



Instagram

@정유경 @김연지 @강창균

#이태원 맛집

#맛스타그램

#이태원

#ITEWON



인스타그램 태그를 이용한 실제 상권과
인스타그램 맛집의 위치 비교 분석



프로젝트 개요

- 프로젝트 배경
- 프로젝트 목표

분석 계획 수립

- 분석 환경 및 사용기술
- 분석 절차

데이터 분석 및 결과

- 데이터 수집
- 전처리
- 라벨링
- 최종 데이터 구성도
- 모델링
- EDA
- 결론 및 시사점

프로젝트 마무리

- 한계점 및 개선방향
- Q & A





Instagram



프로젝트 개요



INSTAGRAM



Insta_matjip

...

최근 한달 내 이용 및 주 이용 소셜미디어

■ 주 이용 ■ 최근 한달 내 이용	전체			Gap	성별		연령별				
	2018년	2019년	2020년		남	여	10대	20대	30대	40대	50대
	Base (500)	(500)	(500)		(310)	(310)	(120)	(124)	(126)	(124)	(126)
유튜브	27.6 78.8	32.6 83.8	37.6 89.2	+5.0	49.7	34.2	60.0	42.7	38.9	33.9	34.9
네이버 블로그	17.0 65.8	21.8 71.8	22.2 78.2	+0.4	15.2	21.3	1.7	12.9	19.0	29.8	27.0
인스타그램	14.8 48.4	19.2 53.0	19.0 58.6	-0.2	11.6	26.1	18.3	28.2	27.8	2.9	7.1
페이스북	15.6 60.2	9.0 55.2	7.0 53.0	-2.0	12.3	4.8	15.0	8.1	7.1	4.0	8.7
밴드	11.2 50.4	6.2 45.4	6.0 47.2	-0.2	6.1	3.5	0.0	1.6	4.8	4.8	12.7

평소 자주 보거나 공유하는 게시물

	전체		성별		연령별				
	남	여	10대	20대	30대	40대	50대		
Base	(383)	(174)	(209)	(90)	(93)	(89)	(58)	(53)	
맛집/음식	36.8	45.5	25.6	52.7	40.4	43.1	49.1		
여행	31.0	26.3	20.0	33.3	21.3	29.3	45.3		
일상생활	25.3	30.1	34.4	28.0	29.2	20.7	22.6		
패션/유희/잡화	16.7	33.0	31.1	24.7	27.0	27.6	13.2		
러빙/라이프스타일	14.9	23.0	4.4	19.4	27.0	24.1	26.4		
TV/연예	19.0	17.2	27.8	20.4	10.1	20.7	7.5		
뷰티	2.9	23.0	22.2	15.1	10.1	12.1	5.7		
동물	10.3	15.8	17.8	8.6	16.9	12.1	9.4		
스포츠	21.8	3.3	12.2	8.6	13.5	12.1	13.2		
유머/개그	13.2	9.1	14.4	15.1	9.0	6.9	5.7		
음악/뮤직비디오	13.2	8.6	18.9	10.8	10.1	1.7	7.5		
운동/건강	12.6	8.6	6.7	8.6	13.5	12.1	13.2		
도서/공연/문화	6.3	9.1	7.8	7.5	5.6	12.1	7.5		
육아/교육	5.2	10.0	2.2	4.3	21.3	6.9	1.9		

▲ [Base : 최근 1개월 내 인스타그램 이용자, N=383, 단위 : %, 순위형 응답(1+2+3순위)] * 5% 이하인 경우 제시하지 않음 * 하늘색 음영 : 평균 대비 +5%p 이상인 데이터

문제점 :

길 건너 가면 가게가 있는데 지리적인 문제로 상권으로 포함되지 않은 경우가 다수이다.

즉, 실제 상권과 실제 이용 맛집 장소가 차이가 존재

사실 :

1. 20-30대가 많이 이용하는 SNS : 인스타그램
2. 인스타그램에서 자주 공유하는 게시물 1위 : 맛집,음식 관련 게시물 사진을 게시하기 위해 맛집을 일부러 찾아가는 경우가 있을 정도이다.

가정 :

진짜 맛집 데이터는 인스타에 더 많이 있지 않을까?

좋아요





이태원상권



그림 1 이태원 해시태그/위치정보를 활용한 상권지도

A구역 : 현재 #이태원맛집, #이태원핫플 등의 가장 많은 해시태그가 밀집된 지역

B구역 : #한남동맛집 이라는 신규 상권으로 #이태원디저트 #이태원브런치 등의 해시태그 밀집

C구역 : #우사단길이라는 떠오르는 상권으로 이태원상권과 이어지는 트렌드 상권으로 부상

1. 인스타그램 내 실제 이용 맛집 지역과 국가 지정 상권 지역 차이 확인
2. 실질적인 맛집 지도 제작
3. 맛집 지도의 비즈니스적 활용 방안 도출

