STEPHEN RO, Ph.D.

Data Scientist
stephensro@gmail.com | github.com/royourboat
Qualify for TN Visa (Canadian citizen); Authorized to work in the US

Technical Skills

Data Science: Deep Learning/ Neural Networks (TensorFlow | Image Processing CNNs) | Machine Learning (K-means | Clustering | PCA | Trees) | Big Data (PySpark) | Natural Language Processing | Linear & Logistic Regression (Scikit-Learn) | Data Processing & Visualization (Pandas | Matplotlib |Flask) | Analysis (NumPy | SciPy)
Programming Languages: Python | SQL | Bash | Spark | YAML | C++ | FORTRAN | HTML / CSS | Mathematica
Industry Knowledge: Mathematical Modeling | Statistical Analysis | Predictive Modeling | Quantitative Analysis | Data Management | Data Collection | Large Data Sets | Scientific Coding | Fluid Dynamics | Dynamical Systems | Fluid Dynamics | Course & Curriculum Design | Lecturing, Teaching & Mentoring | Communication at all levels

Projects

The Data Incubator Capstone Project: "Best-valued wine recommender for beginners: OVINO" February 2023

- Wrote Github Actions workflows to scrape 13.6k alcohol products and inventories across 800 store locations. Managed tables in PostgreSQL database with product features, prices; store hours, locations, inventory.
- Scraped 9 million Vivino reviews. Created training data of reviews with prices paid and price-sentiment. Developed 91%-acc price-sentiment model using NLP and machine learning techniques.
- Technologies used: Scikit-learn, PostgreSQL, Pandas, GitHub Actions, Flask, Render, spaCy, Google API.

The Data Incubator Certification Projects September 2022 - March 2023

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

Experience

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

- Designed and taught a fully online course on scientific computing practices, e.g., large data, good visuals, time-series, numerical simulations, data management, scientific writing, and their applications to astrophysics (e.g., Pluto surface features, N-body, galaxy spectral features, satellite burnup sims, accelerating universe).

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

- Derived new analytic acoustic shock model using fluid equations and 3 TB of simulation data. Deployed simulations on a 100-core computer cluster.
- Improved accuracy of shock model predictions by a factor of 50 across 3-orders of magnitude in param space.
- Discovered flaws in gold standard hydrodynamic astrophysics code used worldwide. Consulted code developers on correctly reproducing basic wave solutions and building new applications.

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

- Wrote and designed a spectrographic image synthesizer with real-world statistics from instruments and radiation physics. Used for future world-class instrument testing and grant proposal preparation.
- Built a data pipeline (python, C++) simulating data cleaning, data management, and feature extraction.
- Wrote a PDEs relaxation solver to model stellar winds (Python, FORTRAN, C++). Analyzed thousands of stellar wind solutions and extracted parameters to disprove leading ideas on stellar surfaces.

Education and Teaching

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

University of Toronto, Department of Astronomy & Astrophysics May 2022 - April 2023
Course Manager and Administrator; AST 101 *Toronto, Ontario*

- Managed 36 teaching assistants and administered one of the largest courses in NA (1,500 students).
- Discovered multi-million dollar fraudulent business selling illegal solutions to hundreds of students.