

# 맛집 추천 시스템

곽민지, 김미소



## 요약

회사가 보유한 레스토랑 예약 및 방문 DB를 바탕으로 적절한 레스토랑을 추천하는 시스템을 만들어내는 프로젝트였습니다.

우선 데이터를 preprocessing하였습니다. 여러개의 data table이 있었고, 이를 의도에 맞게 정리하는 것에 많은 노력이 필요하였습니다. 레스토랑의 정보와 예약 및 방문 정보가 서로 다른 data table에 있었기에 적절하게 join하였고, 어마어마한 data를 보다 더 손쉽게 다루기 위해 특정 지역별, 인기도 순으로 data를 filtering하였습니다.

다음은 모델 구축이었습니다. 저는 추천 시스템에 관련된 학술보고서 및 논문을 참고하여 몇 가지 모델을 구성하였습니다. 그중에서 가장 효과적인 모델은 '협력적 필터링 모델'이었습니다. 이것은 다른 사용자와의 유사도를 계산해낸 뒤 그 유사도를 가중치로 계산하여 특정 레스토랑에 대한 어떤 사용자의 선호도를 예측해내는 계산 엔진입니다. 즉, 내가 선호하는 레스토랑을 선호하므로 나와 유사하다고 할 수 있는 다른 사용자들이 선호하는 레스토랑을 나에게 추천하는 것입니다. 이 방법 외에도 선호하는 레스토랑의 유형을 파악하여 새롭게 구성한 데이터를 바탕으로 '컨텐츠 기반 모델'을 접목하여 부족한 데이터를 보충함으로써 하이브리드 모델을 만들어냈습니다. 그 외에도 Kmeans clustering, random forest 등 데이터 마이닝 기법을 적용하여 추천 모델을 완성하였습니다.



## 요약

마지막으로는 웹형식으로 시각화를 구현하였습니다. 정식으로 웹 개발을 배운 적은 없었기 때문에 혼자 힘으로 독학하였습니다. html, css 언어를 통해 화면을 구성하였고 지도 api와 javascript를 끼어서 지도 위에 추천 레스토랑을 표시하는 기능을 구현하였습니다. 이렇게 구성한 화면을 database와 연동하기 위해 python 언어와 flask 라이브러리로 프로그램을 코딩하였습니다. 그 결과 특정 사용자가 자신의 구매 내역 및 자신이 선택한 선호 옵션을 가지고 자유롭게 레스토랑을 검색하고 추천받을 수 있게 하였습니다.

웹 형식으로 되어 있고 맛집 검색 버튼을 누르면 추천하는 맛집의 목록이 뜨는 동시에 네이버 지도 위에 위치가 나타나는 형식입니다. 여러 옵션을 조정할 수 있는데 기존 회원의 경우 데이터베이스에 저장된 사용자 id 중에 하나를 선택하도록 되어있습니다 해당 사용자 id가 전에 방문하였던 맛집 내역들을 기반으로 추천이 이루어지게 됩니다 신규 회원의 경우 방문했던 맛집 내역이 없으므로 단순 인기 순으로 추천이 이루어지게 됩니다 식당 필터 버튼을 누르면 오른쪽과 같이 다양한 식당 업종 별로 필터링해서 추천을 받을 수 있습니다.

## ② 프로젝트 목표

네이버 지도 검색에 업체를  
등록할 수 있는 서비스



업체를 예약할 수 있는 서비스



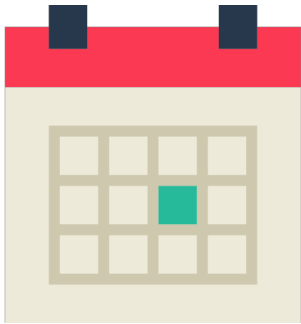
최종목표 :

“당신만을 위한 예약 하기 좋은 맛집 추천”

