

Nour Hatem

Data Scientist | Machine Learning Engineer

nour.hatem117@gmail.com | +201282417851 | Giza, Egypt

[linkedin.com/in/nour-hatem-](https://www.linkedin.com/in/nour-hatem-) | github.com/nour-hatem | kaggle.com/nourelhashash | nour-hatem.web.app

Professional Summary

Data Scientist and ML Engineer in training, building intelligent systems that move businesses from guessing to deciding. Skilled across the full data lifecycle from statistical modeling and ML pipelines to BI dashboards using Python, SQL, Scikit-learn, XGBoost, and Power BI. Not here to analyze. Here to deliver.

Education

Helwan University

B.Sc. in Computer Science and Artificial Intelligence

Major: Artificial Intelligence

September 2023 – Present

Helwan, Egypt

Relevant Experience

Data Analysis Trainee

National Telecommunication Institute (NTI)

March 2026 – Present

120 Hours

- Analyzed real-world datasets across e-commerce, healthcare, and finance using Python, SQL, R, and Excel; built end-to-end workflows from raw SQL extraction to final dashboard delivery in Power BI and Tableau
- Developed a freelance portfolio with client-ready proposals, personal branding strategy, pricing, and project scoping documents

Generative AI & Deep Learning Trainee

Information Technology Institute (ITI) – in partnership with NVIDIA DLI

February 2026

90 Hours

- Earned **two NVIDIA DLI certifications** in a 90-hour AI program; applied prompt engineering using NVIDIA NIM and LangChain
- Built **deep learning models from scratch** in PyTorch for CV (MNIST, ASL) and NLP (BERT Q&A)

Machine Learning Trainee

Digital Egypt Pioneers Initiative (DEPI)

November 2025 – Present

- Completed **20-week ML program** applied 10+ supervised and unsupervised algorithms and built an end-to-end ML project with a Dockerized application, REST API, and production-ready model serving
- Gained hands-on experience in Deep Learning (CNNs, RNNs, LSTMs, Transformers), NLP, Computer Vision, Azure AI services, and MLflow for experiment tracking

AI & Machine Learning Trainee

Huawei HCIA-AI V3.5 Certification Program

October 2025

- Implemented **4 ML algorithms** using Python, Scikit-learn; trained deep learning models in MindSpore for CV (MNIST, MobileNetV2, ResNet-50) and NLP (TextCNN) with CNN, RNN
- Applied **regularization (L1, L2, Dropout)** and cross-validation for bias-variance

Machine Learning Specialist Trainee

National Telecommunication Institute (NTI)

July 2025 – August 2025

- Completed **120-hour ML scholarship** (90 technical) with **93.5%**; applied supervised (Linear, Logistic, SVM, KNN, Naive Bayes, Trees, Random Forest) and unsupervised (K-Means, DBSCAN, PCA) algorithms on real-world datasets
- Executed **full ML lifecycle** (feature engineering, model selection, tuning, evaluation) using Colab, Kaggle GPU for scalable training

Projects

Financial Database Analysis

SQL Server · NTI Data Analysis Track

[GitHub Repo](#)

April 2026

- Designed a normalized 11-table financial database on SQL Server ingesting 50K+ transactions across 1,100 customers and 318 loans; engineered a full cleaning pipeline handling duplicate PKs, null values, and future-dated records using CTEs and ROW_NUMBER()
- Delivered 20 analytical queries across customer segmentation, loan risk, and fraud detection; flagged 1,035 high-risk accounts and 5 velocity fraud pairs using Z-score anomaly detection and LAG() window functions; optimized with 9 nonclustered indexes

NYC Taxi Trip Duration — Prediction Pipeline

Personal Project

[GitHub Repo](#) | [Kaggle Notebook](#)

February 2026

- Built an end-to-end regression pipeline on 1.46M NYC taxi trips achieving R^2 of 0.78–0.82 and RMSE of ~260s; engineered 8+ features including Haversine distance and cyclical time encodings encapsulated inside the pipeline to eliminate train/serve skew
- Applied domain-driven data cleaning removing ~15K corrupt records; detected and documented data leakage from a target-derived feature using correlation analysis, demonstrating production ML best practices

