
Restaurant
Recommendation

Background

- Yelp is a popular online platform providing user-generated reviews and ratings for various businesses, including restaurants

- Building a recommendation system using Yelp's dataset can help users make informed decisions while selecting restaurants

- The project aimed to explore machine learning techniques such as collaborative filtering, content-based filtering, and hybrid modeling to develop an efficient restaurant recommendation system

Motivation

- To Utilize the review data available in Yelp dataset to develop a more effective and efficient restaurant recommendation system.

- Review data can improve restaurant recommendations by extracting information about users' preferences and needs from reviews, which can lead to more personalized and relevant recommendations.

- Our approach aims to improve upon the limitations of existing recommendation systems, such as data sparsity and cold start problems, by leveraging the review data to generate more accurate and diverse recommendations.

Literature Review

Technique Used	Dataset Used	Approach	Pros	Cons
Content-based Similarity calculation Weighted based calculations (rating & price)	Data collected from: Google maps, Facebook, Food panda. Initial 3000 restaurant data. After cleaning 1600 restaurant data used.	Content-based filtering model to recommend top similar restaurants. Weighted score = (avg_rating x 0.90 + avg_price x 0.10)	Personalized recommendation based on user's preference. Avoids cold-start problems	Limited quantity & quality of available data may affect the accuracy and scope of recommendation.
Rating LDA model for collaborative filtering	MovieLens 100K dataset	User Based Collaborative Filtering	Incorporates semantic information through topic modeling using LDA	Lack comprehensive comparison of the hybrid approach with other state-of-the-art recommendation algorithms.
(FLAME) Aspect-Based Opinion Mining Collaborative filtering	Yelp, Tripadvisor	Recommendation Aspect Based Opinion Mining, Consider rating as well as reviews on items	The FLAME model combines two powerful techniques, aspect-based opinion mining and collaborative filtering, to enhance recommendation accuracy.	FLAME's probabilistic modeling approach is computationally intensive, particularly when handling large-scale datasets.
Memory Based Collaborative Filtering	MovieLens Dataset	Uses exponential functions for determining the optimal number of common items	Reduction in sparsity problems	Adoption of exponential similarity in practical recommendation systems is limited and not preferred.

Dataset Description

yelp Dataset

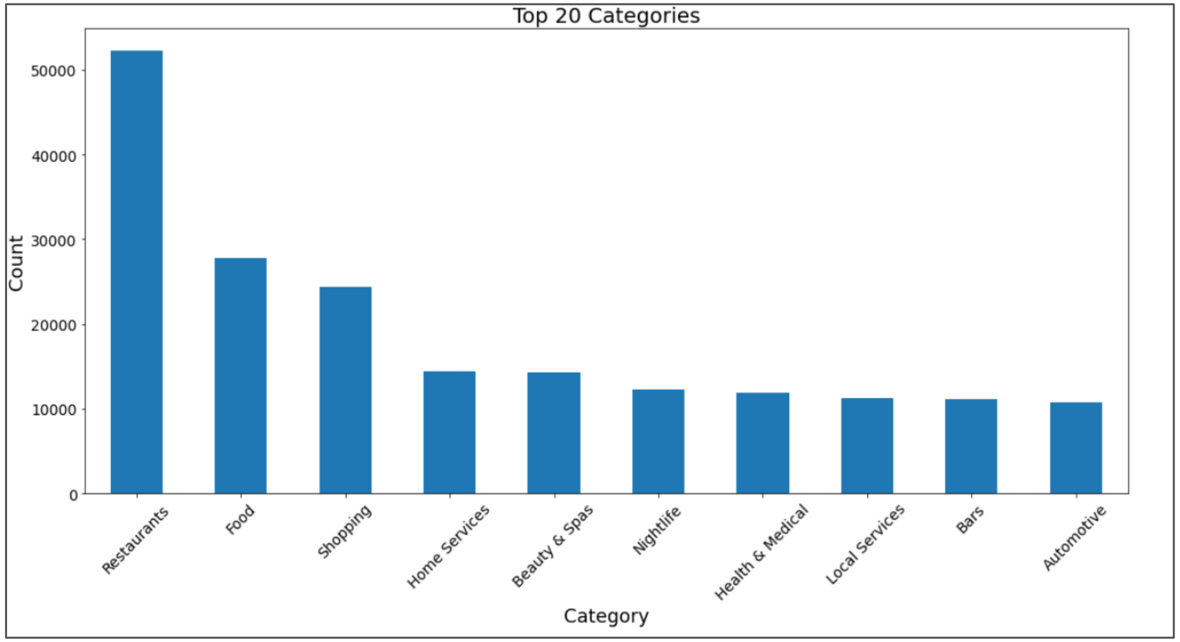
- Acquired data from Yelp Website. It consists of a subset of businesses, reviews, and user data
- Business.json
    - Consists of information about 150,346 businesses
    - Includes attributes like location, ratings, review count, price range, ambience, type of cuisine, etc.

