

배달의 민족 리뷰 분석을 통한 다속성별 추천 시스템

2022년 5월 18일

Imagine AI 팀

경제금융학부 팀장 유민균

컴퓨터공학과 팀원 강주영

휴먼지능정보공학과 팀원 구화목

Contents

1. Introduction 회사소개
2. Problem 문제점
3. Solution 해결방법 Our Advantages 강점
4. System Flowchart 시스템 플로우차트
5. 기술 구현 상세 설명
6. Future Perspective 발전계획
7. 일정 계획표

1.Introduction

회사소개

(주)우아한 형제들

문 앞으로 배달되는 일상의 행복

우아한형제들은 배달이 일상을 조금 더 행복하게 하도록

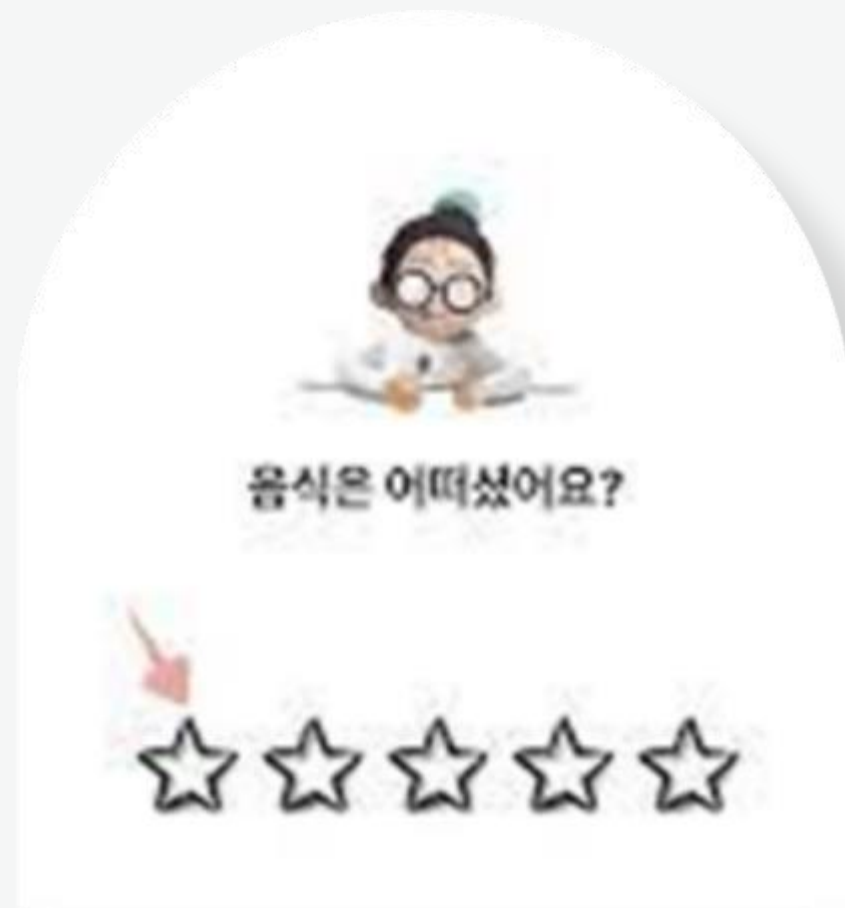
오늘도 달리고 있습니다.

