Tunis, Tunisia +216 92 536 640 yosra.karoui@esprit.tn

EDUCATION

ESPRIT school of Engineering (CGE Member)

Master's in Computer Science Engineering, Specialization in <u>Data Science</u> (with Honors)

Tunis, Tunisia Oct 2024

SKILLS SUMMARY

Programming Languages:

• Database Management Systems:

· Big Data:

• Data Analytics and Visualization:

• ETL and Data Warehousing:

DevOps:

Machine Learning and AI:

Languages:

Python, SQL, Java ,Scala ,R, Shell

NoSQL(Mongo DB), MySQL server, PostgreSQL

Spark, Hadoop, Kafka, Databricks

Power BI , Pandas, NumPy, Tableau, Seaborn , Matplotlib , Excel

Airflow, Talend, AWS Glue, Azure Data Factory CI\CD ,Docker, Git, Jenkins, Kubernetes , SonarQube. TensorFlow, Keras, Scikit-learn, NLTK, Langchain

Arabic(Native); French (C1), English (C1)

EXPERIENCE

Data Engineer & Analyst

Dassault Systèmes | Paris, France

Feb 2024 - Jul 2024

- Developed a real-time streaming data pipeline that transferred data from a Data Lake to Data Factory, resulting in a 25% increase in data ingestion speed and improved processing efficiency.
- Crafted and launched six real-time interactive dashboards using Data Perspective, enabling the visualization of critical data insights that informed strategic decisions for ongoing R&D projects.
- Fine-tuned Transformer models, enhancing its performance and response accuracy by 90% for a chatbot prototype.

Data Scientist Intern

ESPRIT-Tech | Tunis. Tunisia

Jun 2023 – sept 2023

Implemented convolutional neural networks (CNNs) for an image denoising system, achieving high-performance noise reduction. Conducted accuracy evaluation across multiple noise models to optimize system effectiveness.

RELEVENT PROJECTS

Data streaming pipeline | Link

Tools: [Airflow / Python / Kafka /Zookeeper / Spark / Cassandra / Docker]

Optimized a data engineering pipeline that orchestrates the ingestion, processing, and storage of data. Containerization of components with Docker for easy deployment and scalability.

ETL pipeline for Twitter data extraction | Link

Tools: [Airflow / Python / AWS /S3 Bucket]

• Designed an ETL pipeline to extract data from Twitter, preprocess it, and load it into an AWS S3 bucket.

DevOps End-To-End Project | Link

Tools: [Jenkins / Nexus / Docker / Git / Spring boot / SonarQube / Prometheus / Grafana]

Streamlined CI/CD for a Spring Boot backend by automating integration, enforcing SonarQube quality checks, deploying with Docker, and adding real-time monitoring via Prometheus and Grafana

COVID-19 Data Exploration with Tableau and SQL | Link

Tools: [Tableau / SQL Server / Excel]

- Extracted, cleaned, and transformed large datasets using SQL for accurate analysis and reporting.
- Created interactive dashboards in Tableau to visualize key metrics such (deaths, cases, and vaccination rates.)
- Performed in-depth trend analysis by country, comparing critical indicators to support data-driven decision-making.

HR Data Analytics with Power BI | Link

Tools: [Power BI / Power Query / Excel/ DAX]

- Cleaned data with Power Query and created three Key HR indicators (headcount, leave, salaries)
- Developed advanced dashboard using DAX Calculations

CERTIFICATIONS

AWS Cloud Practitioner, NLP Specialization, Pwc Power BI Job Simulation, Accenture Data Analytics and Visualization KPMG Data Analytics, Python for Data Science, NVIDIA AI,