좋아요





Instagram



분석 계획 수립



처리툴



분석툴

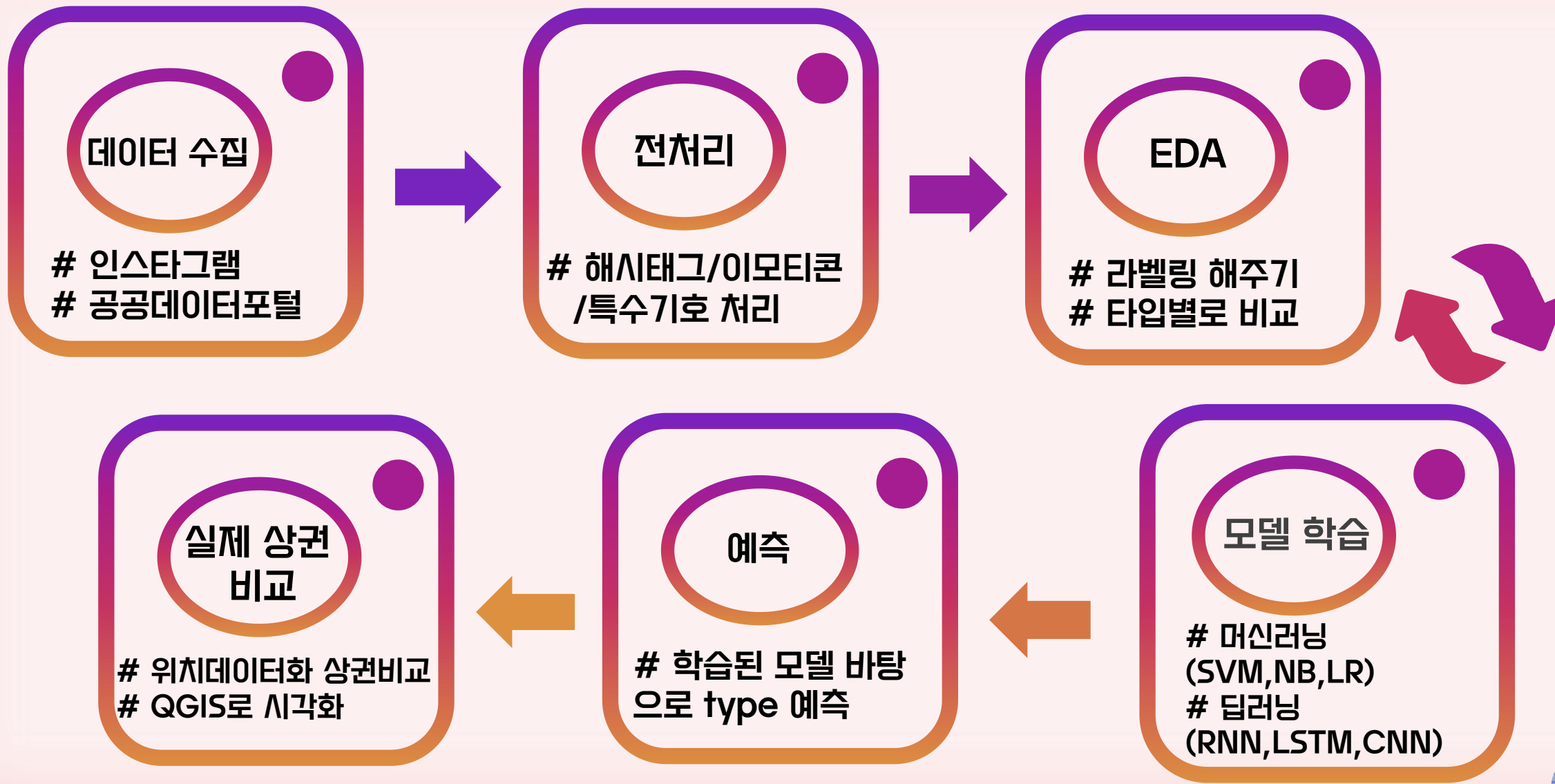


시각화툴



Folium







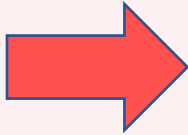
Instagram



데이터 분석 및 결과

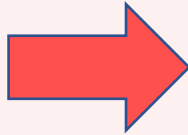


멘토님



Instagram

DATA 공공데이터포털
. GO . KR



실제 상권 데이터

좋아요





“#” 해시태그
빼주기



이모티콘
빼주기
예시:
<U+653C>



특수기호
들어간 것들
빼주기





"#" 해시태그 포함되지 않은
데이터

#

이태원맛집, 경리단길맛집, 해
방촌맛집 등 이태원 주변 맛집
을 해시태그 하지 않은 관련 없
는 데이터

관련X

라벨링 0

선정적

쩍벌녀,호빠,출장마사지, ...등
선정적인 글

광고성

1. 상담, 문의, 전화, 프로필링크,
택배, 플러스친구,공구,대관, ..
2. 다이어트, 디톡스, 젤네일, 속눈썹,
강아지옷, 맛집외 다른 광고글
3. 다른 지역 태그 EX) 홍대,강남,
부산 건대, 마포, 광화문, ...

라벨링 0에 해당되지 않은
데이터들은 라벨링 1

나머지

라벨링 1

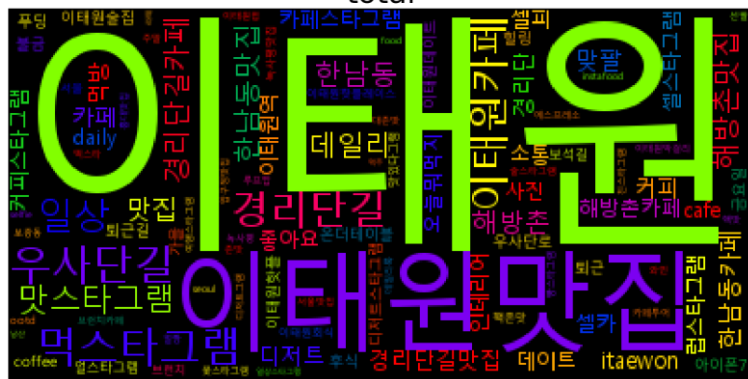
좋아요





Insta_matjip [Follow me](#)

