

CODESTATES

---

# VACCINE PASS SENTIMENT ANALYSIS

---

DATA SCIENTIST SONG MIN HYE



조선비즈 PICK | 1일 전 | 네이버뉴스

## "비과학적" vs "백신 독려"... 해외서도 찾기 힘든 백화점 방역패스

美, 유럽 등 방역패스에 백화점 거의 없어 프랑스는 면적 2만㎡ 넘을 때만 의무 적용 "대규모 전포 화기 작 돼 간염 가능성 작아" 10일부터 면적이 3000㎡가 넘는



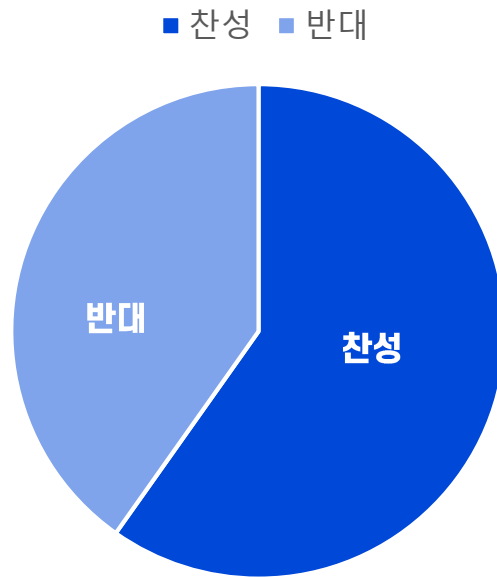
# “방역 패스, 사람들이 어떻게 생각 할까?”



우려(42.8%) △여전한 돌파감염 가능성(35.2%) △접종 대상자의 고위험군 제한 필요성(4.3%) 등을 이유로 꼽았다. 백신패스에 대한 의견은 6대 4 비율로 찬성이 높았다. 식당이나 카페, 실내체육시...  
한국일보 | 14시간전 | 다음뉴스

