

RECOMMENDER SYSTEMS

김재훈 김은하 김정현 박기태 안수빈 이혜연

1

Summary of
recommender systems

2

Finalizing model

3

Model Evaluation

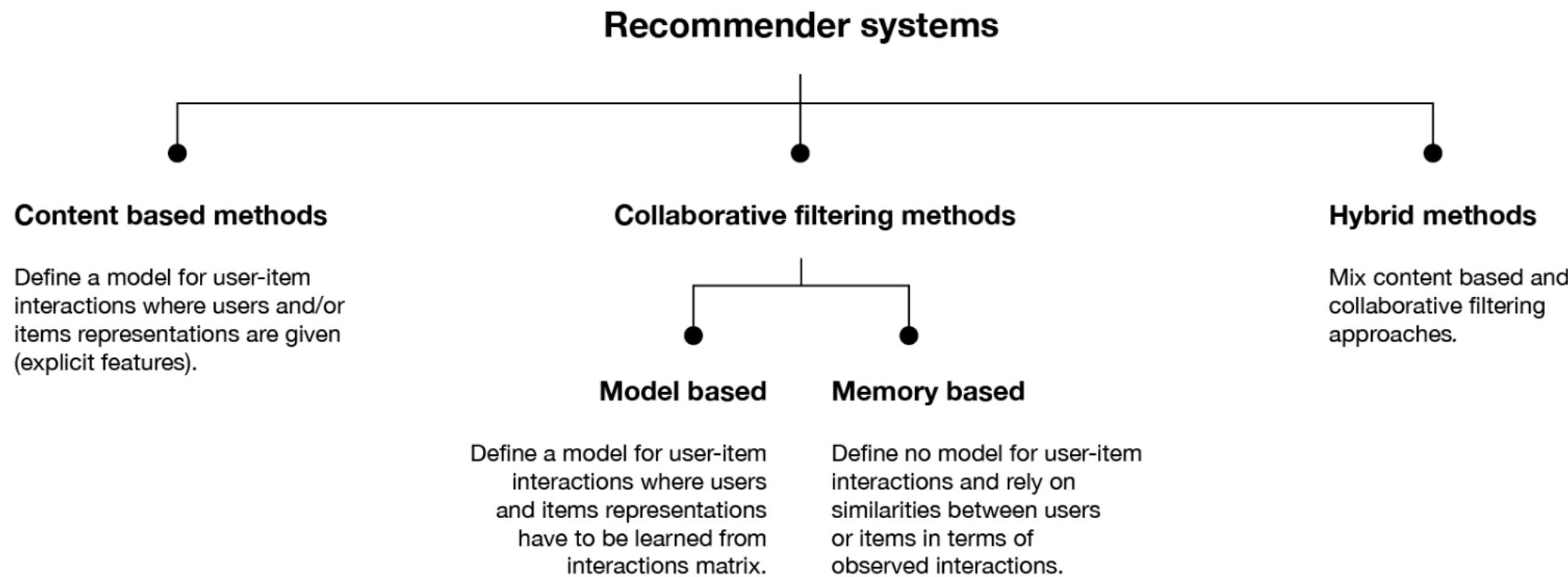
4

Recommendation
for you

PART(1)

