



Drug Recommendation



21. 09. 08 (수)

김준기 배홍직 이상목 이준병 이현준

발표자 : 이상목



Contents

2학기 전체 일정

01

모델 개선
챗봇 개발

모델 개선 계획

02

BERT를 이용한 문장 분류

챗봇 개발 계획

03

사용자가 사용할 챗봇 개발



01

「 2학기 전체 계획 」

- 모델 개선
- 챗봇 개발

일정



	9월				10월				11월				12월			
	1주차	2주차	3주차	4주차	1주차	2주차	3주차	4주차	1주차	2주차	3주차	4주차	1주차	2주차	3주차	4주차
계획 수립																
Model																
bert 공부																
model 구현																
model 테스트																
ChatBot																
scenario 제작																
Backend																
Frontend																
문서화																

9월 10월에는 모델 개발에 집중하고, 11월 12월은 배포가 가능할 수준으로 구현해보려 합니다.



02

「
모델 개선 계획
」

9 ~ 10월

계획

- input 정의

학습	배포
데이터셋의 리뷰 데이터	사용자가 챗봇과 상담한 내용

- 전체 모델 구조

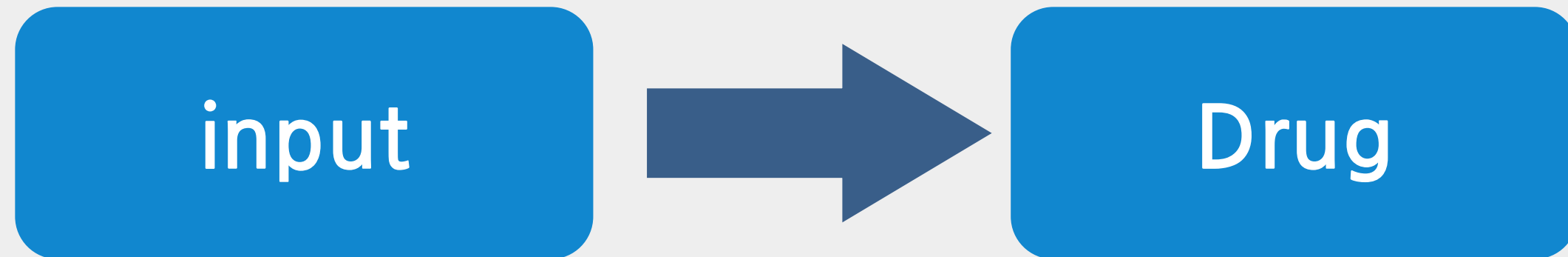


- review 데이터를 사용한 문장 임베딩
- 임베딩 된 문장을 input으로 하여 condition 분류
- output으로 나온 condition 벡터를 이용해 drugname 예측
 - > 1학기에는 단순 rating을 기준으로 상위에 위치한 약을 추천하였지만, 2학기에는 ML 적용 예정

모델 구조

input을 바로 약물로 분류하지 않고, 조건을 거치는 이유

- input을 통해 바로 약물 분류



데이터 셋은 약에 대한 리뷰 데이터로, 부정적인(rating이 낮은) 리뷰가 혼재되어 있음

- 해당 데이터의 경우 정답(추천해줄 약물)으로 분류하기는 위험(좋지 않은 약 추천 해줄 가능성)하다고 판단
왜 모델이 이 약물을 추천해주는지에 대한 이유를 알 수가 없음