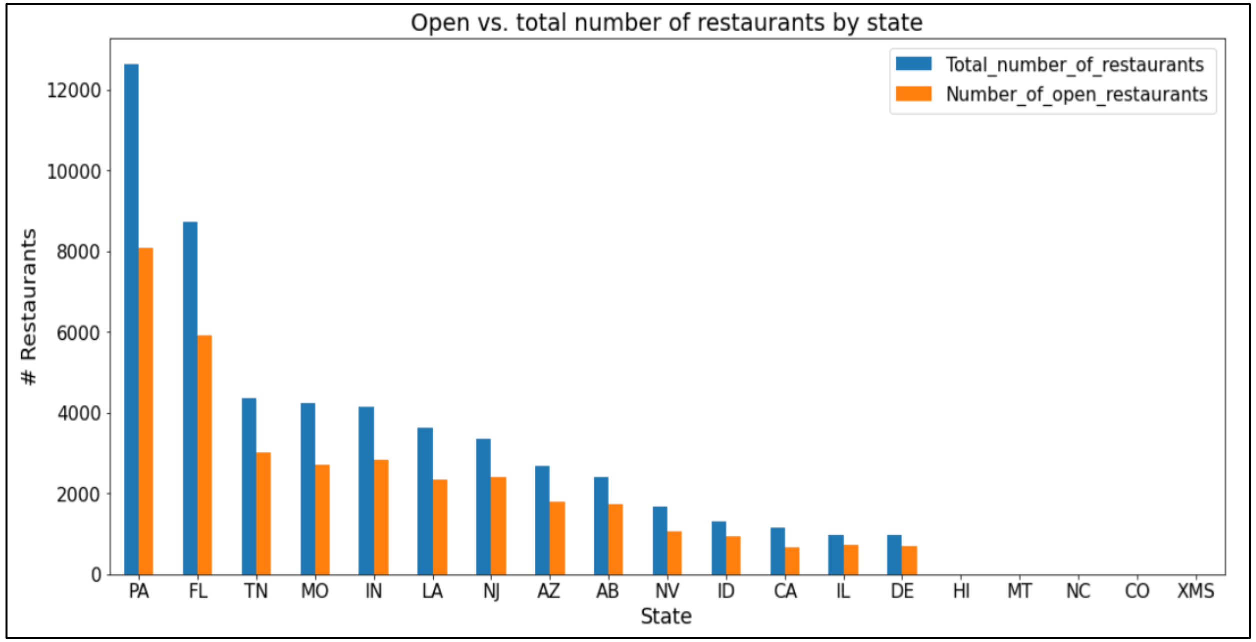
150,346 businesses

- Reviews.json
    - Consists of 6,990,280 reviews
    - Includes details about the user, ratings, business id, user-generated reviews, etc.

