SEPEHR SADIGHPOUR

SENIOR DATA ENGINEER

hi@sepehr.xyz
(510) 759-4850
Oakland, CA

Profile

I enjoy helping socially-conscious organizations deliver high-quality insights using data and machine learning. My experience as a data engineer, ML engineer, and web developer, enable me to understand complexities of data and ML pipelines across the entire stack.

Strengths

ML DevOps

Creating workflows for data scientists to rapidly train, interpret, productionize, and monitor ML models.

Data Engineering

Building ETL and ML pipelines using Spark and SparkML. Empowering analysts to validate data quickly and surface insights.

Data Science

Implementing appropriate ML algorithms to solve a variety of business challenges.

Experience

Clarify Health Solutions — Nov 2017 - Mar 2020

Senior Data Engineer

Built Spark pipelines for validating, feature engineering, and training models on 10s of billions of rows of healthcare claims data.

Sped up model training 20x, enabling scaling to thousands of models per day.

Architected an extensible, scalable, and efficient service for training and organizing hundreds of thousands of models, serving predictions and model interpretations in batch and real-time modes. Managed external team of developers in charge of implementation.

Designed statistical metrics for detecting data drift.

Omada Health — Apr 2016 - Jul 2017

Data Scientist

Guided product direction via analyses of trends in participant behavior, leading to 6% lift in user engagement

Designed and implemented A/B tests for new features in close collaboration with designers

Managed 4-person data team's research into areas such as chatbots, reinforcement learning, and machine vision

BayThrive — Jan 2010 - Apr 2016

Owner, Developer, Project Manager

Increased traffic 20x and conversion rates 4x for biotech sales firm through SEO and iterative design

Built development and production infrastructure for early-stage startup using Docker and Ansible

Developed heuristic for financial research firm to estimate degree of self-similiarity in time-series data

Mepkin Abbey — Jun 2008 - Jan 2010

Webmaster, Go-To Guy

Directed development of custom registration system from need-assessment to hiring and project management

Built website, e-commerce solution, and established processes for inventory management and online sales

Education

Galvanize University — Oct 2015 - Jan 2016

Data Science Fellow

Capstone Project: Built information-retrieval app for discovering scientific articles by semantic similarity using document vectors

Duke University Class of '08

Bachelors of Science, Physics

<u>Capstone Project</u>: Trained neural networks and genetic algorithms to solve energy states of quantum systems

---

Technical

<u>Modeling</u>: Spark ML, scikit-learn, H2O, Vowpal Wabbit, LIME

<u>NLP</u>: Spacy, gensim, fastText, WordNet, NLTK

<u>ETL</u>: Spark, Airflow, Luigi

<u>Infra</u>: Terraform (AWS EMR, Batch, ECS, S3, Redshift), Docker

<u>Software</u>: Flask/Django, PHP, git, bash, RESTful API design, some JS

<u>Analysis</u>: Pandas, R, SQL (Postgres, Presto)

<u>Experimentation</u>: Experiment design, multi-armed bandits

<u>Viz</u>: Plotly, leaflet.js, seaborn, some D3.js