Connor Scott McManigal

Newport Beach, CA 92660 • (949) 630-5208 • mcmanigc@uci.edu • linkedin.com/in/connormcmanigal/ •
https://connormcmanigalporfolio.netlify.app/

SUMMARY

➢ Master of Data Science student with strong analytical and problem-solving skills, a diverse computational background, and an insatiable passion and drive to learn.
➢ Experienced in data manipulation and visualization, statistical analysis, and machine learning to develop predictive models, generate data-driven insights, and drive actionable business strategies.
➢ Four year Division 1 scholar-athlete, honing leadership, communication, work ethic, and time management skills.

EDUCATION

UNIVERSITY OF CALIFORNIA IRVINE DONALD BREN SCHOOL OF ICS • Irvine, CA   *2023-2024*
❖ Master of Data Science (*Expected 12/2024*)
  ● *GPA: 4.0/4.0*
UNIVERSITY OF CALIFORNIA SAN DIEGO • La Jolla, CA   *2019-2023*
❖ B.S. Cognitive Science with specialization in Machine Learning and Neural Computation
  ● *GPA: 3.8/4.0; 12x Provost Honors List; 4x ACWPC All-Academic Award*

TECHNICAL SKILLS

Programming: Python, R, SQL (including BigQuery SQL), Java
Libraries: Pandas, NumPy, Seaborn, Matplotlib, SciPy, Scikit-learn, PyTorch, TensorFlow, ggplot2, tidyverse
Machine Learning: Supervised/Unsupervised/Deep Learning, LLMs, Generative AI, Statistical Modeling
Tools: Git/Github, Google Cloud Platform, Jupyter, Google Colab, VSCode, Microsoft Office Suite
Mathematics: Probability, Statistics, Hypothesis Testing, Linear Algebra, Calculus

EXPERIENCE

CORELOGIC • San Diego & Irvine, CA   *6/2024-9/2024*
Data Science and Analytics Intern • Data Science and Analytics Team
● Collaborated with Director of DS&A and product managers to deliver Department of Insurance reports.
● Leveraged Python and BigQuery for wrangling and conducted validation analysis in GCP for resiliency models, evaluating scoring metrics against loss metrics as segmentation plots using new data.
● Delivered statistics validating resiliency models and generated descriptive flat files to streamline future filings and support sales and marketing efforts, enhancing the product's potential revenue.
● Conducted R&D using large language models (LLMs) and generative AI to automate tax file ingestion, establishing a proof of concept projected to save ~$3-5 million annually.
● Explored prompt engineering techniques to optimize LLM outputs, improving model performance through iterative testing, adjustments, and analysis of prompts.
● Presented to senior executives, sales teams, product managers, and the Chief Data Officer, refining the translation of technical insights into business strategies for all organizational levels and stakeholders.

AMERICAN MEDICAL ASSOCIATION • Newport Beach, CA   *6/2022-9/2022*
Integrated Health Model Initiative(IHMI) Intern • Informatics and Digital Products Team
● Engaged with CMIO and VP of IDPT to report progress and present insights for strategic planning.
● Conducted extensive analysis of the AI landscape in healthcare, researching industry standards to deepen personal understanding of augmented AI applications in medicine.
● Audited "FDA nutrition label" algorithm evaluation criteria, offering insights on integrating a risk framework to ensure ethical, equitable, and transparent outcomes for patients, physicians, and providers.
● Executed UI analyses for Sensely mobile app from user-centered design perspective, highlighting the need for more intuitive button placement and tutorials to improve interaction with upcoming health tools.
● Authored white paper analyzing machine learning for predicting successful patient engagement in healthcare programs, proposing pre-program surveys to capture quality user data for optimal predictions.

PROJECTS

Internet Traffic VAE Network Attack Detection • Python Anomaly Detection   *3/2024-6/2024*
- 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   *2/2024-3/2024*
- 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   *3/2023-6/2023*
- 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   *3/2023-6/2023*
- 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 (2010-2015 Seasons) • Python Data Analysis   *3/2022-6/2022*
- 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.