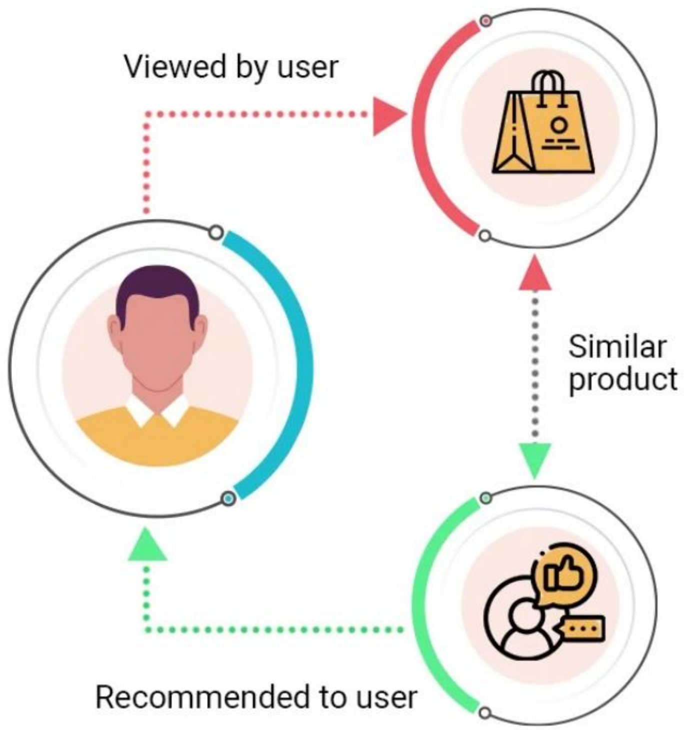
RESTAURANT RECOMMENDATION SYSTEM

WHAT IS A RECOMMENDATION SYSTEM?

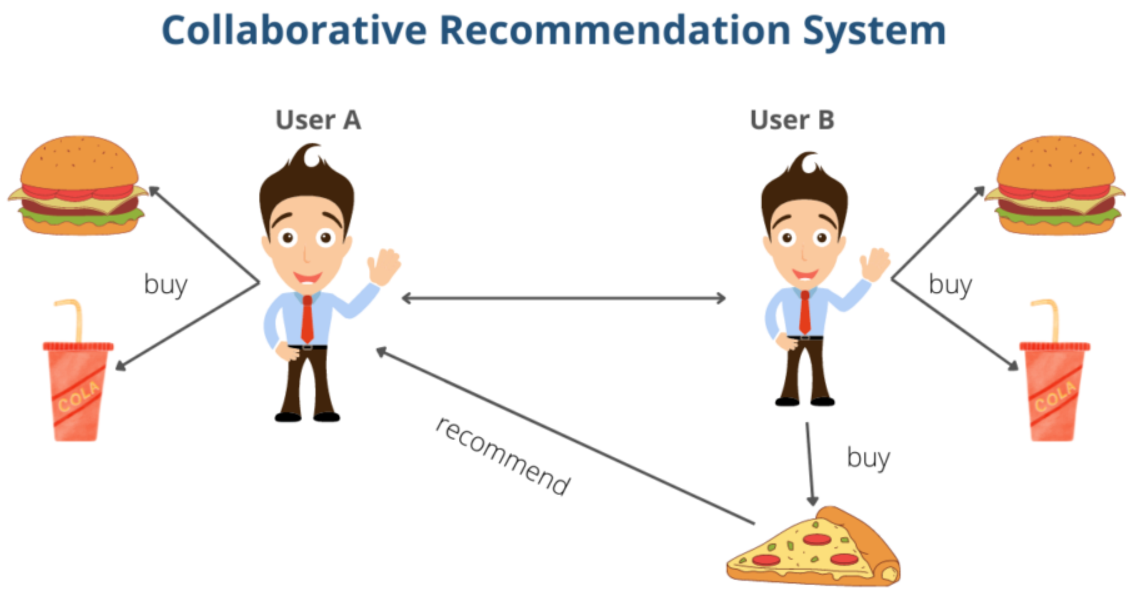
RECOMMENDATION SYSTEM IS AN INFORMATION FILTERING SYSTEM THAT SEEKS TO PREDICT THE RATING OR PREFERENCE A USER WOULD GIVE TO AN ITEM

IT ASSISTS USERS IN DISCOVERING NEW ITEMS THAT THEY MAY FIND APPEALING



APPLICATIONS

- Streaming Service
- E-Commerce Service
- Tourism Service
- Healthcare Service
- Education Service
- Food and Beverages



