

# 목차



## 01. 지난주 피드백

지난주 피드백



## 02. 추천 기준 정의

설문조사  
상황/음식  
특별함 점수/라벨 색 추출



## 03. 유사도 행렬

리뷰 임베딩 & pca  
코사인 유사도  
피어슨 상관계수



## 04. 추천시스템

추천시스템 원리  
추천시스템 구현



## 05. 한계 및 의의

한계/의의  
소감

발표 흐름 체크!

맥주 추천 ?!

## 콘텐츠 기반 추천 시스템

: 콘텐츠 내용을 기반으로 콘텐츠를 추천!



### 02. 추천 기준 정의

: 어떤 기준으로 추천해줄까?

음식에 따라 추천하는 건 어때?

기분에 따라 추천하는 건?

또 다른 기준은 없을까?





## 추천기준 정의

설문조사

# 특정 상황과 가장 어울리는 맛의 맥주는?

토픽모델링의 결과인  
맥주 스타일별로 대표되는 맛을 활용하면  
상황과 맛의 관계를 파악할 수 있지 않을까?



# 추천기준 정의

설문조사

"저번 주량은 뭐가 다른 거야?"

토픽 모델링?

: 문서의 집합에서 토픽을 찾아내는 프로세스



각 문서의 집합에서 숨겨진 주제를 찾아내  
주제끼리 키워드별로 묶어주는 비지도학습

1. 맛을 표현하는 단어 추출 위해

stop\_words에서 색깔을 표현하는 단어를 추가  
(‘white’, ‘golden’)

2. Perplexity와 Topic Coherence로

파라미터 튜닝을 실시해 모델 성능 향상



# 추천기준 정의

설문조사\_유의성 검증

“기분에 따라 맥주의 맛과 도수를 선택하는 사람이  
그렇지 않은 사람보다 더 많을까?”



“맛과 도수 선택의 차이가  
사람들의 상황과 연관이 있는 걸까?”



# 추천기준 정의

설문조사\_유의성 검증

"기분에 따라 맥주의 맛과 도수를 선택하는 사람이  
그렇지 않은 사람보다 더 많을까?"

모비율 Z-test

&

RC table 독립성 검정

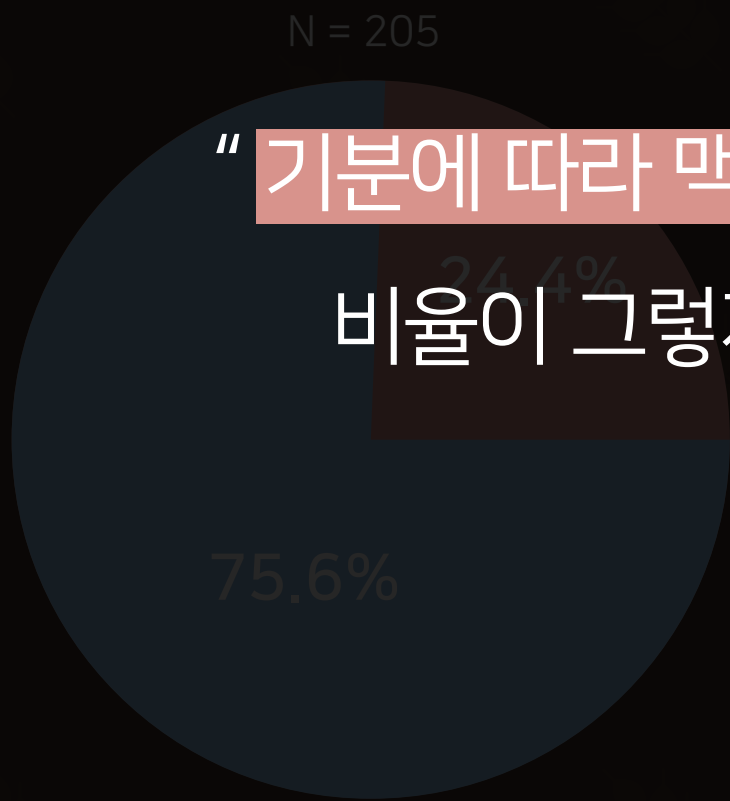
을 통해 설문조사의 결과를 해석해 보자!