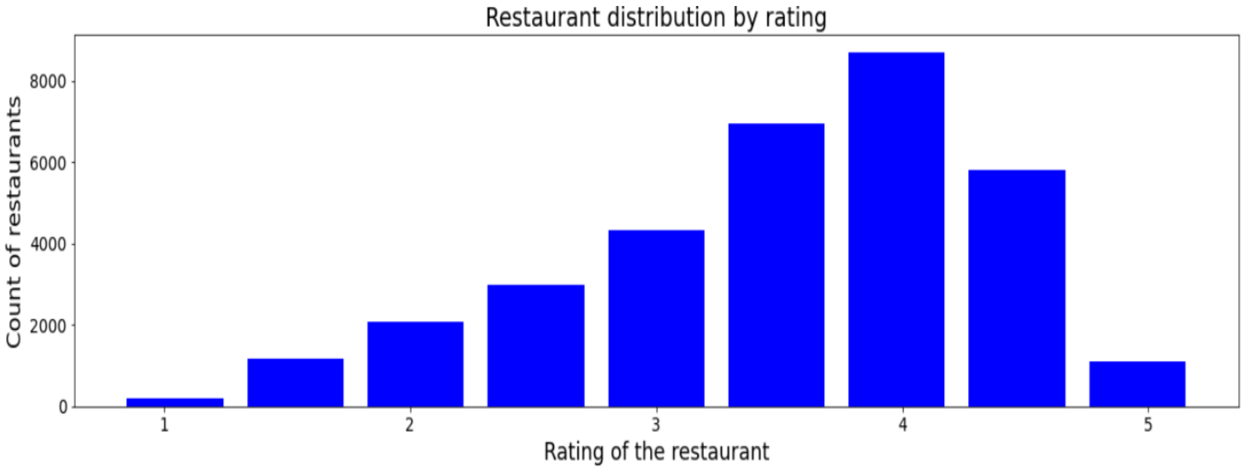
6,990,280 reviews

Exploratory Data Analysis – Business dataset

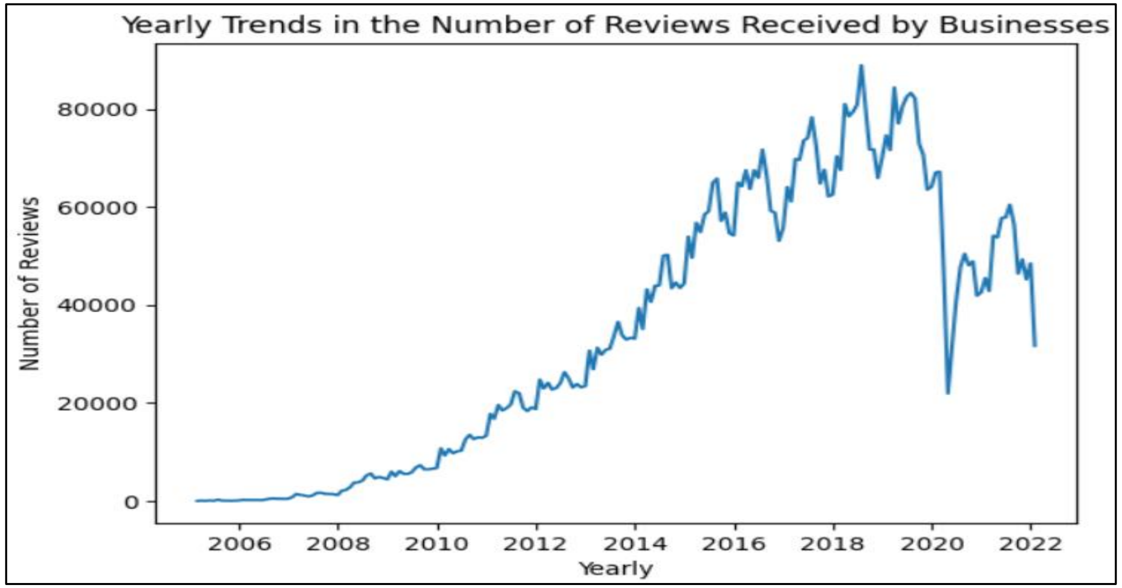


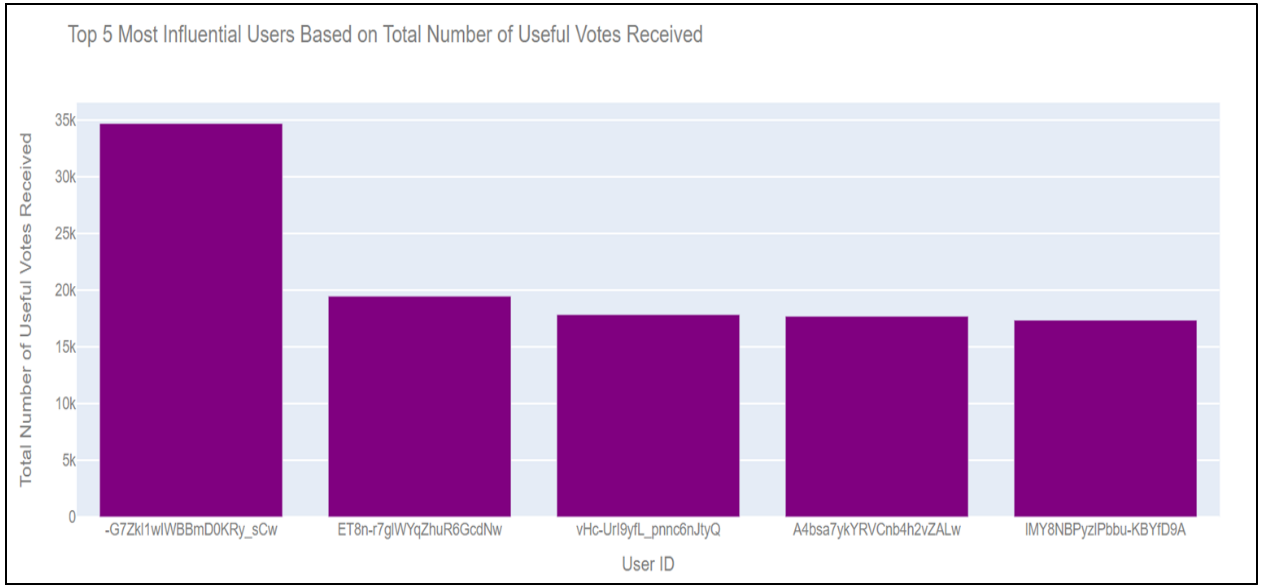


Restaurants in USA



Exploratory Data Analysis – Reviews Dataset

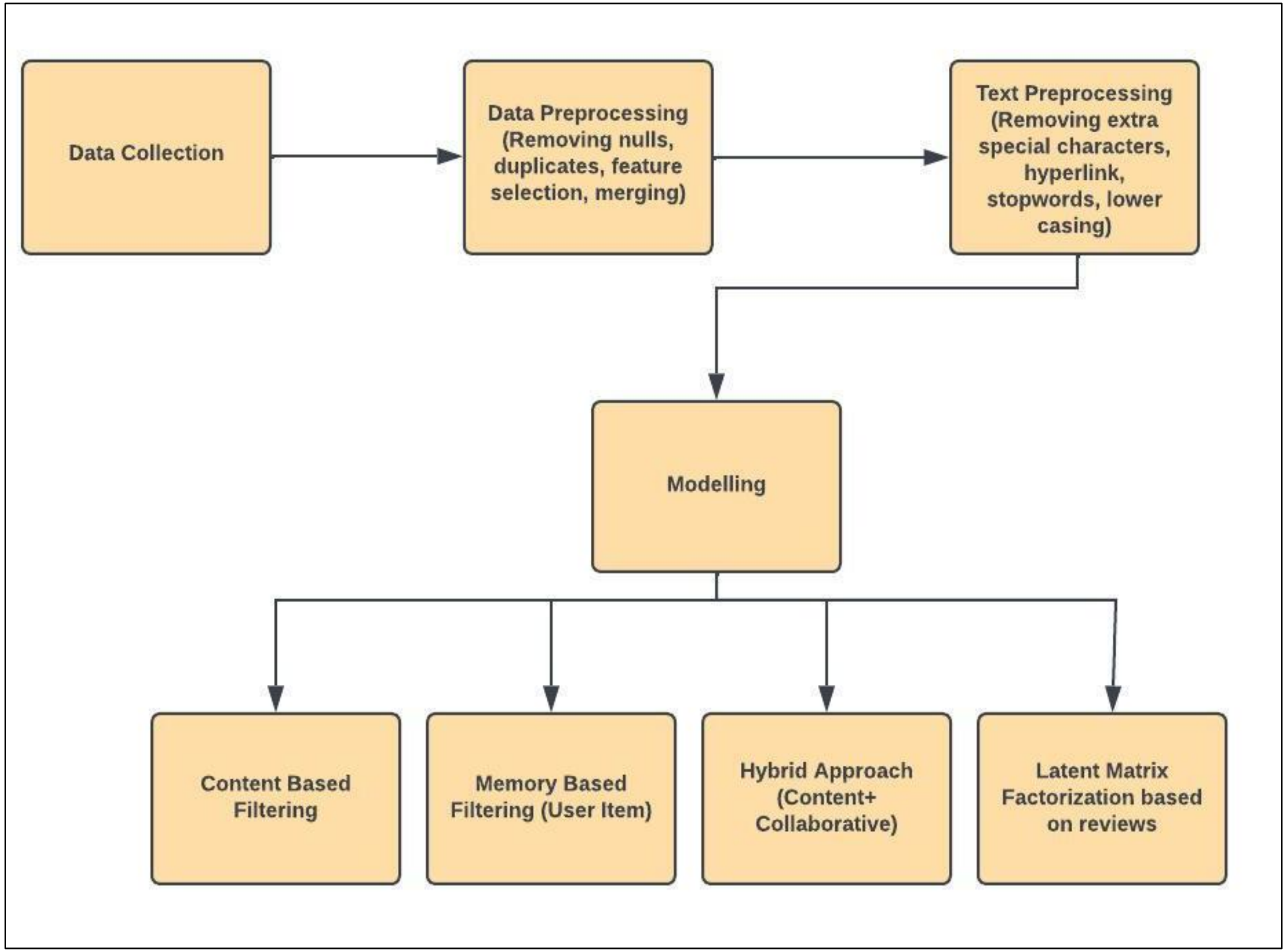




Chinese Cuisine

Italian Cuisine

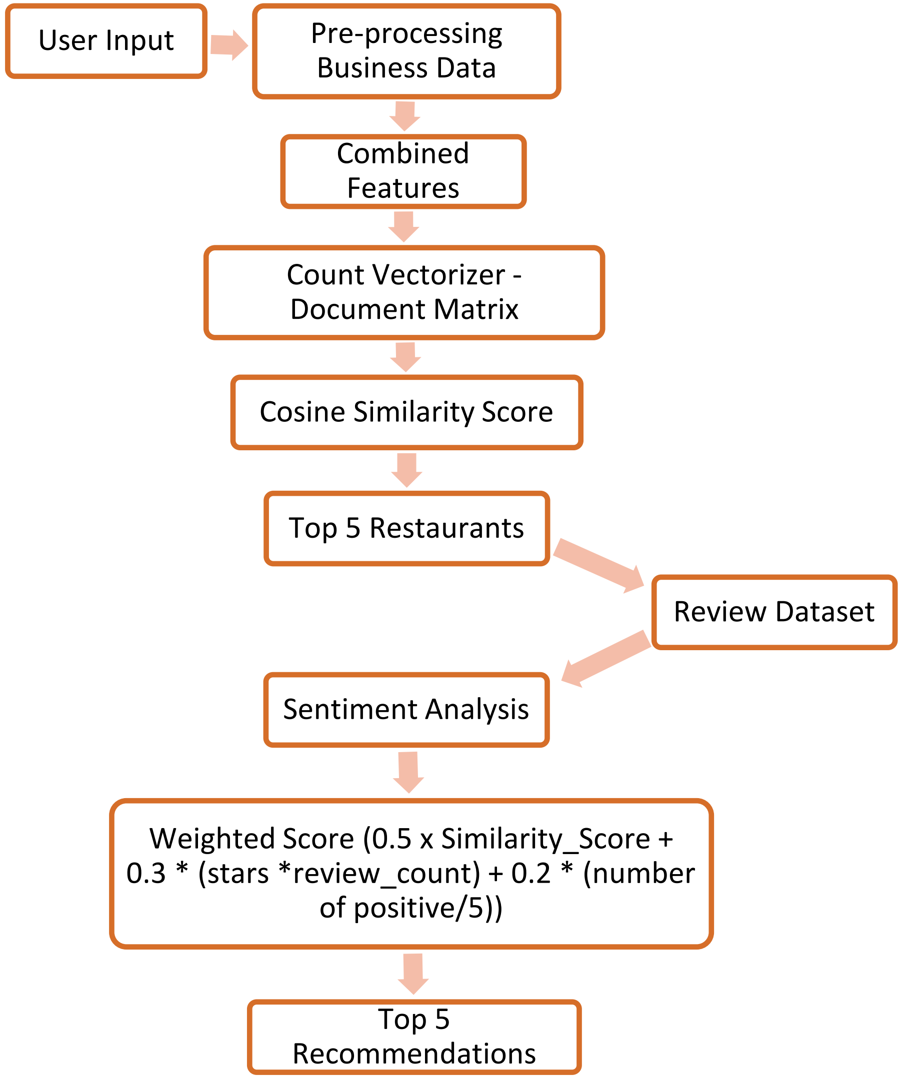
Methodology



Data Pre-processing

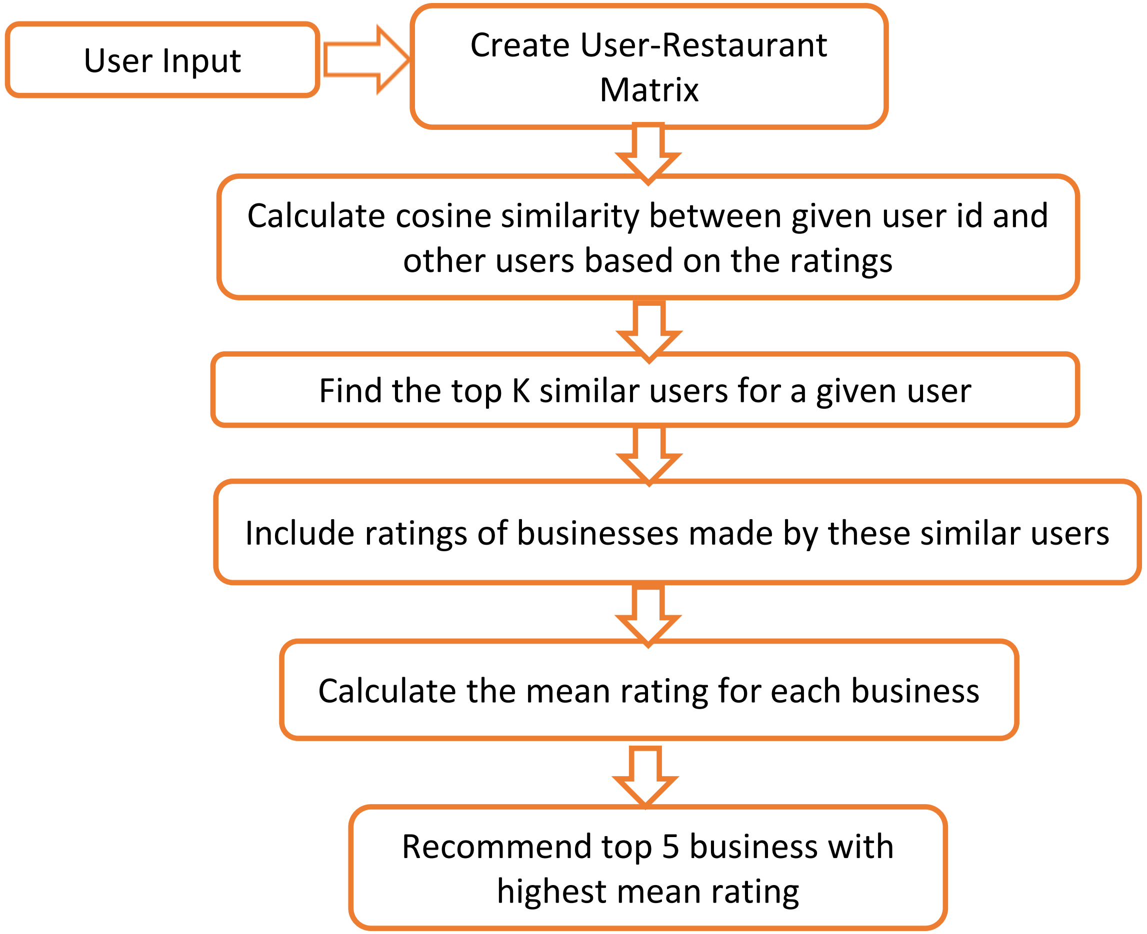
Drop the instances with null categories	(150243, 14)
Filtered the data to keep only the Restaurant businesses	(52286, 14)
Dropped the closed Restaurants	(35004, 14)
Retained the businesses present in the USA	(33267, 14)
Filtered the businesses for the state of Pennsylvania as it contains majority count of restaurants	(8072, 14)
Splitted the category attribute into separate columns to extract the cuisine type information	(8072,54)
