Brian Lewis
brianburkelewis@gmail.com
Oakland, CA
tetraptych.github.io

Summary

I'm a passionate, experienced, and results-oriented data scientist with a proven ability to deliver on high-impact initiatives, from developing personalized data-driven products to implementing ML models end-to-end using the latest technical frameworks. Expert in the entire Python DS stack with significant experience in Spark.

Professional Experience

Mindstrong Health
Senior Data Scientist 2020-present

- ❏ Built and deployed machine learning models to help our clinical team provide therapeutic and psychiatric services to under-served populations
- ❏ Estimated clinically-meaningful quantities (sleep, physical activity) from passively-collected phone data using approaches adapted from the biometric research literature
- ❏ Helped junior members of the team deploy their models at scale, promoting a culture of collaboration and best programming practices throughout the data organization

KeepTruckin
Data Scientist 2019-2020

- ❏ Designed, validated, and deployed ML models on massive, disaggregated datasets produced by IoT devices installed in 200,000+ vehicles
- ❏ Conducted experiments to understand user behavior and improve engagement outcomes
- ❏ Analyzed the feasibility of new product offerings and delivered my findings to the executive team, quantifying the opportunities to guide product strategy

Bayes Impact
Data Scientist 2017-2019

- ❏ Led the research arm of the organization, investigating disparities in access to social services
- ❏ Designed and implemented models, algorithms, and APIs to support in-house initiatives to bring evidence-based decision making to government policymaking
- ❏ Worked in partnership with the Centers for Medicare & Medicaid Services to measure physician quality from the nation's largest claims data warehouse as part of a congressionally-mandated effort to implement value-based care at a national scale

MedeAnalytics
Data Engineer 2014-2016

- ❏ Developed two web applications and backend databases for a large state Medicaid agency:
  - ❏ A big-picture analytics dashboard tracking important public health issues and agency costs
  - ❏ A portal for health professionals to view patients' health information at the point of care
- ❏ Overhauled the ETL process for Medicaid claims to dramatically reduce data latency

Education

- ❏ 2014 Masters of Arts, Mathematics, University of Wisconsin-Madison
- ❏ 2012 Bachelors of Science, Mathematics, Stanford University

Awards

- ❏ 2014 Mathematics Graduate Teaching Award, University of Wisconsin-Madison
- ❏ 2012 Phi Beta Kappa Inductee, Stanford University

Professional Development

- ❏ 2016 Data Science Immersive, Galvanize Academy

Technical Skills & Proficiencies

Languages

- ❏ Expert-level Python; proficient in Java, C++
- ❏ Expert in Spark, both via PySpark and Spark SQL
- ❏ SQL (Snowflake, Redshift, Vertica, Postgres, Hive) and NoSQL (MongoDB, DynamoDB)

Machine Learning & Statistics

- ❏ pandas, numpy, scikit-learn, statsmodels, scipy, seaborn
- ❏ SparkML, XGBoost, PyTorch, BERT via 🤗 transformers
- ❏ regression, classification, clustering, hyperparameter tuning, feature selection
- ❏ hypothesis testing, bootstrapping, cross-validation, Bayesian inference
- ❏ statistical data analysis and causal inference

Miscellaneous

- ❏ AWS ecosystem, Docker, Kubernetes
- ❏ GIS and spatial algorithms, mapping and data visualization (Looker, Tableau, Redash)
- ❏ Data modeling, API design, high-performance Python, web-scraping, multiprocessing