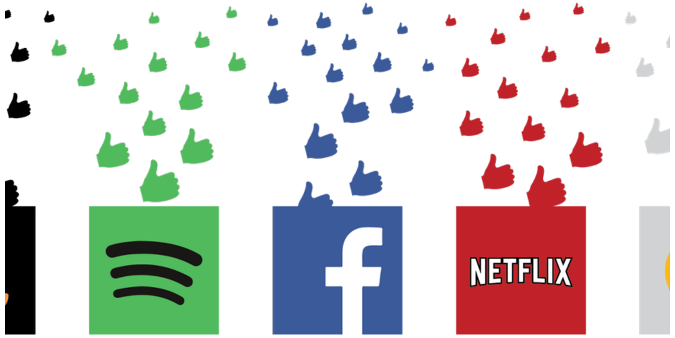
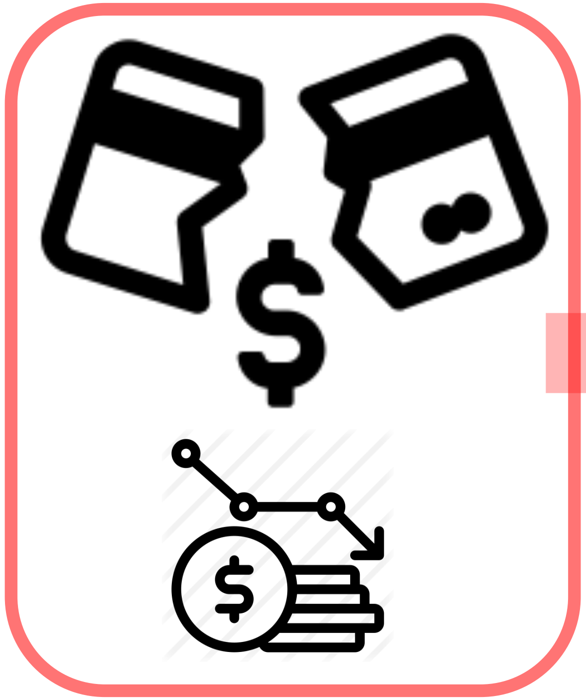
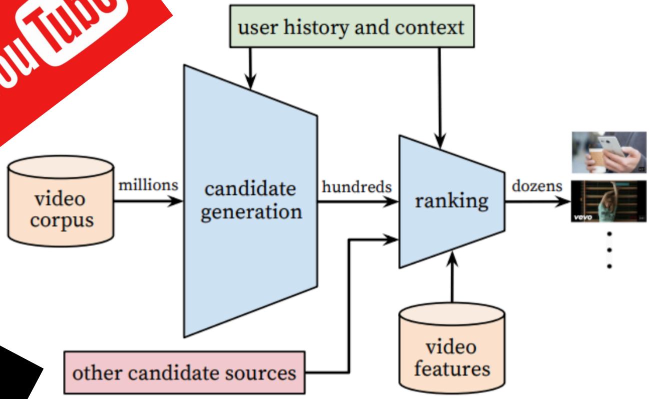



