



# 전국 반려동물 시설 리뷰 감성평가

2조 냥아치

2024.05.02

NyangAchi

# Contents

## 1. 프로젝트 개요

1-1) 배경, 1-2) 목표

## 2. 팀 소개 및 일정

2-1) 팀 소개, 2-2) 개발 환경, 2-3) 프로젝트 일정

## 3. 데이터 선정 및 수집

3-1) 데이터 선정, 3-2) 데이터 수집

## 4. 자연어 처리

4-1) 데이터 전처리  
4-2) 비지도 학습, 4-3) 지도 학습

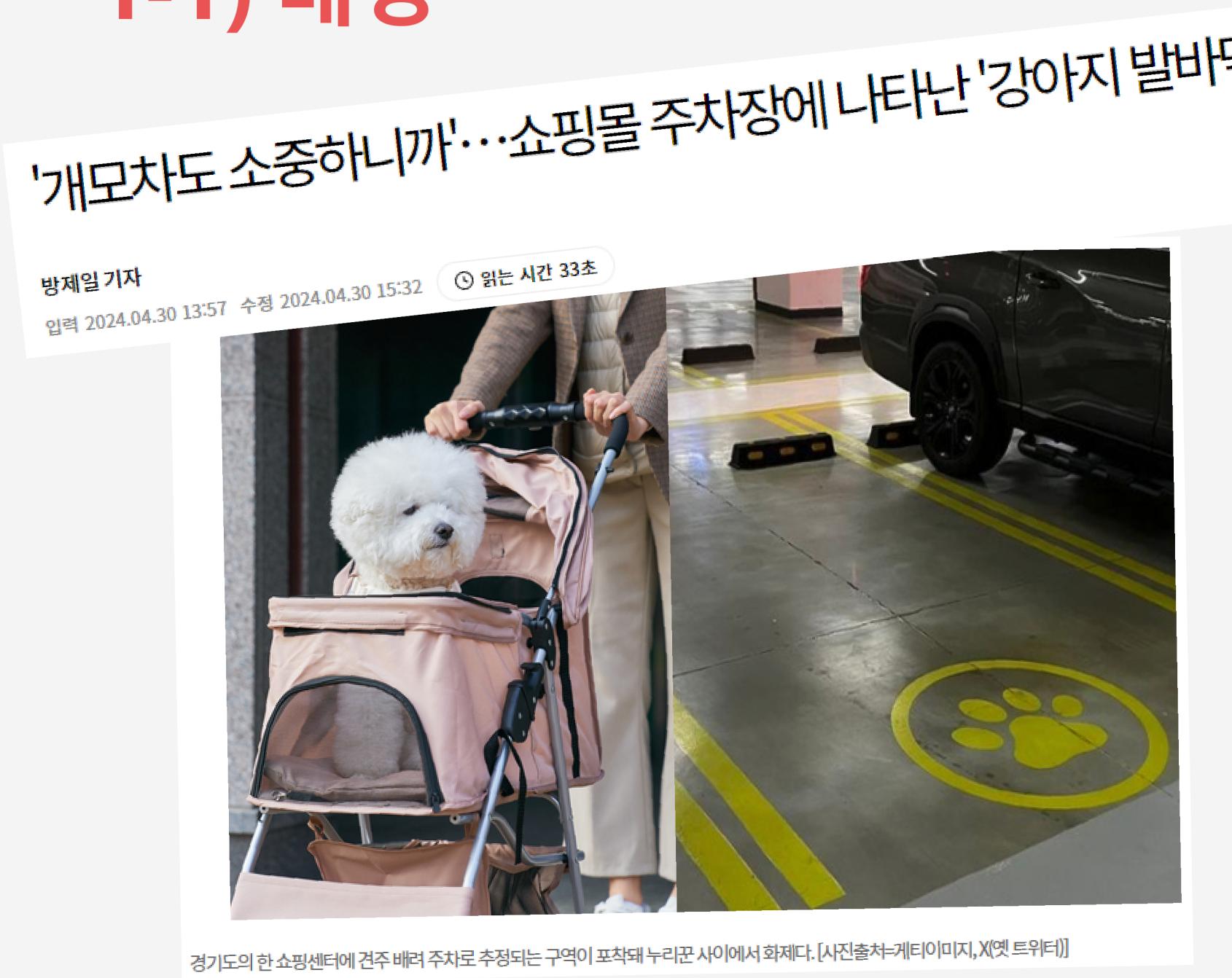
## 5. 프로젝트 성과 및 느낀점

5-1) 프로젝트 성과, 5-2) 향후 계획, 5-3) 느낀점

## 6. 질의 응답

# 프로젝트 개요

## 1-1) 배경



반려동물을 위한 시설이 증가하고 있다.



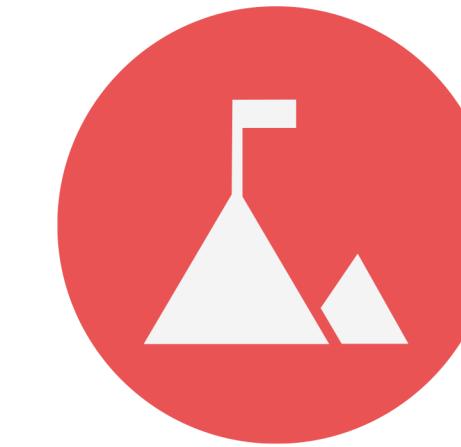
감정? 긍정? 부정?

