :** Secteur d’activité (ex. technologie, santé).



* + **Actualité des Données :** Date de publication, mise à jour.
  + **Représentativité :** Catégorie professionnelle, niveau d’expérience.
  + **Précision des Données :** Détail des qualifications, descriptif du poste.



1. **Phase Pilote :**
   * **Objectifs de la Phase Pilote :** Tester la méthodologie de collecte des données et identifier les problèmes potentiels.
   * **Méthodologie de la Phase Pilote :** Sélection des sites, tests de web scraping, évaluation des données, comparaison préliminaire, ajustements.



1. **Collecte Principale :**
   * **Objectifs de la Collecte Principale :** Collecter des données exhaustives et représentatives sur les offres d’emploi en ligne.
   * **Méthodologie de Collecte :** Élargissement de la collecte, automatisation du web scraping, vérification des données, intégration des données, comparaison systématique.
2. **Bulgelisation :**
   * **Objectif de la Bulgelisation :** Harmoniser et comparer les données issues de différentes sources.
   * **Processus de Bulgelisation :** Standardisation des formats, harmonisation des variables, normalisation des nomenclatures, intégration des données, analyse des écarts.

### 1.3 Internet comme Nouvelle Source de Données

#### a. Une Source Diversifiée

Internet fournit une source de données diversifiée pour les offres d'emploi, incluant des sites spécialisés, des réseaux sociaux professionnels et des portails d'emploi. Cette diversité permet de capter une large gamme d'offres couvrant différents secteurs et niveaux de qualification.

#### b. Potentiel et Limites des Données en Ligne

Les données en ligne offrent un potentiel immense pour l'analyse du marché du travail, permettant une mise à jour en temps réel et une couverture étendue. Toutefois, elles posent des défis, notamment en ce qui concerne la déduplication des offres et la garantie de la représentativité des données. Une attention particulière est nécessaire pour éviter les biais et s'assurer que l'analyse ne sous-représente pas certains secteurs ou types d'emplois.

### 1.4 Champs d’Étude et Représentativité

Les offres d'emploi en ligne constituent un champ d'étude riche, mais il est crucial d'assurer leur représentativité par rapport à l'ensemble du marché du travail. Certaines professions, notamment celles impliquant un recrutement informel ou ayant peu de visibilité en ligne, peuvent être sous-représentées. L'analyse doit donc prendre en compte ces disparités pour offrir une vue complète et précise du marché.

## II. Production de Données Statistiques depuis les Offres en Ligne

### 2.1 Collecte des Données : Approche par Web Scraping

#### a. Sélection des Sites Sources

La sélection des sites sources est cruciale pour la qualité des données collectées. Les plateformes choisies pour ce projet incluent des sites d'emploi populaires en Côte d'Ivoire tels que emploi.educarriere.ci, novojob.ci, et emploi.ci, parmi d'autres. Ces sites offrent une large couverture des offres d'emploi disponibles dans le pays.

#### b. Web Scraping : Méthodologie et Alternatives

Le web scraping est une méthode efficace pour collecter des données en ligne, permettant d'automatiser l'extraction des informations des pages web, offrant ainsi une solution rapide et scalable pour la collecte de données. Cependant, d'autres méthodes, comme l'utilisation d'API des sites d'emploi, peuvent également être envisagées en fonction de la disponibilité et des restrictions des plateformes.

#### c. Mise en Pratique du Web Scraping

En pratique, le web scraping nécessite des compétences techniques pour écrire et maintenir les scripts de collecte de données. Il est important de respecter les conditions d'utilisation des sites sources et d'implémenter des mesures pour éviter les blocages et garantir la qualité des données collectées.

* Téléchargement et Installation des Outils :
* Téléchargement des Outils : Les outils nécessaires pour la collecte de données, tels que les bibliothèques de web scraping (par exemple, BeautifulSoup, Scrapy) et les environnements de développement (comme Anaconda pour Python), ont été identifiés et téléchargés.
* Installation et Configuration : Les outils ont été installés et configurés sur les systèmes locaux, en s'assurant de leur compatibilité avec les environnements existants

**Étapes du téléchargement et Installation des Outils :**

1. **Téléchargement des Outils :** Les outils nécessaires pour la collecte de données, tels que les bibliothèques de web scraping (par exemple, BeautifulSoup, Scrapy) et les environnements de développement (comme Anaconda pour Python), ont été identifiés et téléchargés.
2. **Installation des Outils :** Les outils téléchargés ont été installés sur les systèmes de travail. Cette phase a parfois rencontré des problèmes liés à la connectivité et à la compatibilité des logiciels avec le matériel existant.

