



# 음향 데이터 Covid-19 검출 AI 기반 서비스

한경닷컴 IT교육센터

Team : 누가 기침소리를 내었조  
팀장 : 김영원  
팀원 : 정호영, 김혜원

# Contents



## 1. 프로젝트 개요

- 01. 팀 구성 및 역할
- 02. 기획 의도
- 03. 개발 흐름 및 일정
- 04. 기술 스택
- 05. 시연 동영상



## 2. 인공지능

- 01. 설계 및 구조도
- 02. 데이터 확인 및 전처리
- 03. 모델 (resnet, vgg)
- 04. 히스토리



## 3. 챗봇

- 01. 설계 및 구조도
- 02. 데이터 수집 및 전처리
- 03. Dialogflow 구성



## 4. 플라스크

- 01. 플라스크 설계
- 02. 화면 구성
- 03. 코로나 접종 현황판
- 04. 목소리 유사도
- 05. 코로나 예측 그래프



## 5. 마무리

- 01. 보완할 점
- 02. 느낀 점
- 03. 출처



# Part. 1

## 프로젝트 개요

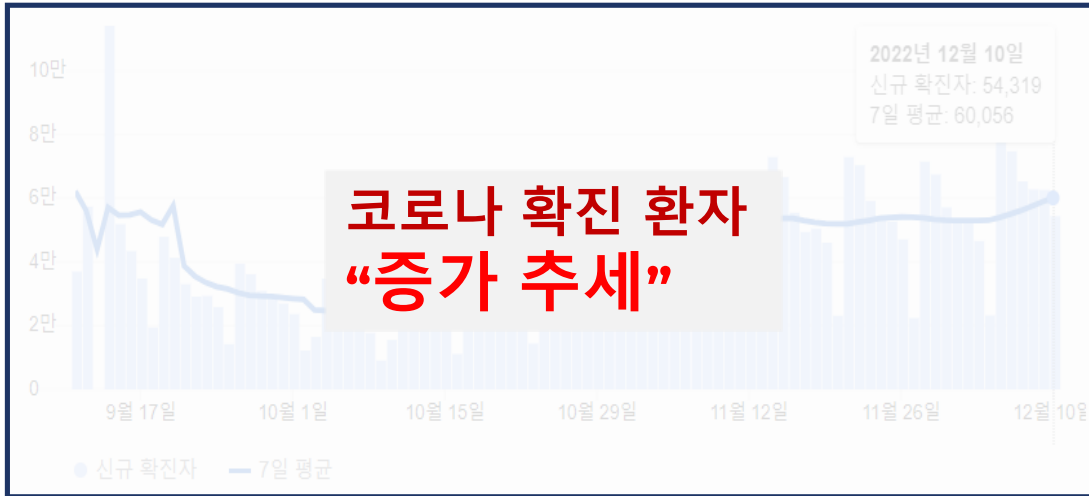


- 01. 팀 구성 및 역할
- 02. 기획 의도
- 03. 개발 흐름 및 일정
- 04. 기술 스택
- 05. 시연 동영상

## 1) 팀 구성 및 역할

팀원	역할	주요 담당	세부 업무
김영원	팀장	머신러닝 / 데이터분석	- EDA - 인공지능 학습 모델 설계 및 테스트
정호영	팀원	챗봇 / RPA	- 챗봇 알고리즘 설계 - 데이터 크롤링
김혜원	팀원	웹 프레임워크 / DB	- Flask 웹 프레임워크 개발 - Bootstrap 디자인

## Part 1 2) 기획 의도



코로나19 확진자가 늘어나며 자가진단키트 수요도 급증하고 있지만 가격이 비싸 국민 부담이 커지고 있다.

무증상자에 대한 검사 수요가 늘고 있다. 전국 자가진단키트 매출은 전 두 배 이상 급증한

**자가 진단 키트  
“수요에 따른 가격 증가”**

그러나 현재 검사 1회분당 자가진단키트 가격은 약국이나 편의점 구매 시 5000~6000원, 온라인 구매 시 2000~3000원 선에서 형성돼 있다.

의료계는 방역당국이 전문가를 신속 항원검사 결과를 만장키로 지원할 바꾼 데 대해 검사를 받기 위해 진료검진소에서 몇 시간씩 대기해야 하는 국민의 불편을 줄이고, 항바이러스제 투여 시기를 앞당길 수 있으며 PCR 비용은 물론 선별검사에 대한 의료진의 소진 문제도 다수 완화할

하지만 중대본의 영향이 나오는 등 위용성분이 높은 전문가들은 의료

**“의료 인력의 소진”  
신속항원검사와 자가키트의  
“정확도 차이 발생”**

일반을 자기 신속항원검사는 코구멍 안쪽에 1.5~2cm 가량 면봉을 넣고 10회 정도 동글게 문질러 채취하는 반면에 동네 병원에서 시행하는 전문가를 신속항원검사는 코구멍 끝은 비인두까지 면봉을 넣어 검체를 채취하므로 정확성이 더 높다는 것

**Q 코로나 검사 고통과 인아프게 받는법** **Q 코로나 자가진단 키트...**

코로나 검사 고통을 막는 법도 있고 인아프게 받는법도... 오늘기준 어제부터 조금씩

코로나 검사 고통을 막는 법도 있고 인아프게 받는법도... 오늘기준 어제부터 조금씩

**코로나 검사 고통**

제가 곧 코로나자가키트 해야하는데요 주사보다 아플까 진짜 코에 넣는거 진짜 못하겠는데...

**코로나 검사 고통**

메전에도 하다가 중간에 포기했거든요... 진짜 독은아픈데 자가키트를 할 자신이 없어서그런데 어렵 경우 뭐 어떻게 해야하나요...

## Part 1 2) 기획 의도

### “코로나, 자가 진단” 키워드 검색 연관어

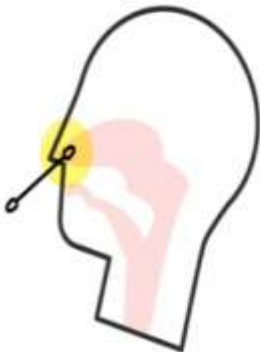
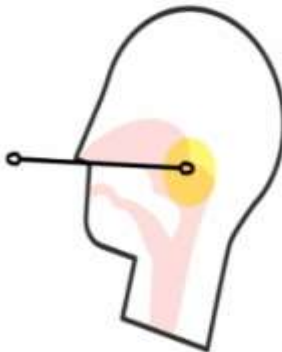


출처 : 메디칼타임즈, 동아일보, 연합뉴스 등 2022.12.09 최근 10건 내 기사

순위	단어	언급량	순위	단어	언급량
1	검사	0.163043	16	지불	0.032609
2	키트	0.097826	17	수요	0.021739
3	자가진단	0.086957	18	급증	0.021739
4	코로나	0.076087	19	안팎	0.021739
5	가격	0.065217	20	회피	0.021739
6	국민	0.054348	21	일주일	0.021739
7	부담	0.054348	22	현재	0.021739
8	무증상	0.043478	23	편의점	0.021739
9	진자	0.043478	24	최근	0.021739
10	구매	0.043478	25	직장	0.021739
11	비용	0.032609	26	모씨	0.021739
12	대비	0.032609	27	휴가	0.021739
13	경우	0.032609	28	크게	0.021739
14	발생	0.032609	29	때문	0.021739
15	직장인	0.032609	30	여행	0.021739

## Part 1 2) 기획 의도

### 기존 코로나 검사 방법

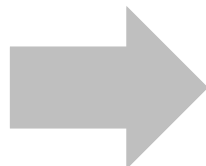
	개인용	전문가용
채취·검사·판독	사용자가 스스로 수행	의료진이 모두 수행
채취 방법	콧구멍 안쪽에 1.5~2cm가량 넣고 10회 정도 둥글게 문질러 채취	콧구멍 깊은 비인두까지 면봉을 넣어 채취
		

진단 부위

코, 목

소요 시간

15~30분



### 새로운 코로나 검사 방법

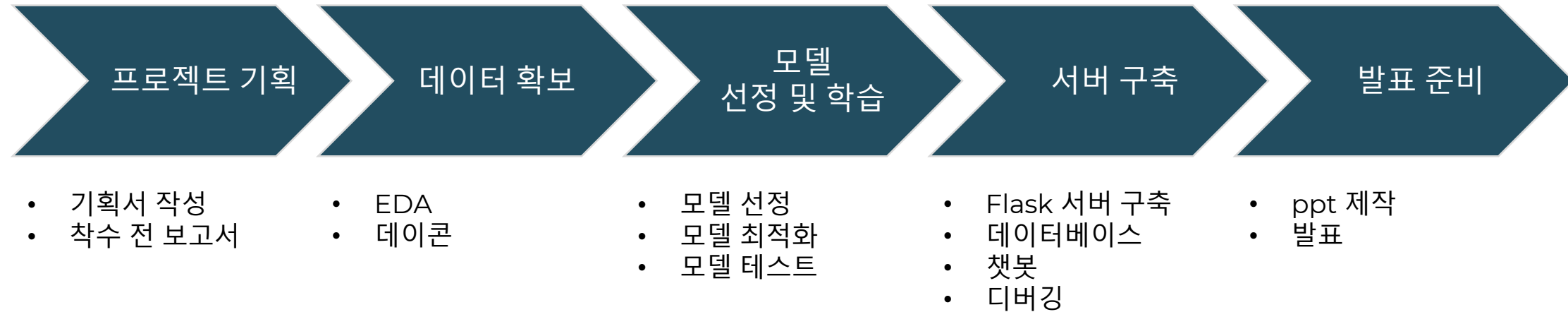


진단 부위

목소리(음성)

소요 시간

즉시 결과 도출





## Part 1 3) 개발 흐름 및 일정

단계	세부정보		주차							
			1	2	3	4	5	6	7	8
1	프로젝트 기획	주제 선정								
		기획서 작성								
2	데이터 확보	EDA								
3	모델 선정 및 학습	모델 선정 및 최적화								
		모델 테스트								
4	서버 구축	Flask 서버 구축								
		데이터베이스								
		챗봇								
		디버깅								
5	발표 준비	ppt 제작								
		발표								

- 프로젝트 기간 : 11월 18일 ~ 1월 13일

## Part 1 3) 개발 흐름 및 일정

단계	세부정보		업무 내용	진행율
1	프로젝트 기획	주제 선정	1차 주제 선정 (기후에 따른 농작물 추천 시스템)	11-25 ~ 11-28 (100%)
		일정표 작성	일정표 작성	11-28 (100%)
		웹 화면 구상	웹 화면 구상	11-29 ~ 12-02 (100%)
		기획서 작성	기획서 작성	11-30 (100%)
		주제 선정	2차 주제 선정 (스마트온실시스템 시뮬레이터)	12-01 (100%)
		주제 선정	3차 주제 선정 (음향 데이터 COVID-19 검출)	12-05 (100%)
		착수 전 보고서	ppt 작업 및 발표	12-05 ~ 12-12 (100%)
		역할 배분	역할 배분	12-06 (100%)
		기획의도	기획의도 검색	12-06 ~ 12-08 (100%)

- 프로젝트 기간 : 11월 18일 ~ 1월 13일

## Part 1 3) 개발 흐름 및 일정

단계	세부정보		업무 내용	진행율
2	데이터 확보	데이터 수집	1차 데이터 수집	11-29 ~ 11-30 (100%)
		EDA	1차 데이터 EDA	11-29 ~ 11-30 (100%)
		데이터 수집	2차 데이터 수집	12-01 (100%)
		EDA	2차 데이터 EDA	12-02 (100%)
		데이터 수집	3차 데이터 수집	12-12 (100%)
		EDA	3차 데이터 EDA	12-13 ~ 12-20 (100%)
		데이터 전처리	기침 소리 데이터 전처리	12-13 ~ 12-20 (100%)
		챗봇	선별진료소, 호흡기진료센터 수집 및 전처리	12-14 (100%)
		UiPath	코로나 접종 현황 데이터 크롤링	12-20 ~ 12-21 (100%)
		챗봇	Q&A 데이터 수집 및 전처리	12-22 ~ 12-29 (100%)

- 프로젝트 기간 : 11월 18일 ~ 1월 13일

## Part 1 3) 개발 흐름 및 일정

단계	세부정보		업무 내용	진행율
3	모델 선정 및 학습	모델 실행	데이콘 공유 코드	12-08 ~ 12-09 (100%)
		모델 실행	Resnet 34	12-13 ~ 2023-01-09(100%)
		모델 연결	플라스크에 모델 연결	12-15 ~ 12-16 (100%)
		1차 모델 수정	출력 뉴런수, jpg 파일	12-20 (100%)
		모델 변수값 조정	Resnet 34 Learning Rate 조정	12-22 (100%)
		모델 실행	Vgg 16	12-22 ~ 2023-01-09(100%)
		모델 seed값 고정	모델 seed값 고정	12-23 ~ 12-28 (100%)
		데이터 증강	White noise, Shifting, Stretching	12-28 ~ 12-29 (100%)

- 프로젝트 기간 : 11월 18일 ~ 1월 13일

## Part 1 3) 개발 흐름 및 일정

단계	세부정보		업무 내용	진행율
4	서버 구축	Flask 공부	템플릿, form, 부트스트랩	12-05 ~ 12-09 (100%)
		챗봇 공부	Dialogflow, Intents, Entities Fulfillment, Integrations	12- 08 ~ 12-12 (100%)
		Flask 서버 구축	Flask 서버 구축	12-09 (100%)
		Flask 웹페이지	로그인 페이지	12-09 (100%)
		Flask 웹페이지	회원가입 페이지	12-12 (100%)
		Flask 웹페이지	AI 검증 페이지	12-14 ~ 12-27 (100%)
		Flask 웹페이지	결과 페이지	12-15 ~ 12-19 (100%)
		Flask 웹페이지	코로나 접종 현황판 페이지	12-20 ~12-27 (100%)
		Flask 웹페이지	확진시 조치 안내 페이지	12-20 (100%)
		Flask 웹페이지	근처 진료 병원 지도 페이지	12-21 ~ 12-27 (100%)
		Flask 웹페이지	Q&A 페이지	12-27 ~ 12-28(100%)

