

# Mohamed Hassan

## Data Scientist — Data Engineer

🏠 Riyadh | 📞 0574345011 | ✉️ [mo7amed7assan1911@gmail.com](mailto:mo7amed7assan1911@gmail.com)  
in [linkedin.com/in/mo7amed1911](https://www.linkedin.com/in/mo7amed1911) | 🐙 [github.com/mo7amed7assan1911](https://github.com/mo7amed7assan1911)

### Summary

---

One who likes to manipulate data, focused on creating innovative data solutions. Proficient in extracting insights from complex datasets through machine learning, deep learning, and statistical analysis. Skilled in designing ETL pipelines, data modeling, cloud platforms, and Big Data technologies such as Apache Spark for developing robust data solutions that drive impactful decisions. Currently expanding skills in Large Language Models (LLMs) through projects like machine translation systems and chatbot development.

### Education

---

**Faculty of Engineering | Tanta University** Sep. 2018 – June. 2023

- Bachelor of Engineering specializing in Computer Engineering and Control Department.
- Grade:** Very good with honors (82.8 %)

### Experience

---

**AI Engineer | ADRI | Riyadh - Saudi Arabia** Aug. 2024 – Present

- Developed a machine translator by fine-tuning a HuggingFace LLM and provided my work with an API that **increased the translation speed** from 1 minute per PDF page to 3 seconds per page.
- Designed and implemented a ChatBot using LLama 3.1 70B, integrated with open webUI.
- Conducted research and analysis on Arabic LLMs such as ALLaM and Jais.

**Data Engineer | PetroDayen | Alexandria - Egypt** Feb. 2024 – May. 2024

- Reduced expense calculation time by 75%** with interactive KPI visualizations.
- Built an ETL pipeline to extract data from Google Sheets into a Data Warehouse to handle relationships between customers, employees, projects, and resources
- Integrated the Data Warehouse with Power BI to visualize real-time KPIs and insights.
- Was responsible for documenting all company projects and services, and building a company WordPress website.

**Data Science Intern | Banque Misr 🏦 | Cairo - Egypt** Aug. 2023 – Sep. 2023

- Demonstrated leadership on a **time series analysis** end-to-end project, distributing tasks through **Agile** methodologies using **Jira**.
- Developed a generalized predictive model with XGBoost for **ATM transaction forecasts**.
- Developed a **predictive model** to detect loan defaulters following EDA.

**AI & ML Instructor (Freelance) | Tabuk University - Saudi Arabia** Sep. 2022 – Feb. 2023

- Taught fundamental concepts in Artificial Intelligence (AI) and Machine Learning (ML), with a focus on explaining linear algebra principles, searching algorithms, and ML models.

### Courses and Certificates

---

• <b>NLP with Transformers</b> — Hugging Face	Feb 2024 – Present
• <b>Data Warehousing Concepts</b> 🏦 — DataCamp	June 2024
• <b>Data Engineering ZoomCamp</b> — DataTalksClub	Feb 2024 – June 2024
• <b>Deep Learning Specialization</b> 🏦 — Coursera	Oct 2022 – Dec 2022
• <b>Machine Learning Specialization</b> 🏦 — Coursera	Sep 2022 – Sep 2022
• <b>Data Analysis Nanodegree Program</b> 🏦 — Udacity	Jul 2022 – Aug 2022
• <b>Machine Learning Engineer Track with Scikit-learn</b> 🏦 — DataCamp	Sep 2023 – Present

## Selected Projects

---

### Floor Plan Generation using DL techniques — Graduation project 🐦

Sep. 2022 – June. 2023

- **USED:** *Graph Neural Networks (GNN), Convolutional NN (CNN), PyTorch, PyTorch Geometric*
- Provided **home design** based on input ground boundaries, generating horizontal views and 3D architecture.
- **Studied papers** like DeepLayout (2019) and ArchiGAN (2019), and collected and processed floor plan images into Shapely polygons and PyTorch Geometric graphs.
- **Built a deep learning architecture** consisting of **Convolution Neural Network (CNN)** models to take the boundary as images and get centroids of rooms, and **Graph Neural Network (GNN)** to take centroids as graphs and then predict the width and height of rooms.

### ETL and Data Warehouse for Fantasy Premier League 🐦

June. 2024 - Jul 2024

- **Used:** *Docker, Apache Spark, Workflow Orchestration (Mage), PostgreSQL*
- Set up a development environment using **Docker & Docker-Compose**.
- Developed an **ETL pipeline** with **Apache Spark** to load data into the **Data Warehouse**.
- Designed a **data model** and implemented it with **PostgreSQL**.
- Utilized Mage for **workflow orchestration** and automated ETL runs each game week.

### Loan Default Study 🐦

Aug. 2023 – Jan. 2024

- **Used:** *Python, Tableau, XGBoost*
- Merged self-aggregated previous dataset with current application data to create the final dataset for analysis.
- Conducted in-depth analysis of loan applications, identifying insights like age distribution and occupation trends.
- Developed a predictive model for default risk assessment using advanced techniques like PCA and XGBoost, achieving high recall and AUC scores.

### Analysis for flight delays and cancellations in USA 🐦

Aug. 2022 – Aug. 2022

- **Used:** *Python, Tableau*
- Cleaned the time columns using Python, and merged datasets to get one dataset.
- Built a story for analysis using Tableau from 5 slides, including a dashboard containing all info about flight cancellations and a motion graph to show the rank of cancellation reasons over months.

## Technical Skills

---

### Tools:

- **ML & DL frameworks:** PyTorch, PyTorch Geometric, TensorFlow, Keras, Scikit-learn.
- **LLM frameworks:** Hugging Face Transformers, Ollama, LangChain.
- **Databases & Data Warehouse:** SQL (PostgreSQL, MySQL), NoSQL (MongoDB).
- **Containerization tools:** Docker.
- **Workflow Orchestration tools:** Mage.
- **Cloud Platforms:** AWS (S3, EC2, RDS).
- **Visualization tools:** Tableau, Power BI.
- **Python libraries:** PySpark, Pandas, NumPy, Matplotlib, BeautifulSoup, OpenCV, Shapely, GeoPandas.

### Concepts and Techniques:

- **RAG, ETL, Data modeling, Big Data, Web Scrapping, containerization, Git, Linux, Agile(Jira)**
- **Machine Learning:** Linear models, tree-based models, clustering techniques, and ensemble methods.
- **Deep Learning:** CNNs, GANs, and GNNs (Graph Neural Networks), NLP tasks with LLMs
- **Statistical Analysis:** Descriptive and inferential statistics, hypothesis and A/B testing, and regression analysis.

## 🏆 Achievements

---

**First place winner:** Our Graduation Project won first place in a **nationwide Hackathon competition against 35 universities**, aiming to revolutionize city planning in Egypt through intelligent design solutions.

[View the celebration here](#) 🐦