

ANJALI SHRIVASTAVA

vastava.github.io • anjali.shrivastava99@gmail.com • (phone number redacted)

SKILLS

Programming Languages and Frameworks

Python (PySpark, Pandas, NLTK, NumPy, scikit-learn, matplotlib, plotly, seaborn) • R (ggplot2, dplyr) • Javascript (d3, Angular, React) • CSS (Bootstrap) • HTML • Java • SQL • Scala • MATLAB

Database Management

MySQL • Postgres • Redshift • DBeaver • AWS Athena

Other Software and Tools

Cit • Databricks • Jupyter Notebooks • IntelliJ/PyCharm • Tableau • Qlik Sense • Microsoft Excel • MLFlow • Hyperopt

Statistics

Experimental Design • Bayesian Statistics and Probability • Time Series Analysis • Econometrics

EDUCATION

University of California, Berkeley

August 2017 - May 2021

- B.A. Data Science (with a concentration in Geospatial Information and Technologies)
- B.S. Industrial Engineering and Operations Research

Relevant Courses

Data Structures, Algorithms, Machine Learning, Databases, Stochastic Processes and Markov Chains, Principles of Data Science, Nonlinear and Discrete Optimization, Forecasting, Linear Algebra

Cumulative GPA: 3.7

THANKS FOR READING!

ABOUT

Product-minded data scientist with experience in e-commerce and FinTech industries. Capable of creating, deploying and validating scalable machine learning models and/or data pipelines in Python, PySpark and SQL. Passionate about data visualization and finding the 'story' in the data. Also have a knack for scraping data, and generally tinkering with messy tables. [Check out my online portfolio!](#)

EXPERIENCE

Thermo Fisher Scientific — *Data Scientist II, Digital Commerce*
July 2021 - present

- Improved market basket product recommendations by introducing sequential pattern mining to determine association rules—increased model output to 85% of SKUs sold within past year, compared to previous rate of 50%
- Built click-through-rate prediction model to better enhance our recommendation engines on digital storefronts
- Used MLFlow API to capture post-deployment metrics and evaluate model performance; since implementing, team has caught >5 bugs

HiGeorge — *Data Engineer and Visualization Specialist*
September 2020 - July 2021

- Built data pipelines to publicly available data and APIs in Python
- Developed live, updating data visualizations in d3 and Angular

Elin.ai — *Data Science Intern*
September 2019 - December 2019

- Used NLP to analyze correspondence (ie. Slack messages, emails) to measure company culture and employee satisfaction levels
- Performed validation of sentiment analysis model

Intuit — *Business Systems Analyst Intern, Finance*
May 2019 - August 2019

- Worked on development of interactive website that features daily KPIs and charts using React, d3
- Led initiative to redesign dashboards in Tableau and Qlik Sense to make them more user-friendly

RESEARCH

Human Rights Center
September 2019 - May 2021

- Used OSINT and web scraping techniques to archive online evidence of human rights abuses; performed data investigations for Amnesty International and other human rights organizations

D-Lab at Berkeley

January 2018 - January 2019

- Worked with PhD candidates and researchers in the history department to supplement their social science research with data analysis