JASON LU

Data Scientist • 626 354 3033 • jlu7922@gmail.com • linkedin.com/in/jasonflu/ • github.com/jlu7922

PROFESSIONAL PROFILE

As a data scientist with a background in computational neuroscience, I am passionate about leveraging data to uncover insights and drive innovation. Through my expertise in data analysis, statistical modeling, and machine learning, I am dedicated to solving complex problems and empowering teams to make informed decisions. My experience in computational neuroscience provides a unique perspective and skill set that enables me to understand and model complex biological systems, enabling me to create innovative solutions that optimize outcomes. With a focus on collaboration, communication, and continuous learning, I strive to transform data into valuable insights for businesses and organizations.

SKILLS

Programming: Python, C++, R, NumPy, SciPy, Pandas, PySpark, Data Structures and Algorithms, Jupyter
Data Analytics: SQL, Hadoop, MapReduce, Scikit-learn, AWS, MongoDB, Tableau, SAS, SAP BI, Database systems
Statistics: Statistical Inference and Data Analysis, Structural Bioinformatics

WORK EXPERIENCE

Data Science Immersive Fellow   03/2023 – 06/2023
General Assembly
Successfully completed 500+ hours of expert-led instruction in Python, Data Visualization, and statistics, with hands-on learning of data manipulation fundamentals and the industry's most in demand technologies. Developed projects, including:

- **NLP Classification on Reddit Posts** - Utilized web scraping through the Reddit Pushshift API to collect a comprehensive dataset of over 10,000 subreddit posts. Applied Natural Language Processing techniques to process and analyze the text. Build a classification model for accurate prediction of a subreddit post's origin.
- **Predicting Housing Price** - Developed multiple linear regression models to accurately predict sale prices for homes in Ames, IA within a $30,000 margin of error.
- **Standardized Test Analysis** - Analyzed how location and intended college major of high school students majorly correlates with projected SAT/ACT performance, utilizing the Pandas and Matplotlib data analysis libraries.

Web Database Developer   05/2021 – 09/2021
Project Compass (New York, NY)

- Designed interface for backend to server data interaction on a website built for analyzing and communicating how children with autism solve problems.
- Created procedural learning algorithm from ground-up utilizing AWS and Google's Speech to Text and Text to Speech API to determine language processing of users.
- Implemented techniques like KNN, logistic regression, and SVM to generate word-pairs for text matching problems.

Machine Learning Engineer Intern   11/2020 – 01/2021
Siemens PLM (Los Angeles, CA)

- Successfully designed and implemented an API for accessing and converting XML file data into CSV files to aid root cause data mining and to prepare datasets for training machine learning models.
- Improved an internal configuration tool to help users solve validation failures within Siemens NX software
- Optimized model to better predict error-prone features and provide insight for automating solutions

EDUCATION

Bachelor's Degree in Computational Neuroscience, Minor Degree in Data Science   Fall 2023
University of Southern California, Los Angeles, CA