# 프로젝트 개요

## 1-2) 목표



전국 반려동물 동반가능한 시설의 시설분류별  
SNS 리뷰를 수집,  
감성 분석 모델 생성 및 성능평가 시각화



수집한 시설분류별 데이터  
감성 분석 후 시각화

# 팀소개 및 일정

## 2-1) 팀 소개

김재현

- 총괄
- 리뷰 데이터 수집
- 데이터 전처리



최은석

- 리뷰 데이터 수집
- 감성분석 모델 생성  
(지도 학습) 및 시각화



정소라

- 리뷰 데이터 수집
- 불용어 처리
- 시설별 감성분석
- 감성분석 모델 생성  
(비지도 학습) 및 시각화



# 팀소개 및 일정

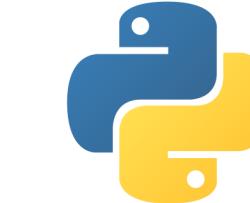
## 2-2) 개발 환경

IDE



Ver. 6.5.4

Language



Ver. 3.11.5

데이터 수집



Selenium

Ver. 4.19.0



Selenium  
WebDriver

Ver. 4.19.0

BeautifulSoup

Ver. 4.12.2

분석 및 시각화

Pandas  
Ver. 2.0.3



NumPy  
Ver. 1.24.3



Ver. 3.7.2



Ver. 1.9.3



Ver. 0.12.2

자연어 처리



KoNLPy  
Ver. 0.6.0



scikit  
learn  
Ver. 1.3.0

TensorFlow™  
Ver. 2.16.1

# 팀소개 및 일정

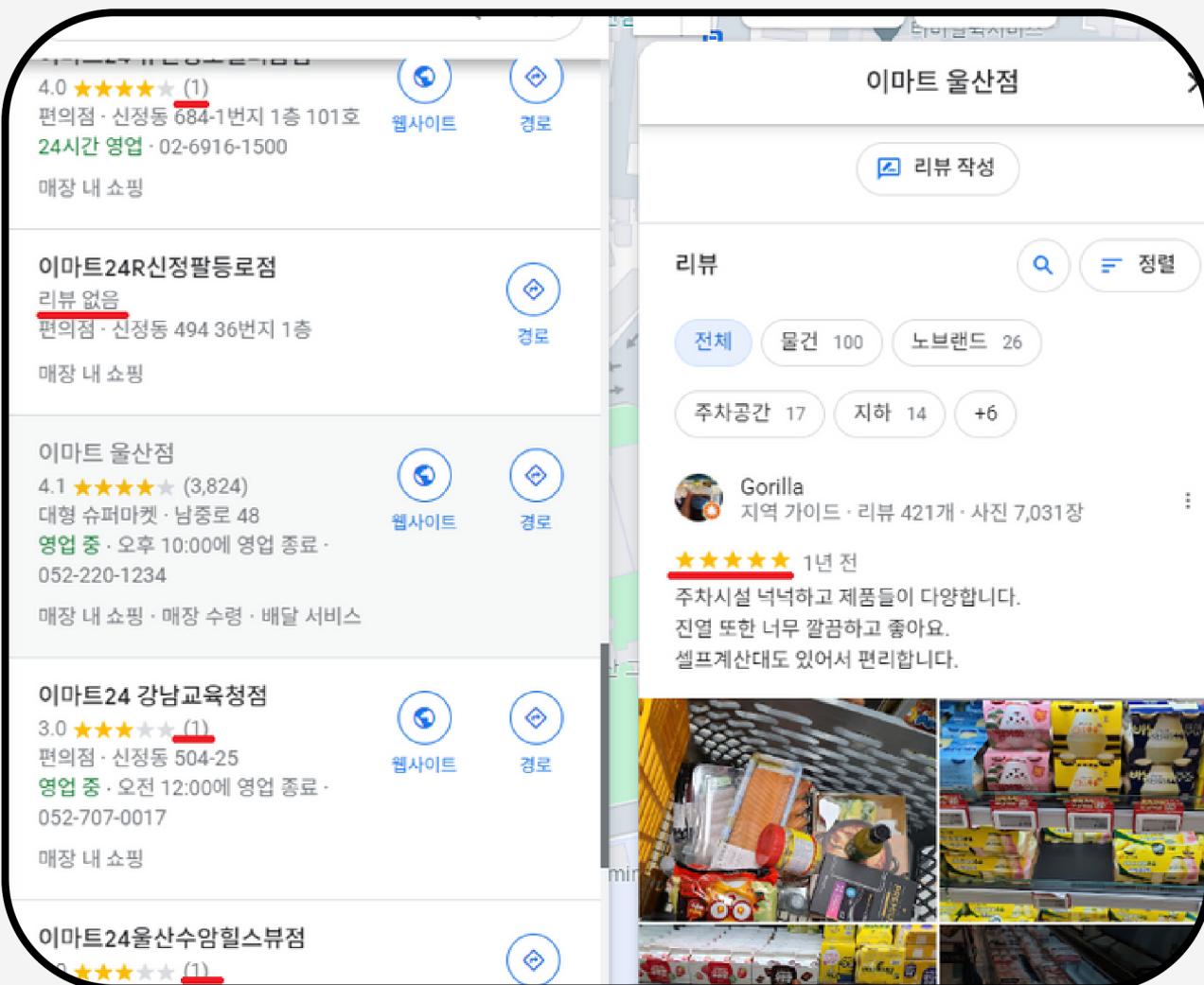
## 2-3) 프로젝트 일정



# 데이터 선정 및 수집

## 3-1) 데이터 선정

### 구글 리뷰



### 네이버 리뷰



- 별점 시스템
- 적은 리뷰

- 많은 리뷰
- 2021년 10월 이후 사라진 별점 시스템

# 데이터 선정 및 수집

## 3-2) 데이터 수집 (1/2)

문화체육관광일반 한국문화정보원

[활용승인 / 무료] 한국문화정보원\_전국 반려동물 동반 가능 문화시설 위치 데이터

```
1 # 행정구역+시군구가 포함된 시설명을 변수로 지정  
2 items = name_data['행정구역'] + ' ' + name_data['시군구'] + ' ' + name_data['시설명']
```

### 1. 공공데이터포털 데이터 이용

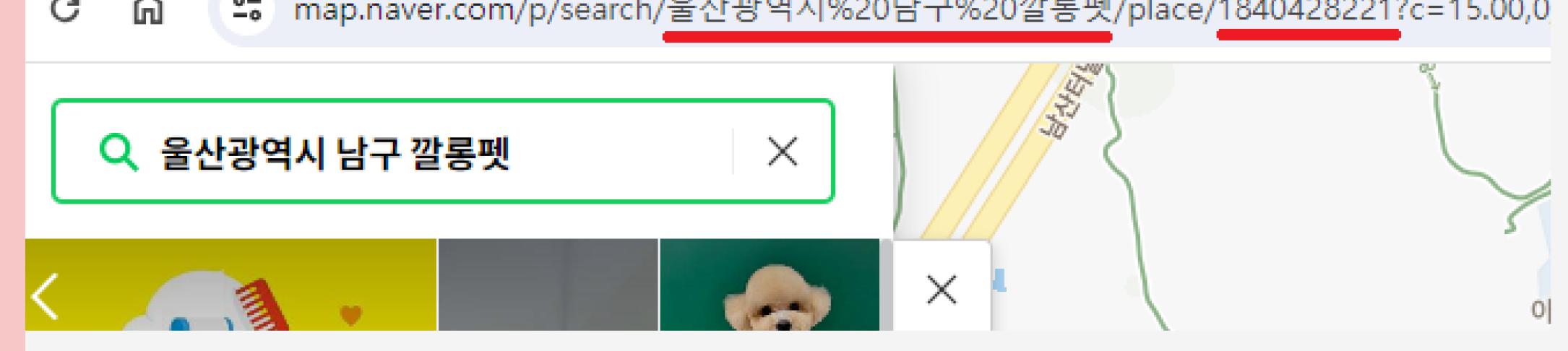
### 2. 행정구역+시군구+시설명 네이버Map URL로 검색(Search) 후

### 3. 네이버 Place URL로 변환

```
final_url = 'https://pcmap.place.naver.com/place/' + res_code[0] + '/review/visitor#'
```

Search

시설코드

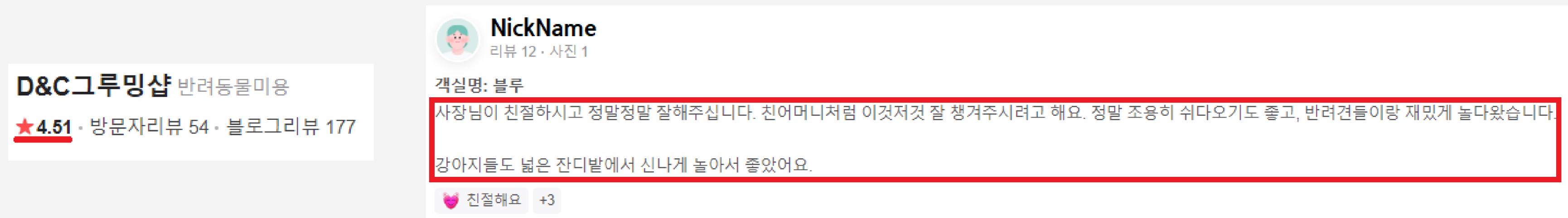


시설코드

	name	naverURL
0	대구광역시 달성군 119동물병원	<a href="https://pcmap.place.naver.com/place/80672995/r...">https://pcmap.place.naver.com/place/80672995/r...</a>
1	전라북도 전주시 덕진구 21세기 동물병원	<a href="https://pcmap.place.naver.com/place/17095445/r...">https://pcmap.place.naver.com/place/17095445/r...</a>
2	서울특별시 을지2동 21세기동물병원	<a href="https://pcmap.place.naver.com/place/1571398284...">https://pcmap.place.naver.com/place/1571398284...</a>
3	전라남도 목포시 21세기동물병원	<a href="https://pcmap.place.naver.com/place/1382096574...">https://pcmap.place.naver.com/place/1382096574...</a>

# 데이터 선정 및 수집

## 3-2) 데이터 수집 (2/2)



D&C그루밍샵 반려동물미용  
★4.51 · 방문자리뷰 54 · 블로그리뷰 177

NickName  
리뷰 12 · 사진 1

객실명: 블루

사장님 친절하시고 정말정말 잘해주십니다. 친어머니처럼 이것저것 잘챙겨주시려고 해요. 정말 조용히 쉬다오기도 좋고, 반려견들이랑 재밌게 놀다왔습니다.

