Zeyad Mohamed Ahmed Ahmed Refaey

Zeyad.mo.refaey@gmail.com

Linked In
GitHub

+201026887363

Summary

Data Scientist in training with strong foundations in data analysis, machine learning, and data engineering. Skilled in Python, SQL, and ML frameworks (scikit-learn, TensorFlow) with hands-on experience in regression, classification, and analytics projects. Passionate about building predictive models and turning data into actionable insights.

Skills

Programming & Databases: Python, SQL
Machine Learning & AI: Scikit-learn, TensorFlow (Beginner), Regression Models, Classification, Model Evaluation.
Data Analysis & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Power BI, Tableau.
ETL & Data Engineering: ETL Pipelines, Data Cleaning, Schema Design
Tools & Platforms: Jupyter Notebook, VS Code, SQL Server Management Studio (SSMS), Visual Studio.
Version Control: Git, GitHub.
Soft Skills: Analytical Thinking, Problem-Solving, Communication, Team Collaboration.

Projects

Multiple Linear Regression Analysis | link — July 2025
Python (scikit-learn, Pandas, Matplotlib)
Developed a custom regression pipeline to predict continuous variables with high accuracy. Implemented data preprocessing (missing value handling, outlier detection, feature engineering), multicollinearity checks (VIF), scaling, and model evaluation (R^2 score). Automated predictions on new datasets and generated regression summaries to enhance interpretability.

Iris Flower Classification | link — September 2025
Python (scikit-learn, RandomForest, Seaborn)
Built and evaluated a Random Forest classification model to predict Iris flower species. Conducted exploratory data analysis (pairplots, correlations), applied feature scaling, and achieved **high accuracy**. Delivered a complete ML workflow including training, testing, and visualization of results with a confusion matrix.

Logistic Regression for Binary Classification | link — August 2025
Python (statsmodels, Pandas, Seaborn)
Implemented a logistic regression model from scratch to classify binary outcomes, including manual feature encoding, probability predictions, and evaluation. Designed a reusable pipeline to preprocess categorical variables, fit models, generate predictions, and export results, achieving strong classification accuracy.

Customer Segmentation with K-Means Clustering link — August 2025
Python (scikit-learn, Seaborn, Matplotlib)
Performed customer segmentation using K-Means clustering on satisfaction and loyalty data. Applied feature scaling, elbow method for optimal cluster selection, and visualized customer groups to support targeted marketing strategies and data-driven decision-making.

Experience

Generative AI Intern (NVIDIA DLI Program) | link September 2025
Information Technology Institute (ITI)
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).

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.

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**.

Education

B.Sc. in Computer Science – Data Science Track Egyptian Chinese University(ECU)

Last Term GPA: 3.06 **(Expected Graduation:** 2027**)**