



Carnegie Mellon University

# Productionizing a Recommendation System for a Large Scale use

---

Team4 : The-land-of-harry-potters-amazing-endgame-part-2

***Dhanraj Kotian David Zagardo Ayush Gautam***

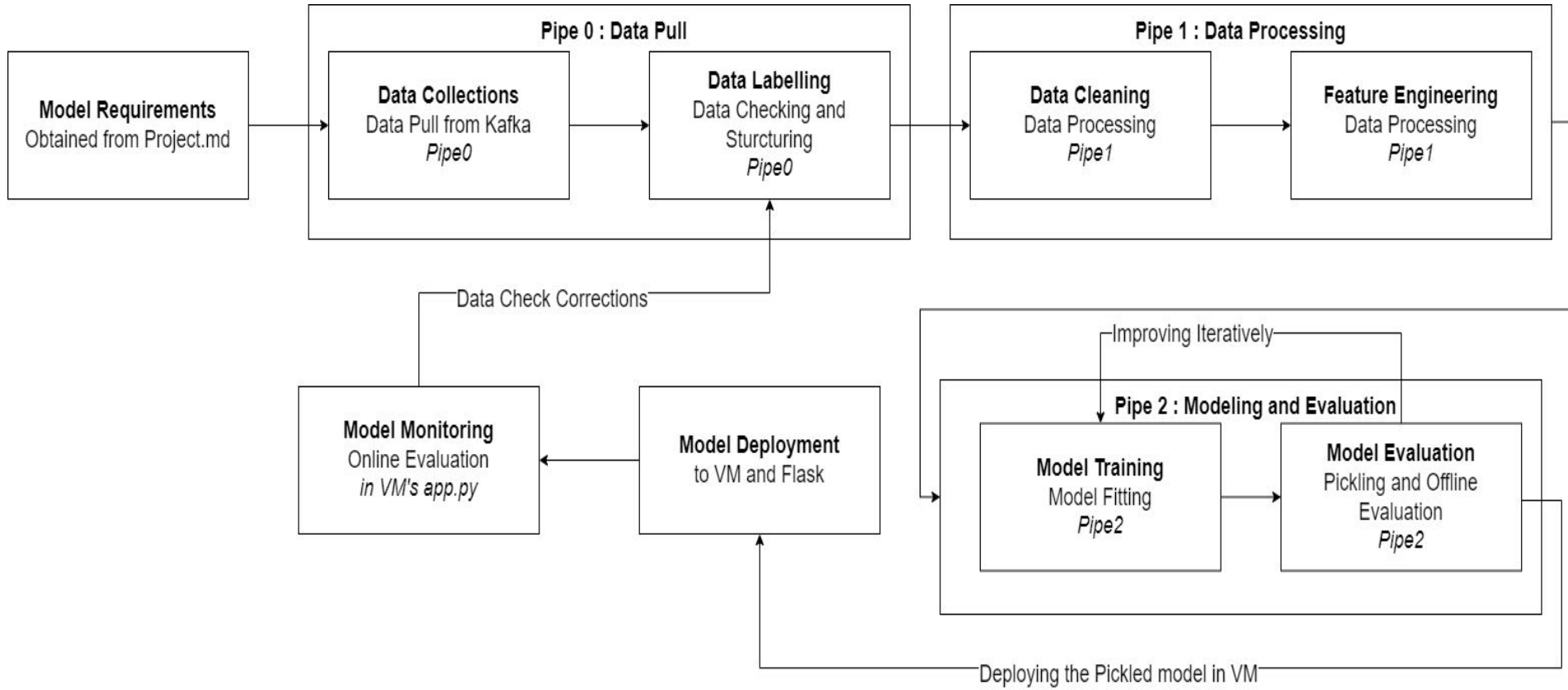
***Davidson Siga Stefan Mason***

*17-645 Machine Learning in Production Fall 2022*

# Agenda

- Design and Reflection
  - Infrastructure
  - Software
  - Model
  - Teamwork

# Pipeline



# Asynchronicity

- Some code for telemetry gathering is executed when a recommendation request is received
- Asynchronous reading/writing speeds up response time, prevents blocking