강아지들도 넓은 잔디밭에서 신나게 놀아서 좋았어요.

❤ 친절해요 +3

	name	rate	review
0	대구광역시 중구 525그루밍		전체적으로 다 마음에 들어요!!\n감사합니다 또 방문하겠습니다~^^
1	대구광역시 중구 525그루밍		우리복주를 늘 잘 케어 잘해주시고 이뻐해주시고이쁘게 사진도 한번씩 보내주시구ㅠㅠ 글...
2	대구광역시 중구 525그루밍		미용 너무 잘 해주십니다! 재방문 의사 있어요 감사합니다!! 다음에 또 들리겠습니다!!
3	대구광역시 중구 525그루밍		목욕 맡겼었는데 사장님이 진심으로 강아지 좋아하시더라고요 ~~ 저희 강쥐도 편했는지...
4	대구광역시 중구 525그루밍		집에서 미용할 컨디션이 아니라서 맡겼는데\애가 미용도 싫어하고 미용매너도 없다하셨...
...	...	...	...
4234	강원도 평창군 힐링하시개 애견펜션		아가들과 너무 즐겁게 놀다왔어요 공기도 좋고 바베큐도 즐겁게 먹었어요 옆방 멍멍미...
4235	강원도 평창군 힐링하시개 애견펜션		우리 가족 오랜만에 여행에 다녀왔어요ㅎㅎ 강아지동반 팬션 알아보다가 운동장이 좋아보...
4236	강원도 평창군 힐링하시개 애견펜션		운동장에서 멍이를 뛰어다니기좋구요! 사장님 뭐필요한거없는지 물어봐주시고 잘챙겨주세요구...
4237	강원도 평창군 힐링하시개 애견펜션		사장님께서 시골할머니처럼 잘챙겨주시고 엄~청 친절하세요~!\n잔디 마당도 넓어서 ...
4238	강원도 평창군 힐링하시개 애견펜션		입실할때부터 친절하게 짐도 옮겨주시고 혼자와서 많이 신경도 써주시는 사장님ㅠㅠ 강아...
4239 rows × 3 columns			

- name 컬럼 : 검색 시 text
- rate 컬럼 : 시설 전체 평점
- review 컬럼 : 방문자 리뷰

# 자연어 처리

## 4-1) 데이터 전처리

name	rate	review	시설분류
충청북도 옥천군 가화쌈지숲공원		?	문화
충청북도 옥천군 가화쌈지숲공원		친절해요	문화
충청북도 옥천군 가화쌈지숲공원		좋아요 만족스러워요	문화
충청북도 옥천군 가화쌈지숲공원		와우~~	문화
충청북도 옥천군 가화쌈지숲공원		굿	문화
충청북도 옥천군 가화쌈지숲공원		굿	문화

1. 리뷰 데이터 정규화
2. 형태소 분석 Okt(), Mecab()
3. 어간 추출 KNU 감성사전
4. 감성 분석

	review	시설분류	review2	review_okt	okt_stemmed	score	label
0	?	문화				0	0
1	친절해요	문화	친절해요	친절해요	친절하다	2	1
2	좋아요 만족스러워요	문화	좋아요 만족스러워요	좋아요 만족스러워요	좋다 만족스럽다	4	1
3	와우~~	문화	와우	와우	와우	0	0
4	굿	문화	굿	굿	굿	1	1

1. 리뷰 데이터 정규표현식을 이용하여 한글만 추출
2. 형태소 분석 : 조사만 제외
3. 어간 추출 : okt.morphs(text, stem=True)
4. KNU 감성 사전으로 감성 분석

# 자연어 처리

## 4-1) 데이터 전처리

형태소 분석

감성 분석

**pecab()**

[('리뷰' - NNP'),  
 ('많' - VA'),  
 ('있' - VA'),  
 ('귀엽' - VA'),  
 ('재밌' - VA'),  
 ('챠' - NNP'),  
 ('내라' - NNP'),  
 ('정리' - NNP'),  
 ('없' - VA'),  
 ('넓' - VA'),  
 ('예쁘' - VA'),  
 ('좋' - VA'),  
 ('크레' - NNP'),  
 ('알려' - NNP'),  
 ('갈' - VA')]

**okt() - 지도학습**

[('설사', 'Noun'),  
 ('를', 'JKO'),  
 ('해서', 'VV+EC'),  
 ('병원', 'NNG'),  
 ('갔', 'VV+EP'),  
 ('어요', 'EF'),  
 ('배', 'NNG'),  
 ('를', 'JKO'),  
 ('만져', 'VV+EC'),  
 ('보', 'VX'),  
 ('고', 'EC'),  
 ('꼼꼼히', 'MAG'),  
 ('살펴보', 'VV'),  
 ('시', 'EP'),  
 ('더니', 'EC'),  
 ('약', 'NNG'),  
 ('은', 'JX'),  
 ('안', 'MAG'),  
 ('먹', 'VV'),  
 ('어도', 'EC'),  
 ('되', 'VV'),  
 ('겠', 'EP'),  
 ('다고', 'EC'),  
 ('하', 'VX'),  
 ('셔', 'EP+EC')]

**mecab() -  
비지도 학습**

**kkma()**

[('퀴업' - VA'),  
 ('다르' - VA'),  
 ('아힐' - VA'),  
 ('높' - VA'),  
 ('예쁘' - VA'),  
 ('이쁘' - VA'),  
 ('미끄럽' - VA'),  
 ('좋' - VA'),  
 ('그렇' - VA'),  
 ('안되' - VA'),  
 ('길' - VA'),  
 ('갈' - VA'),  
 ('많' - VA'),  
 ('미심쩍' - VA'),  
 ('재밌' - VA'),  
 ('괜찮' - VA'),  
 ('심하' - VA'),  
 ('없' - VA'),  
 ('넓' - VA'),  
 ('조마조마하' - VA'),  
 ('아프' - VA'),  
 ('나쁘' - VA')]

**hannanum() komoran()**

[('감사드리' - P'),  
 ('쉬' - P'),  
 ('먹' - P'),  
 ('주' - P'),  
 ('많' - P'),  
 ('그러' - P'),  
 ('나쁘' - P'),  
 ('들' - P'),  
 ('웬만하' - P'),  
 ('무섭' - P'),  
 ('대' - P'),  
 ('쓰' - P'),  
 ('끌나' - P'),  
 ('약하' - P'),  
 ('아니' - P'),  
 ('찾' - P'),  
 ('않' - P'),  
 ('챙기' - P'),  
 ('느끼' - P'),  
 ('안되' - P'),  
 ('사' - P'),  
 ('말' - P'),  
 ('고르' - P'),  
 ('트' - P'),  
 ('남기' - P'),  
 ('찰' - P'),  
 ('가' - P'),  
 ('이쁨' - P')]

# 자연어 처리

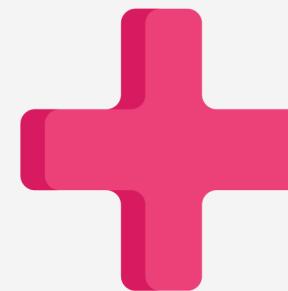
## 4-1) 데이터 전처리

형태소 분석

감성 분석

	word	word_root	polarity
0	(;	(	1
1	(_.)	(_.)	-1
2	(^^)	(^^)	1
3	(^~^)	(^~^)	1
4	(^^*)	(	1
...	...	...	...
14849	반신반의하다	반신반의	0
14850	신비롭다	신비	1
14851	아리송하다	아리송	-1
14852	알쏭하다	알쏭하	-1
14853	알쏭달쏭하다	알쏭달쏭	-1
14854 rows × 3 columns			

