

# Beltagy Mohamed

Data Scientist

## Summary

Data Analyst & AI Specialist with expertise in machine learning, deep learning, big data engineering, and cloud-based AI solutions. Proficient in Python, R, SQL, Power BI, Tableau, and AWS SageMaker for data cleaning, visualization, predictive modeling, and scalable ML pipeline development. Experienced in Hadoop, Spark, Kafka, and ETL workflows, delivering measurable results such as 96% model accuracy and 99.7% reduction in AI diagnosis time. Strong collaborator with proven leadership in cross-functional teams to solve complex real-world challenges.

## Education

### Bachelor of Computer Science at Fayoum University , Fayoum

September 2025 - July 2025

**Grade:** Good

Earned A+ grade and formal recognition from the Dean for developing an AI-based medical diagnosis application, demonstrating advanced skills in machine learning model development, data preprocessing, and application deployment.

Secured 1st place in a Software Engineering course project, which was selected for special review and discussion by the course professor for its innovation and quality.

Organized and managed university-level events, workshops, and conferences, enhancing project management and stakeholder coordination abilities.

## Professional experience

### AWS Machine Learning Engineer Intern at National Telecommunications Institute , Cairo

November 2024 - May 2025

- Completed intensive training in Data Engineering, Machine Learning, Deep Learning, and AWS Cloud fundamentals, alongside continuous soft skills development in communication, teamwork, and public speaking.
- Executed multiple technical tasks and real-world projects, culminating in a capstone project that integrated all learned skills into a cloud-based AI solution.
- Applied supervised and unsupervised learning techniques with advanced feature engineering and model optimization, improving model performance and scalability.

### Vice President — Data Analysis Team at Enactus Fayoum, Fayoum

January 2025 - July 2025

- Led the Data Analysis Team in delivering an environmental sustainability project, winning 1st place in the Schneider Electric Competition.

## Info

[biltaji.mohamed@gmail.com](mailto:biltaji.mohamed@gmail.com)

+201096383395

Cairo, Egypt

## Links

[LinkedIn](#)

[Medium](#)

[GitHub](#)

## Skills

### Programming Languages:

Python, C++, SQL

### Data Analysis & Machine

**Learning:** Pandas, NumPy, Scikit-learn, TensorFlow, Keras, PyTorch, R (ggplot2, dplyr), Statistics, Linear Algebra, Feature Engineering, Model Optimization

**Data Visualization:** Tableau, Power BI, Matplotlib, Seaborn, DAX, Power Query

### Big Data & Cloud

**Technologies:** Hadoop, Hive, HBase, Spark, Kafka, Flink, AWS SageMaker, AWS Lambda, AWS Glue, AWS Redshift, AWS S3

**Database Management:** SQL, MySQL, Data Modeling, ETL

**Operating Systems:** Linux, Windows

**Development Tools:** Git, GitHub, Flask, Jupyter Notebook, Microsoft Office Suite

**Soft Skills:** Team Leadership, Public Speaking, Problem Solving, Strategic Planning, Collaboration, Communication

## Volunteering

### Content Creator at Sabq, Online

August 2025 - Present

- Produced written and visual content for online audience engagement

- Designed and executed data collection, cleaning, and visualization strategies that improved decision-making efficiency for the project by 35%.
- Organized technical workshops and strategy camps to train team members on data analytics tools (Python, Excel, Power BI), improving team skill levels and project delivery time.

### **Artificial Intelligence Intern at Information Technology Institute, Online**

August 2024 - September 2024

- Gained solid foundations in statistics, linear algebra, and mathematical concepts essential for AI.
- Developed expertise in machine learning and deep learning through hands-on projects and class assignments, including supervised/unsupervised algorithms and neural networks.
- Developed expertise in machine learning and deep learning through hands-on projects and class assignments, including supervised/unsupervised algorithms and neural networks.
- Utilized python for exploratory data analysis, visualization, and predictive modeling, achieving measurable accuracy improvements.
- Designed data workflows that increased model accuracy by 15%.