**Étapes de la collection et Traitement des Données :**

1. **Collection des Données :** Les données ont été collectées à partir des sites web ciblés. Les scripts de scraping ont été exécutés pour extraire les informations nécessaires des offres d'emploi.
2. **Traitement des Données :** Les données brutes ont été nettoyées et organisées. Cela comprenait la suppression des doublons, la correction des erreurs, et la normalisation des données pour garantir leur qualité et leur fiabilité.

* Processus d'Extraction des Données :
* Script de Web Scraping : Un script de web scraping a été développé pour extraire automatiquement les données des sites d'emploi sélectionnés. Ce script parcourt les pages web, identifie les éléments d'intérêt (par exemple, titre de l'emploi, description, localisation), et extrait les informations pertinentes.
* Manipulation des Données : Les données extraites sont ensuite manipulées pour être nettoyées et formatées, afin d'être prêtes pour l'analyse ultérieure. Cela inclut la gestion des doublons, la normalisation des formats, et l'intégration des données dans une base structurée.

**Étapes de la réalisation des Tests et Défis Techniques**

1. **Tests des Outils :** Les outils installés ont été testés pour vérifier leur bon fonctionnement et leur capacité à scraper les données efficacement.
2. **Problèmes de Connectivité :** Des défis liés à la connectivité ont parfois ralenti les tests, notamment en raison de la lenteur des réseaux ou des restrictions d'accès aux sites web ciblés.
3. **Défis Techniques :** Certains sites web utilisent des mesures anti-scraping, rendant la collecte des données plus complexe. Des solutions techniques ont été mises en place pour contourner ces obstacles, comme l'utilisation de techniques de rotation d'IP et de gestion des cookies.

### 2.2 Transformation des Données Brutes en Données Statistiques

#### a. Harmonisation des Données

L'harmonisation des données est essentielle pour garantir leur cohérence et leur comparabilité. Cela implique la normalisation des formats et la standardisation des champs, tels que les intitulés de postes et les localisations géographiques.

**Étapes de la** **Mise en Place des Syntaxes:**

1. **Mise en Place des Syntaxes :** Des syntaxes spécifiques ont été développées pour extraire les informations nécessaires à partir des pages web. Cela inclut la configuration de scripts pour collecter des données sur des offres d'emploi.
2. **Mise en Place des Indicateurs :** Des indicateurs clés ont été définis pour structurer les données collectées. Cela permet une analyse plus approfondie et pertinente des données.

**Étapes de l’Analyse Textuelle:**

1. **Analyse Textuelle :** Les données textuelles collectées ont été analysées pour identifier les tendances et les patterns significatifs. Cela inclut l'extraction de mots-clés et l'identification des compétences recherchées.
2. **Identification des Compétences :** Les compétences mentionnées dans les offres d'emploi ont été extraites et classées pour identifier les qualifications les plus demandées sur le marché.

#### b. Classification des Métiers par Machine Learning

L'utilisation du machine learning pour classifier les métiers permet de structurer les données de manière cohérente et automatisée. Les algorithmes de traitement du langage naturel sont utilisés pour analyser les descriptions des offres et les catégoriser selon des nomenclatures professionnelles standardisées.

#### c. Déduplication des Offres

La déduplication est un processus clé pour éliminer les doublons dans la base de données, assurant ainsi la précision et la fiabilité des informations. Les techniques de correspondance de texte et d'empreintes digitales sont utilisées pour identifier et supprimer les offres d'emploi en double.

**Étapes du Nettoyage des Données**

1. **Nettoyage des Données :** Les données collectées ont été nettoyées pour éliminer les erreurs et les incohérences. Des méthodes comme le TF-IDF ont été utilisées pour identifier et supprimer les doublons.
2. **Structuration des Données :** Les données ont été structurées pour faciliter leur analyse et leur interprétation. Cela inclut le renommage des colonnes et la standardisation des formats de données.

#### d. Passage de la Nomenclature des Métiers aux Familles Professionnelles

Le passage des métiers individuels aux familles professionnelles permet une analyse plus globale et cohérente des tendances du marché. Cette étape implique l'agrégation des données selon des catégories professionnelles plus larges, facilitant ainsi l'identification des tendances et des besoins en compétences.