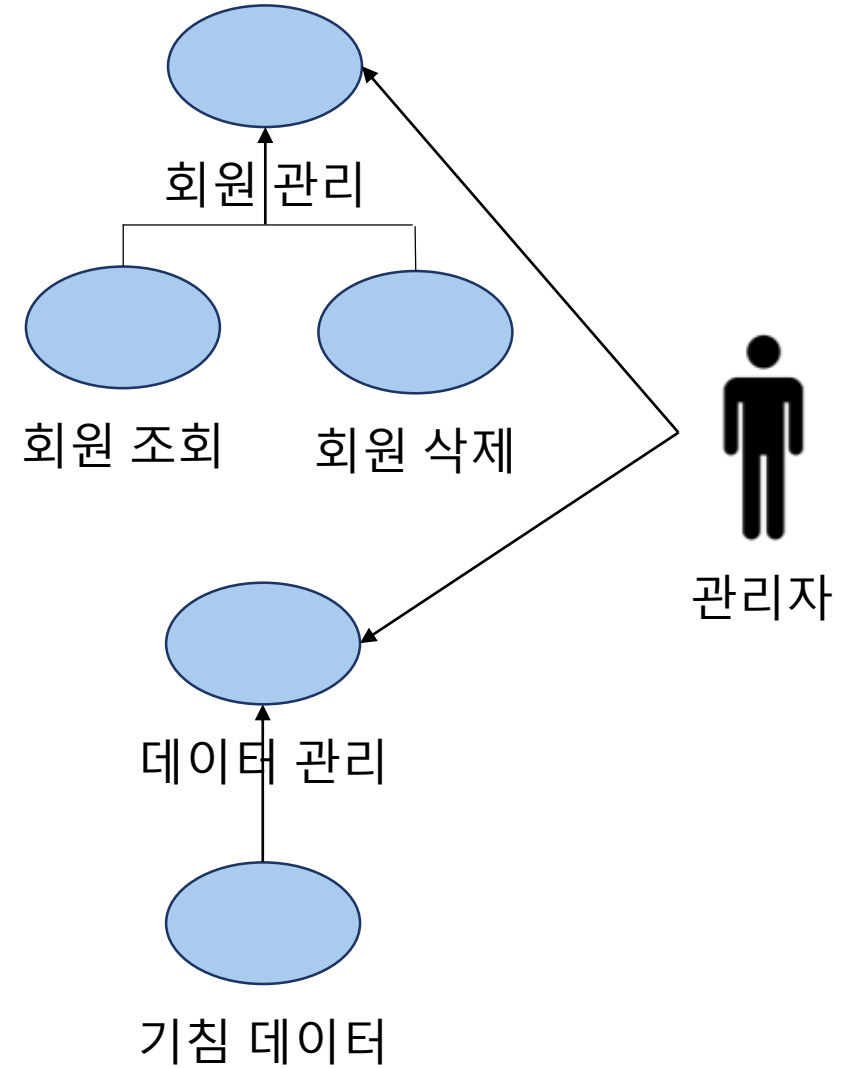
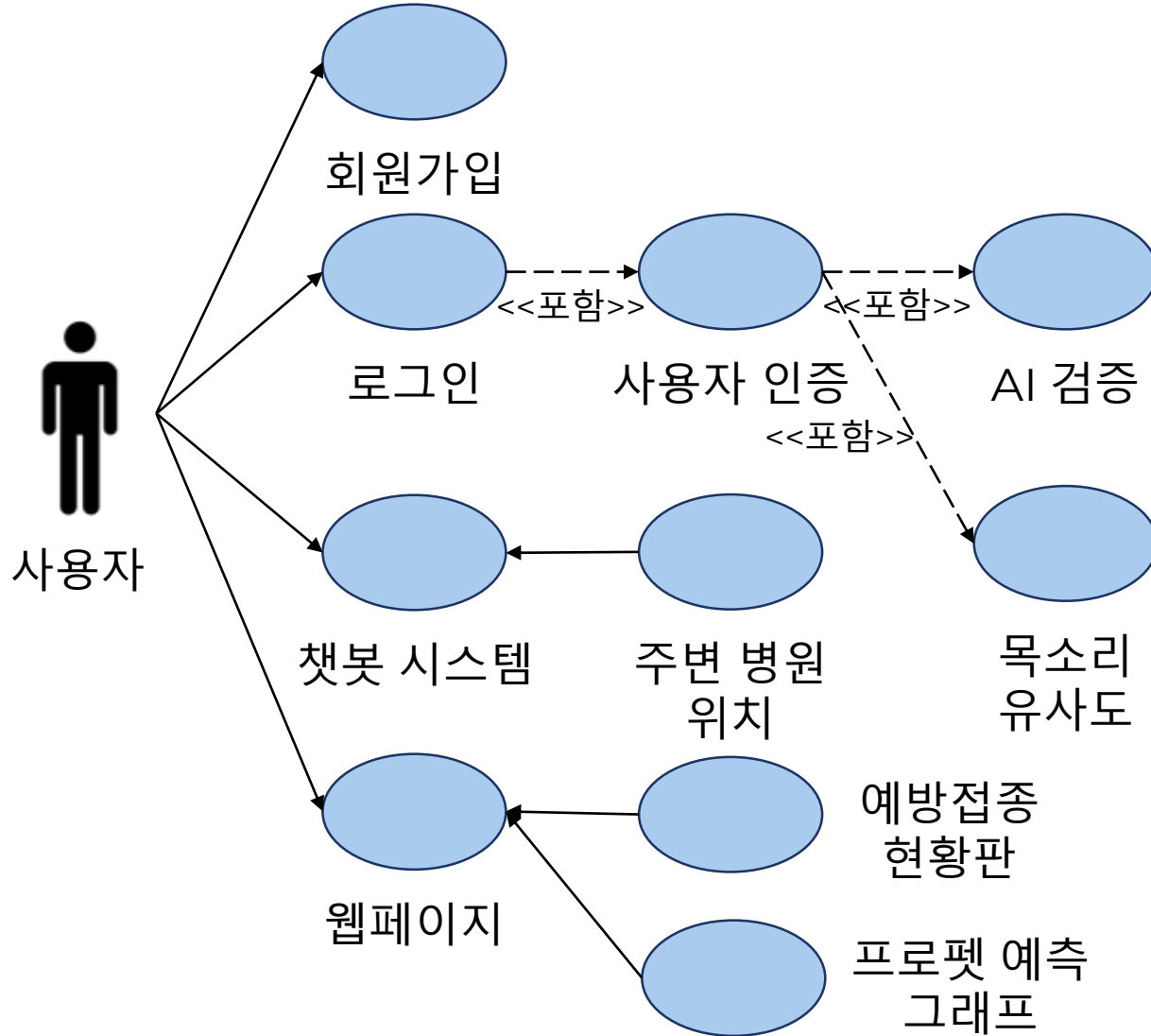
- 프로젝트 기간 : 11월 18일 ~ 1월 13일

## Part 1 3) 개발 흐름 및 일정

단계	세부정보		업무 내용	진행율
4	서버 구축	데이터베이스	회원가입 DB	12-13 (100%)
		챗봇	인텐트 구성	12-13 (100%)
		챗봇	DialogFlow 적용	12-15 (100%)
		챗봇	엔티티 구성	12-16 ~ 12-19 (100%)
		챗봇	플라스크에 챗봇 연결	12-19 ~ 12-20 (100%)
		챗봇	AI 연동 (cos 유사도)	12-21 ~ 12-22 (100%)
		UiPath	오케스트레이터 설정	12-23 ~ 12-29 (100%)
		목소리 유사도	목소리 유사도 평가 (cos 유사도)	12-28 ~ 2023-01-05 (100%)
		챗봇	버튼 구성	12-29 (100%)
5	발표 준비	ppt 제작	ppt 제작	12-27 ~ 2023-01-11 (100%)
		발표 연습	발표 연습	01-11 ~ 01-12 (100%)

● 프로젝트 기간 : 11월 18일 ~ 1월 13일

Part 1 3) 개발 흐름 및 일정



## Part 1 4) 기술 스택

### 하드웨어



OS : Windows10 64bit  
CPU : 12<sup>th</sup> Gen Intel®  
Core™ i7-12700  
RAM : 64GB  
GPU : NVIDIA GeForce  
RTX 3070

### DB



### 버전 관리



git



### 데이터 수집



### 웹 프레임워크



Flask



JAVASCRIPT



Bootstrap



Dialogflow



HTML



CSS

### 개발 도구



python



jupyter



PC

### ML / DL



NumPy

Pandas



librosa



TensorFlow



Keras



PyTorch



## Part 1 5) 시연 동영상



# Part. 2

## 인공지능



- 01. 설계 및 구조도
- 02. 데이터 확인 및 전처리
- 03. 모델 (resnet, vgg)
- 04. 히스토리

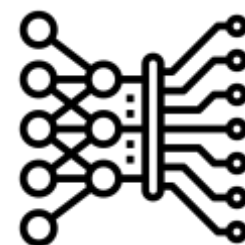
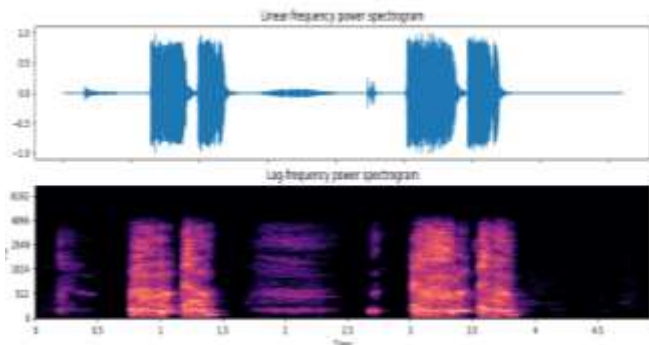
## Part 2 1) 인공지능 설계 및 구조도

코로나 음성 양성 AI 분류 모델

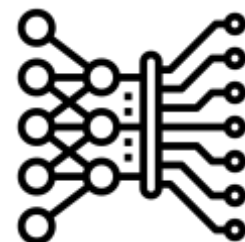
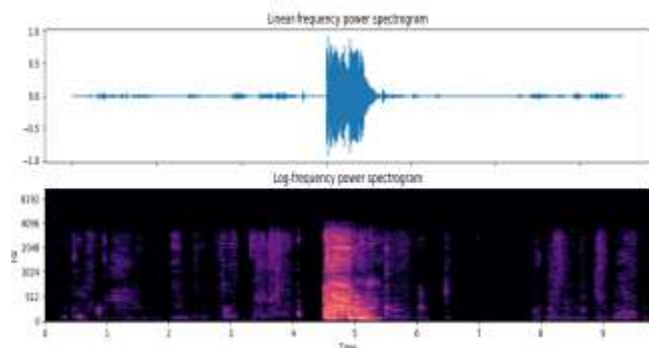
기침 소리 녹음



음성 데이터



양성 데이터



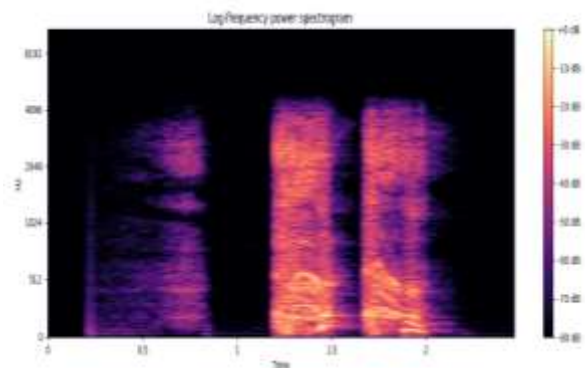
코로나 예측 결과



## Part 2 2) 데이터 확인 및 전처리

### 데이콘 제공 데이터셋

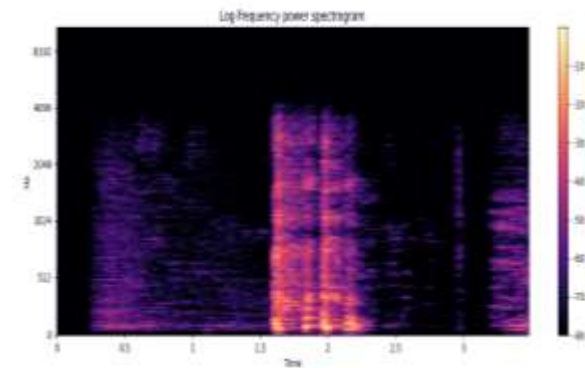
제공내용	예시	제공방식
기침소리	00001.wav	wav 파일
대상자 ID	1	csv 파일
대상자 나이	24	csv 파일
대상자 성별	female	csv 파일
대상자 호흡기 질환 여부	0	csv 파일
대상자 발열 혹은 근육통 증상 여부	1	csv 파일
코로나19 음성/양성 여부 (음성:0, 양성:1)	0	csv 파일



음성 변환  
이미지



양성 변환  
이미지

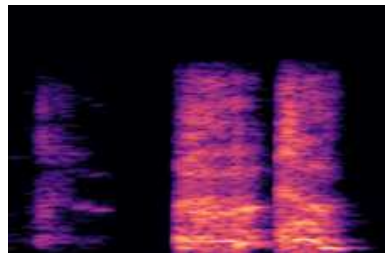


## Part 2 2) 데이터 확인 및 전처리

데이터 원본

00052.wav  
00053.wav  
00054.wav  
00055.wav  
00056.wav  
00057.wav  
00058.wav  
00059.wav  
00060.wav  
00061.wav

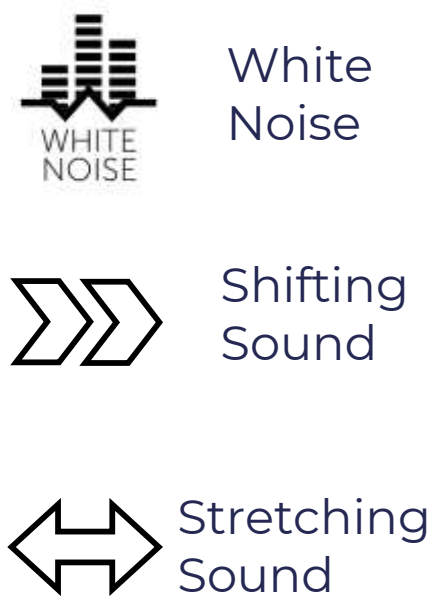
멜스펙트로그램



결측치/이상치



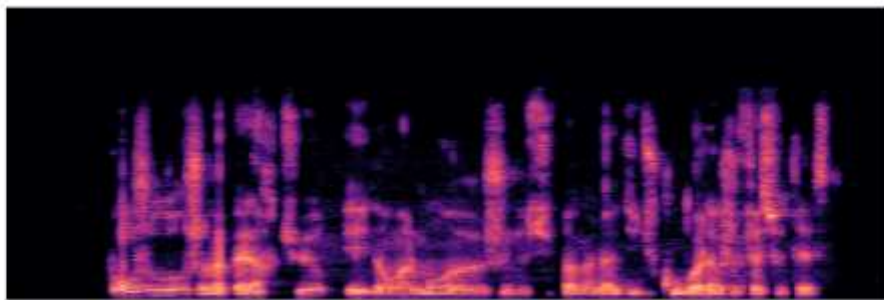
데이터 증강



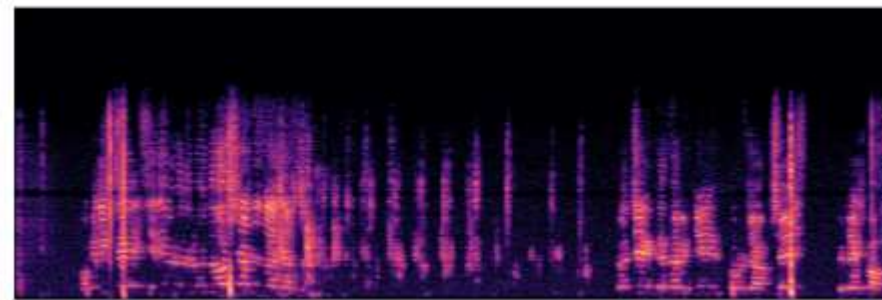
## Part 2 2) 데이터 확인 및 전처리

결측치, 이상치 제거 (음성: 304개, 양성: 43개)

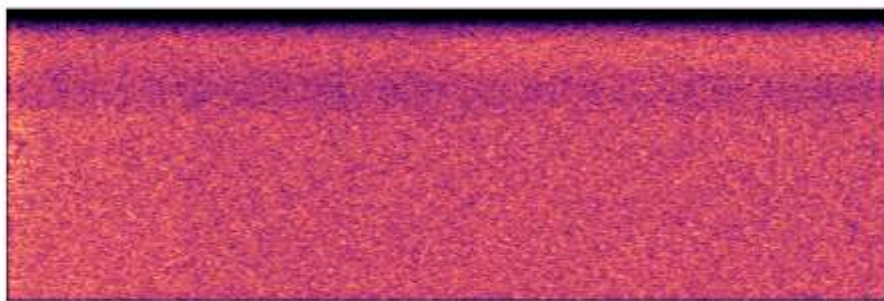
대화하는 소리



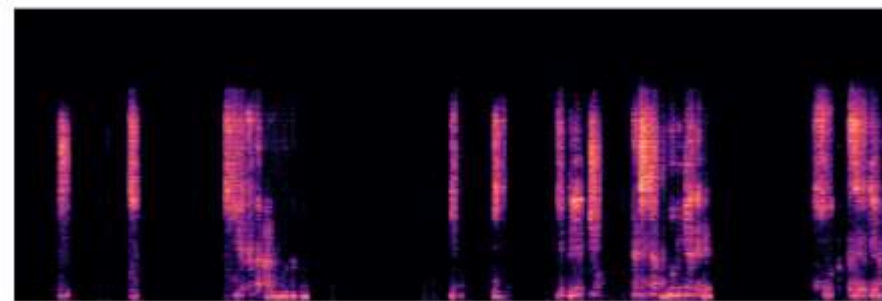
생활 소음



무음



이상치 (나이 1세, 성별 other)

