Omar Hazem Ahmed

01064001251 omarhazemhassan@gmail.com
linkedin.com/in/omar-hazem-ahmed GitHub Giza, El-Haram

Profile

- Data Science and Artificial Intelligence (DSAI) student at Zewail City of Science and Technology (expected graduation: 2027)
- Strong foundation in machine learning, data analysis, and Python programming
- Proficient in libraries and tools such as NumPy, Pandas, scikit-learn, TensorFlow, and Matplotlib
- Experienced in developing and evaluating predictive models and performing statistical analysis
- Passionate about applying data-driven solutions to real-world problems
- Strong communicator with experience working independently and collaboratively in project-based environments
- Eager to contribute to impactful projects in data science, ML research, or AI development

Education

University of Science and Technology, Zewail City **2023 – Present**
B.Sc. in Computational Science and AI, Concentration in Data Science and AI

Skills

- **Languages**: Python, SQL, C++
- **Technical Skills**: Machine Learning, Data Analysis, Data Visualization, scikit-learn
- **Soft Skills**: Problem Solving, Self-Learner, Presentation Skills

Languages

- English B2
- Arabic

Professional Experience

Freelancer, Mostaqel 04/2025 – Present

- Engaged in freelance data science and machine learning projects

IBM Data Scientist, Digital Egypt Pioneers Initiative (MCIT) – DEPI Internship 10/2024 – 06/2025
Cairo, Egypt

- Participated in a national-level data science internship supported by IBM and the Digital Egyptian Pioneers Initiative (DEPI)
- Focused on building, evaluating, and optimizing machine learning and AI models using real-world datasets
- Worked with supervised learning algorithms: Decision Trees, Random Forest, XGBoost, Support Vector Machines, and Neural Networks
- Applied hyperparameter tuning, cross-validation, and feature selection to improve model performance
- Developed end-to-end machine learning pipelines using Python, scikit-learn, Pandas, and Jupyter Notebooks
- Completed over 30 hands-on projects covering data preprocessing, exploratory analysis, model evaluation, and reporting
- Demonstrated technical communication skills by documenting project work and publishing on GitHub
- Collaborated with mentors and peers through group assignments and live sessions to simulate industry envi-ronments

PROJECTS

Tissue Insight Mega Project

- Built a full-stack web application for multi-modal cancer detection, comprising:
- Gene Expression Analysis: Processed CSV-based gene expression data and trained a tree-based classifier, improving prediction accuracy by 35%, achieving a 91% F1 score
- Image Classification: Implemented a CNN model to analyze microscopy images of cells, distinguishing cancerous from non-cancerous samples
- Interactive RAG Chatbot: Integrated a Retrieval-Augmented Generation system to guide users through data uploads and interpretation, enhancing user engagement and support

Multiclassification Model

- Developed a machine learning model to perform multiclass classification on network traffic data, distinguishing between 6 different categories of activity or behavior
- Preprocessed raw network data, including encoding categorical features, handling imbalanced classes, and normalizing inputs
- Trained and evaluated multiple classification algorithms (e.g., Random Forest, XGBoost, Neural Networks) to identify the most accurate model
- Achieved high accuracy and F1-score in identifying and differentiating network behavior classes for anomaly detection and threat monitoring
- Optimized model performance using hyperparameter tuning and cross-validation techniques

Auto Learn Model

- Developed a binary-classification pipeline to detect anomalies in network traffic by comparing 4 distinct machine learning models
- Trained and evaluated 4 classifiers—Random Forest, Support Vector Machine, Gradient Boosting, and Logistic Regression—using cross-validation
- Optimized hyperparameters and performed model comparison based on precision, recall, and F1 score to ensure robust anomaly detection and now it detects 20% more accurate

DEPI-Datascience-AI

- Curated a comprehensive GitHub repository documenting completed assignments, labs, and mini-projects from the Digital Egyptian Pioneers Initiative (Data Science & AI track)
- Covered topics such as data preprocessing, exploratory analysis, feature engineering, and model evaluation

Webscraper Fast People Search

- Developed a Selenium-based scraper to extract rental and sale listings for apartments and houses from an online real-estate platform
- Navigated dynamic pages, handled pagination, and bypassed simple anti-bot measures
- Parsed HTML content to extract key fields (e.g., price, location, property size, amenities) and performed basic data cleaning achieving a 50% score above the avg cleaning projects across my university teams
- Stored the cleaned dataset into CSV format for downstream analysis and reporting

Olivia Search Engine

- Developed a modular search engine implementing foundational information retrieval techniques
- Built an inverted index and Boolean query parser supporting AND/OR/NOT operations for precise document filtering
- Implemented TF-IDF and BM25 scoring to rank documents, tuning parameters for optimal relevance
- Incorporated query expansion methods (synonym lookup, pseudo-relevance feedback) to enhance recall
- Integrated ELMo and BERT embeddings for semantic retrieval, improving contextual relevance
- Evaluated system performance using precision, recall, and Mean Average Precision (MAP), demonstrating a 12% increase in retrieval accuracy

Credit Card Using SQL

- Developed a Streamlit application to visualize credit card transaction data, backed by a PostgreSQL database
- Designed a normalized PostgreSQL schema to store customer and transaction records
- Ingested and cleaned large datasets (50,000+ rows) using Python and Pandas
- Built interactive Streamlit dashboards featuring key metrics (e.g., spending trends, fraud risk scores)

Visualizing Citations in Research Paper

- Built a citation network visualization tool that extracts and maps research paper citations
- Automated retrieval of metadata and citation links using web scraping and scholarly APIs
- Utilized NetworkX to generate interactive visualizations of the citation network
  Enabled users to input custom search queries and dynamically explore academic influence

Plotly Dash Dashboard

- Designed and developed interactive dashboards using Plotly Dash to visualize datasets in real-time
- Automated retrieval of metadata and citation links for scholarly articles
- Parsed citation relationships to construct directed graphs
- Utilized NetworkX for interactive visualizations of citation networks

Simple Dashboard

- Built a responsive and interactive web dashboard using AMCharts 5 to visualize key performance metrics
- Created various chart types (bar, pie, scatter) to represent multidimensional data
- Customized themes, tooltips, and animations for enhanced user engagement
- Integrated live or static datasets to dynamically render charts

Library Management System

- Developed a console-based Library Management System using Object-Oriented Programming in C++
- Designed core classes (Book, Member, Librarian, LibrarySystem) for managing inventory and operations
- Applied OOP concepts—encapsulation, inheritance, polymorphism—for modularity and scalability
- Built a user-friendly command-line interface for administrators and users

CERTIFICATES

- Introduction to Data Visualization with Plotly in Python, DataCamp
- Intermediate Python, DataCamp
- Crash Course on Python, Coursera
- Building Dashboards with Dash and Plotly, DataCamp
- Foundations: Data, Data, Everywhere, Coursera
- Introduction to SQL, DataCamp
- Introduction to Python, DataCamp
- Ask Questions to Make Data-Driven Decisions, Coursera
- Supervised Learning with scikit-learn, DataCamp