### 3.2 Couverture et Représentativité des Données

1. **Matching des Variables :** Les données collectées ont été comparées avec celles fournies par d'autres sources, comme les informations d'Elodie, pour assurer la cohérence et la pertinence des variables.
2. **Réorganisation des Syntaxes :** Les syntaxes de collecte ont été réorganisées pour faciliter la maintenance et la mise à jour des scripts, garantissant une collecte continue et fiable des données.

## III. Caractéristiques de la Base d’Offres en Ligne 2024

### 3.1 Composition et Temporalité des Données

#### a. Dynamique de la Source de Données

La base de données d'offres d'emploi en ligne est dynamique, reflétant les fluctuations du marché du travail au fil du temps. Les données collectées en 2024 permettent de suivre les variations saisonnières et les tendances à long terme des offres d'emploi disponibles en Côte d'Ivoire.

#### b. Répartition des Offres par Secteur et Région

La répartition des offres d'emploi par secteur et région révèle des disparités géographiques et sectorielles. Certaines régions peuvent avoir une plus grande concentration d'offres dans des secteurs spécifiques, offrant des insights précieux pour les politiques de développement régional.

### 3.2 Couverture et Représentativité des Données

#### a. Analyse de la Représentativité

L'analyse de la représentativité des données est essentielle pour évaluer leur fiabilité et leur utilité pour la prise de décision. Cette section examine dans quelle mesure les données en ligne reflètent le marché du travail dans son ensemble et identifie les secteurs ou les professions sous-représentées.

#### b. Comparaison avec les Autres Sources de Données

La comparaison avec d'autres sources de données, telles que les enquêtes traditionnelles et les données administratives, permet de valider la qualité et la représentativité des offres d'emploi en ligne. Cette analyse comparative aide à identifier les forces et les limites des données en ligne pour l'analyse du marché du travail.

