

Rory M. Flynn

Data Science, Statistics, Machine Learning



Location:
Phone :
Email :
Website:
Github:

REDACTED
REDACTED
REDACTED
RMFlynn.co
github.com/rmFlynn

Education

Masters of Computer Science Graduated: Dec. 12, 2020

University of Colorado, Denver

Concentration: Data Science in Biomedicine, **Read My Thesis**

Thesis Adviser: Ashis Kumer Biswas

Bachelors of Science in Mathematics Graduated: Dec. 11, 2015

Metropolitan State University of Denver

Concentration: Probability and Statistics

Minor: Chemistry

Associates of Science Graduated: Dec. 15, 2013

Community College of Denver

Previous/Current Employment

Research Software Engineer Apr. 2021 - Now

Colorado State University

- Responsible for developing DRAM, an Open Source framework for exploring microbiome genetic data.
- Lead developer and designer of DRAM2, a massive upgrade to DRAM with 50%+ original code and 30% original features, huge 10x speed improvements, and distributed computing support.
- Responded to and closed 100+ issues on GitHub and Jira.
- Maintained Code base, Documentation, Conda environment, and Continuous Integration.
- Developer of KB-dram, a separate version of dram meant to run on the K-base platform in collaboration with the KBase team. Restoring functionality in a changing API.
- Presented work in posters and presentations.
- Created 5 large scale complex ETL pipelines in Python, Snakemake, and BASH. Automated multi-day process, standardizing protocols, and maximizing use of HPC systems.
- Co-maintained tools on the CSU HPC, and other HPC systems, and provided troubleshooting via GitHub.
- Worked with students and experts specializing in multiple disciplines, eg. microbiologists, biochemists, and ecologists.
- Supported students and postdoctoral staff with data science, machine learning, and statistical tasks.

Machine Learning Engineer Dec. 2020 - Apr. 2021

Biointellisense

- Developed machine learning and time series models, processing data for wearable devices.

Machine Learning Graduate Apprentice Jan. 2020 - Dec. 2020

University of Colorado Denver

- Created documentation and tools to securely manage and analyze a large genetic data set.
- Designed an original algorithm, implemented in Python, to address the challenge of imbalanced classes in genetic data.
- Utilized HPC resources to build, fit, and evaluate custom and off-the-shelf models.
- Communicated findings and research to advisors of diverse backgrounds through presentations and reports.

Teaching Assistant (Algorithms) Aug. 2019 - Dec. 2019

University of Colorado Denver

- Assisted students with class comprehension, graded written and Python programming assignments.

- Created a complex Python package for importing, cleaning, and transforming data from partner institutions to ensure compatibility with NREL tools.

Research Associate Mar. 2016 - Aug. 2018

BBC Research and Consulting

- Performed various system administration, automation, and scientific programming tasks using Python, SQL, Perl, and VBA.
- Created complex web applications for data entry, tracking, and manipulation using R-shiny.
- Designed a small server environment for internal hosting from scratch, using Ubuntu Linux and Docker. Maintained said server and software stack, while training others to do the same.
- Automated statistical analysis, including multi-variable regression, ANOVA, and Monte Carlo Simulation using R and Perl.
- Worked with large relational databases, cleaning, proofing, and analyzing the data therein.

Cloud Operations Intern May 2014 - Jan. 2015

Recondo Technology

- Created various administrative programs using Ruby, BASH, PowerShell, and Puppet scripts.
- Maintained stage and prod Linux servers, implemented hardware and software, and managed databases.

Publications, Presentations

Conference Presentations.....

2022

Poster & Talk

"Distillation and Refinement of Annotations of Metabolism Enables Rapid Understanding of Microbial Functional Potential," 18th International Symposium on Microbial Ecology, Aug. 2022. Lausanne, Switzerland

2022

Poster & Talk

"DRAM: Distillation and Refinement of Annotation of Metabolism," Front Range Microbiome Symposium, Apr. 2022. Fort Collins, Colorado

Publications.....

2023

In progress
First Author

R. M. Flynn, M. A. Borton, B. Bolduc, *et al.*, "Dram2: Improved annotating and functional profiling of genomes with phylogenetic trees," 2023

2023

Coming Soon
Co-Author

M. Shaffer, M. A. Borton, B. Bolduc, *et al.*, "Kb-dram: Annotating and functional profiling of genomes with dram in kbase," 2021

2023

Coming Soon
Co-Author

B. McGivern, B. Woodcroft, J. Ellenbogen, *et al.*, "A polyphenol metabolism cache discovered in thawing permafrost microbiomes," 2021

2022

Preprint
Co-Author

I. Lelewi, J. Rodriguez-Ramos, M. Shaffer, *et al.*, "Exposing new taxonomic variation with inflammation – a murine model-specific genome database for gut microbiome researchers," *bioRxiv*, 2022. doi: 10.1101/2022.10.24.513540

2021

Masters Thesis
First Author

R. M. FLYNN, "Predicting autism spectrum disorder from genome-wide association data with genetic balancing generative adversarial network," Ph.D. dissertation, proquest, 2021

Recent Projects

DRAM2 Aug. 2022

Work assignment

In progress

Large re-write of the DRAM metagenome annotation toolkit and pipeline. Intended to bring the tool kit up to modern standards of Python development, add revolutionary features based on probabilistic graphs, provide massive speed improvements, and fully modularize design.

Read Mapping No Problem Sept. 2022

Work assignment

In progress

Professional implementation of a complex ETL pipeline, integrated with HPC scheduling software. Written in Python and Snakemake.

Talking GAN Project Nov. 2021

Graduate Project

Finished

A Generative Adversarial Network, featuring a U-Net design, that generates video of a speaking human mouth using only voice audio and a still picture of a mouth.

Contagion View GTK Project Jun. 2020

Graduate Project

Finished

A desktop application, created in GTK3, that can be used to visualize the spread of the COVID-19 outbreak. It features an interactive map, linked time series plots, and multiple data filters.

Skills

Programming Languages

Python	Expert	7+ years of heavy use in academia and corporate environments.
BASH	Advanced	5+ years of use in academia and corporate environments.
Java/Type Script	Intermediate	4+ years of use for graphics and visuals, p5.js and d3.js.
SQL	Intermediate	4+ years of use primarily in Postgres and MySQL.
C++	Intermediate	3+ years of use in academia.
R	Intermediate	4+ years of use in academia and corporate environments.

I also have experience with Java, Perl, Ruby, and SAS. Interested in experiencing more programming languages, especially Rust and Julia.

Operating Systems

Linux:	Expert	9+ years heavy use
Deep understanding of the Linux tool kit, packaging formats, file structure, and ecosystem. Highly effective in a command line environment, with a long history using Docker, Vi/Vim, ssh and other system tools.		
Windows:	Intermediate	9+ years of use
Troubleshooting and programming, using tools like PowerShell, ODBC, and VBA. Looking forward to using WSL.		

Disciplines

Data Management	Automation	Genetic Data Analysis	Machine Learning
Web Scraping	Regression Analyses	Data Visualization	Web Applications

Office Software

Microsoft Office (Word, Excel, Access)	LibreOffice	Google Docs
--	-------------	-------------

Other Software

Git	Linux Command Line	VirtualBox	Docker
TEX/LaTeX			

Interests

Generative AI	Active Learning	Linux App Development	Meta Learning
Protein Synthesis	Speech Recognition	Genetics	Data Rights & Privacy