Michael Baluja

Machine Learning and Data Science

michael@michaelbaluja.com
michaelbaluja.com
linkedin.com/in/michaelbaluja
github.com/michaelbaluja
twitter.com/mlbaluja
760-330-6385

Education

M.S. in Electrical and Computer Engineering

University of California, San Diego (September 2021 – March 2023)

Focus in Machine Learning and Data Science (GPA: 3.85)

B.S. in Cognitive Science

University of California, San Diego (June 2019 – June 2021)

Specialization in Machine Learning and Neural Computation (Major GPA: 3.91)

Work Experience

Data Scientist Intern

Experian – North America DataLabs (June 2022 – September 2022)

- Analyze and refactor existing development code to deliver up to 92.6% speed increase.
- Assemble Python package for statistical analysis of disparate impact to protected groups from six metrics.
- Integrate explainable machine learning methods to ensure fairness across model predictions.

Data Assistant

UC San Diego – Library Data Curation Team (June 2021 – June 2022)

- Collect 125k research object records across eight public research repository APIs and websites.
- Create and maintain open-source user interface for big data collection via supported repositories.
- Collaborate within data curation team to integrate user feedback into application development.

Research Assistant

UC San Diego – CMRR STAR Group (January 2020 – Present)

- Improve deep reinforcement learning framework to design error protection weight masks for text classification networks trained on two different classification tasks for multiple data sets.
- Increase model performance by up to 300% in the presence of simulated weight errors.
- Communicate experimental results in conference papers, posters, and presentations for four undergraduate conferences.

Technical Skills

Python **SQL** **Scikit-Learn** **Unix** **Tkinter** **Selenium** **Pandas** **git** **PyTorch**

Projects

PyCurator (June 2021 – Present)

- Standardize data collection pipeline for web scraping and API queries via Object-Oriented multiple inheritance approach.
- Design public user interface to allow streamlined access to eight common online research repositories.
- Implement best practices to ensure computational simplicity, reducing runtime by up to 72%.

Empirical Comparison for Natural Language Processing Classifiers (October 2020 – January 2021)

- Explored data sets such as Amazon and Yelp Review data and refined for model training and validation.
- Tested Logistic Regression, Random Forest, and Support Vector Machine Classifiers against Deep Neural Network performance on two-class text classification problems.
- Reported performance across three metrics, most predominately the F1-score, and performed hypothesis testing to understand the performance trade-off of neural networks against off-the-shelf models.