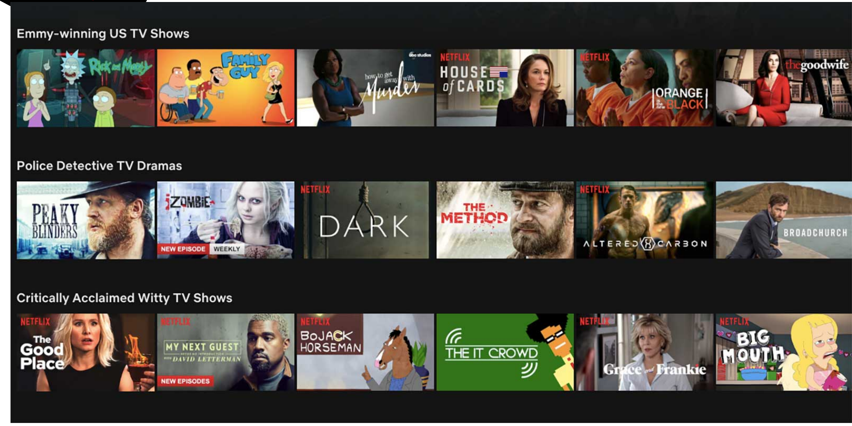

# Recommender System on Amazon Game data

Jae Choi
Contact Info: Jaehyuk0325@gmail.com

# PROBLEMS

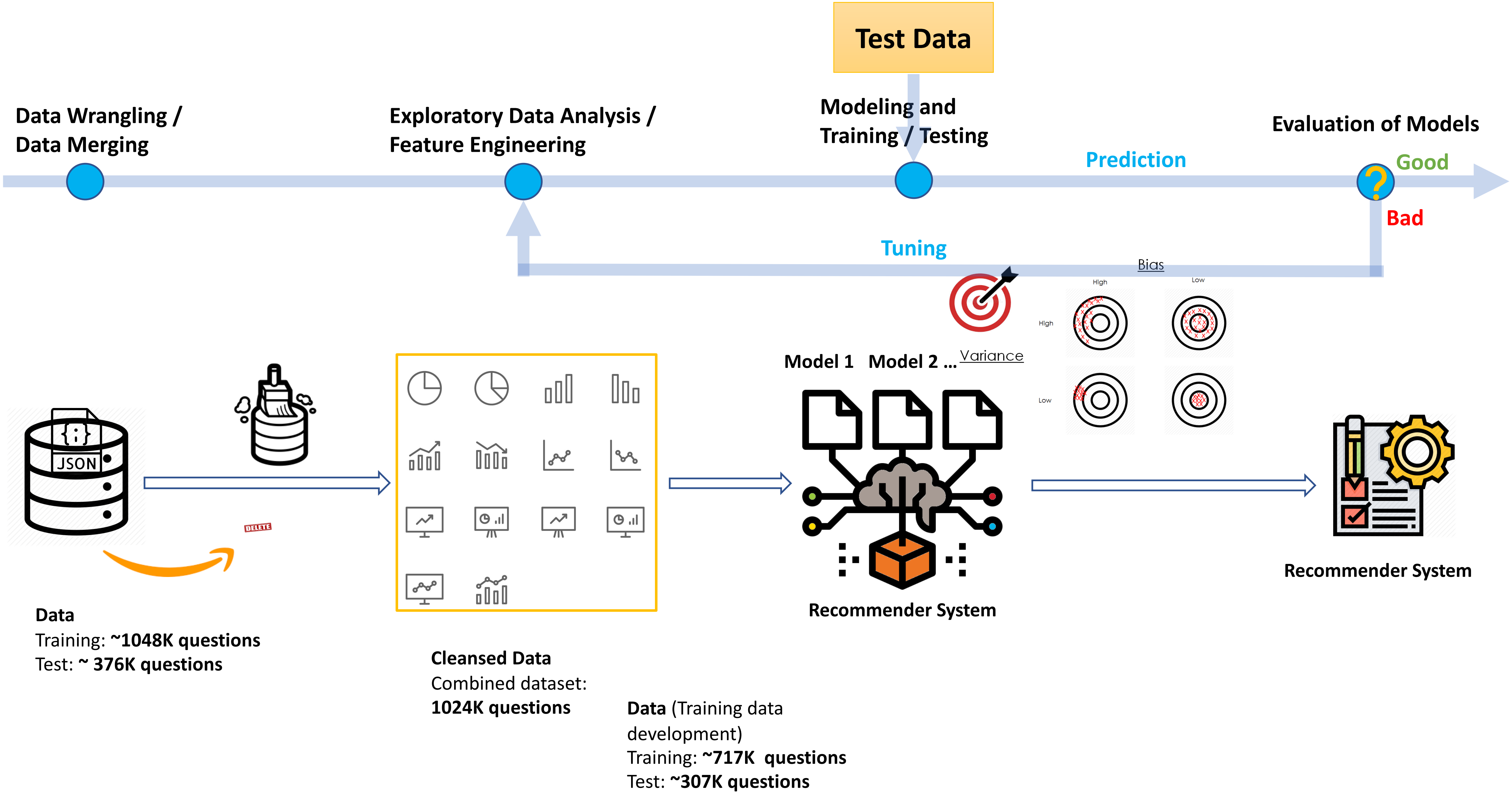



user history and context
video corpus
millions
candidate generation
hundreds
ranking
dozens
other candidate sources
video features


Emmy-winning US TV Shows
Police Detective TV Dramas
Critically Acclaimed Witty TV Shows


amazon.com
Help | Close window
Recommended for You
Inside Apple: How America's Most Admired--and Secretive--Company Really Works
Our Price: $9.99
Used & new from $9.99
See all buying options
Rate this item
I own it
Not interested
Because you purchased...
The Toyota Way : 14 Management Principles from the World's Greatest Manufacturer
(Kindle Edition)
This was a gift
Don't use for recommendations

# STEPS


Test Data
Data Wrangling / Data Merging
Exploratory Data Analysis / Feature Engineering
Modeling and Training / Testing
Evaluation of Models
Prediction
Good
Bad
Tuning
Bias
High
Low
High
Low
Variance
JSON
DELETE
Model 1 Model 2 ...
Recommender System
Recommender System
Data
Training: ~1048K questions
Test: ~ 376K questions
Cleansed Data
Combined dataset:
1024K questions
Data (Training data development)
Training: ~717K questions
Test: ~307K questions

# DATA WRANGLING


Metadata | Ratings

**Merged Data:**
Total Rows: ~2.56 million
# of Users: ~72,000,
# of Items: ~1.54 million

**Data Extraction:**
- 5 years valid data (2014 – 2018)
- Filtering out the reviews of non-purchasers

pandas

**Data Imputation**
**Data Type Conversion**

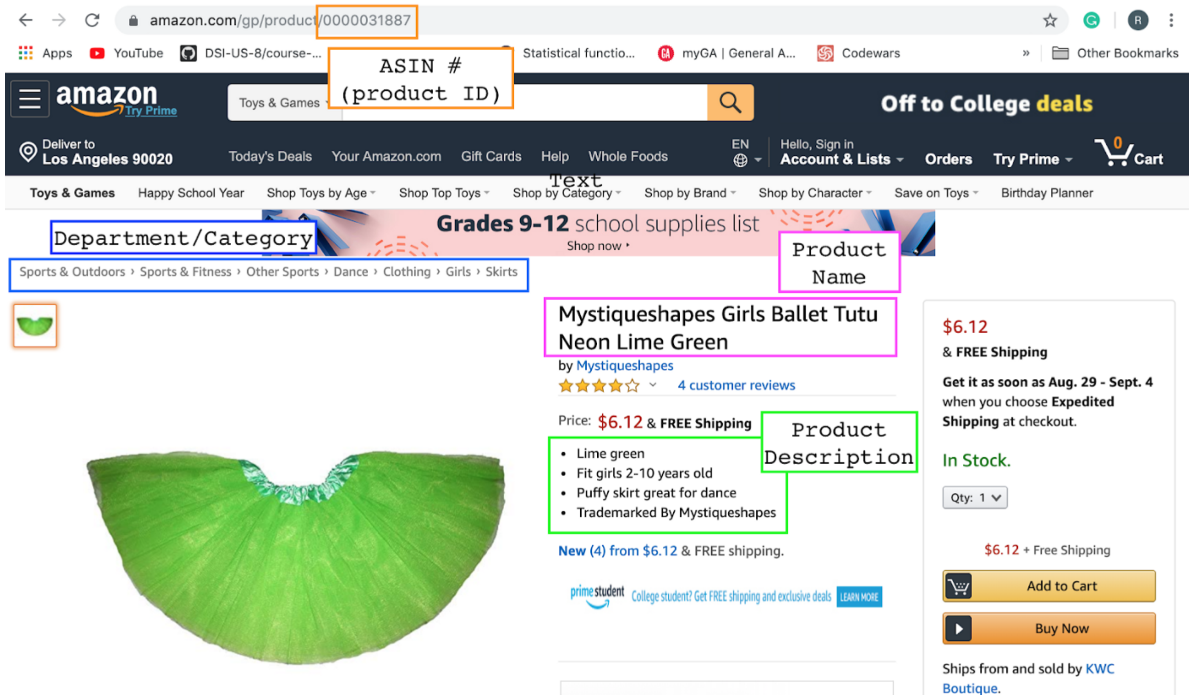
**Cleaned Data:**
Total Rows: ~49,535
# of Users: ~23,789,
# of Items: 241

**User-Item Matrix**

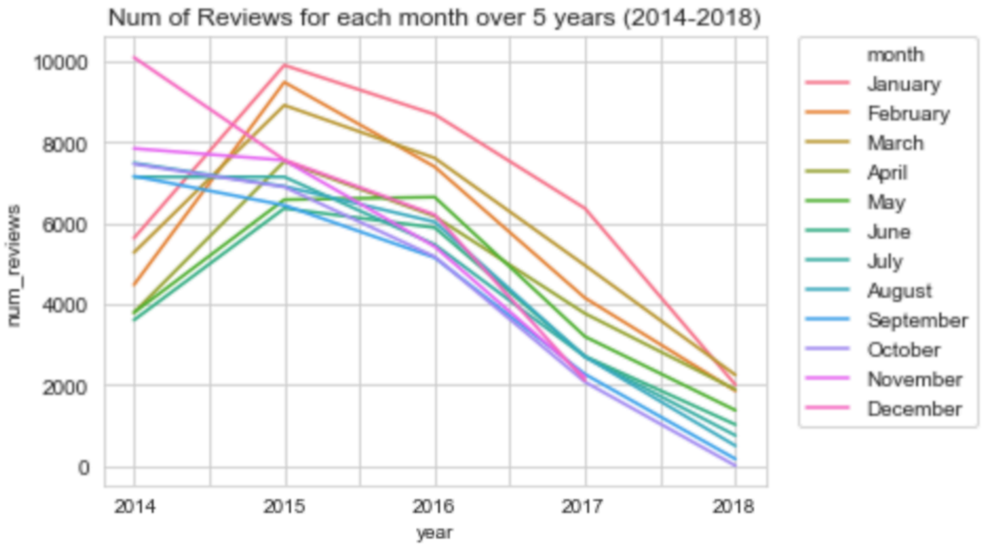
| | Item1 | Item2 | Item3 | Item4 | Item5 |
|---|---|---|---|---|---|
| Alice | 5 | 3 | 4 | 4 | ? |
| User1 | 3 | 1 | 2 | 3 | 3 |
| User2 | 4 | 3 | 4 | 3 | 5 |
| User3 | 3 | 3 | 1 | 5 | 4 |
| User4 | 1 | 5 | 5 | 2 | 1 |

