Ruby Cheung

(612) 481-6071 • in/rubyccheung • rcheung763.github.io/dsportfolio/

Relevant Work Experience

Institutional Research Analyst
Wabash College, Crawfordsville, Indiana — November 2022 – Present

- Designed and automated institution-level dashboards and reports with **R** to support data-driven decision-making.
- Conducted research on registration tracking, retention, and other institutional priorities directed by the Dean.
- Coordinated external reporting and managed ad hoc data requests for grant applications, accreditation, process improvements, and other campus needs utilizing the **SQL queries**.
- Developed an institutional research (IR) handbook to document processes, improve reporting standards, and reduce the time it takes to complete reporting tasks by approximately **50%**.

Graduate Research Assistant
Seattle University, Seattle, Washington — June 2024 – September 2024

- Collaborated with the Math and Biology departments to develop and deploy **NLP** models.
- Processed a high volume of unstructured data, performing **feature engineering** and vectorization.
- Conducted **sentiment analysis** on the comments from carnivorespotter.com, designed a robust evaluation process to test model performance under class imbalance conditions, achieving **81%** accuracy.

Economic Research Specialist
St. Catherine University, St. Paul, Minnesota — June 2020 – August 2022

- Led economics research teams on projects focused on development economics and labor markets in North Africa and Middle East.
- Conducted labor analysis using regression to predict employment rates and employment status using **logistic regression.**
- Designed a census map for Sudan using **GIS** software.
- Designed Android-based survey instruments using **ODK-X** with integrated **SQLite** for data management.

Technical Competencies

Programming languages: R, Python, Stata, SQL, html, CSS
Techniques/Methods: Supervised/Unsupervised Learning, Neural Networks, Regression Analysis, Statistical Analysis, Hypothesis Testing, Time Series Analysis, Sentiment Analysis, Text Classification, Web Scraping
Technology/Frameworks: RStudio, Scikit-Learn, TensorFlow, Spark, MySQL, Tableau, AWS, NoSQL Databases, Selenium, BeautifulSoup, PyTorch, Matplotlib, MongoDB, Prophet, Keras, SARIMA

Education

M.S. Data Science, Seattle University — June 2025
B.S. Economics, St. Catherine University — May 2020

Selected Projects

Predicting Recidivism Among Georgia Parolees | Python, scikit-learn, SVM, XGBoost

- Utilized data from the Georgia Department of Community Supervision, built a supervised and an unsupervised model to identifying key variables that can help to predict the likelihood of recidivism, producing actionable insight to mitigate risk
- Developed and evaluated **Gradient Boosting** and **SVM** models, achieving a 73.6% accuracy for binary classification and a 72.6% accuracy with linear SVM, with a focus on reducing false positive rates.
- Applied variable importance techniques that identified correlations between continuous employment and low recidivism rates, highlighting the significance of stable employment for reducing reoffending.

Evaluating Forecasting Methods for Gross Gaming Revenue Prediction | R, Python, ARIMA, ETS, GARCH, Prophet

- Scraped Louisiana Gaming Reports to create dataset
- Conducted **ADF tests** for stationarity, performed **residual analysis** using **Ljung-Box tests** and **Shapiro-Wilkes tests**
- Addressed COVID-era and hurricane shocks by incorporating Prophet features and creating exogenous variables to ensure accuracy, the best model achieved low RMSE and a MAPE of 12.3%.

Identifying Substance Abuse Patterns among Youth | R, Random Forest, Boosting, Regression

- Utilized Random Forest and **Boosting algorithms** to predict substance abuse among youth.
- Identified key factors influencing substance abuse through variable importance analysis.
- Achieved an accuracy of 85% with the best-performing model, highlighting the effectiveness of the approach.