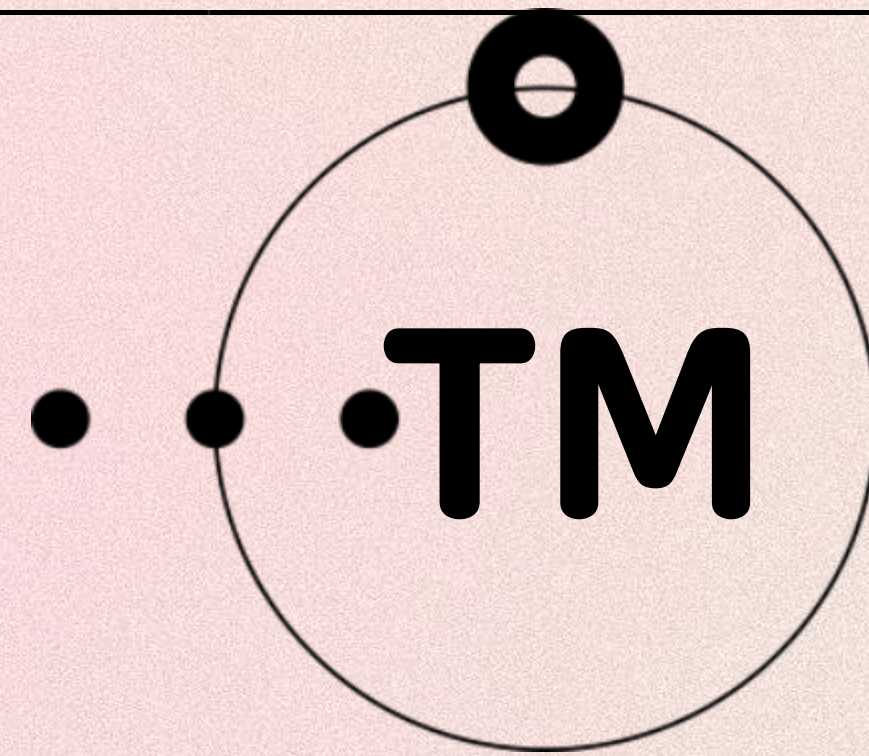


POINTER TEST

POINTER TEST

2022.05.30

팀 프로젝트 제안발표



< TM-WDM – Topic modeling , we do modeling >

2018204009 김기수

2018204072 김재윤

2018204077 김태영

Presentation INDEX

01

주제 선정 및 배경

주제를 소개하고, 주제를 선정하게 된 배경을 설명합니다.

02

전반적인 프로젝트 소개

이 프로젝트에 대한 설명을 하고, 소개합니다.

03

실행 계획

이 프로젝트를 하는 과정을 도식화하여 나타냅니다.

