

Eros Rojas

☎ (403) 467-0298 | ✉ eros.rojasv@gmail.com | 🏠 erosrojas.github.io | 📱 erosrojas | 🌐 eros-rojas

Skills

Programming	Python (standard + vectorized), R, Julia, SQL, Java (OOP), C++, MatLab, LaTeX
Packages/Libraries	pandas, TensorFlow, NLP (mUSE, top2vec, BERTopic), Vectorizers, scikit-learn, ray multiprocessing
Data Visualization	matplotlib, seaborn, plotly, Shiny, umap-learn, ggplot2, Jupyter Notebook

Experience

Data Scientist Co-op

Ottawa, Ontario

GOVERNMENT OF CANADA - APPLIED RESEARCH

Jan. 2023 - Present

- Developed an end-to-end NLP pipeline that fully embeds, topic models, and evaluates text metrics to satisfy our clients needs.
- Utilized and trained semi-supervised TensorFlow-Hub models to increase the efficiency and accuracy of our text embeddings.
- Implemented Ray multiprocessing to distribute NLP model to a Kubernetes cluster, reducing run-time from 10hrs to 18 minutes.
- Participated in bi-weekly reading groups and local conferences regarding novel data science breakthroughs and technologies.

Data Science Teaching Assistant (TA)

Vancouver, British Columbia

UNIVERSITY OF BRITISH COLUMBIA DEPARTMENT OF STATISTICS

Sep. 2021 - Present

- Organize and deliver a range of teaching activities including lectures emphasizing the key concepts of data science and statistical modeling at an undergraduate level in R.
- Review and give feedback on semester-long data analysis projects based around a predictive algorithm, which includes offering constructive criticism on both code mechanics and rationale.

Data Analyst

Calgary, Alberta

VERGE AGRICULTURE

Jan. 2021 - Apr. 2021

- Developed a mathematical model using Taylor series that quantified the non-linear relationship between soil compaction, and varying soil types with fluctuating moisture, clay, and porosity levels.
- Analyzed worldwide soil data in order to determine various methods of mitigating annual crop yield losses due to soil compaction.

Personal Projects

Statistical Analysis into Heart Disease

WRITTEN IN R + GIT

Oct. 2022 - Dec. 2022

- Trained a LASSO regularized Logistic Regression model to predict the presence of binary heart disease. Optimized the model sensitivity (to reduce false-negatives) through cross-validation, variable selection, outlier detection, and exhaustive threshold setting techniques.

Personal Data Science Blog

WRITTEN IN PYTHON + R + HTML + GIT

July. 2022 - Present

- Building a personal blog through Quarto, a technical publishing system based on Pandoc, and GitHub Pages, to create a platform that hosts, compiles, and automatically renders Python + R code through specialized markdown files.

S&P 500 Forecasting through Time-Series Analysis

WRITTEN IN R

Dec. 2020 - Feb. 2021

- Utilized the ARIMA algorithm to forecast S&P 500 time-series trends and future stock prices, while fitting and optimizing the model through ACF and PACF statistical plots.

Awards

ROBOTICS

2020 **Qualified**, VEX World Championships (Cancelled due to COVID)

Calgary, Alberta

2020 **Winner**, Design Award and Think Award at VEX Alberta Provincial Championships

Calgary, Alberta

Education

Bachelor of Science (Major in Mathematics, Minor in Data Science), 3rd Year

Vancouver, Canada

UNIVERSITY OF BRITISH COLUMBIA

Sep. 2020 - Present

- Completed 6 Computer Science courses including CPSC 330 (Applied Machine Learning) and CPSC 340 (ML and Data Mining)