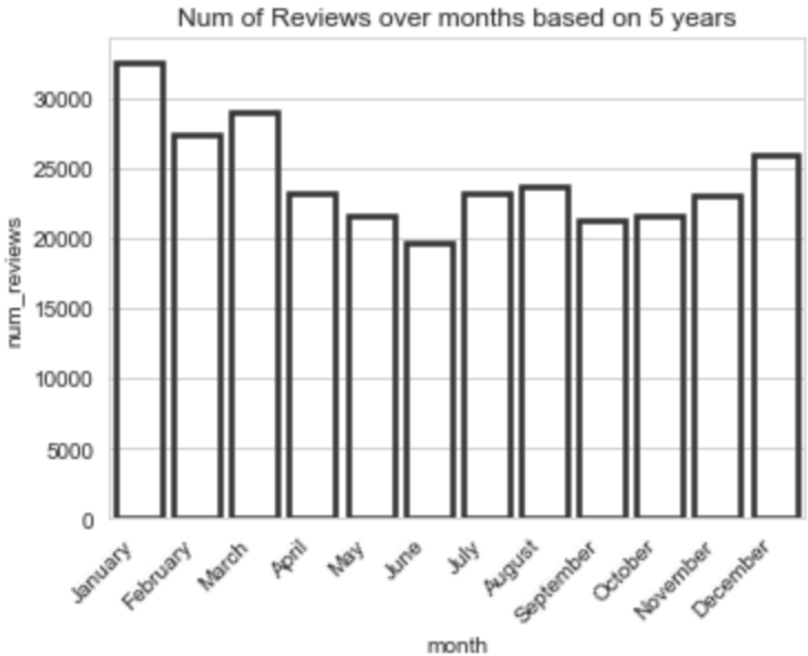
**Sample Review:**

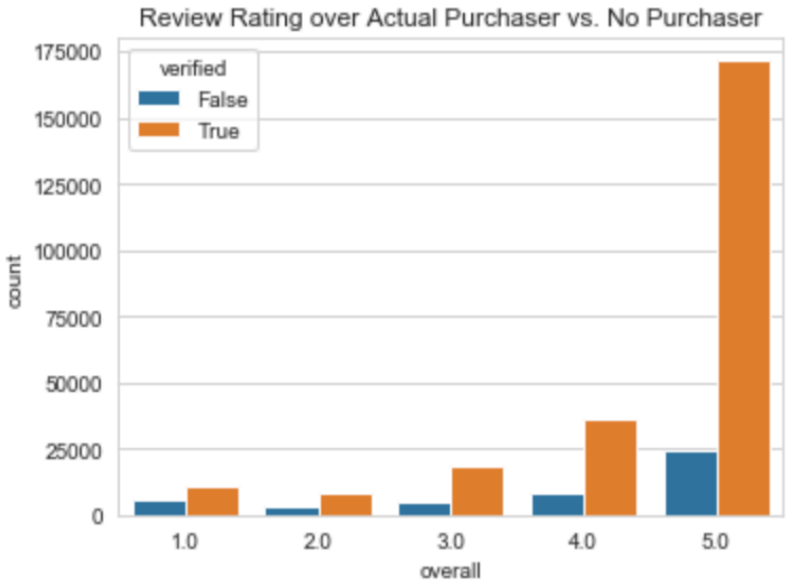
{
"reviewerID": "A2SUAM1J3GNN3B",
"asin": "0000013714",
"reviewerName": "J. McDonald",
"vote": 5,
"style": {"Format:": "Hardcover"},
"reviewText": "I bought this for my husband who plays the piano. He is having a wonderful time playing these old hymns. The music is at times hard to read because we think the book was published for singing from more than playing from. Great purchase though!",
"overall": 5.0,
"summary": "Heavenly Highway Hymns",
"unixReviewTime": 1252800000,
"reviewTime": "09 13, 2009"
}



# EDA


Yearly Reviews (2014 – 2018)

- Rising trend 2014 – 2015
- Falling trend 2015 – 2018
- # of reviews over WINTER season tends >> that over SUMMER season.

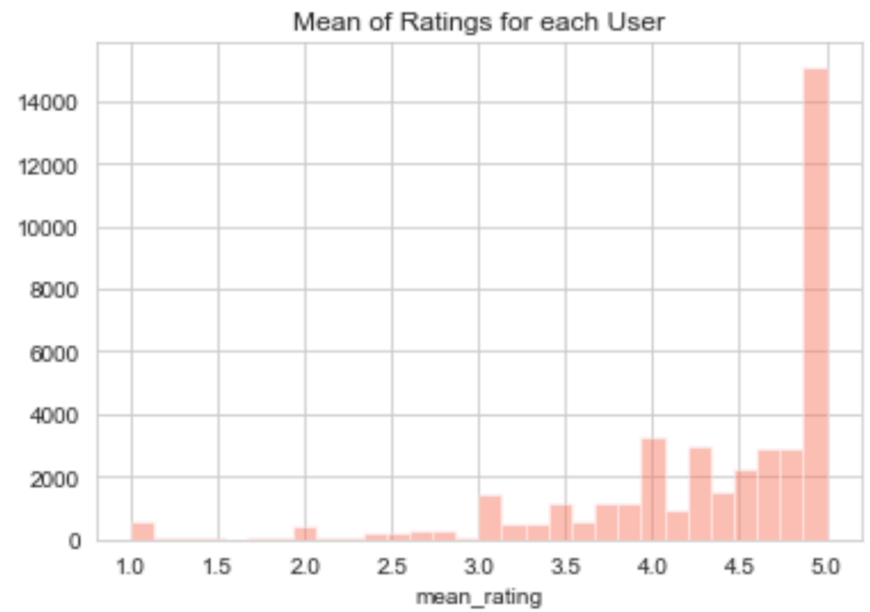
Monthly Reviews (2014 – 2018)