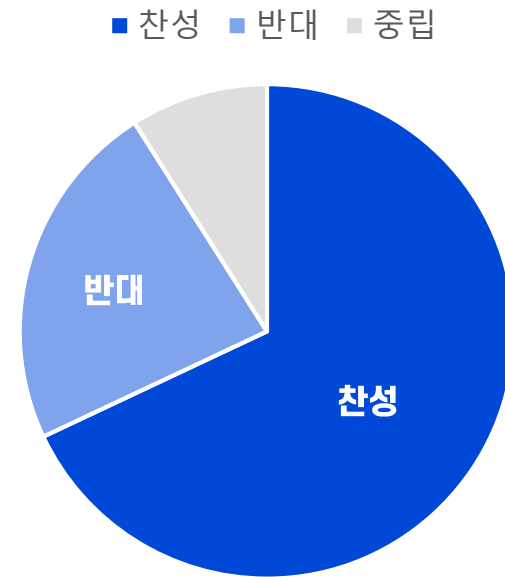
# HYPOTHESIS

## 안산시 방역 패스 관련 찬반 그래프



백신 패스·백신 3차 접종...안산시민에게 물어보니, 한국일보

## 서울시 방역 패스 관련 찬반 그래프

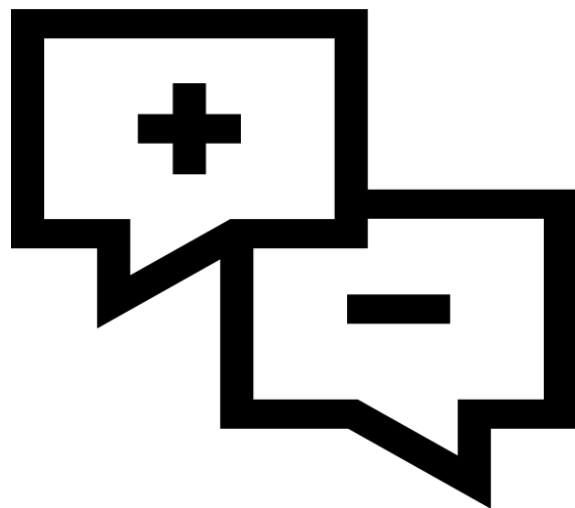


서울시민 68% '방역 패스' 찬성...76% "K-방역 모델이 대 확산 방지 효과 있었다", 세계일보

## HYPOTHESIS

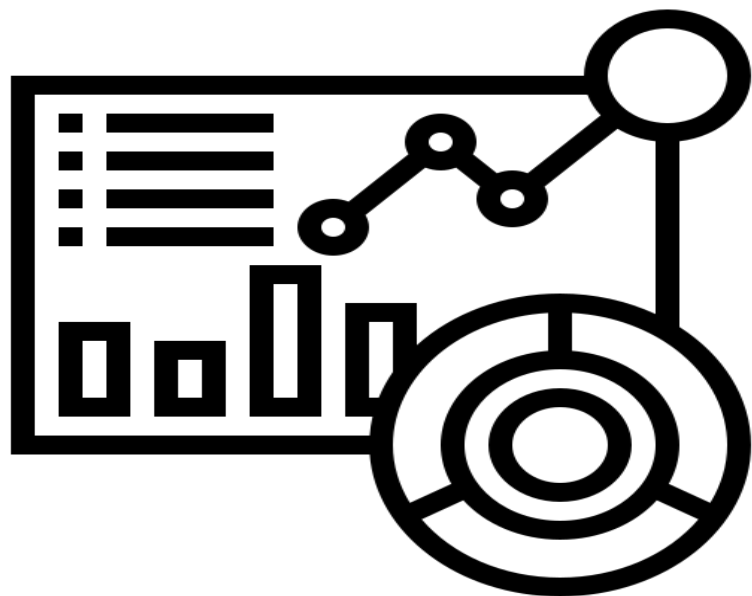
### 가설 설정

---



**“실제 여론은 설문조사의 결과와 다른 방향일 것이다.”**

### 최종 분석 목표



“후에 새로운 대안 마련 시, 데이터로 사용 가능하도록”

# DATASET

## DATASET INFORMATION

- Train Labeling Data (1208 \* 2)
- Total Data

## USING CRAWLER

- NAVER 뉴스 기사 댓글 수집
- 수만휘, Orbi 청소년 커뮤니티 댓글 수집

## LABELING TRAIN

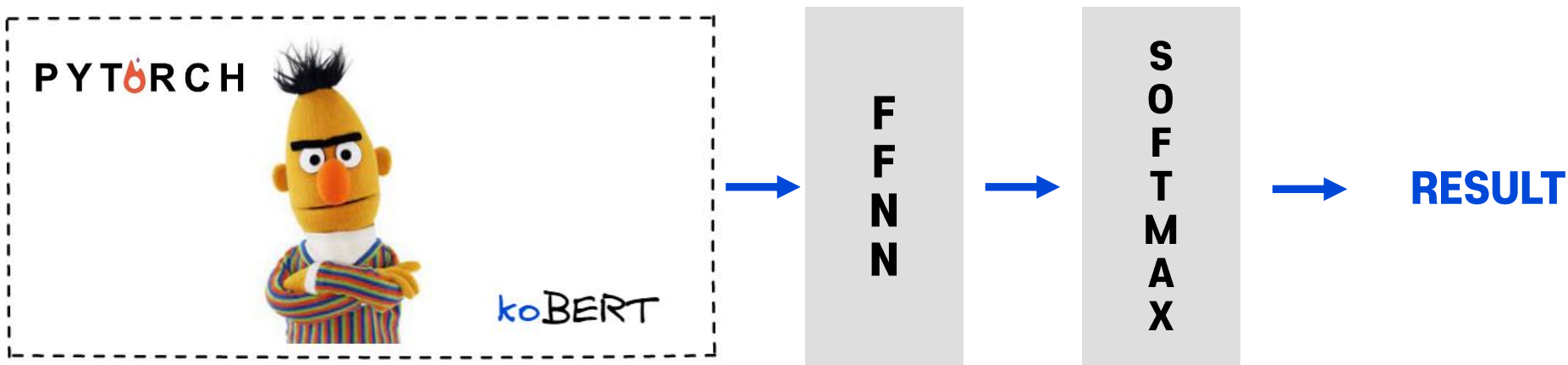
- 수작업으로 소수의 데이터 라벨링
- 긍정 : 1 / 부정 : 2 / 중립 : 0



## MODEL

### 선택한 모델 : KoBERT

Korean BERT (Bidirectional Encoder Representations from Transformers)



