MATT KAYE

mrkaye97.github.io | mrkaye97 | mrkaye97@gmail.com | (646) 853-5997 | kayem20

EXPERIENCE

Data Scientist

CollegeVine

Sept 2020 – Present

- First member of the CollegeVine data science team. Responsible for establishing data science best practices, building out modeling pipelines, helping teammates do rigorous data analysis, and explaining data scientific methods to other members of the team to democratize data scientific work at CollegeVine
- Transitioned our chancing algorithm (used by hundreds of thousands of students to get their chances at over 1,500 colleges and universities) away from a rules- and heuristics-based model to an ML-forward approach. Built expert domain knowledge into the model features and made product-first modeling choices to ensure intuitive model behavior for users
- Built a dashboard with sequential testing tools and helped explain the benefits of sequential testing to our PMs to let them run A/B tests more quickly without sacrificing statistical rigor
- Responsible for owning all components of the data science process, including wrangling data, doing data analysis, building and validating models, deploying models to production (generally as Dockerized microservices on Heroku), and monitoring models and APIs in production
- Serving as a subject matter expert for everyday statistics- and data science-related questions and problems

Open-Source R Developer

slackr & fitbitr

Oct 2020 – Present

- Current author and maintainer of *slackr*, an R package for connecting R to Slack with 215k+ downloads
- Author and creator of *fitbitr*, an R package that streamlines pulling Fitbit user data via the Fitbit API
- Responsible for all aspects of package development and maintenance, including implementing new methods, improving error handling and messaging, writing unit tests, establishing and maintaining a CI/CD pipeline, writing descriptive documentation, helping users work through issues and bugs, reviewing PRs, and more

Baseball Operations Fellow

Baltimore Orioles

Mar 2020 – Sept 2020

- Created a fully Bayesian, simulation-based projection system for MLB player performance over a six year time horizon
- Modeled free agent salaries with a gamma hurdle regression framework
- Contributed to a Markov Chain Monte Carlo approach to determining optimal shifts against opposing hitters
- Worked on a variety of day-to-day data science tasks related to game strategy and player evaluation

SKILLS

Programming Languages:

R	●●●●●
Python	●●●●○
SQL	●●●●○
Shell Scripting	●●●○○
Haskell	●●○○○
Java	●●○○○

Frameworks, Software, and Tools:

AWS (S3, Redshift, RDS, etc.) | brms | Docker | Heroku | Numpy | Pandas | R Shiny | Scikit-Learn | Tidyverse + Tidymodels

Data Science:

Bayesian Modeling and Inference | Data Wrangling | Data Visualization | Machine Learning | Time Series Modeling

EDUCATION

Bachelor of Arts
Economics and Mathematics

Carleton College

Sept 2016 – Nov 2019

Choate Rosemary Hall

Sept 2013 – June 2016

INTERESTS

Baseball | Browsing r/AskReddit | Cooking | Running | Minimalism | Skiing | Data Visualization | Fantasy Novels | Solo Travel | Nature & Architecture Photography