세상은 넓고 맛집은
배달의민족



2. Problem

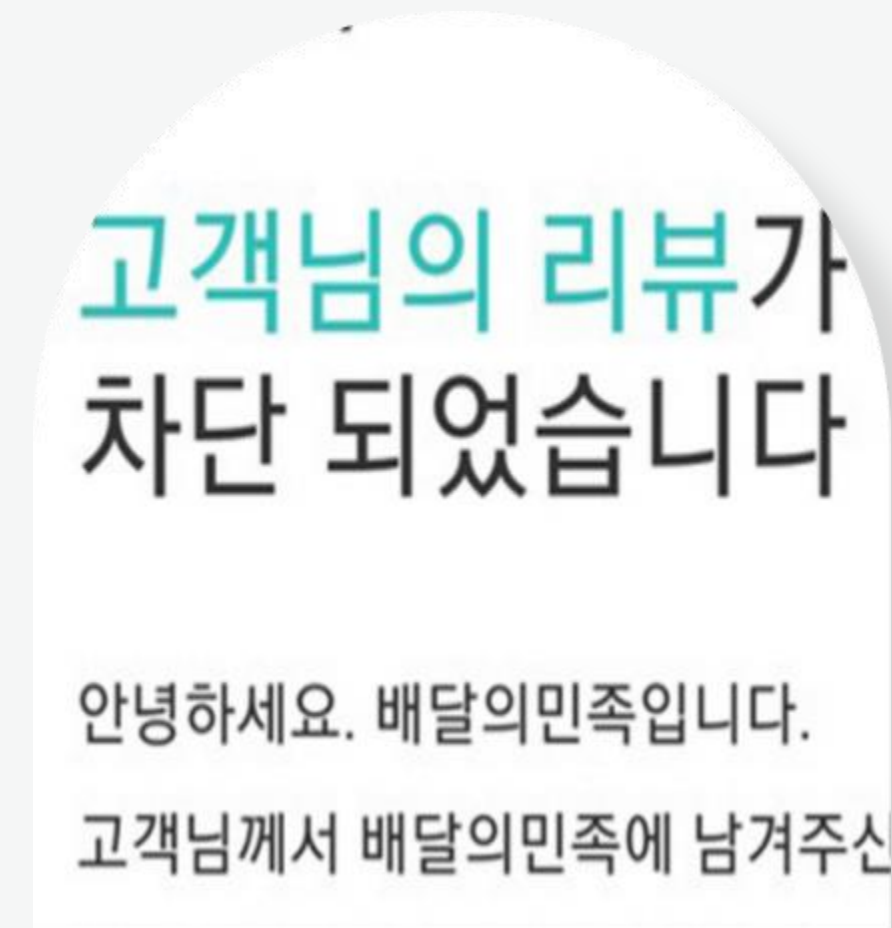
문제점



1. 속성별 리뷰가 없음



2. 무질서하며 신뢰도 낮은 리뷰

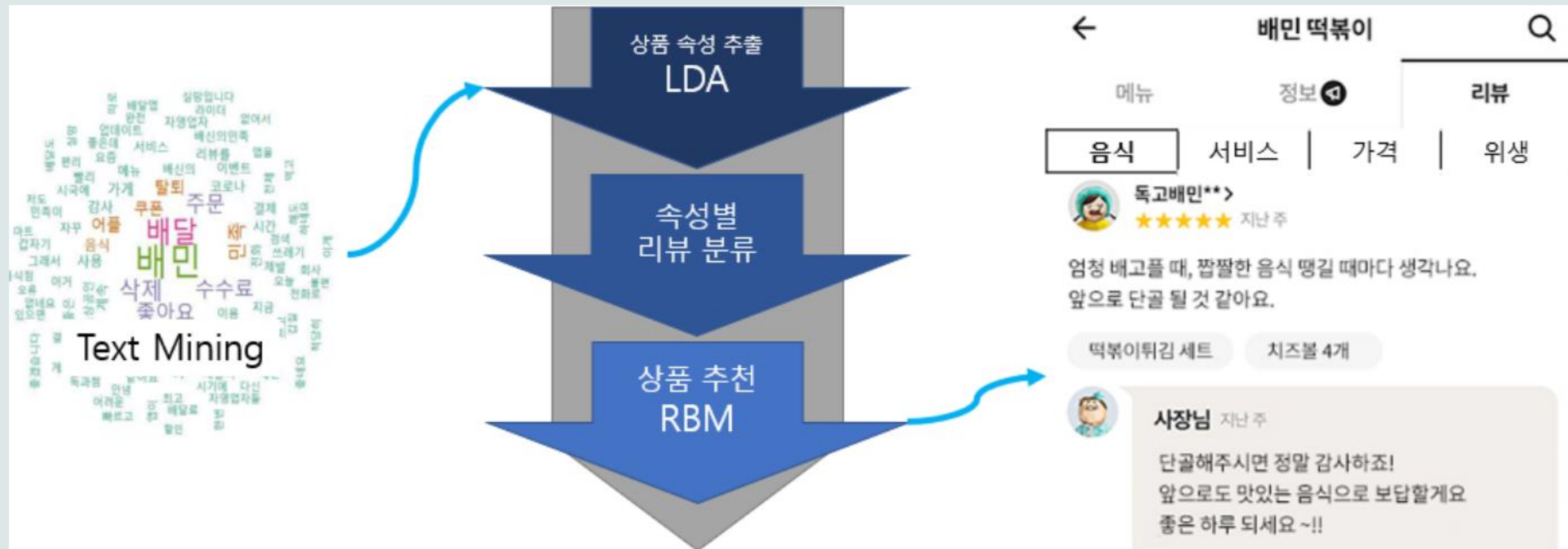


3. 마구잡이 소비자 리뷰 차단 문제

3. Solution 해결방법 Our Advantages 강점

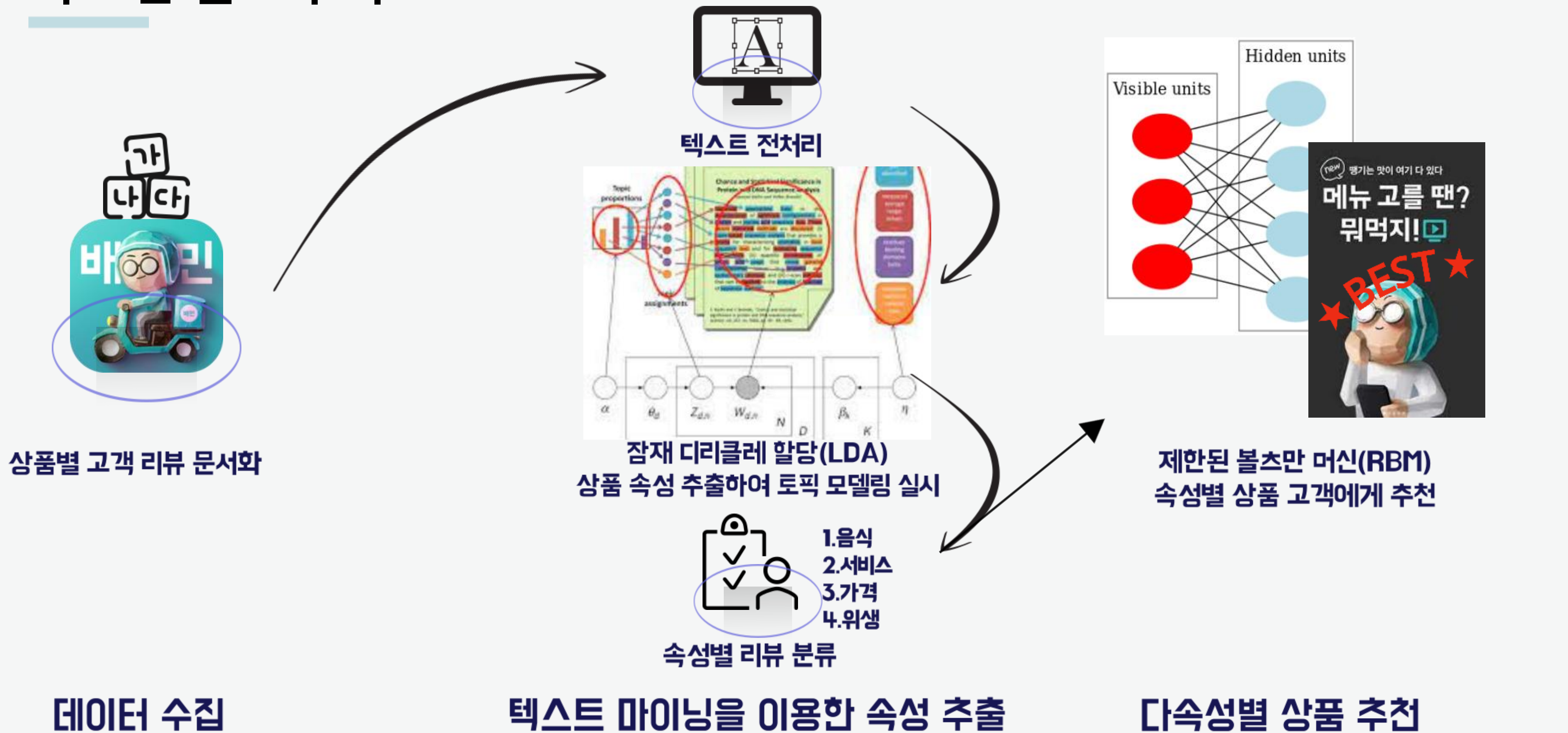
☐ 다속성별 추천 시스템

1. 상품의 속성별 평가 정보를 제공함으로써 리뷰추천 정확도를 높일 수 있다
2. 허위 리뷰 탐지 AI 오류를 줄일 수 있다



4. System Flowchart

시스템 플로우차트



5. 기술 구현 상세 설명

1. 리뷰 DB 구축

2. 텍스트 마이닝

1. 데이터 전처리

2. 토픽모델링(LDA)
상품 속성 추출

3. 속성별 리뷰 분류

3. 점포 추천

1. 속성 별 평점 부여

2. 속성별 점포 평점
매트릭스 구성

3. 점포 추천(RBM)

1. 리뷰 DB 구축

짱구 >
★★★★★ 이번 달

무척 맛있었어요
아쉬운 건 수저세트를 주지 않은거요

(맛집랭킹)우시장순대국밥(공기밥포함)

김민성 >
★★★★★ 지난 달

술먹은 다음날 해장하기 딱좋네요ㅜㅜ
자주 시켜먹고있습니다ㅎㅎ 설렁탕 존맛탱!

(맛집랭킹)우시장설렁탕(공기밥포함) **추천**

유니 >
★★★★★ 지난 달

(맛집랭킹)우시장돼지국밥(공기밥포함) 공기밥추가

절대미식가 >
★★★★★ 어제, 수정됨

돼지국밥에 돼지고기는 별루없고 고기도 딱딱 국물만! 잔뜩~이런 돼지국밥 처음임 별 세개도 아까움!

