Cale Williams

San Antonio, TX | 830.267.0728 | calerwilliams@gmail.com | williamscale.github.io

Education

M.S. Analytics | May 2024 | Georgia Institute of Technology

- **Relevant coursework topics:** Machine learning methods; Linear & logistic regression; Random forests; k-means clustering; k-nearest neighbor, support vector machine, & Naïve Bayes classification; Principal component analysis; Stepwise regression, LASSO, & elastic net feature selection methods; Linear/integer/convex optimization methods

B.S. Aerospace Engineering | May 2016 | University of Texas at Austin

Experience

Data Analysis Intern | National Renewable Energy Laboratory | 10/2022 – 08/2023 | Golden, CO

- Built pipeline to ingest and process complex data sets and deployed into online interactive dashboards for stakeholders to view model outputs

Stress Engineer | Sierra Nevada Corporation | 07/2018 – 12/2020 | Louisville, CO

Stress Engineer | The Spaceship Company | 07/2016 - 07/2018 | Mojave, CA

Projects

biRds: National Park & Bird Sighting Dashboard

- Queried, cleaned, and joined multiple complex datasets for ML model ingestion; Implemented data quality checks to ensure smooth model tuning & development; Built visualizations summarizing data features and model results within dashboard

Sports Media Technology 2022 Data Challenge Finalist

- Evaluated baseball infielder arm strength and accuracy using statistical significance tests

Pitcher Roster Optimization

- Cleaned datasets and built ML and optimization models to quantify baseball pitcher arsenals and construct optimal roster subject to skill and financial constraints

NBA Analysis

- Tracking Data Classification: Processed large dataset and built unsupervised classification model to improve temporal labels
- Field Goal Dashboard: Scraped & cleaned large datasets and integrated into dashboard
- Rim Protection Analysis: Evaluated player defensive performance and summarized results into succinct plots & tables
- Bayesian Logistic Regression: Trained model using data science techniques, predicted on test dataset, and calculated model performance metrics for model evaluation

Tools & Software

- **R:** tidyverse, Shiny, Plotly, R Markdown
- **Python:** pandas, NumPy, Matplotlib, scikit-learn, SciPy, SQLite, PyMC, CVXPY, NetworkX