total



전체 데이터
이태원
이태원맛집
먹스타그램
우사단길
이태원카페
일상



Insta_matjip Follow me

type1

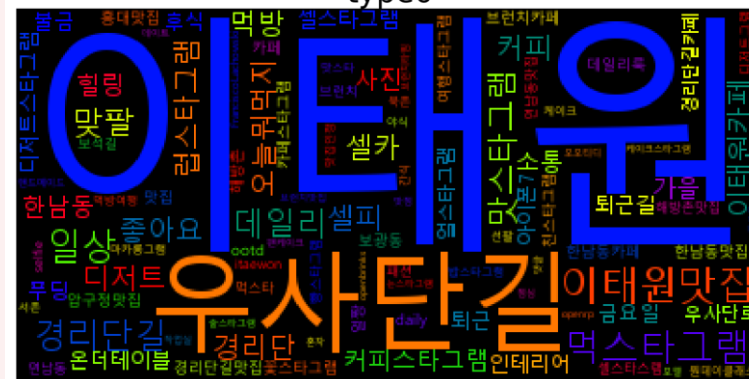


정확한 데이터
이태원
이태원맛집
이태원카페
먹스타그램
한남동맛집
경리단길카페



Insta_matjip Follow me

type0



광고 데이터
이태원
우사단길
이태원맛집
일상
맛팔
셀카

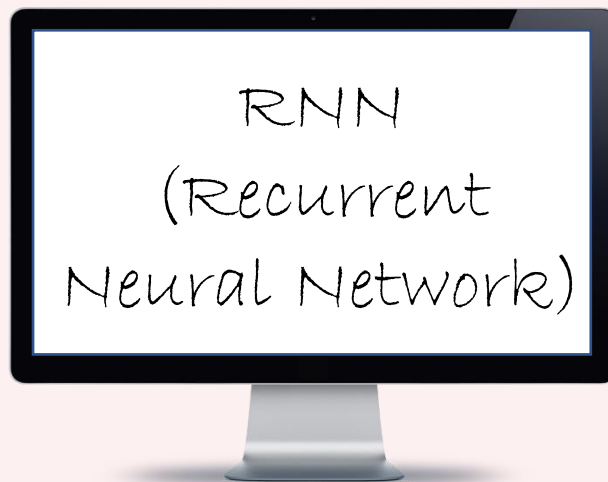




Train : Validation = 8 : 2



accuracy : 0.9275
precision : 0.9413
recall : 0.9601
f1 score : 0.9542



accuracy : 0.9531
precision : 0.9796
recall : 0.9601
f1 score : 0.9692

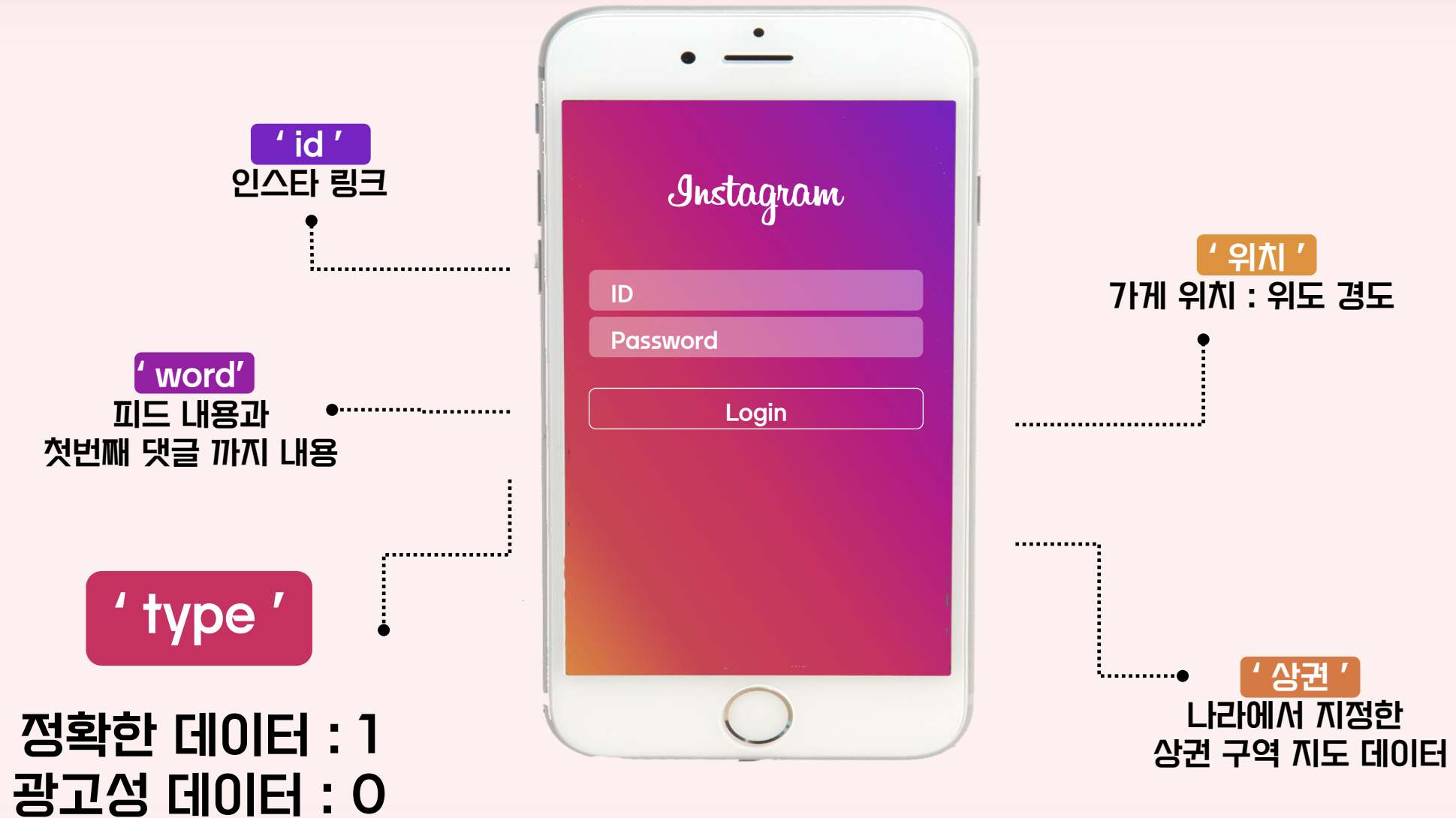


accuracy : 0.9541
precision : 0.9723
recall : 0.9585
f1 score : 0.9693



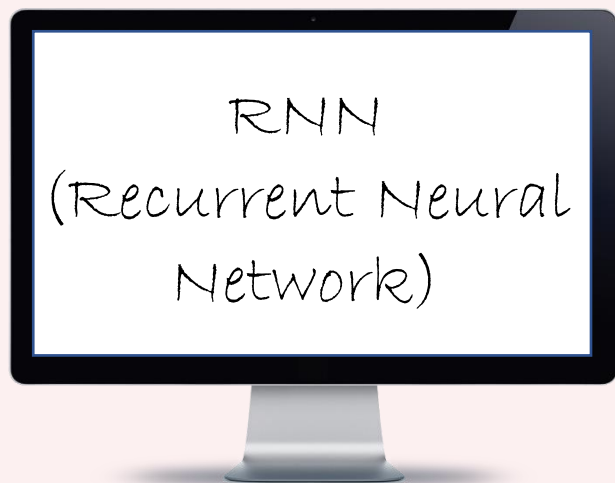


최종 데이터 구성도



좋아요

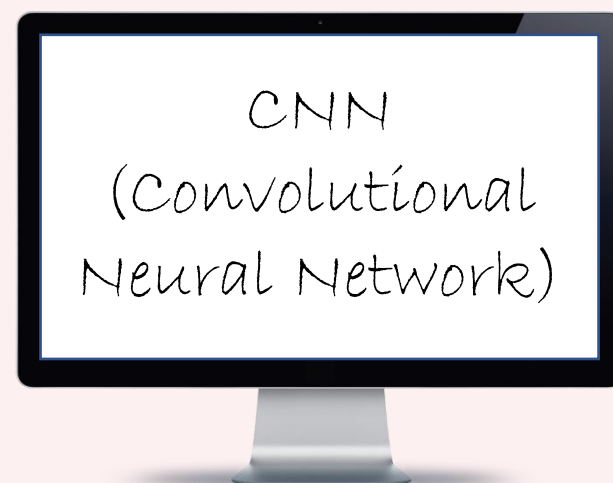




accuracy: 0.8682
precision: 0.8938
recall: 0.9008
f1 score: 0.8948



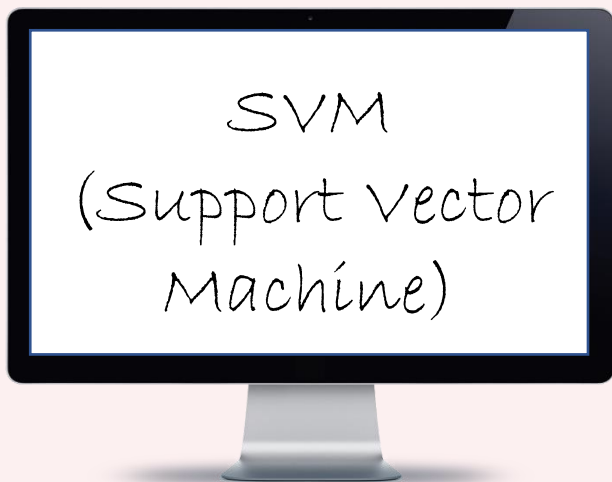