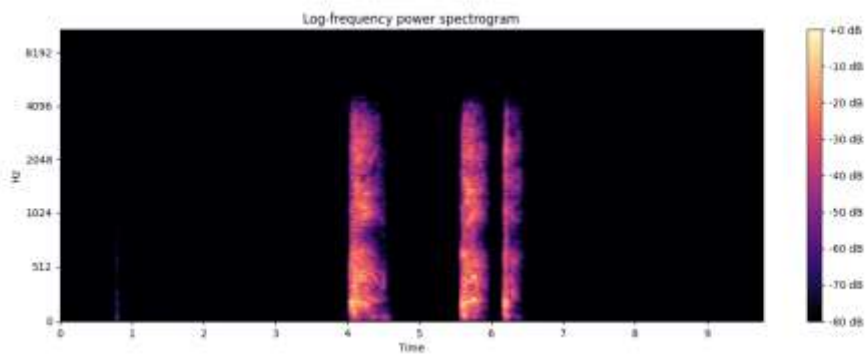




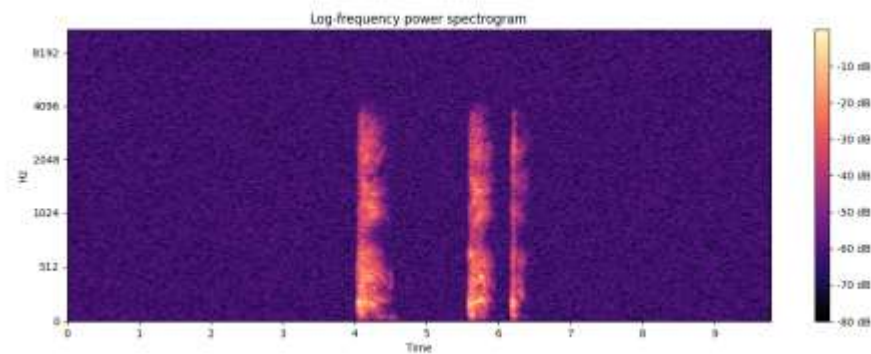
## Part 2 2) 데이터 확인 및 전처리

### 음향 데이터 증강

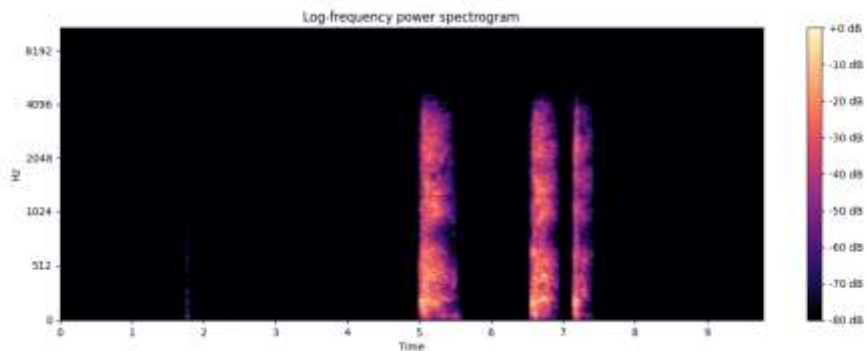
이미지 원본



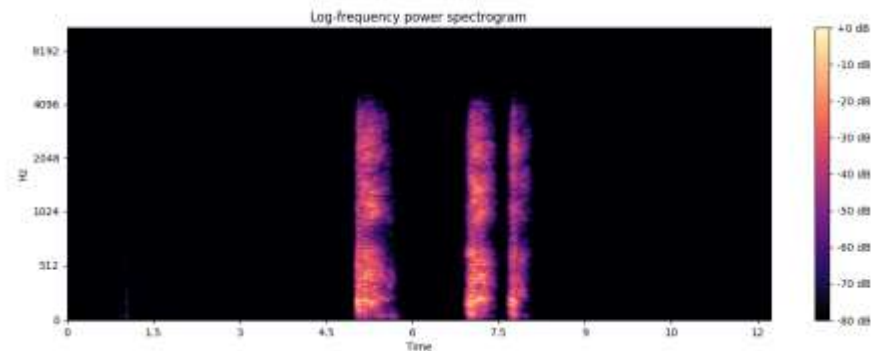
White\_noise



Shifting



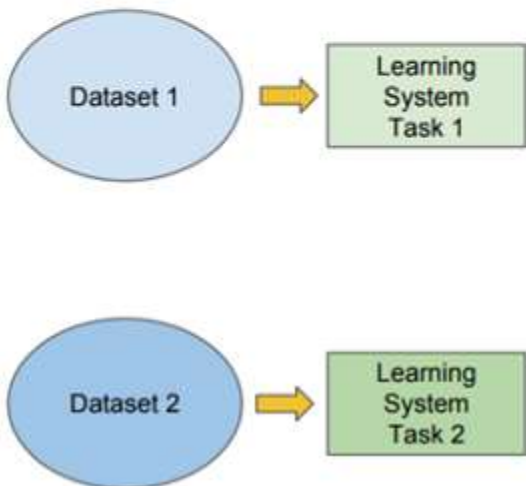
Stretching



## 전이학습이란

### Traditional ML

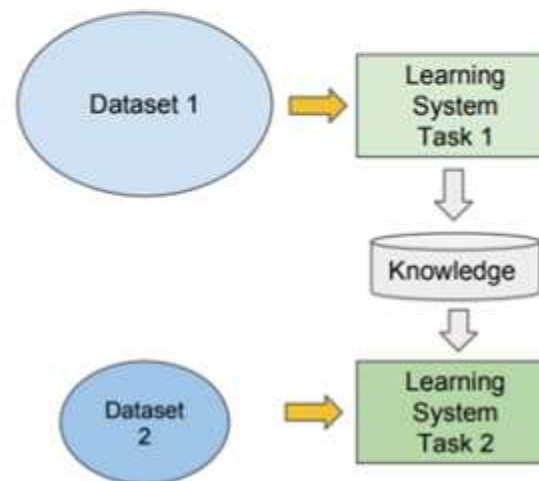
- Isolated, single task learning:
  - Knowledge is not retained or accumulated. Learning is performed w.o. considering past learned knowledge in other tasks



vs

### Transfer Learning

- Learning of a new tasks relies on the previous learned tasks:
  - Learning process can be faster, more accurate and/or need less training data





# ImageNet Challenge

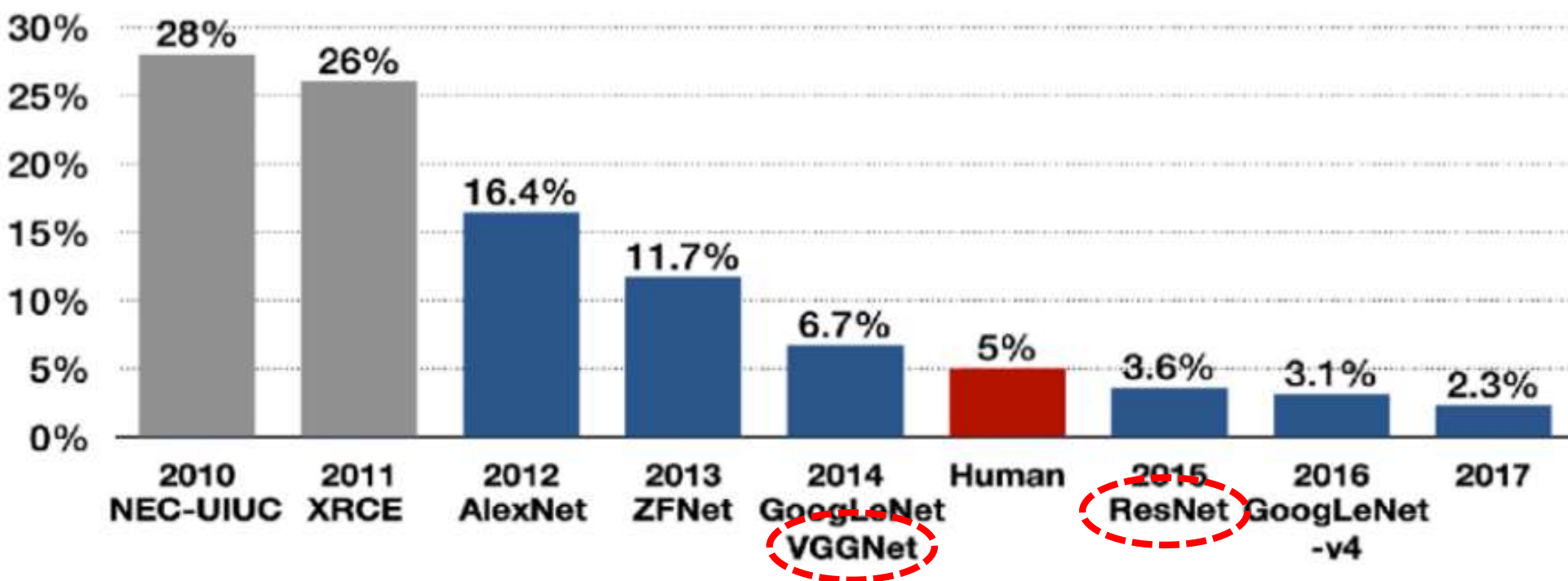
IMAGENET

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.



## Part 2 3) 인공지능 모델 (resnet, vgg)

Top-5 error



## Part 2 4) 히스토리

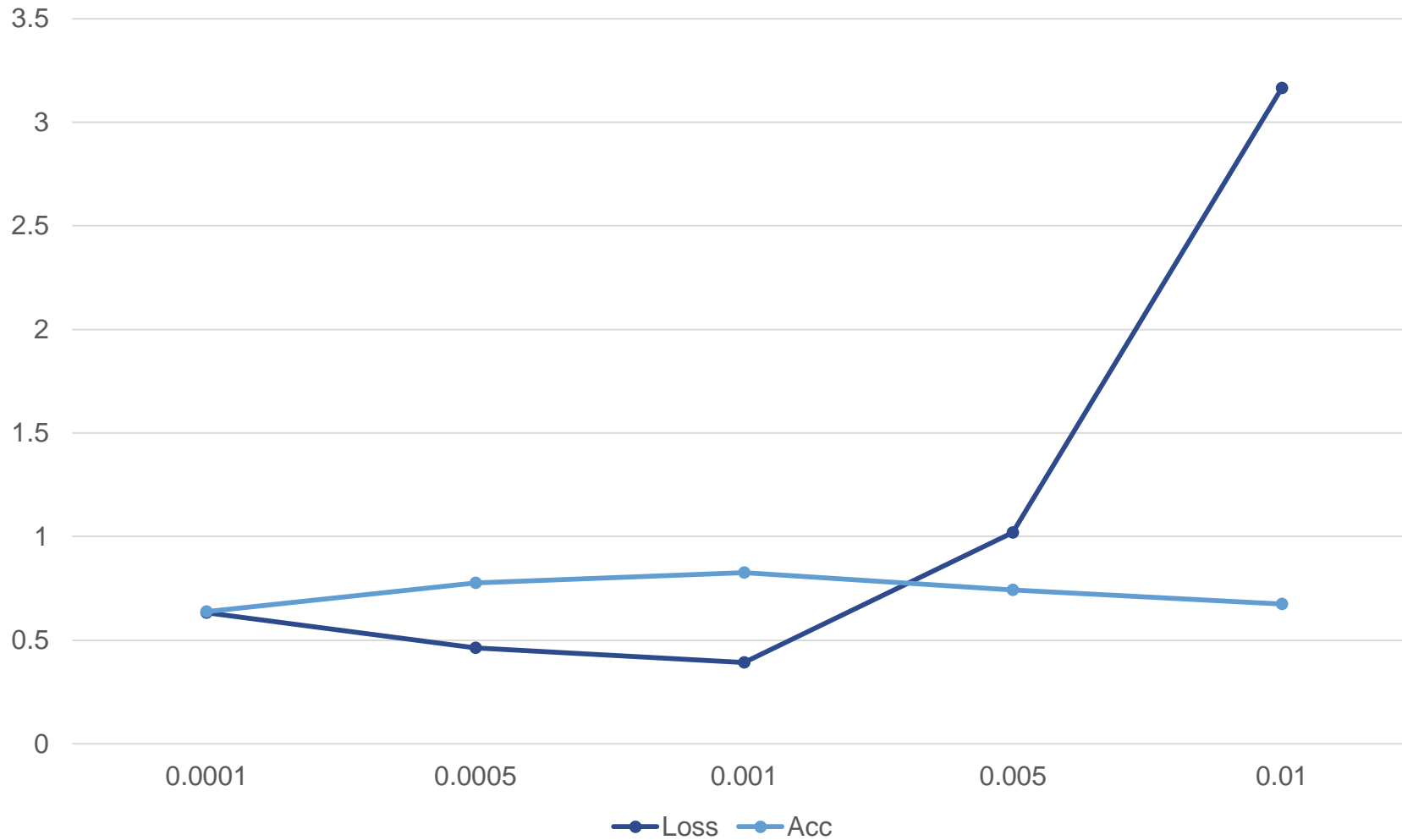
실행 횟수	모델	train 데이터셋	test 데이터셋	하이퍼 파라미터	결과	추가사항
1	resnet34	음성 : 30개 양성 : 30개	음성 : 3469개 양성 : 276개	num_epochs = 50 batch_size=4 lr=0.001 momentum=0.9	Loss: 1.7728 Acc: 31.1348%	
2	resnet34	음성 : 270개 양성 : 270개	음성 : 3229개 양성 : 36개	num_epochs = 50 batch_size=4 lr=0.001 momentum=0.9	Loss: 1.1847 Acc: 43.9510%	train 데이터셋 늘리기
3	resnet34	음성 : 270개 양성 : 270개	음성 : 3229개 양성 : 36개	num_epochs = 50 batch_size=4 lr=0.001 momentum=0.9	Loss: 1.3266 Acc: 38.4380%	출력 뉴런수=2
4	resnet34	음성 : 210개 양성 : 210개	음성 : 2985개 양성 : 53개	num_epochs = 50 batch_size=4 lr=0.001 momentum=0.9	Loss: 0.5739 Acc: 72.6794%	결측치 이상치 제거
5	resnet34	음성 : 210개 양성 : 210개	음성 : 2985개 양성 : 53개	num_epochs = 50 batch_size=4 lr=0.001 momentum=0.9	Loss: 0.3914 Acc: 82.6531%	.png -> .jpg 이미지 사용

## Part 2 4) 히스토리

실행 횟수	모델	train 데이터셋	test 데이터셋	하이퍼 파라미터	결과	추가사항
6	resnet34	음성 : 210개 양성 : 210개	음성 : 2985개 양성 : 53개	num_epochs = 50 batch_size=4 lr=0.01 momentum=0.9	Loss: 3.1657 Acc: 67.2811%	
7	resnet34	음성 : 210개 양성 : 210개	음성 : 2985개 양성 : 53개	num_epochs = 50 batch_size=4 lr=0.005 momentum=0.9	Loss: 1.0190 Acc: 74.2265%	
8	resnet34	음성 : 210개 양성 : 210개	음성 : 2985개 양성 : 53개	num_epochs = 50 batch_size=4 lr=0.005 momentum=0.9	Loss: 0.4618 Acc: 77.5181%	
9	resnet34	음성 : 210개 양성 : 210개	음성 : 2985개 양성 : 53개	num_epochs = 50 batch_size=4 lr=0.005 momentum=0.9	Loss: 0.6321 Acc: 63.6932%	

## Part 2 4) 히스토리

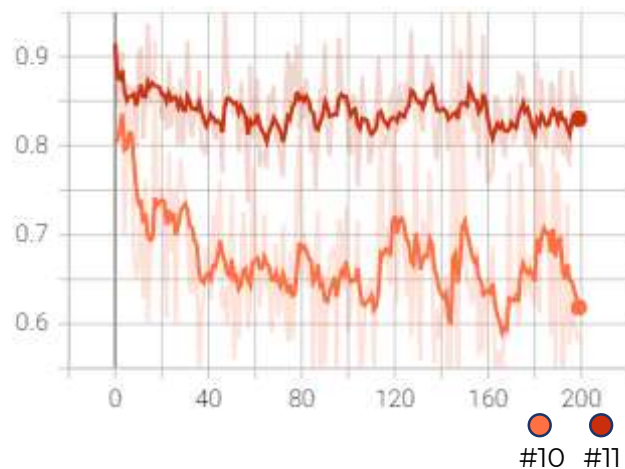
학습률 변화에 따른 Loss/Acc 추이



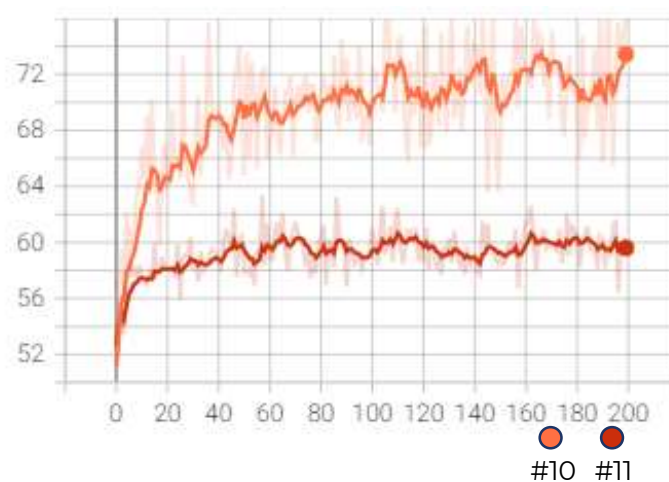
## Part 2 4) 히스토리

실행횟수	모델	train 데이터셋	test 데이터셋	하이퍼 파라미터
10	resnet34	음성 : 210개 양성 : 210개	음성 : 2985개 양성 : 53개	num_epochs = 200 batch_size=4 lr=0.001 momentum=0.9
11	resnet34	음성 : 1005개 양성 : 999개 (데이터 증강)	음성 : 2985개 양성 : 53개	num_epochs = 200 batch_size=4 lr=0.001 momentum=0.9
12	resnet34	음성 : 1005개 양성 : 999개	음성 : 2985개 양성 : 53개	num_epochs = 1000 batch_size=4 lr=0.01 momentum=0.9