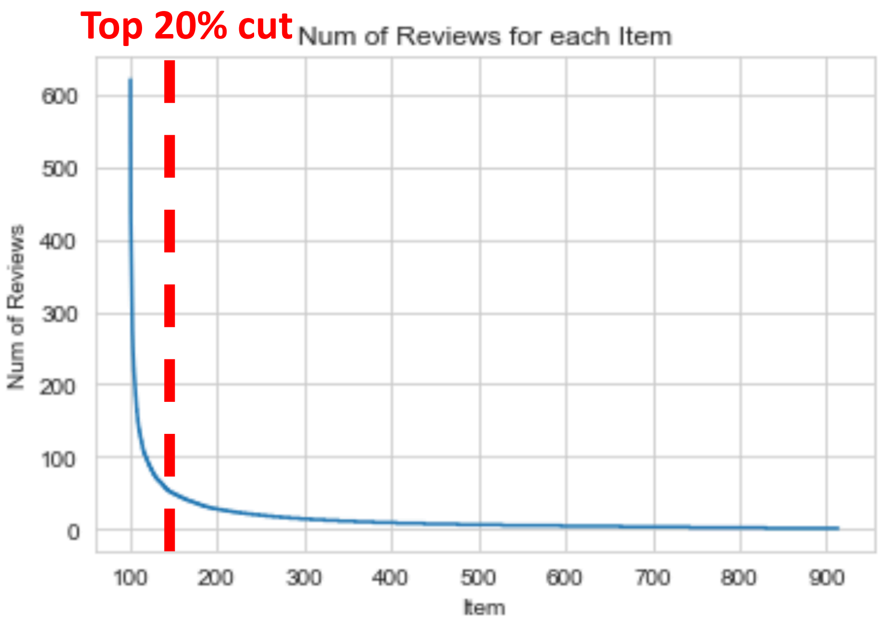
- Increment over SUMMER break.
- Falling trend Jan – June

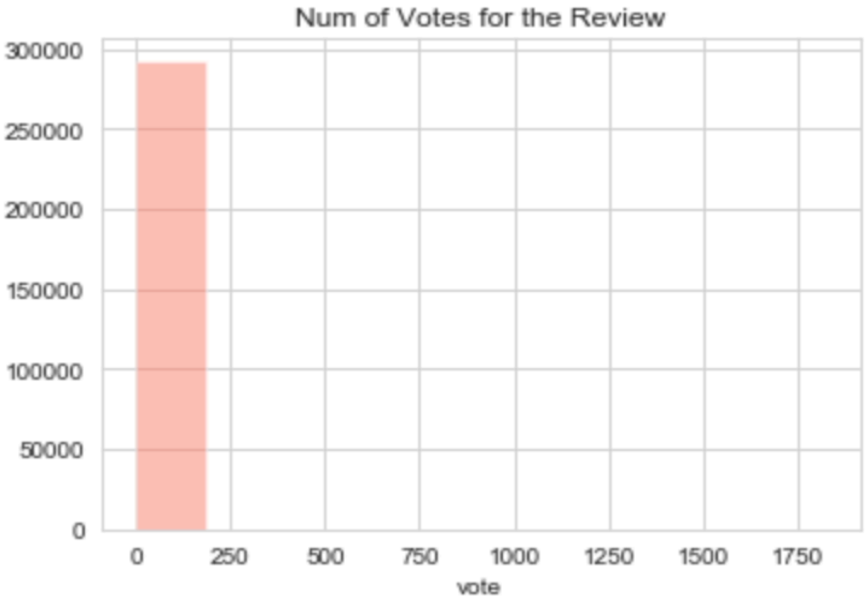
Real Buyer vs. No Buyer

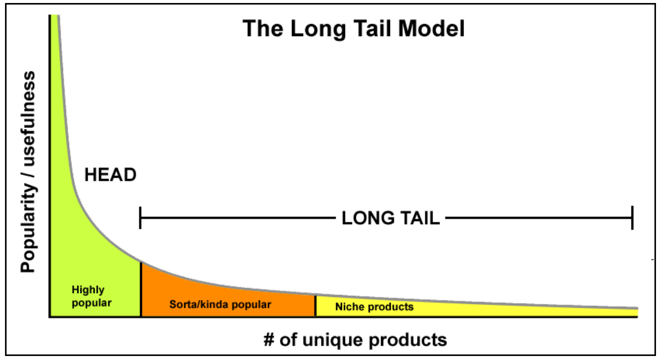
- Both are LEFT skewed
- Data of real buyers is skewed more dramatically.

# EDA




Mean of Reviews (2014 – 2018)

- Trend over only real buyers.