*SUMMARY of
RECOMMENDER SYSTEMS*

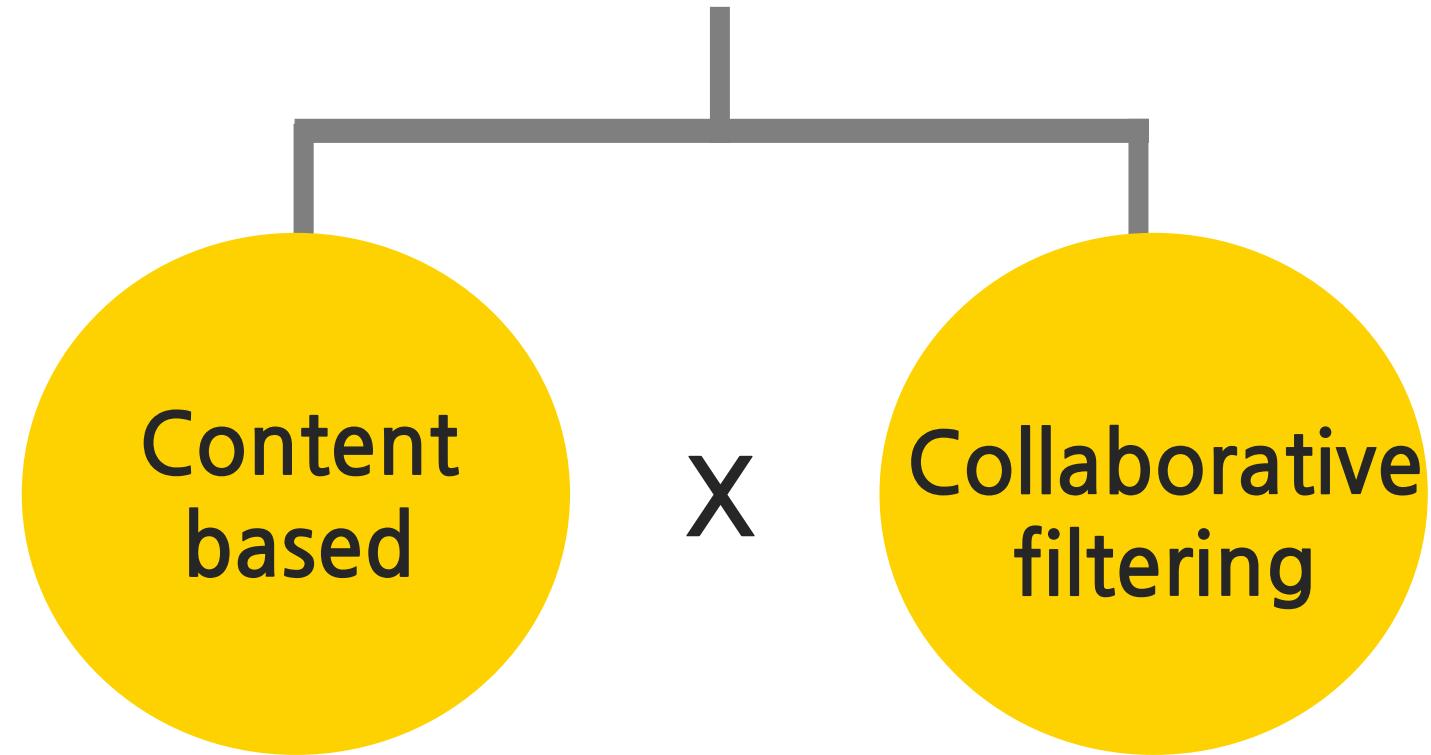
Part(1) Summary of Recommender systems



PART(2)

FINALIZING MODEL

“ Hybrid Method ”



Contents-based 기법

특정 식당과 유사한 식당들을 추출하기 위해 3가지 방식 사용
특정 유저의 평점 히스토리를 참고하여, 높은 평점의 식당들과 유사한 식당 추천

'Rake' 'TF-IDF' '단순 count-vectorization'

위 3가지 기법들의 결과에 크게 차이점이 존재하지 않음
식당 정보 데이터가 단순했던 것이 이유였을 것이라고 추정
⇒소포게시판에 한해서 세가지 기법 중 아무거나 활용해도 될 것 같음

Collaborative filtering 기법

추천 시스템 트렌드에서 가장 중요한 기법

Matrix Factorization 모델을 사용 결정 !!

다양한 CF 기법들을 학습하였으나 데이터 feature의 한계
평점 데이터만을 활용한 단순한 Matrix Factorization 모델 사용
딥러닝에 대한 이해가 부족하여 기존 딥러닝 모델에서의 발전은 없었음

Matrix factorization

두 개의 latent factor matrix, bias를 학습하여
최종 prediction rating matrix 예측

$m = \text{number of users}$, $n = \text{number of items}$
choose d , the number of features

$$\begin{matrix} & \begin{matrix} ? & ? \\ ? & ? \\ ? & ? \\ ? & ? \end{matrix} \\ \begin{matrix} ? & ? \\ ? & ? \\ ? & ? \\ ? & ? \end{matrix} & \times \begin{matrix} ? & ? & \text{Feature 1} \\ ? & ? & \text{Feature 2} \end{matrix} \\ m \times d & d \times n \end{matrix} = \begin{matrix} & & 1 \\ & 5 & \\ 2 & & \\ & 3 & \\ 2 & & \\ & 1 & \end{matrix} \quad m \times n$$
$$\hat{r}_{ui} = q_i^T p_u$$

