JARED CHOY

Alamo, CA 94507 | jaredchoy9@gmail.com | 925.822.6465 | github.com/jared-choy

ASPIRING DATA SCIENTIST

Master's (MS) in Statistics with 2+ years of experience in machine learning, predictive modeling, and data science. Expert in Python, PyTorch, and statistical analysis with proven success building and deploying ML algorithms for business optimization. Skilled in extracting insights from large, complex datasets and presenting actionable recommendations to stakeholders. Background in cloud-based ML solutions, cross-functional collaboration, and translating data science into real world insights.

Data Analysis • Data Cleaning • Data Processing • Data Mining • Data Management • Data Warehousing • Data Engineering • PyTorch • ETL • Business Acumen • Azure • A/B Testing • Key Metrics • Root Cause • Machine Learning • SQL • Python • R • Supply Chain • Excel • Tableau • Power BI • Looker • Pipelines • Hypothesis Testing • Ad-hoc • API • Process Improvement • Survival Analysis • Git • Information Systems • Scripting • SAS • React • Classification • Modeling • TensorFlow • AWS • Statistical Analysis • Regression Analysis • Time Series Analysis • Predictive Modeling • Prescriptive Modeling • NoSQL • PostgreSQL • Spark

EDUCATION

Masters of Science (MS) in Statistics | Emphasis in Data Science | University of California, Davis | December 2024
Bachelor of Science (BS) in Statistics | Statistical Data Science Track | University of California, Davis | June 2023

Relevant Coursework: Applied Statistical, Regression Analysis, Nonparametric Statistics, Introduction to Probability Theory, Computational Statistics, Data & Web Technologies, Statistical Methods in Machine Learning, Applied Time Series Analysis

PROFESSIONAL EXPERIENCE

***Sacramento Area Council of Governments (SACOG)* | Sacramento, CA** **April 2024 – April 2025**

Research and Data Intern

- Partnered with city and county leadership teams to understand data requirements and standardize datasets across multiple agencies
- Built automated ETL pipelines and data processing solutions in Python, improving data consistency and accessibility for decision-making
- Analyzed large-scale public health and infrastructure datasets to identify population trends, service gaps, and community needs
- Created interactive Tableau dashboards and presented findings to internal teams, planners, and community stakeholders
- Proposed data-driven strategies to address regional planning challenges and improve public service delivery

***Vuemix* | San Jose, CA** **May 2023 – March 2024**

Software Development and Machine Learning Intern

- Designed and implemented machine learning algorithms using Python (PyTorch, scikit-learn) to identify business optimization opportunities in product development.
- Built and deployed predictive models on AWS and Azure cloud platforms, creating scalable ML solutions for client-facing applications.
- Manipulated and cleaned large datasets to develop classification and prediction models that improved business decision-making processes.

***UC Davis School of Education* | Davis, CA** **October 2023 – January 2024**

Student Research Data Assistant (Seasonal Role)

- Cleaned, managed, and organized multi-year educational and demographic data for program evaluation.
- Translated complex education metrics into visual and written summaries for leadership and community outreach.

KEY PROJECTS

- **NBA Draft Pick Career Success Correlation Analysis**: NBA Draft Pick Career Success Correlation Analysis: Utilized Regression Modeling, VIF, and Hypothesis Testing techniques to merge and analyze player and award data (1947–2021), exclusively coding the project in R and presenting findings through data visualizations.
- **Restaurant Recommendation System**: Developed a collaborative filtering-based restaurant recommender system using Python and scikit-learn. Built similarity matrices, tuned hyperparameters, and tested model performance with RMSE metrics. Demonstrated model outputs using Tableau dashboard to support UX decisions.
- **Computer Vision Art Analysis**: Performed mixed-methods analysis with color analysis techniques (GrabCut algorithm, K-means clustering) to investigate Warhol's themes around media influence and fame.
- **UC Transfer Admittance Analysis:** Used web scraping to analyze UC admissions data by ethnicity and major. Conducted non-parametric tests, logistic regression, KNN, and decision trees to explore GPA and ethnic diversity in admission rates.
- **Sentiment Analysis of Domestic Airline Tweets:** Achieved 90% accuracy with a BERT model on airline tweets. Used NLP techniques, Elastic Net, SVM, and RNNs/CNNs to handle sentiment classification.
- **MNIST Handwritten Digit Classification and Augmentation**: Executed logistic regression, SVM, and KNN for MNIST digit classification, applying dimensionality reduction and a KDE-based bootstrap pipeline for data augmentation.
- **Optimal Bounds for Eigenvalue Estimation**: Researched the Power and Lanczos algorithms for eigenvalue approximation, analyzing performance metrics (ARE, PRF), computational sustainability, and simulating results with random initialization.