Save as csv file	

Content-based Filtering




```
(YelpContentModeling) shilpashivarudraiah@Shilpas-MacBook-Air-2 YelpContentBasedModeling % python
run.py
/Users/shilpashivarudraiah/YelpContentBasedModeling/run.py:17: SyntaxWarning: "is not" with a lite
ral. Did you mean "!="?
  if _ is not -1:
[nltk_data] Downloading package stopwords to
[nltk_data]     /Users/shilpashivarudraiah/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
loading Review Dataframe
Review Dataframe Loaded
Enter a business name to get recommendations or press 'q' to exit: Starbucks
100%|██████████████████████████████████████████████████| 25/25 [00:49<00:00,  1.97s/it]
{'ReAnimator Coffee': 126.62, 'Baltic Bakery': 16.82, 'Fine Wine & Good Spirits': 9.06, 'Snacks By
 The Pound': 8.98, 'Liberty City Roasters': 11.499999999999998}
['ReAnimator Coffee', 'Baltic Bakery', 'Liberty City Roasters', 'Fine Wine & Good Spirits', 'Snack
s By The Pound']
[0, 1, 4, 2, 3]
Recommended Businesses are

Business Name:  ReAnimator Coffee
Stars:  4.0
Review Count:  105
Combined Features:  Food
Weightage Scores:  126.62
Business Name:  Baltic Bakery
Stars:  4.5
Review Count:  12
Combined Features:  Food
Weightage Scores:  16.82
Business Name:  Liberty City Roasters
Stars:  4.5
Review Count:  8
Combined Features:  Food
Weightage Scores:  11.499999999999998
Business Name:  Fine Wine & Good Spirits
Stars:  3.5
Review Count:  8
Combined Features:  Food
Weightage Scores:  9.06
Business Name:  Snacks By The Pound
Stars:  3.5
Review Count:  8
Combined Features:  Food
Weightage Scores:  8.98
Please enter the number of recommendations which you find relevant: 3
Precision is:  0.6
Enter a business name to get recommendations or press 'q' to exit: q
Mean Average Precision is:  0.6
```