NU K 감성사전



사전 단어  
추가 작업

- 감성사전으로 score 계산
  - label컬럼에
    - score > 0 return 1 (긍정)
    - score < 0 return -1 (부정)
    - score = 0 return 0
- (0이면 중립 OR Stemming값 NaN)

```
1 df_review_data[df_review_data['label'] == 0].shape  
(22660, 7)  
  
1 df_review_data[df_review_data['label'] == 1].shape  
(112536, 7)  
  
1 df_review_data[df_review_data['label'] == -1].shape  
(6049, 7)
```

# 자연어 처리

## 4-2) 비지도 학습 - Topic Modeling

- 반려동물 동반시설 리뷰데이터 - Topic Modeling

- 클래스 균형화

시설분류	
미용	997
동물병원	997
동물약국	997
문화	997
반려동물용품	997
위탁및숙박	997
카페및식당	997
Name:	count, dtype: int64

- 벡터화 : CounterVectorizer 와 TF-IDF 실행
  - CounterVectorize로 시각화 -> 긍정 데이터만 존재
  - TF-IDF로 선택

```
2 tfidvect = TfidfVectorizer(max_features=3000) # tf-idf 벡터라이즈 (중요도)
```

- 불용어 추가

```
1 # 불용어 리스트
2 stop_words = ['하다', '이다', '맞다', '가다', '먹다', '되다', '보고', '싶다', '정도', '있다', '오다', '너무', '없다', '자다',
   '들다', '되어다', '넓다', '않다', '같다', '타다', 'ㅎㅎ', '해주다', '다니다', '분풀', '저희', '항상', '자주', '우리', '여기', '아이',
   '매기', '엄청', '지다', '다른', '나오다', '그렇다', '—', 'ㅠㅠ', '비다', '조금', '아주', '으쌰으쌰', 'ㅋㅋ', 'ㅋㅋㅋ', 'ㅎㅎㅎ']
```

- LDA 모델 생성

```
1 n_components = 4 # 토픽 수
2 LDA_model_tfidf = LatentDirichletAllocation(n_components=n_components, random_state=2405)
```

# | 자연어 처리

## 4-2) 비지도 학습 - Topic Modeling

- Topic 모델링 : TF-IDF (단어 중요도)



1. 의료 및 케어  
서비스



2. 고객 서비스,  
상호작용



3. 제품 관련



4. 이용 시설

# | 자연어 처리

## 4-2) 비지도 학습 - Topic Modeling

- Topic 모델링 : TF-IDF (단어 중요도)



1. 의료 및 케어  
서비스

★전문적인  
서비스



2. 고객 서비스,  
상호작용

★고객 서비스의 질



3. 제품 관련

★가격  
★제품의 다양성



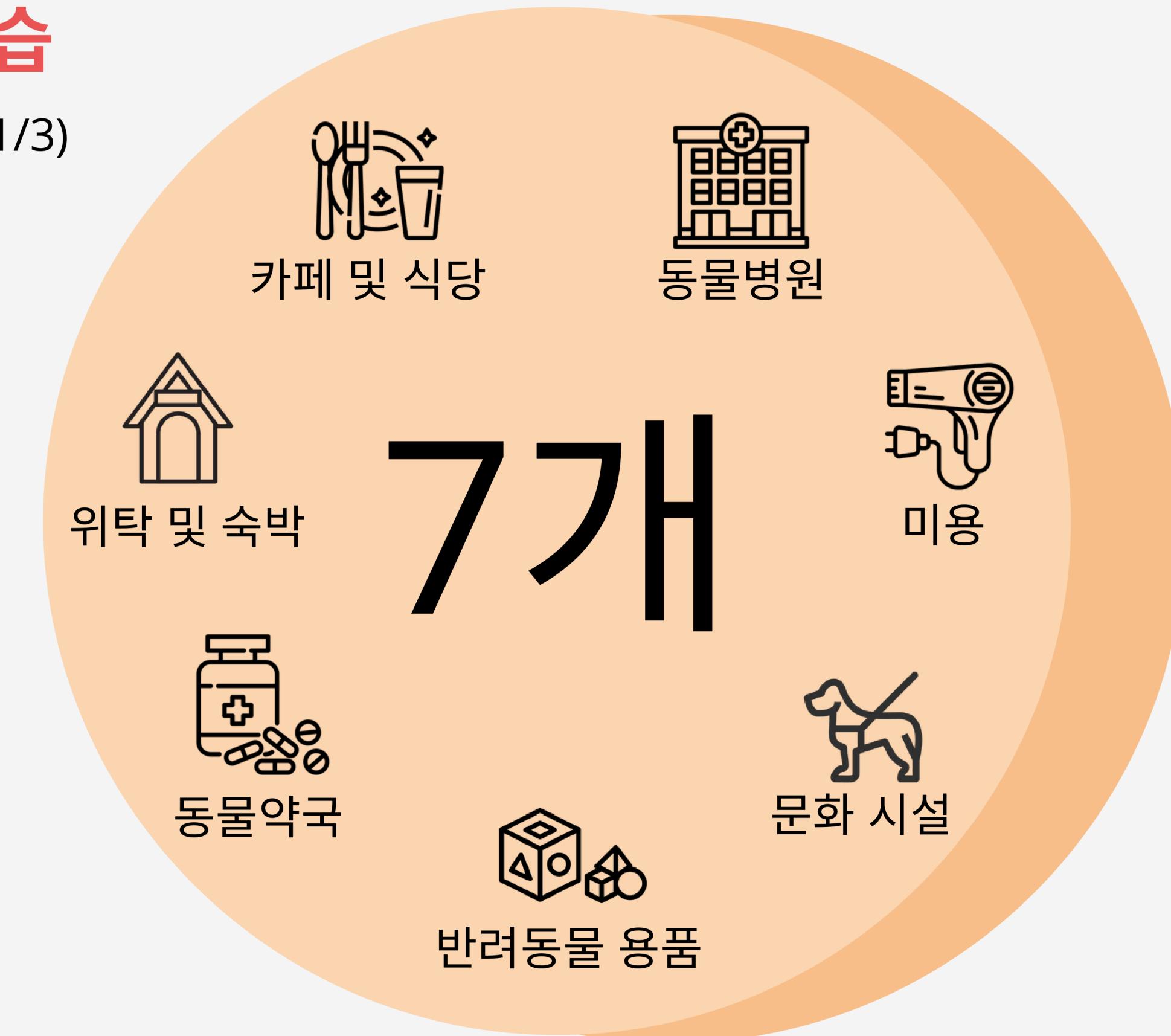
4. 이용 시설

- ★ 청결도
- ★ 맛
- ★ 분위기

# | 자연어 처리

## 4-2) 비지도 학습

- 시설분류별 WordCloud (1/3)