OUR PROBLEM STATEMENT

The system utilizes the user's historical ratings of multiple restaurants

Based on this data, the system recommends new restaurants that the user has not tried before and predicts that the user will give a high rating

DATASET – YELP

Popular platform for restaurant and business ratings and reviews

5 json files – information about users, businesses, reviews, checkins, tips

~8GB data with many US states and cities

Huge dataset with high sparsity

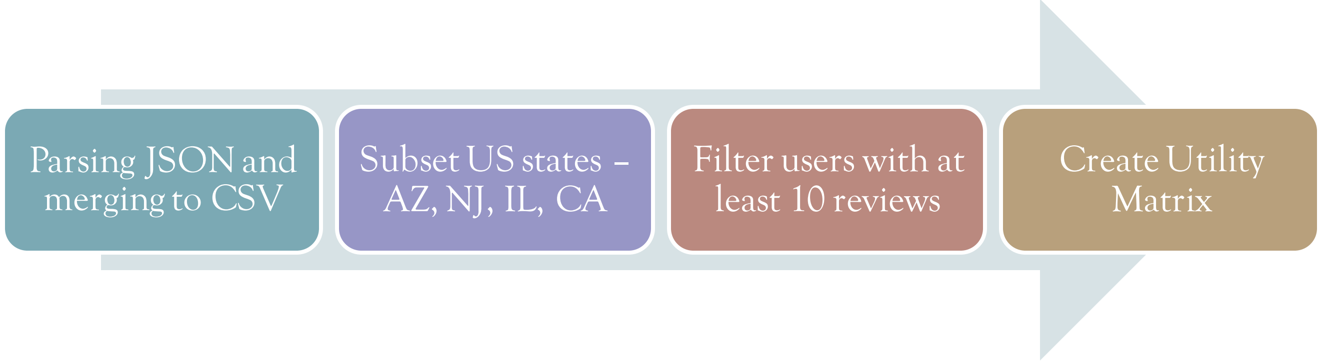
DATA PROCESSING

User id	Item id	Rating
1	a	1
1	b	2
2	a	3
2	c	2
3	c	5

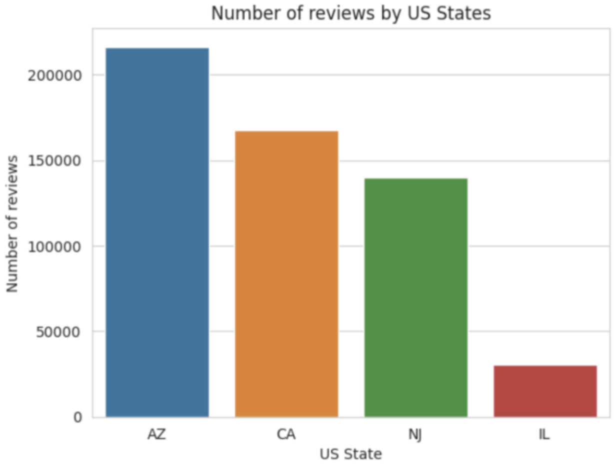
Data

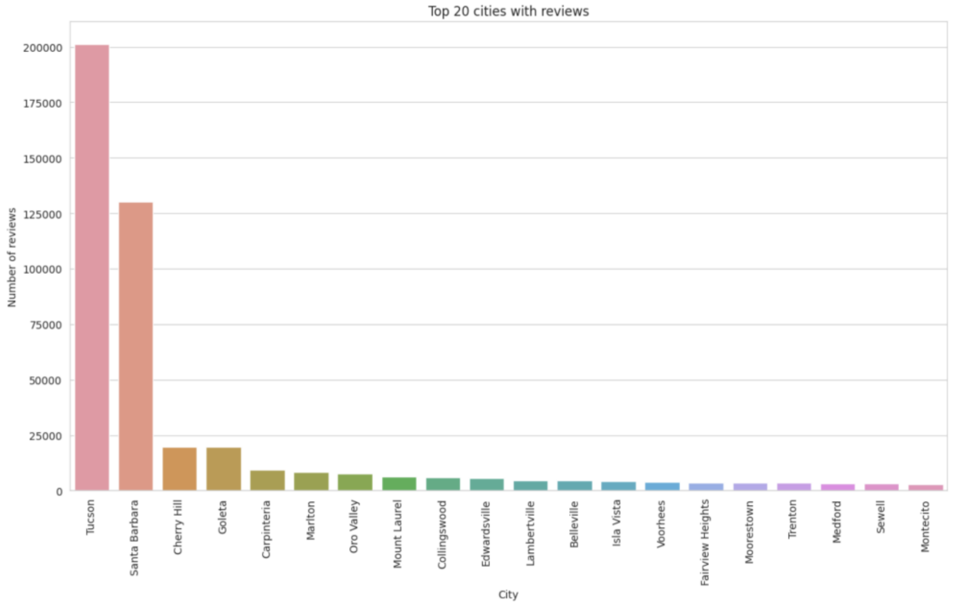
	item a	item b	item c
user 1	1	2	nan
user 2	3	nan	2
user 3	nan	nan	5