출처:<https://towardsdatascience.com/recsys-series-part-4-the-7-variants-of-matrix-factorization-for-collaborative-filtering-368754e4fab5>

Bias

- 평점 데이터의 variation로 인한 학습 성능 저하 방지
- Bias 학습을 통해 평점을 후하게/짜게 주는 유저의 특징을 해결
 - 3가지 bias들을 통해, rating variation을 보정
 1. global average bias: 모든 평점 값들의 평균 값
 2. user bias: 특정 유저에 대한 bias 벡터
 3. Item bias: 특정 아이템에 대한 bias 벡터
- * 최종 예측 평점 값 R_{ui}
= Global bias + User bias + item bias
+ 두 latent factor vector의 내적 결과 값(user-item interaction)

PART(3)

MODEL EVALUATION

주요 Hyperparameter 설정

- Number of latent factors
: 10 => 학습하고자 하는 latent factor matrix들의 차원을 결정.
=> user latent factor matrix : 2524x10
item latent factor matrix : 410x10
 - Learning rate : 0.05
- 100 epochs 학습 결과, MSE = 0.04 정도의 성능을 보여주었음.

최종 결과

- 기존 딥러닝 기반 Matrix factorization 모델보다 성능이 좋았음
: 동일한 hyperparameter 상에서 100 epochs 학습 결과 $MSE = 1.98$ 정도의 성능
우리 모델은 $MSE = 0.04$!
- 기존 딥러닝 모델과 유사하게 두 latent factor matrix를 학습하는 과정을 거쳤지만
학습 성능이 더 좋은 이유는 자세하게 파악 불가..
 - > latent factor CF 기법 자체가 블랙박스 모델이므로,,
 - > Bias 추가해서 그런 것이라고 조심스럽게 예측해 봄
- 하지만, 딥러닝 모델과 달리 MF with biases 모델에서는 따로 validation loss를
계산하지 않아서 객관적인 성능 비교가 어려움

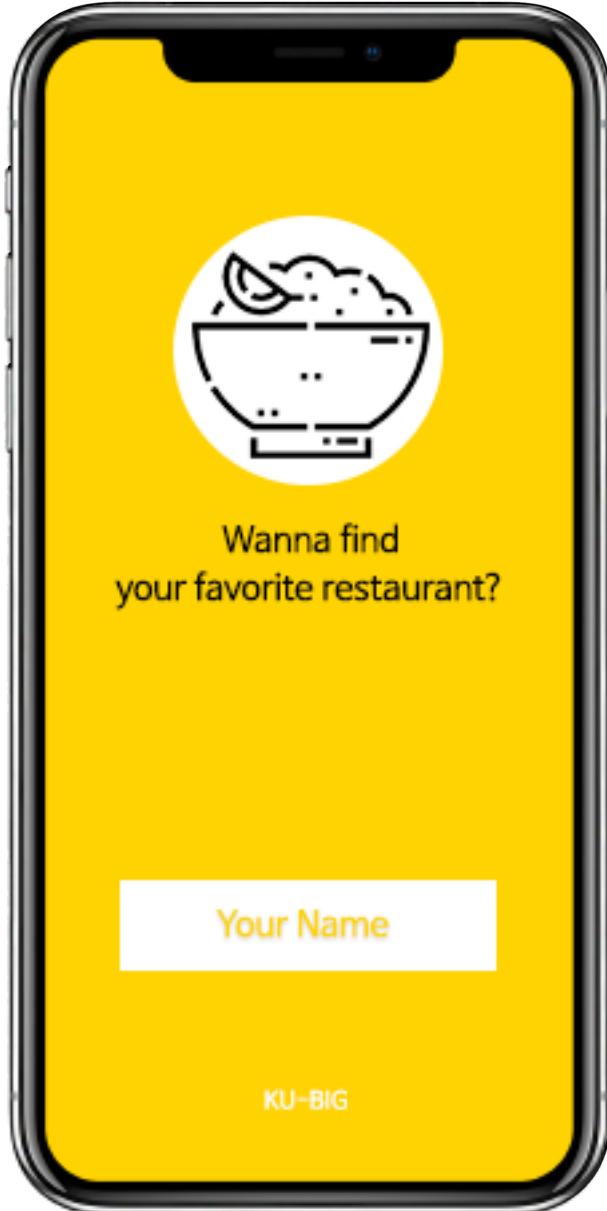
그래서 과연, 우리 모델은 마음에 쏙~들게 맛집 추천을 했을까?

