

Data Scientist Resume - French

AYYOUB EL-HYANI

elhyani.ayyoub@gmail.com

Résumé Professionnel

- Professionnel en Data Science avec expertise en modélisation prédictive, apprentissage automatique et NLP
- **Actuellement à la recherche d'un poste en alternance de Data Scientist** à partir de septembre 2025
- Expérience en analytique de santé, solutions IA pour l'agriculture et maintenance prédictive

Formation

Université Paris Cité (À partir de septembre 2025)

Master 2 – Systèmes Distribués et Technologies de la Science des Données (alternance) | Paris, France - Programme en alternance axé sur les systèmes distribués, l'intelligence artificielle, le big data et les architectures cloud

École des Sciences de l'Information (2022 – 2025)

Diplôme d'Ingénieur – Science des Données et Ingénierie des Connaissances | Rabat, Maroc - Cours spécialisés en apprentissage automatique avancé, ingénierie des données et modélisation statistique

Classes Préparatoires aux Grandes Écoles (2020 – 2022)

Filière MP – Mathématiques, Physique, Sciences Industrielles (MPSI/MP) | Marrakech, Maroc

Expérience Professionnelle

Somei Maroc (Fév. 2025 – Présent)

Stagiaire Data Science – Maintenance Prédictive | Rabat, Présentiel - Développé des modèles prédictifs (Random Forest, XGBoost) dans **Dataiku** pour prédire les pannes de canalisations d'eau - Implémenté des **méthodes d'analyse de survie** (Random Survival Forest) pour estimer les temps de défaillance - Collaboré avec des experts du métier pour valider les modèles et créer des tableaux de bord pour la planification de la maintenance

Nucleon Security (Juin 2024 – Août 2024)

Stagiaire Data Science | Rabat, Présentiel - Mené la migration de plus de 10M enregistrements de malware de MongoDB vers PostgreSQL, améliorant l'efficacité - Développé des pipelines ETL en Python avec pandas et SQLAlchemy, réduisant le temps de traitement de 35% - Conçu des tableaux de bord Apache Superset pour la surveillance en temps réel des menaces

Broccoli Data (Juil. 2023 – Sep. 2023)

Stagiaire Data Engineering | Télétravail - Développé des pipelines ETL modulaires avec **dbt** pour automatiser les transformations statistiques - Converti des jeux de données non structurés en formats parquet optimisés, améliorant l'accessibilité et la vitesse de requête

Projets en Machine Learning

Prédiction de l'Évolution du Neuroblastome

- Développé des modèles ML pour prédire la survie et la récurrence en utilisant des données cliniques et génomiques intégrées
- Réalisé des analyses de données à haute dimension avec PCA, t-SNE et clustering
- Évalué les modèles avec validation croisée, ROC-AUC et métriques de survie

Chatbot Agricole avec RAG (Retrieval-Augmented Generation)

- Construit un chatbot adapté aux agriculteurs marocains utilisant des LLMs et un pipeline RAG
- Effectué l'indexation vectorielle (FAISS + SentenceTransformers) sur des documents et guides officiels

- Généré des réponses fiables via l'API OpenAI et l'ingénierie de prompts personnalisée

Compétences Techniques

- **Langages de Programmation:** Python, R, MATLAB, Scala, Java
- **Machine learning:** scikit-learn, XGBoost, TensorFlow, PyTorch, Random Forests, Analyse de Survie
- **NLP/LLM:** Transformers, RAG, SentenceTransformers, FAISS, API OpenAI
- **Ingé des Données:** SQL, PostgreSQL, MongoDB, dbt, Dataiku, Apache Superset
- **Visualisation:** Power BI, Matplotlib, Seaborn, Tableau
- **Outils:** Jupyter, Git, Docker, APIs REST

Langues et Compétences Transversales

- **Langues:** Français (C2), Anglais (C1), Arabe (natif)
- **Soft skills:** Communication claire, Pensée critique, Autonomie, Travail en équipe, Adaptabilité, Gestion du temps