(맛집랭킹)우시장돼지국밥(공기밥포함)

○○ >
★★★★★ 지난 달

(맛집랭킹)우시장돼지국밥(공기밥포함) 물냉면

kimran >
★★★★★ 지난 달

신고하기





닉네임	가게이름	평점	리뷰
링링링	설렁탕	1.0	처음시켜봤는데 리뷰이벤트 참여 한다 했는데 무시를 하셨네요.
절대미식가	설렁탕	3.0	돼지국밥에 돼지고기는 별루없고 고기도 딱딱 국물만! 이런 돼지
○○	설렁탕	3.0	
짱구	설렁탕	4.0	무척 맛있었어요 아쉬운 건 수저세트를 주지 않은거요
김민성	설렁탕	4.0	술먹은 다음날 해장하기 딱좋네요ㅜㅜ 자주 시켜먹고있습니다ㅎ
천생연분먹부림	설렁탕	3.0	소면추가.. 음.. 너무.. 맛은 평범
2738ker속자	설렁탕	3.0	
천공성라퓨터	설렁탕	2.0	
야마모토	설렁탕	2.0	고기냄새가 좀 역했어요... 제가 민감해서 그럴지 모르겠지만
곤쥬님	설렁탕	2.0	
공주	설렁탕	2.0	국은 맛있었는데 밥이 너무 질어서.. 거의 죽같아서 별로였어요ㅠ
지성인	설렁탕	1.0	고기가 잘게 부서져 있어요. 정육점에 남은 고기 사용한 것처럼
신희	설렁탕	1.0	리뷰 보고 시켰는데 그닥 맛이 좋지 않았음
			배달 기사님 연락도 안주시고 그냥 두고가셔서 음식 몇분이나 밖에
과연그렇까	설렁탕	1.0	음식은 고추가루 부었나봅니다. 콩나물국박이 먹고싶어지더라고
맛집을 찾아서	설렁탕	1.0	영업시간 마감전에 주문했는데 너무 성의 없는거 아닌가요 후기
			배달시간도 늦고 냉면다데기도 안오고 냉면 다불고 설렁탕 소금도
이에진	설렁탕	1.0	무슨맛인지 모르겠음 새우젓 양념장 다 넣었는데 아무맛도 안나
먹자꾸나	설렁탕	1.0	다시는 안 시켜먹을 거 같음;;;
오진성	설렁탕	1.0	배고파서 먹었네요 두번 다시 시켜먹지 않을래요. 캡사인 넣었나
오진성	설렁탕	1.0	음식맛은 좋은데 만안에서 수메시!?? 은박지?? 같은게 나왔어요.
넉넉	설렁탕	5.0	갈비만두 감사합니다~ 항상 맛있게 잘 먹고 있어요~~ ^^ 국물이
4025	설렁탕	5.0	매콤하고 맛있어요! 해장에 딱이네요:) 근데 여러 메뉴를 시키면
kimkim	설렁탕	5.0	설탕탕도 맛있네요 너무 맛있게 잘 먹었습니다.
rlad	설렁탕	5.0	국물진하고 건더기실하고 너무잘먹었습니다.

웹 크롤링 등을 통한 리뷰 데이터 수집 및 정규화

1. 리뷰 DB 구축

2. 텍스트 마이닝

1. 데이터 전처리

2. 토픽모델링(LDA)
상품 속성 추출

3. 속성별 리뷰 분류

3. 점포 추천

1. 속성 별 평점 부여

2. 속성별 점포 평점
매트릭스 구성

3. 점포 추천(RBM)

2. 텍스트 마이닝 - 데이터 전처리

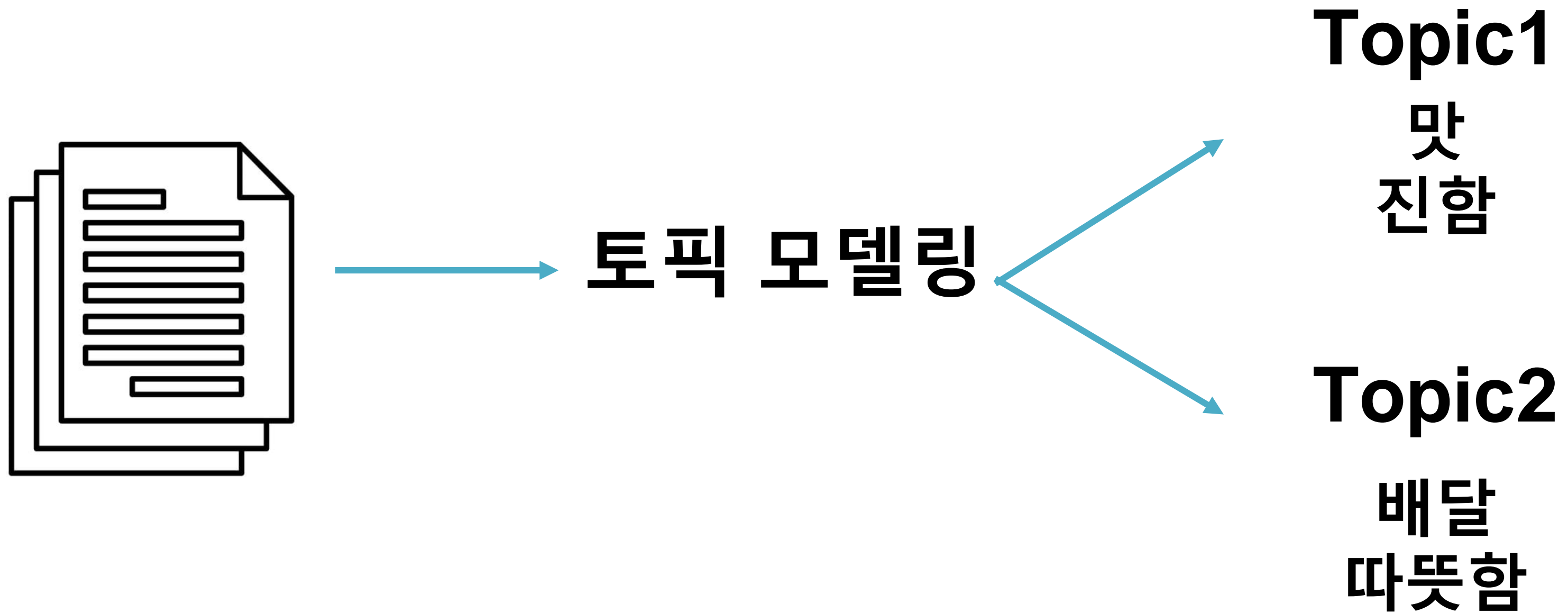
링링링	설렁탕	1.0	처음시켜봤는데 리뷰이벤트 참여 한다 했는데 무시를 하셨네요. 불쾌하네요
절대미식가	설렁탕	3.0	돼지국밥에 돼지고기는 별루없고 고기도 퍽퍽 국물만! 이런 돼지국밥 처음임 별 세개도 아까움!
??	설렁탕	3.0	
짱구	설렁탕	4.0	무척 맛있었어요 아쉬운 건 수저세트를 주지 않은거요
김민성	설렁탕	4.0	술먹은 다음날 해장하기 딱좋네요ㅌㅌ 자주 시켜먹고있습니다ㅎㅎ 설렁탕 존맛탱!