### **Big Data Engineer Intern at National Telecommunications Institute, Smart Village**

July 2024 - August 2024

- Completed foundational training in Python, Linux, and SQL before advancing to Huawei Big Data curriculum.
- Acquired proficiency in HDFS, Zookeeper, Kafka, Spark, HBase, MapReduce and related big data tools, applying them to large-scale data storage and processing tasks.
- Managed and optimized big data infrastructure using Hive, HBase, MapReduce, and Spark, improving query efficiency by 25%.
- Implemented distributed data processing pipelines for high-volume datasets, enhancing system scalability and performance.

### **Data Analyst – Call Center Project (Power BI) at Qafza, Online**

December 2024 - December 2024

- Designed an interactive Power BI dashboard to analyze call center KPIs, tracking metrics such as average handling time, first call resolution, and customer satisfaction scores.
- Performed data cleaning, transformation, and DAX optimization to improve dashboard refresh performance by 40%.
- Delivered actionable insights that helped improve operational efficiency in the call center.

### **Business Problem-Solving Intern at McKinsey & Company (Forward Program), Online**

May 2023 - November 2023

- Collaborated with an international, cross-functional team to develop a transportation optimization strategy for London, addressing cost, efficiency, and sustainability.
- Applied structured problem-solving, root cause analysis, and data-driven recommendations.

### **Volunteer at Resala Charity Association, Fayoum**

March 2025 - April 2025

- Assisted in preparing and distributing food, clothing, and medicine to underserved communities during Ramadan.

### **Languages**

**Arabic** - Native

**English** - Conversational

### **Certificates**

**AI& Data Science - AWS Machine Learning Engineer from Ministry of Communications and Information Tecnology**

May 2025 • [↗](#)

**Data Analysis-Skills for the Freelance Professional from National Telecommunications Institute**

September 2024 • [↗](#)

**Big Data Associate from National Telecommunications Institute / Huawei Egyptian Talent Academy**

August 2024 • [↗](#)

**Power-BI Intensive Course from Qafza**

December 2024 • [↗](#)

**Machine Learning Pipline from Great Learning**

December 2024 • [↗](#)

**Mastering critical SKILLS in Data Structures using C++ from CSGetSkilled Academy**

March 2023 • [↗](#)

**Digital Marketing from Udacity**

June 2022 • [↗](#)

**Forward from Mckinsey**

November 2023 • [↗](#)

- Enhanced communication, leadership, and strategic thinking skills through McKinsey's structured learning framework.

## Projects

### Clear View App – Graduation Project • 🔗

Fine-tuned YOLOv8s achieving 96% accuracy on an 8K-image dataset, reducing diagnosis time from 1 hour to 10 seconds (99.7% reduction) for near real-time image analysis.

Eliminated patient costs for eye disease diagnosis, reducing expenses from 150 EGP per visit to zero by enabling remote AI-based scanning.

Improved model robustness under limited data conditions by benchmarking EfficientNetB3, MobileNetV2, and Xception architectures, significantly reducing overfitting.

Developed and deployed a medical RAG-based Q&A chatbot using LLaMA 4 integrated with Flask, achieving ~90% validated accuracy over 100+ medical queries.

Implemented real-time response capability by sourcing reliable medical information from WebMD, CDC, and MedlinePlus.

### Startup Expansion Data Analysis • 🔗

Conducted exploratory data analysis (EDA) on U.S. startup expansion dataset containing metrics on sales, marketing spend, revenue, and profit across multiple cities and states.

Analyzed geographic expansion trends, identifying top 10 states for both new and existing startup locations.

Calculated Return on Marketing Spend (ROMS) and ROMS%, revealing marketing strategies that achieved up to X% higher returns in specific sales regions.

Compared performance metrics between new vs. old expansions, uncovering revenue patterns that guided targeted marketing investment.

Investigated correlation between marketing spend and profit/revenue, optimizing future budget allocation.

Tools & Technologies: Python, Pandas, NumPy, Matplotlib, Seaborn, Jupyter Notebook.

### Comprehensive Diabetes Analysis and Prediction • 🔗

Performed exploratory data analysis (EDA) to identify key trends, distributions, and correlations within diabetes datasets.

Executed data cleaning, preprocessing, and feature engineering to prepare data for predictive modeling.

Developed and trained machine learning models (e.g., Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting) to predict diabetes outcomes.

Evaluated model performance using accuracy, precision, recall, and F1-score, achieving up to X% prediction accuracy.

Visualized findings to support better understanding of risk factors and potential prevention strategies.

Tools & Technologies: Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Jupyter Notebook.