- input을 통해 조건을 예측 후, 약물 분류

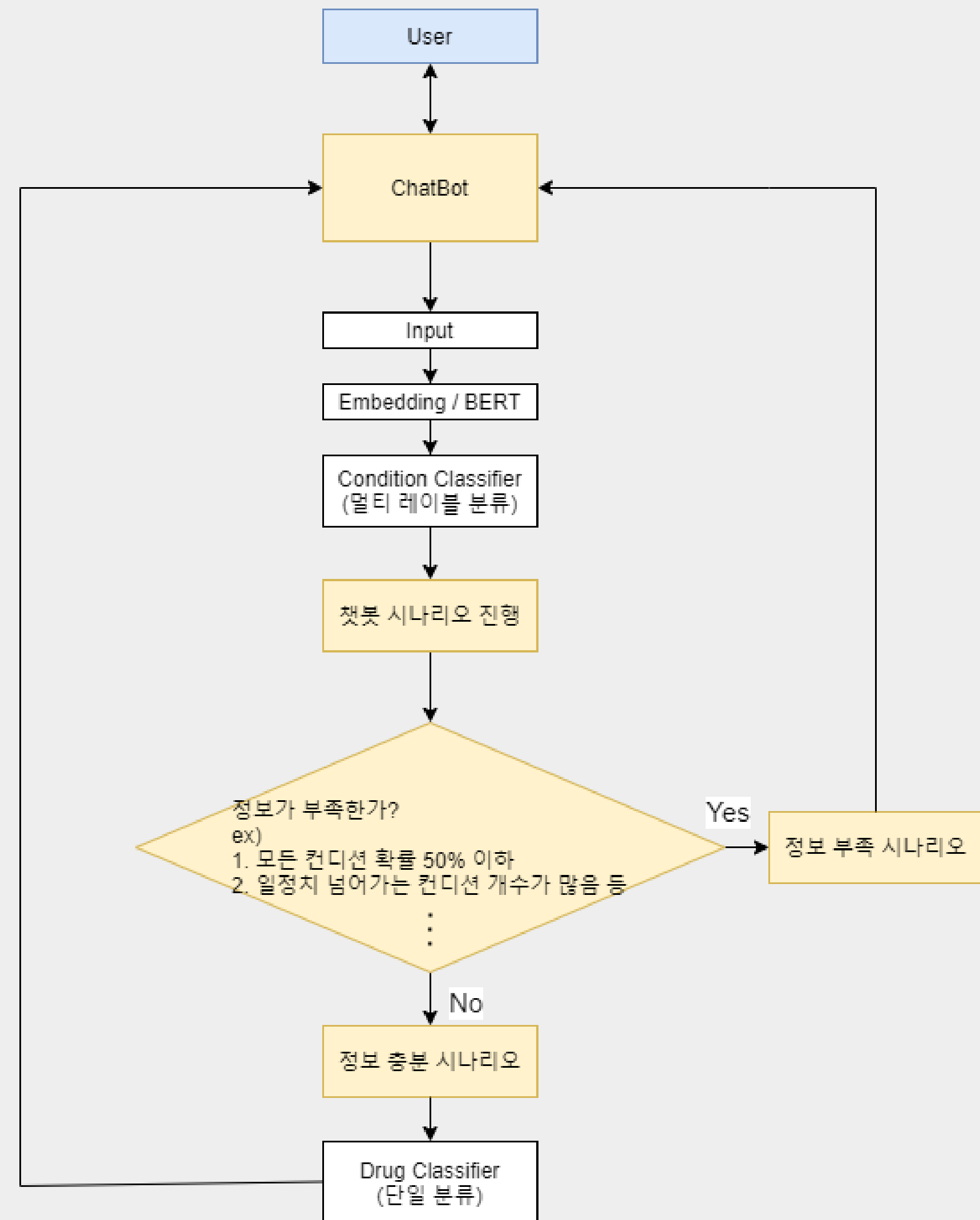


부정적인 리뷰 역시 조건을 분류하는 데 활용하여도 문제가 되지 않음

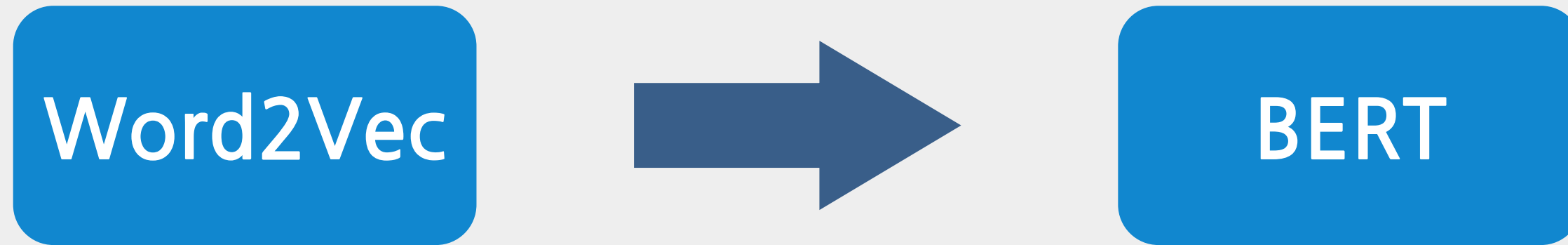
모델의 중간 과정으로 조건을 예측하기에, 모델이 왜 이 약을 추천하였는지 유추 가능

