

Yosra KAROUI



Junior Data Scientist

✉ yosra.karoui@esprit.tn |  LinkedIn |  Github |  Website

COMPÉTENCES

- **Langages de programmation et scripts** : Java, Python, JavaScript, Scala, SQL, Linux (Shell Scripting)
- **Bases de données** : MySQL, PostgreSQL, MongoDB, Cassandra
- **Data Warehousing** : Snowflake, BigQuery, AWS Redshift , Data Factory Studio
- **Big Data** : Kafka, Spark, Hadoop
- **DevOps** : Jenkins, Docker, SonarQube, Grafana, Prometheus, Git
- **Cloud Computing** : AWS, Azure
- **ETL Pipelines** : Apache Airflow, SSIS, Talend, AWS Glue
- **Analyse de données** : Power BI, Tableau, Data Perspective Studio, Pandas, Numpy, Matplotlib
- **Machine Learning** : PyTorch, XGBOOST, Keras TensorFlow, Scikit-Learn, NLTK, SpaCy, Transformers
- **Web Scraping** : BeautifulSoup , Scrapy , Requests
- **Langues** : Anglais (C1), Français (C1)






EXPÉRIENCE

- **Dassault Systèmes**  05/2024 - 02/08
R&D Data Engineer/Analyst - Stage de fin d'études Vélizy-Villacoublay, France
 - Conception d'un pipeline de données en streaming en temps réel, transférant des données d'un Data Lake vers Data Factory et la création de tableaux de bord et de rapports avec data perspective.
 - Développement d'un prototype de chatbot avec fine-tuning de Transformers pré-entraînés .
- **ESPRIT-Tech**  06/2023 - 09/2024
Data Scientist - Stage Tunis - Tunisie
 - Implémentation de réseaux de neurones convolutifs pour un système de débruitage d'images .

EDUCATION

- **École supérieure privée d'ingénierie et de technologie : membre de la CGE** 10/2024
Diplôme d'ingénieur en informatique avec spécialisation en Data Science (Bac+5) Tunis, Tunisie

PROJETS ACADÉMIQUES CLÉS

- **Développement d'un Pipeline ETL pour l'Extraction de Données Twitter** 10/2024
Outils: [Airflow / Python / AWS /S3 Bucket] 
 - Création d'un pipeline ETL pour extraire des données de Twitter, les prétraiter et les charger dans un bucket AWS S3.
- **Développement d'un Pipeline de Streaming de Données** 10/2024
Outils: [Airflow / Python / Kafka /Zookeeper / Spark / Cassandra / Docker] 
 - Un pipeline de data engineering qui orchestre l'ingestion, le traitement et le stockage des données .
 - Conteneurisations des composants avec Docker pour un déploiement et une évolutivité faciles.
- **Projet DevOps : Optimisation du Pipeline CI/CD pour une Application Spring Boot** 11/2023 - 12/2023
Outils: [Jenkins / Nexus / Docker / Git / Springboot / SonarQube / Prometheus / Grafana] 
 - Rationalisation du pipeline CI/CD pour un backend Spring Boot avec automatisation du code, contrôle qualité, déploiement Docker, surveillance et alertes par e-mail.
- **MindMatch : Career Center Platform** 02/2023-05/2023
Outils: [BeautifulSoup / SpaCy / NLP /KMean / Logistic Regression / Power BI] 
 - Développement d'une plateforme de centre de carrière avec extraction de données (web scraping), gestion des données (Python) et algorithmes de machine learning pour profilage, régression et classification. Création d'un tableau de bord Power BI pour visualiser les tendances du marché du travail.
- **Prédiction de la maladie rénale chronique** 10/2022-12/2022
Outils: KNN / KMean / SVM / Naive Bayes 
 - nettoyage, visualisation et préparation des données.
 - Comparaison de modèles de classification : KNN, Naive Bayes, SVM, Random Forest, Arbre de décision..