

Sebastian Algharaballi-Yanow

Machine Learning Engineer and Data Scientist

Irvine, CA, 92614 | (619) 370-2204 | salghara@uci.edu | [GitHub](#) | [LinkedIn](#)

Technical Skills

Programming Languages/Libraries:

- Python, R, SQL. Extensive experience with Pandas, Numpy, Seaborn, Sci-kit Learn, Matplotlib, PyTorch, Tensorflow, NLTK, Spacy, OpenCV, SciPy, Transformers, and the Tidyverse.

Machine Learning:

- Supervised/Unsupervised Learning, Deep Learning, Predictive Modeling, Emotional Intelligence, Cloud Services (AWS, Google Cloud, Azure).

Data Science:

- Data Wrangling, Data Mining, Exploratory Data Analysis (EDA), Data Visualization, Statistical Analysis, Relational Databases.

Tools/Technologies:

- Git/Github, Jupyter Notebook, R Studio, Anaconda, VSCode, MySQL, Power BI, Tableau.

Education

UC Irvine – *Master's of Data Science*

September 2023 - December 2024.

UC San Diego - *Cognitive Science: Machine Learning & Neural Computation (BS)*

September 2020 - June 2023. Cum Laude. GPA of 3.876.

UC San Diego Extension - *Machine Learning Methods (Certification)*

June 2022 - June 2023.

Experience

Scale AI - *Generative Artificial Intelligence Prompt Engineer*

May 2024 - Present

- Edited, refined, and trained ChatGPT responses to enhance accuracy and coherence in real-time conversations.
- Applied regression modeling to predict user satisfaction scores from conversation metrics, resulting in an increase in predictive accuracy and overall model performance.
- Analyzed OpenAI datasets and leveraged clustering techniques to identify user patterns, informing GPT-model adjustments and improved response relevancy.

MoodMe - *Machine Learning Engineer & Co-Founder*

December 2023 - Present

- Leader in the creation and implementation of machine learning models for innovative and unbiased emotion detection applications.
- Collaborated closely with engineering and marketing teams to integrate AI-powered technologies with our services, focusing on user engagement and unique insights.
- Successfully enhanced our emotion detection model through transfer learning, boosting, and noise injection, improving its ability to correctly identify emotions from 75% to 91.5% of cases.
- Led the development of a customized BERT text emotion detection model that is the backbone of EmotionTrak, a well-being diary that helps users track and understand their emotional patterns over time. Model achieved an F1 score of 0.89 in detecting emotions across 8 categories after analyzing over 500,000 text samples during testing.

Donald Bren School of ICS - *Master of Data Science Ambassador*

October 2023 - Present

- Student representative for the MDS Program.
- Supported the creation and development of ideas for marketing campaigns.
- Conducted time series analysis using R's "forecast" library and Python's "statsmodels" to track and predict trends in data science master's programs across the U.S., providing valuable insights for internal program development and competitor analysis.

Sportradar US - *Sports Data Operator*

September 2022 - Present

- Attended NCAA/Pro sporting events (mens and women's basketball, baseball, men's and women's volleyball, soccer) and recorded game-related statistics (using company software).
- Communicated with supervisor on-site regarding the specifics of data.
- Successfully operated 200+ games with 5-star performance.

Featured Data Science Projects

Autoencoder Anomaly Analysis for NBA Players: ([Anomaly Analysis Project](#))

- **Problem:** Delved deeper into understanding the fluctuations in NBA player game-to-game performances, particularly focusing on Kobe Bryant's 2008-2009 season.
- **Action:** Created an autoencoder model in an effort to uncover any non-linear relationships and nuances indicative of performance trends.
- **Result:** Achieved 100% evaluation metrics (precision, recall, F1-score) when labeling Bryant's games as above or below average, potentially discovering factors that contributed to a range of performances.

Advanced NBA Referee Analysis: ([NBA Referee Project](#))

- **Problem:** Explored decision-making patterns of NBA referees to enhance fairness and integrity of the sport. Aimed to identify areas where officiating can be optimized.
- **Action:** Engineered a 1410x23 data frame that utilized cleaned referee statistics. Created advanced neural network models (three FNNs and one CNN) for comprehensive analysis.
- **Result:** Obtained 92% test accuracy through innovative CNN application, confirming the effectiveness of the approach in identifying patterns and areas of improvement in referee standards.

Predicting CCRB's Outcome on Civilian Complaints: ([CCRB Project](#))

- **Problem:** Addressed predicting officer remediation amid the complexity of police misconduct complaints.
- **Action:** Extracted pivotal data features using the "Eli5" Python library, directing focus on variables affecting disciplinary outcomes. Developed and evaluated logistic regression, K-NN, Random Forests, and Gradient Boosting models (using F1-score).
- **Result:** Unveiled logistic regression as the optimum predictive algorithm, enabling accurate forecasts of officer remediation outcomes.

R Research Report

- **Case Study ([Case Study](#)):**
 - **Problem:** Investigated the prevailing trends in tobacco usage among American youth.
 - **Action:** Wrangled/analyzed a comprehensive CDC dataset and produced multiple visualization models to illuminate detailed patterns in youth tobacco and e-cigarette usage.
 - **Result:** Unveiled pivotal insights, such as the popularity of sweet-flavored vapes and a zero-sum relationship between e-cigarettes and traditional tobacco usage.