세부 모델 구조

- 유저는 챗봇과 소통하며, 챗봇은 유저의 인풋을 통해 다음과 같은 시나리오로 모델을 구동
- 현재 생각중인 시나리오로는 정보가 부족한지 중간에 확인해보며 유저와 소통하는 것
정보가 부족하다면 챗봇이 유저에게 추가 정보를 요구하는 형태 구상
- 컨디션 분류기는 사용자가 한 번에 여러 증상을 가지고 있을 수 있다고 생각하여, 멀티 레이블 분류가 가능하도록 함
약 분류기는 예측된 컨디션 확률을 input으로 결과로 하나의 약을 추천하도록 함



BERT



- 대량의 corpus로 학습되어 단어 Embedding 품질 향상
- 양방향 학습을 진행해 문맥을 고려한 단어 벡터 생성
- 학습이 안된 단어가 나오는 OOV 문제 해소

BERT TEST



```
In [82]: f1 = f1_score(test_encoded_datasets['label'], topic, average='macro')
          f1
Out [82]: 0.8119580396137055
```

상위 20개 condition에 대해서
BERT를 review컬럼의 데이터를 이용해 Fine-tuning 한 후
review를 condition으로 분류한 결과
1학기때 만들어 본 모든 모델의 성능을 뛰어넘는
약 **81%**의 F1-score를 보임.

condition 분류

Input

review

Sigmoid

Output

adhd: 0.6
birth control: 0.7
pain: 0.3
depression: 0.9
anxiety: 0.8
...
bipolar disorder: 0.5

review 문장을
적절한 condition으로
분류한다.

사용자에게 condition을 표
시해주는 경우
가장 높은 확률을 가지는
condition을 표시한다.

drugname 예측

Input

Output

adhd: 0.6
birth control: 0.7
pain: 0.3
depression: 0.9
anxiety: 0.8
...
bipolar disorder: 0.5

Softmax

drug

이전 단계의 Output인
condition 분류 벡터를

Input으로 사용해
drugname을 예측한다.