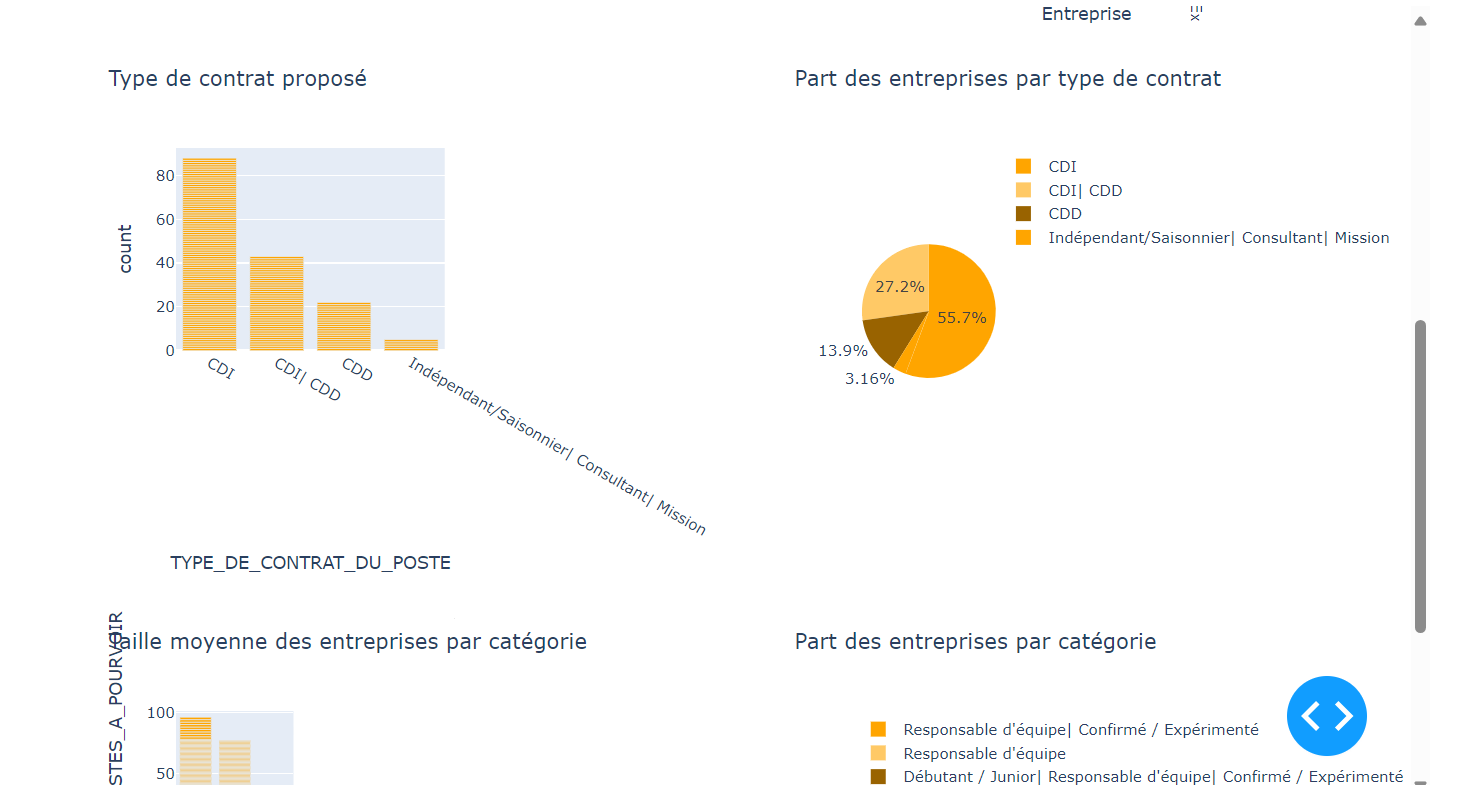
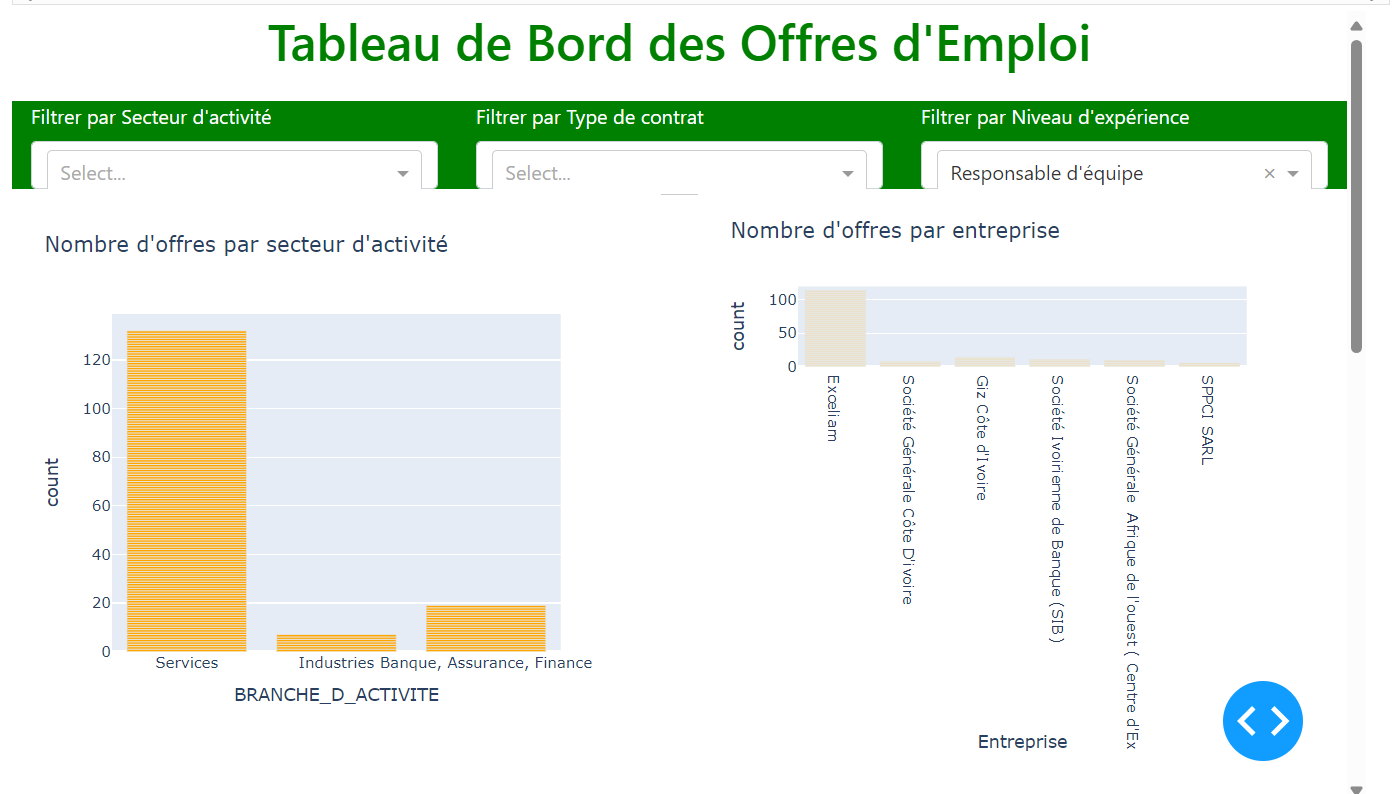
### 3.3 Utilisations de la Base de Données