Budget vs Actual Performance Dashboard

Personal Project

[GitHub Repo](#)

January 2026

- Delivered a 5-page executive Power BI dashboard analyzing 9,500+ transactions across 6 departments and 5 regions over 3 years, replacing 50+ manual Excel reports with a single interactive solution
- Uncovered 112% budget utilization (\$87.9M variance) and reduced monthly reporting time from 40 hours to 2 hours by automating KPI tracking with DAX-powered measures for variance, utilization rates, and trend forecasting

HR Analytics Dashboard

Personal Project

[GitHub Repo](#)

September 2025

- Built an interactive Power BI dashboard analyzing 2,845 employees across engagement, performance, and training cost metrics; flagged a critical engagement risk with a score of 2.94/6 despite 86% active headcount
- Identified workforce concentration risk with 67% of employees in a single department and \$1.59M training spend with unvalidated ROI

End-to-End Heart Disease Prediction System

[Live App](#) | [Kaggle](#) | [GitHub](#)

Final Project – Sprints x Microsoft Summer Camp

August 2025

- Benchmarked 5 supervised models on 13 clinical features; optimized XGBoost via hyperparameter tuning achieving **86.6% accuracy, 89.6% F1-score, and 94.8% recall**
- Deployed the final model as an interactive Streamlit web application for real-time heart disease risk assessment with confidence scores

Alzheimer's Disease Classification: Supervised ML for Early Diagnosis

[Kaggle](#) | [GitHub](#)

Final Project – NTI Machine Learning Scholarship

August 2025

- Engineered an end-to-end supervised ML pipeline on 2,149 clinical records and 34 features; benchmarked 10 classifiers and achieved top performance with **XGBoost: 92.6% F1-Score and 94.8% Accuracy**
- Identified Age, MMSE score, and functional assessments as the strongest predictors through in-depth EDA and correlation analysis

Technical Skills

- **Data Science & ML:** Scikit-learn, XGBoost, Supervised & Unsupervised Learning (Regression, Classification, Clustering), Deep Learning (TensorFlow, Keras, PyTorch), NLP, Model Evaluation (Precision, Recall, F1), Hyperparameter Tuning, PCA
- **Data Analysis & Preprocessing:** Python (Pandas, NumPy), Data Cleaning, Feature Engineering, EDA, Excel (Pivot Tables, Power Query)
- **Data Visualization & BI:** Matplotlib, Seaborn, Plotly, Tableau, Power BI, Interactive Dashboards
- **Programming & Scripting:** Python, SQL (MySQL, MS SQL Server), C, C++, Java, JavaScript, HTML5, CSS3
- **Databases & Data Engineering:** Relational Database Design, Query Optimization, MySQL, MS SQL Server, Git, GitHub
- **MLOps & Deployment:** MLflow, Docker, FastAPI, Streamlit, REST APIs
- **Cloud & Environment:** Azure ML, AWS, Jupyter Notebook, Google Colab, VS Code, Linux CLI, Conda

Certifications & Courses

- **[Machine Learning Specialization](#)** December 2025
Coursera – DeepLearning.AI & Stanford Online | Instructor: Andrew Ng
- **[Building LLM Applications with Prompt Engineering](#)** February 2026
NVIDIA Deep Learning Institute (DLI)
- **[Getting Started with Deep Learning](#)** February 2026
NVIDIA Deep Learning Institute (DLI)
- **[Career Essentials in Data Analysis](#)** October 2025
Microsoft & LinkedIn Learning
- **[Data Analyst in Python Track](#)** September 2025
DataCamp | 36 Hours
- **[AI & Machine Learning Foundations](#)** August 2025
Sprints x Microsoft | 40 Hours
- **[MySQL Data Analysis](#)** April 2025
Maven Analytics

Languages

Arabic: Native

English: Conversational