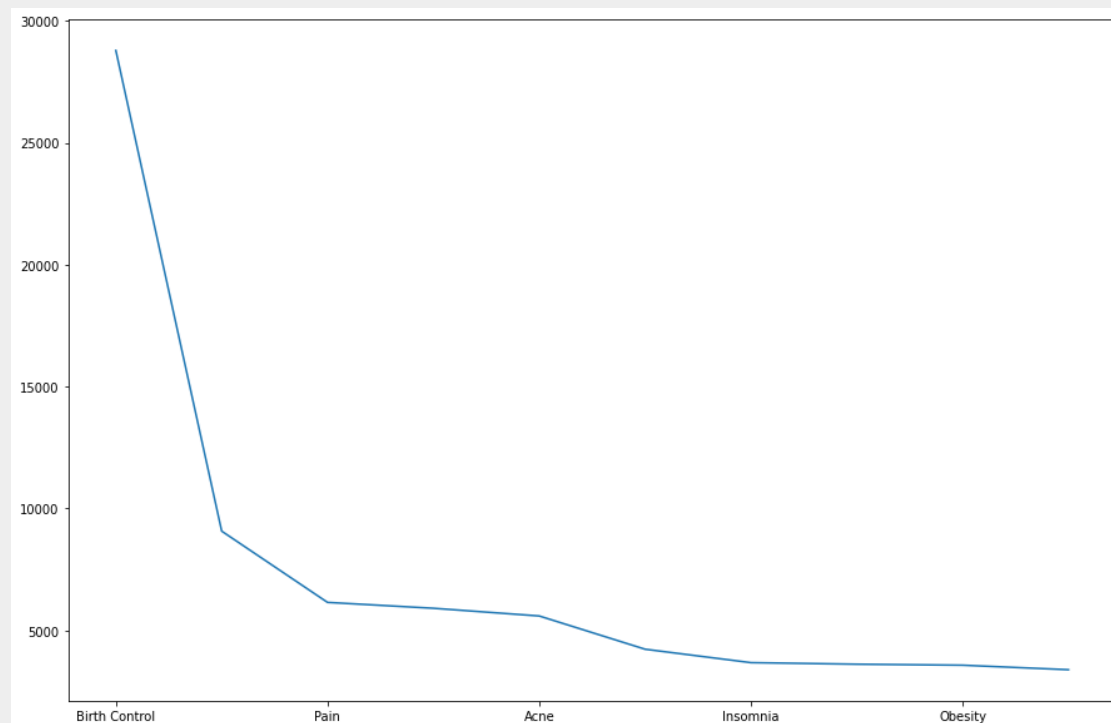
컨디션 압축

데이터 셋에서 일부 유사한 컨디션이 존재하는 것으로 파악

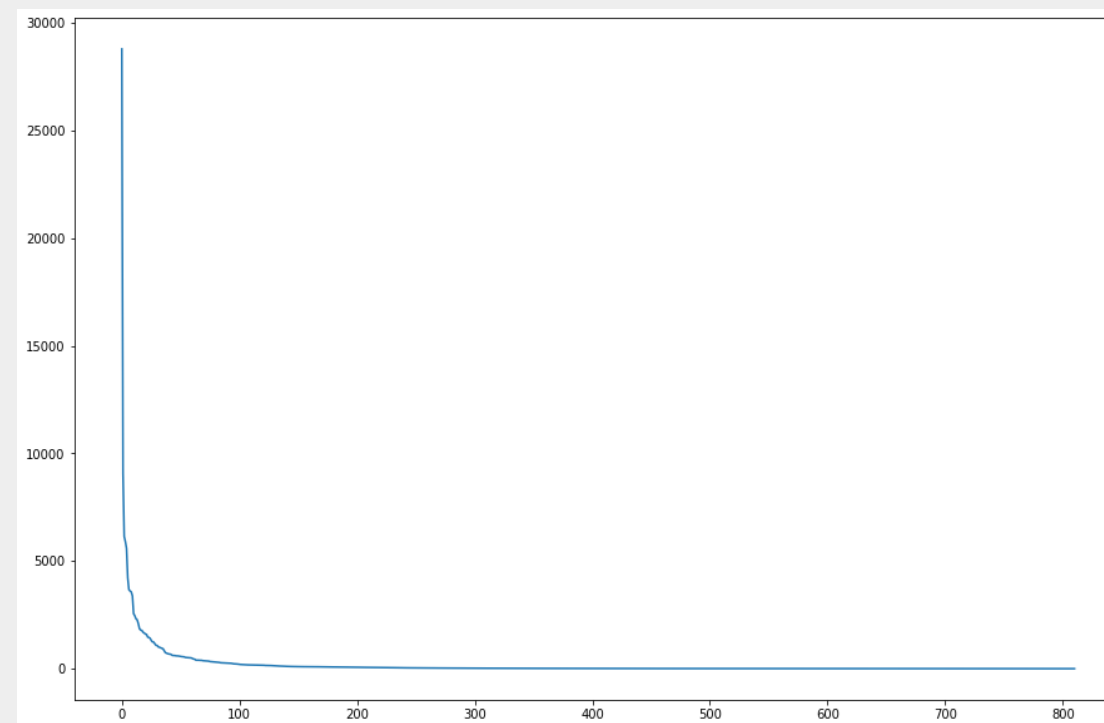
Bronchitis	기관지염
Pneumonia	폐렴

Major Depressive Disorder	우울증
Depression	우울증

상위 20개의 컨디션만 예측할 경우, 버려지는 컨디션이 많아지기에 활용할 방법 고민



상위 10개 컨디션 count



전체 컨디션 count

따라서, 유사한 컨디션을 묶어줄 방법에 대해 고민한 결과,
word2vec를 이용하여 embedding 된 컨디션 벡터들에 대해
K-means를 이용하여 그룹화하여 유사한 컨디션별로 압축을 시도해 보려 합니다.