목표 : 다양한 알고리즘을 적용하여 예측의 정확도 향상, 추천 리스트 평가

## ③ 데이터

1. 식당에 **예약**을 한 데이터



약 300만 건의 예약

2. 식당에 **전화**를 건 데이터



약 400만 건의 전화

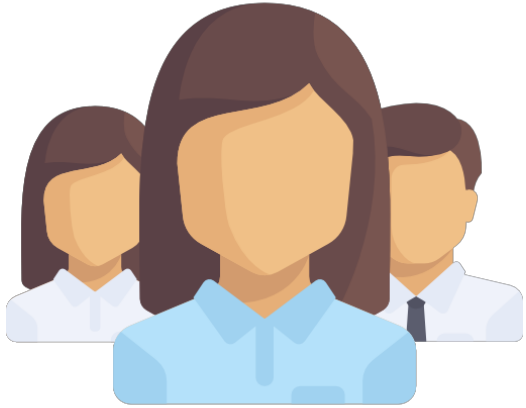
3. 업체 정보



약 3만 개의 업체

### ③ 데이터

4. **성별, 연령**의 데모 데이터



약 80만 명의  
성별/나이 데이터

5. 사용자의 **검색** 데이터.



약 80만 명의  
상위 10,000개 검색어 데이터

## ② 데이터 전처리

### 1 지역 샘플링



Toy 모델 빌드를 위해  
서초구, 강남구  
관악구 지역 사용

### 2 중복 제거



같은 사람, 같은 업체,  
같은 날짜인 경우  
하나로 처리

### 3 예약 3회 이상 사용자



아주 적은 예약을 한  
사용자 제거

### ③ 기본 알고리즘 : CF (협업 필터링)

아이템 기반의 협업 필터링



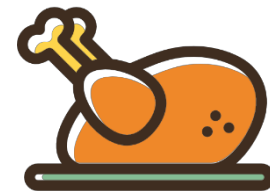
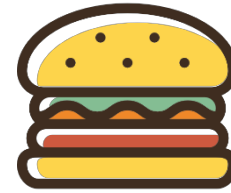
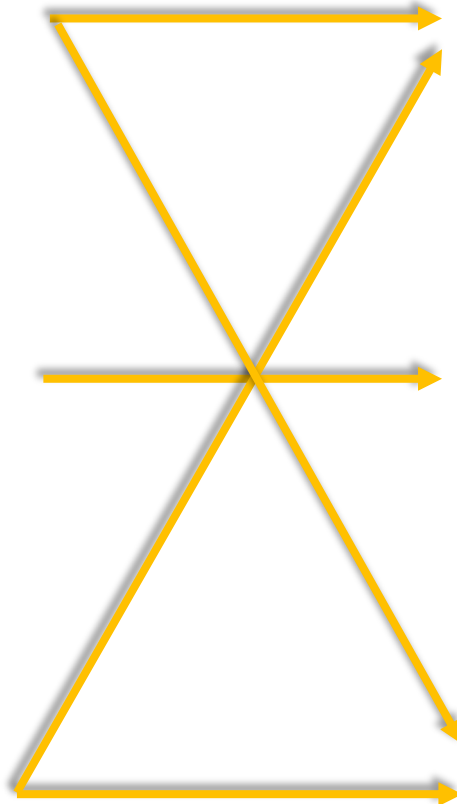
음식점 예약



