

AbderRahman Sobh

169 Charles St
Sunnyvale, Ca 94086
(217) 979-4127
abbysobh@gmail.com

EXPERIENCE

Manager/Tech Lead, Machine Learning Engineering – Earnin, Inc

April-2019 – Present

Engineering Tasks:

Architected an in-house platform for auto-deployment of machine learning models, reducing model training/deployment times by at least 10x.

Designed and developed multiple Kubernetes microservices providing real-time, business-critical ML model predictions with 24/7 uptime.

Management Tasks:

Set OKRs and initiatives for the team, resulting in focused projects which provide maximum impact to the company's objectives.

Provided 1:1 mentoring to all team members for specific uses of applied Data Science, Machine Learning, and Data Engineering.

Planned and supported individual career growth and knowledge growth paths for all reports.

Set clear standards for engineering quality and practice across the team's codebase.

Freelance Data Scientist – Self Employed

June-2018 – Present

Data Science/Engineering Tasks:

Developed tools for private clients: ETL construction for data ingestion, NLP Tagging for automated Topic Detection, analytics/forecasting tools for insights.

Authored a set of learning materials covering the Data Science workflow: Data Harvesting, Data Cleaning, Feature Engineering, ML Algorithms, NLP Algorithms.

Provided 1:1 guidance with private clients and students for specific uses of applied Data Science.

SKILLS

Python,
Kubernetes,
Spark, SQL,
Docker, Bash,
AWS (EC2,
Lambda, S3,
Sagemaker, EMR,
Redshift),
NoSQL, MySQL,
Tableau, R, SAS,
Matlab,
Mathematica,
C, C++, TCL,
VBA/Excel,
Fortran90,
HTML, PHP,
JavaScript, CSS,
XML

GROUPS

Data Scientist
Engineers Project
(2016-2017) –
Burlingame, Ca

New York Open
Statistical
Programming
(2013-Present) –
Manhattan, NY

LANGUAGES

English, Arabic,
Japanese

Lead Data Scientist – Life 360, Inc.

Dec-2017 – June-2018

Data Science Tasks:

Lead development of a high-level Roadmap of Data Science projects using User Location Data, Driver Behavior Data, App Usage Behavior Data, and various internal platform metrics.

Developed a pipeline for refining location data, enriching it with additional insightful transformations, and storing it in a format readily available for predictions.

Built a tool for predicting and ranking the most relevant user locations based on user/group behavioral history (Location clustering, time series analysis).

Product/Project Management Tasks:

Planned in-app delivery of location prediction as a product which suggests interesting events and locations to users.

Aligned product delivery within company vision and user-persona relevance in order to maintain feature consistency and a familiar user experience.

Coordinated between departments to secure resources and review proposed initiatives.

Provided reasonable ticket scoping and timelines for objective completion.

Data Scientist – AJ+ / Al Jazeera International

Dec-2016 – Dec-2017

Data Science/Data Engineering Tasks:

Planned a comprehensive Roadmap of potential Data Science projects in Social Media

Developed Data Science tools from end-to-end: Product concept design, database/schema design, data pipeline development (i.e. ETL, Data Cleaning, Scoring/Ranking/Tagging, Weight Corrections), Dockerized pipeline deployment, insight delivery (Tableau dashboards, Slackbot).

Built a tool for live scoring of Social Media Engagement (i.e. Likes, Shares, Comments) for media produced by the company on Facebook, YouTube, Instagram, Twitter.

Authored and published an article under the AJ+ brand, documenting the scoring tool's creation process for the benefit of the public community:

<<https://medium.com/aj-platforms/re-thinking-engagement-at-aj-69a35e0a38c>>

Implemented tools for text-feature extraction using analysis of grammar used, content topic modeling, word vector embedding, and sentiment analysis.

Successfully overhauled legacy Data Science tools and inferences used by the company.

Collaborated with Data Engineers to define requirements, integrate pipeline development, and release tools into a production environment.

Software Engineer – University of Illinois at Urbana–Champaign

Jan-2012 - Dec-2016

Data Science Platform:

Designed and built distributed Data Science pipelines using Jupyter Notebooks, Python ML libraries, and Spark, to provide single-click analysis tools for educating future data scientists on the use of NLP and Text feature extraction.

Collaborated with Full Stack developers to integrate data science pipelines using Docker containers to encapsulate data science software stacks which consistently provide end-users with quick usage of data-science tools and visualizations.

Simulation Workflow Development:

Integrated software applications within openVZ containers providing rapid deployment and accessibility of custom built applications.

Developed GUIs using TCL and RAPPTURE (XML framework) for containerized applications to ease application usage for users.

Optimized simulation memory-handling and runtime on millions of data points using Python, TCL by using associative arrays/hash mapping.

Connected application workflows to remote cluster systems and utilized parallel threading on serial procedures to accommodate scaling.

Integrated visualized workflow results using the VTK toolkit for robust manipulation on custom displays.

EDUCATION

University of Illinois at Urbana–Champaign *Bachelors of Science in Statistics*

August 2007 - May 2012

Dean's Honor List (2007–2008), (2012)

PROJECTS

Data Science Tutor / Kaggle Competition Advisor — Mentor (2018)

Kaggle continues to be one of the best places for Data Scientists to find datasets to investigate. As a result, I combined a tutoring program with solving some of the challenges available on Kaggle. During the course of study, we covered topics such as: “conceptually approaching data’s potential”, “data cleaning, feature analysis, and feature engineering”, “picking the right model to use”, and “explaining the final model in a business setting”.

PORTFOLIO

<https://itg-abby.github.io/portfolio/>