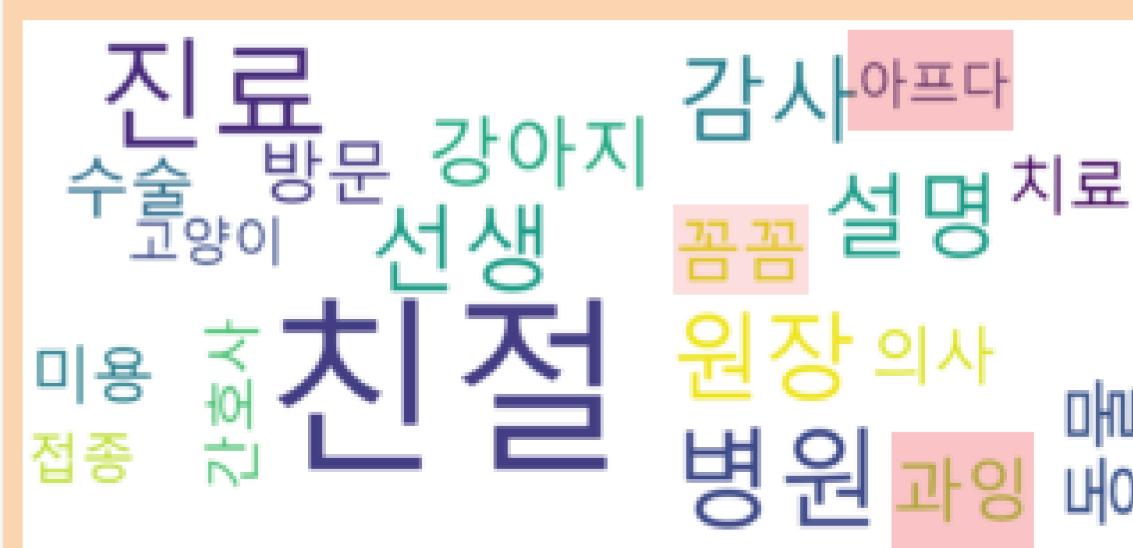


# 자연어 처리

## 4-2) 비지도 학습

- 시설분류별 WordCloud (2/3)

- Topic modeling
- 의료 및 케어 서비스
- 고객서비스 상호작용
- 제품관련
- 이용 시설



### 동물병원

- 전문성 보여주기
- '과잉' 진료 개선
- > 구체적인 진료내역



### 동물약국

- 올바른 복약지도와 약의 부작용에 대해 설명
- 사용자의 위치기반 서비스



### 미용

- 고객의 신뢰도 중점
- 미용 전후 사진 비교

# 자연어 처리

## 4-2) 비지도 학습

- 시설분류별 WordCloud (3/3)

- Topic modeling
  - 의료 및 케어 서비스
  - 고객서비스 상호작용
  - 제품관련
  - 이용 시설



문화시설



- 문화시설을 이용하는 다양한 목적  
> 사용자의 선호를 바탕으로 맞춤형 추천을 제공



위탁 및 숙박



- '시설'의 청결함 강조
- 가족 단위 고객을 타깃으로 한 프로모션



카페 및 식당



- 분위기 및 인테리어 강조
- 반려동물 동반 고객 유치
- 반려동물 메뉴



반려동물 용품



- 제품의 품질과 만족도
  - 식품 - 기호성 강조
  - 의류, 악세사리 - 디자인 강조

# 자연어 처리

## 4-3) 지도 학습

- 긍정, 부정 데이터 -> 클래스 균형화

```
label
1    6049
-1   6049
Name: count, dtype: int64
```

- 모델 생성 (LSTM, GRU, RNN)

```
# LSTM 모델 구성
model = Sequential([
    Embedding(input_dim=len(tokenizer.word_index) + 1, output_dim=128),
    LSTM(units=60),
    Dense(1, activation='sigmoid')
])
```

- 벡터화 : 임베딩(Embedding)
- 하이퍼 파라미터 : 각 모델 동일

- 학습 데이터 분할



```
# 데이터 분할
X_train, X_temp, y_train, y_temp = train_test_split(data, labels, test_size=0.2, random_state=42)
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=0.5, random_state=42)
```

# 자연어 처리

## 4-3) 지도 학습

- LSTM 성능 테스트 - 혼동 행렬 (confusion matrix)



$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

	precision (정밀도)	recall (재현율)	f1-score (F1 점수)	support
부정	0.95	0.97	0.96	609
긍정	0.97	0.95	0.96	601

- 정밀도 : 모델이 True로 예측한 것 중 실제 True인 비율
- 재현율 : 실제 True인 것 중 모델이 True로 올바르게 예측한 비율
- F1 점수 : 정밀도와 재현율의 평균