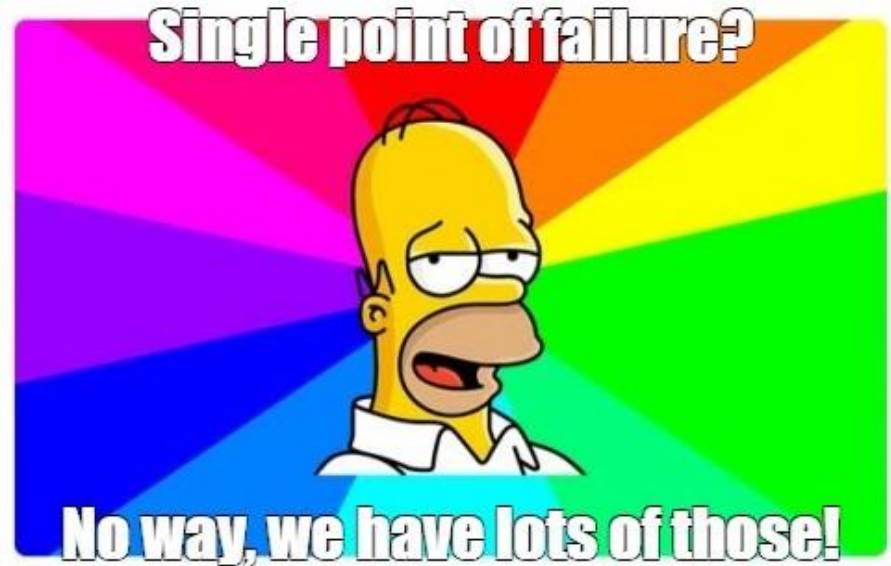
```
import asyncio

async def main():
    print('Hello ...')
    await asyncio.sleep(1)
    print('... World!')

asyncio.run(main())
```

# Infrastructure challenges and learnings

- Single point of failure - Load balancer
- Log file disaster - separate data and application
  - Scalability
  - Maintainability
  - Security

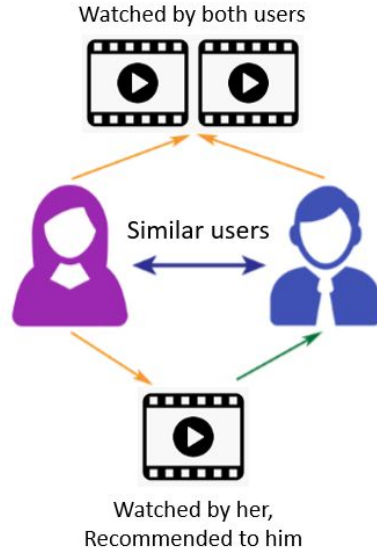


# Model Design Choices

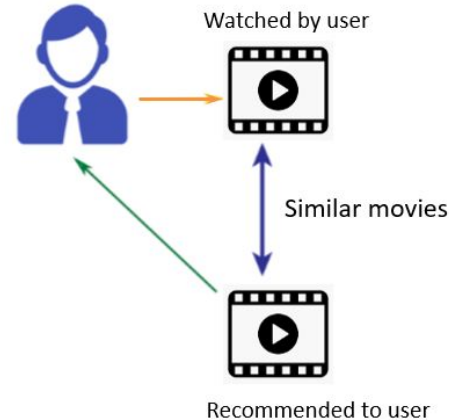
We choose Collaborative Filtering as the recommender system design



