

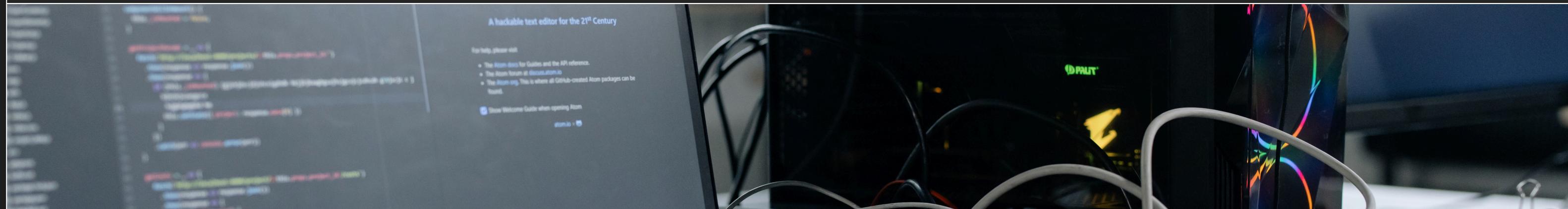


Search



17-445 MACHINE LEARNING IN PRODUCTION

Team 17: Substack Redemption



Anushka

Jeremy

Jessie

Praneet

Soroosh



01

Milestone 1

Key Design Decisions

- Produced 3 csv files from Kafka for logging and training
- Tested KNN and SVD models, choosing SVD due to its lower training and inference costs, smaller model size, and comparable prediction accuracy to KNN
- Service deployed using Flask

Milestone 2

Key Design Decisions

- Evaluate offline performance using RMSE, training time, inference time, and size
- Online evaluation consisted of tracking user engagement, the quality of recommendations, and user satisfaction based on ratings (inconclusive results)
- Data tests focused on column formats, rating values, and thresholds for drift
- CI Pipeline testing through Jenkins, integrated with GitHub webhooks ensuring consistent validation of code changes, with comprehensive test coverage reports accessible for monitoring



Milestone 3

Key Design Decisions

- Containerized with Docker, using a load balancer for A/B testing
- Automated model retraining and deployment using a Jenkins CI/CD pipeline
- Monitored system performance through Prometheus and Grafana, tracking service availability, model accuracy, and recommendation adoption rates.
- Implemented A/B testing with two SVD models, analyzing user behavior and ensuring balanced traffic distribution between models.

Milestone 4

Key Design Decisions

- Potential fairness issues may come from sampling methods (time specifically)
- Equal quality recommendations across user groups and minimizing bias from data collection periods as requirements
- Explored popularity and filter bubble loops that could limit diversity in recommendations and engagement
- Security focused on Kafka stream poisoning and denial-of-service attacks, recommending anomaly detection, rate limiting, and user authentication.

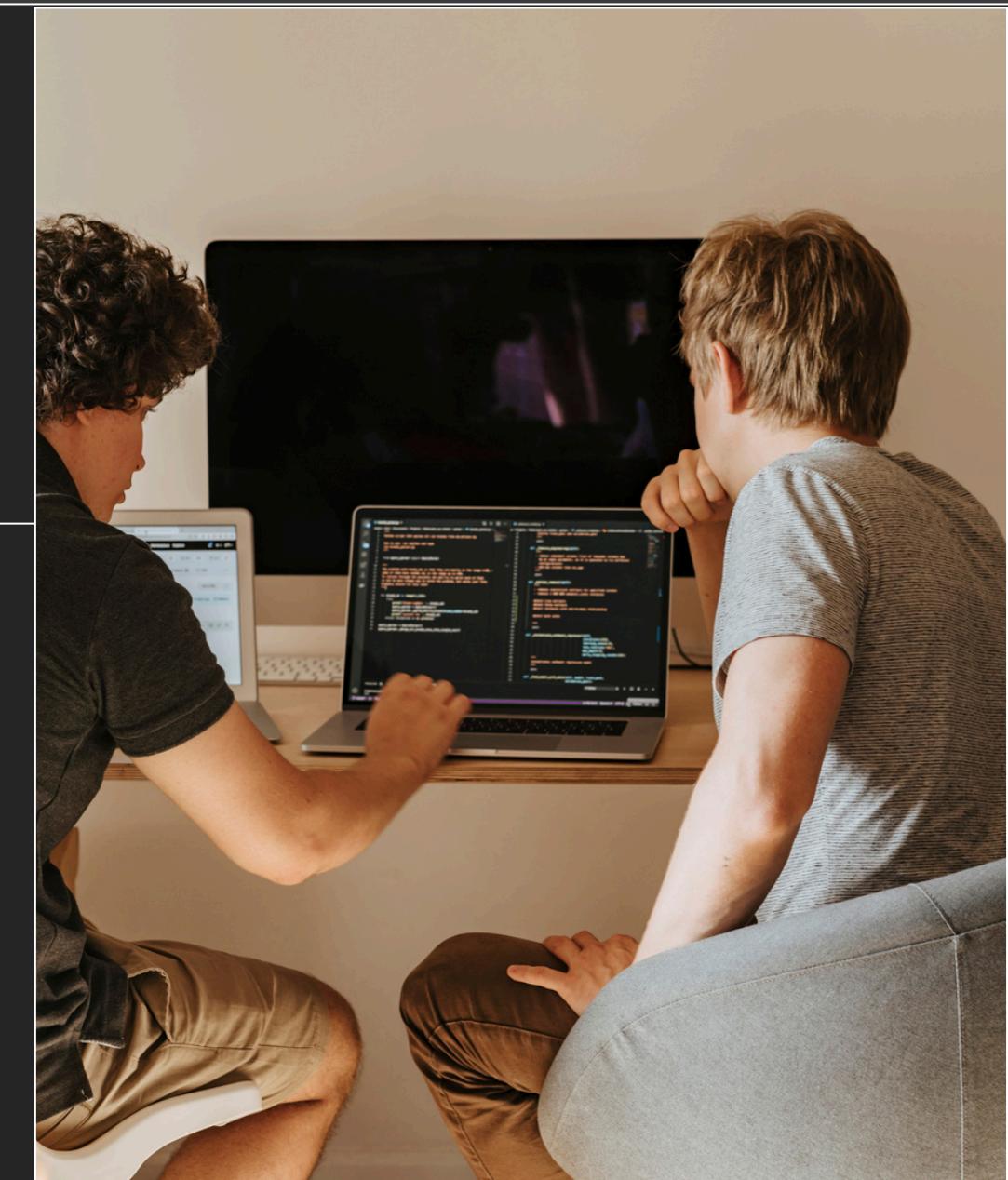
Reflection on Work

Challenges

- Service uptime
- Training pipeline setup
- Tool integration (Kafka, service, load balancer, Prometheus, Grafana)

Potential Fixes

- Multiple Jenkins jobs for each section of pipeline
- 2 approvers per PR
- Explore lightweight tools beyond MLFlow
- Use DVC for dataset management
- Experiment with different models for accuracy



Reflection on Team

Challenges

- Late start on milestones
- Coordinating team meetings in person
- Gathering feedback from TA

Potential Fixes

- Assign 'product manager' for each milestone
- Read report early and start sooner
- Leverage TA for questions and guidance





Questions

