MATT KAYE

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

EXPERIENCE

Data Scientist

CollegeVine

Sept 2020 – Present

- 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 model to generate affinity scores to enable colleges to more intelligently recruit high school students.
- Designed and built out an analytics warehouse, supercharging data science and analytical capabilities at CollegeVine.
- Set up and managed an MLFlow instance, dramatically improving data science developer experience (DX) at CollegeVine.
- Spearheaded the transition of our workflows to Apache Airflow, dramatically reducing the complexity of scheduling and running jobs
- 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 Contributor

mlflow, slackr, lightMLFlow, fitbitr

Oct 2020 – Present

- Contributor to MLFlow, an open-source platform for managing the machine learning lifecycle.
- 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:

Bash | Python | R | SQL

Frameworks, Software, and Tools:

Airflow | AWS (Batch, Redshift, S3, etc.) | CI/CD (Circle, GHA) | dbt | Docker | Git | Heroku | MLFlow | {{plumber}}

HOBBIES

- Distance running
- Learning Haskell and Rust
- Low and slow cooking
- Nature and architecture photography
- Reading fantasy novels
- Skiing chop and powder
- Solo traveling + hostel hopping