Collaborative Filtering

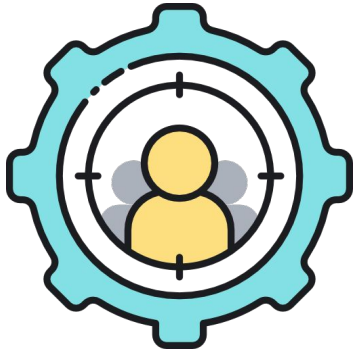


Content-Based Filtering



# Model Design benefits

## Benefits of Collaborative Filtering as compared to Content Based



**Personalization**

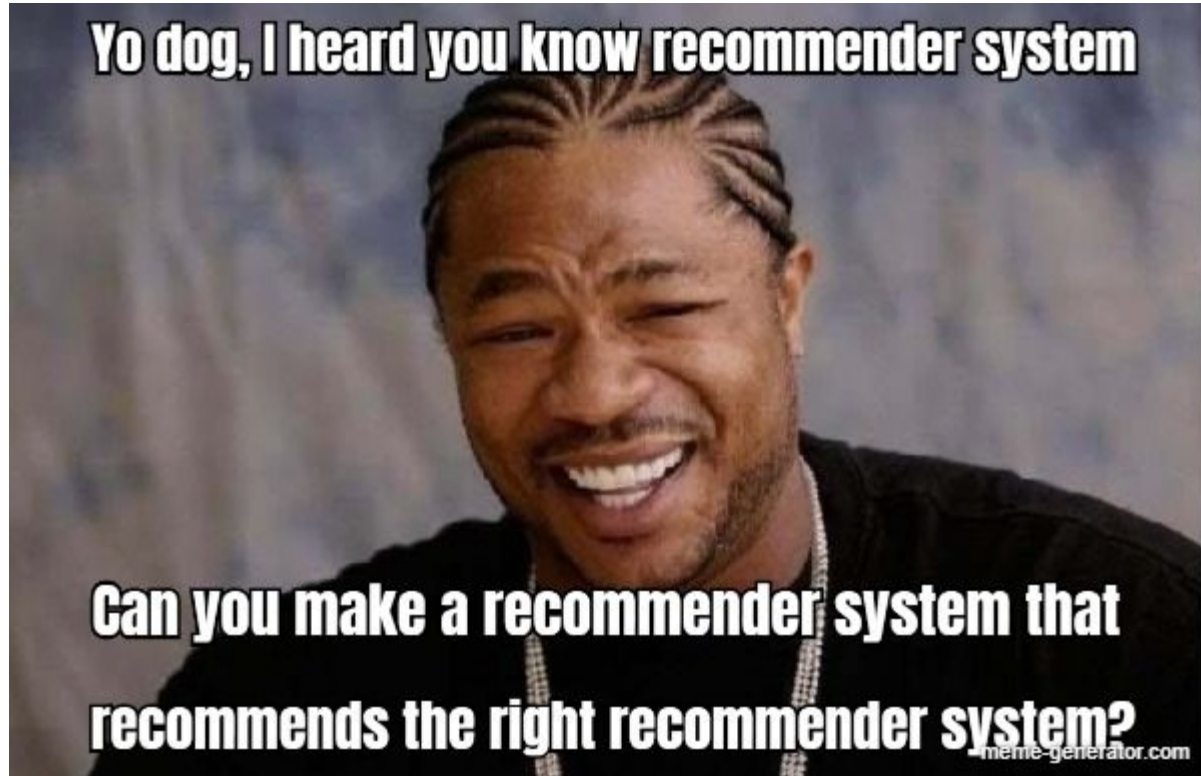


**Scalability**



**Improved diversity**

So many options of Collaborative Filtering! SVD?KNN?SVDPP?NMF?!!





# Model Design Challenges



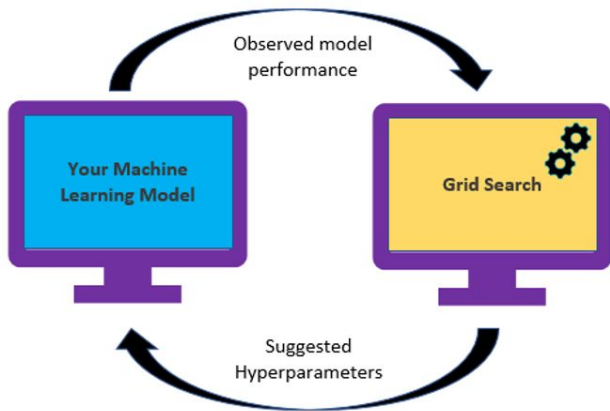
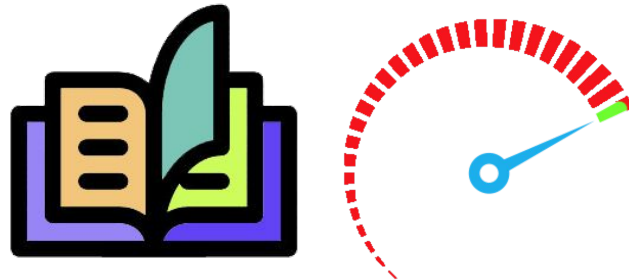
## Dilemma of Time Complexity vs Space Complexity

Movielens 1M	RMSE	MAE	Time
SVD	0.873	0.686	0:01:07
SVD++ (cache_ratings=False)	0.862	0.672	0:41:06
SVD++ (cache_ratings=True)	0.862	0.672	0:34:55
NMF	0.916	0.723	0:01:39
Slope One	0.907	0.715	0:02:31
k-NN	0.923	0.727	0:05:27

**SVD has the best time and space complexity for 1M+ Data**

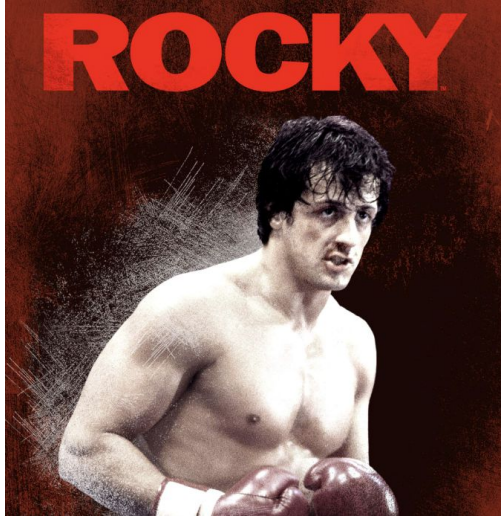
# Interesting Modelling Design Choice Outcomes

Hashmaps/Dictionary based storage retrieval can **drastically improve model inference times**



Grid Search for Hyperparameter tuning helped **improve Model Performance substantially**

# Teamwork





Carnegie Mellon University

Thank you  
Open to Questions