학습 모델의 성능은 숫자로 표현 가능하지만, 추천 시스템 자체의 성능은 결국 **유저의 주관적인 판단**에 달려있습니다,, 그래서 저희는 팀원들이 각자 안암 맛집에 대한 평점 데이터를 추가적으로 생성한 뒤 기존 데이터에 추가하여 학습하였지요.. 우리 팀원들은 각각 어떤 맛집을 추천 받았을까요?!?!? 만족했을까요? 다음 페이지를 함 보시쥬~

PART(4)

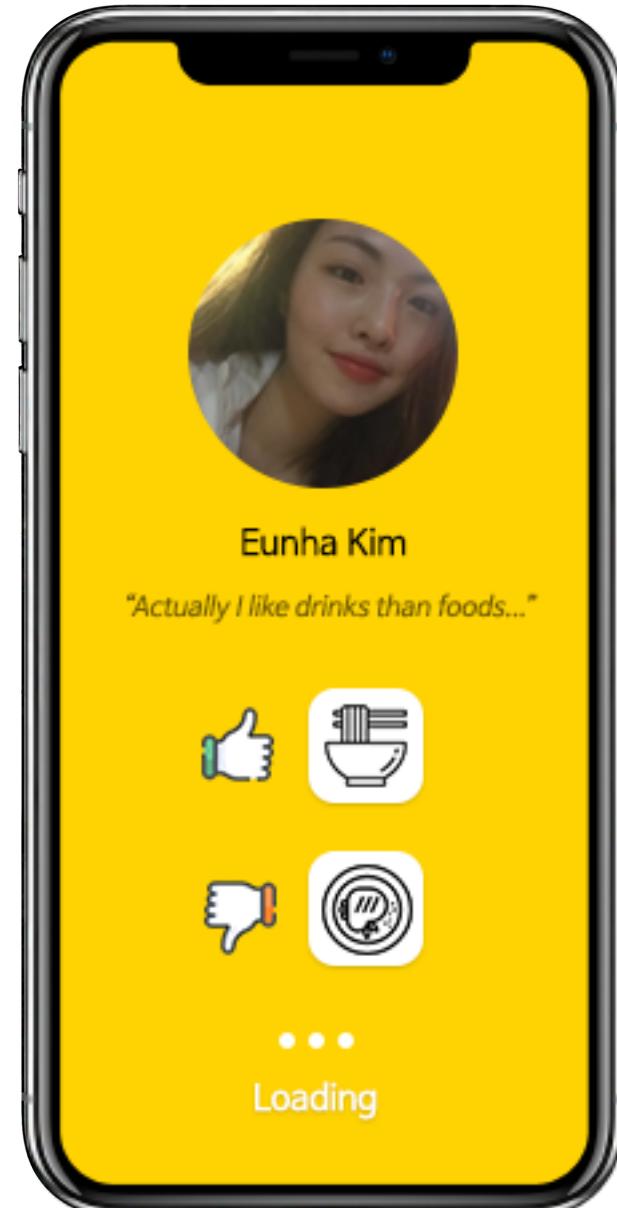
RECOMMENDATION FOR YOU

Recommendation For You



Putting Data

이름	place_id	태그	user_id	평점	예상평점
칠기마라탕	175	#먹어야산다#중식#마라탕#마라샹궈#정문앞#꿔바로우#포장#혼밥	2519	5.0	5.007120
서울쌈냉면 고대점	16	#먹어야산다#한식#냉면#정대후문#포장#혼밥#저렴#가성비	2519	5.0	5.028863
아미가	136	#먹어야산다#한식#백반#법대후문#포장#혼밥	2519	5.0	4.991504
치즈를 사랑한 짬닭	25	#먹어야산다#한식#짬닭#정대후문#포장	2519	5.0	4.899266
모이리타	31	#먹어야산다#양식#파스타#리조또#피자#참살이길#데이트#회식#모이리타	2519	5.0	5.016447
프로마치	83	#먹어야산다#양식#다이어트#샌드위치#샐러드#수프#참살이길#옆살이길#혼밥#포장	2519	5.0	4.889597
철남 (정대후문)	80	#먹어야산다#한식#두루치기#정대후문#회식#철판남매	2519	5.0	4.948096
찌개집	79	#먹어야산다#한식#백반#김치찌개#참살이길#뒷살이길#회식	2519	4.0	4.028410
미농	167	#먹어야산다#양식#제기동#브런치#감성	2519	4.0	3.995782
용조수	20	#먹어야산다#중식#사천식#꿔바로우#정대후문#포장#가성비	2519	4.0	4.030871
다손김밥	123	#먹어야산다#분식#김밥#수제비#정문앞#포장#혼밥	2519	4.0	3.987672
금룡짬뽕	42	#먹어야산다#중식#중화요리#짜장면#짬뽕#탕수육#참살이길#포장#혼밥	2519	4.0	4.007429
언니네반점	3	#먹어야산다#중식#정대후문#사천식#혼밥#가성비#저렴#포장#회식	2519	3.0	3.019057
본샤브칼국수	57	#먹어야산다#참살이길#한식#버섯칼국수#샤브샤브#혼밥#가성비	2519	3.0	3.037692
밀플랜비 (Meal Plan B)	55	#먹어야산다#정대후문#멕시칸#부리또#핫도그#혼밥#저렴#포장	2519	3.0	3.087279
고른햇살	9	#먹어야산다#분식#정대후문#포장#혼밥#가성비	2519	3.0	3.480282