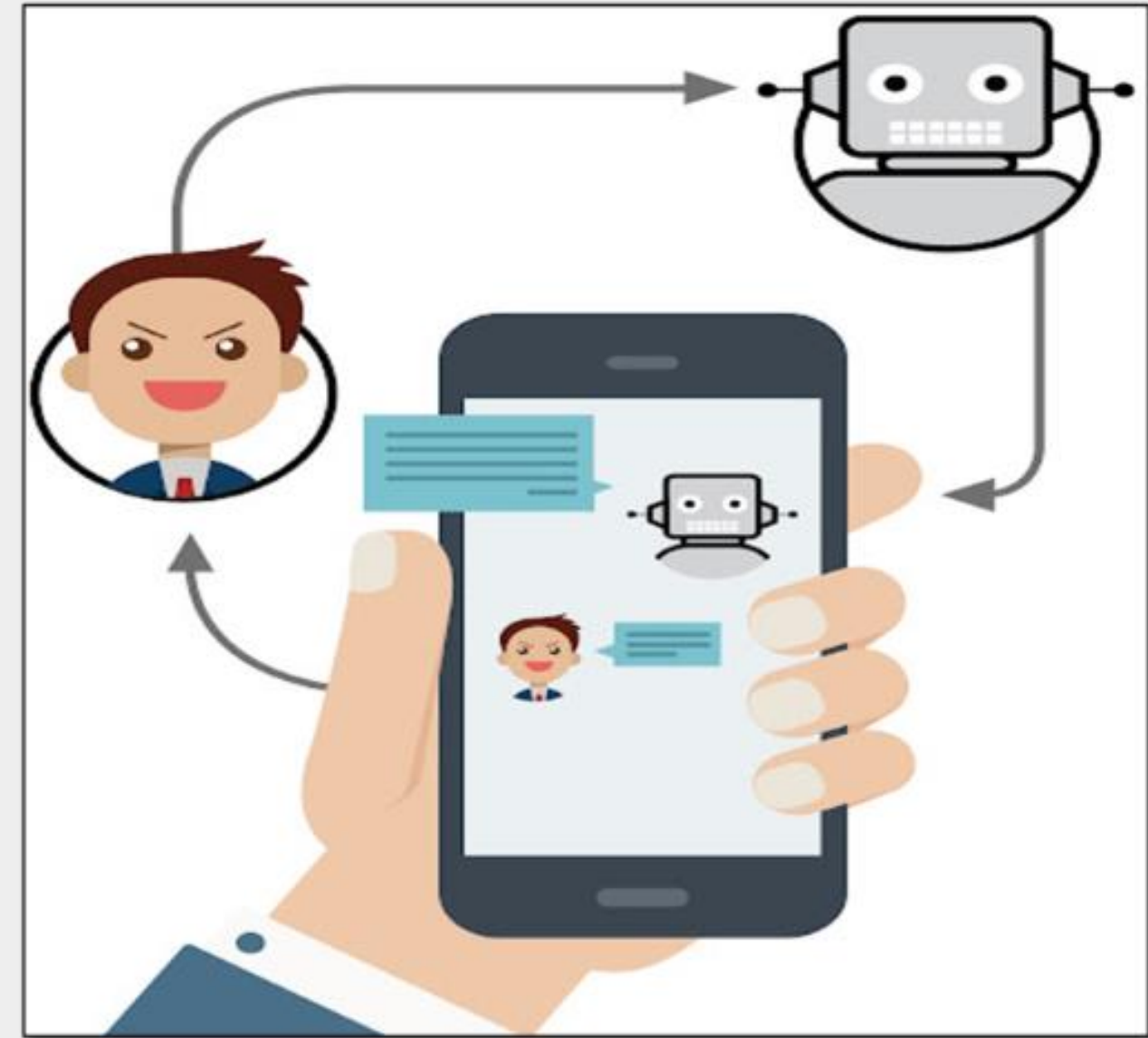
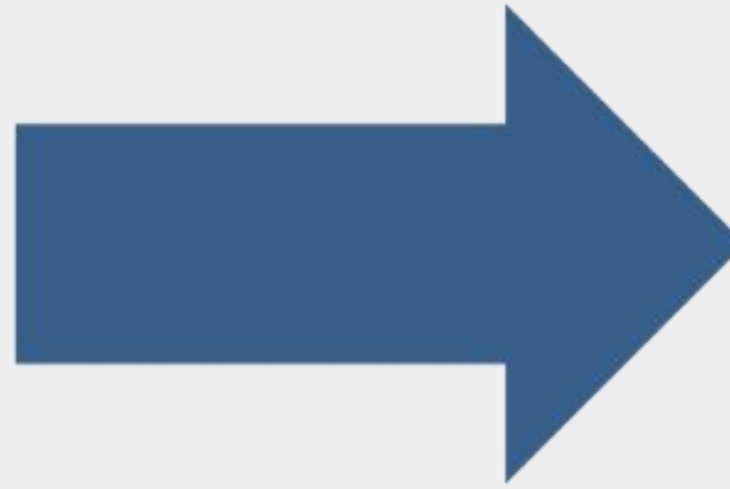
