

RYAN S. BAE

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

<https://ryanbae89.github.io/> ♦ <https://www.linkedin.com/in/ryanbae89/> ♦ <https://github.com/ryanbae89>

Data scientist with strong interest in applied machine learning/natural language processing to build analytics/ML infrastructure and improve products. Areas of interest include:

- Data science/applied machine learning
- Applied statistics and experimental design
- Natural language processing
- Big data frameworks and ML engineering

Education:

University of Washington at Seattle – M.S Data Science 2017 – 2019

- Data Science Merit & Opportunity Scholarship Recipient

University of Michigan at Ann Arbor – M.S.E. Aerospace Engineering 2012 – 2014

University of California at Los Angeles – B.S. Aerospace Engineering 2007 – 2012

Technical Suite:

Programming:

Python, R, SQL

Data Visualization:

Tableau, matplotlib

Deep Learning:

PyTorch

Big Data/Cloud:

Azure, Databricks, PySpark

Professional Experience:

Microsoft (Data Scientist – CSX Data)

Feb 2019 – Current

- Designed and built pattern-less personal data detector using statistical testing at scale on Windows telemetry.
- Designed and built ML pipeline to relate Windows 10 feedback with diagnostic telemetry, increased actionability of feedback and reduced time to resolution of bugs.
- Developed ML text classification pipeline for Windows user feedback, resulted in improvements in bug/issue discovery time from user feedbacks.
- Trained domain adapted FastText and BERT models on Windows feedback data, evaluated performance on downstream classification tasks, which showed gains in F1-scores.
- Put all project models into production using AzureML, Databricks, and PySpark.
- Mentored summer interns to successful internship experiences.

Microsoft (Data Scientist Intern - COSINE D&I)

Jun 2018 – Aug 2018

- 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 in Python.
- Drove future direction of Windows Insider program by recommending changes to the net promoter question based on results found above.

Clobotics (Machine Learning Engineer Intern)

Nov 2017 – May 2018

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

Space Systems/Loral (Propulsion Development/Analysis Engineer)

Sep 2014 – Feb 2017

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

Relevant Projects:

Video Similarity Search Engine (MS Data Science Capstone Project)

Sep 2018 - Mar 2019

- Built video similarity search engine using extracted features from the videos. Implemented the feature extraction portion using 3D CNN and similarity search using Facebook's FAISS architecture.

News Article Recommender

Mar 2018 – Jun 2018

- Designed the architecture and led a team of data scientists to build a news articles recommender. Built LDA topic models to perform topic modeling and tagging on 120,000 news articles.

UFO Sightings Dashboard/Visualization

Oct 2017 – Dec 2017

- Performed data engineering and cleaning using pandas on 80000+ UFO sightings around the world, built visualizations using Tableau and Python and hosted on a website.