청년다방 고대안암점	23	#먹어야산다#참살이길#꿀꺽#안암로터리#분식#퓨전#떡볶이#감자튀김	2519	3.0	3.023647
나정순할매쭈꾸미	44	#먹어야산다#쭈꾸미#이공계#안암로터리	2519	2.0	2.074459
고래돈까스	29	#먹어야산다#일식#돈까스#정문앞#포장#혼밥#회식#가성비	2519	1.0	1.040772
돈까스하우스	11	#먹어야산다#경양식#돈까스#정대후문#포장#혼밥#가성비#돈까	2519	1.0	1.072481



Collaborative Filtering

	이름	place_id	태그	예상평점
1534	안암역 봉어빵	17	#먹어야산다#정대후문#봉어빵#계란빵#버블호떡#포장마차#테이크아웃	10.743619
7065	밥이랑 고기랑 (류 본점)	396	#먹어야산다#한식#백반#고기#정문앞#혼밥	10.533425
4662	배달의 쌀국수 동대문점	129	#먹어야산다#제기동#쌀국수#배달	10.098119
6235	노랑통닭	238	#먹어야산다#치킨#참살이길#옆살이길#맥주#배달#포장	9.776213
5647	비야 고대본점	188	#먹어야산다#한식#부대찌개#정대후문#회식#저렴	9.297986
6366	59쌀피자 제기동	252	#먹어야산다#피자#제기동#고대사거리#포장#배달#오구쌀피자	9.123276
7059	멕시카나치킨	393	#먹어야산다#치킨#정문앞#배달#포장#고대사거리	9.039389
7008	쎄오쎄오	356	#먹어야산다#성신여대입구#동남아#쌀국수#분짜#볶음밥	8.407972
7040	모심	378	#먹어야산다#한식#정대후문#코스#회식	8.078314
7036	로라방앗간	375	#먹어야산다#떡볶이#납작만두#튀김	7.946490
6871	폭립하우스	317	#먹어야산다#이공계#보문#고기#폭립#삼겹살구이#배달	7.930813
6933	마라천사	337	#먹어야산다#중식#마라탕#마라샹궈#정대후문	7.676685
7033	전통만두국이상조	373	#먹어야산다#만두#빈대떡	7.660775
7045	누리수제닭강정	383	#먹어야산다#이공계#닭강정	7.587776
5571	김밥천국 정대후문점	180	#먹어야산다#분식#정대후문#포장#혼밥	7.553258
6068	고대고기집	223	#먹어야산다#고고집#고기#참살이길#회식#가성비	7.414649
6445	돌냄비 열우동	261	#먹어야산다#경양식#분식#돈까스#우동#정문앞#고려대역#포장#혼밥#저렴	7.310137
6795	맥도날드 법대후문	301	#먹어야산다#햄버거#법대후문#배달#포장#혼밥#저렴#가성비	7.288534
5597	남이네 분식	183	#먹어야산다#한식#분식#백반#칼국수#참살이길#뒷살이길#혼밥#저렴#가성비	7.263210
6077	놀부부대찌개&철판구이 고대점	224	#먹어야산다#한식#부대찌개#참살이길#포장	7.261297