# of Reviews for Each Item

- Popularity cut top 20 %

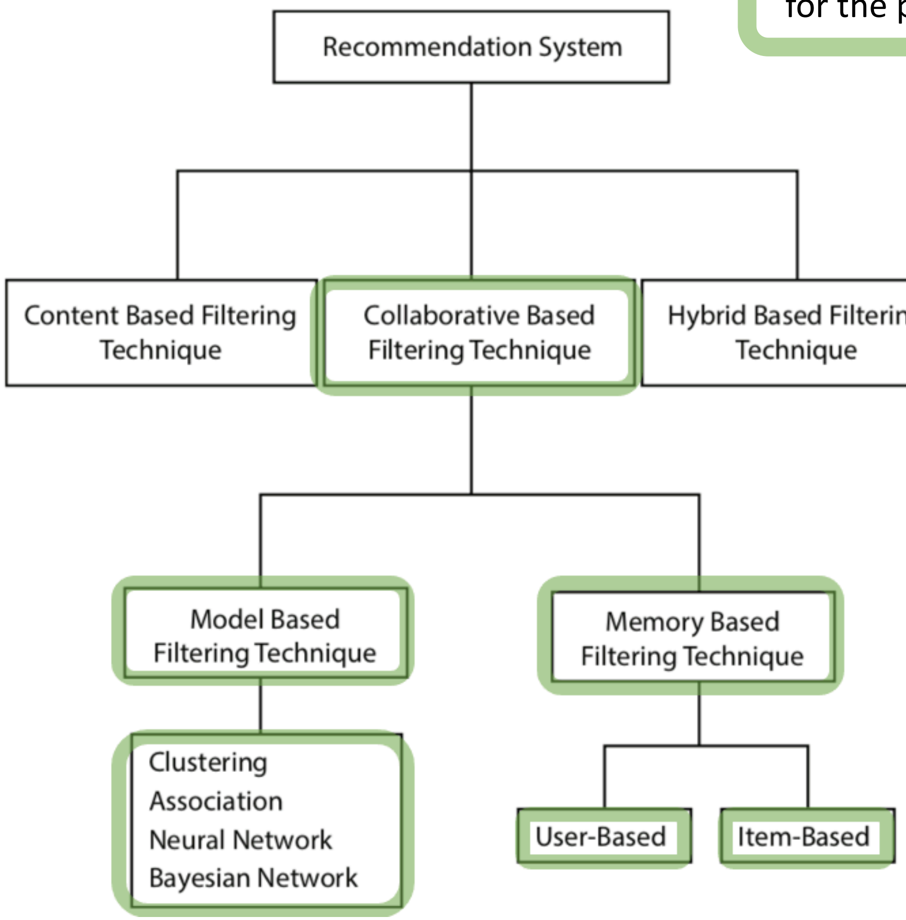
# of Votes for Reviews

- Range of # of votes (Mostly < 250).
- Some OUTLIERs over 1500 votes

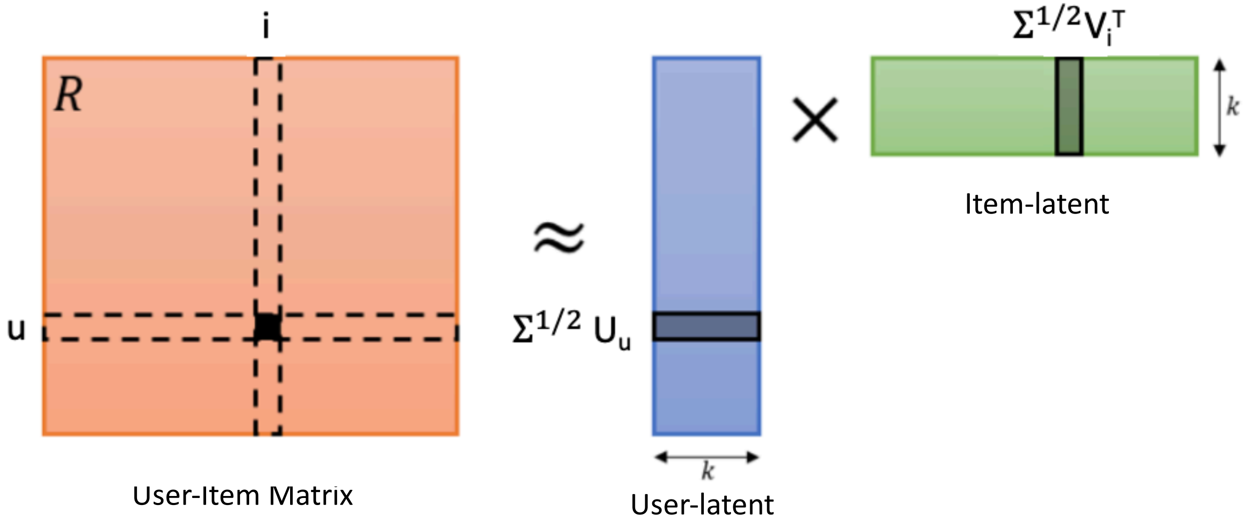
# MODELING

# Recommender System Types

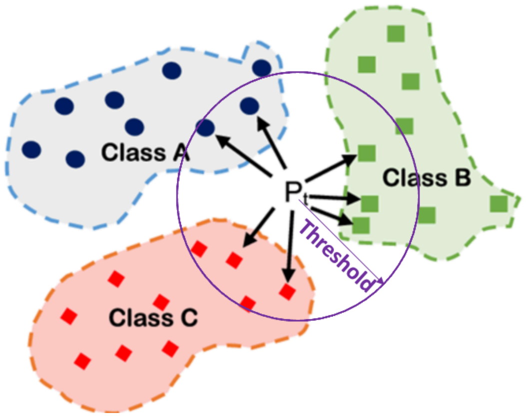
Models that I used for the project




Model-based:
**Matrix Factorization**




Memory-based:
**KNN**




| | Item1 | Item2 | Item3 | Item4 | Item5 |
|---|---|---|---|---|---|
| Alice | 5 | 3 | 4 | 4 | ? |
| User1 | 3 | 1 | 2 | 3 | 3 |
| User2 | 4 | 3 | 4 | 3 | 5 |
| User3 | 3 | 3 | 1 | 5 | 4 |
| User4 | 1 | 5 | 5 | 2 | 1 |

Similarity

| | Item1 | Item2 | Item3 | Item4 | Item5 |
|---|---|---|---|---|---|
| Alice | 5 | 3 | 4 | 4 | ? |
| User1 | 3 | 1 | 2 | 3 | 3 |
| User2 | 4 | 3 | 4 | 3 | 5 |
| User3 | 3 | 3 | 1 | 5 | 4 |
| User4 | 1 | 5 | 5 | 2 | 1 |