accuracy: 0.8875
precision: 0.8872
recall: 0.9414
f1 score: 0.9117



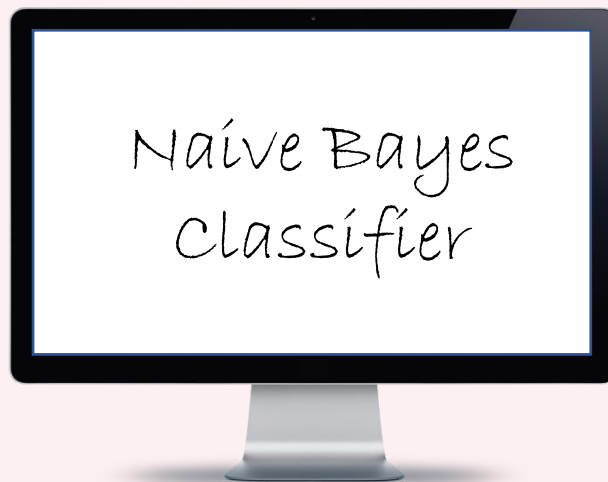
accuracy: 0.8907
precision: 0.9007
recall: 0.9302
f1 score: 0.9132

좋아요





accuracy : 0.8905
precision : 0.8942
recall : 0.9395
f1 score : 0.9163



accuracy : 0.8770
precision : 0.8820
recall : 0.9320
f1 score : 0.9063



accuracy : 0.8848
precision : 0.8911
recall : 0.9335
f1 score : 0.9118

좋아요





SVM
(Support Vector
Machine) +
CountVectorizer

f1 score : 0.9163

CNN
(Convolutional
Neural Network)
+ Tokenizer

f1 score : 0.9132

Naïve Bayes Classifier

각 토큰의 확률값을 곱셈한 것으로 라벨을 결정하는 알고리즘으로 토큰이 많아지면 0으로 판단할 확률이 높아진다.

Logistic Regression

선형으로만 분류가 가능하다.

2차원으로만 데이터를 분류할 수 있다.

→ SVM 활용

RNN & LSTM

시계열 데이터에 적합한 알고리즘을 가지고 있어, 시계열 데이터가 아닌 태그 데이터에 적합하지 않다.