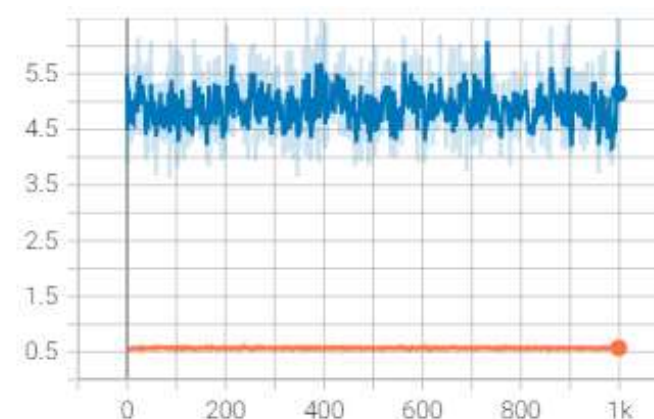
Train Loss



Train Acc



Train Sum

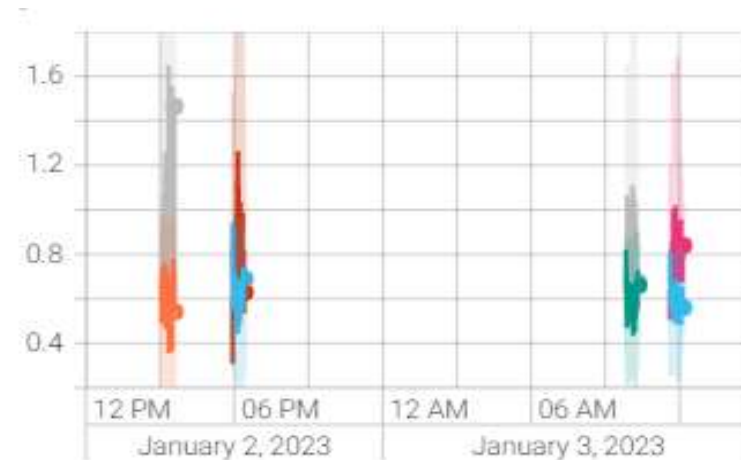
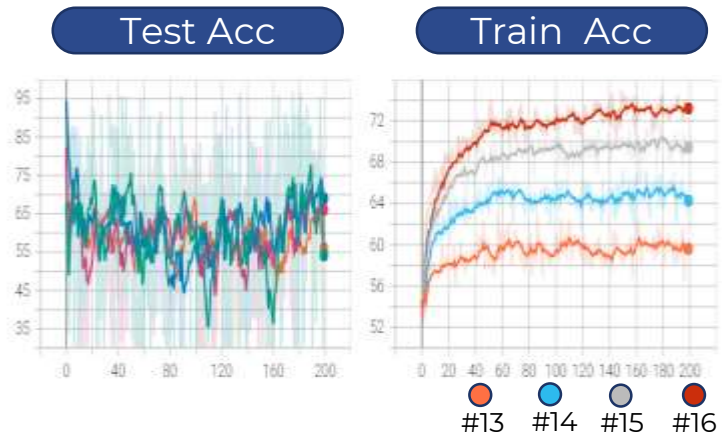
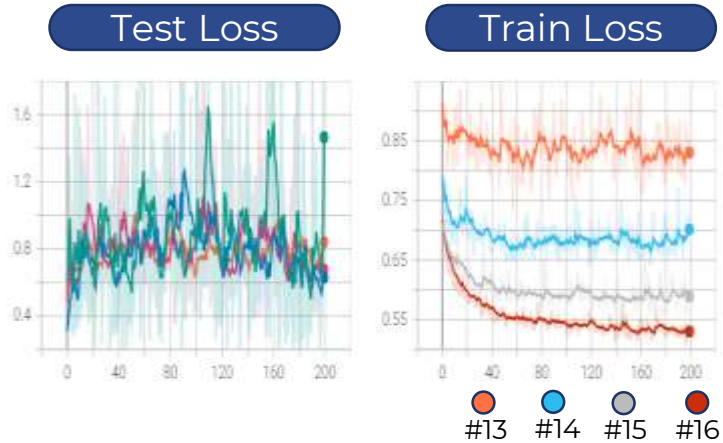




## Part 2 4) 히스토리

모델	train 데이터셋	test 데이터셋
resnet34	음성 : 1005개 양성 : 999개	음성 : 2985개 양성 : 53개

실행횟수	하이퍼 파라미터
13	num_epochs = 200 batch_size=4 lr=0.001 momentum=0.9
14	num_epochs = 50 batch_size=8 lr=0.001 momentum=0.9
15	num_epochs = 50 batch_size=16 lr=0.001 momentum=0.9
16	num_epochs = 50 batch_size=32 lr=0.001 momentum=0.9



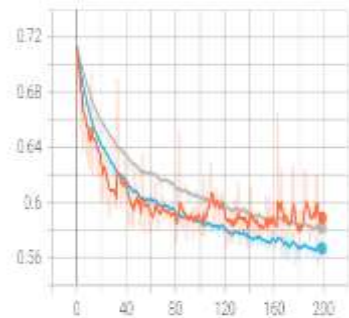
실행결과 batch\_size  
(16,32)에서 Test의 변동성이  
줄어드는 결과를 확인함  
➔ 배치 16,32로 작업결정함

## Part 2 4) 히스토리

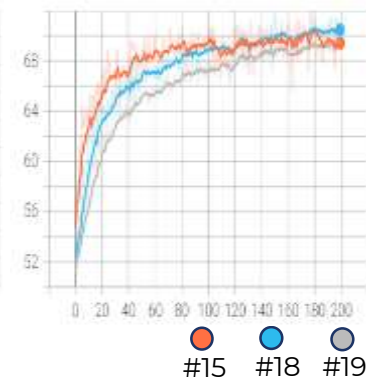
모델	train 데이터셋	test 데이터셋
resnet34	음성 : 1005개 양성 : 999개	음성 : 2985개 양성 : 53개

실행횟수	하이퍼 파라미터
17	num_epochs = 200 batch_size=32 lr=0.0001 momentum=0.9
18	num_epochs = 50 batch_size=16 lr=0.0001 momentum=0.9
19	num_epochs = 50 batch_size=16 lr=0.00005 momentum=0.9
20	num_epochs = 50 batch_size=32 lr=0.00005 momentum=0.9