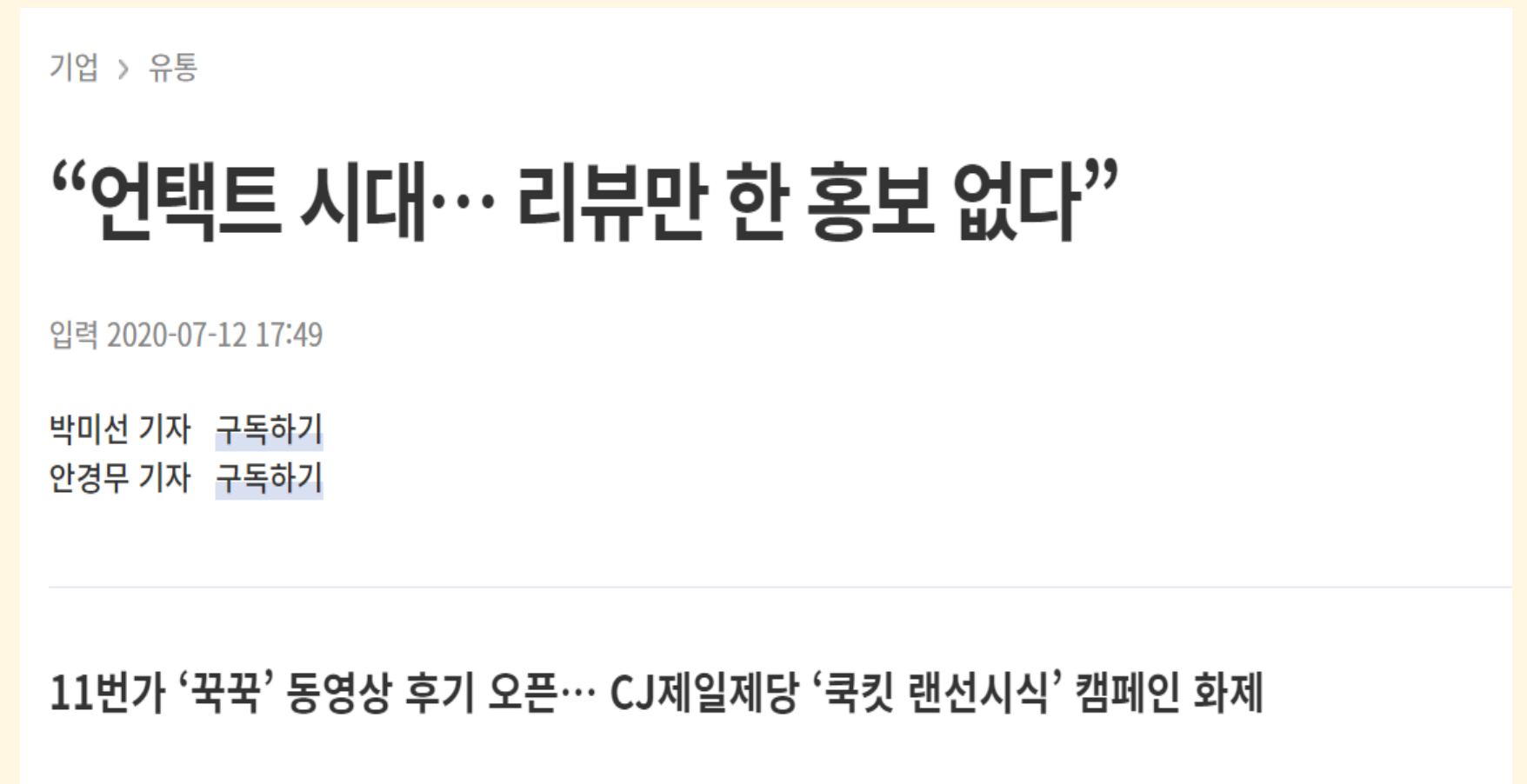
04

기대효과

이 프로젝트로 인해 나타나는 기대효과에 대해 설명합니다.



E커머스, 배달, 언택트 관련 앱 증가



AI기반으로 한 마케팅 방식의 활성화

온라인 서비스의 간편함 경험과 확산
⇒ 기업의 리뷰를 활용한 마케팅 전략의 중요성 대두

대학생 58.8% '나 혼자 산다'

※남녀 대학생 1,048명 조사 / 자료제공:잡코리아X알바몬

나 혼자산다 58.8%

Q. 혼자 사는 이유

학교와 집이 멀어서 (59.3%)

혼자 살고 싶어서 (33.6%)

기타
(7.1%)



JOBKOREA x albamon

그래픽=잡코리아&알바몬

- 2020년 기준 약 60%가 자취를 하고 있음
=> 코로나 종식에 따라 비율은 늘어날 것으로 예상
- 자취용품과 관련한 구매 수요 증가 예상
=> 기존 제품에 대한 단점 보완 및 장점 극대화가 필요

자취용품을 판매하는 사이트인
'오늘의 집' 리뷰 평점별 토픽 모델링 분석

[2020년 2월 기준]

