MATT KAYE

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

EXPERIENCE

Data Scientist

CollegeVine

Sept 2020 – Present

- First member of the CollegeVine data science team. Responsible for owning all parts of the data scientific process and for establishing data engineering and MLOps infrastructure.
- Devised and implemented an acceptance probability model used by hundreds of thousands of students every year to understand their chances at their favorite schools.
- Built a sequential testing toolkit to enable our team to call A/B tests more quickly without sacrificing statistical rigor.
- Created a recommender system to suggest new colleges to students. Additionally, utilized the recommender system to suggest "top candidates" to our college partners, driving leads and engagement.
- Built out a feature store, enabling our team to iterate more quickly, repeat less code, and easily perform BI tasks that were previously infeasible.
- Set up and managed an MLFlow instance, dramatically improving data science developer experience (DX) at CollegeVine.
- Owned the deployment and monitoring of machine learning models in production, generally as REST APIs packaged up as Dockerized microservices and deployed on Heroku.

Open-Source R Developer

slackr, lightMLFlow, fitbitr

Oct 2020 – Present

- Current author and maintainer of $slackr$, an R package for connecting R to Slack with 250k+ downloads.
- Author and creator of $lightMLFlow$, a lightweight, user-friendly R wrapper for the MLFlow REST API.
- 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
- Devised 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

EDUCATION

Bachelor of Arts
Economics, Mathematics

Carleton College

Sept 2016 – Nov 2019

Choate Rosemary Hall

Sept 2013 – June 2016

SKILLS

Programming Languages:

R | Python | Shell Scripting | SQL

Frameworks, Software, and Tools:

Airflow | AWS | CircleCI | Docker | Git | Github Actions | Heroku | MLFlow | {{plumber}} | {{shiny}}

INTERESTS

Baseball analytics | Browsing r/AskReddit | Low and slow cooking | Distance running | Skiing chop and powder | Reading fantasy novels | Solo traveling + hostel hopping | Nature & architecture photography