Data Scientist

AHMED ALI

(49) 1748908135|AhmedOAli177@gmail.com
https://www.kaggle.com/ahmedoraby/
https://www.linkedin.com/in/ahmed-ali-091892146/
https://ahmedoraby7.github.io/devportfolio/
https://leetcode.com/AOAli77/
https://github.com/AhmedOraby7
Germany, Aachen, 52074

Education

MASTER OF DATA SCIENCE – RWTH Aachen – Germany, Aachen — July 2023
Majors: Computer science, Data science, Machine Learning
Thesis: (BIG DATA) Data integration in a data lake system

BACHELOR OF COMPUTER SCIENCE – Alexandria University – Egypt, Alexandria — September 2017
Majors: Computer science

Skills

- **Programming languages**: **SQL** (SQL Server, MYSQL, PostgreSQL), **Python**, **Java**
- **Machine Learning**: Statistical Modeling, Language Models (BERT, DeBERTa, RoBERTa, ELECTRA), XGBoost, A/B Testing
- **Deep Learning:** DNN, GANs, RNN, LSTM, GRU, VAEs
- **Cloud:** GCP, AWS
- **Python Libraries**: Pandas, NumPy, SciPy, MatPlotLib, Scikit-learn, XGBoost, NLTK, PySpark, PyTorch
- **MLOps**: TFX, Kubeflow, FastAPI, Docker, SHAP, CI/CD
- **Other Tools**: Tableau, Git

Work Experience

DATA SCIENTIST – DXFACTURE – Germany, Aachen — February 2023 - Present

- Developed a comprehensive **Tableau** dashboard for SweetConnect, showcasing CO2 emission and product pricing data, enhancing the company's ability to make environmentally conscious and cost-effective decisions.
- Conducted correlation and regression analysis to investigate the impact of external temperatures on factory machinery power consumption, thereby enhancing power usage estimates through accurate weather forecasting.

SOFTWARE ENGINEER – Exporto – Germany, Aachen — April 2022 - November 2022

- Optimized **SQL** queries, significantly enhancing the user experience for Exporto customers.
- Designed and developed a global category search feature, boosting parcel processing efficiency for warehouse staff.

SOFTWARE ENGINEER INTERN – Amazon Web Services – Germany, Berlin — September 2021 - February 2022

- Designed, implemented (**React**), and tested (**Cypress – Enzyme**) advanced filters for AWS QuickSetup services, boosting customer engagement.
- Streamlined the development process by implementing a new workflow for **AWS Synthetics**, reducing the team's development time by 5% by enabling local testing without deploying code to gamma first.

DATA SCIENTIST – EJADA – Saudi Arabia, Riyadh — July 2017 - March 2020

- Leveraged machine learning models to detect unusual tax filing patterns, providing an early warning system for potentially **fraudulent** activities, thus ensuring the integrity of tax collection processes.
- Developed significant **Tableau** dashboards, providing insights on tax payments across various groups that represented **millions** in Saudi Riyal, influencing strategic decision-making.
- Acted as an external resource for the data engineering team, successfully delivering 20+ **ETL** packages to production using Microsoft **SQL** and SSDT, contributing to the efficient management and transformation of data for enhanced business analytics and decision-making processes.

Selected Projects

PREDICTIVE RISK MODELING FOR PROPERTY TITLES Link June 2023

- Developed a predictive machine learning model using Random Forest with Principal Component Analysis (PCA) to estimate the overall title risk for properties, based on property information and historical title defect data.
- Achieved a precision of 0.879 and recall of 0.889, providing a robust basis for making profit-maximizing recommendations for decision threshold adjustments during production.

CHURN-PREDICTION Link May 2023

- Analyzed extensive user data to predict user churn, a crucial metric in the gaming industry. Utilized exploratory data analysis and feature engineering to develop significant features, and applied machine learning models like Logistic Regression and Random Forest, achieving a high accuracy of 96% in churn prediction.

ESTIMATE DELIVERY TIME Link May 2023

- Conducted **exploratory data analysis** to identify key trends, including peak demand times and differences in delivery times across various markets, thereby gaining insights into factors influencing delivery times.
- Implemented machine learning algorithms like **Linear Regression** and **Random Forest** to predict delivery times, optimizing model performance through **feature selection** and hyperparameter tuning.
- Utilized statistical methods such as **best subset selection** and **p-value** calculations to identify and select the most relevant predictors for the model, improving model accuracy and interpretability.

SENTIMENT ANALYSIS ON REVIEWS Link February 2023

- Utilized **Python** to perform sentiment analysis on 14,000 reviews on IMBD.
- Applied deep learning to produce a final model with an accuracy of 88%

A/B TESTING EMAIL-SIGN-UP Link December 2022

- Designed and executed an **A/B test** for Urban Wear's pre-launch email sign-up page, using **Python** and **statistical methods** to compare the effectiveness of different submit button colors.
- Analyzed test results to provide data-driven recommendations, enhancing decision-making processes and maximizing email collection efforts.

MUSIC RECOMMENDATION ENGINE Link December 2022

- Designed and **developed** a personalized music recommendation engine, leveraging public Spotify datasets and a **machine learning (KNN)** algorithm, and utilized **Python** for **data cleaning** and **scraping**.
- Hosted the application on Streamlit, enabling personalized music recommendations for users.

Achievements

- Unsupervised Training of Language Models (Seminar) Link
- Machine Learning Engineer specialization (Coursera)
- 3rd place at Meta hackathon 2021 (3/10 teams)
- NLP Specialization (Deeplearning.ai)
- NLP Nanodegree (Udacity)
- Leetcode Monthly badge (3x)