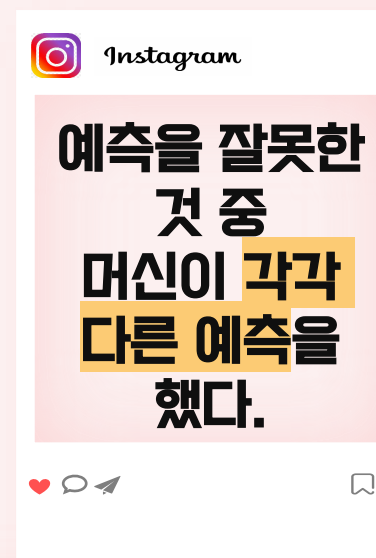
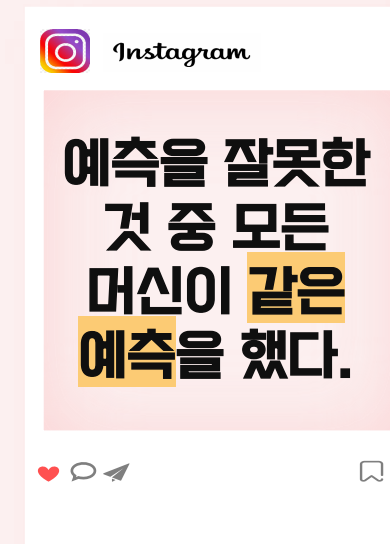
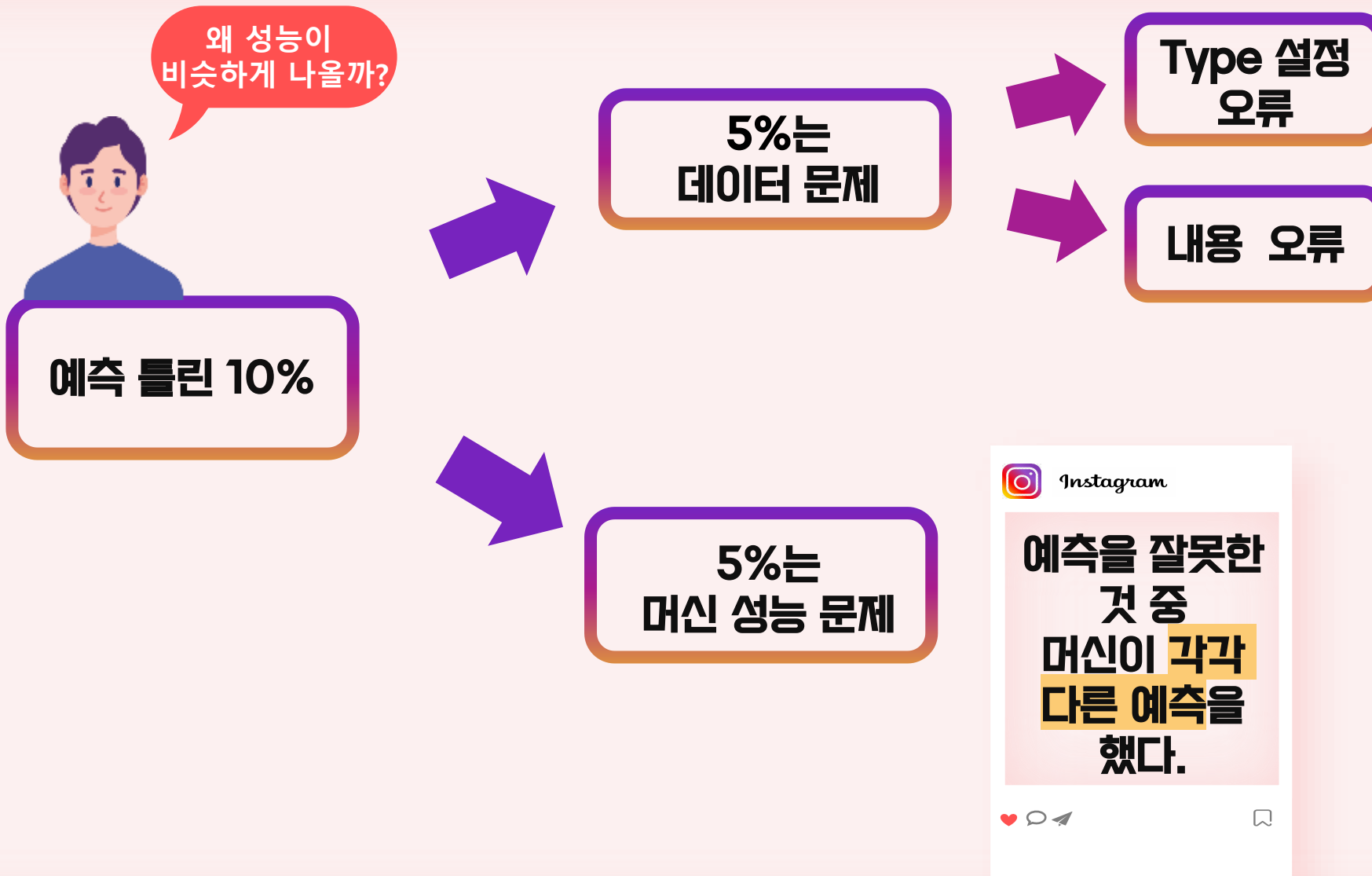
→ CNN 활용

정밀도와 재현율은
trade off 관계이므로
둘 다 좋은 점수를 맞아
야 좋은 모델이므로
f1score를 선택!

CountVectorizer는 먼저 단어 사전 벡터를 만들고, 카운팅
할 문장을 확인하며 그 단어 사전의 횟수를 카운팅하여 수치
벡터화 하는 것이다

Tokenizer는 단어를 여러개의 Token으로 나누는 것이다





좋아요





Instagram

Type 설정 오류 예시

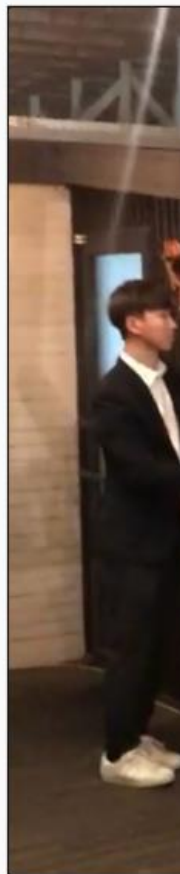


↳ type은 "0"이라 되어있고 예측은 "1" 혹은 type은 "1" 예측은 "0"이라 해서 예측값이 오히려 맞는 상황

Instagram



Instagram



Instagram



Q 검색



seungaeer • 팔로우



seungaeer #macrame with #beachcombing #exhibition

제주도 바닷가에 버려진 병, 전구 등의 물건들과 함께한 매듭공예 전시

#이태원 #우사단길 #milelab #매듭공예 #전시

403주



좋아요 10개

9월 22, 2014



댓글 달기...