불쾌	무시	퍽퍽	맛	냄새
질음	리뷰	돼지	고기	새우젓
갭사인	국물	배달시간	배달	건더기

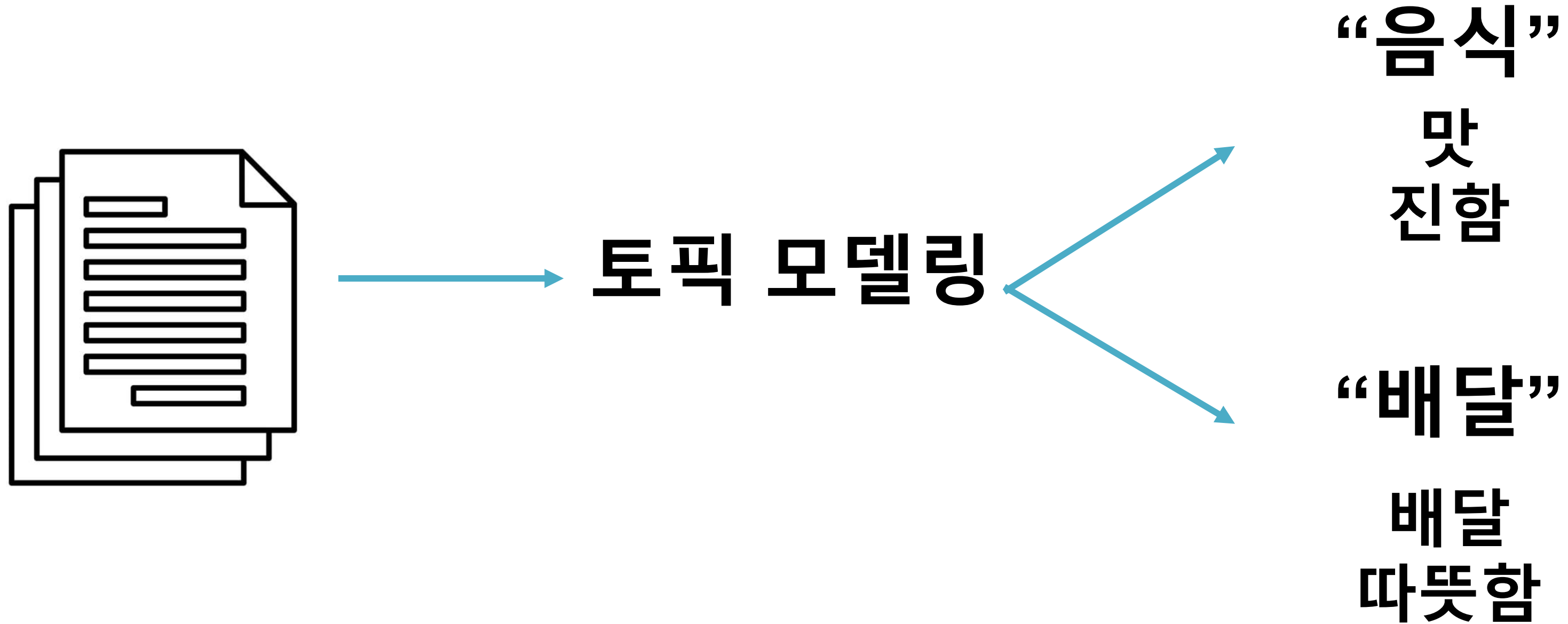
결측치 제외, 장난 리뷰 제외, 용어 추출, 불용어 처리

2. 텍스트 마이닝 – LDA 이용한 토픽 모델링



토픽 모델링을 통해 속성 추출

2. 텍스트 마이닝 – LDA 이용한 토픽 모델링



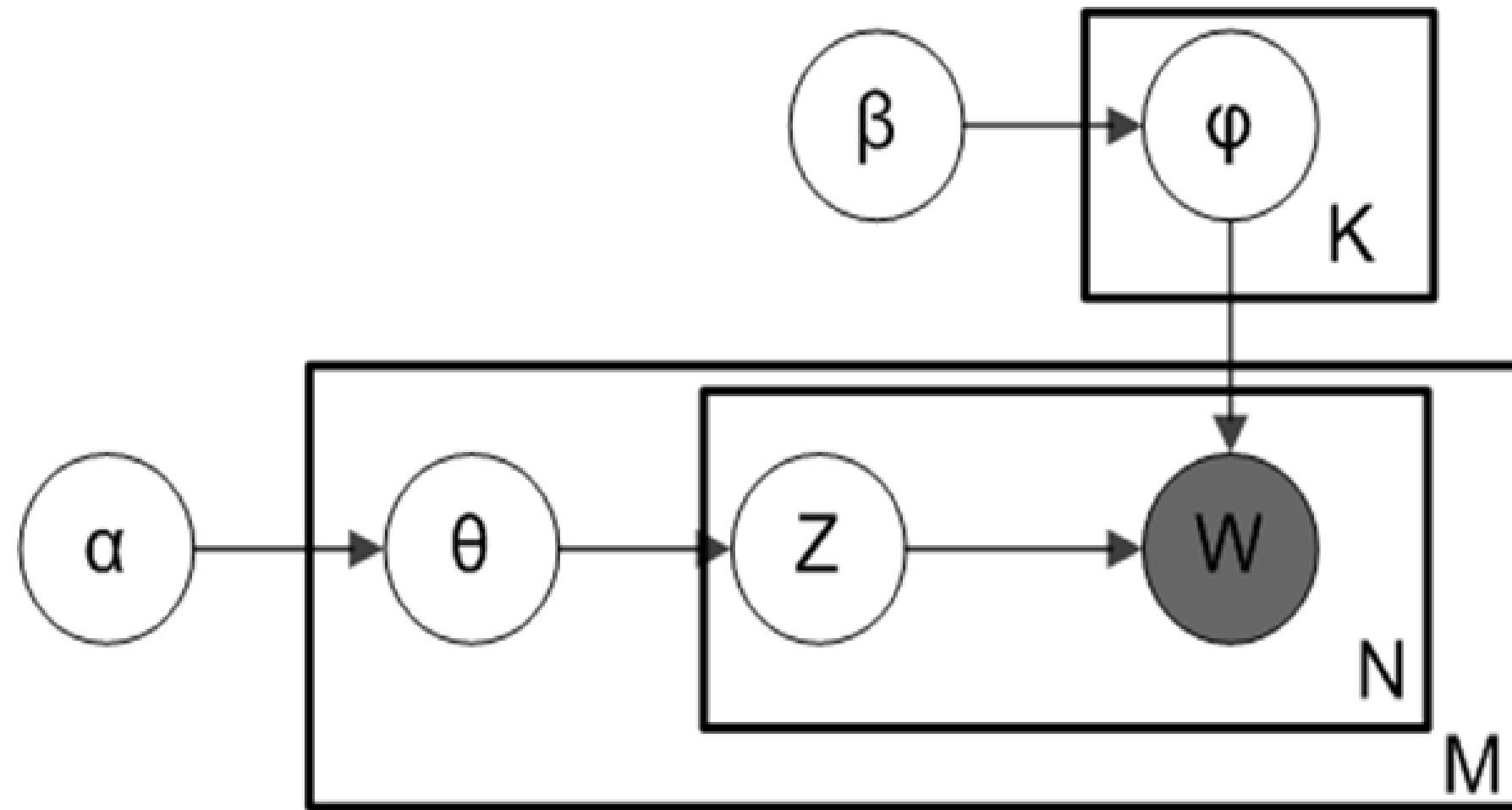
속성 키워드 네이밍

2. 텍스트 마이닝 – LDA 이용한 토픽 모델링

속성	키워드
음식(food)	sauce, pasta, meatball, veal, chicken
가격(price)	price, reasonable, fair, good, quality
음식/서비스(food/service)	good, food, service, choice, nice
서비스(service)	quick, service, friendly, fast, bite
위생(cleanliness)	floor, restaurant, clean, door, bathroom
음식/분위기/서비스 (food, atmosphere, service)	excellent, service, food, superb, ambience
위치(location)	street, walk, easy, station, find

속성 별 키워드 선정

2. 텍스트 마이닝 – LDA(Latent Dirichlet Allocation)



- α : 문서들의 토픽 분포를 얼마나 밀집되게 할 것인지에 대한 설정 값
- θ : 문서의 토픽 디리클레 분포
- z : 해당 단어가 속한 토픽의 번호
- w : 특정 문서의 n 번째 단어, 실제 관측 가능한 값
- ϕ : 주제의 단어
- β : 문서내 단어들의 토픽 분포를 얼마나 밀집되게 할 것인지에 대한 설정 값입니다.
- k : 토픽을 몇개로 설정할 것인지에 대한 설정 값으로 되어 있습니다.

2. 텍스트 마이닝 – LDA 과정

1. 모든 단어를 k개 중 하나의 토픽에 할당

	1번문서		2번문서				3번문서		
단어	문고리	거래	가방	나눔	문고리	드림	비대면	거래	택배
주제	topic1	topic2	topic1	topic1	topic2	topic2	topic3	topic2	topic3

2. 토픽-문서의 단어 분포 계산 (토픽 + α)

토픽-문서	1번문서	2번문서	3번문서
topic1	1.01	2.01	0.01
topic2	1.01	2.01	1.01
topic3	0.01	0.01	2.01

2. 텍스트 마이닝 – LDA 과정

3. 토픽-단어 분포 계산 (토픽 + β)

토픽-단어	문고리	거래	가방	나눔	드림	비대면	택배
topic1	1.001	0.001	1.001	1.001	0.001	0.001	0.001
topic2	1.001	2.001	0.001	0.001	1.001	0.001	0.001
topic3	0.001	0.001	0.001	0.001	0.001	1.001	1.001

4. 단어 하나 제외하고 토픽 단어, 문서의 분포 고정

1번문서				2번문서			3번문서		
단어	문고리	거래	가방	나눔	문고리	드림	비대면	거래	택배
주제	미분류	topic2	topic1	topic1	topic2	topic2	topic3	topic2	topic3

2. 텍스트 마이닝 – LDA 과정

5. 토픽 선정

1번 문서 내 topic1이 있을 확률 : $1.01/3.03 = 0.333$
(이 때 분모의 경우 1번 토픽 내 문서 분포 합입니다.)

