

Yosra KAROUI



Data Engineer & Analyst

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

COMPÉTENCES

- **Langages de programmation et scripts** : Java, Python, Scala, SQL, Linux (Shell Scripting)
- **Bases de données** : MySQL Server , PostgreSQL, MongoDB, Cassandra
- **Data Warehousing** : Snowflake, BigQuery, AWS Redshift , Data Factory Studio
- **Big Data** : Kafka, Spark, Hadoop , Cloudera , Pig, Hive, Flume , Elasticsearch
- **DevOps Et Cloud** : Jenkins, Docker, SonarQube, Grafana, Prometheus, Git , AWS, Azure
- **ETL Pipelines** : Apache Airflow, 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** 
Outils: [Airflow / Python / AWS / S3 Bucket]
 - Création d'un pipeline ETL pour extraire, prétraiter et charger des données de Twitter dans AWS S3.
- **Développement d'un Pipeline de Streaming de Données** 
Outils: [Airflow / Python / Kafka / Zookeeper / Spark / Cassandra / Docker]
 - Orchestration de l'ingestion, du traitement et du stockage des données.
 - Conteneurisation des composants avec Docker pour un déploiement et une évolutivité faciles.
- **Exploration des données COVID-19 avec Tableau et SQL** 
Outils: [Tableau / SQL / SQL server / Excel]
 - Extraction et nettoyage des données avec SQL.
 - Création d'un tableau de bord interactif sur Tableau pour visualiser cas, décès et vaccinations.
 - Analyse des tendances par pays et comparaison des indicateurs clés.
- **Projet DevOps : Optimisation du Pipeline CI/CD pour une Application Spring Boot** 
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.
- **Projet Data Science : Plateforme intelligente pour centre de carrière** 
Outils: [BeautifulSoup / SpaCy / NLP / KMean / Logistic Regression / Power BI]
 - Extraction de données via web scraping, application d'algorithmes de machine learning pour le profilage, la régression et la classification, et création d'un tableau de bord Power BI pour visualiser les tendances du marché du travail.