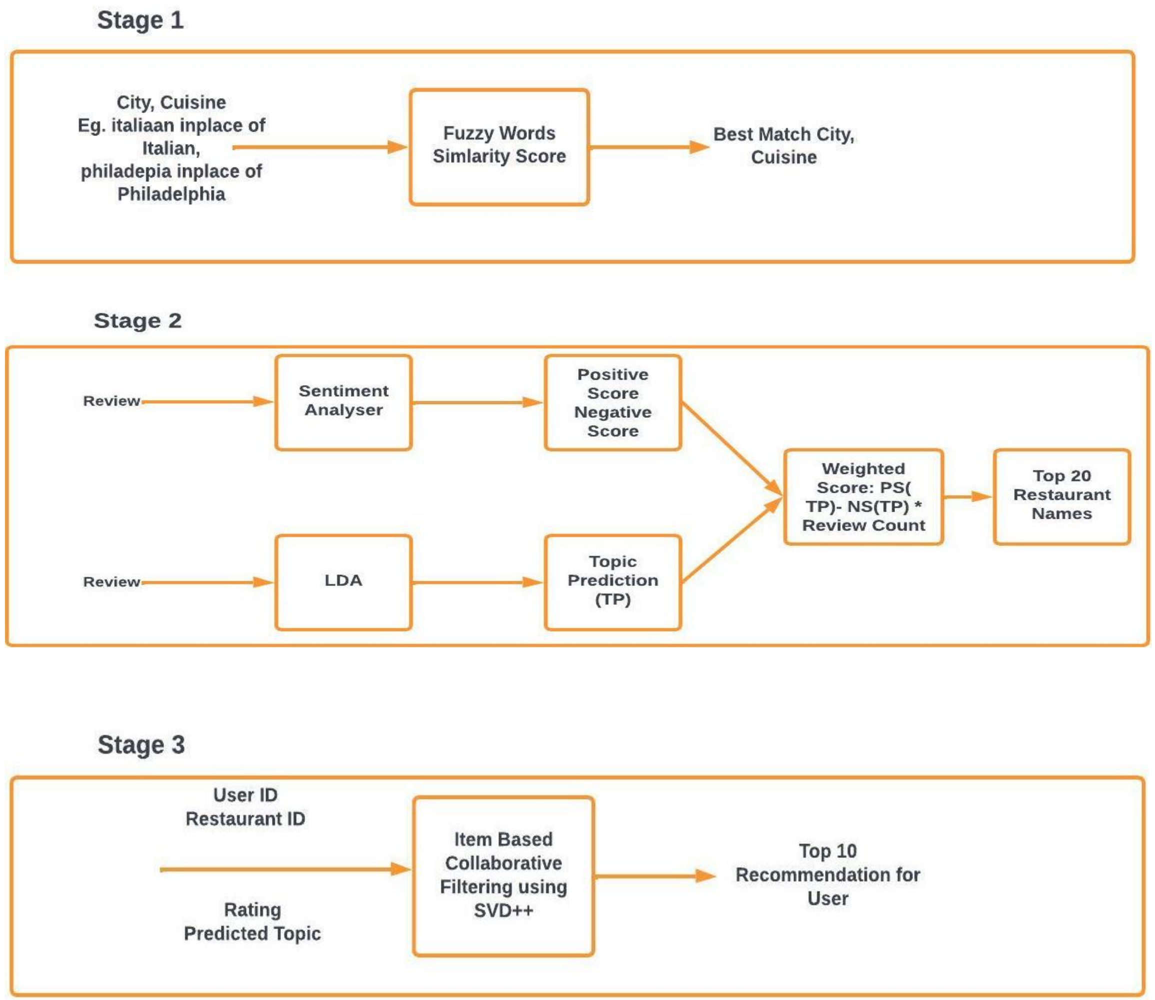
Memory-based Collaborative Filtering




```
Recommended businesses for user _7bHUi9Uuf5__HHc_Q8guQ:
- The Sweet Life Bakeshop
- John's Roast Pork
- Terakawa Ramen
- Zahav
- Barbuzzo
- ICI Macarons & Cafe
- Morimoto
- Talula's Garden
- Reading Terminal Market
- Cafe La Maude
```

Hybrid Model

Model Architecture



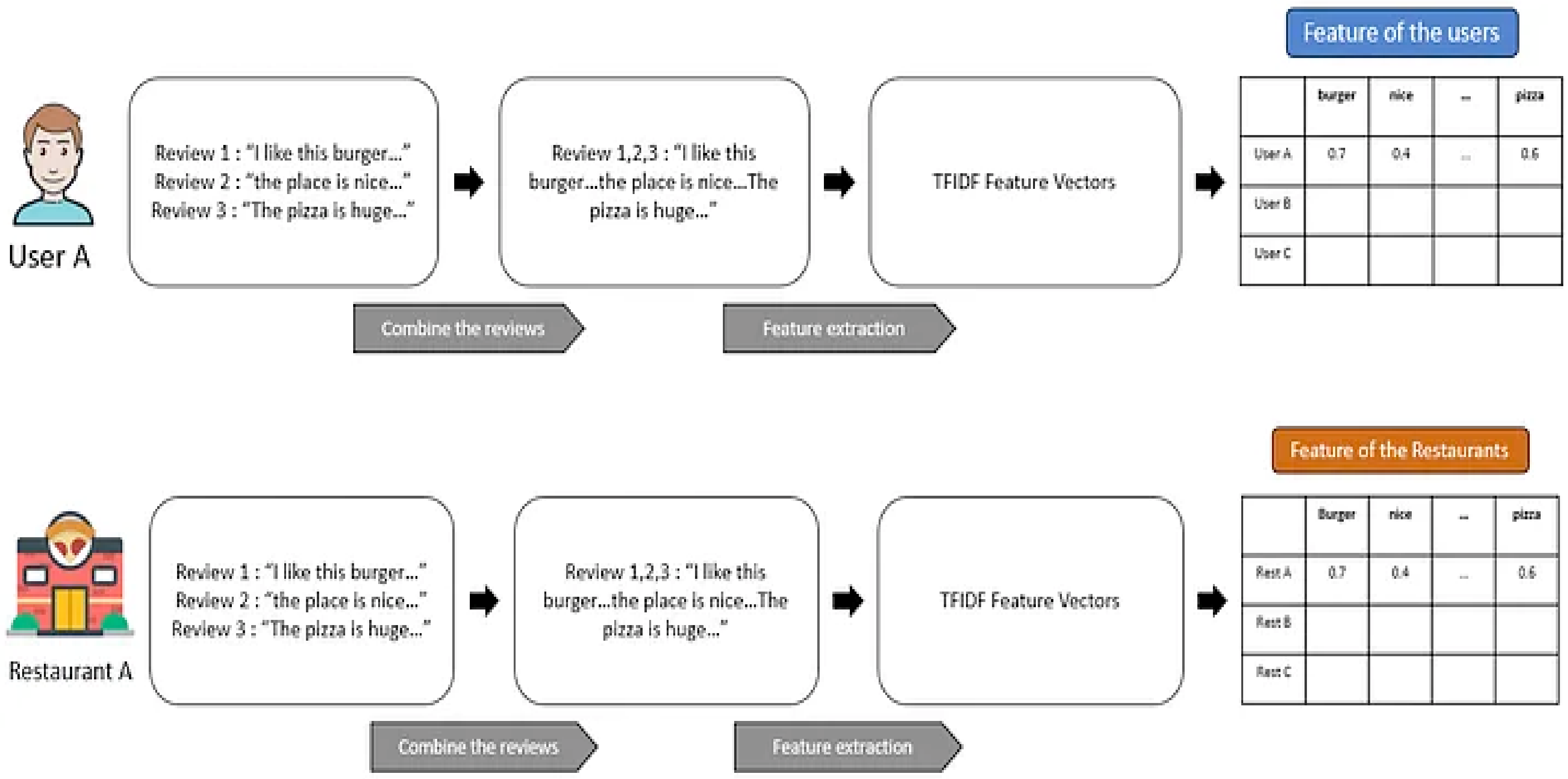

```
[(0,
  '0.033*"pizza" + 0.015*"good" + 0.013*"place" + 0.010*"great" + 0.008*"like" + 0.007*"time" + 0.007*"service" +
0.006*"cheese" + 0.006*"ordered" + 0.006*"delicious"'),
 (1,
  '0.011*"good" + 0.010*"place" + 0.009*"great" + 0.008*"like" + 0.006*"time" + 0.006*"really" + 0.005*"service" +
0.005*"beer" + 0.005*"cheese" + 0.005*"dont"'),
 (2,
  '0.015*"good" + 0.013*"place" + 0.010*"chicken" + 0.008*"great" + 0.008*"service" + 0.007*"ordered" + 0.007*"like" +
0.007*"time" + 0.006*"order" + 0.005*"got"'),
```