1번 토픽 내 단어가 '문고리'일 확률 : $1.001/3.007 = 0.332$
(이 때 분모의 경우 1번 토픽 내 키워드 분포 합입니다.)

마지막으로 1번 문서의 '문고리'가 topic1일 확률 $0.333*0.332=0.110$ 이 됩니다.

위처럼 1번 문서 내 topic2이 있을 확률,
1번 문서 내 topic3이 있을 확률 모두 구해줍니다.

해당 과정을 모두 반복하면 가장 높은 확률을 가진 토픽에 해당 단어와 문서가 분류됨으로써 LDA학습은 완료가 됩니다.

6. 반복

2. 텍스트 마이닝 – 속성별로 리뷰 분류

전체 리뷰	속성	분류된 리뷰
Very small bistro nearby my hotel. excellent food and service. Typical French cooking. The owner is the host and is very attentive to his guests. Prices are reasonable.	음식 (food)	excellent food and service. Typical French cooking.
	분위기 (atmosphere)	
	서비스 (service)	excellent food and service. The owner is the host and is very attentive to his guests.
	가격 (price)	Prices are reasonable.
	위생 (cleanliness)	
	위치 (location)	Very small bistro nearby my hotel.



음식 (food)	분위기 (atmosphere)	서비스 (service)	가격 (price)	위생 (cleanliness)	위치 (location)
256,399	141,388	215,632	90,604	41,687	133,896

2. 텍스트 마이닝 – 속성별로 리뷰 분류

음식 (food)	분위기 (atmosphere)	서비스 (service)	가격 (price)	위생 (cleanliness)	위치 (location)
256,399	141,388	215,632	90,604	41,687	133,896



1. 리뷰 DB 구축

2. 텍스트 마이닝

1. 데이터 전처리

2. 토픽모델링(LDA)
상품 속성 추출

3. 속성별 리뷰 분류

3. 점포 추천

1. 속성 별 평점 부여

2. 속성별 점포 평점
매트릭스 구성

3. 점포 추천(RBM)

3. 점포 추천 - 속성별 평점 부여

속성	데이터 개수
음식	14,831
분위기, 음식	6,426
음식, 위치	7,036
음식, 가격	3,560
음식, 서비스	26,345
분위기, 음식, 위치	4,013
분위기, 음식, 서비스	28,910
음식, 위치, 서비스	21,612
음식, 가격, 서비스	13,620
분위기, 위생, 음식, 서비스	4,701
분위기, 음식, 위치, 서비스	34,248
분위기, 음식, 가격, 서비스	14,338
위생 음식, 위치, 서비스	3,593
음식, 위치, 가격, 서비스	13,123
분위기, 위생, 음식, 위치, 서비스	8,926
분위기, 위생, 음식, 가격, 서비스	2,985
분위기, 음식, 위치, 가격, 서비스	22,867
분위기, 위생, 음식, 위치, 가격, 서비스	8,167



속성	평점
음식	4.3
음식, 배달	4.5
음식, 위생	3.2
음식, 가격	3.4
음식, 양	4.2
음식, 배달, 위생	4.6
음식, 배달, 가격	3.4
음식, 배달, 양	4.2
음식, 배달, 위생, 가격	4.5

