STEPHEN RO, Ph.D.

stephensro@gmail.com | GitHub | Website | LinkedIn

Researcher and instructor leveraging 12 years of experience on large data sets, analysis and modeling, simulations, and presentations. Independent starter, experienced team collaborator, and mentor.

Projects

The Data Incubator Capstone Project

- Independently created an end-to-end wine app that recommends best-priced wines: OVINO. Selected highest quality DS projects amongst two DS cohorts.
- Deployed with XGBoost classifier with 0.90 PR AUC trained on wine reviews with price-sentiment using sentence transformers (S-BERT) and NLP techniques.

Wine Scraper Bots [Link]

- Designed 3 automated web scrapers to acquire live features of 13.6k products across 800 store locations. Automated data cleaning, uploads to SQL database, and updates to OVINO recommender.
- Scraped 31 million wine reviews, isolated 150k with prices, curated 6k price-sentiment training data.

The Data Incubator Certification Projects — September 2022 - March 2023

- Stack Overflow veteran-user classifier (PySpark + SQL + ML); NYC cuisine-borough violations (SQL); CNNs and transfer learning an inception network (TensorFlow); time-series temperature, and Yelp rating predictions (random forest, linear/SGD regressors, FFTs); Polarizing word identifier in Yelp reviews (NLP & Naive Bayes);

Experience

UC Berkeley, Astronomy Department — July 2020 - August 2020
Course Instructor: Introduction to Scientific Python for Astronomers — *Berkeley, CA*

- Designed and taught a data-driven python course for astronomers. Topics: large data, time-series, numerical simulations, good visuals, scientific writing, and their applications to astrophysics.

UC Berkeley, Theoretical Astrophysics Center — September 2017 - August 2020
Postdoctoral Research Fellow — *Berkeley, CA*

- Designed state-of-the-art shock models. Reduced shock evolution prediction errors by 50x across 3-orders of magnitude in param space using time-dependent analyses and mathematical models.
- Decreased computational resources by 10,000x for shock evolution calculations.
- Deployed thousands of explosion simulations on 100-core clusters and analyzed 3 TB of data.

University of Toronto, Department of Astronomy & Astrophysics — September 2011 - August 2017
Doctoral Candidate — *Toronto, Canada*

- Wrote and designed a telescope image synthesizer that uses real-world instrument statistics and uncertainties. Used by 4 graduate students for writing world-class instrument proposals.
- Wrote a multi-dimensional relaxation solver on a mesh grid. Analyzed thousands of high-dimensional datasets, extracting features, and reshaped leading models on stellar surfaces.

Education

Data Scientist Certification | The Data Incubator Fellowship Program | March 2023

PhD in Astrophysics focus on Computational Fluid Dynamics | University of Toronto | November 2017

BSc in Mathematical Physics | Queen's University | May 2011

Technical Skills

Data Science: Supervised Learning (Linear | Logistic | KNN | Trees | SVM | Naive Bayes) | Unsupervised Learning (K-means |HDBSCAN | UMAP | PCA) | Big Data (PySpark) | Natural Language Processing | Processing & Visualization (Pandas | Matplotlib) | Analysis (NumPy | SciPy)

Programming Languages: Python | SQL | Bash | YAML | C++ | FORTRAN | HTML / CSS

Industry Knowledge: Mathematical Modeling | Statistical Analysis | Predictive Modeling | Quantitative Analysis | Data Management | Data Collection | Dynamical Systems | Lecturing, Teaching & Mentoring | Communication at all levels | Fluid Dynamics | Course & Curriculum Design