음식점 예약



음식점 예약



예약 기록의 유사도를 반영하여 음식점을 추천



### ③ 기본 알고리즘 : CF (협업 필터링)

사용자	업체	빈도수
미소	업체A	5
미소	업체B	2
민지	업체C	1
민지	업체B	2

사용자-빈도수 Table

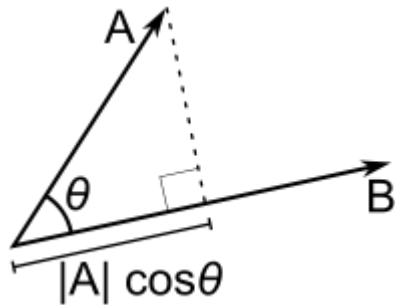


사용자-빈도수 Pivot Table

업체 사용자	업체A	업체B	업체C
미소	5	2	0
민지	0	2	1
소민	1	1	3

### ③ 기본 알고리즘 : CF (협업 필터링)

$$A = [5 \ 0 \ 1]$$
$$B = [2 \ 2 \ 1]$$



사용자-빈도수 Pivot Table

업체 \ 사용자	업체A	업체B	업체C
미소	5	2	0
민지	0	2	1
소민	1	1	3

음식점-음식점 유사도 Table

	업체A	업체B	업체C
업체A	1	0.001	0.027
업체B	0.001	1	0.013
업체C	0.027	0.013	1

### ③ 기본 알고리즘 : CF (협업 필터링)

**목표**  
미소가  
C음식점을  
몇 번 갈 건지  
추정하기

사용자-빈도수 Pivot Table  
미소는 업체A와 업체B를 얼마나 선호했는가

	업체A	업체B	업체C
미소	5번	1번	?번

음식점-음식점 유사도 테이블  
업체C는 업체A,업체B와 얼마나 유사한가

	업체 A	업체 B	업체 C
업체 A	1	0.001	0.4
업체 B	0.001	1	0.013
업체 C	0.4	0.013	1

$$1) \frac{\sum_{j \in N} W_{ij} r_{ui}}{\sum_j |W_{ij}|} = \frac{5\text{번} * 0.4 - 1\text{번} * 0.013}{0.4+0.013} = \text{숫자 로 예측.}$$

- 2) 위와 같은 방식으로 모든 음식점에 대한 추정 값을 구하고  
값이 높은 N개의 음식점 추천

### ③ 기본 알고리즘 : CF (협업 필터링)

**목표**  
미소가  
C음식점을  
몇 번 갈 건지  
추정하기

사용자-빈도수 Pivot Table  
미소는 업체A와 업체B를 얼마나 선호했는가

	업체A	업체B	업체C
미소	5번	1번	4.87번

음식점-음식점 유사도 테이블  
업체C는 업체A,업체B와 얼마나 유사한가

	업체 A	업체 B	업체 C
업체 A	1	0.001	0.4
업체 B	0.001	1	0.013
업체 C	0.4	0.013	1

$$1) \frac{\sum_{j \in N} W_{ij} r_{ui}}{\sum_j |W_{ij}|} = \frac{5\text{번} * 0.4 - 1\text{번} * 0.013}{0.4+0.013} = 4.87 \text{ 로 예측.}$$

- 2) 위와 같은 방식으로 모든 음식점에 대한 추정 값을 구하고  
값이 높은 N개의 음식점 추천

## ④ 성능 평가 지표

$$\text{Recall (재현율)} = \frac{\text{옳게 추천한 아이템 수}}{\text{고객이 실제로 구매한 아이템 수}}$$

### Recall at top 20 – 추천 리스트 평가지표

(2) N= 20

- 데이터 의 Sparsity고려
- 실제 추천 서비스 제공 시에 사용자에게 노출되는 상위 아이템만을 평가

#### (1) Precision이 아닌 Recall 채택

고객이 실제구매한 아이템 로 수에 더 가중치를 주기 위함

\* Precision : 옳게 추천한 아이템 수 / 추천한 아이템 수

# 4 실험 설계

실제 데이터

	식당1	식당2	식당3
사람1	1	1	0
사람2	0	1	1
사람3	1	0	1

6건의 예약

Drop-rate  
1/2

구멍 뚫어서 만든 학습 데이터

	식당1	식당2	식당3
사람1	1	0	0
사람2	0	1	0
사람3	0	0	1

3건의 예약

예측(추천)

예측 결과(추천)