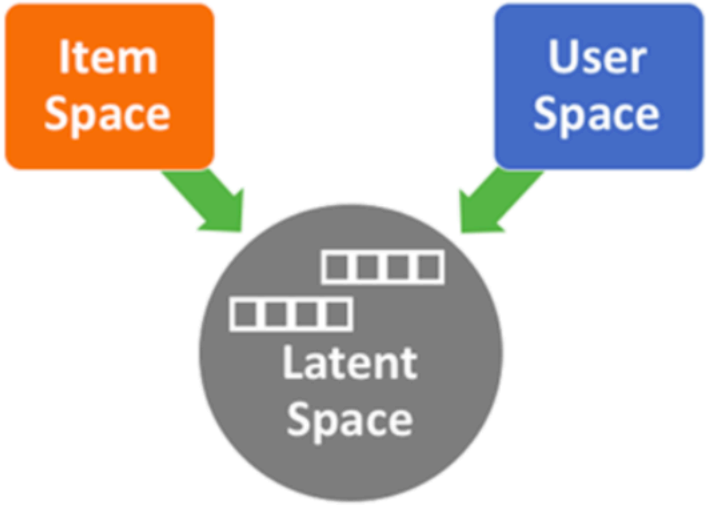
Similarity

# MODELING

**Content-based Filtering**

**Hybrid Filtering**

### Model-based CF




Latent Model

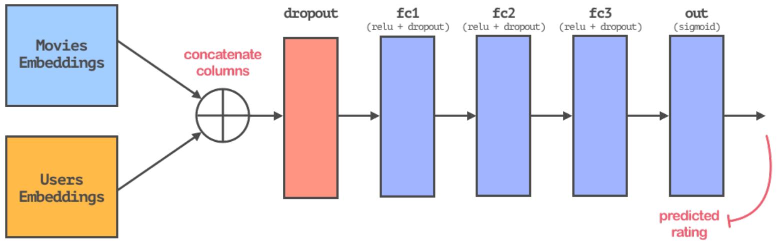
- SVD
- SVD++
- NMF

### Memory-based CF


Neighborhood Model

- KNN
- KNN with Means
- KNN with Z-score

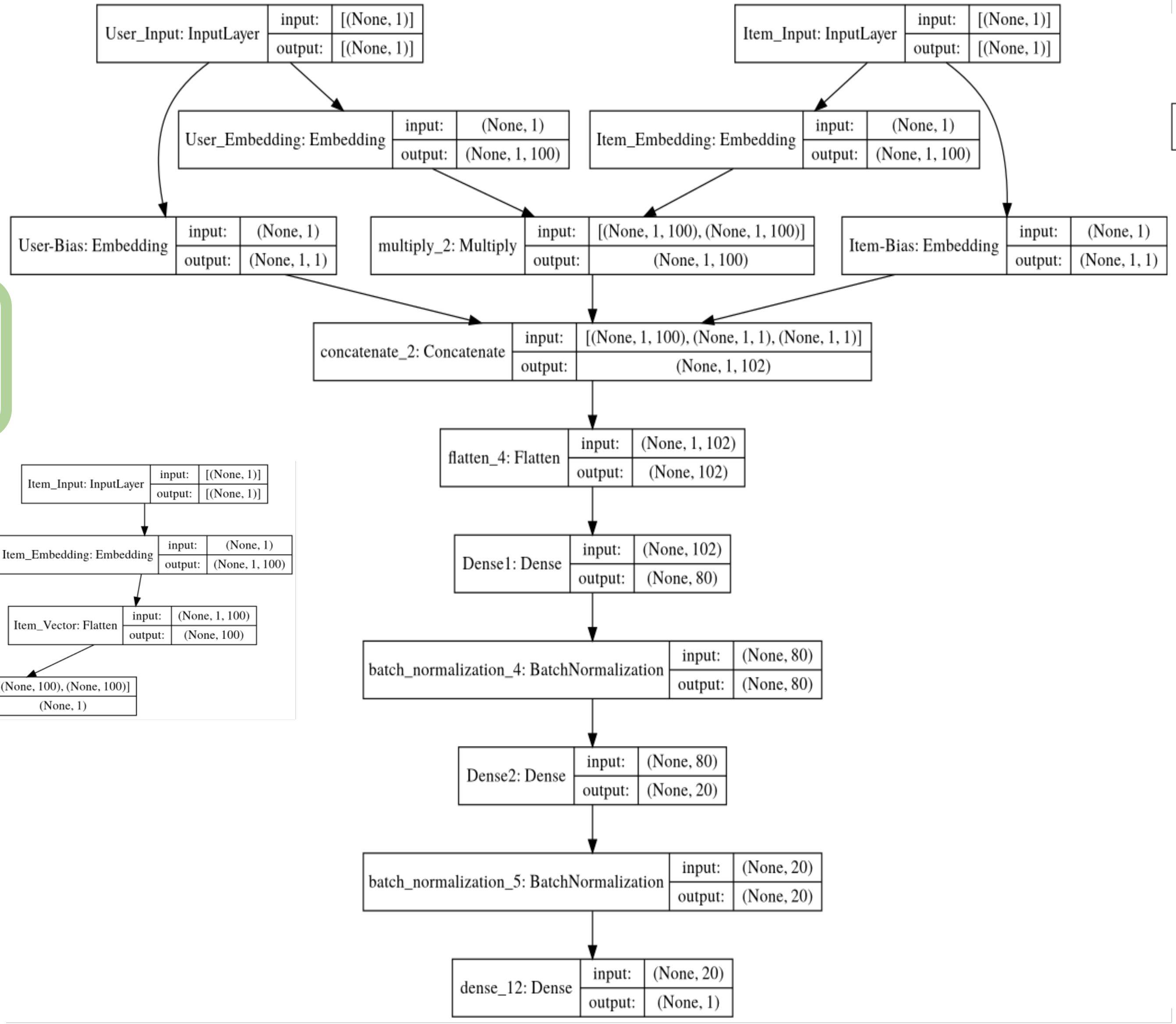
### Deep Learning based Recommender System



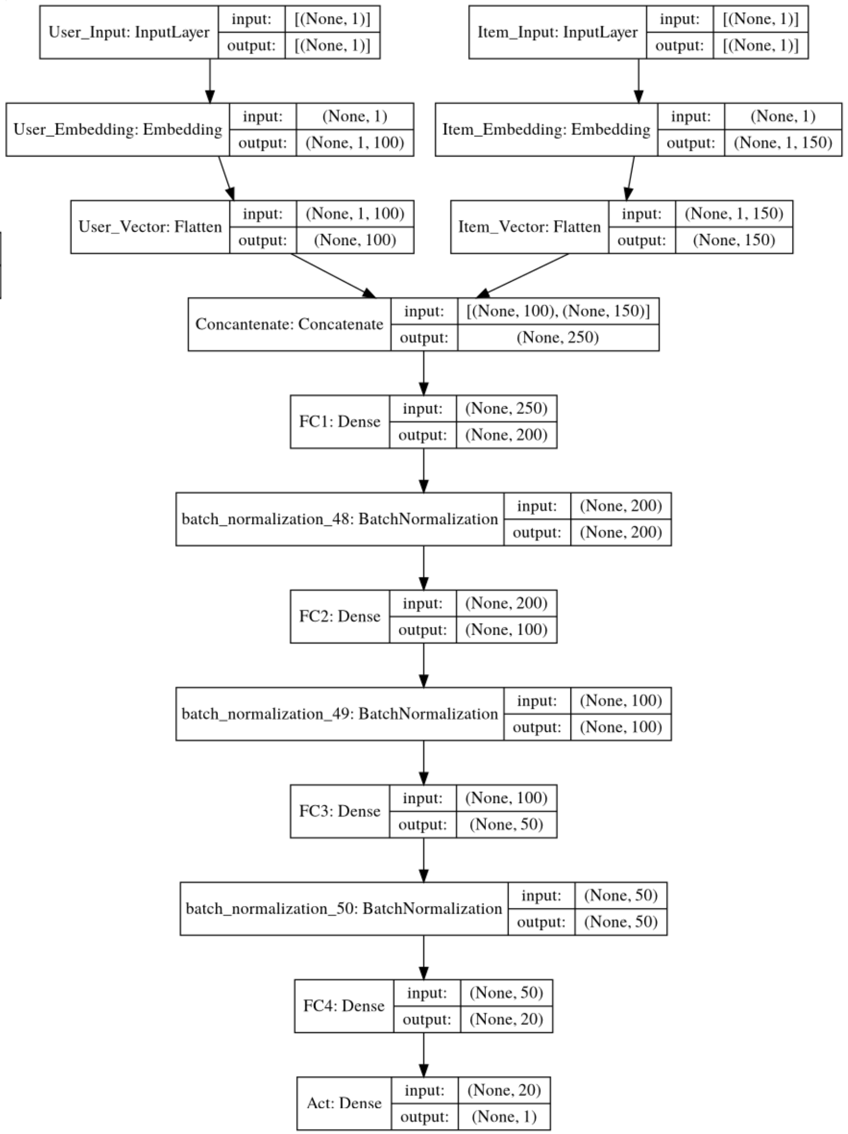

- Neural Matrix Factorization 1
- Neural Matrix Factorization 2
- Neural Matrix Factorization 3