### Reason

- Pre-trained Model
- 기존 BERT의 한국어 성능 한계를 극복하기 위해 개발된 모델
- 문맥의 의미 파악 가능

# RESULT

## PERFORMANCE

```
100% ██████████ 16/16 [00:07<00:00, 2.42it/s]
epoch 1 batch id 1 loss 1.0529940128326416 train acc 0.484375
epoch 1 train acc 0.6794621394230769
100% ██████████ 4/4 [00:00<00:00, 5.87it/s]
epoch 1 test acc 0.8594460227272727
100% ██████████ 16/16 [00:07<00:00, 2.42it/s]
epoch 2 batch id 1 loss 0.5980338454246521 train acc 0.84375
epoch 2 train acc 0.8357121394230769
100% ██████████ 4/4 [00:00<00:00, 6.47it/s]
epoch 2 test acc 0.8594460227272727
100% ██████████ 16/16 [00:07<00:00, 2.40it/s]
epoch 3 batch id 1 loss 0.5614994168281555 train acc 0.84375
epoch 3 train acc 0.8357121394230769
```

test acc 0.8718039772727273

```
100% ██████████ 4/4 [00:00<00:00, 6.42it/s]
epoch 4 test acc 0.8594460227272727
100% ██████████ 16/16 [00:07<00:00, 2.41it/s]
epoch 5 batch id 1 loss 0.559780478477478 train acc 0.84375
epoch 5 train acc 0.8357121394230769
100% ██████████ 4/4 [00:00<00:00, 6.42it/s]
epoch 5 test acc 0.8594460227272727
100% ██████████ 16/16 [00:07<00:00, 2.41it/s]
epoch 6 batch id 1 loss 0.5625834465026855 train acc 0.84375
epoch 6 train acc 0.8357121394230769
100% ██████████ 4/4 [00:00<00:00, 6.36it/s]
epoch 6 test acc 0.8594460227272727
100% ██████████ 16/16 [00:07<00:00, 2.42it/s]
epoch 7 batch id 1 loss 0.500923752784729 train acc 0.84375
epoch 7 train acc 0.8473557692307693
100% ██████████ 4/4 [00:00<00:00, 6.42it/s]
epoch 7 test acc 0.8718039772727273
100% ██████████ 16/16 [00:07<00:00, 2.40it/s]
```

### 모델 성능 평가

Channel Level :  $100/3 = 33.3333$   
>> 0.33333

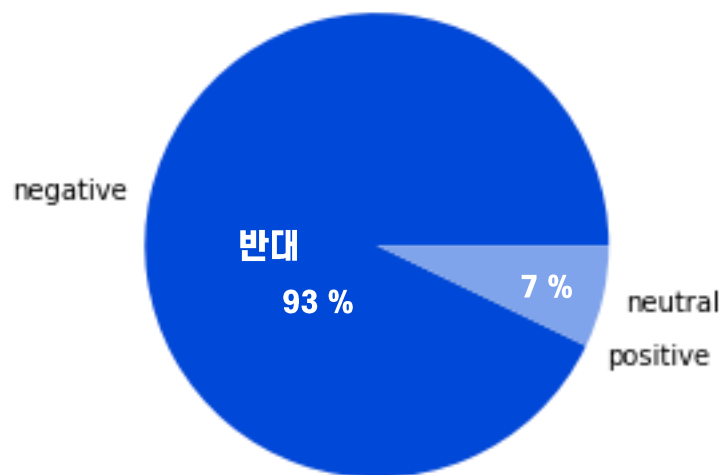
Best Test Accuracy : 0.8718

Best Train Accuracy : 0.8473



## RESULT

### PERFORMANCE



#### 가설 검정 결과

예측 검정을 해 보았을 때,  
방역 패스에 대한 **부정적인 감정이**  
긍정적인 감정보다 **압도적으로 우세하였다**

**긍정의 예측이 거의 되지 않았다는 점에서, 성능을 제대로 발휘하는 모델이라고 보기 어렵다**

# LIMIT

## Data Part

- 모집단의 한계, 여론을 대변했다고 보기 어려움
- 데이터의 불 균형성
- Train Label Data의 부족

## Model Part

- 용량의 한계
- 시간의 한계

# RETROSPECTIVE



CODESTATES

---

**THANK YOU  
FOR  
LISTENING**

---