	식당1	식당2	식당3
사람1	0.99	0.33	0.19
사람2	0.45	0.88	0.32
사람3	0.56	0.23	0.87

사용자마다 상위 20개 추천

결과 비교:  
recall=맞춘 건수/실제 건수=5/6

성능 평가 결과

고객 클러스터링 사용

컨텐츠 기반 모델을 접목

Random Forest

성능 평가 결과

	채택
데이터 전처리 방식	예약했으면 1 예약하지 않으면 0
사용 알고리즘	사용자 기반 협업 필터링
유사도 계산 방식	자카드 유사도

	식당1	식당2	식당3
사람1	1	1	1
사람2	1	0	0
사람3	1	0	1

Recall = 0.3016

성능 평가 결과

	채택
데이터 전처리 방식	예약했으면 1 예약하지 않으면 0
사용 알고리즘	사용자 기반 협업 필터링
유사도 계산 방식	자카드 유사도

취향이 비슷한 다른 사용자가 선호하는 아이템을 추천

Recall = 0.3016



성능 평가 결과

고객 클러스터링 사용

컨텐츠 기반 모델을 접목

Random Forest

성능 평가 결과

	채택
데이터 전처리 방식	예약했으면 1 예약하지 않으면 0
사용 알고리즘	사용자 기반 협업 필터링
유사도 계산 방식	자카드 유사도

Recall = 0.3016

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

코사인 유사도,  
유클리드 유사도에 비해  
높은 성능을 보여줌

# 5 실험 결과

## 성능 평가 결과

## 고객 클러스터링 사용

## 컨텐츠 기반 모델을 접목

## Random Forest

## 단순 협업 필터링의 한계

두 고객이 모두 구매를 한 상품에 대해서만 유사도를 구할 수 있어 추천 정확성을 낮출 우려



**K-means** 클러스터링으로 가까운 이웃 집단을 찾아 추천 시행

## 단순 협업 필터링 결과

전체 사용자 협업 필터링 (클러스터 5 사용자)	전체 사용자 협업 필터링 (클러스터 12 사용자)
0 더 리버사이드 호텔 더가든키친	0 더 리버사이드 호텔 더가든키친
1 오레노 강남역점	1 오레노 강남역점
2 울프강 스테이크하우스	2 울프강 스테이크하우스
3 몽중헌	3 몽중헌
4 빙스 도곡역점	4 빙스 도곡역점
5 아웃백 도곡점	5 아웃백 도곡점

## 고객 클러스터링 결과

클러스터5 협업 필터링 (클러스터 5 사용자)	클러스터 12 협업 필터링 (클러스터 12 사용자)
0 도교등심 선릉 스카이라운지	0 페어링룸
1 일일향2호점 (논현점)	1 스시선수
2 도교등심트리플 청담점 [1]	2 스시코우지
3 마코토	3 스시초희
4 수담 한정식	4 우오
5 투뿔등심 가로수길점9	5 몽중헌

## 5 실험 결과

K- means 클러스터링 후  
**거대 클러스터 문제 발생**    원인 : 높은 sparsity로 인함

클러스터1	0.6789	클러스터7	0.6790
클러스터2	0.6939	클러스터8	0.6658
클러스터3	0.6701	클러스터9	0.6534
클러스터4	0.6746	클러스터10	0.6626
클러스터5	0.7925	클러스터11	0.6778
클러스터6	0.7964	클러스터12	0.7187
클러스터0	0.2659	평균	0.6638

18000명의 사용자가 하나의 클러스터 0에 쏠림  
9000명은 고루 분포 → **거대 클러스터 문제 발생**

# 5 실험 결과

성능 평가 결과

고객 클러스터링 사용

컨텐츠 기반 모델을 접목

Random Forest

Sparse한 사용자

-> 식당의 feature 데이터를 활용(주요 메뉴, 토픽, 분위기, 테마 등)

식당이름	feature
식당A	한정식 튀김 불고기 점심특선
식당B	샐러드바 신선한 디저트 초밥



feature들을 선별하여  
아이템 - feature 매트릭스

	갈비	초밥	f3	f4	f5
식당1	0	1	0	0	0
식당2	1	1	1	0	1
식당3	0	1	1	1	1



	식당1 식당2
사용자1	
사용자2	

 $\times$ 

	특성1 특성2
식당1	
식당2	

 $=$ 

	특성1 특성2
사용자1	
사용자2	

사용자- feature의 매트릭스가 도출된다(사용자가 어떤 feature의 식당을 주로 이용했는지)