Contents based

이름	코드	태그	score
사랑마라탕	Num=a_nam_0357	#먹어야산다#중식#마라탕#마라샹궈#정대후문#포장#혼밥	0.801784
라화쿵부(법후)	Num=a_nam_851	#먹어야산다#중식#마라탕#마라샹궈#법대후문#혼밥#포장	0.801784
신룽푸마라탕	Num=a_nam_0094	#먹어야산다#중식#마라탕#마라샹궈#참살이길#배달#포장#혼밥	0.750000
신마라탕	Num=a_nam_1038	#먹어야산다#정문앞#사천식#중식#마라탕#마라샹궈#혼밥#훠궈뷔페	0.750000
이름	코드	태그	score
투고샐러드 고려대점	Num=a_nam_0208	#먹어야산다#다이어트#샐러드#참살이길#뒷살이길#포장#테이크아웃#혼밥	0.670820
서브웨이 고려대점	Num=a_nam_0045	#먹어야산다#샌드위치#참살이길#포장#테이크아웃#혼밥	0.645497
본죽 고려대점	Num=a_nam_0024	#먹어야산다#한식#죽#다이어트#참살이길#포장#혼밥	0.645497
히포크라테스 스포	Num=a_nam_0067	#먹어야산다#베이커리#참살이길#뒷살이길#샐러드#스튜#마카롱#포장#다이어트#혼밥#감성	0.572078
싸움의고수 고대점	Num=a_nam_0003	#먹어야산다#보쌈#참살이길#포장#혼밥	0.565685
샌디앤샐리 (sandy&sally)	Num=a_nam_857	#먹어야산다#샐러드#정문앞#다이어트#샌드위치#포장#테이크아웃#커피	0.559017
베트남 노상식당	Num=a_nam_812	#먹어야산다#참살이길#옆살이길#동남아#쌀국수#배달#포장#혼밥	0.559017
이즈 스테이크	Num=a_nam_908	#먹어야산다#경양식#참살이길#샐러드#스테이크#배달#포장#혼밥#가성비	0.527046
히포크라테스 스포 2호점	Num=a_nam_969	#먹어야산다#정문앞#수프#스튜#다이어트#샐러드#마카롱#감성#혼밥	0.527046
사보텐 고대점	Num=a_nam_0054	#먹어야산다#일식#돈까스#참살이길#포장#혼밥	0.516398



Putting Data



	이름	place_id		태그	user_id	평점	예상평점
170	우승식당	4	#먹어야산다#불고기#이공계#제육볶음	2520	5.0	4.857970	
2836	나정순할매쭈꾸미	44	#먹어야산다#쭈꾸미#이공계#안암로터리	2520	5.0	4.891445	
6841	투고샐러드 고려대점	308	#먹어야산다#다이어트#샐러드#참살이길#뒷살이길#포장#테이크아웃#혼밥	2520	5.0	5.052972	
6794	샐러데이	300	#먹어야산다#다이어트#샐러드#정대후문#테이크아웃	2520	5.0	5.000110	
6757	서대문 꼼장어	296	#먹어야산다#꼼장어#안암로터리#포장	2520	5.0	4.981842	
...
4337	육쌈냉면 고대점	107	#먹어야산다#한식#냉면#참살이길#포장#혼밥#회식#저렴	2520	2.0	2.014873	
341	고대맛집ㅋㅋ	8	#먹어야산다#분식#안암오거리#이공계#혼밥#회식#백반#떡볶이	2520	2.0	3.341917	
5274	BHC치킨	164	#먹어야산다#치킨#정문앞#배달#포장	2520	1.0	0.966398	
5596	끼니와 새참	182	#먹어야산다#분식#한식#비빔밥#제육덮밥#교내#중앙광장	2520	1.0	1.004024	
6644	신풍푸마라탕	280	#먹어야산다#중식#마라탕#마라샹궈#참살이길#배달#포장#혼밥	2520	1.0	1.090027	



