

Zeyad Mohamed Ahmed Ahmed Refaey

[Linked In](#)

[GitHub](#)

Zeyad.mo.refaeys@gmail.com

+201026887363

Summary

Data Scientist in training with a strong foundation in data analysis, machine learning, and deep learning. Experienced in Python and SQL, with hands-on expertise using scikit-learn, TensorFlow (Keras), and PyTorch to build supervised and unsupervised learning models, including regression, classification, and clustering. Familiar with large-scale data processing using PySpark, and passionate about developing predictive models and transforming data into actionable, business-driven insights.

Education

B.Sc. in Computer Science – Data Science Track Egyptian Chinese University(ECU)
Last Term GPA: 3.06 (**Expected Graduation: 2027**)

Skills

Programming & Databases: Python, SQL, PySpark.

Machine Learning: Scikit-learn, Supervised Learning (Regression, Classification), Unsupervised Learning (Clustering), Feature Engineering, Model Evaluation, Hyperparameter Tuning.

Deep Learning: TensorFlow (Keras), PyTorch, Neural Networks, Convolutional Neural Networks (CNNs), Transfer Learning, Regularization Techniques

Big Data: PySpark, Distributed Data Processing, Large-Scale Data Analysis

Data Analysis & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Power BI, Tableau

Version Control: Git, GitHub

Soft Skills: Analytical Thinking, Problem Solving, Communication, Team Collaboration

Projects

Bank Customer Segmentation & Interactive Deployment | [GitHub Link](#)

Dec 2025

Python (Streamlit, Scikit-Learn, PCA, Joblib)

- Engineered a clustering pipeline to segment 45,000+ bank customers into 5 distinct behavioral profiles (e.g., "High-Value Managers", "Young Professionals").
- Reduced feature dimensionality using PCA and optimized clusters through K-Means, DBSCAN, and Hierarchical clustering comparisons to improve marketing targeting.
- Developed and deployed an interactive Streamlit web application allowing marketing teams to input client data and receive real-time segment classification.

XGBoost Regression for Consumer Spending Prediction | [GitHub Link](#)

Nov 2025

Python (XGBoost, Scikit-Learn, Seaborn, Pandas) |

- Built a high-performance Gradient Boosting (XGBoost) model to forecast total customer expenditure based on demographic and financial attributes.
- Implemented an end-to-end regression pipeline including Label Encoding, feature scaling, and outlier management to ensure data quality.
- Evaluated model accuracy using R² and Mean Absolute Error (MAE), visualizing performance through Actual vs. Predicted scatter plots and feature importance rankings.

MNIST Digit Recognition with Deep Learning & Grad-CAM | [GitHub Link](#)

Jan 2026

Python (TensorFlow/Keras, OpenCV, Matplotlib, NumPy) |

- Designed and trained a Convolutional Neural Network (CNN) to achieve high-accuracy handwritten digit recognition on the MNIST dataset.
- Integrated Grad-CAM (Gradient-weighted Class Activation Mapping) to provide model interpretability by generating heatmaps that visualize the CNN's focus areas.
- Conducted misclassification analysis and utilized heatmaps to debug model weaknesses, enhancing the transparency of the deep learning architecture.

Predictive Classification for Bank Term Deposits | [GitHub Link](#)

Dec 2025

Python (Random Forest, Scikit-Learn, Pandas) |

- Developed a Random Forest classification model to predict customer subscription likelihood for financial products (term deposits).
- Streamlined data preprocessing by automating the removal of low-impact features and encoding complex categorical variables like job type and marital status.
- Analyzed model performance using Confusion Matrices and Classification Reports to maximize precision, helping the bank reduce campaign costs.

Experience

Data Science Intern | Fawry:

Dec 2025 – Jan 2026

Onsite / Cairo, Egypt

During this onsite internship, I developed a diverse range of machine learning and deep learning solutions, starting with an NLP-powered sentiment analysis system using LSTM and GRU architectures in TensorFlow/Keras, which was deployed via a REST API using Flask or FastAPI. I performed financial customer segmentation by applying K-Means clustering and the Elbow Method to identify distinct client behavioral patterns. Additionally, I built a feedforward neural network for handwritten digit classification on the MNIST dataset, focusing on data normalization and categorical cross-entropy optimization. My work also involved engineering predictive models for retail customer spending using Linear Regression and conducting extensive exploratory data analysis with Pandas, NumPy, and Seaborn to clean datasets and visualize complex transaction behaviors.

Data Analysis Intern (NTI) | [link](#):

June – September 2025

(Learning Program) | Remote

Completed a structured internship focused on practical data analysis and visualization skills.

Worked on hands-on projects covering data cleaning, exploratory analysis, and dashboard creation. Practiced tools such as **Python (Pandas, NumPy, Matplotlib)**, **SQL**, **Power BI**, and **SSIS**.

Generative AI Intern Beginner Level (Information Technology Institute (ITI)) | [Link](#)

September 2025

(NVIDIA DLI Summer Program)

Completed a 35-hour summer training focused on Generative AI fundamentals and LLM development.

Gained hands-on experience with Python, Generative AI concepts, Prompt Engineering, and Retrieval-Augmented Generation (RAG).

Licenses & Certifications

AI & Machine Learning Summer Camp | [link](#)

September 2025

Sprints x Microsoft

Hands-on training covering Python programming, OOP, data structures, and exception handling. Built ML models for supervised and unsupervised learning tasks.