

SARAH ASSAAD

AI Engineer & Data Scientist

Machine Learning • NLP & AI Pipelines • Data Engineering • Full-Stack AI Systems

sarahassaad797@gmail.com | +961-71780532 | Koura, North Lebanon | [GitHub](#) | [LinkedIn](#)

PROFESSIONAL SUMMARY

AI Engineer and Data Scientist with a mathematical foundation (BSc Mathematics — Data Science, GPA 3.52) currently deepening expertise through an MSc in Artificial Intelligence. I take full ownership of AI systems — from raw data and pipeline design through model development to deployed, user-facing applications. My work spans machine learning, deep learning, NLP, computer vision, and time-series forecasting, with a consistent focus on explainability, data quality, and results that hold up on unseen data. I work well across technical and non-technical teams, communicate findings clearly, and bring rigour and genuine curiosity to every problem I tackle.

EDUCATION

Master in Artificial Intelligence	Bachelor of Mathematics – Data Science Option
- Université Saint-Joseph (USJ), Beirut - GPA: 3.15 Sep 2025 – May 2027 - Machine Learning, Natural Language Processing, Computer Vision, Statistics for AI & Machine Learning, Optimization for AI, Game Theory, Parallel Computing, Foundations of Decision Modeling.	- Université Saint-Joseph (USJ), North Lebanon - Final CGPA: 3.52 Sep 2022 – May 2025 - Core coursework: Statistical Inference, Big Data Technologies (Spark/Hadoop), Probability, Optimization

SKILLS

ML & AI	Scikit-learn, XGBoost, LightGBM, Random Forest, ARIMA, Prophet, supervised & unsupervised learning, cross-validation, hyperparameter tuning, class imbalance handling, anomaly detection
Deep Learning	PyTorch, TensorFlow, Keras, CNNs, neural network design and training
NLP	spaCy, NLTK, TF-IDF, TextBlob, Naive Bayes, LDA topic modeling, sentiment analysis, text classification
Computer Vision	OpenCV, Kalman Filters, Lucas–Kanade Optical Flow, Hungarian Matching, IoU-based association, multi-object tracking
Data Engineering & Integartion	Python (Pandas, NumPy), SQL, ETL/ELT pipelines, REST APIs, JSON/XML/CSV data processing, PostgreSQL, MySQL, Spark, Hadoop, Excel VBA
Software Development	Flask, SQLAlchemy, Streamlit, HTML/CSS, authentication & authorization, role-based access control, SMTP integration, relational database design
Model Evaluation & Explainability	SHAP, ROC-AUC, residual analysis, train/validation curve analysis, model evaluation
Tools & Platforms	Git, Matplotlib, Seaborn, Plotly, Tableau, Power BI, Jupyter Notebook
Languages	Arabic (Native) • English (Professional) • French (Conversational)

EXPERIENCE

Planning & Gate / Logistics Intern

Jan 2025

CMA CGM – Tripoli, Lebanon

- Automated reporting workflows using Excel VBA, reducing manual processing and improving reporting reliability.
- Integrated data from planning systems, gate operations, and shipping records into unified reporting datasets.
- Collaborated with Planning, Operations, and IT teams to deliver actionable business insights.
- Selected as a CMA CGM scholarship recipient for academic performance and technical potential.

PROJECTS

Electricity Consumption Forecasting for Smart Cities

Python • XGBoost • LightGBM • ARIMA • Prophet • SHAP • Streamlit | Final Year Project

- Built a fully automated end-to-end ML pipeline over four years of hourly electricity and weather data (35,000+ records): data cleaning, feature engineering, model training, evaluation, and live Streamlit dashboard deployment — zero manual steps.
- Implemented and benchmarked seven model families: Linear Regression, Decision Tree, Random Forest, XGBoost, LightGBM, ARIMA, and Prophet; best model (XGBoost) achieved $R^2 = 0.90$, MAE = 1,096 on unseen data.

The Text Psychologist — AI-Powered NLP Analytics Application

Python • spaCy • NLTK • TF-IDF • Scikit-learn • TextBlob • Streamlit

- Designed and deployed a complete multi-stage NLP pipeline: text preprocessing → TextBlob sentiment analysis → TF-IDF + Naive Bayes emotion classification → spaCy POS/dependency writing style analysis → LDA topic modelling → contextual advice generation.
- Built a live Streamlit dashboard with real-time inference: word frequency bar charts, parts-of-speech pie charts, and emotion strength visualisations rendered on demand.

Skincare E-Commerce Platform — Full-Stack Web Application

Python • Flask • SQLAlchemy • Jinja2 • SQL • SMTP API

- Built a production-ready full-stack web application from scratch: user authentication, session management, role-based access control (admin vs. customer), complete product management admin dashboard, and shopping experience.
- Designed and managed a relational database (SQLAlchemy ORM) covering user accounts, product catalogue, and order management; ensured data consistency and integrity throughout.
- Implemented automated password recovery via SMTP email integration — end-to-end API and third-party service connection.
- Demonstrates full ownership from schema design through backend logic to frontend rendering — a complete, deployable product.

Loan Default Prediction — Risk Modelling & Data Quality Auditing

Python • Scikit-learn • XGBoost • Streamlit

- Identified class imbalance as the root cause of misleading accuracy metrics; corrected with weighted learning and rebuilt the evaluation framework around ROC-AUC and recall.
- Applied systematic cross-validation to select the most generalizable model and deployed results via dashboard for stakeholder review.

Multi-Object Tracking — Real-Time Computer Vision System

Python • OpenCV • NumPy • Kalman Filters • Hungarian Matching

- Implemented a real-time multi-object tracking system using Kalman Filters for motion prediction and Lucas-Kanade Optical Flow for estimation; evaluated on the MOT15 benchmark using MOTA and MOTP.