게시



좋아요





Instagram

내용 오류 Test type 0 예측 1 해시태그 키워드



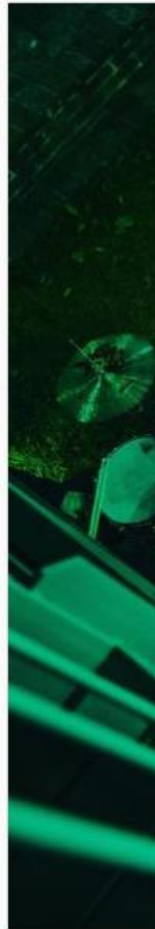
↳ 해시 태그에는 광고내용이 적어서 "1"이라 잘못 예측
↳ 맛집 데이터 아는데 이태원 태그해서 "1"이라 잘못 예측
↳ 익선동 맛집인데 이태원맛집, 한남동맛집 태그해서 "1"이라 잘못 예측

Instagram

Instagram

Instagram

Q 검색



changhwa_official • 팔로우



창화당 익선동



changhwa_official 10월 첫날 ... 그리고 연휴의 시작
즐거운 한가위 되세요 🍁🍂🍃
@chang.hwa.dang
#창화당 #만두잘하는집 #만두맛집 #만두 #먹스타그램 #한남동맛집 #고대맛집 #익선동맛집 #종로맛집 #이태원맛집

246주



denverhotchic 🍷👍🍷



245주 답글 달기



insajonggajib 안녕하세요! 인스타 잘 보고 갑니다~



좋아요 224개

10월 1, 2017



댓글 달기...

게시



좋아요

게시



Instagram

내용 오류 Test type 1 예측 0 해시태그 키워드



↳ 내용은 카페 맛집이지만 태그는 셀카에 관련 이라서 "0"이라 잘못 예측

↳ 댓글에는 태그내용이 추가되어 있지만 피드 내용에는 이태원 관련 없어 "0"이라 잘못 예측

↳ 실제로는 "1"이지만 댓글에는 광고성 태그를 달아서 "0"으로 잘못 예측

Instag Instagram Instagram

Q 검색



nailmoood · 팔로우
Studio Concrete



nailmoood -
솔직하게 할 말 다 하는게 마음에 들어
#유아인 #유아인카페
-
@hongsick
수정됨 · 238주



nailmoood #청주#네일#청주네일
#청주네일샵#내성발톱#봉명동네일
샵#청주맘#한스네일#청주한스네일
#청주발톱교정#사창동네일샵#청주
내성발톱#청주파고드는발톱#오창
내성발톱#오송내성발톱#소통#맛팔
#선팔#nail #nails #nailart #daily#청
주봉명동네일#fff#f4f#follow#한남
동카페#스튜디오콘크리트



좋아요 44개

11월 26, 2017



댓글 달기...