# 5 실험 결과

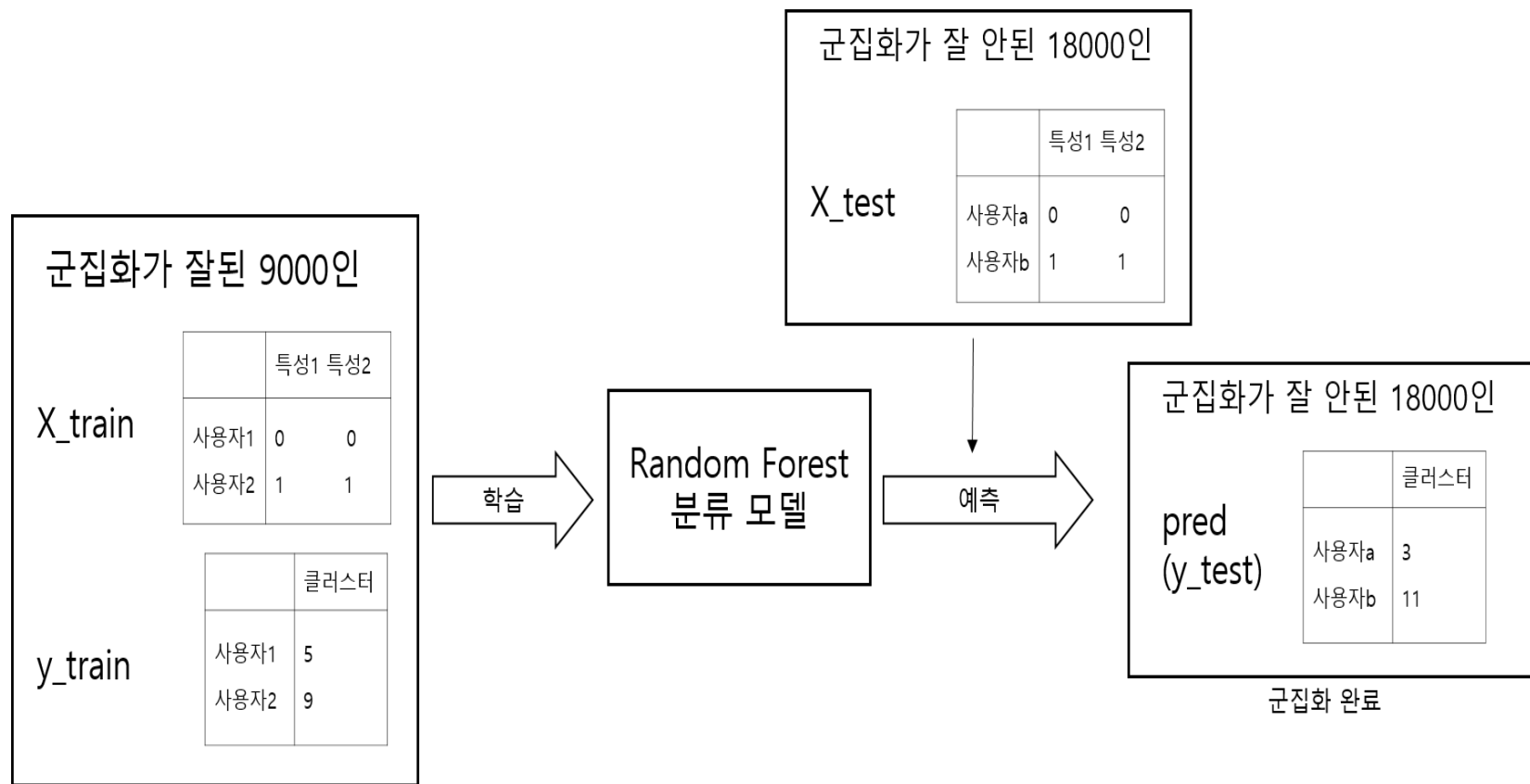
성능 평가 결과

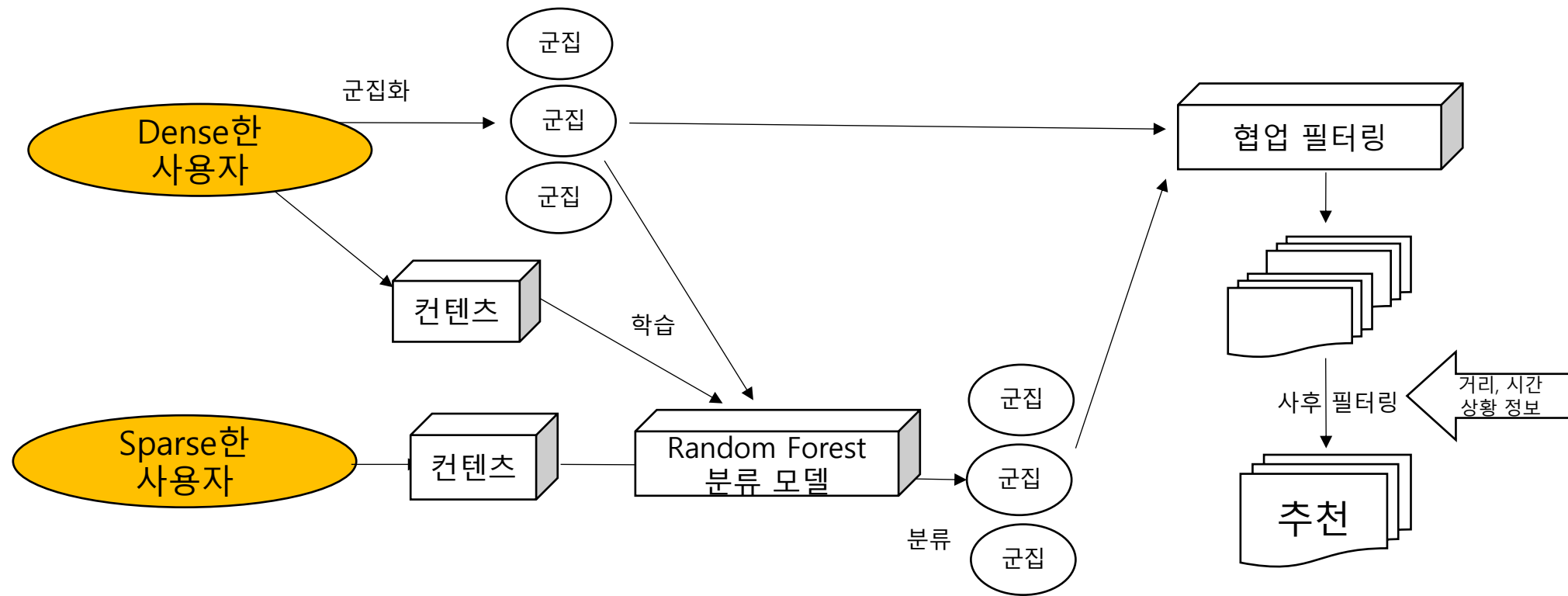
고객 클러스터링 사용

컨텐츠 기반 모델을 접목

Random Forest

18000명을 재분류(9000명의 K means 클러스터링 수행결과 기준)