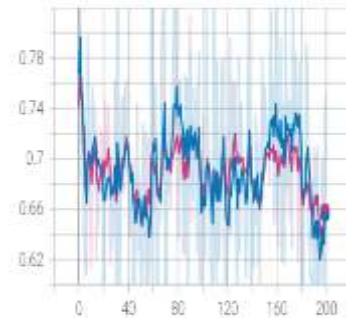
Train Loss



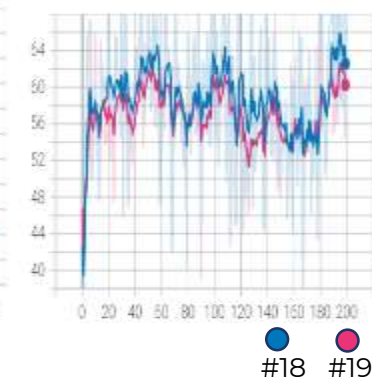
Train Acc



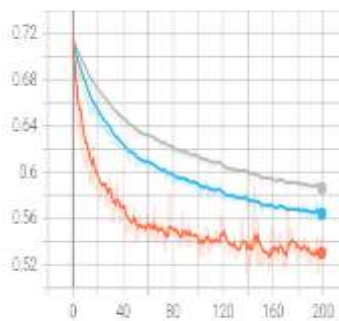
Test Loss



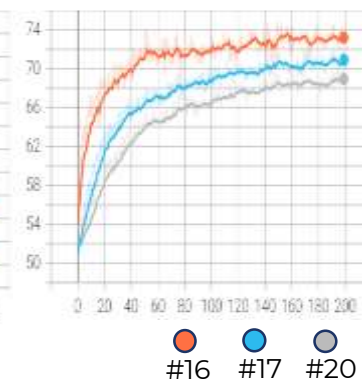
Test Acc



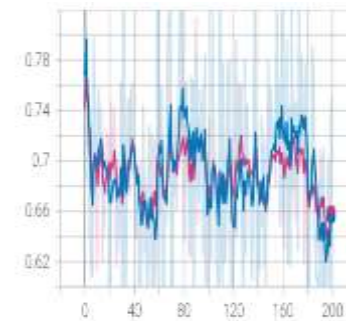
Train Loss



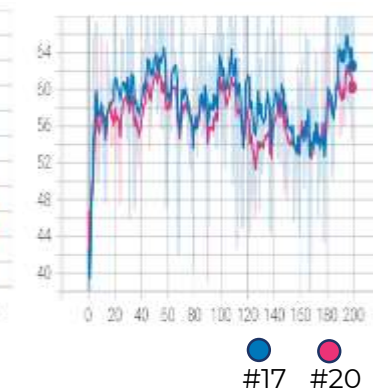
Train Acc



Test Loss



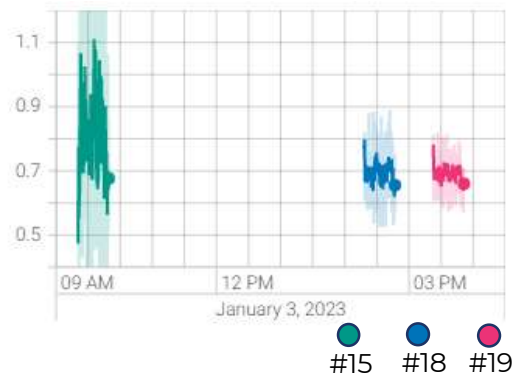
Test Acc



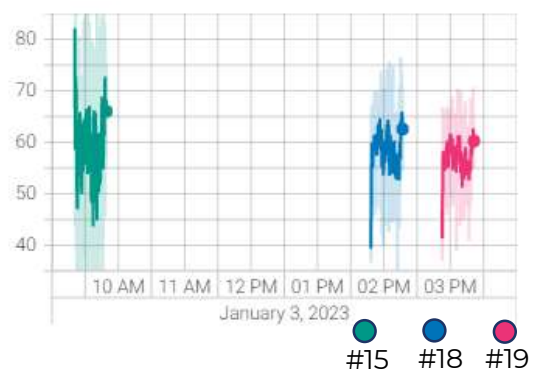


## Part 2 4) 히스토리

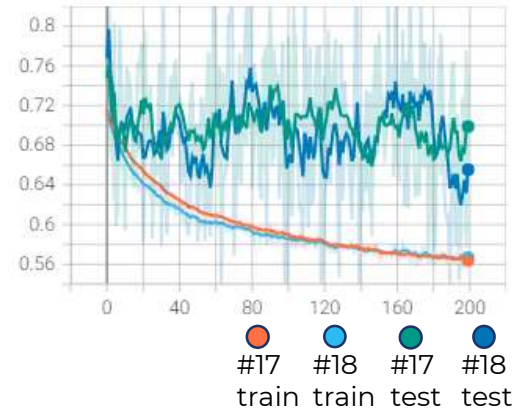
Test Loss



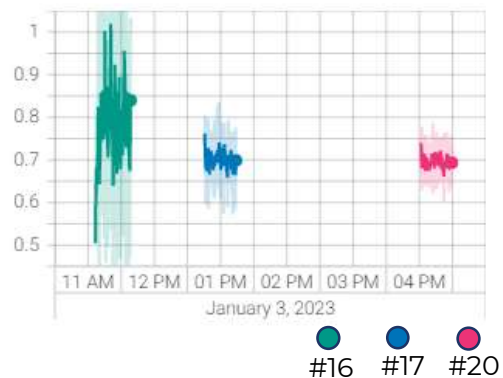
Test Acc



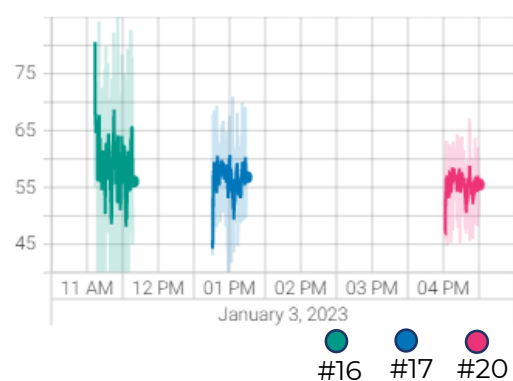
Sum Loss



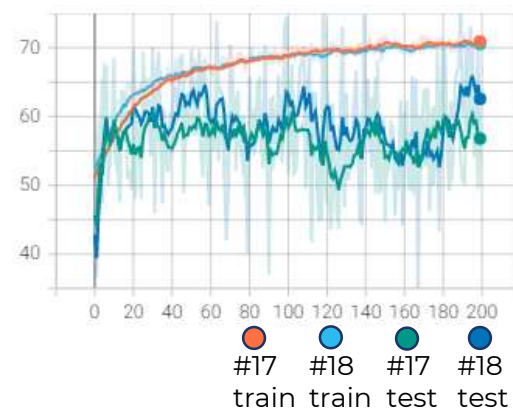
Test Loss



Test Acc



Sum Acc

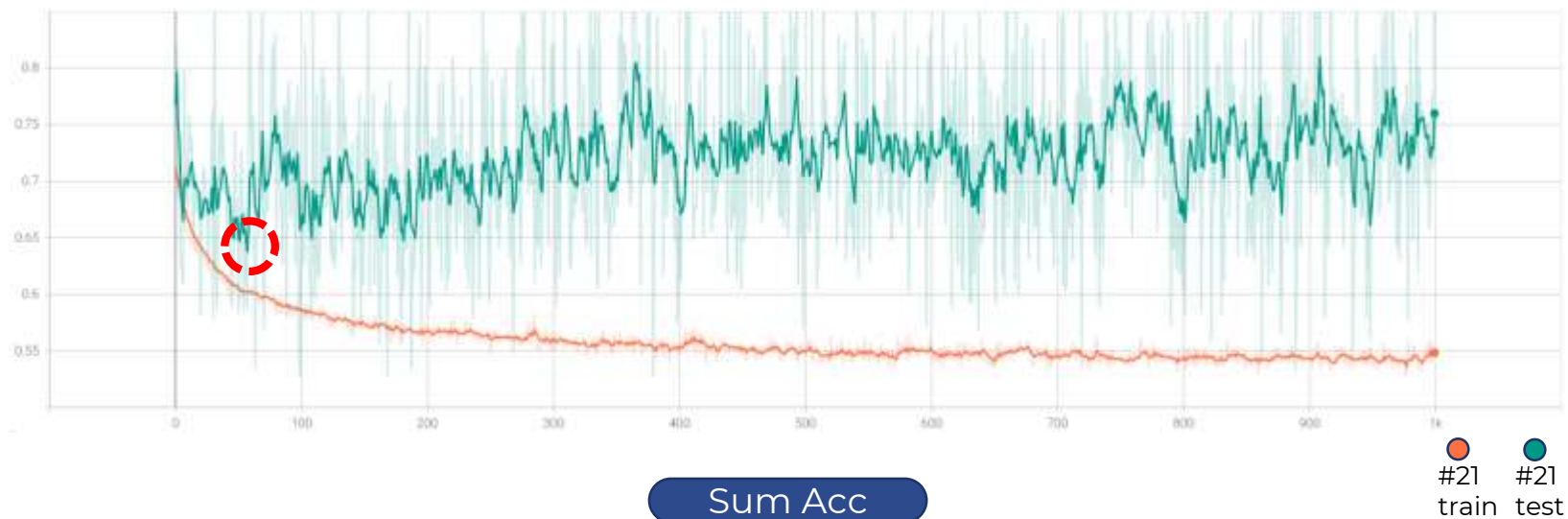


실행결과 학습률  
0.0001에서  
최적의 결과를  
확인함

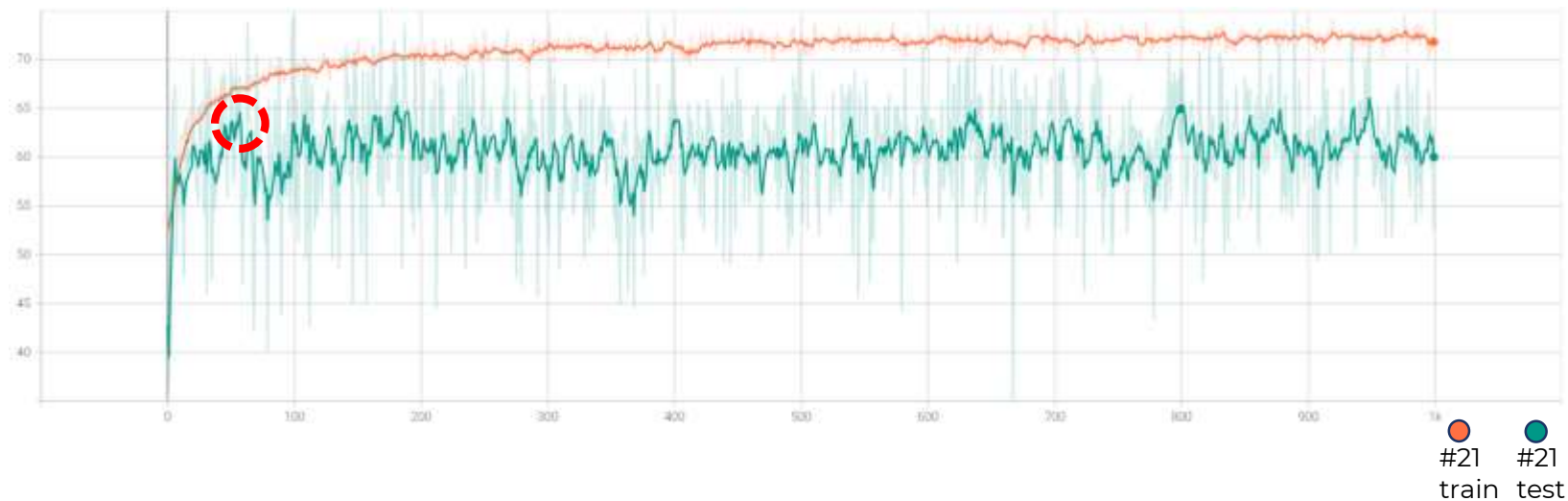
➔ 배치사이즈 16,  
lr 0.0001  
로 작업결정함

## Part 2 4) 히스토리

Sum Loss



Sum Acc



실행결과 에포크  
57이후 일정한  
변화를 확인할 수  
없었음

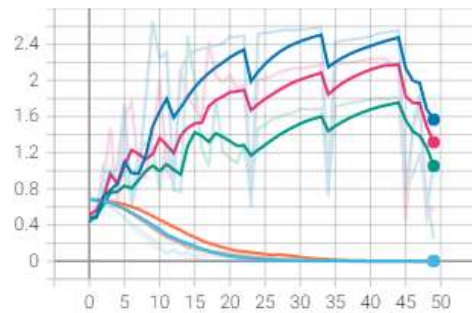
➔ 배치 16,  
lr 0.0001,  
에포크 57로  
최종 결정함

## Part 2 4) 히스토리

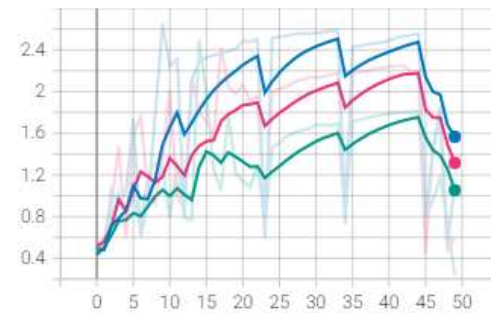
모델	train 데이터셋	test 데이터셋
vgg16	음성 : 1005개 양성 : 999개	음성 : 2985개 양성 : 53개

실행횟수	하이퍼 파라미터
9	num_epochs = 50 batch_size=4 lr=0.001 momentum=0.9
12	num_epochs = 50 batch_size=8 lr=0.001 momentum=0.9
15	num_epochs = 50 batch_size=16 lr=0.001 momentum=0.9

Sum Loss

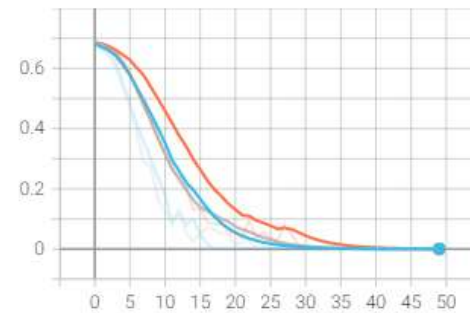


Test Loss



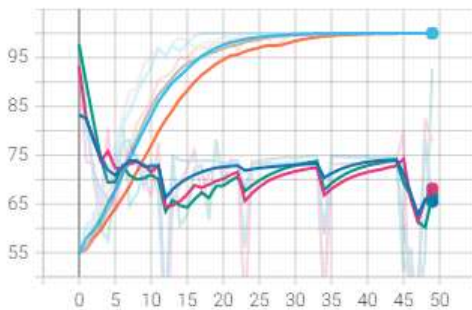
#9 #12 #15

Train Loss

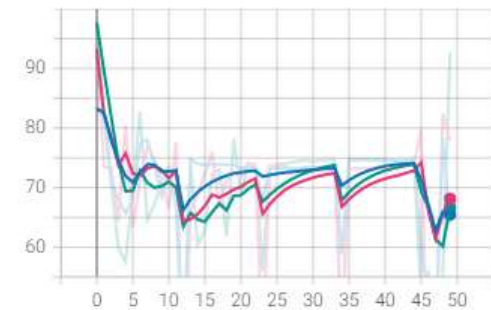


#9 #12 #15

Sum Acc

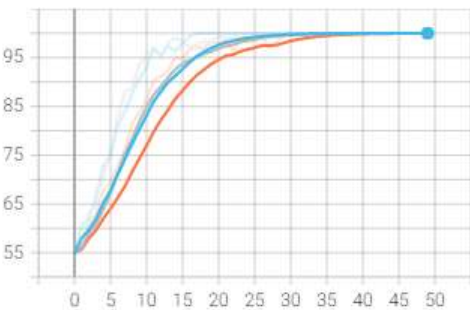


Test Acc



#9 #12 #15

Train Acc



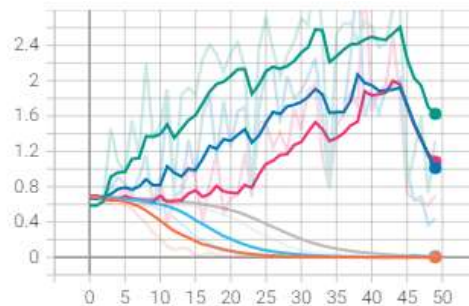
#9 #12 #15

## Part 2 4) 히스토리

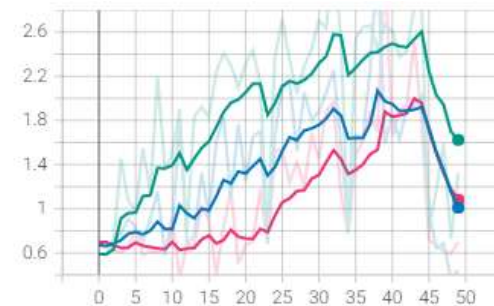
모델	train 데이터셋	test 데이터셋
vgg16	음성 : 1005개 양성 : 999개	음성 : 2985개 양성 : 53개

실행횟수	하이퍼 파라미터
10	num_epochs = 50 batch_size=4 lr=0.0001 momentum=0.9
13	num_epochs = 50 batch_size=8 lr=0.0001 momentum=0.9
16	num_epochs = 50 batch_size=16 lr=0.0001 momentum=0.9

Sum Loss

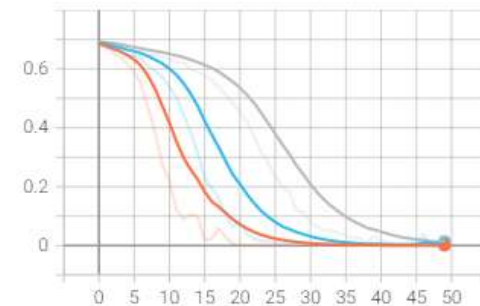


Test Loss



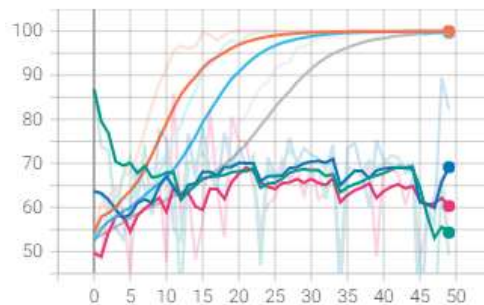
#10 #13 #16

Train Loss

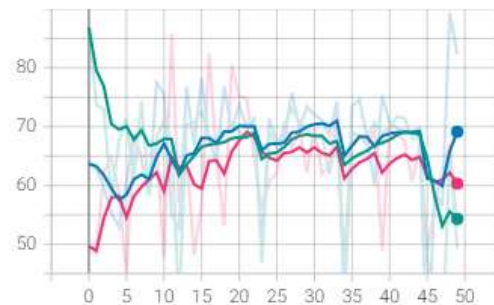


#10 #13 #16

Sum Acc

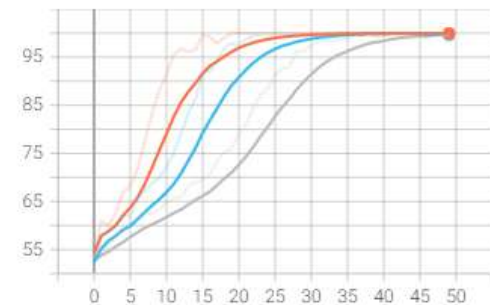


Test Acc



#10 #13 #16

Train Acc



#10 #13 #16



## Part 2 4) 히스토리

모델	train 데이터셋	test 데이터셋
vgg16	음성 : 1005개 양성 : 999개	음성 : 2985개 양성 : 53개

실행횟수	하이퍼 파라미터
11	num_epochs = 50 batch_size=4 lr=0.00001 momentum=0.9
14	num_epochs = 50 batch_size=8 lr=0.00001 momentum=0.9
17	num_epochs = 50 batch_size=16 lr=0.00001 momentum=0.9



## Vgg16 실행 결과

train의 학습이 빠르고 이에 따른 과적합 문제가 발생  
과적합 문제로 인해  
전체적으로 Test의 Acc값이 하락, Loss 값이 상승함



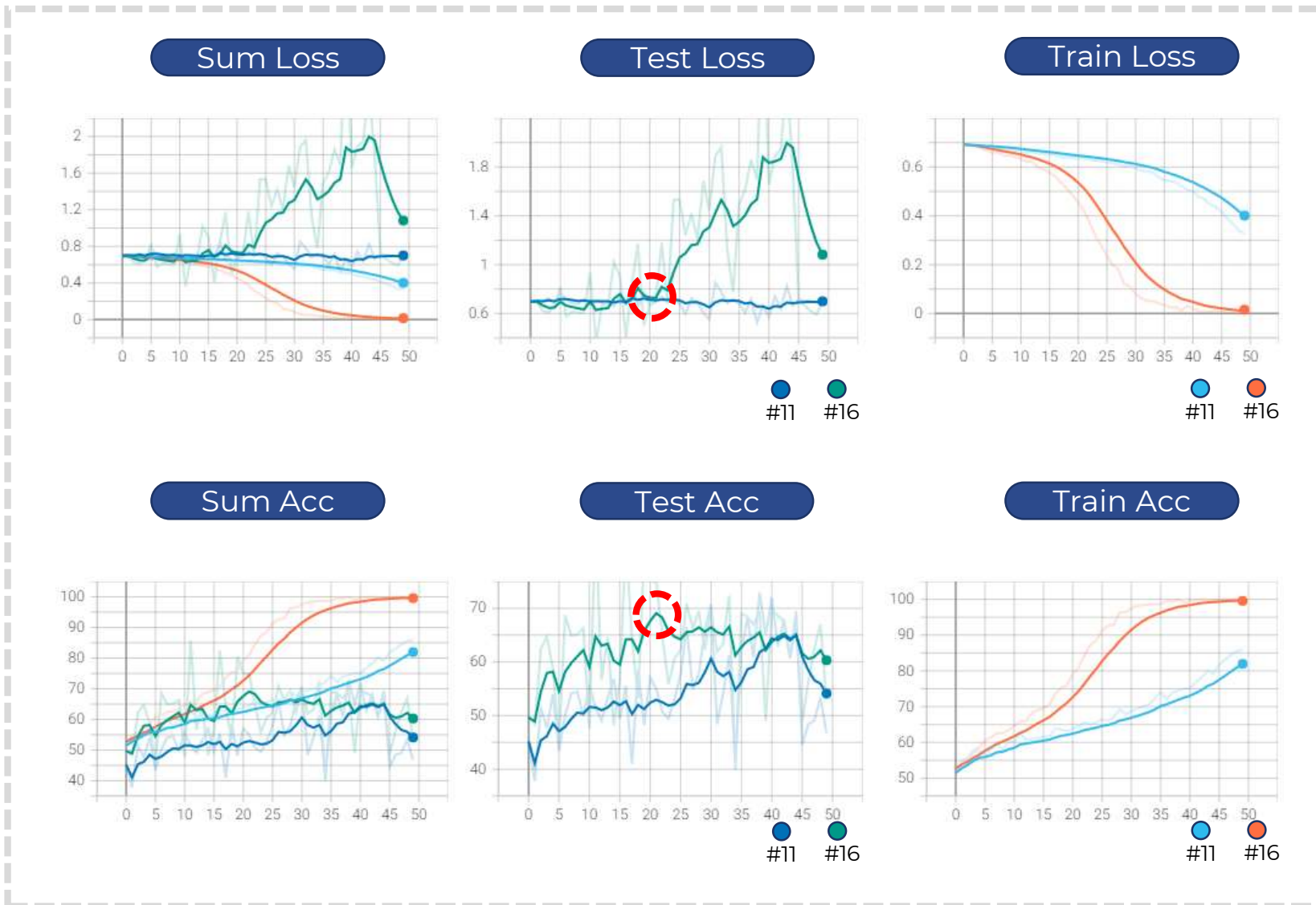
Learning rate의 값이 매우 작아야 안정화를 보이고  
그에 맞는 배치사이즈를 맞춰야 원하는 결과를 보였음

## Part 2 4) 히스토리

모델	train 데이터셋	test 데이터셋
vgg16	음성 : 1005개 양성 : 999개	음성 : 2985개 양성 : 53개

실행횟수	하이퍼 파라미터
11	num_epochs = 50 batch_size=4 lr=0.00001 momentum=0.9
16	num_epochs = 50 batch_size=8 lr=0.00001 momentum=0.9