Utility Matrix



EXPLORATORY DATA ANALYSIS

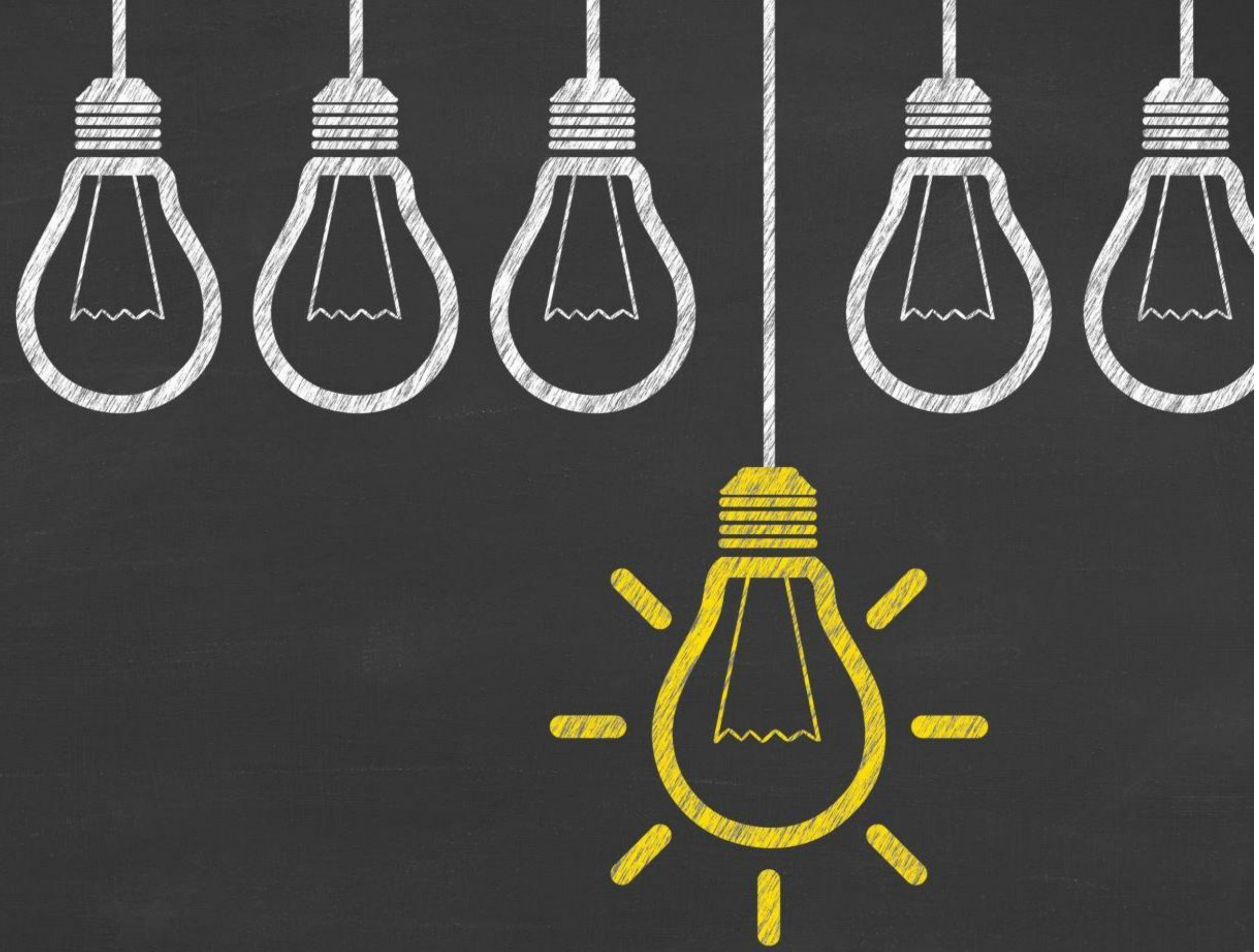


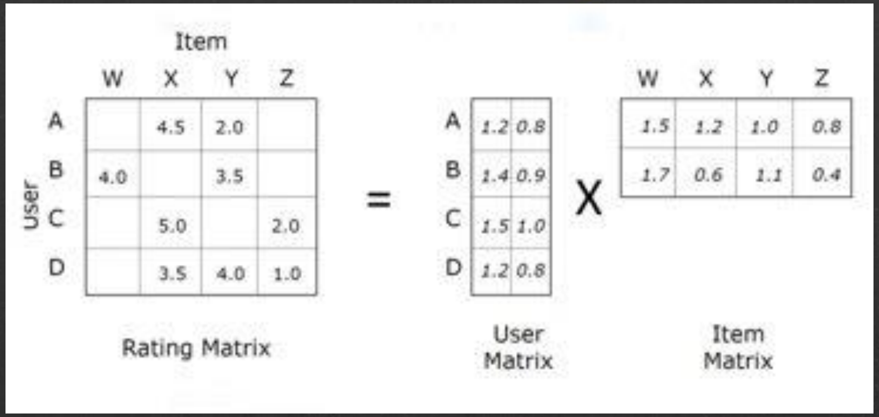


EVALUATION CRITERIA - RMSE

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