Query

```
python trainedmodel.py chiese pihiladelpdia
```

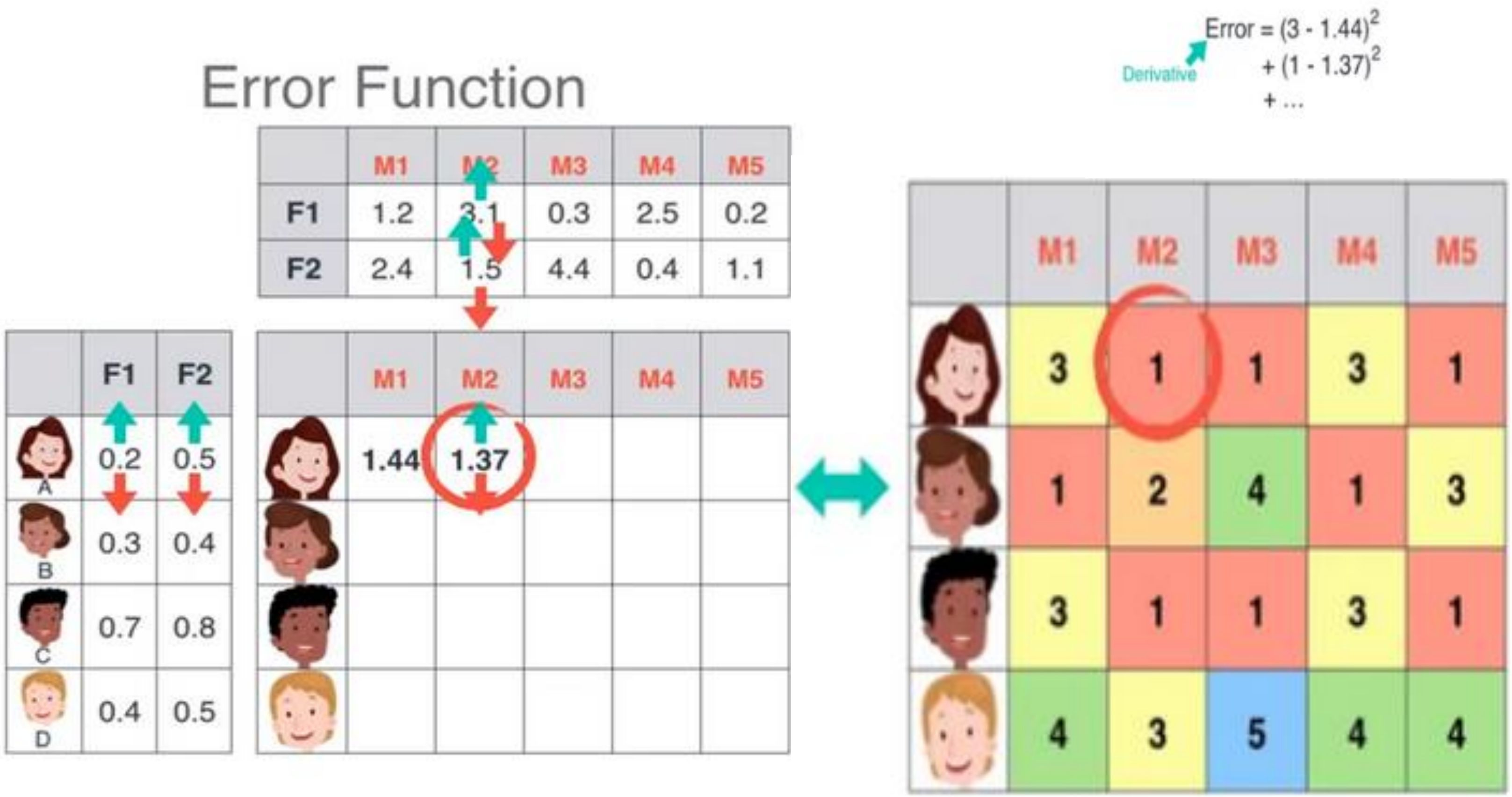
```
1.  Unit Su Vege Location:  Philadelphia Rating:  4.534926292245101

2.  Dump-N-Roll Location:  Philadelphia Rating:  4.482285375049951

3.  Chubby Cattle Location:  Philadelphia Rating:  4.472334279154627

4.  Wm Mulherin's Sons Location:  Philadelphia Rating:  4.386406652491579

5.  Saloon Restaurant Location:  Philadelphia Rating:  4.356117278482909
```

