zmosleymis@gmail.com
(209) 487-4826

Zack Mosley

Portfolio
LinkedIn

Summary

Graduate student in Data Science at UC San Diego with hands-on experience in machine learning, statistical modeling, and distributed computing gained through coursework and applied projects. Built predictive models and recommendation systems using Python, scikit-learn, and Apache Spark. Seeking an entry-level data scientist role applying ML techniques to solve business problems.

Education

M.S. in Data Science 2024 - 2026
University of California, San Diego La Jolla, CA
Graduating: March 2026 | University GPA: 3.80 /4.00

B.S. in Management Information Systems 2022 - 2024
California State University, Sacramento Sacramento, CA
University GPA: 3.80 /4.00 (Magna Cum Laude)

A.S. in Accounting | A.S. in Business Administration 2020 - 2022

Projects

Car Value Predictor (2024) (Personal Project, View Project)
Pricing model to support used-car valuation decisions by estimating market prices across 10 major manufacturers using live marketplace data.

- Collected and standardized 84K+ vehicle listings via custom web scraper and automated pipeline for cleaning, outlier handling, and feature engineering
- Selected LightGBM after benchmarking regression approaches, achieving MAE $1.7K (6.3% of average price) with $R^2 > 0.96$ across manufacturers
- Validated robustness on new data collected one month later, maintaining $R^2 > 0.95$ despite market volatility
- Delivered interactive Dash dashboard with price distributions and SHAP-based explanations to support transparent pricing decisions

Amazon Sports & Outdoor Products Recommender System (2025) (Academic Project, View Code)
Hybrid recommendation system designed to surface relevant products from a sparse 4M+ Amazon review dataset.

- Mitigated 99.998% sparsity by combining collaborative filtering (SVD) with content-based features (Ridge regression) to improve cold-start performance
- Benchmarked five recommendation strategies and optimized hybrid weighting (60% SVD, 40% Ridge) to balance personalization and coverage
- Achieved strong ranking quality (nDCG@10: 0.9981) with 31.5% error reduction over baseline methods

NYC Yellow Taxi Fare Prediction (2025) (Group Project, My Contributions | View Code)
Fare prediction model built from 150M+ NYC taxi records, earned 3rd place in class competition.

- Engineered time/efficiency features from 20GB+ trip data using Apache Spark and Dask
- Trained regressor achieving MAE $1.91 and R^2 0.94, improving RMSE by 49% over baseline
- Scaled training with distributed computing (15 workers, 60 cores) and validated performance using residual analysis

Skills

Programming & Libraries: Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn), SQL, JavaScript (D3.js)
ML & Analytics: Regression, Classification, Recommendation Systems, Hyperparameter Tuning, A/B Testing
Data Engineering: ETL Pipelines, Apache Spark, Dask, Data Cleaning, APIs
Tools & Platforms: Git, Tableau, Excel, Azure

Experience

DBI Beverage 2014 - 2020
Lead AM Warehouse Associate Stockton, CA

- Managed 7-person team distributing 50K+ cases/day for Coors, Modelo, and Corona brands
- Spearheaded new Warehouse Management Software (VIP EasyOps) implementation, partnering with vendors to modernize workflows
- Improved inventory accuracy 12% through better labeling, FIFO procedures, and discrepancy investigation