#16의 Loss 값이 상승하기 전 가장 높은 정확도를 확인함  
→ #16모델 에포크 21로 결정함



## Part 2 4) 히스토리

### resnet34

	예상(양성)	예상(음성)
실제(양성)	29	24
실제(음성)	939	2046

정확도	0.683
정밀도	0.029
재현율	0.547
F1-score	0.056

### vgg16

	예상(양성)	예상(음성)
실제(양성)	50	3
실제(음성)	760	2225

정확도	0.748
정밀도	0.061
재현율	0.943
F1-score	0.115



# Part. 3

## 챗봇

01. 설계 및 구조도

02. 데이터 수집 및 전처리

03. Dialogflow 구성



## 1) 설계 및 구조도

안녕하세요 고객님,  
저희 홈페이지를 방문해주셔서 감사합니다.  
아래 유형을 선택해주세요



## 1) 챗봇 설계 및 구조도

### 챗봇 사용 목적

사용자에게 **주변의 선별진료소**를 알려줌

사용자가 원하는 서비스에 대한 **접근이 용이**

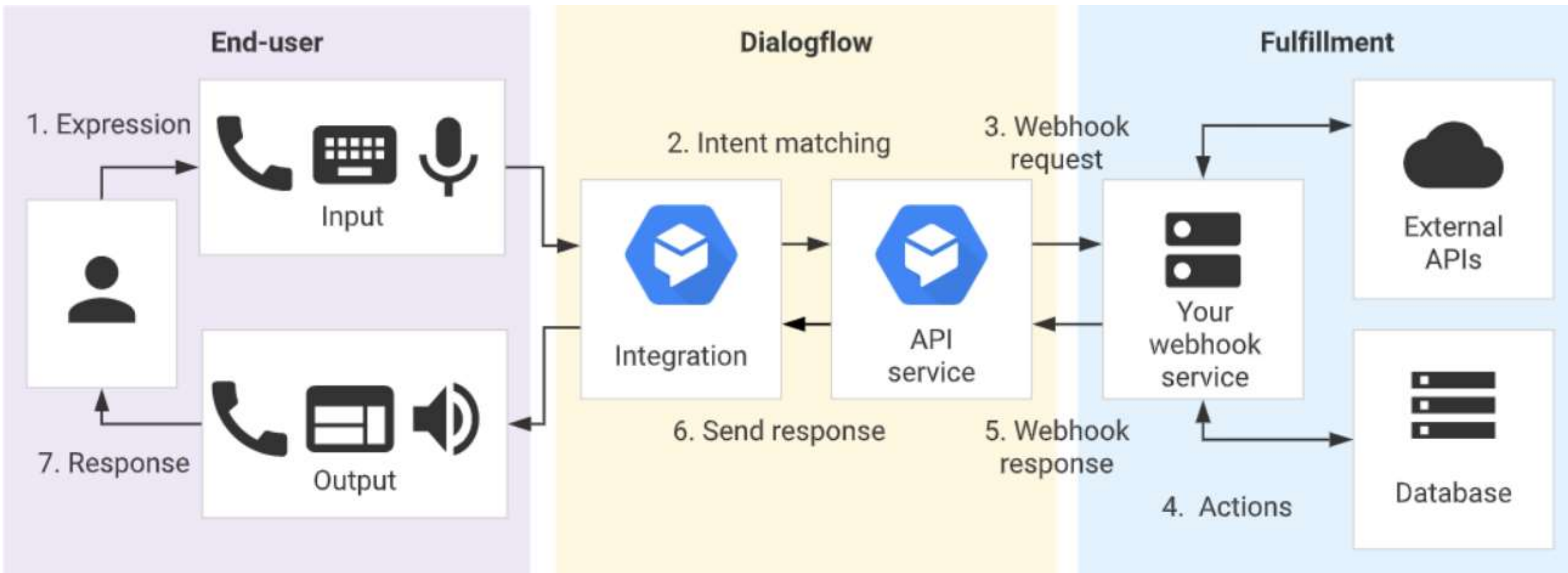
스마트 기기를 통해 **언제 어디서나** 시간과 장소에 구애 받지 않는 자유로운 상담 가능

업무 시간 외 야간, 주말, 공휴일에도 **24시간 문의 접수 및 상담 가능**

상담원 별 주관적 오류를 배제한 보다 **전문적이고 일괄적인 정보 전달 가능**

민원 데이터 분석을 통한 **피드백과 대처 방안 수립**으로 서비스 편의성 확대 및 만족도 향상

## Part 3 1) 챗봇 설계 및 구조도



사용자 문구 입력

문장의 형태 구분 후 자연어 처리

외부 서버에 저장된 자료와  
비교 후 답변

## Part 3 2) 데이터 수집 및 전처리

### 데이터 원본

A	B	C	D	E	F
시	구	동	의료기관명	주소	전화번호
서울	강남구	삼성동	강남구보건소	서울 강남구 삼성동(삼성2동) 8 강남구보건소	02-3423-5555
서울	강남구	일원동	삼성서울병원	서울 강남구 일원로81 삼성서울병원	02-1599-3114
서울	강남구	도곡동	강남세브란스병원	서울 강남구 언주로211 강남세브란스병원	02-1599-6114
서울	강남구	도곡동	강남베드로병원	남부순환로2637	02-1544-7522
서울	강동구	성내동	강동구보건소	서울 강동구 성내로45	02-3425-6713
서울	강동구	둔촌동	중앙보훈병원	서울 강동구 진황도로 61길 53	02-2225-1114
서울	강동구	상일동	강동경희대병원	서울시 강동구 동남로 892	02-440-7000
서울	강동구	길동	강동성심병원	서울시 강동구 성안로 150	1588-4100
서울	강북구	변동	강북구보건소	서울 강북구 한천로 897, 강북구보건소	02-901-7704,7706
서울	강서구	염창동	강서구보건소	서울 강서구 염창동 275-12 강서보건소	02-2600-5465
서울	관악구	정릉동	관악구보건소	서울 관악구 정릉동 1570-1	02-879-7133/7136
서울	관악구	신림동	에이치플러스양지병원	서울특별시 관악구 남부순환로 1636	1877-8875
서울	광진구	자양동	광진구보건소	서울특별시 광진구 자양1동 777	02-450-7090
서울	광진구	화양동	건국대학교병원	서울특별시 광진구 능동로 120-1	02-1588-1533
서울	구로구	구로동	구로구보건소	서울 구로구 구로동(구로제5동) 109-4 구로보건소	02-860-3000
서울	금천구	시흥동	금천구보건소	서울 금천구 시흥제1동 1020 금천구보건소	02-2627-2717
서울	금천구	시흥동	희명병원	서울특별시 금천구 시흥대로 244	02-2219-7231
서울	노원구	상계동	노원구보건소	서울 노원구 상계6.7동 노원구보건소, 701	02-2116-4900
서울	노원구	상계동	상계백병원	서울 노원구 동일로 1342	02-950-1114
서울	도봉구	쌍문동	도봉구보건소	서울 도봉구 쌍문동 565	02-2091-4490
서울	도봉구	쌍문동	한일병원	서울시 도봉구 우이천로 308(쌍문동)	02-901-3114
서울	동대문구	용두동	동대문구보건소	서울 동대문구 용두동 39-9	02-2127-4283
서울	도대문구	회계동	사육서울병원	서울 동대문구 망우로 87	1577-3675



### 데이터 전처리

Q	A
서울 강남구 삼성동	병원명은 강남구보건소 주소는 서울 강남구 삼성동(삼성2동) 8 강남구보건소 전화번호...
서울 강남구 일원동	병원명은 삼성서울병원 주소는 서울 강남구 일원로81 삼성서울병원 전화번호는 02-1...
서울 강남구 도곡동	병원명은 강남세브란스병원 주소는 서울 강남구 언주로211 강남세브란스병원 전화번호는...
서울 강남구 도곡동	병원명은 강남베드로병원 주소는 남부순환로2637 전화번호는 02-1544-7522 ...
서울 강동구 성내동	병원명은 강동구보건소 주소는 서울 강동구 성내로45 전화번호는 02-3425-671...
...	...
서울 종로구 평동	병원명은 강북삼성병원 주소는 서울시 종로구 새문안로 29 전화번호는 02-2001-...
서울 종로구 평동	병원명은 서울적십자병원 주소는 서울시 종로구 평동 164 전화번호는 02-2002-...
서울 중랑구 신내동	병원명은 중랑구보건소 주소는 서울 중랑구 신내2동 662 중랑구청 전화번호는 02...
서울 중랑구 신내동	병원명은 서울의료원 주소는 중랑구 신내로 156 전화번호는 02-2276-8333 ...
서울 중랑구 면목동	병원명은 녹색병원 주소는 중랑구 사가정로 49길 53 전화번호는 02-490-200...

## Part 3 2) 데이터 수집 및 전처리

임베딩이란? 컴퓨터가 문자를 해석하는 방법

### ASCII TABLE

love는 live보다 like와 더 유사하다. 하지만 컴퓨터가 보기에는 love는 live와 더 비슷해 보인다.

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

단어	1	2	3	4
love	0x6C	0x6F	0x76	0x65
live	0x6C	0x69	0x76	0x65
like	0x6C	0x69	0x6B	0x65

A -> [0.1, 0.0, 0.0]

B -> [0.0, 0.0, 1.0]

C -> [0.0, 0.2, 1.0]

D -> [0.2, 0.2, 0.8]

D -> [0.5, 0.5, 0.8]

컴퓨터는 ASCII, 유니코드, UTF-8 encoding 등으로 문자를 표현하고 저장한다.



## Part 3 2) 데이터 수집 및 전처리

### 질문 데이터 임베딩

```
import csv
with open('embedding_SB.csv', 'w') as file : #선별진료소
    mapdata = []
    write = csv.writer(file)
    for temp in df['Q']:
        t = model.encode(temp)
        mapdata.append( t )

    write.writerows(mapdata)

#csv 불러오기
df = pd.read_csv('./병원정보/서울 선별진료소 질문답변1.csv')
df1 = pd.read_csv('embedding_SB.csv', header=None)
```




### 입력 텍스트와 임베딩 데이터 비교

```
def chatbot_text(text):
    em_result = model.encode(text)
    co_result = []
    for temp in range(len(df1)):
        data = df1.iloc[temp]
        co_result.append(cosine_similarity([em_result],
        [data])[0][0])
    df['cos'] = co_result
    df_result = df.sort_values('cos', ascending=False)
    r = random.randint(0,5)
    return df_result.iloc[0:5]

chatbot_text("서울 관악구 신림동")
```

Q	A	cos
서울 관악구 신림동	병원명은 에이치플러스양지병원 주소는 서울특별시 관악구 남부순환로 1636 전화번호는 ...	1.000000
서울 영등포구 대림동	병원명은 명지성모병원 주소는 서울특별시 영등포구도림로 156 전화번호는 02-829...	0.996786
서울 관악구 청룡동	병원명은 관악구보건소 주소는 서울 관악구 청룡동 1570-1 전화번호는 02-87...	0.996690
서울 영등포구 신길동	병원명은 한림대학교 강남성심병원 주소는 서울특별시 영등포구 신길로 1 전화번호는 1...	0.996552
서울 영등포구 신길동	병원명은 의료법인 성애병원 주소는 서울특별시 영등포구 여의대방로 53길22 전화번호...	0.996552

## Part 3 3) Dialogflow 구성

 **Entities**

Custom

System

@ keyword

@ name\_dong

@ name\_gu

@ name\_seoul



name\_dong SAVE

상암동

상암동

중현동

중현동, 중정로2가, 중정로3가, 합동, 미군동, 북이현동

천연동

천연동, 내천동, 오천동, 영천동, 영지동

연희동

서울 자치구

홍제동

426

홍은동

467

남가좌동

북가좌동

서초동

잠원동

잠원동

반포동

반포동, 반포본동, 반포재1동, 반포재2동, 반포재3동, 반포재4동

방배동

방배동, 방배본동, 방배재1동, 방배재2동, 방배재3동, 방배재4동

양재동

양재동, 양재재1동, 양재재2동



## Part 3 3) Dialogflow 구성

### 대화 흐름 유도

**Text Response**

- 1 안녕하세요 고객님,  
저희 홈페이지를 방문해주셔서 감사합니다.  
아래 유형을 선택해주세요.
- 2 Enter a text response variant

**Custom Payload**

```
1 {  
2   "richContent": [  
3     {  
4       "options": [  
5         {  
6           "text": "주변 병원 정보"  
7         },  
8         {  
9           "text": "코로나 관련 Q&A"  
10        },  
11        {  
12         "text": "지도맵 보기",  
13         "link": "https://d19b-t12-221-224-125.jp.ngrok.io/map"  
14        },  
15        ],  
16      },  
17    ],  
18    "type": "chips"  
19  }  
20 }
```



**Chatbot**

안녕하세요 고객님,  
저희 홈페이지를 방문해주셔서 감사  
합니다.  
아래 유형을 선택해주세요

주변 병원 정보    코로나 관련 Q&A

지도맵 보기

질문을 입력하세요...



**Chatbot**

주변 병원 정보

현 주소를 입력해 주세요.

처음으로

서울 관악구 신림동

**선별진료소**  
병원명은 에이치플러스양지병원 주소는 서울특별시 관악구 남부순환로 1636 전화번호는 1877-8875 입니다.

**호흡기 진료센터**  
병원명은 수정형외과의원 주소는 서울특별시 관악구 난곡로 349, 하모니타워 2층 (신림동) 전화번호는 02-6671-8275 입니다.

질문을 입력하세요...

## Part 3 3) Dialogflow 구성

**Chatbot**

주변 병원 정보

현 주소를 입력해 주세요.

처음으로

서울 관악구 신림동

**선별진료소**  
병원명은 에이치플러스양지병원 주소는 서울특별시 관악구 남부순환로 1636 전화번호는 1877-8875 입니다.

**호흡기 진료센터**  
병원명은 수정형외과의원 주소는 서울특별시 관악구 난곡로 349, 하모니타워 2층 (신림동) 전화번호는 02-6671-8275 입니다.

질문을 입력하세요...

**Fulfillment** ?

☒ Enable webhook call for this intent

☐ Enable webhook call for slot filling

**Webhook** ENABLED

Your web service will receive a POST request from Dialogflow in the form of the response to a user query matched by intents with webhook enabled. Be sure that your web service meets all the [webhook requirements](#) specific to the API version enabled in this agent.

URL\* https://d6b3-112-221-224-125.jp.ngrok.io/Question

BASIC AUTH

Enter username

Enter password

HEADERS

Enter key

Enter value

## Part 3 3) Dialogflow 구성

Custom Payload


```
1 {  
2   "richContent": [  
3     [  
4       {  
5         "rawUrl": "https://i.postimg.cc/nz5RjMmP/254938.jpg",  
6         "type": "Image",  
7         "accessibilityText": "Dialogflow across platforms"  
8       },  
9       {  
10        "title": "코로나 관련 Q&A",  
11        "actionLink": "https://d19b-112-221-224-125.jp.ngrok.io/QnA",  
12        "type": "Info",  
13        "subtitle": "클릭하시면 해당 주소로 이동합니다."  
14      },  
15      {  
16        "options": [  
17          {  
18            "text": "처음으로"  
19          }  
20        ],  
21        "type": "chips"  
22      }  
23    ]  
24  }  
25 }
```



**Chatbot**

자신의 플레이아웃을 방문해주시기 바랍니다.  
아래 유형을 선택해주세요

코로나 관련 Q&A



**코로나 관련 Q&A**  
클릭하시면 해당 주소로 이동합니다.

처음으로

질문을 입력하세요...

