# 목차



## 주제 선정배경

수제맥주, 수입맥주 등
다양한 맥주에 대한 소비자의 니즈 증가

이에 대응해 편의점, 식당 등 맥주 소매 업체에서는

국내맥주 4캔 만 원, 수입맥주 3캔 만 원 등 다양한 맥주 소비 트렌드에 맞춰
소비자에게 맥주의 다양성에 대한 옵션을 제공하기 시작

여러가지 맥주가 존재하고,
또 그것을 구매할 수 있다는 **다양성에 대한 선택권**이

소비자에게 주어진 것은 긍정적이다!

그러나 이러한 장점은 종종

**"어떤 기준을 두고 맥주를 골라야 하는지에 대한 결정을 어렵게 하는"**

문제를 야기하곤 한다.

분명 다음과 같은 상황을 마주한 적이 있을 것이다...

이진모

특이 사항: 아는 척 잘 함

# 데이터 수집

데이터 설명

## 컨텐츠 기반 추천 시스템 구현

맥주 자체에 대한 정밀 분석

설문을 통한 유저 프로필 구축

### 1. 맥주 리뷰 데이터

- 맥주의 맛/색깔과 같은 각 요소들은 어떤 감성과 맞닿아 있나?

- '단어' 단위에서의 감성 분석, 그리고 '리뷰' 단위에서의 감성 분석

### 2. 맥주 메타 데이터

-맥주의 도수를 나타내는 ABV 지수

-맥주의 쓴 맛을 나타내는 IBU 지수

-맥주캔 (또는 병)의 형태를 표현한 이미지 데이터

# 데이터 가공

데이터 설명

## 컨텐츠 기반 추천 시스템 구현

맥주 자체에 대한 정밀 분석 ⇒ 맥주 리뷰 데이터

자연어처리 끝장내자

### 리뷰 데이터 요약

**Keyword Extraction:**

BERT를 활용한 tri-gram 데이터 셋 생성

**Topic Modeling:**

LDA 방식을 활용한

유사 semantic cluster 생성

### 감성 분석

**단어(unigram)단위:**

Corpus 기반의 긍/부정어 추출 &

Transfer Learning을 이용한 ML approach

**리뷰 단위:**

긍/부정 리뷰 분류

## 감성분석?

: 텍스트에 들어있는 의견이나 감성, 평가, 태도 등의
주관적인 정보를 컴퓨터를 통해 분석하는 과정

⇊

N개의 단어 뭉치 단위로 끊어
각각을 하나의 토큰으로 간주하는 N-gram 모델 사용해 자연어 처리

맥주 리뷰에 맞는 감성 사전을 구축할 순 없을까?

⇓

## 머신 러닝 기반 감성 분석

# 전이학습
# (Transfer-Learning)

: 데이터로 충분히 훈련된 모델을 가져와

기존 데이터에 맞게 적합 시키는 방법

⇒ 적은 양의 데이터로도 학습이 가능함

# 데이터 가공

Bert (N-gram)

## BERT

: BERT는 '사전 훈련 언어모델' 중 하나로

문서 요약이나 키워드 추출에서 좋은 성능을 보이는 Sentence BERT가

다양하면서도 정확도가 높은  n-gram 키워드 군을 추출하려는

우리의 목표에 적합해 이를 활용

# 데이터 가공

Bert (N-gram)

코사인 유사도를 사용하는 방법으로,
MMR은 반복적으로 일어나는 일련의 과정
MSS는 필터링하듯 단계 by 단계로 진행

## < Bert기반 3-gram 적용 결과 >

| | name_kor_x | mss | mmr |
|---|---|---|---|
| 1 | 트롤브루 자몽 | ['grapefruit slightly sour', 'sweet grapefruit background', 'red grapefruit hazy'] | ['juicy grapefruit grapefruit', 'weizen 500ml supermarket', 'bread spicy weizen'] |
| 2 | 트롤브루 레몬 | ['taste moderately sweet', 'sugar slightly refreshing', 'lemon sugar grains'] | ['lemonade flavors sweet', 'yellowish foam aroma', 'lemon lime malts'] |
| 3 | 스팀 브루,<br>임페리얼 스타우트 | ['roasted aroma caramel', 'taste sweet malts', 'hint chocolate roasted'] | ['raisins slightly chocolate', 'beer trends heavy', 'light toasty coffee'] |