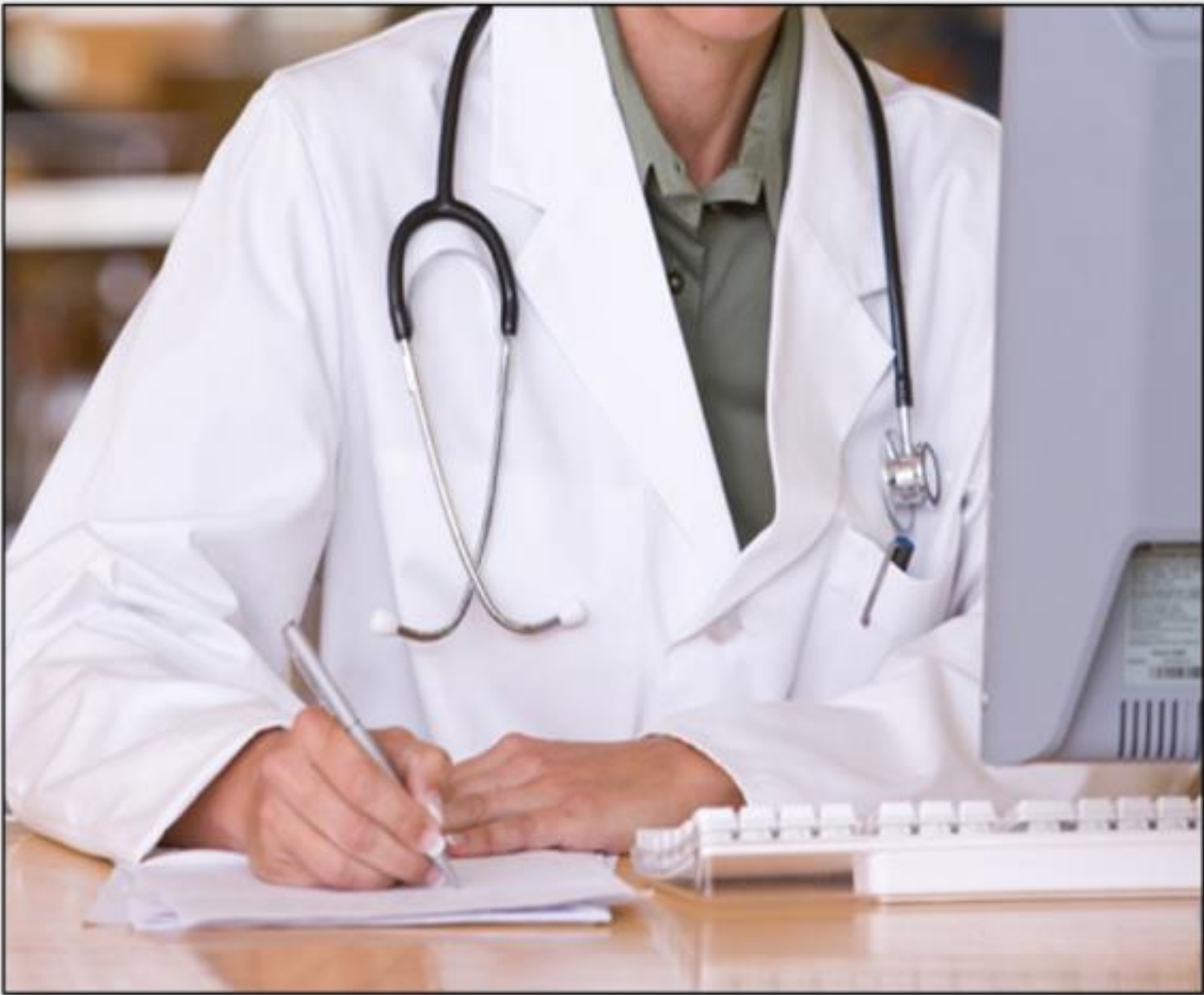


