Saleh Alkhalifa, M.S.

Data Scientist • ML Engineer • Project Manager

Data Scientist, Machine Learning Engineer and AWS Cloud Architect with 5+ years' experience in the science and technology sectors focusing on AI, ML, NLP and DevOps.

Experience

Amgen, Boston MA

Senior Data Scientist, Team Lead – OTDS Advanced Analytics — 08/2022—Present

- Developed **roadmaps**, set **goals** and defined **metrics** for the group to set our strategy and vision and align with Amgen operational goals while promoting collaboration, respect and inclusion.
- Led a team of 3 data scientists and contributed to developing novel **deep learning** models to translate unstructured data to useful insights for the Predictive Technologies business unit.
- **Developed** and **deployed** novel ML engineering methods using **event-driven architectures** within **AWS** (Sagemaker, API Gateway, Lambda, S3)
- Developed robust **APIs** (Flask, Nodejs, Express) to cater to custom business cases requiring complex logic relating to real-time data processing.

Data Scientist, Team Lead – Operations Advanced Analytics — 02/2020—08/2022

- Led the development of novel **NLP** semantic **searching** tools (BERT, **Transformers**, Similarity Searching) to move our organization from data-rich to decision-smart.
- Developed **anomaly detection LSTM** models using customer data to **discover defects**; presented results to technical and non-technical stakeholders at various levels.
- Developed, tested, and implemented large-scale **ETL pipelines** using Python, Spark, and SQL to support projects relating to product complaints, customer data, and health communications.
- Partnered with tech and business leads to gather requirements, identify opportunities, and solve problems with ML & NLP models (**LSTM**, **ARIMA**, **AutoEncoder**, **RNN**, **XGB**, etc.).
- Built and maintained dashboards (**Dash**, **Tableau**) and reports for **stakeholders** allowing them to gain **insights** at scale and make data-driven decisions based on performance and **metrics**.
- Established **best practices** to scale up predictive models within Amgen to increase discoverability and maintain integrity of data in millions of healthcare documents.

Sr Associate Data Scientist – Digital Integration and Predictive Technologies — 06/2019—02/2020

- Led a team of 2 data scientists developing new forecasting, **ML** and **NLP** models.
- Developed and deployed scalable **APIs** using **Java**, **Javascript**, and **Python**.
- Developed **business intelligence NLP** monitoring platforms to detect relevant content (linguistics) in **IP** & **patent** filings of external competitors and advised on internal strategies.
- Designed, developed, and deployed a platform to track and report **technical** and **business metrics** to the data science senior leadership team through automated data pipelines.
- Engineered **multivariate statistical** methods and stochastic models (including **A/B** testing) for anomaly detection operating with a 94% accuracy (patent pending).
- Collaborated with software engineers to **validate** and deploy models to **production**.

Associate Data Scientist – Attribute Sciences — 11/2017—06/2019

- Developed a wide range of machine learning models to address unmet scientific and business needs using **Keras**, **Tensorflow**, **Spark**, **Pandas** and **Sklearn**.
- Developed a timeseries forecasting model to predict instrument LCAP operating at **88% accuracy** to reduce development timelines.

NovaLyse Biosolutions, Villanova PA

Research Assistant — 11/2015—08/2017

- Developed a wide range of ML (Machine Learning) models and MD (Molecular Dynamics) simulations to predict chemical properties using Python, MATLAB, Keras, and RDKIT.
- Developed, purified, and analyzed hundreds of Quaternary Ammonium Compounds (QACs) to design and generate machine learning datasets.

Education

Northeastern University, MS, Computer Science (Virtual, Part-Time)		2019-Present
Villanova University, MS, Computational Science		2015-2017
Emmanuel College, BS, Computational Chemistry		2011-2015

Recent Publications / Patents

- Multi-objective Deep Data Generation w/ Correlated Property Control, SIGKDD **2022**.
- Machine Learning in Biotechnology and Life Sciences, Book, *Packt*, **2022**.
- Intelligent Mitigation of Equipment Performance Deficiencies, Patent, *USPTO*, **2021.**
- Analysis of the Destabilization of Bacterial Membranes, Article, *CBC*, **2019**.
- Quaternary Ammonium Compounds with Multiple Cationic Moieties, Article, *CBC*, **2019**.

About Me

Phone
+1 (617) 971-6559

Email
salehesam@gmail.com

LinkedIn
linkedin.com/in/saleh-alkhalifa

Github
github.com/alkhalifas

Website
www.salehalkhalifa.com

Technology

Python / JavaScript
●●●●●

AWS / GCP
●●●●○

Tensorflow / Keras
●●●●○

NodeJS / Express
●●●●○

React / React Native
●●●○○

Skills

A.I. / Deep Learning / NLP

M.L. Engineering / Development

Big Data / Analytics

AWS / Cloud Architecting

Project Management

Agile / Scrum / Jira

English / Arabic / German

Certifications

Certified Scrum Product Owner

The Machine Learning Pipeline on Amazon Web Services, AWS

Managing Complex Technical Projects, MIT

Cloud Practitioner Essentials, AWS

Data Science and Big Data Analytics, MIT