3. 점포 추천 - 속성별 점포 평점 매트릭스 구성

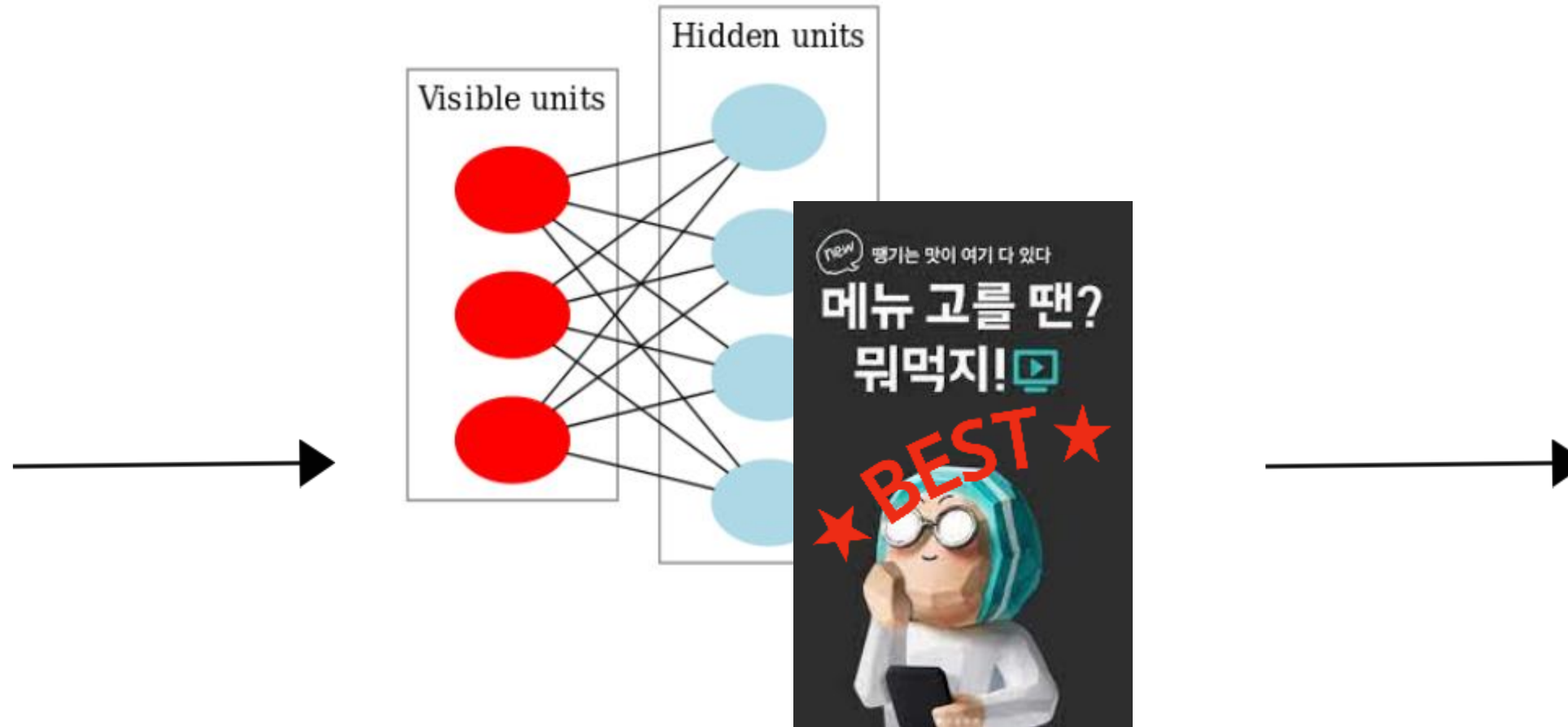
속성	평점
음식	4.3
음식, 배달	4.5
음식, 위생	3.2
음식, 가격	3.4
음식, 양	4.2
음식, 배달, 위생	4.6
음식, 배달, 가격	3.4
음식, 배달, 양	4.2
음식, 배달, 위생, 가격	4.5



	음식	위생	가격	양	배달
가게1	4.2	3.4	4.2	4.5	4.2
가게2	4.5	4.2	4.6	4.3	4.6
가게3	2.3	4.5	4.5	3.2	4.5
가게4	4.2	4.3	4.2	4.5	3.4
가게5	3.4	4.5	4.5	4.5	4.2
가게6	4.3	3.2	2.3	2.3	4.5

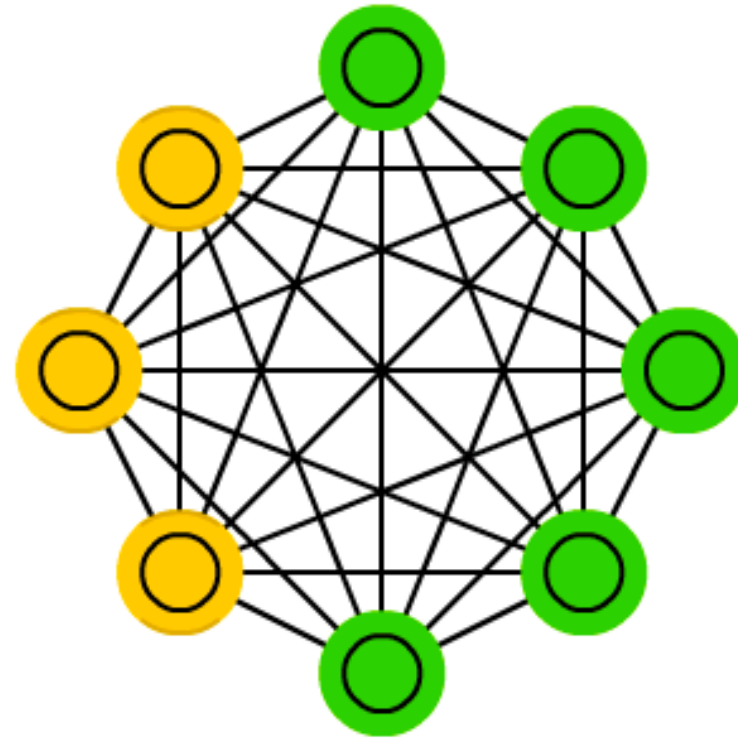
3. 점포 추천 - 추천모델 적용 (RBM)

	음식	위생	가격	양	배달
점포1	4.2	3.4	4.2	4.5	4.2
점포2	4.5	4.2	4.6	4.3	4.6
점포3	2.3	4.5	4.5	3.2	4.5
점포4	4.2	4.3	4.2	4.5	3.4
점포5	3.4	4.5	4.5	4.5	4.2
점포6	4.3	3.2	2.3	2.3	4.5

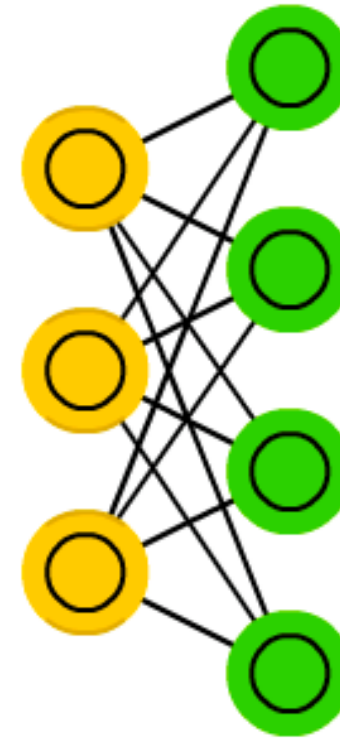


3. 제한된 볼츠만 머신(RBM: Restricted Boltzmann Machine)

