

Mohammed Babaqi

Data Science

mbabaqi2020@gmail.com • +967 778558924 • Yemen, Sana'a

[LinkedIn](#) • [GitHub](#) • [Kaggle](#)

Profile

Data professional focused on data engineering, analytics, and applied machine learning. Hands-on experience building Python-based data workflows, automated reports, and Power BI dashboards, including an operational monitoring solution at StarPlus Pharma. Developed end-to-end projects involving multilingual news processing, predictive modeling, and FastAPI-based model deployment. Skilled in Python, SQL, and turning raw data into reliable, actionable insights.

Education

Bachelor of Data Science at Al-Nasser University, Sana'a, Yemen July 2023 - July 2027
Applied academic projects in data analysis, machine learning, NLP, and data engineering.

Experience

Data Analyst at [StarPlus Pharma](#), Sana'a, Yemen May 2025 - October 2025

- Automated reporting workflows and developed data pipelines using Python and Excel.
- Built a Power BI operations monitoring dashboard covering customers, sales, and financial performance.

Projects

DiaClassifier • [GitHub](#)

Engineered DiaClassifier, an end-to-end, Dockerized machine learning solution for early diabetes risk detection. Designed and benchmarked multiple classifiers with optimized hyperparameters and imbalance mitigation techniques, prioritizing high recall for medical sensitivity. Built a scalable FastAPI REST API for model serving and an interactive Streamlit interface for user-friendly predictions. Implemented modular architecture and containerized deployment to simulate real-world production environments

InsightBot • [GitHub](#)

InsightBot is an automated news scraping and analytics system that collects, processes, and visualizes news from over 30 websites. It uses Python for web scraping and NLP to extract topics, generate headlines, and detect languages across multilingual content. Processed data is stored in MongoDB for structured access, while a Streamlit-based interface with Admin and User roles enables interactive exploration. Advanced insights and trend visualizations are delivered through Power BI dashboards, providing a complete end-to-end solution for real-time news intelligence and analysis.

Technical Skills

- Programming Languages:** Python, R, SQL
- Data Manipulation & Analysis:** pandas, NumPy, Dask, Polars, Matplotlib, Seaborn, Plotly
- Machine Learning & AI:** BERT, LLM, scikit-learn, XGBoost, LightGBM, mlxtend
- Big Data Technologies:** Apache Spark, Hadoop, Dask, Databricks,
- Databases & Storage:** PostgreSQL, Microsoft SQL Server, MongoDB, Neo4j
- BI Tools:** Power BI, Tableau, Looker, Excel
- Tools:** Git, GitHub, Docker
- AI Tools & Platforms:** Hugging Face, Kaggle, Gemini, ChatGPT

Languages

- Arabic** - Native
- English** - Proficient