Malak Soula

Egypt, Alexandria
(+20) 1551021696 malaksola1606@gmail.com linkedin.com/in/MalakAbdelfattah github.com/MalakSoula

Education

University of Sadat City **Oct. 2021 – Jun. 2025**
B.Sc. in Computer Science and Artificial Intelligence – Bioinformatics Department *Sadat City, Egypt*

- **Graduated with Distinction — GPA:** 3.71 / 4.00

Relevant Coursework

- Algorithms Analysis
- Data Analysis
- Data Structures
- Object-Oriented Programming
- Genetic Algorithms
- Database Management Systems
- Machine Learning
- Bioinformatics
- Mathematical Biology
- Next-Generation Sequencing (NGS)
- Molecular Biology

Technical Skills

Programming Languages: Python, SQL (PostgreSQL), R

Libraries & Frameworks: Scikit-learn, TensorFlow, SMOTE, Pandas, NumPy, Seaborn, Matplotlib

Tools & Platforms: Power BI, Tableau, Excel, Jupyter Notebook, Google Colab, VS Code, Git, GitHub

Bioinformatics Tools: Galaxy, Mothur, Krona, Phinch

Awards

SAS Curiosity Cup 2025 – Global Finalist (Top 15/112 Teams) **Apr. 2025**
SAS – Remote International Competition

- Ranked in the top 13% globally among 112 teams.
- Developed and pitched a healthcare analytics model using SAS tools.
- Presented insights via a technical video and narrative-based storytelling.
- Gained hands-on experience with forecasting and sustainability-driven analytics.

SAS Forecasting Hackathon for Sustainability & Data4Good – 3rd Place **Dec. 2024**
The Arab Academy for Management, Banking and Financial Sciences (AAMBFS) *Al Sheikh Zayed, Egypt*

- Awarded 3rd place for a data-driven solution to optimize medical inventory management.
- Utilized advanced forecasting models in a collaborative, real-world healthcare scenario.
- Recognized for innovation by SAS and the ESG & Data4Good Center of Excellence.

Academic Excellence Awards (Top of Class) **2022 – 2024**
Faculty of Computer Science and Artificial Intelligence, University of Sadat City *Sadat City, Egypt*

- Achieved 1st place in the Bioinformatics Department for 3 consecutive academic years.
- GPA: 3.7 (Year 1), 3.5 (Year 2), 3.67 (Year 3), Final GPA: 3.71 with Distinction.
- Received formal recognition for outstanding academic and research performance.

Experience

Data Science and Analytics Intern **Apr. 2025 – Present**
The National Council for Women – Sprints Learning Journey *Remote*

- Completed the "Data Dynamo" module, focused on data querying and filtering using PostgreSQL.
- Completed the "Data Analysis and Visualization" module, with hands-on training in Excel, Power BI, and Tableau.
- Developed interactive dashboards to support gender equity research and policy insights.
- Currently enrolled in the extended Python programming module for automation and advanced data handling.

Data Science and Machine Learning Trainee **Oct. 2024 – Feb. 2025**
Microsoft Student Club – EELU *Remote*

- Completed a 5-month intensive program covering Python, Pandas, Matplotlib, Power BI, and ML fundamentals.

- Built and deployed machine learning models for NASA NEO classification and customer segmentation.
- Designed interactive dashboards using Power BI:
  - Pizza Sales Analysis: Visualized sales trends by category, size, and time.
  - Customer Segmentation: Combined clustering with business KPIs for strategic targeting.
- Presented solutions through live demos and collaborative coding sprints.

Cancer Data Science Intern **Sept. 2024 – Oct. 2024**
HackBio *Remote*

- Applied ML models to classify genomic data using R and Python.
- Explored gene expression patterns in large-scale cancer datasets.
- Documented findings in collaborative GitHub repositories.

Healthcare Research Intern **Aug. 2023 – Sept. 2023**
Hamad Medical Corporation *Doha, Qatar*

- Co-authored a systematic review with the surgical research team at HMC.
- Performed literature screening, data extraction, and quality assessment.
- Explored the intersection of digital health and AI in surgical research.

Graduation Project

FSP: Football Scene Predictor | *ML, Computer Vision, Medical AI, HTML, CSS, PHP, Flutter* **Jun. 2025**

- Collaborated in a cross-functional team to develop a smart system aiding referees and coaches in real time.
- Implemented two core modules:
  - **Match Analysis Tool:** Detected offside, goals, handballs, and out-of-play moments using computer vision.
  - **Player Health Checker:** Predicted cardiac risk and assessed player readiness using ML on historical medical and performance data.
- Led the development of the medical AI module, building predictive models for cardiovascular health using Python-based ML techniques.
- Built a responsive front-end with HTML/CSS and dynamic backend with PHP; deployed cross-platform using Flutter for both mobile and web access.

Projects

JobPulse – Hiring Trends Dashboard | *Power BI, Excel, Power Query, Wuzzuf Dataset* **May 2025**

- Developed a Power BI dashboard analyzing 25,000+ Wuzzuf job postings to uncover hiring trends, in-demand skills, and job categories.
- Used slicers and filters to display job roles, experience levels, and skill frequency.
- Applied Power Query for data cleaning, null handling, and skill normalization.
- Documented the ETL pipeline and insights in a technical summary report.

Customer Segmentation Analysis | *K-Means, Python, Scikit-learn, Pandas, Power BI* **Feb. 2025**

- Segmented 5,000+ customers into 4 distinct behavioral clusters using K-Means, improving targeting precision by 30%.
- Visualized clusters and customer insights using Power BI dashboards.

NASA NEO Hazard Classifier | *Python, Scikit-learn, Pandas, NASA Dataset* **Jan. 2025**

- Achieved 94% precision in classifying Near-Earth Objects as hazardous using ML on 338,000+ records.
- Conducted EDA, preprocessing, and trained ML classifiers for impact risk prediction.

Genetic Age Prediction Model | *Python, ML Models, Gene Expression* **Jan. 2025**

- Predicted biological age from gene expression profiles using regression-based ML algorithms.
- Performed EDA and feature selection on high-dimensional genetic data.

Used Vehicle Data Insights | *Python, Kaggle* **Dec. 2024**

- Analyzed used car prices, mileage, and brand popularity using Python.
- Identified pricing trends and optimal resale value ranges for top 10 brands across 10,000+ used car listings.

Pizza Sales Analytics Dashboard | *Power BI* **Dec. 2024**

- Designed a Power BI dashboard showcasing sales by date, size, and product category.
- Created interactive visuals for performance comparison and trend tracking.

Cherry Tomato Gene Expression Classification | *Python, Bioinformatics, Kaggle* **Nov. 2024**

- Distinguished organic vs. non-organic cherry tomatoes via ML classification on gene expression data.
- Used EDA, preprocessing, and classifiers to achieve high predictive accuracy.

Falcon Football Club Management System | *SQL, ERD, Database Design* **Dec. 2023**

- Built a DBMS for managing club operations including news, rosters, and training schedules.
- Designed ER diagrams and normalized relational schemas to ensure data integrity.

Sub-optimal Health Systematic Review | *Research Project, Literature Analysis* **Aug. 2023 – Sep. 2023**

- Assisted in systematic review project at Hamad Medical Corporation focused on sub-optimal health indicators.
- Contributed to screening, review synthesis, and research documentation.