Boltzmann Machine (BM)



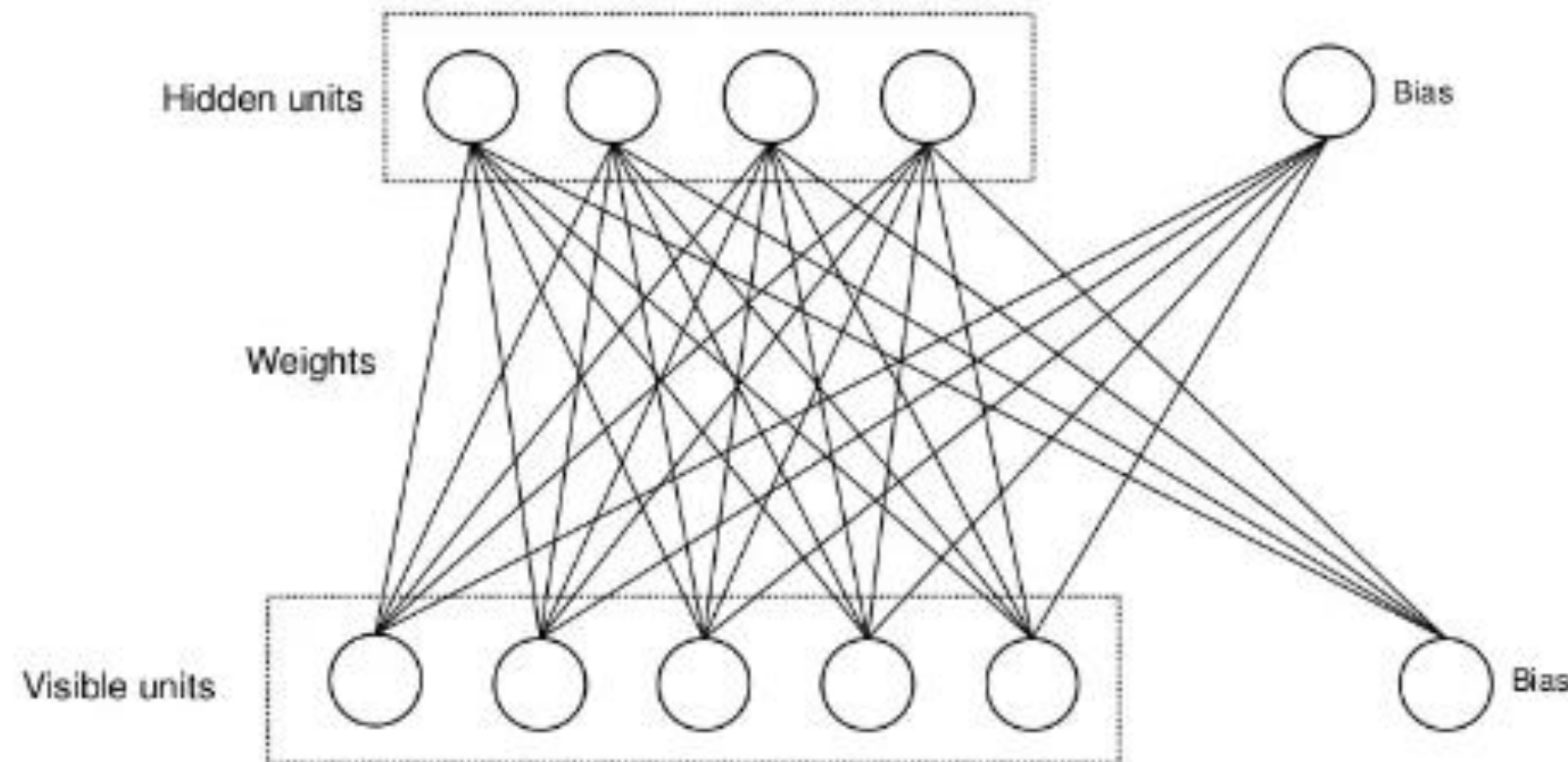
Restricted BM (RBM)



두 개의 층(입력층 1개, 은닉층 1개)으로 구성된 머신

3. RBM 모델을 사용하는 이유

RBM(제한된 볼츠만 머신, Restricted Boltzmann machine)은 차원 감소, 분류, [선형 회귀 분석](#), 협업 필터링(collaborative filtering), 특징값 학습(feature learning) 및 주제 모델링(topic modelling)에 사용할 수 있는 알고리즘



3. RBM 이용한 협업 필터링

협업 필터링이란?

협업 필터링(collaborative filtering)은 많은 **사용자**들로부터 얻은 기호정보(taste information)에 따라 사용자들의 관심사들을 자동적으로 예측하게 해주는 방법이다.

협력 필터링 접근법의 근본적인 가정은 사용자들의 과거의 경향이 미래에서도 그대로 유지 될 것이라는 전제에 있다.

이 시스템은 특정 사용자의 정보에만 국한 된 것이 아니라 많은 사용자들로부터 수집한 정보를 사용한다는 것이 특징이다.

고객들의 선호도와 관심 표현을 바탕으로 선호도, 관심에서 비슷한 패턴을 가진 고객들을 식별해 내 비슷한 취향을 가진 고객들에게 서로 아직 구매하지 않은 상품들은 교차 추천하거나 분류된 고객의 취향이나 생활 형태에 따라 관련 상품을 추천하는 형태의 서비스를 제공하기 위해 사용된다.

