# Abdelrhman Ahmed Ezzat

## Data Scientist

Kafr el-Sheikh, Egypt +20-101-801-8692 abdalarhmanezzat@gmail.com

www.linkedin.com/in/abdelrhman-a-ezzat github.com/Abdelrhman941

## Summary

Data Scientist proficient in Python and Machine Learning, with strong expertise in Data Science workflows including EDA, predictive modeling, Model Optimization, and data storytelling. Experienced in end-to-end projects with real-world applications. Strong foundation in AI techniques, with hands-on deployment experience (FastAPI, Docker) and a track record of building real-world ML prototypes.

## Education

Menoufia University — B.Sc. Artificial Intelligence & Data Science. Oct 2022 – Expected Jul 2026

## Experience

**Digital Egypt Pioneers Initiative (DEPI) – Generative AI Trainee** Jul 2025 – Present

- Contributed to a national initiative on advanced AI and Data Science technologies, focusing on Generative AI and Large Language Models (LLMs).
- Trained in Generative AI, LLMs, and prompt engineering techniques.
- Developed prototype Machine Learning applications leveraging LLMs and Generative AI models, with focus on scalability and deployment.
- Collaborated with a team of peers to design prompt engineering workflows.

## Skills

### Technical Skills:

- **Programming & Data Handling:** Python, SQL, PostgreSQL, NumPy, Pandas, OOP.
- **Machine Learning:** Scikit-learn, Random Forest, XGBoost, SVM, Decision Trees, Logistic Regression, Feature Engineering, Feature Selection, Model Evaluation, Hyperparameter Tuning.
- **Deep Learning:** PyTorch, NLP (Transformers, NLTK), Computer Vision (OpenCV), Reinforcement Learning.
- **Deployment & Cloud:** Docker, FastAPI, MLOps, AWS (S3, EC2), Azure (ML Studio), GCP (BigQuery).
- **Data Visualization & Storytelling:** Matplotlib, Seaborn, Power BI, Streamlit.
- **Big Data & Forecasting:** Hadoop, Spark, Prophet, ARIMA, SARIMA.

### Soft Skills:

- Problem-Solving, Communication Skills, Adaptability, Collaboration & Teamwork, Time Management, Critical Thinking, Leadership, Creativity.

## Projects

**Traffic Sign Detection and Classification (GTSRB)** Dec 2023

- • Collaborated with a team to design and optimize a custom CNN model trained on the German Traffic Sign Recognition Benchmark (GTSRB) dataset, **achieving 98.5% test accuracy** across 43 classes.
- Reduced inference latency to <50ms per image, enabling potential deployment in real-time autonomous driving systems.
- Addressed robustness challenges by augmenting data under different lighting, rotation, and weather conditions.

***Tools**: Python, TensorFlow, Keras, OpenCV.*

### DQN Agent for 2048 Game (Reinforcement Learning) May 2025

- Designed and implemented a Deep Q-Learning agent with replay buffer, target networks, and epsilon-greedy exploration, trained over 10,000+ episodes on GPU.
- Consistently reached the 2048 tile in 85% of games, with an average score exceeding **15,000 points**.
- Built monitoring dashboards with TensorBoard and Matplotlib to visualize Q-value evolution, convergence stability, and action distributions

***Tools**: TensorFlow, NumPy.*

### Vehicle Detection using Haar Cascades (OpenCV) Dec 2023

- Developed a prototype vehicle detection system using Haar cascade classifiers for real-time traffic monitoring.
- Annotated bounding boxes and integrated live video feed detection pipeline for deployment on CCTV streams.
- Experimented with feature tuning (scale factor, minNeighbors) to balance detection accuracy and false positives.

***Tools**: Python, OpenCV.*

### Sports Popularity Analysis with Web Scraping May 2025

- Engineered a large-scale data collection pipeline scraping Google Trends API and YouTube analytics for 50+ sports categories.
- Processed 10,000+ time-series data points and applied seasonal decomposition to uncover popularity spikes.

***Tools**: Python, BeautifulSoup, Pandas.*

### Auto-correct System using NLP & Edit Distance May 2025

- Designed a spelling correction engine combining edit distance, probabilistic N-gram models, and transformer-based contextual embeddings.
- **Achieved 92% accuracy** on benchmark datasets with a vocabulary size of 50,000+ words.
- Implemented correction strategies for both single-token and multi-word errors, improving usability in search and text input applications.

***Tools**: NLTK, spaCy, Edit Distance, N-gram models, Transformers.*

### House Prices EDA & Regression Modeling Apr 2025

- Performed EDA and feature engineering as part of a Data Science pipeline.
- Built regression Machine Learning models with cross-validation, improving baseline RMSE by 15%.

***Tools**: Pandas, Seaborn, Scikit-learn.*

### Student Performance Prediction Dec 2023

- Developed regression models to predict student grades using socio-demographic and study habit features.
- Visualized data trends with Seaborn, highlighting key predictors of academic success.

***Tools**: Scikit-learn.*