정확도 : 96%

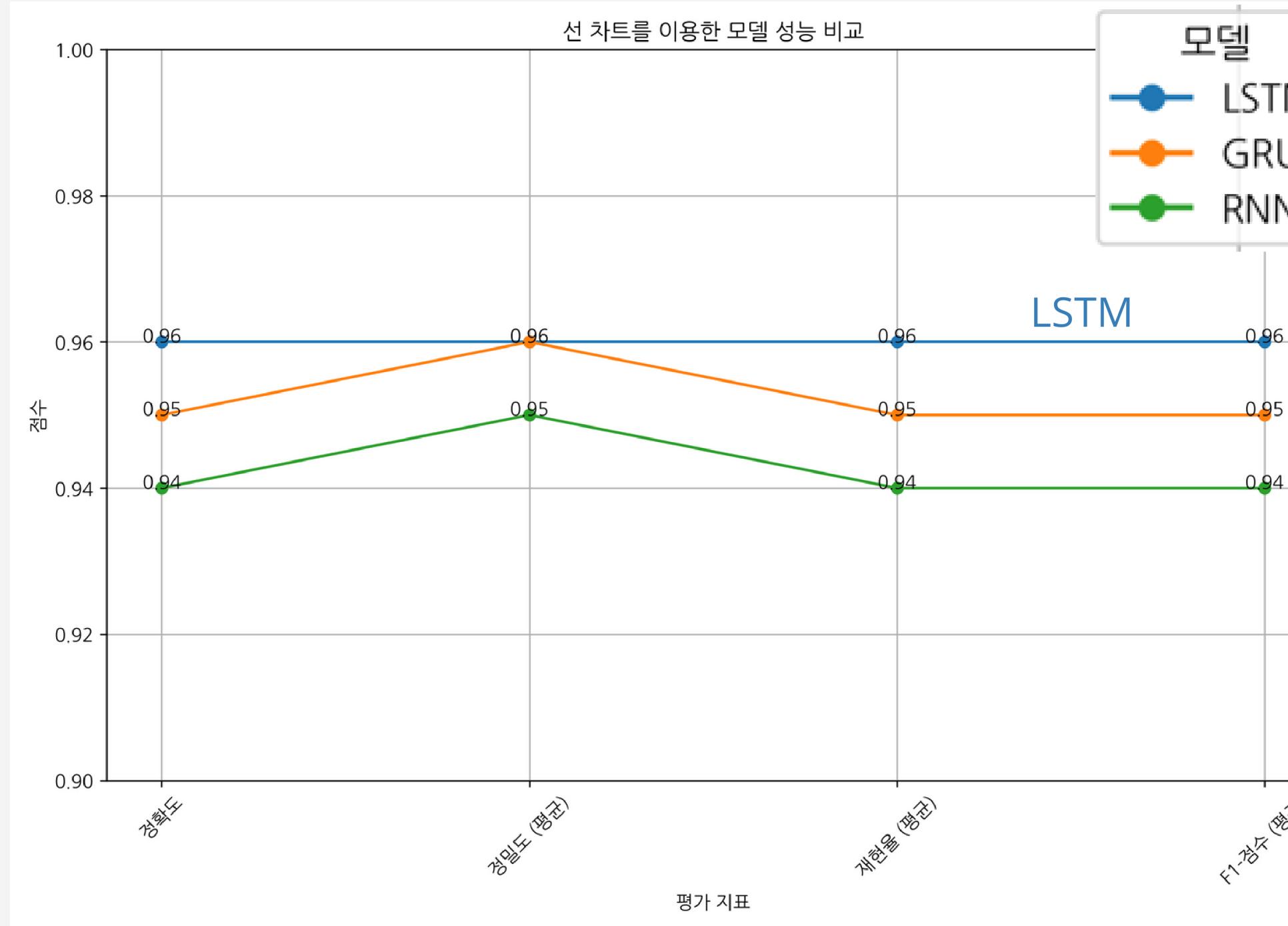
accuracy

0.96

# 자연어 처리

## 4-3) 지도 학습

- 선 차트를 이용한 모델 성능 비교

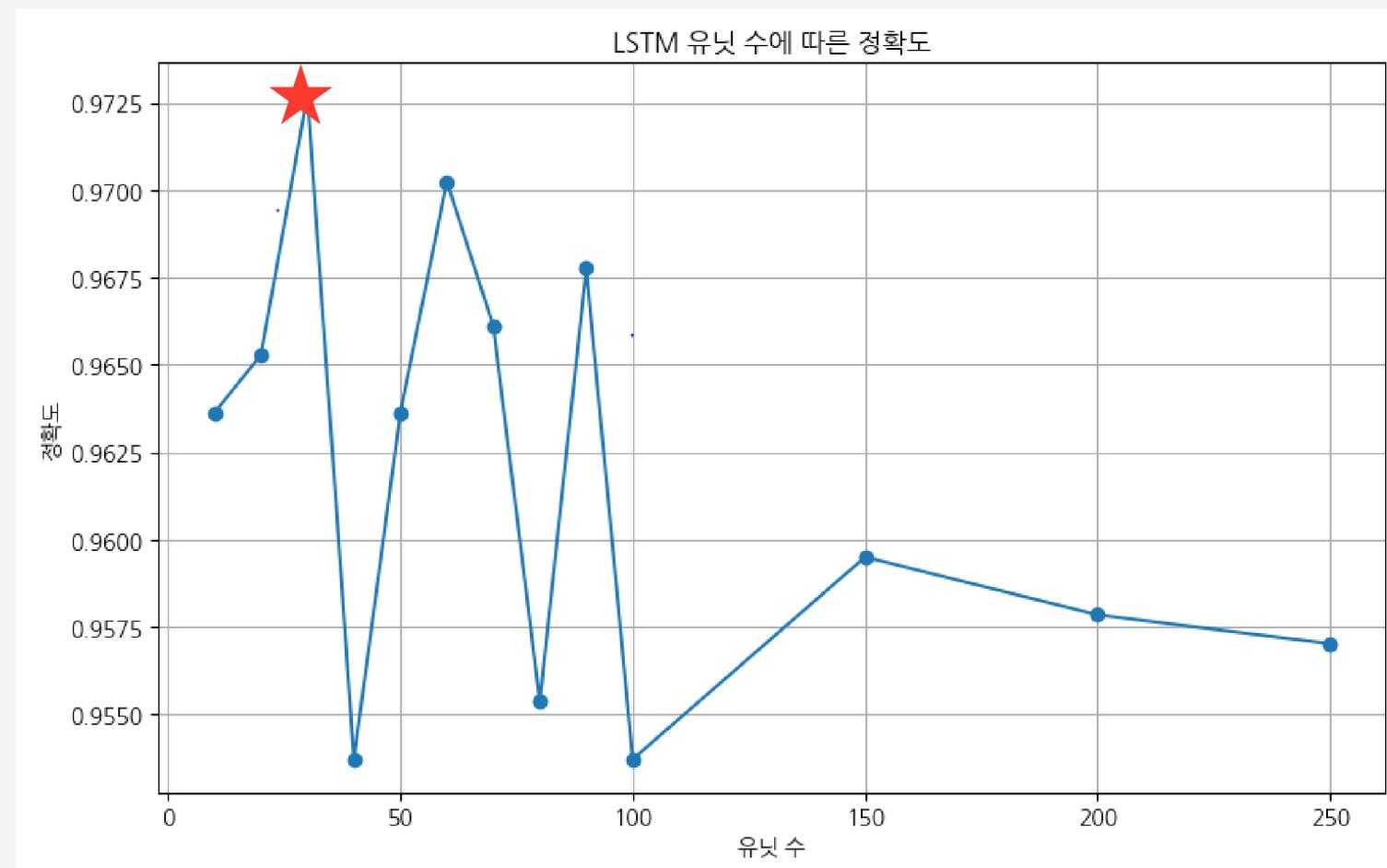


3가지 모델 중 LSTM이 성능이 좋았다!