⋮

# EDA

아프라키, 남미, 아시아는 맥주 종류가 많지 않다

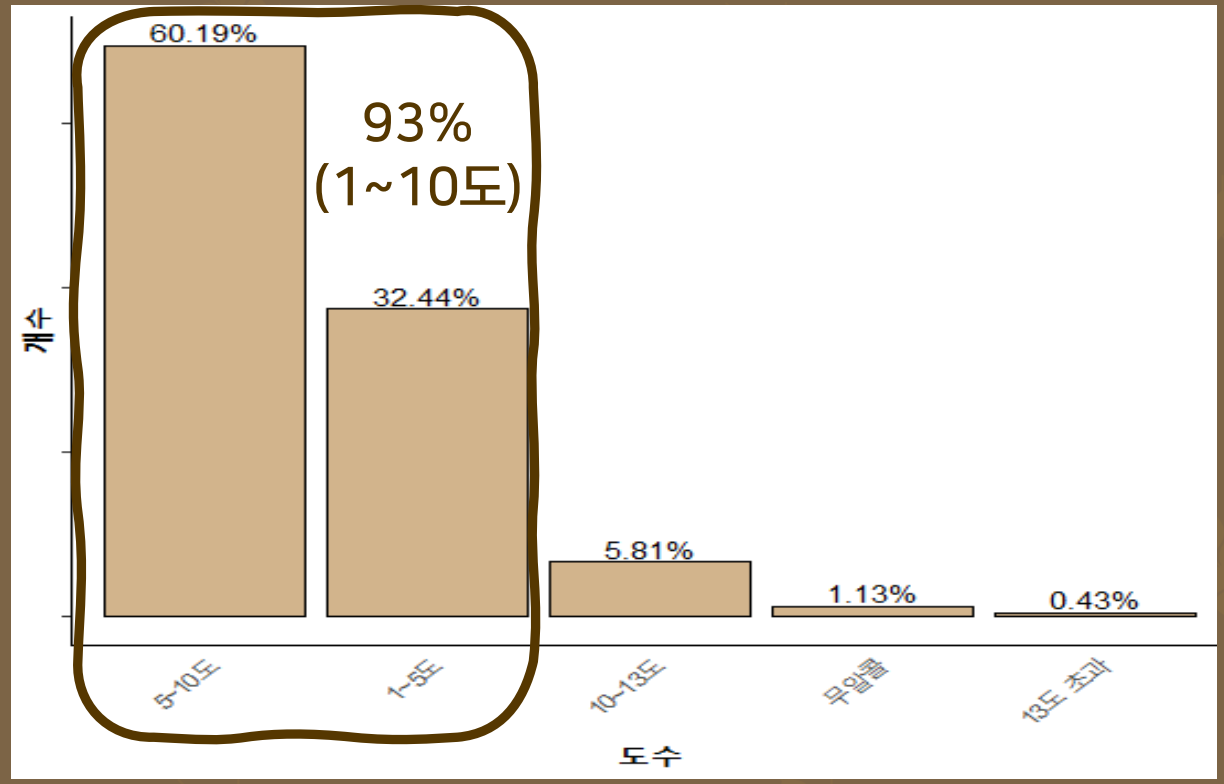
Ex. 이색적인 맥주를 원하는 사용자들에게
남미의 맥주를 추천해줘볼까?


[ 대륙 별 맥주 종류 비율 ]

# EDA

도수별 맥주 비율

## [ 도수별 맥주 비율 ]



# EDA

Topic Modeling

## 토픽 모델링?

: 문서의 집합에서 토픽을 찾아내는 프로세스

→ 각 문서의 집합에서 숨겨진 주제를 찾아내
주제끼리 키워드별로 묶어주는 비지도학습

**Latent**
**Dirichlet**
**Allocation**

# EDA

Topic Modeling

## LDA

: 문서가 토픽들의 혼합으로 구성되며,
이 토픽들은 확률분포에 기반해 단어를 생성한다고 가정

⇒ 주어진 데이터(문서)에 대해
각각의 토픽 내의 단어 분포를 추정한다.

# Topic Modelling | 토픽 모델링







맥주의 종류별로 토픽모델링을 진행하면
각 종류에 따른 특징을 찾아낼 수 있지 않을까?

대표적인 종류인 에일, 라거, 스타우트에 대해
토픽 모델링을 적용해보자!

# 토픽 모델링 결과 평가 지표

1. Perplexity
2. Topic Coherence