03

「
챗봇 개발 계획」

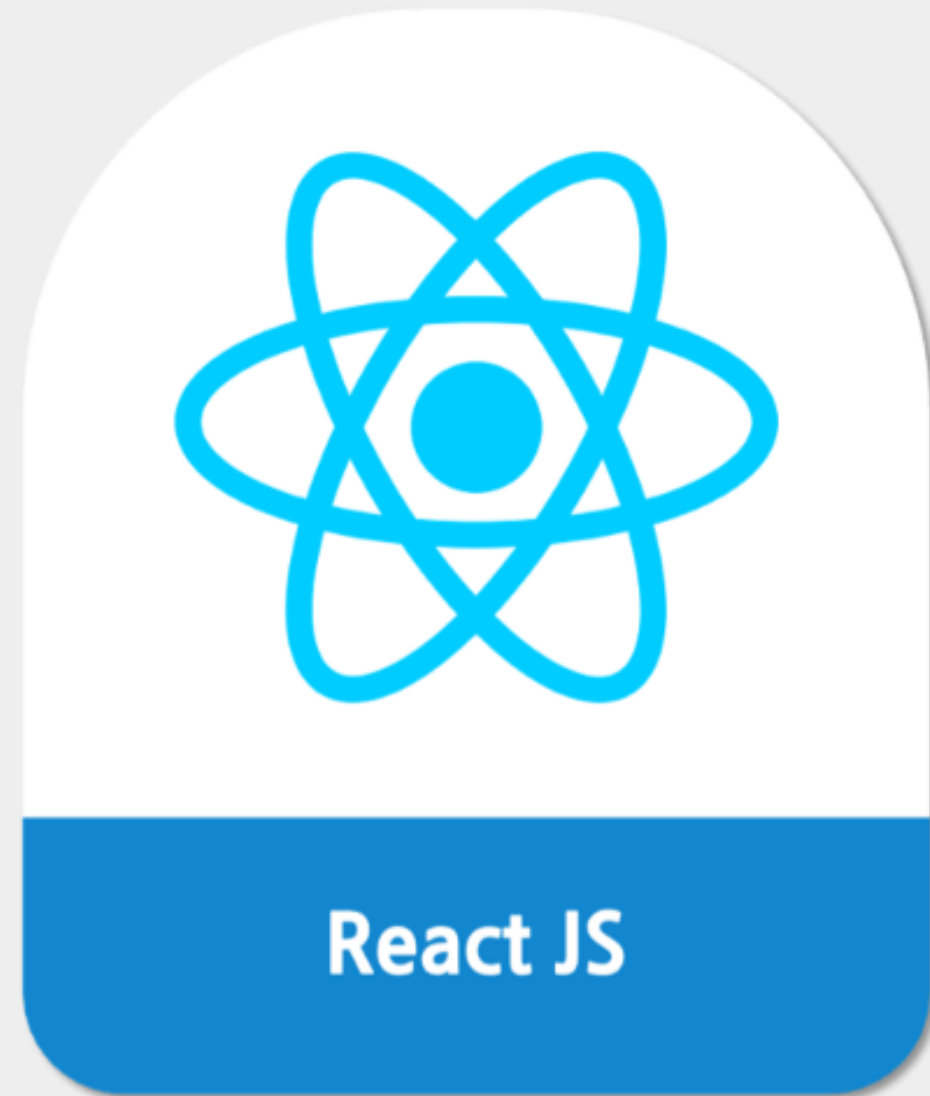
11 ~ 12월

챗봇 개발 계획



- 의사가 condition을 진단할 때, **환자와 대화를** 통해 condition 예측
- 사용자와의 **대화를 위한 챗봇** 인터페이스 사용

챗봇 개발 계획



- 웹, 모바일 챗봇을 통해 **사용자로부터 입력** 받고 Flask 서버로 전송
- Flask 서버로부터 받은 예측된 condition 또는 추천하는 약물을 **사용자에게 출력**

- React 서버로부터 받은 데이터를 전송 받아 개발된 **머신러닝 모델을 통해 예측 수행**
- 예측된 결과와 추천하는 약물을 **React 서버로 전송**