

# HADIL SAHRAOUI

## DATA ENGINEER

+216 52 053 389 ◊ Tunis, Tunisia ◊ [hadir.sahraoui129@gmail.com](mailto:hadir.sahraoui129@gmail.com) ◊ [linkedin.com/in/hadir-sahraoui/](https://linkedin.com/in/hadir-sahraoui/)

### OBJECTIVE

Recent Data Engineer graduate seeking a full-time Entry Level Data Engineer role

### SKILLS

<b>Programming</b>	Python (Pandas, NumPy, Scikit-learn), SQL, Shell scripting
<b>Data Engineering &amp; Pipelines</b>	Apache Spark, Databricks, Delta Lake, Kafka, Airflow, SSIS, Lambda Architecture, Medallion Architecture , ETL/ELT Pipeline Design
<b>Databases &amp; Storage</b>	SQL Server, Snowflake, MySQL, PostgreSQL, Hive, MongoDB
<b>MLOps &amp; Deployment</b>	Docker, MLflow, FastAPI, Django, REST APIs
<b>Analytics &amp; Visualization</b>	Power BI, Tableau, Matplotlib, Seaborn, Excel, KPI Dashboards

### EXPERIENCE

<b>Data Scientist Intern</b> Orange Tunisia	Feb 2025 - Aug 2025 <i>Tunis, Tunisia</i>
<ul style="list-style-type: none"><li>Built churn prediction model with 90% recall and performed customer segmentation using K-Means clustering, achieving a Silhouette Score of 0.85.</li><li>Leveraged Qwen 2.5 LLMs with advanced prompt engineering to generate personalized marketing content.</li><li>Deployed a Django web application with real-time predictions, secure authentication, and Docker containerization for scalability.</li></ul>	
<b>Data Scientist Intern</b> ATB (Arab Tunisian Bank)	Aug 2024 - Sep 2024 <i>Tunis, Tunisia</i>
<ul style="list-style-type: none"><li>Built an OCR-based cheque autocorrection system using KNN and NLP to detect and correct mismatches between digits and words, achieving 92% accuracy.</li><li>Developed and deployed a Django web application with Power BI dashboards for real-time monitoring, reporting, and improved cheque validation efficiency.</li></ul>	
<b>Data Analyst Intern</b> Artibedded	Sep 2023 - Dec 2023 <i>Tunis, Tunisia</i>
<ul style="list-style-type: none"><li>Collected and processed data from web scraping and LLM-driven pipelines, performing data cleaning, preprocessing, and feature engineering for analysis.</li><li>Conducted exploratory data analysis (EDA) and created visualizations using Python (Matplotlib, Seaborn) and Power BI dashboards to derive actionable insights.</li></ul>	

### EDUCATION

<b>Engineering Degree in Data Sciences</b> , National School of Electronics and Telecommunications Machine Learning, Deep Learning, NLP, Big Data, Data Visualization, Decision Systems, Data Analysis, Data Warehousing, ETL Pipelines, Apache Spark	Sep 2025
<b>Preparatory Cycle for Engineering Studies</b> , Preparatory Institute of El-Manar	2020 - 2022

### PROJECTS

**Client Data ETL Pipeline – Orange Tunisia** Automated Lambda ETL pipelines on Databricks using Medallion architecture (Bronze/Silver/Gold). Processed high-volume telecom data and integrated into Snowflake for analytics, customer segmentation, and predictive modeling.

**Social Media Data Engineering for Mental Health** Implemented containerized ETL pipelines to ingest, validate, and normalize Facebook and Twitter data into MongoDB. Orchestrated workflows with Airflow, enabling downstream NLP analysis, including sentiment and zero-shot mental health classification.

**Data Engineering Pipeline – dbt, Snowflake & Airflow** Orchestrated ELT pipelines transforming TPCH raw data into analytics-ready models with dbt, loaded into Snowflake for scalable, reliable data delivery.

**Scalable Telecom ETL Pipeline** Designed a fully automated, scalable ETL pipeline to process raw telecom transactional data from ingestion to final loading. Optimized transformations and storage to support customer behavior analysis, profiling, and downstream analytics using Spark, Delta Lake, and Airflow.