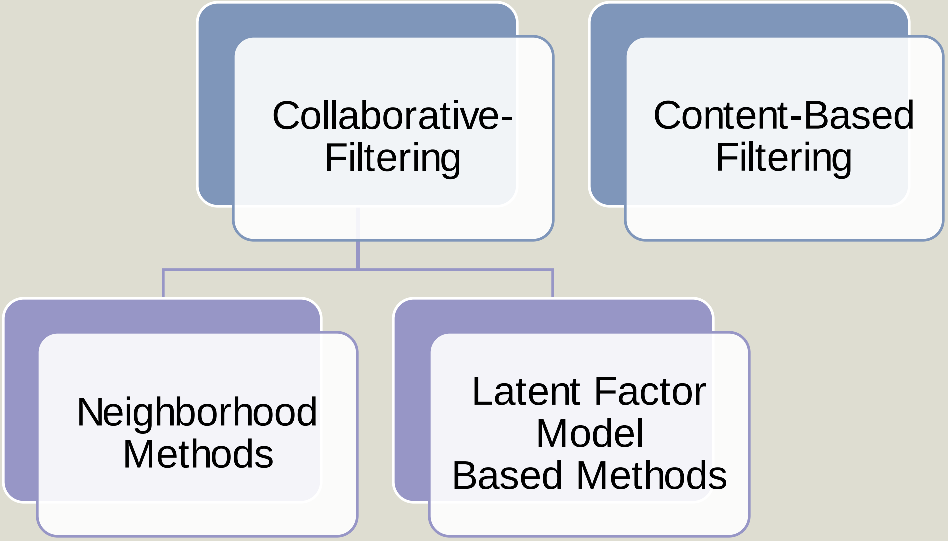
- Train test split: For each user, randomly add 30% ratings to test set
- Root mean square error (RMSE) between predicted ratings and actual ratings