내가 간 맛집 추천 맛집

사용자ID : 신규 회원  
추천개수 : 10

☒ 신규 회원  
☐ 기존 회원

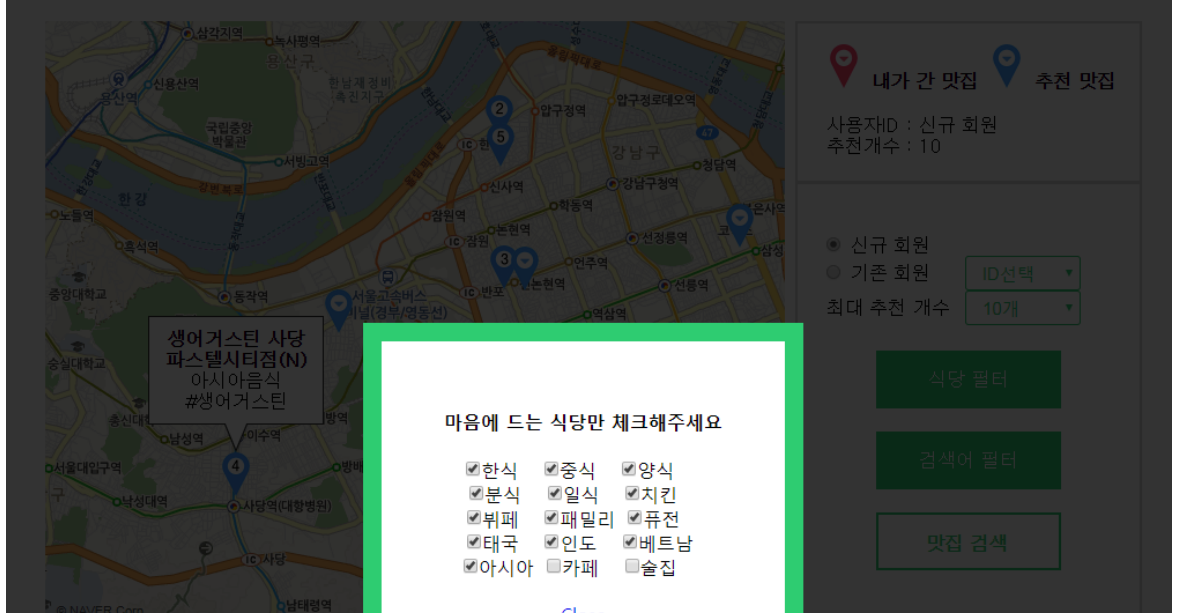
ID선택

최대 추천 개수 10개

식당 필터

검색어 필터

맛집 검색



추천 순위	맛집 이름	분류	해시태그
1	팬아시아 가로수점	아시아음식	-
2	생어거스틴 강남교보타워점(N)	아시아음식	#생어거스틴
3	생어거스틴 사당 파스텔시티점(N)	아시아음식	#생어거스틴
4	생어거스틴 가로수길점	아시아음식	#생어거스틴
5	타이키친	태국음식	-
6	생어거스틴 서래본점(N)	아시아음식	#생어거스틴

추천 순위	맛집 이름	분류	해시태그
1	팬아시아 가로수점	아시아음식	-
2	생어거스틴 강남교보타워점(N)	아시아음식	#생어거스틴
3	생어거스틴 사당 파스텔시티점(N)	아시아음식	#생어거스틴
4	생어거스틴 가로수길점	아시아음식	#생어거스틴
5	타이키친	태국음식	-
6	생어거스틴 서래본점(N)	아시아음식	#생어거스틴

웹 형식으로 되어 있고 맛집 검색 버튼을 누르면

추천하는 맛집의 목록이 뜨는 동시에 네이버 지도 위에 위치가 나타나는 형식입니다.

여러 옵션을 조정할 수 있는데

기존 회원의 경우 데이터베이스에 저장된 사용자 id 중에 하나를 선택하도록 되어있습니다  
해당 사용자 id가 전에 방문하였던 맛집 내역들을 기반으로 추천이 이루어지게 됩니다

신규 회원의 경우 방문했던 맛집 내역이 없으므로 단순 인기 순으로 추천이 이루어지게 됩니다

식당 필터 버튼을 누르면 오른쪽과 같이 다양한 식당 업종 별로 필터링해서 추천을 받을 수 있습니다.