전반적인 프로젝트 소개

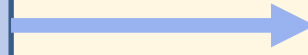
실행 계획

텍스트 수집

전반적인 프로젝트 소개

실행 계획

텍스트 수집



노이즈 캔슬링

전반적인 프로젝트 소개

실행 계획

텍스트 수집

노이즈 캔슬링

토크나이징

전반적인 프로젝트 소개

실행 계획

텍스트 수집

노이즈 캔슬링

토큰나이징

텍스트 벡터화

전반적인 프로젝트 소개

실행 계획

텍스트 수집

노이즈 캔슬링

토큰나이징

텍스트 벡터화

감성 분석

전반적인 프로젝트 소개

실행 계획

텍스트 수집

노이즈 캔슬링

토큰나이징

텍스트 벡터화

텍스트 추출

감성 분석

전반적인 프로젝트 소개

실행 계획

텍스트 수집

노이즈 캔슬링

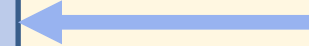
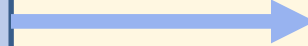
토큰나이징

텍스트 벡터화

토픽 모델링

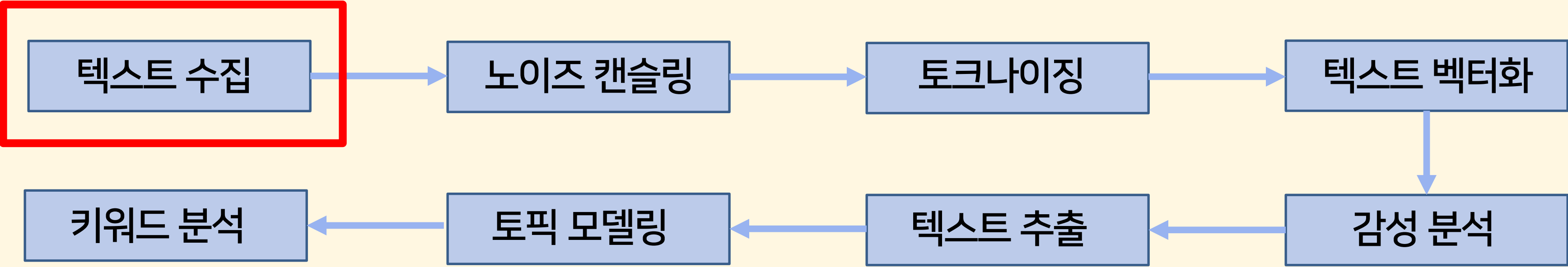
텍스트 추출

감성 분석



전반적인 프로젝트 소개

실행 계획



```
import re
import time

count = 0
#print(len(review_box_list))

html = driver.page_source
soup = BeautifulSoup(html, "html.parser")
class_list = driver.find_element_by_id("product-651554")

start = time.time()
try:
    #name = review_box.find('span', class_ = 'production-item__header__brand').text
    #time.sleep(1)
    #header_name = review_box.find('span', class_ = 'production-item__header__name').text
    #time.sleep(1)
    #price = review_box.find('span', class_ = 'production-item-price__price').text
    #time.sleep(1)
    #rate = review_box.find('strong', class_ = 'avg').text
    #time.sleep(1)
    class_list.click()
    time.sleep(3)
    print(1)
except:
    print(2)
    pass

try:
    button_css = 'body > div.page > div > div > div.production-selling > div.sticky-container.production-selling-navigation__wrap > div > nav > ol > li:nth-child(2) > a'
    review_button = driver.find_element(by = By.CSS_SELECTOR, value = button_css)
    review_button.click()
    time.sleep(3)
    print(1)
except:
    print(2)
    pass
while True:
    if count % 50 == 0:
        item_name.to_csv('item_data.csv')
        print(len(item_name))
    try:
        review_container = driver.find_elements(by = By.CLASS_NAME, value = 'production-review-item__container')
        for k in range(len(review_container)):
            count += 1
            review_xpath = '/html/body/div[1]/div/div/div[2]/div[5]/div/div[1]/div/section[2]/div/div[3]/div[' + str(k+1) + '] /article/p'
            score_xpath = '/html/body/div[1]/div/div/div[2]/div[5]/div/div[1]/div/section[2]/div/div[3]/div[' + str(k+1) + '] /article/div[1]/div/button/span'
            production_xpath = '/html/body/div[1]/div/div/div[2]/div[5]/div/div[1]/div/section[2]/div/div[3]/div[' + str(k+1) + '] /article/div[3]'
            review_comment = driver.find_element_by_xpath(review_xpath).text
            score = re.sub(r'[^0-9]', '', driver.find_element_by_xpath(score_xpath).get_attribute('aria-label'))
            product = driver.find_element_by_xpath(production_xpath).text
            #review_text = driver.find_element_by_class_name('production-review-item__description').text
            #review_writer = driver.find_element_by_class_name('production-review-item__writer__info__name').text
            #review_score = soup.find('span', class_ = 'production-review-item__writer__info__total-star').get_attribute('aria-label')
            #review_detail_item = driver.find_element_by_class_name('production-review-item__description').text
            print('(0)번째 리뷰의 항목 : (1), 별점 : (2), 리뷰의 내용 : (3)'.format(count, product, score, review_comment))
```