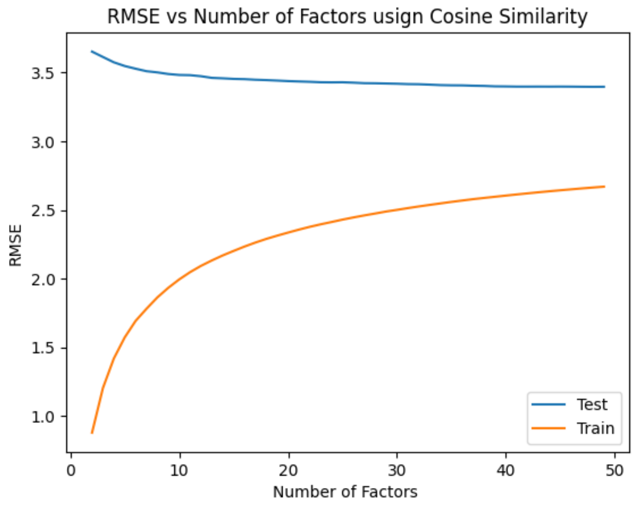


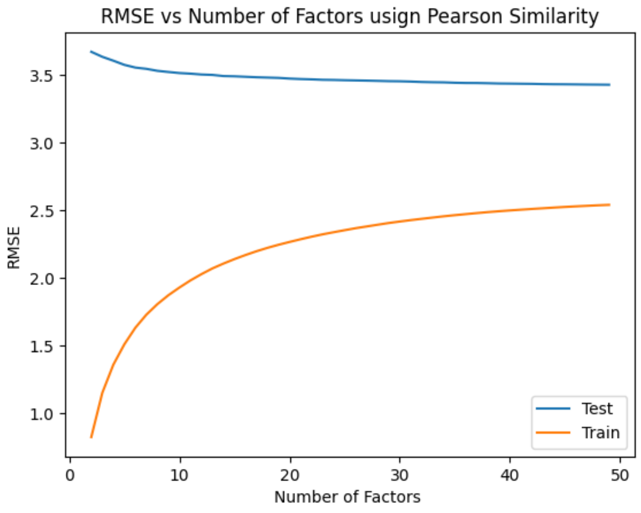
Item
W X Y Z
A 4.5 2.0
B 4.0 3.5
C 5.0 2.0
D 3.5 4.0 1.0
User
Rating Matrix
=
A 1.2 0.8
B 1.4 0.9
C 1.5 1.0
D 1.2 0.8
User Matrix
X
W X Y Z
1.5 1.2 1.0 0.8
1.7 0.6 1.1 0.4
Item Matrix

TECHNIQUES



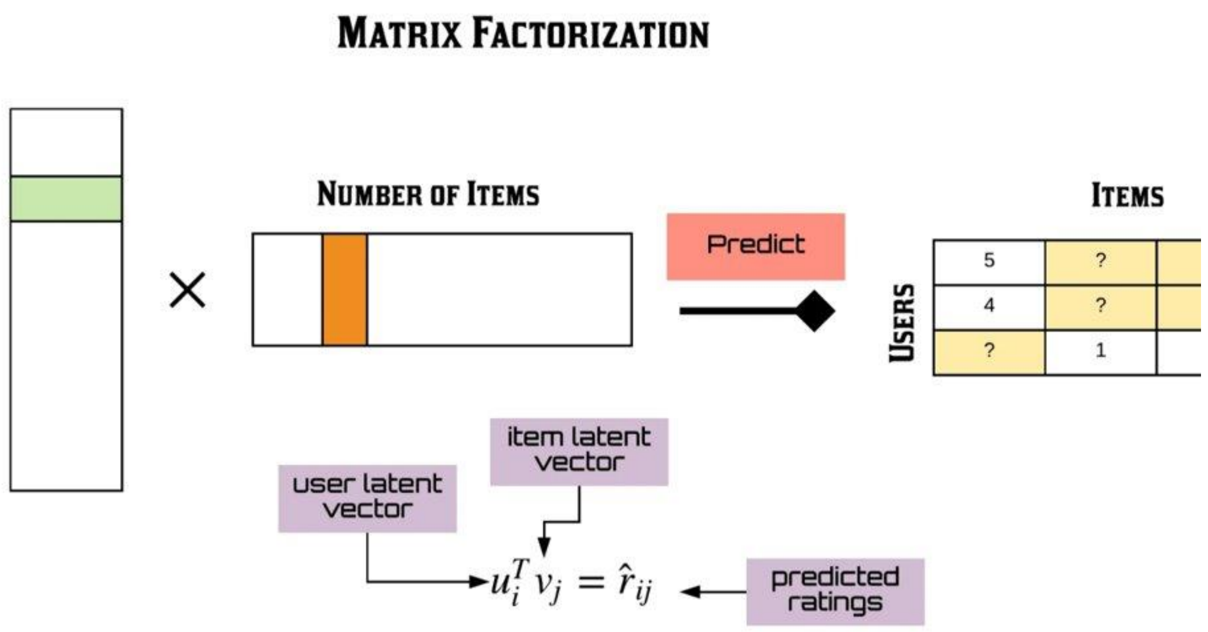
Collaborative-Filtering
Content-Based Filtering
Neighborhood Methods
Latent Factor Model Based Methods





