

JABER MAHMOUD

DATA ENGINEER | DATA PIPELINES | BIG DATA & CLOUD

Tripoli, Lebanon | +961 70 986 883 | jaber_md22@outlook.com

LinkedIn | GitHub | Portfolio

PROFESSIONAL SUMMARY

Mathematics graduate (Data Science, USJ) focused on data engineering: designing data ingestion and ETL pipelines, modelling relational schemas, and turning large, multi-source datasets into clean, analysis-ready tables. Strong in Python and SQL, with hands-on experience integrating satellite, weather, and transactional data (including a 500K+ row geospatial dataset) and automating multi-stage processing workflows. IBM Data Science Professional Certificate holder, currently pursuing the IBM Data Engineering Professional Certificate. Seeking an entry-level Data Engineer role where reliable pipelines and well-structured data power downstream analytics and ML.

TECHNICAL SKILLS

Programming & Querying: Python, SQL, R, C++, C#, JavaScript, HTML/CSS

Data Engineering: ETL/ELT pipeline design, data ingestion & integration, relational schema & data modelling, batch processing, data cleaning & transformation, SQLAlchemy, Alembic

Databases & Storage: Relational databases, SQLite, SQL analytics & query optimization

Cloud & Big Data: Microsoft Azure, Google Cloud, Google Earth Engine, big-data frameworks

Data Science & ML: Pandas, NumPy, scikit-learn, XGBoost, LightGBM, TensorFlow/Keras, SHAP; supervised, unsupervised & ensemble methods; NLP; model evaluation

BI & Visualization: Power BI, DAX, Tableau, Streamlit, Matplotlib, Seaborn, Plotly, Advanced Excel

Tools: Git/GitHub, Flask, Microsoft 365 Copilot, Power Platform, Google Workspace, Canva

EDUCATION

Université Saint-Joseph (USJ), North Lebanon Jan 2023 - May 2026

Bachelor of Mathematics, Data Science Option | Cumulative GPA 3.37/4.0 | Final semester (S7) GPA 4.0/4.0

Relevant coursework: Data Mining, Artificial Intelligence, Natural Language Processing, Big Data Frameworks, Relational Databases, Data Visualization, Statistical Analysis of Data, Probability for Data Science, Data Structures & Algorithms, Advanced C++.

KEY PROJECTS

Lebanon Flood Detection & Risk Intelligence Platform | Geospatial Data Pipeline & ML | Graduation Project

Co-developed with Ibrahim Rajab | Python, XGBoost, MLP, SHAP, Streamlit, Flask, Google Earth Engine

- Built an automated geospatial data pipeline on Google Earth Engine, ingesting and integrating four satellite, radar, and climate sources (NASADEM, ESA WorldCover, ERA5-Land, Sentinel-1) into a 521,664-row labelled dataset across 572 grid cells and five winters (2020-2024).
- Built a 22-feature transformation layer (rainfall, soil, radar, temporal interactions) feeding 11+ classifiers, and delivered a tuned XGBoost + MLP soft-voting ensemble reaching 0.96 accuracy and 0.93 macro-F1 on a time-based 2024 holdout.
- Shipped a Flask web app (live risk map, automated report generator, four-language AI copilot) and a 34-page Streamlit dashboard with hazard indexing and 3D terrain.

Retail Revenue Intelligence Platform | ETL Pipeline & Analytics

Python, Pandas, SQL/SQLite, Power BI, DAX

- Built a 6-stage data pipeline on 19,960 orders (4,338 customers, £10.6M revenue): ingestion, cleaning, SQL transformation in SQLite, RFM segmentation, forecasting, and reporting.
- Modelled 11 RFM customer segments in SQL and surfaced Champions driving £5.7M of £8.9M known-customer revenue; delivered an 8-week revenue forecast and a 15-visual Power BI layer.

AI-Generated Text Detection | NLP Pipeline & Web App

Python, scikit-learn, TensorFlow/Keras, spaCy, NLTK, Flask

- Built a reproducible NLP data pipeline: curated and balanced a 2,400-essay corpus (human PERSUADE 2.0 vs. GPT, Copilot, DeepSeek, Gemini), then ran preprocessing, TF-IDF feature extraction, POS tagging, and clustering.
- Benchmarked 5 classifiers including a GRU; Linear SVM reached 0.80 accuracy, 0.83 F1, and 0.99 ROC-AUC on a prompt-held-out test set, deployed via a Flask app with authentication, quotas, and PDF/DOCX/TXT input.

Flight Delay Intelligence | Data Mining, Clustering & Prediction

Python, scikit-learn, XGBoost, Streamlit

- Cleaned and feature-engineered historical flight records (delay-ratio, seasonal and temporal features), then applied K-Means (K=3, elbow + silhouette validated) to surface operational delay profiles, visualised with PCA.
- Benchmarked regression and classification models for delay prediction and delivered an interactive Streamlit dashboard for exploration.

Elite Kits | Full-Stack App & SQL Data Layer

Flask, SQLite, JavaScript, HTML/CSS

- Designed a relational SQLite schema and seeded 330 products behind a Flask app with CSRF protection, rate limiting, soft-delete audit logs, and an admin dashboard (CSV exports, coupons, application tracking).
- Authored 15 SQL analytics queries covering revenue, top sellers, coupon effectiveness, and customer lifetime value.

Additional projects: SpaceX Falcon 9 Launch Prediction (IBM Data Science Capstone) | LINK vs BTC Financial Analysis (Power BI) | Global Superstore Sales Analysis | Melbourne Rainfall Prediction (IBM).

CERTIFICATIONS

IBM Data Engineering Professional Certificate, IBM / Coursera (in progress)

IBM Data Science Professional Certificate, IBM / Coursera (2025)

Also (Coursera): Google Cloud Fundamentals: Core Infrastructure | Google Prompting Essentials | Microsoft Azure Cloud Services | Tableau | Power BI & Power Virtual Agents | Microsoft Power Platform | Microsoft 365 Copilot.

LEADERSHIP & ACTIVITIES

United Nations Global Compact

- **Ambassador of Change (2026 - Present)**, advocating the UN Global Compact's Ten Principles and Sustainable Development Goals and supporting local engagement initiatives.
- **SDG Brain Lab 5.0, Delegate & Spokesperson.** Led the Fair Trade Lebanon challenge and co-developed LOCAL+, a certification, QR-transparency, and retail-activation system helping Lebanese products compete with imported brands (Business Model Canvas + go-to-market).

RoadRescue AI × RoadChip | USJ BIAT Entrepreneurship Competition | Co-Founder | May 2026

- Pitched a two-product venture (AI roadside assistant + Bluetooth vehicle-monitoring device) targeting 1.5M+ Lebanese vehicle owners; built a working prototype and a business case with four revenue streams and a \$54K+ Year-2 projection.

Global Mentorship Initiative (GMI), completed Dec 2025. Structured mentorship with an international industry professional on personal branding, interviewing, and global business communication.

Social Media Designer | Google Developer Student Club (GDSC), USJ North Lebanon | Nov 2023 - May 2024. Led visual-content strategy and event branding for a campus tech community.

Languages: Arabic (Native) | English (Fluent, Professional Working) | French (A1) | German (A1, in progress).