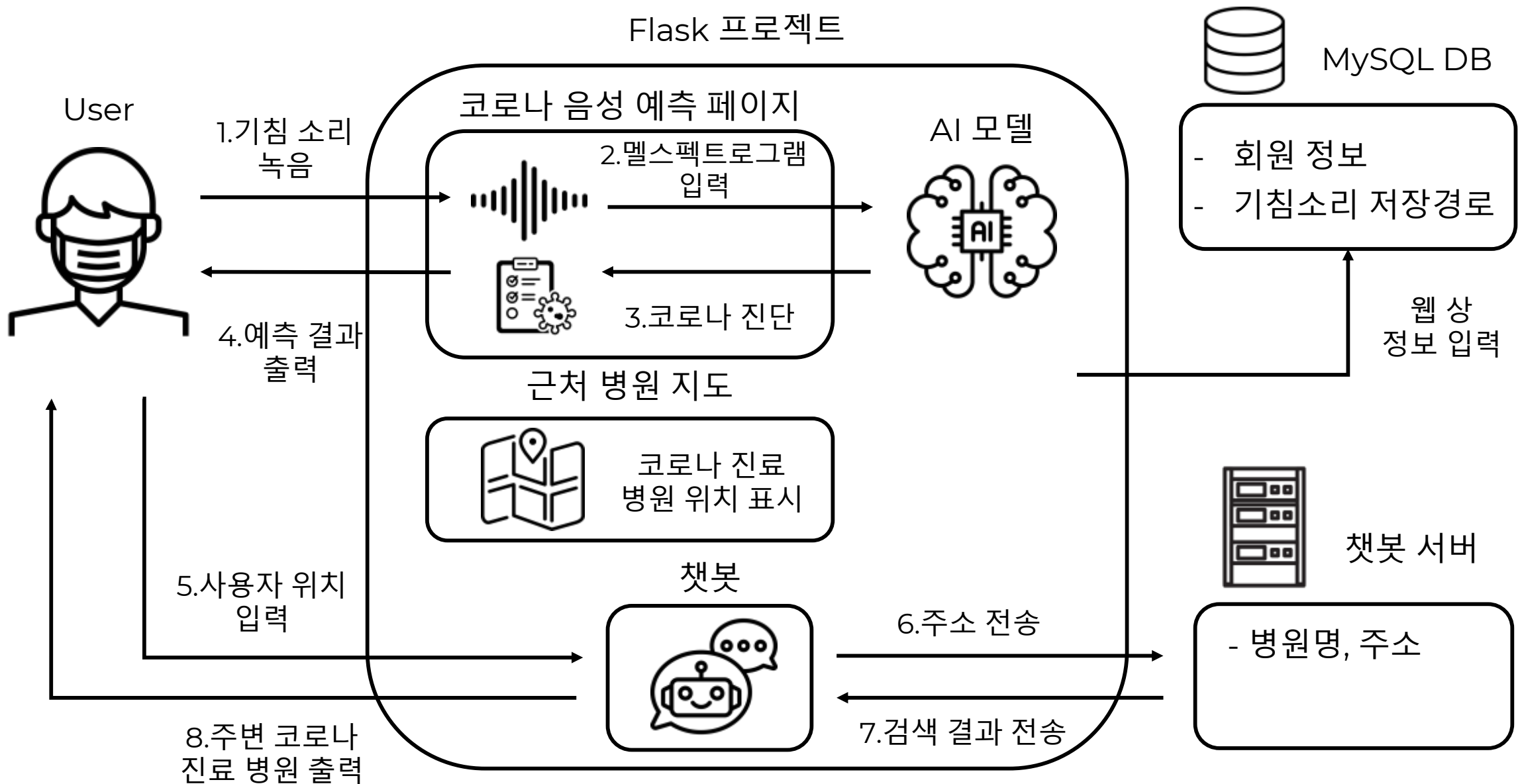
# Part. 4

## 플라스크



- 01. 플라스크 설계
- 02. 화면 구성
- 03. 코로나 접종 현황판
- 04. 목소리 유사도
- 05. 코로나 예측 그래프

## Part 4 1) 플라스크 설계



## Part 4 2) 화면 구성

<홈 화면>



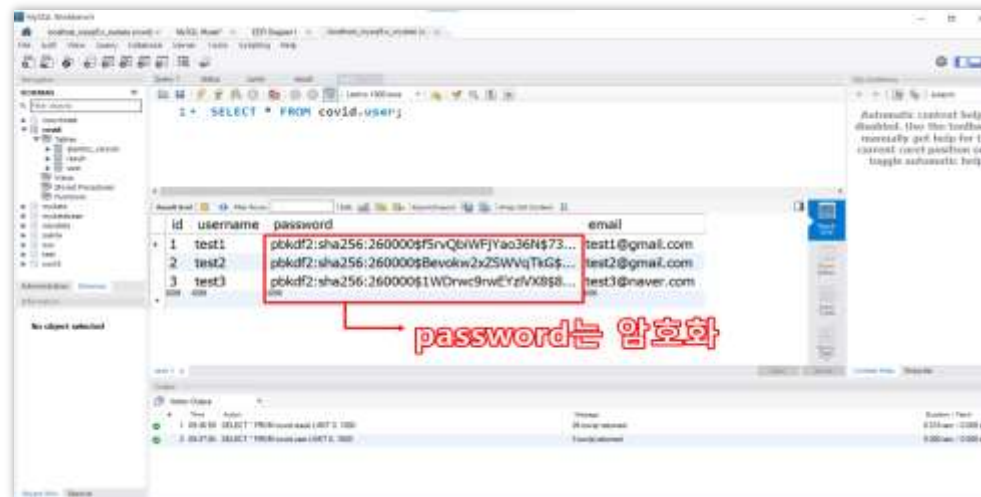
<로그인 화면>



<회원가입 화면>



<회원가입 데이터베이스>



## Part 4 2) 화면 구성

<Covid AI 검증 화면>



<검증 결과(음성) 화면>



<검증 결과(양성) 화면>





## Part 4 2) 화면 구성

### <지도 화면>



### <확진시 조치 안내 화면>



### <코로나 Q&A 화면>

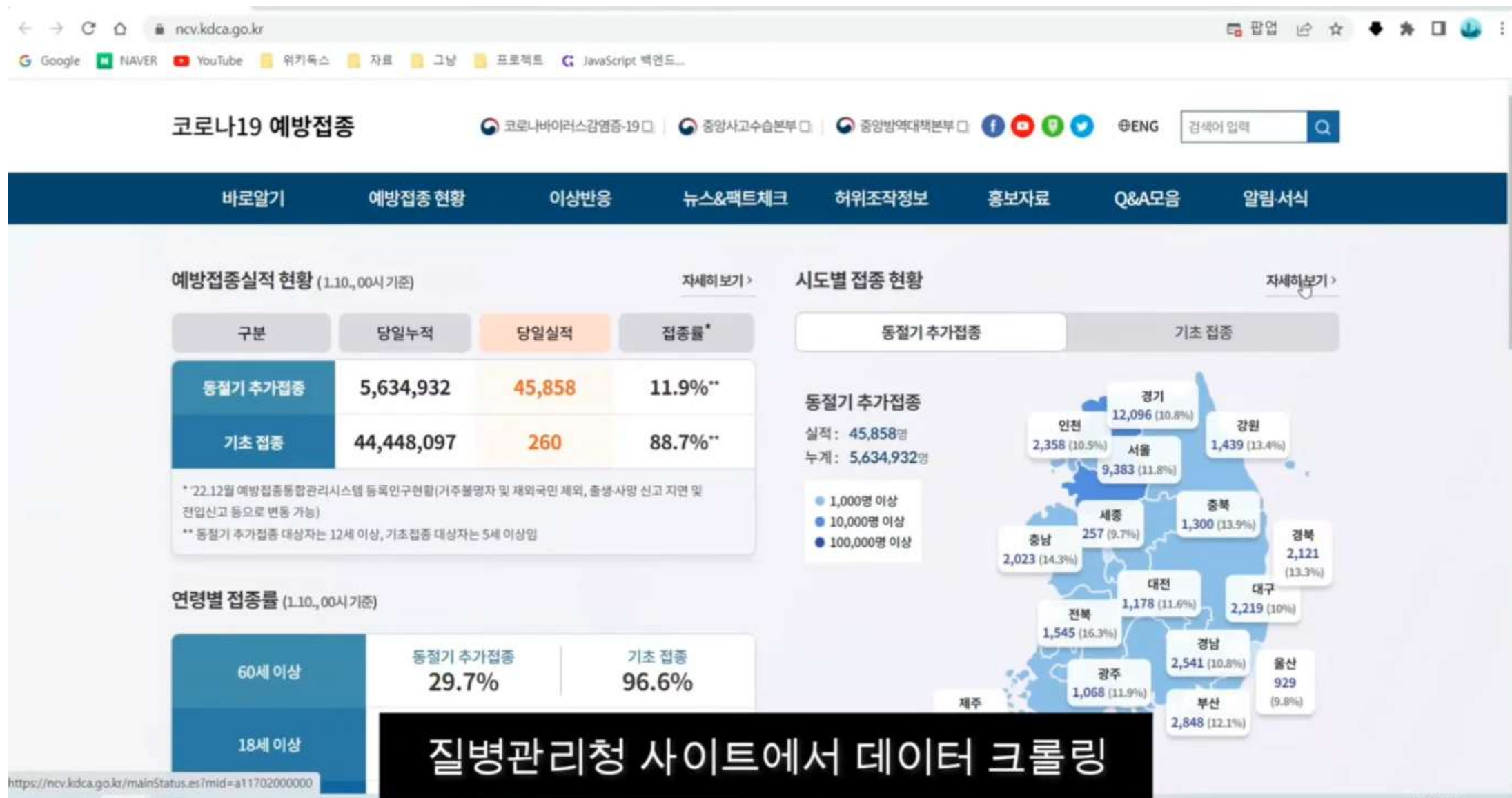


### <코로나 예방접종 현황 화면>



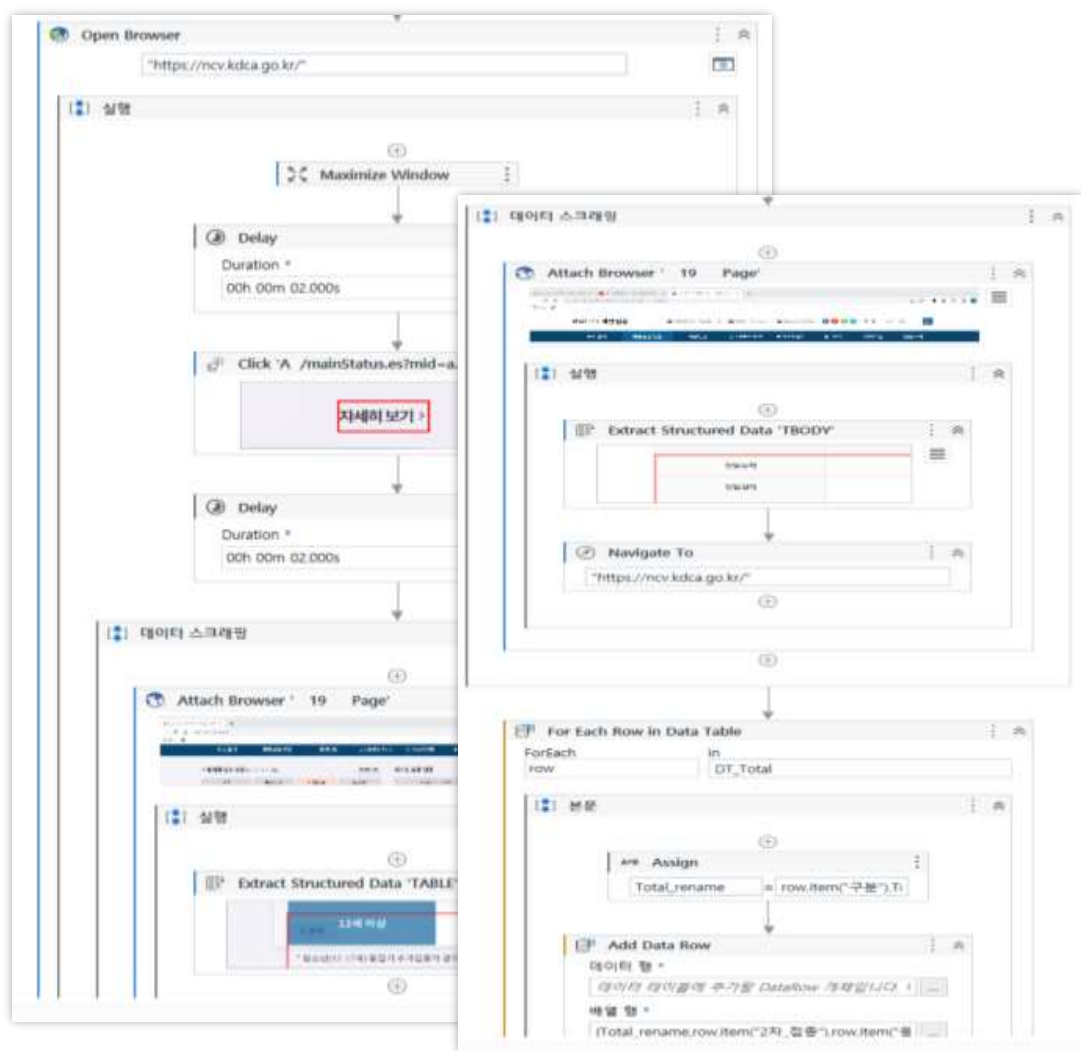


## Part 4 3) 코로나 접종 현황판



## Part 4 3) 코로나 접종 현황판

<UiPath 코드>



<추출 데이터>

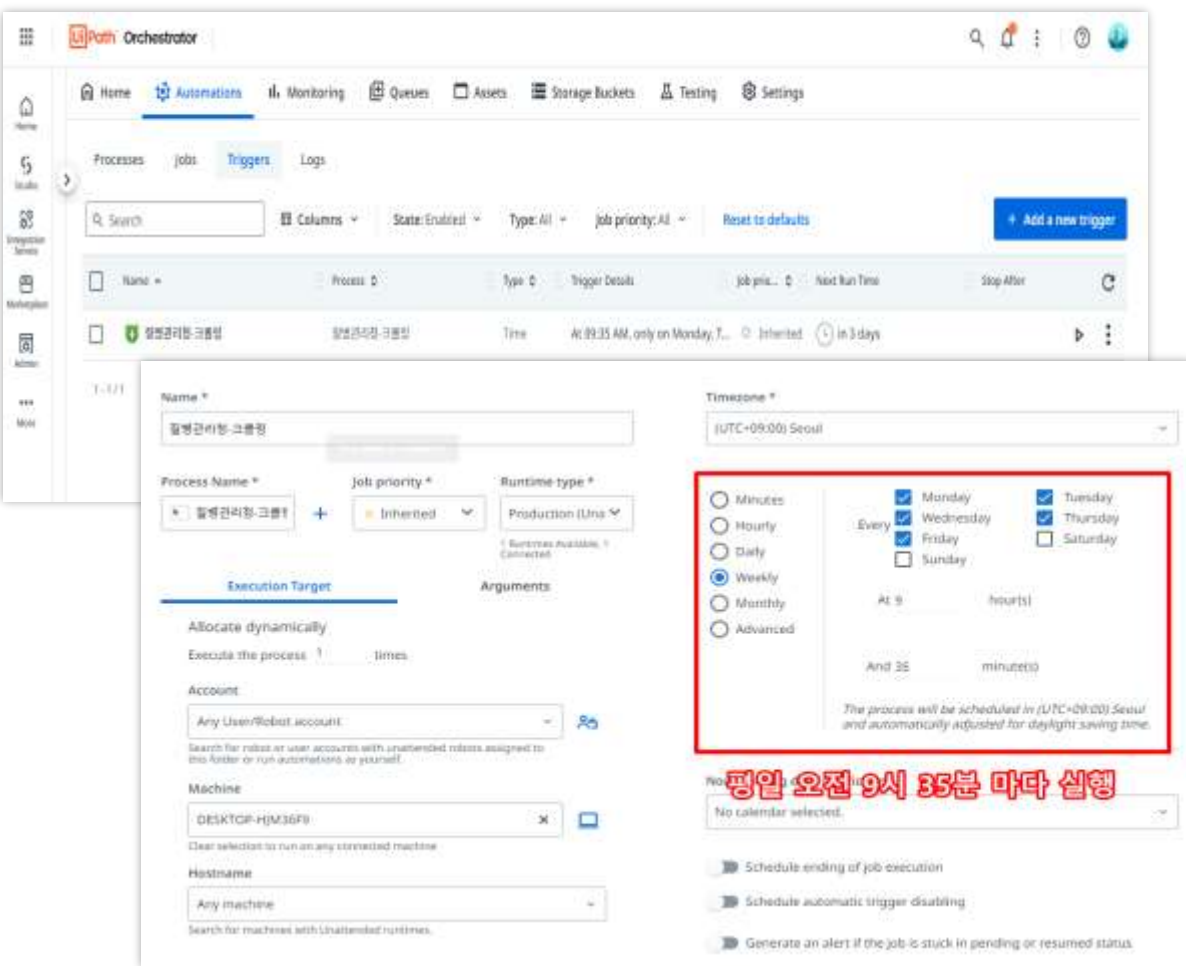
	A	B	C	D	E
1	구분	동절기_신	동절기_접	기초_신규	기초_접종률
2	서울	173	11.70%	1	89.50%
3	부산	23	12%	1	87%
4	대구	6	9.90%	0	85.50%
5	인천	61	10.40%	0	88.50%
6	광주	5	11.80%	0	88.80%