Latent Factor Collaborative filtering with regularization



'reasonable': 1407,
'sister': 1576,
'radishes': 1386,
'pepper': 1251,
'veggies': 1878,
'break': 213,
'sit': 1577,
'shouldnt': 1561,
'planning': 1293,
'given': 724,
'sizes': 1582,
'helped': 810,
'consistency': 391,
'sour': 1621,
'living': 980,
'compared': 377,
'liked': 965,
'seared': 1512,
'platter': 1297,
'particularly': 1231,
'shops': 1558,
'quesadilla': 1377,
'personal': 1259,
'hidden': 815,

Matrix Factorization and Calculate the Error to Optimize Prediction



Error:

$$\min_{P,Q} \sum_{(i,x)\in R} \left(r_{xi} - q_i \cdot p_x^T\right)^2$$

Regularization:

$$\min_{P,Q} \underbrace{\sum_{training} (r_{xi} - q_i p_x^T)^2}_{\text{"error"}} + \lambda \underbrace{\left[\sum_x \|p_x\|^2 + \sum_i \|q_i\|^2\right]}_{\text{"length"}}$$

Output

```
query = "i want to go for brunch place"
```

	Restaurant Name	Category	Rating	Review
0	Cafe La Maude	Sandwiches	4.5	1485
1	Brunch Everyday	Restaurants	4.5	198
2	Green Eggs Café	Restaurants	4.0	2679
3	Five Guys	Food	3.5	84
4	Wm Mulherin's Sons	Restaurants	4.5	610
5	Pizza Shackamaxon	Restaurants	4.0	145
6	Añejo Philadelphia	Restaurants	4.0	132

```
query = "i want to go for vegeterian place"
```

	Restaurant Name	Category	Rating	Review
0	Five Guys	Food	3.5	84
1	Pizza Shackamaxon	Restaurants	4.0	145
2	Cafe La Maude	Sandwiches	4.5	1485
3	Angelo's Pizzeria	Restaurants	4.5	393
4	Stockyard Sandwich	Restaurants	4.5	284
5	Ristorante Pesto	Restaurants	4.5	578
6	Wm Mulherin's Sons	Restaurants	4.5	610

Evaluation Metric

To evaluate the accuracy of the recommendation systems on test data:

- Predict the ratings that a user would give to a business they have not yet predicted
- Calculate RMSE by taking the difference between the actual and predicted ratings

The lower the RMSE the better !!

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

Method	RMSE
User based Memory Collaborative Filtering	1.06
Hybrid Model (LDA ,SVD++)	0.74
Latent Matrix Factorization	0.69

Conclusion

Implemented four models using different approaches including **content based, memory-based, model (matrix factorization), and LDA,** expecting different inputs and give restaurant recommendations

The utilization of the different model allowed for a thorough exploration of all dimensions of the data, leading to the attainment of satisfactory results, thereby concluding the project successfully

This project offers users the accessibility to utilize a recommendation system with a wide range of inputs, enabling them to utilize the application in various aspects

Future Improvements

Enhance recommendation systems by integrating content-based filtering, query-based techniques on reviews, and hybrid approaches for more diverse applications

Implement mechanisms to capture real-time user feedback and adapt the recommendation model accordingly

Experiment with more advanced collaborative filtering techniques, such as deep learning-based models (e.g., Neural Collaborative Filtering), to capture complex user-item interactions

Explore techniques like distributed computing, parallelization, or model compression to ensure the system can handle growing user bases and handle real-time recommendations efficiently

Team Contribution

Team Member	Contributions
Data Collection & Pre-processing	Iqra, Priya
Exploratory Data Analysis	Saniya, Shilpa
Data Transformation + Modelling - Content based	Shilpa
Data Transformation + Modeling- Memory based	Saniya
Data Transformation + Modelling- Knowledge Based + Content Filtering + Collaborative Filtering	Iqra
Data Transformation + Modeling- Collaborative filtering with latent vector and regularization	Priya

References

- Ahmed, T., Akhter, L., Talukder, F. R., Hasan-Al-Monsur, N., Rahman, H., & Sattar, A. (2021). Restaurant Recommendation System in Dhaka City using Machine Learning Approach. In *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*. https://doi.org/10.1109/smart52563.2021.9676197
- Wu, Y., & Ester, M. (2015, February). Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering. In Proceedings of the eighth ACM international conference on web search and data mining (pp. 199-208)
- https://www.youtube.com/watch?v=ZspR5PZemcs
- https://towardsdatascience.com/how-to-build-a-restaurant-recommendation-system-using-latent-factor-collaborative-filtering-ffe08dd57dca

- Chaohui Liu, Xianjin Kong, Xiang Li, Tongxin Zhang, "Collaborative Filtering Recommendation Algorithm Based on User Attributes and Item Score", *Scientific Programming*, vol. 2022, Article ID 4544152, 7 pages, 2022. https://doi.org/10.1155/2022/4544152

- Zhou, X., & Wu, S. (2016). Rating LDA model for collaborative filtering. *Knowledge Based Systems*, *110*, 135–143. https://doi.org/10.1016/j.knosys.2016.07.020

Thank you