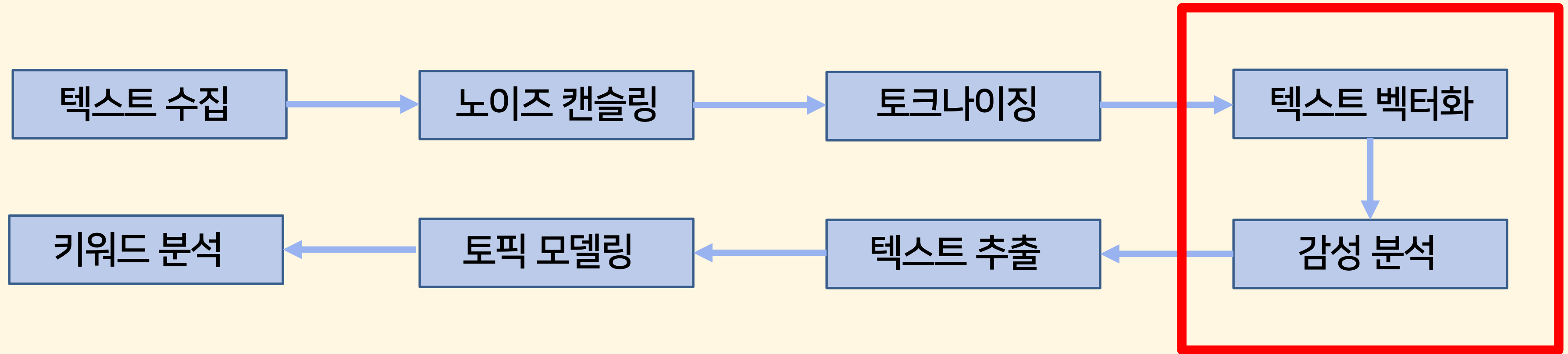
In [23]:

```
df = pd.concat([a, b, c, d, e, f, g]).reset_index().drop(['index'], axis = 1)
df.drop('Unnamed: 0', axis = 1, inplace = True)
df
```

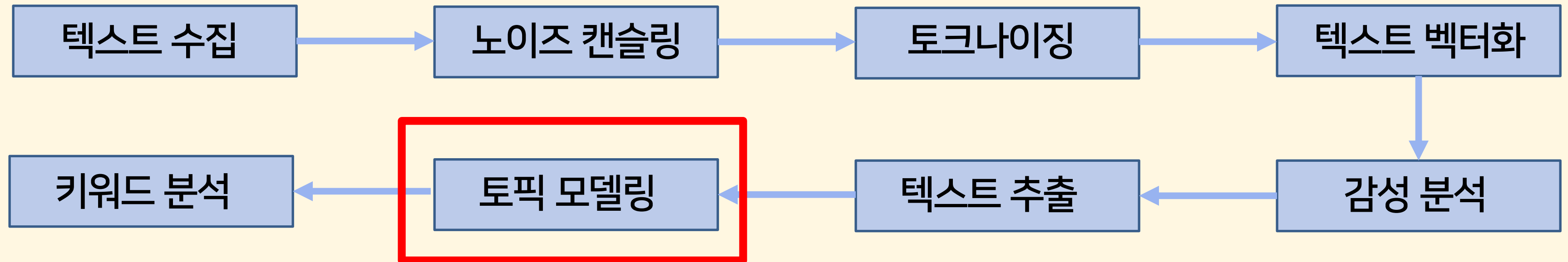
Out[23]:

	product_name	rates	review
0	[5%쿠폰] 어반 아치형 전신거울 3size(거치대 2개증정)\n500 (화이트/메...	5.0	급하지 않아서 배송 ㅏㅏ왔어요 급한사람들은 느리다고 생각할 듯 크기도 마음에 들고 ...
1	블랑 좌식 접이식 거실테이블 3size(보호캡 증정)\n화이트 XL 600X1200...	5.0	여섯명이 써도 널널하고 그렇게 무겁지도 않고 예뻐요 집들이용으로 딱입니다
2	매거진랙 3단 라운드 북선반 2colors\n라운드_오크	4.8	맘에 들어요:) 목재를 사서 만들려고 했는데 자재가격이랑 이 제품가격이랑 비슷하더라...
3	블랑 좌식 접이식 거실테이블 3size(보호캡 증정)\n화이트 L 600X800 (...	4.5	면지도 너무 많이 붙고 변색되면 안당여요 ㅠㅠ
4	HOL 사이드테이블 타공판 협탁 2colors\n메이플	5.0	뒤에 판을 뺄 수 있어서 콘센트를 가렸어요 ㅎㅎ 이게 넘 맘에 드네요 그러구 예뻐욧
...
210249	[7%쿠폰]LEITER 선반 4단 2size\n그레이	4.8	한쪽 면이 찢고려져 왔지만 그래도 좋았습니다
210250	[7%쿠폰]LEITER 선반 4단 2size\n그레이	4.8	그레이랑 화이트 컬러 샀는데 그레이가 생각보다 톤 다운된 그레이라 너무 이뻐요! 화...
210251	[7%쿠폰]LEITER 선반 4단 2size\n매트실버	4.0	잘쓰고 있어요. 오랜만에 어플키니 리부 남겨라하네요 ㅋㅋㅋ 좋아요
210252	[7%쿠폰]LEITER 선반 4단 2size\n화이트	4.8	올려놓는 선반 부분이 힘이없다보니까 가벼운 물건 올려놓으면 지탱하는게 위태위태해보이...
210253	[7%쿠폰]LEITER 선반 4단 2size\n화이트	4.5	깔끔한 디자인으로 어디에 놓아도 이뻐요 근데 밑에 받침? 끼워넣는게 힘이 오지게 필...

210254 rows × 3 columns



- 사람의 리뷰 분위기와 평점 간의 차이 조절
- 학습 데이터 : **Naver Sentiment movie corpus**



- 각 평점별 중요 키워드 탐색
 - 더 좋은 제품을 만들 수 있는 부분에 기여할 수 있을 것이라고 기대
-

1. 소비자의 감정 반영한 리뷰 분석

- 리뷰와 평점의 일치 여부를 판단함으로써, 리뷰의 신뢰성을 높일 수 있음.

2. 소비자리뷰 분석을 통한 제품개선에 기여

- 각 평점 별로 핵심이 되는 키워드를 통해, 서로 간의 비교가 가능하며, 각 제품별 장단점 파악 가능
또한, 토픽 모델링으로 소비자의 니즈를 한 눈에 파악할 수 있으므로, 빠른 시일 내에 문제점을 파악할 수 있음

3. 향후 추천 시스템으로의 연결

- 언택트 시대를 거치며 초개인화 시대로 진입으로 개인에 맞춘 마케팅이 중요해졌음.

구축한 데이터를 이용하여, 협업 필터링(Collaborative Filtering)과 콘텐츠 기반 필터링(Contents-based Filtering)을 활용한 추천 시스템 구현이 가능.

2022.05.30

THANK YOU!

TEXT MINING 3조

궁금한 점이 있으시다면 질문 부탁드립니다.

감사합니다.