## 추천기준 정의

설문조사\_유의성 검증

### 모비올 z-test

Q. 맥주의 도수를 선택할 때, 위의 상황이 영향을 미쳤나요?



“기분에 따라 맥주의 맛과 도수를 선택하는 사람의 비율이 그렇지 않은 사람의 비율보다 크다!”

```
***{r}  
prop.test(x = 155, n = 205, p = 0.5, alternative = 'greater')  
***  
data: 155 out of 205, null probability 0.5  
p-value = 0.0007613  
alternative hypothesis: true p is greater than 0.5
```

$$H_0 : P = 0.5$$

맥주의 도수를 선택할 때 상황은 영향을 미치지 않는다

기각



# 추천기준 정의

상황에 따라

## < 상황에 따른 맥주 >

	상황 1	상황 2	상황 3
에일	기쁜 날	마음이 우울한 날	몸이 지치는 날
라거	몸이 지치는 날		
스타우트	마음이 우울한 날		
포터	기쁜 날	몸이 지치는 날	
밀맥주	몸이 지치는 날		
바이젠	기쁜 날		
필스너	마음이 우울한 날		
람빅	기쁜 날	몸이 지치는 날	
퀵쉬	몸이 지치는 날	기쁜 날	마음이 우울한 날
와인	마음이 우울한 날		



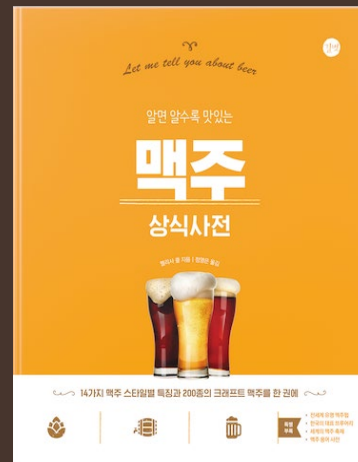


# 추천기준 정의

음식에 따라

## < 음식에 따른 맥주 추천 >

종류	음식 종류
라거	먹태, 쥐포, 치킨, 육류
퀵쉬	치킨, 피자, 족발
필스너	새우감바스, 스파게티
스타우트	굴, 홍합찜, 해산물
람빅	회
포터	초콜릿, 컵케이크와 같은 달콤한 디저트
에일	치즈, 훈제요리(풍미가 강한음식)
와인(스트롱에일)	견과류와, 건조과일
밀 맥주	오렌지,사과, 포도, 빵
바이젠	치즈, 바게트, 빵



‘맥주 상식사전’을 참고해  
각 맥주에 어울리는 음식 종류를 매칭 해줬어



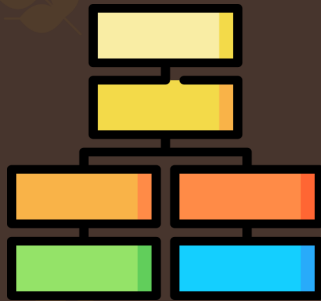
# 추천기준 정의

특별함의 정도

< 특이함 점수 생성에 반영된 변수 >



제조 국가 ↓



스타일 ↓



쓴 맛 정도 ↑



알코올 도수 ↑

특이함 점수 ↑↑





유사도 행렬

리뷰데이터 임베딩

## How to use MSS?

각 리뷰 당 추출한 tri-gram의 결과를  
순서대로 붙여 마치 문장처럼 활용



Ex. < 트롤브루 자몽 >

'grapefruit slightly sour sweet grapefruit background red grapefruit hazy'

이를 통째로 임베딩 해야겠는걸?



## < Word2Vec >

등장하는 단어의 종류 & 등장 빈도만으로 충분히 유의미

3

언어 모델을 활용한 임베딩

• Word2Vec

### Summary

Word2Vec의 학습은 행렬 연산이 전부 !

두 단어 벡터의 내적을 Sigmoid를 통해 예측 값을 산출

두 단어의 유사한 정도를 이용해

주변 단어인지 아닌지 예측하는 모델

2020 2학기 딥러닝팀 교안참고

관희 'Simple is Best'라는 말이 있는게 아니야!





## 추천시스템

추천시스템 원리\_Clustering

### Gaussian Mixture Model

가우시안 분포가 혼합되어 있다고 가정하는 EM 알고리즘 기반 방법론

< 실루엣 계수 0.2 / 클러스터 개수 4 >

타원형 클러스터에 잘 작동하는 경향이 있는데  
우리의 데이터는 타원형이 아니라 부적합

### BIRCH Clustering

계층적 방법으로 대용량 데이터를 빠르게 클러스터링

< 실루엣 계수 0.05 / 클러스터 개수 3 >

사용자에게 입력받는 하이퍼 파라미터의 수가 많아  
파라미터 튜닝의 어려움으로 인해 사용X



## < How to Make Spectral Cluster? >

KNN 분류를 활용해보자~

1. 유사도 행렬 만들기
2. 라플라시안 행렬 만들기
3. 라플라시안 행렬의  
고유값 분해
4. 클러스터링

나뉘지는 군집의 크기를  
균등하게 하기 위해!



## 유저 프로필 구축

Q. 어떤 기준으로 맥주를 추천 받고 싶나요?

오늘의 맥주

추천받고 싶은 기준을 골라주세요

선택

1. 음식

2. 기분

© P-SAT 2021 최강데마

음식 데이터

호출



오늘의 맥주

함께 먹을 음식을 선택해주세요!

마른안주

빵, 치즈류

풍미가 강한 음식

달콤한 디저트

과일안주

회

© P-SAT 2021 최강데마