

Connor Scott McManigal

Newport Beach, CA 92660 • (949) 630-5208 • mcmanigc@uci.edu • [linkedin.com/in/connormcmanigal/](https://www.linkedin.com/in/connormcmanigal/) • <https://connormcmanigalportfolio.netlify.app/>

EDUCATION

University of California Irvine • Irvine, CA

Dec 2024

- ❖ Master of Data Science (GPA: 4.0/4.0)

University of California San Diego • La Jolla, CA

June 2023

- ❖ B.S. Cognitive Science with Specialization in Machine Learning and Neural Computation
 - 4x Division I Water Polo Student-Athlete; 12x Provost Honors List (GPA: 3.8/4.0)

TECHNICAL SKILLS

Programming: Python, R, SQL, Java, Spark, BigQuery

Libraries: Pandas, NumPy, Seaborn, Matplotlib, SciPy, Scikit-learn, PyTorch, TensorFlow, tidyverse, ggplot2

Machine Learning: Supervised & Unsupervised Learning, Deep Learning, LLMs, Statistical Modeling

Tools: Git & Github, Jupyter, Databricks, Google Cloud Platform, Hadoop, Impala

Mathematics: Probability, Statistics, Hypothesis Testing, Linear Algebra

EXPERIENCE

Rockstar Games • Carlsbad, CA

(Ongoing) Jan 2025 - May 2025

Data Science / Data Analytics Intern • Player Insights and Analytics Team

- Partnering with security stakeholders to streamline cheating behavior detection and mitigation.
- Analyzing gameplay data to identify anomalies and patterns indicative of exploitative behavior.

Corelogic • San Diego & Irvine, CA

June 2024 - Sept 2024

Data Science and Analytics Intern • Data Science and Analytics Team

- Collaborated with DS&A leadership and product managers to deliver Department of Insurance reports.
- Used Python, BigQuery, and GCP to wrangle big data, validate resiliency models, and generate statistics, streamlining future filings, supporting sales and marketing, and enhancing revenue potential.
- Conducted R&D on LLM automation for tax file ingestion, applied prompt engineering techniques to optimize performance, and developed a proof of concept projected to save ~\$3-5 million annually.
- Presented insights to senior executives, sales teams, product managers, and the Chief Data Officer, translating technical solutions into business strategies.

American Medical Association • Newport Beach, CA

June 2022 - Sept 2022

Integrated Health Model Initiative Intern • Informatics and Digital Products Team

- Assisted CMIO and VP of IDPT in AI evaluation strategy and healthcare hardware development.
- Audited FDA-style AI evaluation methods, identifying risk framework improvements to ensure fair and transparent outcomes.
- Evaluated Sensely app UI, proposing better button placement and tutorials to enhance user engagement.
- Authored white paper on ML-driven patient engagement, advocating for pre-program surveys to improve data quality and predictions.

UCSD Basement Innovation Sprints Program • La Jolla, CA

Feb 2022 - May 2022

Machine Learning Intern at the San Diego Zoo • ZOOBER 2.0 Team Lead

- Led a team of five to enhance YOLO computer vision algorithm for 24/7 elephant enclosure monitoring.
- Partnered with zoo stakeholders to tackle nighttime adaptation, heat lamp noise, and enclosure coverage.
- Researched industry use cases and recommended solutions like custom datasets, image preprocessing, and toggle-enabled model splitting.
- Presented findings and recommendations at the Innovation Sprints Showcase.

PROJECTS

Internet Traffic VAE Network Attack Detection • Python Anomaly Detection *Mar 2024 - June 2024*

- Constructed variational autoencoders (VAEs) on 123,000 network traces to detect nine types of attacks.
- Developed a Bayesian optimization framework for tuning hyperparameters and trained four PyTorch VAEs, including baseline and Mixed-Loss models.
- Achieved mean area under the curve (AUC) of 0.81 with Mixed-Loss VAE, differentiating from similar studies by using a multinomial approach.

Recipe Review ML Scoring Enhancement • Python ML Sentiment Augmentation *Feb 2024 - Mar 2024*

- Utilized data augmentation to reconstruct recipe review scoring algorithm with limited data, leveraging sentiment analysis libraries VADER and TextBlob to extract review polarity and subjectivity.
- Employed scikit-learn to construct two sets of Multi-Layered Perceptrons and Gradient Boosting Regressors: one using original data and the other combining original and extracted features.
- Achieved a 3-unit reduction in mean absolute error (MAE), with augmented GBR delivering best performance (MAE: 21.44).

San Diego County 2021 Automobile Accidents • R Data Analysis *Mar 2023 - June 2023*

- Investigated the frequency and severity of automobile accidents based on days, months, seasons, and weather conditions, discovering that fair weather, winter, and fall had the highest accident rates.
- Utilized ggplot2 for analysis, visualizing monthly totals and accident distributions by weather and season.
- Applied OLS regression to analyze impact of weather and season on daily totals, finding strong estimates for winter and stormy weather, with adjusted R-squared of 0.43 indicating moderate explainability.

Diabetes Binary Classification Comparison • Python Supervised ML Comparison *Mar 2023 - June 2023*

- Compared performance of logistic regression, decision tree, random forest, K-NN, and SVM models for diabetes classification, utilizing grid and random searches for hyperparameter tuning.
- Evaluated models with recall, precision, accuracy, confusion matrices and ROC-AUC curves, finding that decision tree yielded the best performance with accuracy of 90.9% and false negative rate of 10%.

Effect of NBA Injuries on Team Record (2010-15 Seasons) • Python Data Analysis *Mar 2022 - June 2022*

- Analyzed the impact of injuries on team performance through data cleaning and merging, exploratory analysis, and OLS regression, revealing weak negative relationship between injuries and win percentage.
- Trained linear regressor to assess the relationship between total and returning injured players and win percentage, yielding RMSE of 10.86, underscoring the need for more comprehensive injury data.