6. Future Perspectives

발전 계획

1. 기술 개발 필요성

- 속성별 리뷰 검색 불가능함
- 리뷰가 무질서하며 신뢰도가 낮음
- 소비자 리뷰 차단 오류가 발생함

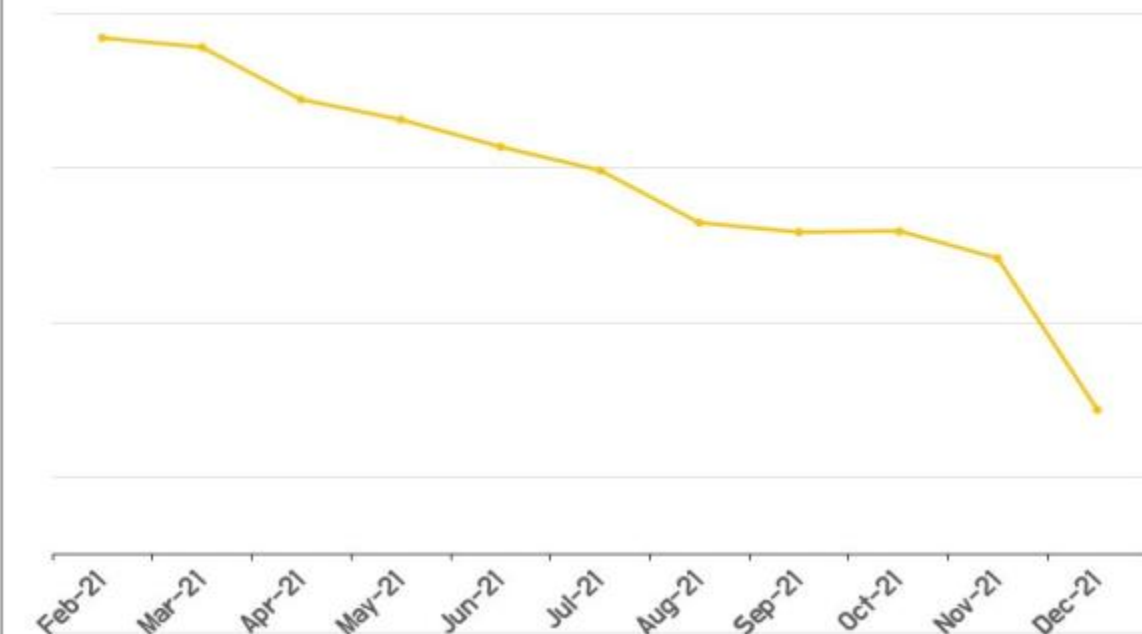
2. 목표 설정

- 리뷰의 신뢰도를 높이하고자 함
- 소비자에게 다속성별 추천 시스템을 제공하고자 함
- 허위 리뷰 오류를 줄이고자 함

회원님은 허위리뷰를 작성하신 것으로
의심되어, 리뷰쓰기 권한이
차단되었습니다. 차단을 해제하려면
고객센터로 연락주시기 바랍니다.

확인

2021년 월별 자전거래 탐지 현황



3. 사업화 방안

- LDA를 이용하여 속성별 모델링을 통해 더욱 정확한 리뷰 분석 모델을 만든다
- 무분별한 리뷰 차단으로 인한 소비자 불만 낮추기 위해 활용한다
- 배달의 민족뿐만 아니라 다른 사업에도 접목 가능하다
- 기업에게 기술을 제공해주고, 일정 사용료를 지급받는다



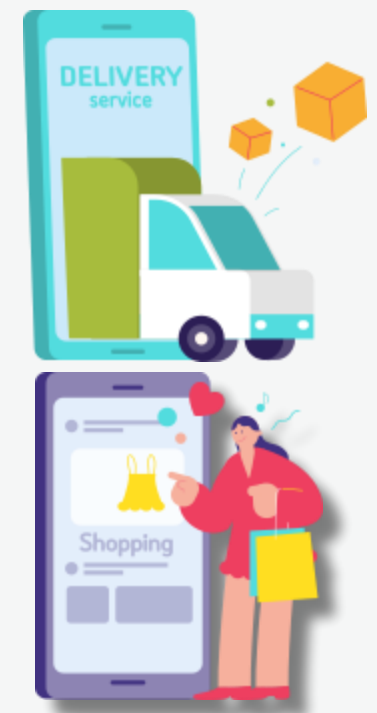
Imagine AI

다속성별 추천시스템개발자



배달의 민족

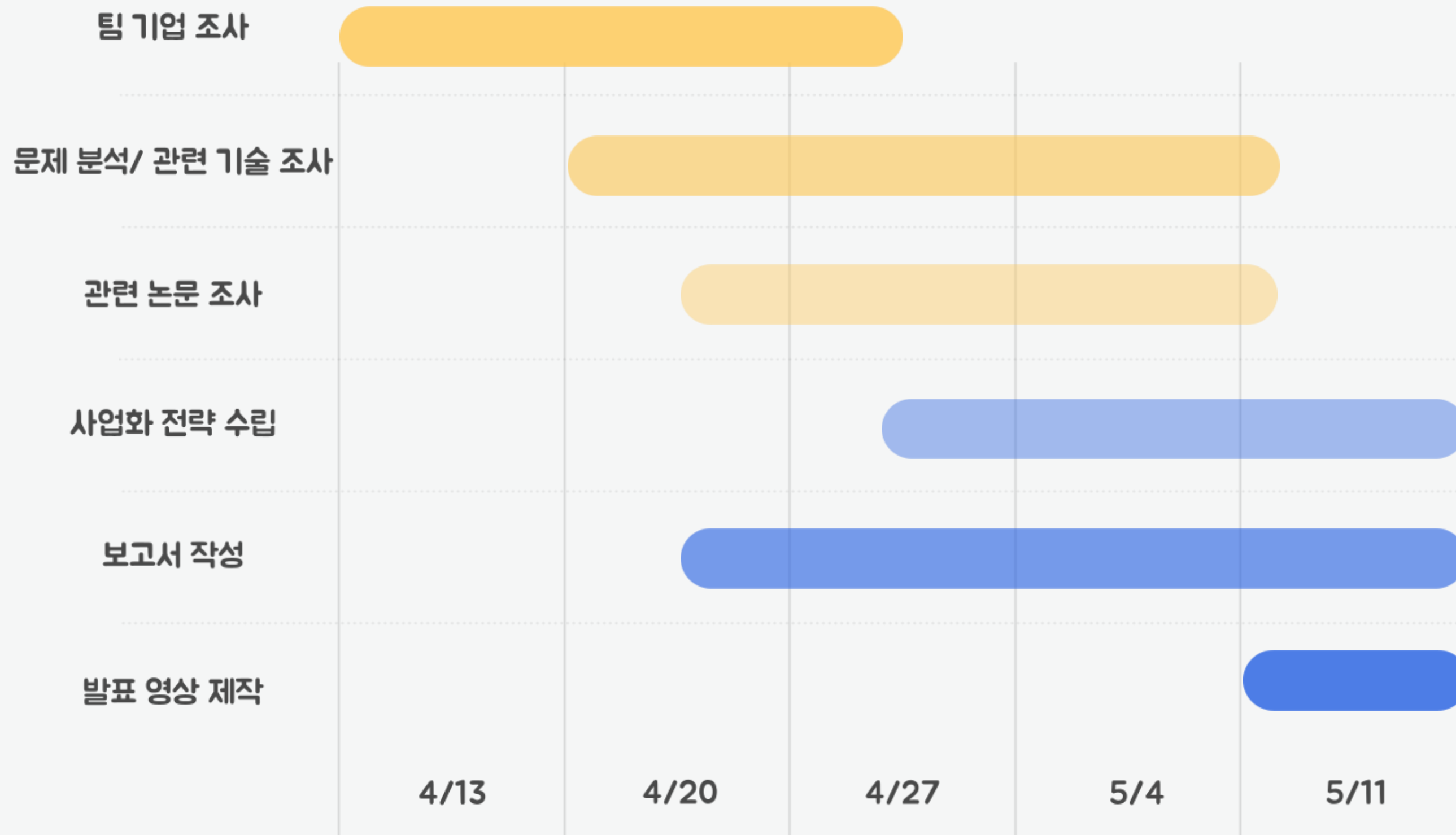
1. 허위리뷰 감별 오류 줄임
2. 다속성별 상품 추천 시스템 제공



다른 사업체에 접목

인터넷 쇼핑몰 혹은 다른 음식 앱에도 접목 가능

7. 일정 계획표



□ 2022년 5월 18일

□ 인공지능 실습 ■ Imagine AI



발표를 들어주셔서

감사합니다 :))

질문이 있으시다면 언제든지 말씀해주세요!
감사합니다.