Ryan Bae
Senior Data/Applied Scientist

(408) 857-1351
Seattle, WA
ryanbae89@gmail.com

Portfolio: https://ryanbae89.github.io/
github.com/ryanbae89
linkedin.com/in/ryanbae89

I am an experienced data scientist providing end-to-end machine learning solutions to improve products, solve business problems, and build AI infrastructure. I love working on all parts of the data science pipeline, from exploratory data analysis to modeling and deployment/ML engineering. I have expertise working with user feedback (NLP), fairness in AI, and privacy.

Skills

Data Science/ML	Natural Language Processing, Applied Stats/Experimentation, Deep Learning
Languages & Tools	Python, R, SQL, PySpark, Pandas, Numpy, Scikit-Learn, PyTorch
Communication	English, Korean (fluent speaker)

Professional Experience

Senior Data & Applied Scientist / W+D Data   Feb 2019 — Present
Microsoft   *Redmond, WA*

- **Patent:** Bae, Ryan. 2022. Correlating Instances of Written Feedback to Common Events in Telemetry Data. U.S. Patent X,XXX,XXX, filed Oct 16, 2022. Patent pending.
- Designed and built ML pipeline to correlate Windows customer feedback with diagnostic telemetry, increased feedback action-ability, increased bug fix-rate, and reduced time to resolution of bugs.
- Evaluated fairness of Windows on-client gaze redirection by predicting error metrics using mixed-effects models, prevented shipping of under-performing models and directly impacted improvements to subsequent models.
- Designed and built novel pattern-less personal data detector for Windows telemetry using statistical testing at scale, work currently being used to satisfy privacy compliance for all Windows telemetry.
- Developed NLP text classification pipeline for Windows user feedback, resulted in improvements in bug/issue discovery time from user feedback.
- Trained domain adapted fasttext and BERT models on Windows feedback data, evaluated performance on downstream classification tasks, which showed gains in F1-scores.
- Mentored summer interns and new-hires to successful projects.

Data Science Intern / W+D Data   Jun 2018 — Aug 2018
Microsoft   *Redmond, WA*

- Built data pipeline in SQL/C calculating net promoter score of Windows pre-release builds.
- Identified Windows quality metrics most associated with detractors using interpretive random forest/logistic regression models.
- Drove future direction of Windows Insider program by recommending changes to the net promoter question.

Machine Learning Engineer Intern   Nov 2017 — May 2018
Clobotics   *Bellevue, WA*

- Implemented parts of a computer vision paper to detect blurriness of retail images in Python. Forward propagation coded from scratch using Python numpy library.

Propulsion Development Engineer   Sep 2014 — Feb 2017
Space Systems/Loral   *Palo Alto, CA*

- Modeled chemical reaction to predict flow decay in spacecraft propulsion system, eliminated propellant waste by $100K per spacecraft, earned company award.

Education

M.S. Data Science, University of Washington at Seattle	2017 - 2019
Data Science Merit & Opportunity Scholarship Recipient	2018
M.S.E. Aerospace Engineering, University of Michigan at Ann Arbor	2012 - 2014
B.S. Aerospace Engineering, University of California at Los Angeles	2007 - 2012

Projects

Yelp Reviews Q&A Bot: A GPT based Q&A bot for information retrieval from a large Yelp customer reviews corpus.	2023
US County Level Gun Violence Model: Predicting gun violence deaths in US at county level using weighted linear regression.	2022
Video Similarity Search Engine: Video similarity search engine using 3D CNN architecture (MSDS Capstone Project).	2019
News Article Recommender: Guided LDA topic modeling and recommendation of 12k news articles (MSDS Class Project).	2018