Collaborative Filtering

	이름 place_id	태그	예상평점
5722	영자's 곰창 193	#먹어야산다#고기#곰창#참살이길#옆살이길#포장#회식	7.149815
7043	남매부대찌개 381	#먹어야산다#부대찌개#월곡역#월곡#등덕여대#배달#포장#배달의민족#배민	7.060417
7046	콩시콩뼈 간자탕 384	#먹어야산다#정대후문#한식#백반#국밥#콩비지	6.849684
7059	멕시카나치킨 393	#먹어야산다#치킨#정문앞#배달#포장#고대사거리	6.617338
6232	427 월곡 닭갈비 237	#먹어야산다#상월곡역#한식#닭갈비#포장#파전	6.524654
6976	오소리순대 346	#먹어야산다#정문앞#한식#순대#국밥#고려대역#포장#혼밥#회식	6.364672
6301	59쌀피자 안암오거리점 247	#먹어야산다#피자#안암로터리#이공계#배달#포장	6.235809
6552	봉추찜닭 고대점 273	#먹어야산다#한식#찜닭#참살이길#배달#포장#회식	6.137538
6737	교촌치킨 294	#먹어야산다#고대사거리#배달#포장	6.063860
5443	우리동네만두가게 173	#먹어야산다#만두#우동#보문#이공계	6.058338
7078	토라 돈부리 404	#먹어야산다#성신여대입구#일식#돈부리	6.050759
6179	공자식탁 235	#먹어야산다#중식#중화요리#짜장면#배달#포장#혼밥#고려대역#가성비	6.036455
6898	치킨소리 안암점 326	#먹어야산다#치킨	5.883708
7000	즐거운맛돈까스 353	#먹어야산다#신설동#일식#돈까스	5.864677
7009	두레식당 357	#먹어야산다#한식#백반#법대후문#혼밥#저렴	5.791893
3290	알래스카에서 온 연어가 맛있는 집 66	#먹어야산다#해산물#정문앞#고려대역#연어#조밥#소주#포장#회식	5.742138
6846	본죽 고려대점 311	#먹어야산다#한식#죽#다이어트#참살이길#포장#혼밥	5.735082
4064	가즈시돈까스 88	#먹어야산다#일식#돈까스#소비#참살이길#뒷살이길#혼밥#가성비	5.652845
5263	Augusty59 163	#먹어야산다#바베큐#플레터#수제맥주#성신여대	5.601607



Contents based

이름	코드	태그	score
비야 이공후문	Num=a_nam_0177	#먹어야산다#한식#이공계	0.57735
누리수제닭강정	Num=a_nam_990	#먹어야산다#이공계#닭강정	0.57735
나정순할매쭈꾸미	Num=a_nam_845	#먹어야산다#쭈꾸미#이공계#안암로터리	0.50000
박가네 빠다귀 해장국	Num=a_nam_0181	#먹어야산다#한식#이공계#배달	0.50000
두부촌	Num=a_nam_720	#먹어야산다	0.50000
스시진	Num=a_nam_0205	#먹어야산다#일식#이공계#배달	0.50000
구룡포전어횟집	Num=a_nam_840	#먹어야산다#회	0.50000
산학관 학생식당	Num=a_nam_1054	#먹어야산다	0.50000
엄자순만두	Num=a_nam_915	#먹어야산다	0.50000
박쉐프의 스키야끼	Num=a_nam_725	#먹어야산다	0.50000

번외편)

프로젝트를 .. 마치며...

프로젝트를 마치며



- 짧은 프로젝트 기간으로 부족한 EDA, 크롤링, 웹개발, 이론 공부
- 예상보다 훨씬 다양한 추천시스템 기법들
- 데이터들의 feature 부족으로 인한 아쉬움

- 실생활에서 접하기 쉬운 데이터를 이용하여 재미있고 의미 있는 분석
- 가시적인 결과물을 통해 직관적인 이해와 평가

프로젝트를 마치며

재훈

정현

은하

혜연

기태

수빈



♥ 좋은 팀원들 ♥