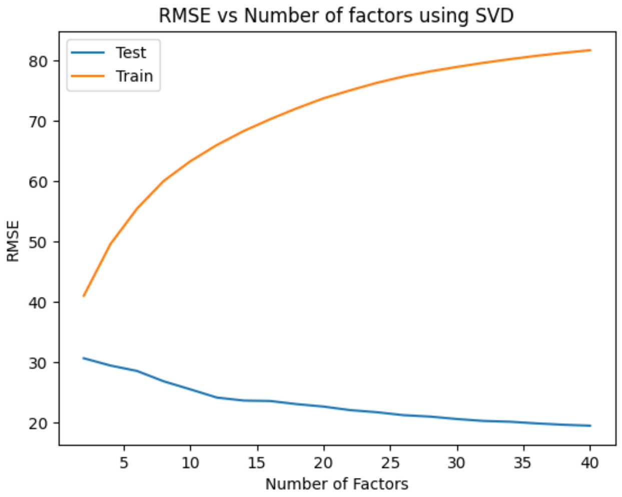
USER SIMILARITY COLLABORATIVE FILTERING

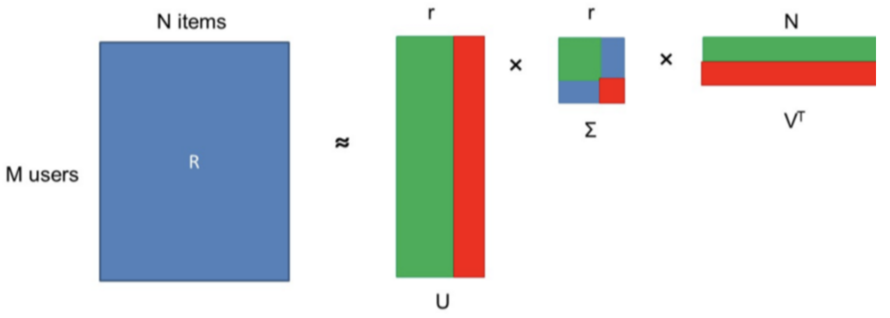
ALGORITHMS FOR LATENT FACTOR MODEL BASED COLLABORATIVE FILTERING



- Singular Value Decomposition (SVD)
- Alternating Least Squares (ALS)
- Stochastic Gradient Descent (SGD)
- Random Forest

SINGULAR VALUE DECOMPOSITION (SVD)





Popular matrix factorization technique

SVD learns low-dimensional representations of users and items that capture their preferences and characteristics

Latent Concept/Factors: Cuisine, price range, atmosphere etc

U: User-concept association

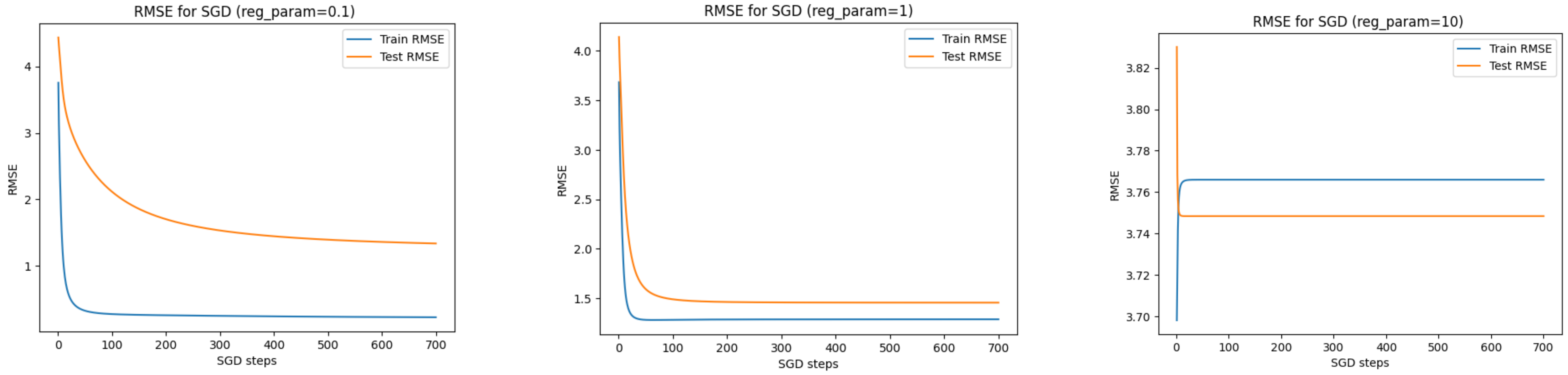
S: Singular value matrix, explains importance of each concept

V: Item-concept association, how much does item have a particular concept

Struggles with sparsity

STOCHASTIC GRADIENT DESCENT (SGD)

- Optimization algorithm for learning latent factors in system
- **Objective Function:** Difference between Predicted **ratings** and **actual ratings**
- Goal: Minimize **objective function**
- Iterative updates to parameters based on gradients
- Result: Learned user and item latent factors