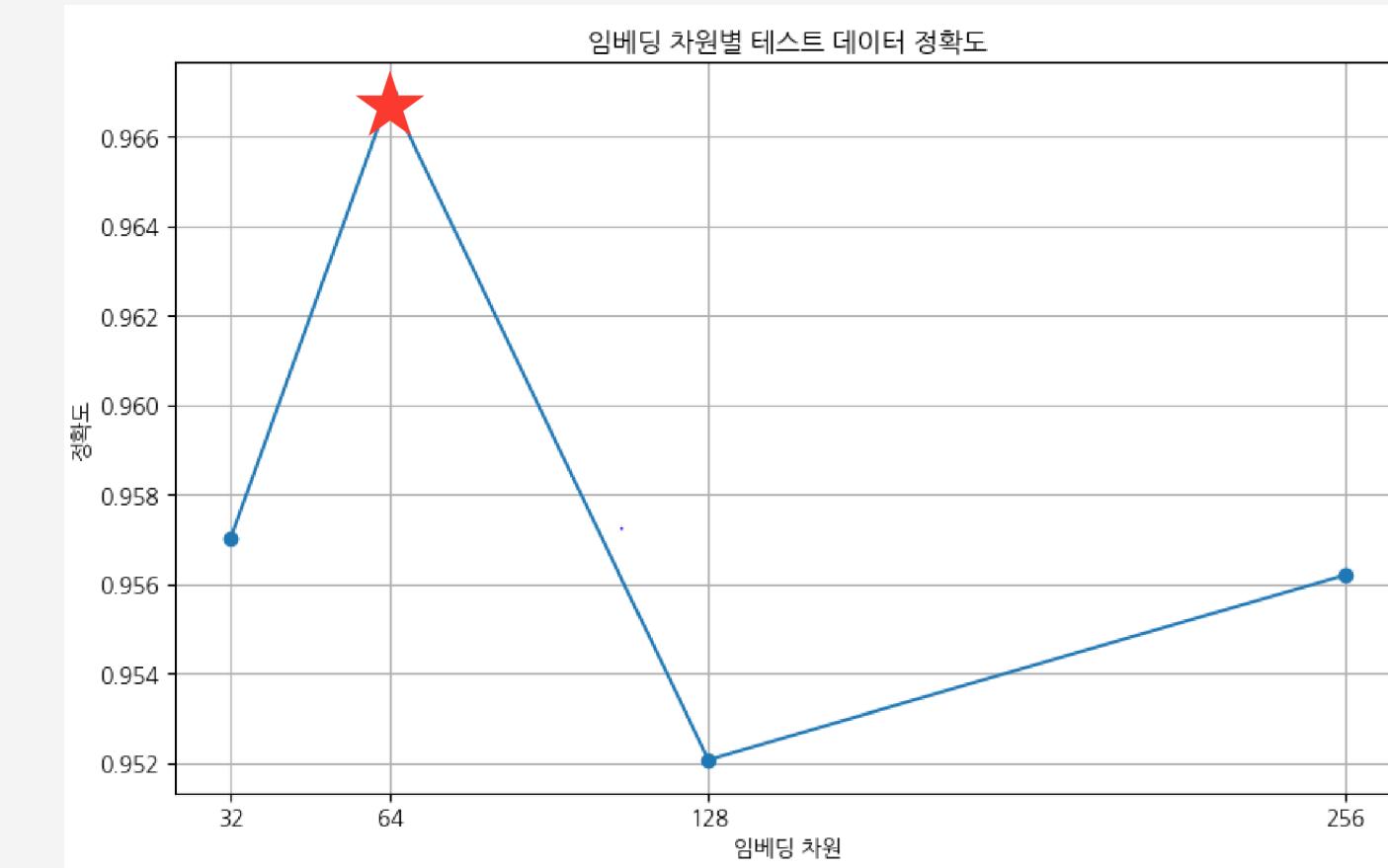
# 자연어 처리

## 4-3) 지도 학습

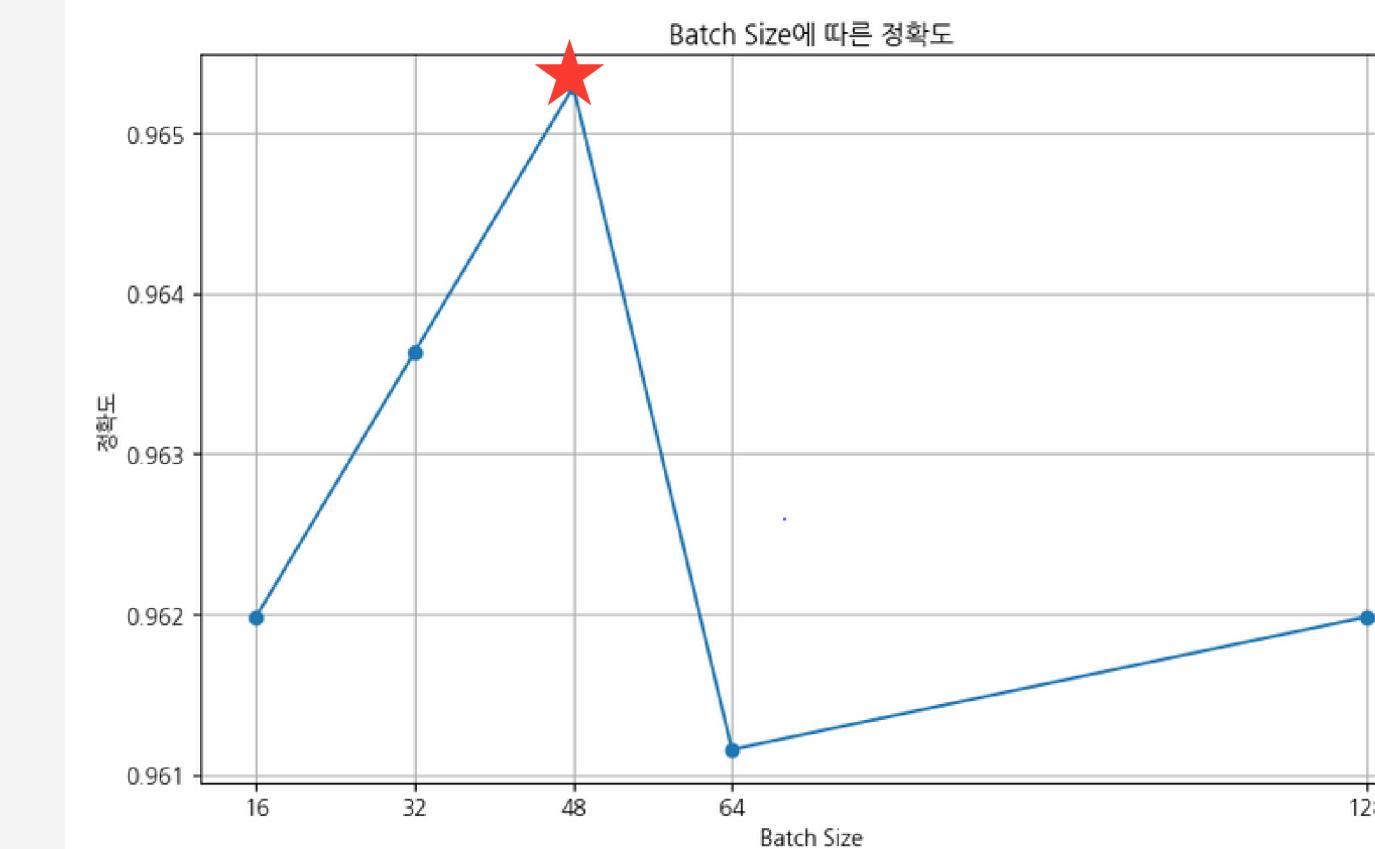
- LSTM 하이퍼 파라미터 성능 측정



- 유닛(units)  
LSTM(units=30),



- 임베딩  
(Embedding)  
output\_dim=64,

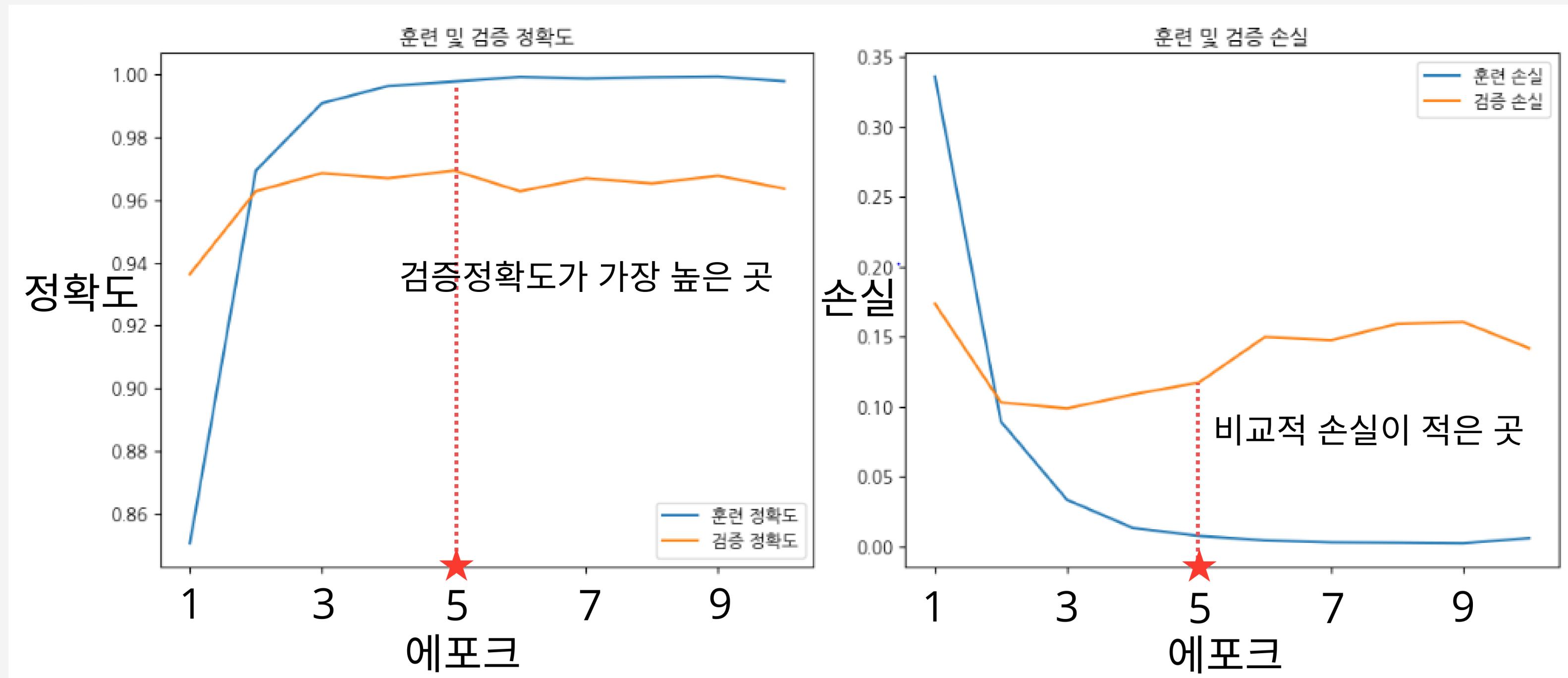


- Batch Size  
batch\_size=48,

# 자연어 처리

## 4-3) 지도 학습

- LSTM 에포크(Epoch) 성능 측정



# 자연어 처리

## 4-3) 지도 학습

- 텍스트 감성 분석을 위한 예측(prediction)

```
# 모델 저장  
# 네이티브 캐라스 포맷 사용  
model_save_path = "mecab_lstm_model.keras"  
model.save(model_save_path)
```

```
# 텍스트 출력  
for i, prediction in enumerate(predictions):  
    sentiment = "긍정" if prediction[0] > 0.5 else "부정"  
    percent = f"{prediction[0] * 100:.2f}%" if sentiment == "긍정" else f"{{(1 - prediction[0]) * 100:.2f}}%"  
    print(f"텍스트: #{new_texts[i]}")  
    print(f"{sentiment} 확률: {percent}\n")
```

