

17-645 Machine Learning in Production

# Group Project: Movie Recommendations

Final Presentation

Team 01 - PopcornTime 

# Our Team



Shazia Parveen



Shun Hang (Alex) Fung



Sreevenk Kovvuri



Venny Ayudiani

# Agenda

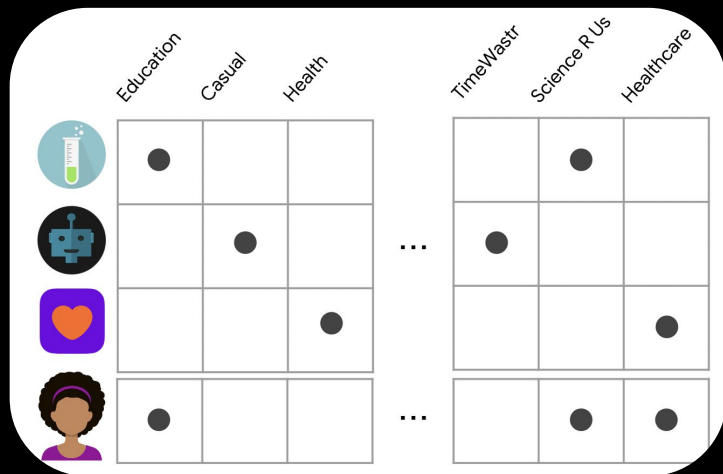
Data Science Key Decisions	Venny
Software Engineering Key Decisions	Sreevenk
Reflection on The Recommender Systems	Shazia
Reflection on Teamwork	Alex
Q & A	

# Data Science Key Decisions

## Content-based Filtering

vs.

## Collaborative Filtering



*Randomly pick 20 movies out of  
top 100 most watched movies*



- + Can help users discover new interests
- + Doesn't need side information
- Cold start problem



## Explicit Collaborative Filtering

based on user's ratings

# surprise

<https://surpriselib.com/>

vs.

## Implicit Collaborative Filtering

based on watch data



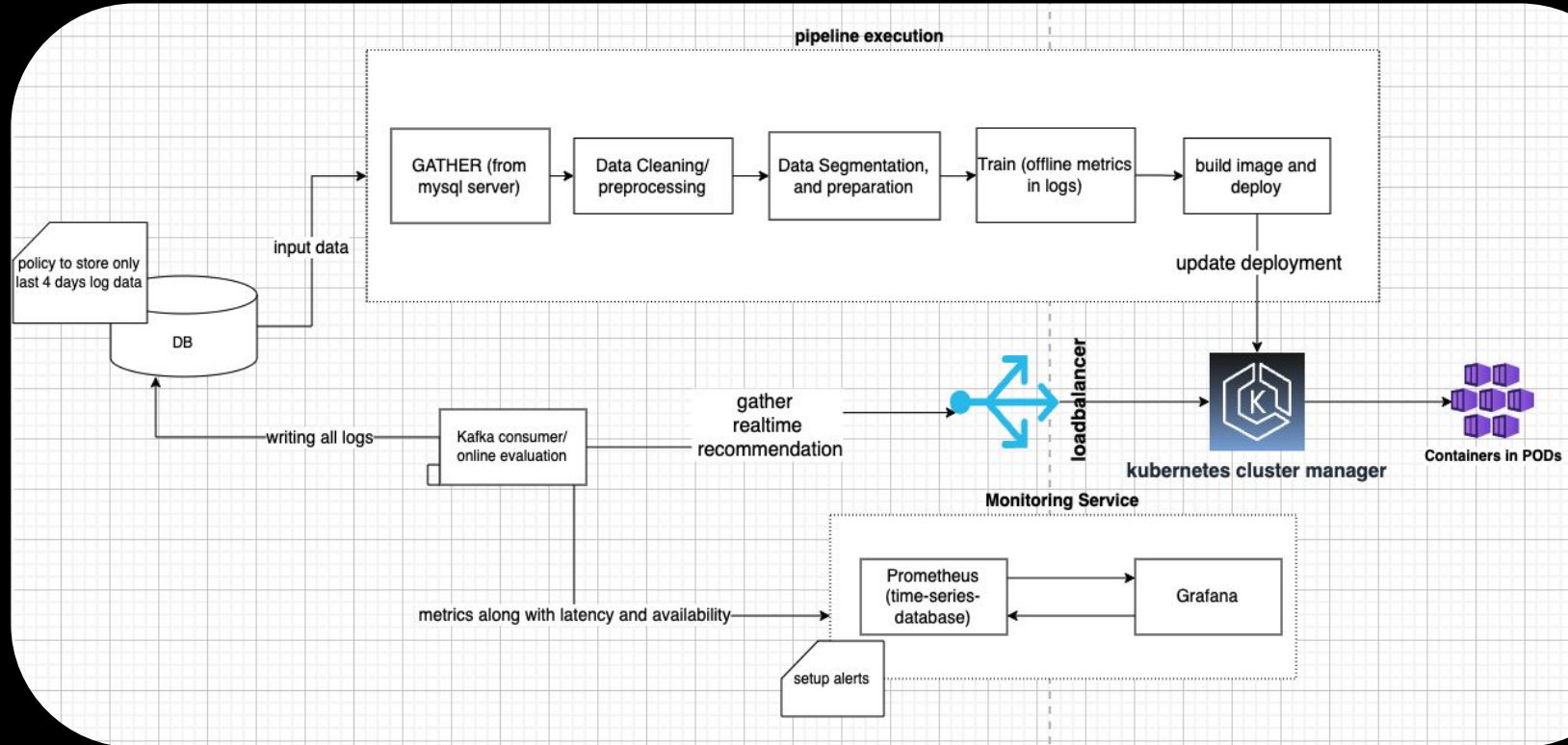
# Implicit

<https://benfred.github.io/implicit/>

- + More data for training
- + Optimized for large data (can handle sparse representation)
- + Faster inference time
- Noisier than rating data (watching does not equal preference)

# Software Engineering Key Decisions

# Architecture Diagram





# Version Control

- DVC - Data and model versioning
- Github - Code versioning, pull requests with code reviews



# Deployment and Load Balancing

- Multiple models with traffic control
- Scale more effectively
- Easier to move to cloud kubernetes services



 FastAPI

# Data Collection

- Unit Tests for each commit
  - Insights with CircleCI
- Training with latest data
  - MySQL server to store latest logs from the kafka stream, with data lifecycle policies to delete old data.
- Monitoring - Grafana & Prometheus
  - Latency, Online Evaluation and Availability metrics are collected from the kafka stream to show up on the dashboard.

# Reflection on The Recommender System

Make trade-offs based on the non-functional requirements of the project

Perform Offline Model Evaluation by  
simulating production environment behavior

Monitoring Service health, availability and  
quality is essential

# Reflection on Teamwork



# Reflection on Teamwork



Ownership of different parts of project



Documents are not organized



Reviewing each other's work



Different preferences of tools



Very Responsive on Slack



Different working hours

**Thank You!**

Q & A