RMSE for SGD (reg_param=0.1)
Train RMSE
Test RMSE
RMSE
SGD steps
RMSE for SGD (reg_param=1)
Train RMSE
Test RMSE
RMSE
SGD steps
RMSE for SGD (reg_param=10)
Train RMSE
Test RMSE
RMSE
SGD steps

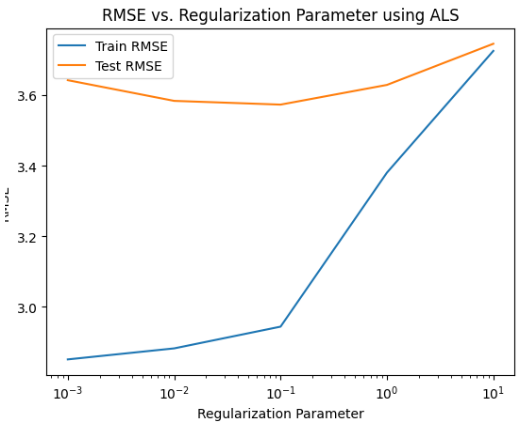
SGD

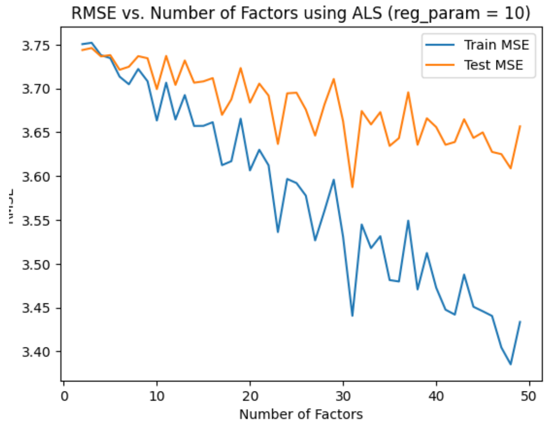
ALTERNATING LEAST SQUARE (ALS)

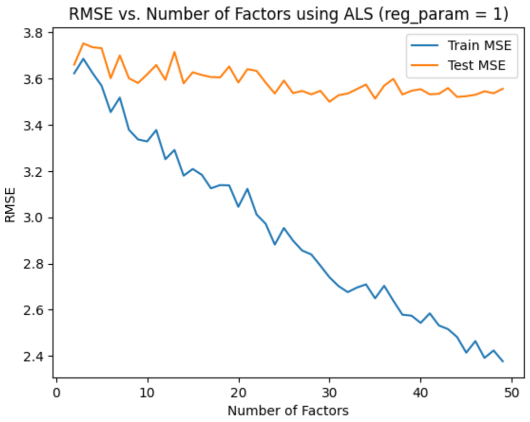
Method for performing matrix factorization by iteratively solving for the user and item factors

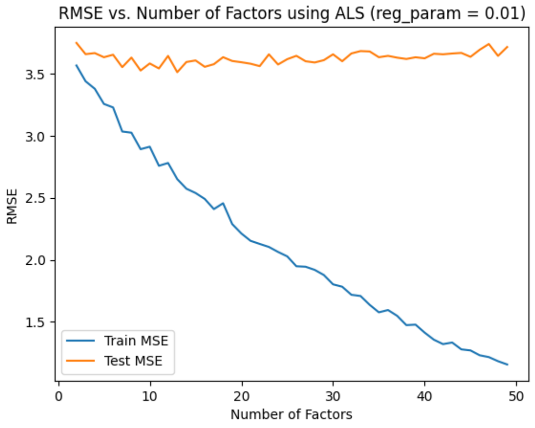
Learns k-dimensional feature vectors for each user and item, where k is the number of latent features

The function that predicts the value of a data instance given the feature vectors is a simple function of the dot product between the corresponding feature vectors









