Saif Allah Ibn El Hadj Mbarek

Ariana-Tunis | saifallah.ibnelhadjmbarek@fsb.ucar.tn | 54999156 | Github

Summary

Data Engineering student with a strong foundation in designing data pipelines, relational data models, and efficient data workflows for analytics and decision-making. Skilled in Python, SQL, and distributed systems, with experience in data preprocessing, transformation, and visualization. Passionate about bridging data engineering and data science by applying statistical analysis and machine learning techniques to extract meaningful insights from complex datasets.

Education

Faculty Of Sciences Of Bizerte-Pre-Engineering Program: Sept 2022 – May 2024

Completed a two-year preparatory program focused on mathematics, physics, and computer science fundamentals, developing excellent analytical and problem-solving skills..

Faculty Of Sciences Of Bizerte - National Engineering Degree in Computer Science Data Engineering: Sept 2024 – July 2027

Pursuing a National Engineering Degree focused on software development, data engineering, and artificial intelligence. Coursework includes databases, distributed systems, cloud computing, real-time systems, and big data processing.

Experience

Faculty of Sciences of Bizerte Sept 2025 – Current

Scrum Master (University Data Engineering Project)

- Leading a team of 6 students in developing an end-to-end data engineering solution under Agile/Scrum methodology.
- Coordinating sprint planning, daily stand-ups, and retrospectives to ensure on-time delivery of project milestones.
- Overseeing data pipeline design, ETL workflow development, and integration of analytical dashboards.
- Collaborating with team members to ensure code quality, version control, and deployment practices.
- **Using**: Python, SQL, Airflow, Docker, Power BI, Git

TAWER DIGITAL GROUP (TDG) June 2025 – Aug 2025

Data Engineering Intern

- Developed core modules of an ERP system for e-commerce operations.
- Designed and implemented data models to structure company operations and reporting.
- Performed KPI analysis to support decision-making and improve business insights.
- Built 5+ interactive dashboards in Power BI to visualize company performance metrics and reducing report generation time.
- Developed a Next.js web application with admin authentication and JWT tokens.
- Deployed the solution using Docker and a VPS for the database, and hosted the frontend on Vercel for scalable access.
- **Using**: Nextjs, Postgresql, Docker, GoogleSheet, PowerBI

Artificial Intelligence Data Engineering July 2024 – Aug 2024

AI/ML Research intern Laboratory

- Worked on implementing and fine-tuning a large language model (LLM), an intelligent model that assists Tunisian students with academic path navigation. Focused on model selection and fine-tuning, resulting in improved accuracy of up to 80.
- integrating speech recognition and Text-to-Speech (TTS) capabilities to enhance accessibility and user interaction.

- **Using**: Pytorch, Hugginface libraries, streamlit.

Projects

Educational chatbot(OrientAI) Github

- Built an AI-powered chatbot based on the TinyLlama-1.1B-Chat-v1.0 model to assist students in educational programs.
- Created a custom QA dataset using web scraping and preprocessing for fine-tuning and evaluation.
- **Using**: Python, Streamlit, Huggingface libraries, Transformers.

End-to-End Machine Learning Analysis Github

- Developed an end-to-end machine learning pipeline on a real-world dataset, covering exploratory data analysis (EDA), data preprocessing, and visualization.
- Achieving 90% accuracy in model selection by fine-tuning hyperparameters, improving the precision of the selected model by 5% through advanced optimization techniques.
- **Using**: Python, scikit-learn, numpy, pandas.

MapReduce Framework in Go Github

- Designed and implemented a MapReduce framework in Go to demonstrate core concepts in distributed data processing—fundamental to big data systems like Hadoop and Spark.
- Simulated large-scale parallel processing by coordinating multiple worker nodes, showcasing fault-isolated, efficient task division across the map and reduce phases.
- Reinforced understanding of how distributed computing frameworks enable high-throughput, fault-tolerant processing of massive datasets.
- **Using**: Go, Goroutines, RPC.

Technologies

Languages: Python,Java(JEE, SE), GO, SQL, NoSQL.

Tools: Apache Airflow, Apache Kafka, Docker, PostgreSQL, Power BI.

Frameworks & Libraries: Transformers, Next.js, Streamlit, Pytorch, scikit-learn.

Version Control & DevOps Git, Github, Vercel, .