

Yosra KAROUI



Junior Data Scientist

✉ yosra.karoui@esprit.tn | [in](#) [Linkedin](#) | [G](#) [Github](#) | [Website](#)

SKILLS

- **Programming languages and scripting:** Java, Python, JavaScript, Scala, SQL, Linux (Shell Scripting)
- **DataBase Management Systems :** MySQL, PostgreSQL, MongoDB, Cassandra
- **Data Warehousing :** Snowflake, BigQuery, AWS Redshift , Data Factory Studio
- **Big Data :** Kafka, Spark, Hadoop , PySpark
- **DevOps :** Jenkins, Docker, SonarQube, Grafana, Prometheus, Git
- **Cloud Computing :** AWS, Azure
- **ETL Pipelines :** Apache Airflow, SSIS, Talend, AWS Glue
- **Data Analysis :** Power BI, Tableau, Data Perspective Studio, Pandas, Numpy, Matplotlib
- **Machine Learning :** PyTorch, XGBOOST, Keras , TensorFlow, Scikit-Learn, NLTK, SpaCy
- **Web Scraping and Extraction :** Beautiful Soup , Scrapy , Requests
- **Languages :** Anglais (C1), Français (C1)






EXPERIENCE

- **Dassault Systèmes**  05/ 2024 - 02/08
R&D Data Engineer/Analyst - End-Of-Studies Internship Vélizy-Villacoublay, France
 - Design of a real-time streaming data pipeline, transferring data from a Data Lake to Data Factory, along with the creation of dashboards and reports using Data Perspective.
 - Development of a chatbot prototype with fine-tuning of pre-trained Transformers.
- **ESPRIT-Tech**  06/2023 - 09/2024
Data Scientist - Summer Internship Tunis - Tunisie
 - Implementation of convolutional neural networks (CNNs) for an image denoising system, with accuracy measurement across different noise models.

EDUCATION

- **Private Higher School of Engineering and Technology: CGE member** 10/2024
Computer Science Engineering Degree With specialization Data Science (Bac+5) Tunis, Tunisie

KEY PROJECTS

- **ETL pipeline for Twitter data extraction.** 10/2024
Tools: [Airflow / Python / AWS /S3 Bucket] 
 - Creation of an ETL pipeline to extract data from Twitter, preprocess it, and load it into an AWS S3 bucket.
- **Data streaming pipeline.** 10/2024
Tools: [Airflow / Python / Kafka /Zookeeper / Spark / Cassandra / Docker] 
 - A data engineering pipeline that orchestrates the ingestion, processing, and storage of data.
 - Containerization of components with Docker for easy deployment and scalability.
- **DevOps End-To-End Project** 11/2023 - 12/2023
Tools: [Jenkins / Nexus / Docker / Git / Springboot / SonarQube / Prometheus / Grafana] 
 - Streamlining the CI/CD pipeline for a Spring Boot backend with code automation, quality control, Docker deployment, monitoring, and email alerts.
- **MindMatch : AI Career Center Platform** 02/2023-05/2023
Tools: [BeautifulSoup / SpaCy / NLP /KMean / Logistic Regression / Power BI] 
 - Development of a career center platform with data extraction (web scraping), data management (Python), and machine learning algorithms for profiling, regression, and classification. Creation of a Power BI dashboard to visualize Job market trends.
- **Prediction of chronic kidney disease.** 10/2022-12/2022
Tools: KNN / KMean / SVM / Naive Bayes 
 - Data cleaning, visualization, and preparation.
 - Comparison of classification models: KNN, Naive Bayes, SVM, Random Forest, Decision Tree.