Aryaman **Sinha**

San Diego, CA 92093

(+1) 858-999-1888 | a2sinha@ucsd.edu | Aryamansinha2704 | aryaman-sinha

Education

University of California, San Diego *La Jolla, CA*

Bachelor of Science in Data Science *Sept 2019 - June 2023 (Anticipated)*

- UC-GPA(at the end of Second Year): 3.578/4.0
- Courses: Data Structures and Algorithms, Practice and Applications of Data Science, Data Management

Skills

Languages Python, R, Java, SQL, LaTeX, Matlab

Packages Pandas, NumPy, Scikit-Learn, Regex, Tensorflow, Matplotlib, Requests, BeautifulSoup, NLTK, Dask, Apache Spark

Concepts Data Cleaning, Machine Learning, Natural Language Processing, Web Scraping, Model Evaluation and Fairness

Tools Text Mining, Robot Parser, Data Imputation, Clustering, Hypothesis Testing, UNIX filesystem operations, Tree/Graph Algorithms, Object Oriented Programming, AWS Cloud Computing

Experience

UC SAN DIEGO *San Diego, CA*

DATA SCIENCE INSTRUCTIONAL ASSISTANT *September 2021 - March 2022*

- Chosen to serve as tutor for class on Theoretical Foundations of Data Science
- Duties included holding biweekly office hours to help struggling students, making creative homework and exam questions, grading assignments and answering student questions
- Re-hired for Winter quarter based on Instructor and student feedback (achieved 100% student recommendation rating)

SCRIPPS RESEARCH TRANSLATIONAL INSTITUTE *La Jolla, CA*

STUDENT INTERN *June 2021 - Sept 2021*

- Prepared a dataset of nearly 3 million data points using a High Performance Computing System for model training
- Developed Baseline model for hearing loss risk prediction
- Created Pipelines for Linear Regression and Artificial Neural Network models

XANE AI *Remote*

MACHINE LEARNING INTERN *June 2020 - September 2020*

- Developed a crawler to create a dataset of images for street sign detection
- Created a Convolutional Neural Network (CNN) for an image classification machine learning model using TensorFlow and PyCaret

DATA SCIENCE STUDENT SOCIETY(DS3) *San Diego*

ASSISTANT PROJECTS DIRECTOR *December 2021 - June 2022*

- Oversaw and ran project committee of 10 project teams each made up of 4 members
- Assessed 150 applicants and conducted interviews to choose most suitable ones
- Mentoring 3 of the 10 projects on Facebook Ads, FIFA player valuations, and NYC Squirrels

Relevant Projects

FIFA soccer match prediction analysis Wrangled and cleaned FIFA videogame data and soccer match data from the last 5 years of around 10,000 matches, developed statistical model using techniques of Bootstrapping and A/B testing. Then, created, tuned and finalized a Gradient Boosting Classifier (GBC) machine learning model to fit the training dataset. And on testing, found that the model had a **higher accuracy than most betting websites**. Finally, applied the statistical and ML model to predict the 2020-21 premier league.

Chicago Police Brutality data analysis Led a team of 4 to wrangle data of Chicago Police and crimes committed over the last 50 years and conduct an Exploratory Data Analysis in search of irregularities in record keeping, identified biases of Race and Gender in the excessive use of force, complaints, awards and demographics of the Chicago PD. The findings of this project were published on several platforms including **Towards Data Science**.

NYPD Allegations Explored ethnic biases in Board Disposition outcomes of allegations registered against Officers. Assessed the type of missingness in the ethnicity and precinct columns and used a permutation test to claim that there are missingness dependencies present. Using a hypothesis test, concluded that the age of a complainant has a significant effect on the board disposition outcome. Developed a baseline model that was later bettered by a fine-tuned Decision Tree Classifier for the task of predicting board disposition outcomes and conducted a fairness assessment.

Political Figure Sentimental Analysis Scraped Wikipedia pages of political figures using beautiful soup and robot parser, checked frequency of key words to detect their political party and conducted a sentimental analysis using Natural Language Toolkit