# MODELING

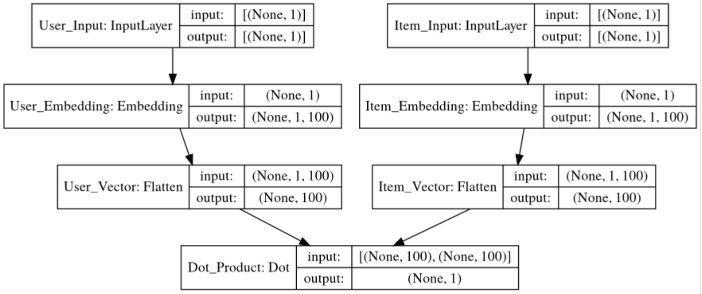
N-MF2




N-MF3




**N-MF1**
(Neural-Matrix Factorization)



# Evaluation of Models

## Model Evaluation Metrics

- **RMSE**: Root Mean Squared Error. Lower value is better.
- **MAE**: Mean Absolute Error. Lower value is better.
- **Precision@k**: number of recommended items that are actually rated within top-over number of top-k. Higher value is better.
- **Recall@**k: number of recommended items that are relevant over number of recommended items. Higher value is better.
- **HR**: Hit Rate; how often we are able to recommend a left-out rating. Higher is better.
- **cHR**: Cumulative Hit Rate; hit rate, confined to ratings above a certain threshold. Higher is better.
- **Coverage**: Ratio of users for whom recommendations above a certain threshold exist. Higher is better.
- **Diversity**:1−S, where Sis the average similarity score between every possible pair of recommendations for a given user. Higher means more diverse.
- **Novelty**: Average popularity rank of recommended items. Higher means more novel.

$$Prec@k = \frac{\text{\# of Relevant Items Recommended in top-k}}{\text{k (\# of Items that Recommended)}}$$

$$Recall@k = \frac{\text{\# of Relevant Items Recommended in top-k}}{\text{\# of Relevant Items}}$$

$$HR = \frac{\text{\# of Hit}}{\text{\# of Users}}$$

$$cHR = \frac{\text{\# of Hits of ratings above threshold}}{\text{\# of Users}}$$

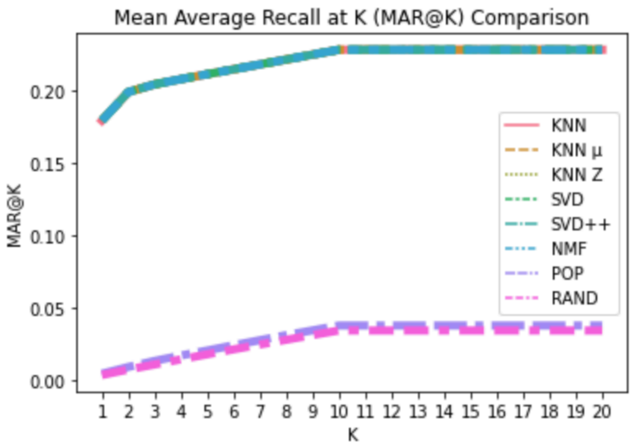
$$Cvg = \frac{\text{\# of Users above threshold}}{\text{\# of Users}}$$

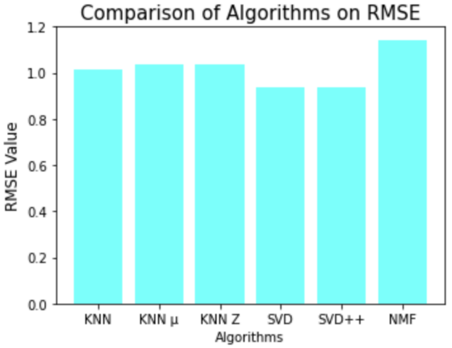
$$Dvr = 1 - S$$

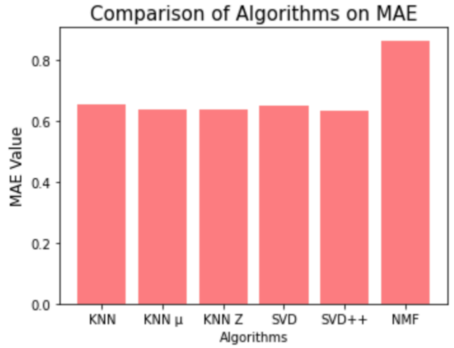
$$Nvl = \sum_{i=0}^{k} \frac{\text{Rating(i) that current User gave}}{\text{Rank(i)}}$$

