AHMED ALI ATYA

Data Analyst & Aspiring Data Scientist

ahmedatiea30@gmail.com LinkedIn|GitHub Giza, Egypt

SUMMARY

I am a results-oriented Data Analyst and aspiring Data Scientist, proficient in utilizing Python, SQL, and statistical methods to extract actionable insights from complex datasets. My hands-on experience includes building predictive models and creating compelling data visualizations. I prioritize detail-oriented work and synergizing with teams to drive solutions that inform data-driven decision-making

EDUCATION

BachelorofScience in Computer Science, Scientific Computing Department

Ain Shams University

09/2023 - 06/2027 Cairo, Egypt

GPA
3.6 / 4.0

SKILLS

Technical Skills

Data Structures | EDA | EDR | GitHub | Java | Jupyter Notebook | MatplotLib | Numpy | Pandas | Python | Scikit-Learn | Seaborn | SQL | C++ | Machine Learning

Soft Skills

Communication | Problem-Solving | Critical Thinking | Leadership | Teamwork | Flexibility | Time Management | Decision-Making | Creativity & Innovation

PROJECTS

ObesityPrediction Model

https://github.com/ahmedali11a/Obesity_Prediction

A project focused on predicting obesity levels using various health metrics and machine learning techniques

- Developed an end-to-end predictive model to classify obesity levels based on 16 lifestyle and health features
- Developed an end-to-end classification model using 16 lifestyle and health features to predict 7 obesity categories.
- Applied data cleaning, exploratory data analysis, feature engineering (BMI, outlier handling, encoding, scaling), and trained multiple machine learning models (Logistic Regression, SVM, Random Forest, Gradient Boosting).
- Achieved up to 100% accuracy with a 98.3% F1-score, with Random Forest and Gradient Boosting outperforming baseline models.
- Visualized feature importance, correlations, and distributions to improve model interpretability and extract key health insights.

water Quality

https://github.com/ahmedali11a/Water_Quality

A project focused on predicting water potability using chemical and physical parameters with machine learning techniques

- Developed an end-to-end predictive pipeline to classify water quality based on nine key features (pH, hardness, solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, turbidity).
- Applied data cleaning, outlier handling, normalization, exploratory data analysis (EDA), and feature visualization to improve model performance.
- Trained and compared multiple machine learning models (Decision Tree, Logistic Regression) with hyperparameter tuning using Grid Search and Randomized Search.
- Deployed an interactive web application using Streamlit for real-time water potability predictions.

Movie Predict

https://github.com/ahmedali11a/movie_details

A machine learning application for predicting movie revenue based on production and performance features

- Built an end-to-end predictive pipeline using budget, popularity, runtime, language, genres, company, country, and ratings as input features.
- Applied data cleaning, outlier detection, feature scaling, and categorical encoding to improve model performance and generalization.
- Trained and compared multiple ML models (Random Forest, Gradient Boosting, Polynomial Regression), achieving up to 0.89 R^2 (train) and 0.84 R^2 (test).
- Developed an interactive Streamlit web app with real-time predictions, ROI calculation, and dataset exploration through visualizations and model insights.

COURSES & CERTIFICATES

Data Manipulation with Pandas

Data Manipulation with Pandas from DataCamp, providing hands-on experience in data manipulation using Python

Introduction to Python

from DataCamp, covering the fundamentals of Python programming with hands-on coding exercises.

Intermediate Python

from DataCamp, expanding Python skills with deeper knowledge of functions, scope, error handling, and best practices.

Exploratory Data Analysis in Python

Exploratory Data Analysis in Python from DataCamp, focusing on discovering patterns and insights in datasets using Python

Pandas

from Kaggle, focusing on preparing, cleaning, and analyzing real-world datasets with Pandas to support data science and machine learning workflows

Intro to machine learning

from Kaggle, focusing on applying machine learning techniques in Python with scikit-learn for classification and regression tasks

Supervised Machine Learning: Regression and Classification

from DeepLearning.AI & Stanford University (via Coursera), focusing on building and training regression & classification models, applying gradient descent and regularization, and evaluating model performance on real-world datasets.

Intermediate machine learning

from Kaggle, focusing on applying advanced machine learning techniques in Python using scikit-learn, including handling missing values, feature engineering, and building robust classification and regression models.

Languages

English (Intermediate)

Arabic (Native)