게시

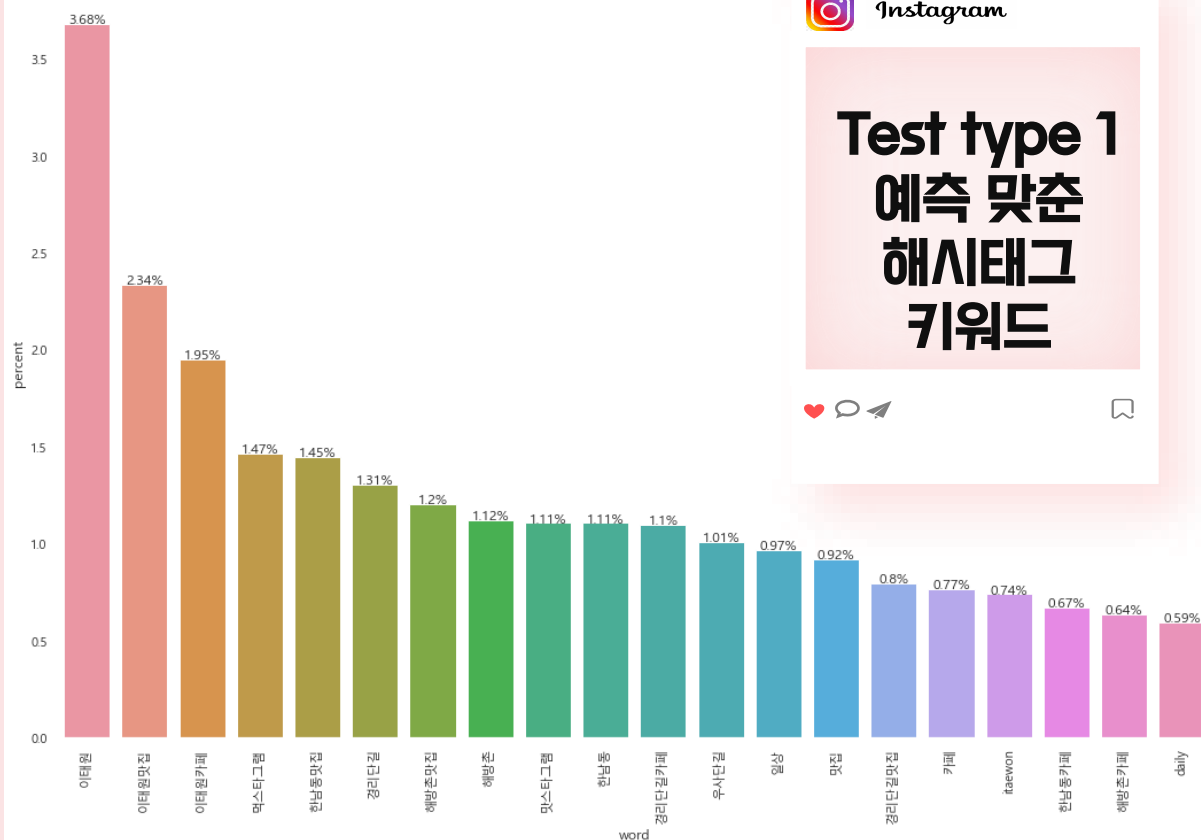


좋아요





Correct_1_word

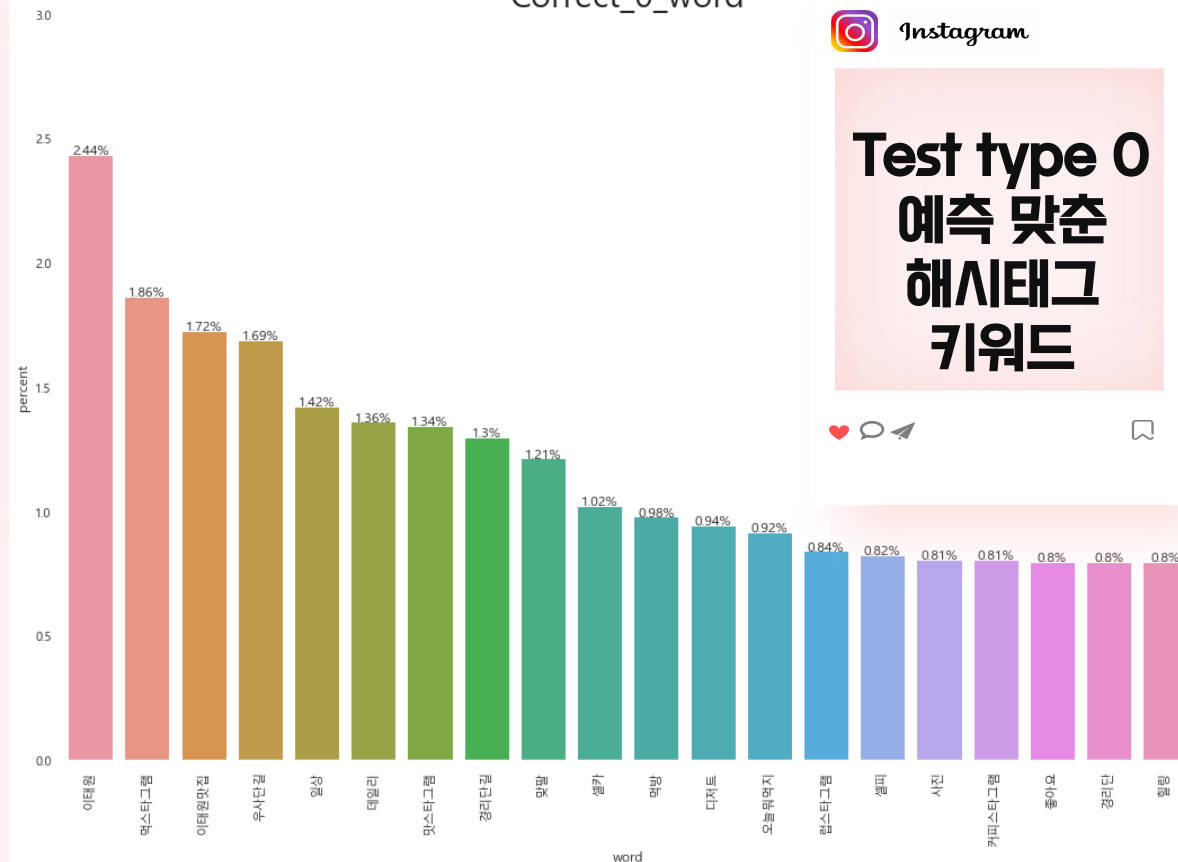


Instagram

Test type 1
예측 맞춘
해시태그
키워드



Correct_0_word



Instagram

Test type 0
예측 맞춘
해시태그
키워드

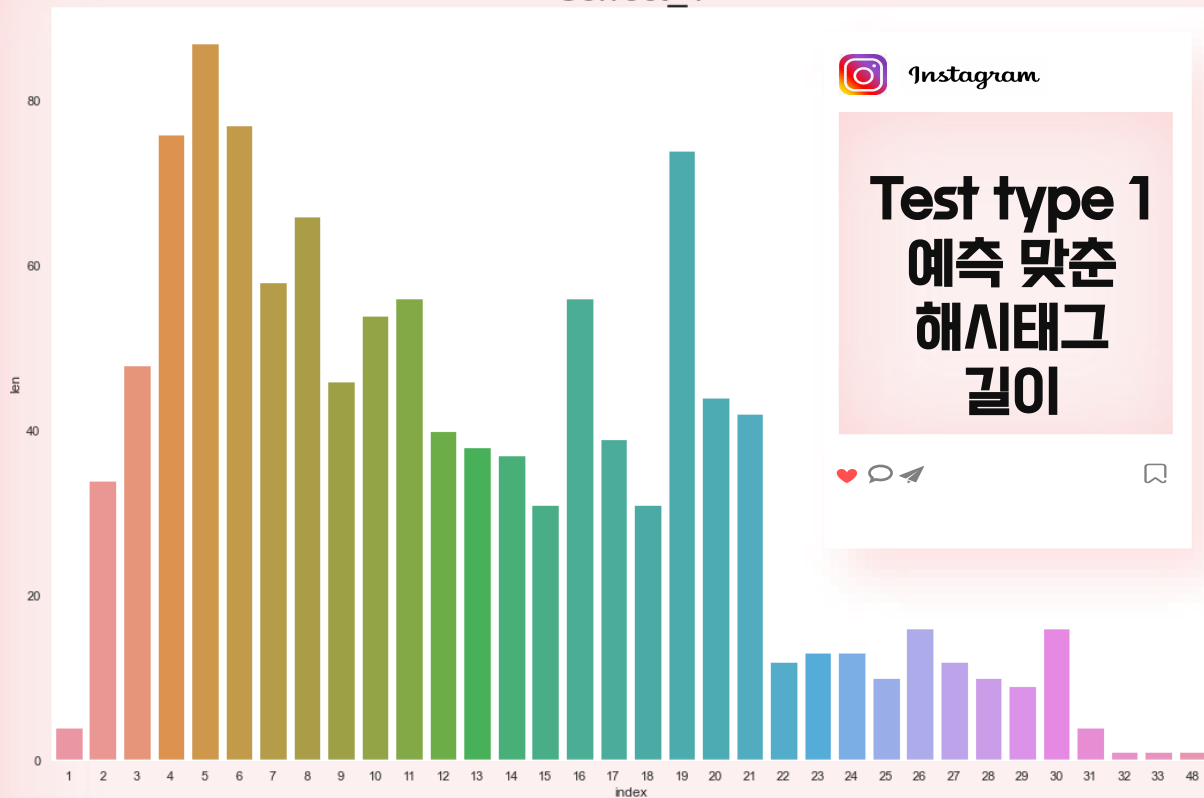


좋아요





Correct_1



Correct_0





결론(실제 상권과 인스타그램 맛집 비교)