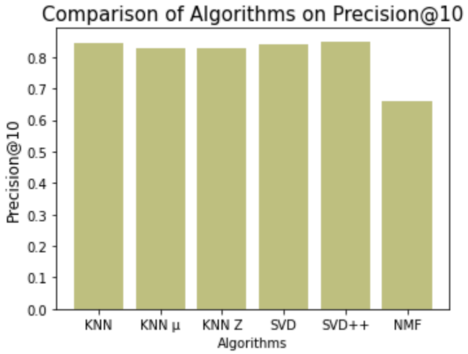
# EVALUATION OF MODELS

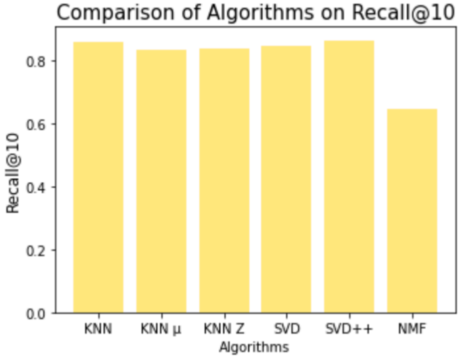
## Model Performance Comparison



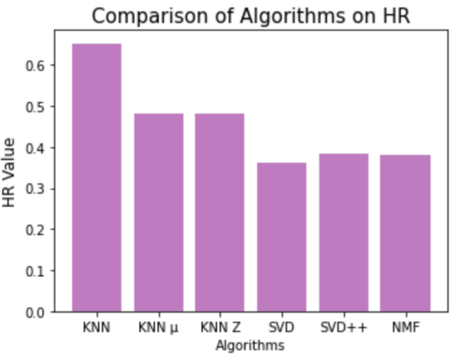



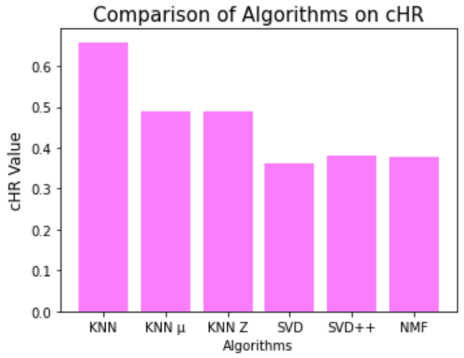



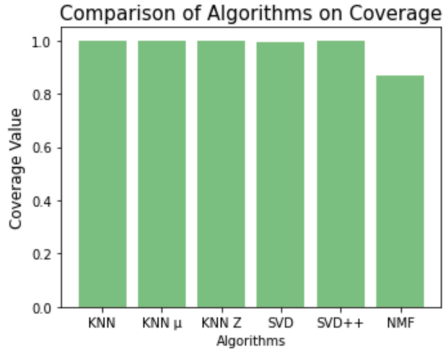



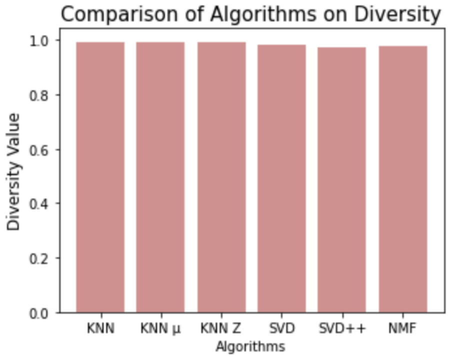



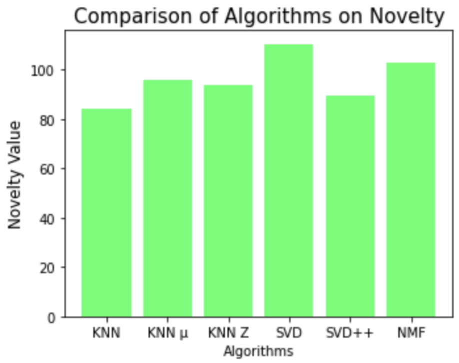

| Model | RMSE | MAE | Pr@10 | Re@10 | HR | cHR | Cvg | Dvs | Nvl |
|---|---|---|---|---|---|---|---|---|---|
| KNN | 1.012 | 0.656 | 0.843 | 0.857 | 0.652 | 0.659 | 1.0 | 0.991 | 84.2 |
| KNN µ | 1.037 | 0.638 | 0.827 | 0.835 | 0.481 | 0.488 | 1.0 | 0.993 | 96.0 |
| KNN Z | 1.037 | 0.641 | 0.829 | 0.836 | 0.480 | 0.488 | 1.0 | 0.993 | 94.1 |
| SVD | 0.935 | 0.651 | 0.838 | 0.848 | 0.360 | 0.361 | 0.99 | 0.980 | 110.5 |
| SVD++ | 0.938 | 0.634 | 0.850 | 0.864 | 0.384 | 0.381 | 1.0 | 0.970 | 89.6 |
| NMF | 1.143 | 0.866 | 0.659 | 0.646 | 0.382 | 0.379 | 0.87 | 0.977 | 103.0 |



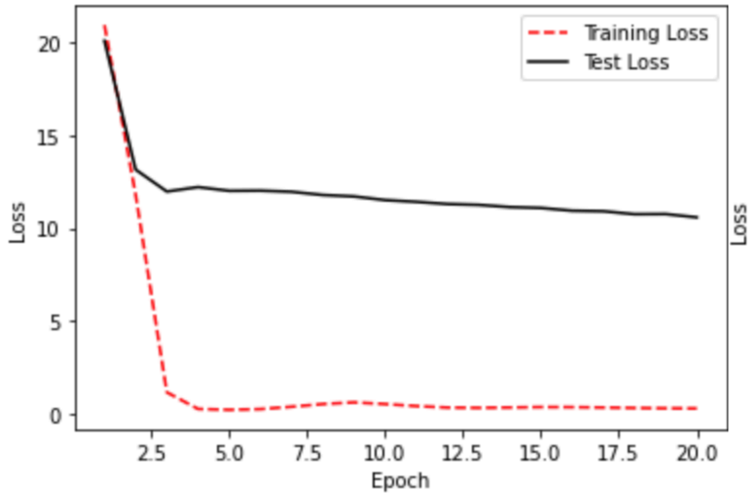



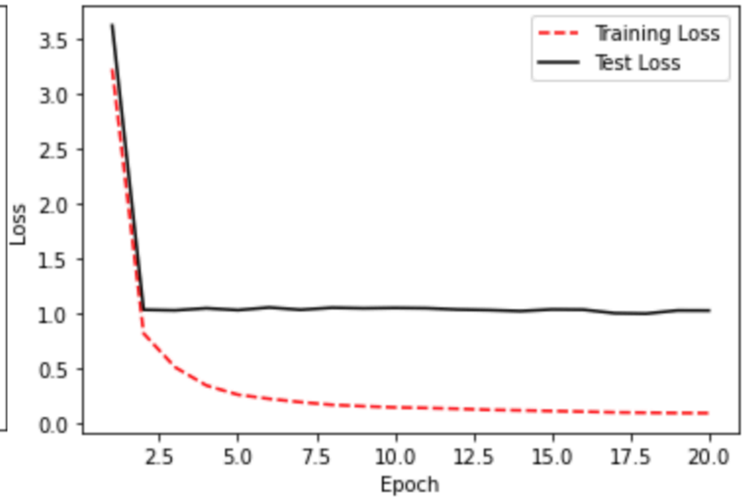
# EVALUATION OF MODELS

## Model Performance Comparison

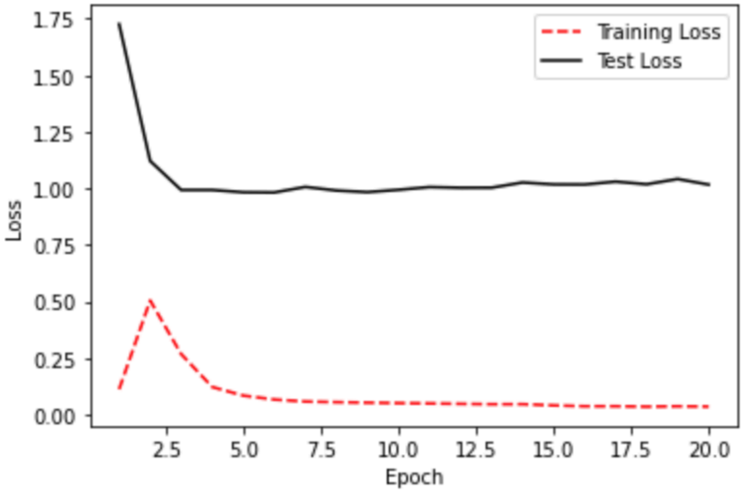
Training Loss vs. Test Loss for N-MF1




Training Loss vs. Test Loss for N-MF3




Training Loss vs. Test Loss for N-MF2




| Model | RMSE |
|---|---|
| KNN | 1.012 |
| KNN µ | 1.037 |
| KNN Z | 1.037 |
| SVD | 0.935 |
| SVD++ | 0.938 |
| NMF | 1.143 |
| N-MF1 | 3.254 |
| N-MF2 | 1.008 |
| N-MF3 | 1.011 |

N-MF1
- Underfitting

He Normal Initialization
Lower Regularization Rate
Add Bias
Add 2 Fully Connected Layers
ReLU
Batch Normalization

N-MF2
- Overfitting

He Normal Initialization
Same Regularization Rate
Concat two latents
add 4 Fully Connected Layers
ReLU
Batch Normalization

N-MF1
- Generalized (compared first two)

# FUTURE WORK

- **Various Metrics**: N-MF models can be evaluated with other metrics including Rank-less and Rank-aware metrics.

NOTE: it is suggested to use Rank-less like Precision@k, Recall@k, and Hit@k, and Rank-aware like MAP(Mean Average Precision) and nDCG (normalized Discounted Cumulative Gain) with penalizing rank; red mark is the unused metrics for this projects.

- **Time Comparison**: Comparison of training time can bring more values on this project.

- **More Model Comparison**: The project can be expanded with Content-based and Hybrid Filtering.