텍스트: "이 동물시설은 정말 깨끗하고 관리가 잘 되어 있습니다. 직원들도 매우 친절하고 좋은 경험을 하였습니다."  
긍정 확률: 99.97%

텍스트: "동물들의 상태가 아주 좋아보였고, 시설도 청결하고 좋았습니다. 다음에도 꼭 방문하고 싶은 곳입니다."  
긍정 확률: 99.74%

텍스트: "이곳은 정말 좋은 동물시설입니다. 동물들이 행복해 보였고, 시설도 잘 갖추어져 있습니다. 만족합니다."  
긍정 확률: 99.96%

텍스트: "직원분들이 너무 친절하셨고, 동물들도 건강하게 잘 키워지고 있어서 기쁩니다."  
긍정 확률: 99.99%

텍스트: "동물들의 복지가 잘 지켜지는 곳이라 더 방문하고 싶습니다. 직원분들도 친절합니다."  
긍정 확률: 99.96%

긍정

텍스트: "직원들이 불친절하고 시설도 청결하지 않았습니다. 이렇게 관리가 안되는 곳은 처음 봤습니다."  
부정 확률: 99.72%

텍스트: "시설이 너무 낡아서 동물들이 안전하게 지내기 힘들었습니다. 관리가 잘 되지 않은 것 같습니다."  
부정 확률: 99.85%

텍스트: "동물들의 상태가 안 좋아보였고, 시설도 관리가 잘 되어 있지 않았습니다. 더 이상 방문하고 싶지 않습니다."  
부정 확률: 60.14%

텍스트: "이 동물시설은 동물들의 복지를 고려하지 않고 관리가 되어 있지 않았습니다. 절대로 다시 방문하고 싶지 않습니다."  
부정 확률: 99.32%

텍스트: "직원들이 무례하고 동물들의 상태도 좋지 않았습니다. 이렇게 관리가 안 되는 곳은 처음이었습니다."  
부정 확률: 75.05%

텍스트: "이곳은 동물들의 복지가 전혀 고려되지 않은 곳 같았습니다. 시설도 관리가 되어 있지 않아서 실망스러웠습니다."  
부정 확률: 99.98%

부정

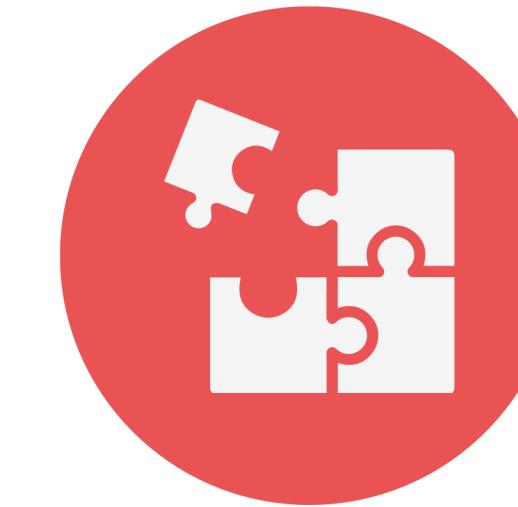
# | 프로젝트 성과 및 느낀점

## 5-1) 프로젝트 성과



### 비지도 학습 모델

사용자의 감정을 이용하여  
시설별 시장의 트렌드와  
소비자의 관심사를 알게 됨

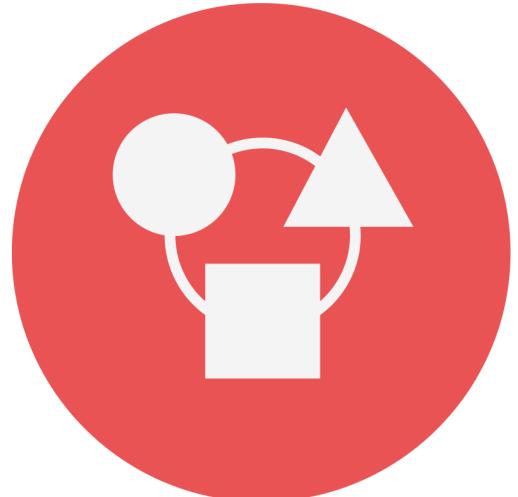


### 지도 학습 모델

비정형 데이터인 SNS 리뷰에  
대한 긍정 / 부정 판단

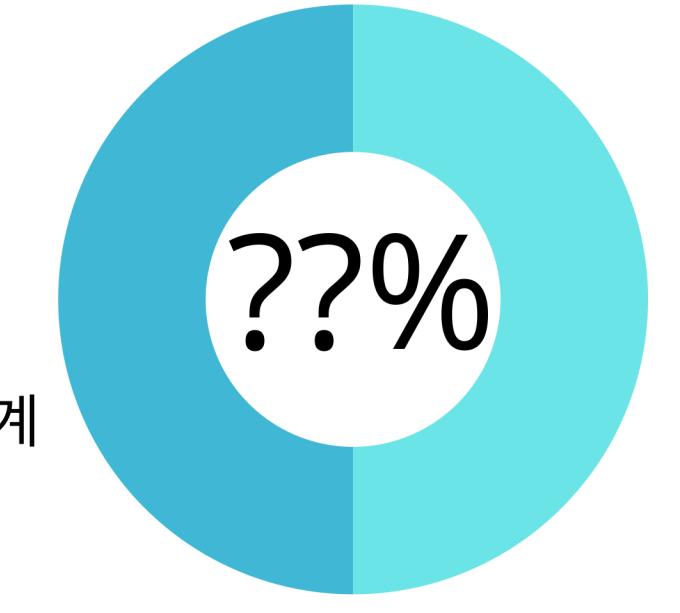
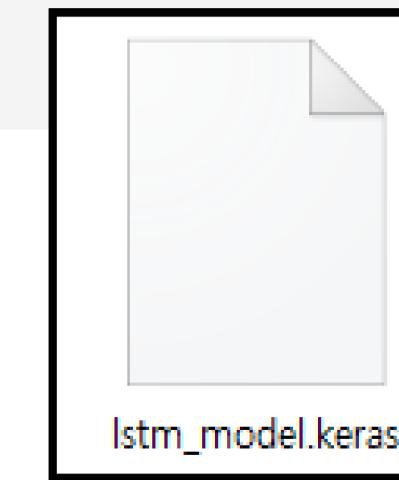
# | 프로젝트 성과 및 느낀점

## 5-2) 향후 계획



### 데이터 유효성 검사기

생성된 `lstm_model.keras`로  
파이널 프로젝트 커뮤니티 댓글에  
데이터 유효성 검사기로 추가



### 긍정/부정 통계

파이널 프로젝트  
시설정보 게시글 긍정/부정 통계



### 맞춤형 시설 추천

사용자의 커뮤니티 내 검색 기록을  
바탕으로 맞춤형 시설 추천  
(위치기반 시설 추천 알고리즘 개발)

# 프로젝트 성과 및 느낀점

## 5-3) 느낀점



- 데이터 수집시 XPath경로 이상함
- 한글 자연어 처리의 어려움  
(형태소 분석 도구, 감성 사전)
- 파일 버전 관리의 실패
- 많은 종류의 library 경험
- 팀원들과 소통의 중요성



- 자연어 처리는 정확한 답이 없어 인상적이었다
- 고객의 감정을 읽을 수 있고, 니즈를 파악한 서비스는 고객의 만족도가 높을 수 밖에 없다



- 한글 자연어 처리의 어려움
- 하이퍼 파라미터 테스트 시 결과가 일정하지 않음

질의 응답



감사합니다

