TF-IDF

궁금?

사용자 유저 별 장소 별

컨텐츠 기반을

사용자에게 우리가 가지고 있는 속성을

900개의 관광지와 숙박, 음식점을 나눈 후 인기순으로 나누고 살펴보기

음식, 카페, 관광지, 숙소,쇼핑

1단계 관광, 숙소, 음식

2단계 -> 관광, 음식 나누는 걸로

1. 관광

* midclass , 구석구석 태그로 분류
* 5개 분류로 나누고 tf-idf
* -> 로지스틱 회귀 사용해서 분류가 정확하게 정확도 검사 가능 -> 지금은 불가능 하지만 나중에 확장할때

1. 음식

* 한식, 양식, 중식 , 일식

리뷰 평점 분석(추천 관광지 특성)

3단계 리뷰 처리

* 네이버 리뷰
* 컨텐츠 기반 모델
* 자연어

구석구석 해시태그 분류

관광

* 자연, 역사, 체험, 테마, 문화, ETC

음식

* 카페, 카페 아닌거

한 번에 나누기 싶지 않음

태그 분류는 두명이서 1차 분류

네이버 리뷰

* 이런점이 좋았어요

-> 관광지 추천 후 맛집 추천?

음식의 갯 수

구석구석 태그

-> 관광지

3단계 -> INPUT ( )

가희님

관광+숙소+음식 -> 중분류부터 선택을 하게 하고, 네이버 리뷰를 통한 처리 후

->관광지 + 숙소 + 음식를 함께 보여줌 , 유사한 관계를 가진 장소를 추천, 위치기반 가까운

(관광지 + 음식)

태영님

사용자 -> 대분류(음식 또는 관광) -> 중분류(관광이라 하면 자연, 테마 등등)를 선택하고

-> 구석구석 태그를 선택하게 하고(대표성을 뽑는다. ) -> 관광지명과 리뷰를 같이 보여주는거

수린님

구석구석 태그를 사용하여 태그 선택을 하는 거, 구석구석 태그를 소분류

리뷰 3개 정도를 보여주자 (리뷰 선택)

리뷰 선택에 따른 해당 리뷰에서 가장 많은 관광지를 보여주자 TOP 5개

관광지 명 + 리뷰 점수

지우님

관광지, 음식 -> 그 안에 중분류 나누고 -> 그리고 input 값을 받아서

결과 = 관광지명, 그 속성에 따른 평점, top 5

ex) 배민을 보면 인기순, 리뷰많은 순, 원하는 구 리스트 형식으로

성원님

1차 관광지, 음식 (음식을 관광의 유형으로)

-> 자연 체험, 문화체험, 음식 등

단계가 적었으면 좋겠다 -> 중분류 라벨링

최종 -> 마지막 결과물은

유사한 관계를 가진 관광지 추천 , 관광지에 대한 리뷰, 평점, 하나를 선택한다면 주위 다른 관광지 추천, 인기순, 리뷰많은 순, 원하는 구 리스트 형식으로

관광을 선택하면 -> 그 주변 음식 추천

구석구석 태그

우선 순위

1. 관광 구석구석 태그 분석 1 - 1455 성원님
2. 음식 구석구석 태그 분석 1 - 1726 수린님

* 음식
* 카페

1. 네이버 리뷰 속성 분석 -

띄어쓰기로 나누고 TF-IDF

문자생성모델학습 -

전처리 -> EDA -> 모델링 -> EDA -> 시각화

네이버

* 음식점
* 관광

-> 속성은 후

각 단어의 빈도

음식과 관광은 분리

음식 관광 모두 속함

1. 음식 내에서 총 리뷰 추출
2. 띄어쓰기로 나누고
3. 각 수 빈도 수 체크

-> 빽다방 : 청결해요 34빈도

맛이 좋아요 21

아웃백 청결 해요 21

서비스가 좋아요 11

청결 55

맛 21

서비스 11

롯데타워 음식 점

정규화

* 정규화 없이 단어에 대한 빈도

->

* 그 가게(관광지) 내에서 각 리뷰에 대한 퍼센트를 구해서 합해서

->

가게마다 리뷰수도 다 다르고 참여자 수도 다르고

총 참여자 수 < 가게의 리뷰 총 선택한 수

정규화 과정을 거친다면 각 가게에서 MIN 값과 MAX 값을 구해서 한다.

표준화를 하게 된다면 -> 가게마다 리뷰수가 다른데 그것을

관광지별 이런점이 좋았어요 총 투표수

리뷰 적은 가게를 버린다면 버린 후에 토큰화시켜서 속성을 알아봐야 할거 같은데

리뷰수 적은 데 만약 인기순이 높다면??

각 속성에 대한 평점

토큰화를 시켜서 각 단어들을 여러속성에 분류해준다면

리뷰 수를 신경을 안써도 되지 않을까?

근데 저희 각 리뷰를 속성에 맞게 분류를 해야한자나요

-> 리뷰를 다 보고

속성을 뽑고 분류를 하고 , 그 후에 이 가게 버릴지말지는 조금 있다가

1. 음식, 관광에서 따로 리뷰만 분리를 해요
2. 그 리뷰를 띄어쓰기로 분리를 해요

* 맛이(index 0) 좋아요(index1)



맛이 좋아요를 선택한 인원수를 다 빈도화 해서 순위를 매긴다

빽다방에 맛이 좋아요면 23for

좋아요 23

아웃백

맛이 13 +23

좋아요 13+23

관광

* 띄어쓰기 나누기 빈도수 체크

음식

* 띄어쓰기 나누기 빈도수 체크

숫자 자르는거

속성

-> 맛 [맛, 맛있어요, ~~~]

-> 청결

a 가게에 커피가 맛있어요. -> 맛

맛이 13

좋아요 26

청결

해요

https://shopitrend.oopy.io/bf0527b1-d980-43ff-9fea-a9ecb88136a8

음식 466 개 중 10860의 리뷰 존재

광광지 354개 중 5037 리뷰 존재

한명당 5,299

1. 음식 0-5299
2. 관광 5037
3. 음식2 5299-

수

목까지

빈도 수 - 합쳐보기