#### a. Indicateurs de Tension et Suivi des Offres

Les données en ligne ont été utilisées pour calculer des indicateurs de tension sur le marché du travail, fournissant des informations sur la demande et l'offre de travail. Un suivi régulier des offres en ligne permet également de surveiller l'évolution du marché en temps réel.

#### b. Nouvelles Exploitations et Perspectives

Les données en ligne offrent de nombreuses possibilités pour de nouvelles exploitations, telles que l'analyse des compétences recherchées, l'étude des dynamiques sectorielles, et la modélisation des tendances futures du marché du travail. Ces perspectives ouvrent la voie à une meilleure compréhension et à une gestion plus efficace des ressources humaines.



## Conclusion

Les offres d'emploi en ligne constituent une source de données précieuse pour l'analyse du marché du travail en Côte d'Ivoire. La collecte et le traitement des données en 2024 offrent une vue dynamique et actualisée des opportunités professionnelles, complétant ainsi les sources traditionnelles de données. La mise en œuvre de techniques avancées de web scraping et d'analyse de texte permet d'extraire des informations utiles pour les chercheurs d'emploi, les recruteurs et les décideurs politiques, contribuant ainsi à une meilleure compréhension des dynamiques du marché et à la promotion du développement économique national.

## Bibliographie

1. [bing.com/ck/a?!&&p=ea6c71797346eedeJmltdHM9MTcyMDQ4MzIwMCZpZ3VpZD0yNDBhODcwNS0wNTJkLTY4MDgtMjVjZS05M2FlMDQ3NDY5MzMmaW5zaWQ9NTIwNg&ptn=3&ver=2&hsh=3&fclid=240a8705-052d-6808-25ce-93ae04746933&psq=ESSnet\_job\_vacancies\_Dares.pdf&u=a1aHR0cHM6Ly93d3cuaW5zZWUuZnIvZnIvc3RhdGlzdGlxdWVzL2ZpY2hpZXIvMzU1OTg4My9FU1NuZXRfam9iX3ZhY2FuY2llc19EYXJlcy5wZGY&ntb=1](https://www.bing.com/search?pglt=2083&q=ESSnet_job_vacancies_Dares.pdf&cvid=6148d4c44b1a40eebfa04c5b1ad49cd8&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIGCAEQRRg80gEHODg1ajBqMagCALACAA&FORM=ANNTA1&PC=DCTS)
2. [Les offres d'emploi en ligne, nouvelle source de données sur le marché du travail : illustration sur l'année 2019 (readkong.com)](https://fr.readkong.com/page/les-offres-d-emploi-en-ligne-nouvelle-source-de-donnees-4120655)
3. [demand for ICT specialists - methodological note (europa.eu)](https://ec.europa.eu/eurostat/documents/7894008/18023605/demand-ICT-specialists-methodological-note.pdf/9b69d43a-0a88-d0c9-afd9-24b140176721?t=1702477607157)

## Annexes

* **ANNEXE 1** - Test de la Matrice de Passage Secteurs-Métiers
  + [Description des tests et résultats]
* **ANNEXE 2** - Analyse de Dissimilarité entre Jocas et les Recrutements en Ligne
  + [Description des méthodes et résultats de l'analyse de dissimilarité]
* **ANNEXE 3** - Indicateur de Dissimilarité de Duncan par Famille Professionnelle
  + [Présentation de l'indicateur et analyse par famille professionnelle]

## Remerciements

Nous tenons à exprimer notre gratitude envers les personnes et institutions qui ont contribué à la réalisation de ce projet, notamment l’équipe de l'Institut National de la Statistique (INS) de Côte d'Ivoire pour la collecte de données, et nos collègues des institutions partenaires pour leur expertise technique. Un remerciement particulier à nos mentors pour leurs précieux commentaires et suggestions.