JohnMatthew Garcia
Machine Learning Engineer | Data Scientist | UC Berkeley Graduate
jm.garcia@berkeley.edu | (415) 802-4846 | https://johnmgarcia.github.io/

EDUCATION

University of California, Berkeley - College of Computing — **Fresno City College**
Bachelor's in Data Science - BA — *Associate in Economics - AA*

Selected Coursework: **CS:** Structure of Computer Programs • Data Structures and Algorithms • Introduction to Artificial Intelligence • Internet: Architecture and Protocols • Machine Learning **Statistics:** • Foundations of Data Science • Techniques of Data Science • Probability for Data Science • Data Mining and Analytics • Data Engineering • Machine Learning and Data Analytics • Intro to Business Analytics

WORK EXPERIENCE

UC Berkeley – Berkeley IT - *Network Engineering Assistant* *Aug 2024 – Present* · 9 mos Berkeley, California

• **Performed** hands-on network support across campus by replacing switches, restoring WiFi, configuring ports and IP addresses, and troubleshooting connectivity issues. Utilized Python/Bash scripts, conducted preventive maintenance on equipment. • **Collaborated** with HelpDesk, IST-NOS, and vendors to resolve technical issues and maintain documentation.

UC Berkeley – D-Lab - *Undergrad Technician* *Jan 2024 – Present · 1yr 4 mos* Berkeley, California

• **Provided Cross-Disciplinary Services**: Delivered in-depth consulting, advising, and access to staff support for software or infrastructure needs at UC Berkeley's D-Lab. Assisted graduate students, faculty, and staff in advancing world-class research in data-intensive social sciences and humanities

PROJECTS

UC Berkeley Sensor Data - Interpolation *PostgreSQL, JSON*
• Implemented a Hampel filter using SQL to identify outliers by calculating the median and median absolute deviation for each group, and applied winsorization by capping outlier values to the nearest boundary defined by 3 scaled MAD intervals.
• Performed interpolation by identifying data gaps, calculating start and end values for missing sequences using forward and backward filling, and generating interpolated values based on linear progression across the gap size.

Fielding Query Performance - Baseball *PostgreSQL, Database Engineering*
• Optimized query performance by setting appropriate indexes, applying filters, clustering data on indexed columns, and creating materialized views to reduce execution time and overall query cost.
• Developed queries from the Lahman's Baseball Database to find various salary metrics of MLB Baseball players; MLB player metrics from Cal and other CA-specific colleges; and player metrics with above 500 "At Bats"
• Analyzed the query optimizer's decision-making process, exploring various join strategies such as hash joins, nested loop joins, and merge joins, and explained how the optimizer's choice impacted query performance.

Disaster Tweet Classification using NLP *Python, Pandas, scikit-learn, NLTK*
• Preprocessed tweet text with NLTK, regex, and stopword removal to clean noisy real-world data.
• Trained and compared logistic regression, random forest, and SVM classifiers on labeled tweet dataset to detect disaster-related content. Achieved F1 score of 0.75 using SVM model, demonstrating strong performance on imbalanced and unstructured data.

TECHNICAL SKILLS

Languages: Python (Pandas, Scikit-learn, TensorFlow, PyTorch), SQL (PostgreSQL), R, Java,
Machine Learning: Regression, Classification, Clustering, Ensembles, Bayesian & Reinforcement Learning, PCA, K-Nearest Neighbors, K-Means, Ordinary Least Squares
Data Science: Data Wrangling, EDA, A/B Testing, Causal Inference, NLP
Databases: PostgreSQL, MongoDB, Query Optimization, Data Modeling, ETL, DB Engineering
Tools & Platforms: Git, Jupyter, Hadoop, Airflow, Spark, Docker, Tableau, Power BI, Analytics Solver, Matplotlib, Seaborn

ADDITIONAL EXPERIENCE

Austin Boom Real Estate
Licensed Real Estate Agent *2008 – 2012 · 5 yrs* Austin, Texas
10+ years of professional sales and customer service experience across real estate, automotive, and hospitality industries.