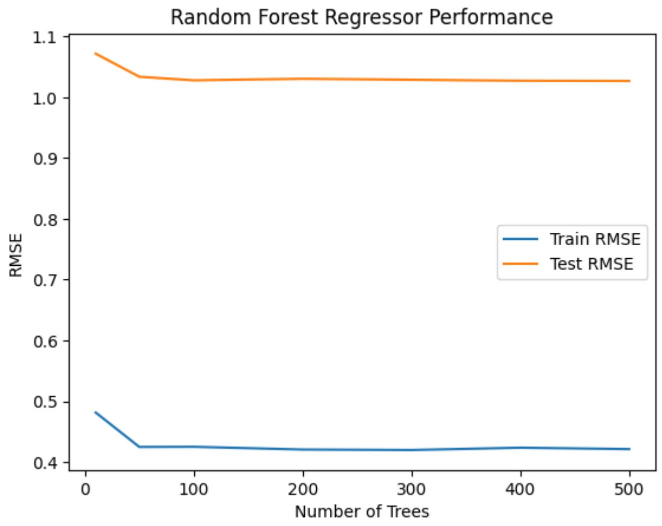
RANDOM FOREST

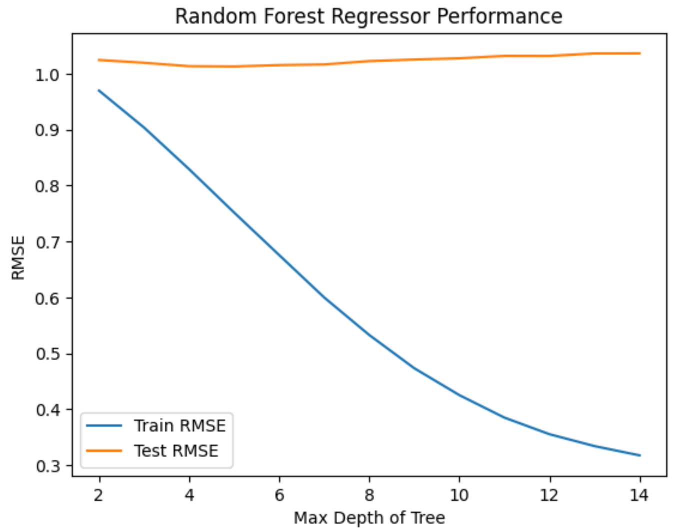
Latent feature vectors for users and items are learned using SGD for predicting item ratings

The learned feature vectors are concatenated for a specific user-item pair and used as an input variable

The target variable is the rating given by the user to the item, and the problem is transformed to a regression problem

The model predicts the rating a user would give to an item based on their past behavior and the behavior of other similar users





RESULTS

Type	Model	Train RMSE	Test RMSE
Latent Factor Model Based Methods	SVD (Singular Value Decomposition)	81.6	19.4
Latent Factor Model Based Methods	ALS (Alternating Least Square)	2.3	3.6
Latent Factor Model Based Methods	SGD (Stochastic Gradient Descent)	1.2	1.5
Latent Factor Model Based Methods	Random Forest Regressor	0.32	1.1
Neighborhood Method	Cosine Similarity	2.5	3.41
Neighborhood Method	Pearson Similarity	2.5	3.45

	user_id	name	stars	review_stars
6407	-BZn63YaADy9GpzHdncDtA	Reytas Filipino Cuisine	4.5	4
27232	-BZn63YaADy9GpzHdncDtA	Trappixx Jamaican Restaurant	3.0	4
50169	-BZn63YaADy9GpzHdncDtA	Jjang Ga Nae	4.0	5
82538	-BZn63YaADy9GpzHdncDtA	Thaimax	2.5	4
262353	-BZn63YaADy9GpzHdncDtA	Dolsot House	4.0	5
329974	-BZn63YaADy9GpzHdncDtA	Outback Steakhouse	2.5	3
350506	-BZn63YaADy9GpzHdncDtA	Phil's Deli And Market	4.0	3
380630	-BZn63YaADy9GpzHdncDtA	Pho Viet	4.5	2
411276	-BZn63YaADy9GpzHdncDtA	Hên Vietnamese Eatery	4.5	4
449253	-BZn63YaADy9GpzHdncDtA	Naked Lunch	4.5	4
497494	-BZn63YaADy9GpzHdncDtA	Kyuramen	4.5	4

User's Rated Places

```
array(["People's Pizza", 'Pho Eden', 'Eden Korean Restaurant',
       'Hanoi Cuisine', 'Zushi Dozo'], dtype=object)
```

Recommended Places

RESULTS - RECOMMENDATIONS

CONCLUSION AND FUTURE SCOPE

Ensemble Techniques

Content-based Recommendation

Addressing Cold Start problem

More Scalable System

REFERENCES

- Yelp, I. (2022, March 17). *Yelp dataset*. Kaggle. Retrieved April 6, 2023, from https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset
- Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal. Retrieved April 6, 2023, from https://www.sciencedirect.com/science/article/pii/S1110866515000341
- *(PDF) collaborative filtering recommender systems - researchgate*. (n.d.). Retrieved April 25, 2023, from https://www.researchgate.net/publication/200121027_Collaborative_Filtering_Recommender_Systems