Page. 21

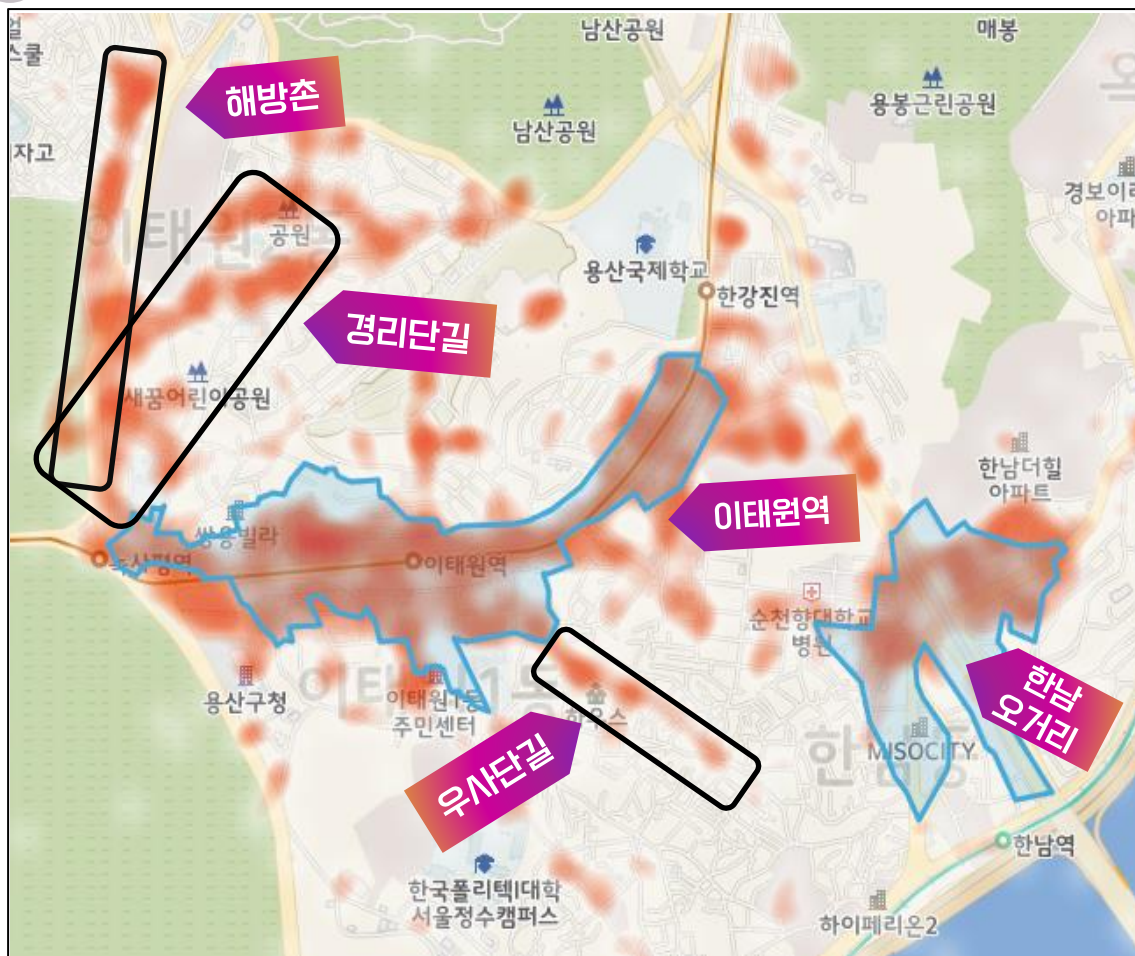


INSTAGRAM



○ Insta_matjip

...



*공데 상권 지도와 인스타 지도 데이터의 차이점

1. 경리단길, 해방촌을 발달 상권으로 추가

2. 이태원역 근처 발달상권 구체적으로 넓게 재구획화

3. 한남동 발달상권 아래부분은 잘라내고 위에 붙어있는 부분을 더 넓게 구획화

4. 우사단길을 발달 상권으로 추가

5. 다른 열은 부분들은 아파트나 상가지역으로 상권보다는 배달로 인한 위치 태그로 판단되어 제외

좋아요





위치기반서비스(LBS), GIS
이용하여 사용하는 스타트업
회사들은 데이터 수가 적기
때문에 공공기반 상권데이터
를 사용할 수 밖에 없다.
공공기반 상권데이터는 실제
와 맞지 않는 부분들이 있기
때문에 우리가 만든 상권 지도
를 사용하면 실질적인 상권 구
역을 이용하게 되는 것이므로
비즈니스적으로 더 높은 활용
가치가 있을 것으로 예상된다.

F1 score보다는 재현율 높여서
사용하기 : 재현율이 실제 중 예측
이 맞는 비율이기 때문에

EX) 맛집 추천 어플 사이트



배달 사이트

데이트 코스 추천 어플



결혼정보회사 등

좋아요





Instagram



프로젝트 마무리



한계점

- 인스타그램을 직접 가지고 오지 못한 것이 아쉬웠다.
- 2017년 인스타그램 데이터였기 때문에 코로나로 인해 폐업한 곳들도 있고 현재와 달라진 곳들이 존재한다.
- 라벨링을 더 많이 못 해서 아쉬웠다.

개선방향

- 단어별, 좋아요수, 댓글 수로 가중치를 두어 계산을 할 수 있었다면 좋았을 것 같다.
- 단어 유사도를 활용한 Word2vec 이용해서 돌려보면 좋을 것 같다.





팔로잉

내 소식



2_조님이 회원님을 팔로우 하기 시작했습니다.

팔로잉



2_조님이 댓글에서 회원님을 언급했습니다 :

@insta_matjip Q&A 시간입니다



2_조님이 댓글을 남겼습니다 :

질문이 있으시다면 질문해주세요



MEDICI님이 회원님을 팔로우 하기 시작했습니다.

팔로우



Instagram

THANK
YOU



#메디치

#수고하셨습니다:D