	A	B	C	D
1	구분	2차_접종	동절기_추가접종	
2	당일_누적	44,447,587	5,588,675	
3	당일_실적		3	694
4	접종률	88.70%	11.80%	
5				

	A	B	C
1	나이	추가접종	기초접종
2	60세 이상	29.50%	96.60%
3	18세 이상	12.60%	96.70%
4	12세 이상	11.80%	94.20%
5			

## Part 4 3) 코로나 접종 현황판

# 오케스트레이터



## Flask 연결 코드

```
from flask import Blueprint, url_for, render_template, flash, request, redirect, escape, jsonify
import pandas as pd
from datetime import datetime
from apscheduler.schedulers.background import BackgroundScheduler

bp = Blueprint('crawling', __name__, url_prefix='/crawling')

data = pd.DataFrame()
data1 = pd.DataFrame()
data2 = pd.DataFrame()

def scheduler():
    global data
    global data1
    global data2

    data = pd.read_csv(f'../dataset/{datetime.now().strftime("%Y%m%d")}_시도별점중변환.csv', index_col=0)
    data1 = pd.read_csv(f'../dataset/{datetime.now().strftime("%Y%m%d")}_면역발점종류.csv')
    data2 = pd.read_csv(f'../dataset/{datetime.now().strftime("%Y%m%d")}_예방접종심적현황.csv', index_col=0)
    print('Scheduler 실행')

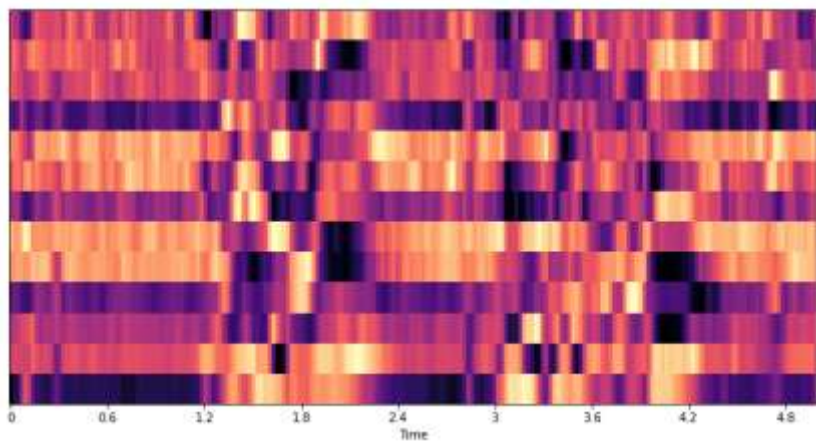
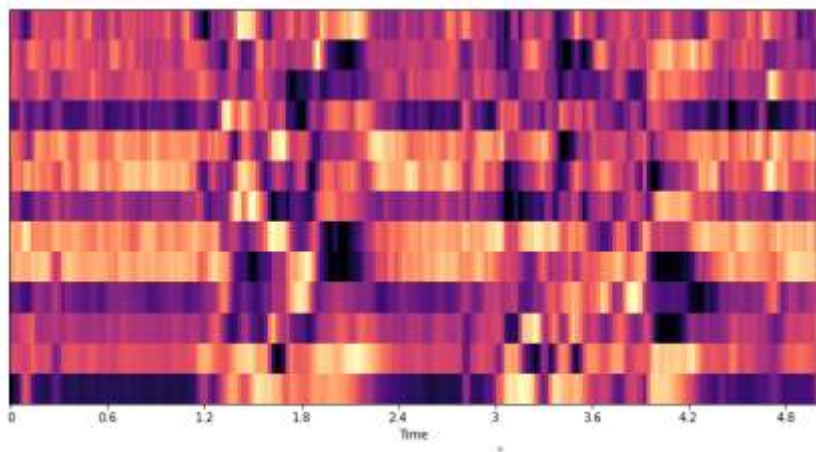
schedule = BackgroundScheduler(daemon=True, timezone='Asia/Seoul')
schedule.add_job(scheduler, 'cron', day_of_week='mon-fri', hour='09', minute='40')
schedule.start()
```

평일 오전 9시 40마다 실행

```
@bp.route('/')
def index():
    return render_template(
        'covidashboard.html', date=datetime.now().strftime("%m.%d"),
        sido=data, age=data1, vaccination=data2.T)
```

## Part 4 4) 목소리 유사도

<비교할 전/후 목소리>



<목소리간 벡터 값>

	0	1	2	3	4
원본	-2.348714	-2.057745	-1.957048	-0.345245	0.454251
테스트	-2.314313	-2.023952	-1.968540	-0.463136	0.428862

<cos 유사도 결과값>

	0	1
0	1.000000	0.974162
1	0.974162	1.000000



## Part 4 4) 목소리 유사도

### <처음 목소리 녹음 화면>



### <비교 목소리 녹음 화면>



### <목소리 비교 화면>



## Part 4 5) 코로나 예측 그래프

프로펫 사용  
목적

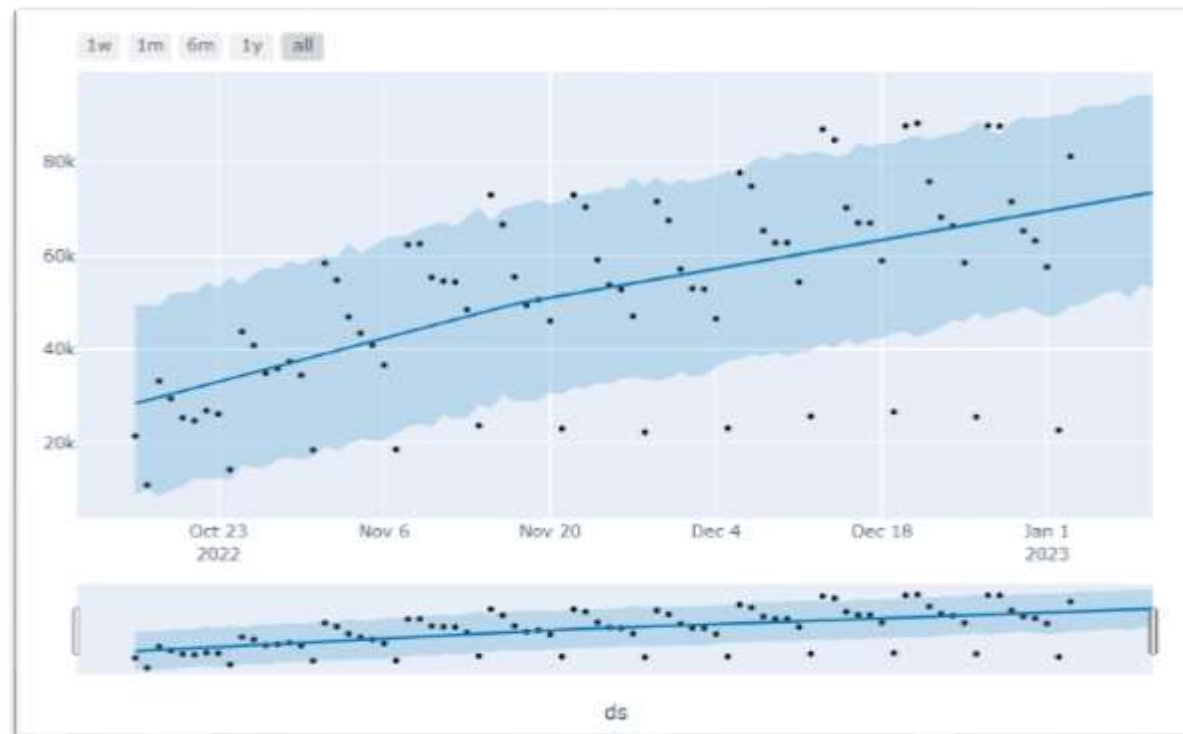
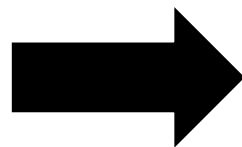
최근 30일을 기준으로 앞으로의 코로나 **확진 동향을 예측**

확진자 예측 데이터를 바탕으로 **앞서 정책 및 계획을 수립**해 사회적 비용을 감소

## Part 4 5) 코로나 예측 그래프

### Prophet 예측 모델 실행

```
1 m = Prophet(  
2     changepoint_prior_scale=0.2,  
3     changepoint_range=0.9,  
4     yearly_seasonality=False,  
5     weekly_seasonality=False,  
6     daily_seasonality=True,  
7     seasonality_mode='additive'  
8 )  
9  
10 m.fit(df_prophet)  
11  
12 future = m.make_future_dataframe(periods=7,  
13 forecast = m.predict(future)  
14  
15 fig = plot_plotly(m, forecast)  
16 py.iplot(fig)
```





## Part 4 5) 코로나 예측 그래프



## Part 4 5) 코로나 예측 그래프



©2022 누가 기침소리를 내었조



# Part. 5

## 마무리



- 01. 보완할 점
- 02. 느낀 점
- 03. 출처

# 보완할 점



음성 데이터를 저장할 때  
재생 시간과 기침의 시작지점을  
맞춰서 학습 성능을 높여보고  
싶습니다.

김영원



평소 챗봇을 잘 안 쓰다  
보니 떠오르는 아이디어가 많지  
않았고 타 사이트 내에 챗봇  
기능들을 살펴보니 추가할만한  
기능들이 많았기에 아쉬움

정호영



음성 녹음을 웹페이지에서  
사용할 때 시간을 지정할 수  
있도록 하거나 기침소리를  
편집하는 기능을 추가해서  
사용자가 편하게 이용하게  
하고 싶습니다.

김혜원



김영원 : <https://github.com/Sodness>

정호영 : <https://github.com/HoYoung12>

김혜원 : <https://github.com/kimhyewon2323>

# 느낀 점



음성데이터는 어떤 방식으로 다루는지에 대해 경험해볼 수 있는 기회여서 다음에 비슷한 프로젝트를 할 때 도움이 될거 같습니다.

김영원



팀프로젝트다 보니 혼자서라면 못했을 것 들이나 놓칠 만한 부분들을 다른 사람이 채워줄 수 있다는 것을 느꼈고, 커뮤니케이션의 중요성 또한 알게 되었습니다.

정호영



팀프로젝트라 협업의 대한 중요성에 대해 느끼게 되었고, 음성 데이터의 대한 경험을 배울 수 있어서 재미있는 프로젝트였습니다.

김혜원



김영원 : <https://github.com/Sodness>

정호영 : <https://github.com/HoYoung12>

김혜원 : <https://github.com/kimhyewon2323>

**7 슬라이드:** <https://www.doctorsnews.co.kr/news/articleView.html?idxno=143658>

**25 슬라이드:** <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a>

**26 슬라이드:** <https://www.slideshare.net/xavigiro/image-classification-on-imagenet-d1l4-2017-upc-deep-learning-for-computer-vision/>

**27 슬라이드:** [https://www.researchgate.net/figure/Algorithms-that-won-the-ImageNet-Large-Scale-Visual-Recognition-Challenge-ILSVRC-in\\_fig2\\_346091812](https://www.researchgate.net/figure/Algorithms-that-won-the-ImageNet-Large-Scale-Visual-Recognition-Challenge-ILSVRC-in_fig2_346091812)

**45 슬라이드:** <https://cloud.google.com/dialogflow/es/docs/fulfillment-overview?hl=ko>

**47 슬라이드:** [https://heung-bae-lee.github.io/2020/01/16/NLP\\_01/](https://heung-bae-lee.github.io/2020/01/16/NLP_01/)

**로그인,회원가입 페이지:** <https://getbootstrap.com/docs/4.0/examples/sign-in/>

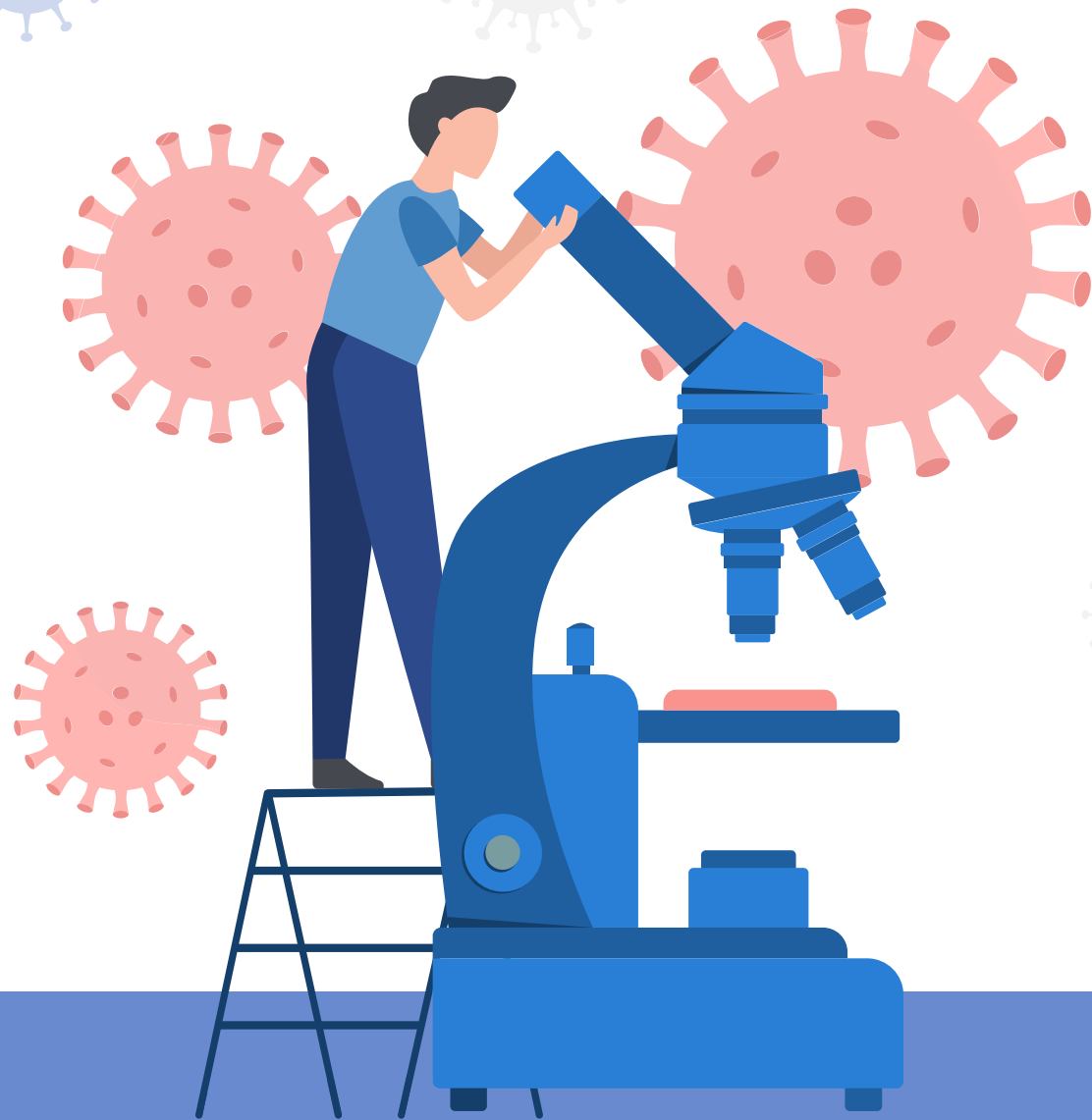
**홈페이지:** <https://nicepage.com/html-templates/preview/stop-covid-19-76644?device=desktop>

**코로나 예방접종 현황 페이지:** <https://ncv.kdca.go.kr/>

**코로나 누적 확진자 데이터:** [https://ncov.kdca.go.kr/tcmBoardList.do?brdId=3&brdGubun=30&dataGubun=&ncvContSeq=&contSeq=&board\\_id=&gubun=](https://ncov.kdca.go.kr/tcmBoardList.do?brdId=3&brdGubun=30&dataGubun=&ncvContSeq=&contSeq=&board_id=&gubun=)

**코로나 QnA 페이지:** <https://ncv.kdca.go.kr/menu.es?mid=a12217000000>

**확진시 조치안내 페이지:** <https://ncv.kdca.go.kr/menu.es?mid=a30401000000>



**Thank you**



# Questions and Answers

