

세미프로젝트 1차

빅데이터 기반 지능형 서비스 개발 A반

INDEX

01 프로젝트 배경

프로젝트 주제
선정이유
개요

02 프로젝트 팀소개

팀명 소개
구성 및 역할

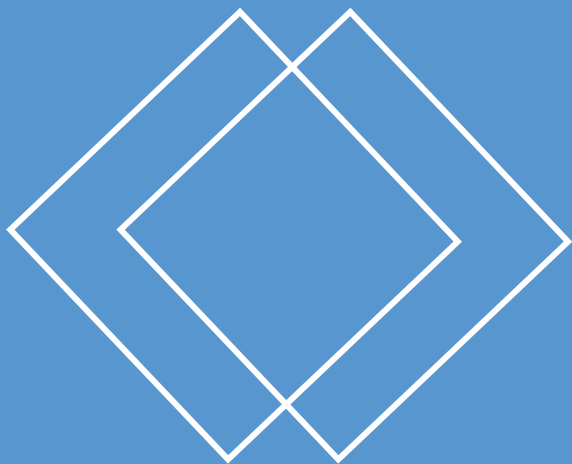
03 프로젝트 수행절차 및 방법

수행절차
데이터 수집
데이터 전처리
데이터 분석
시각화

04 프로젝트 수행 결과 및 활용방안

워드클라우드
예측 모델
지도 시각화

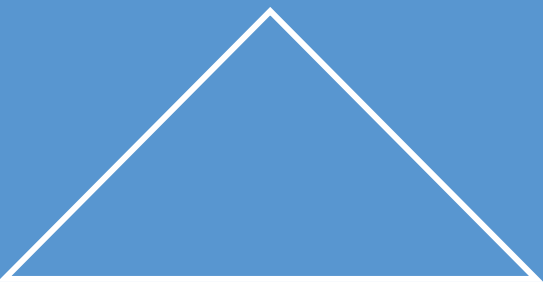
05 프로젝트 후기 및 느낀점



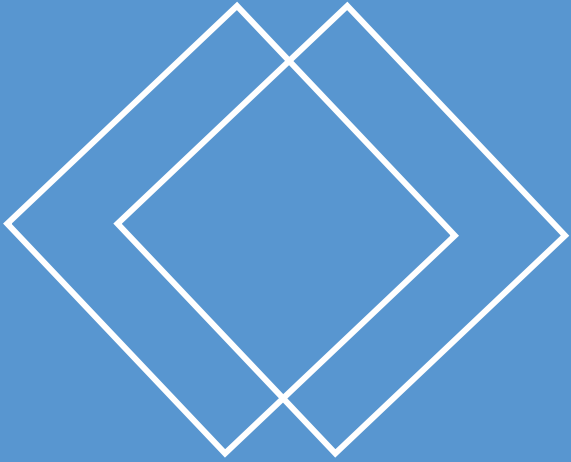
01

프로젝트 배경

프로젝트 주제
선정이유
개요



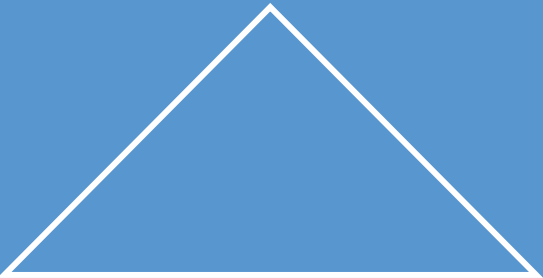




02

프로젝트 팀소개

팀명 소개
구성 및 역할



프로젝트 팀소개


빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

1. 팀명 소개



일 칠 오 공

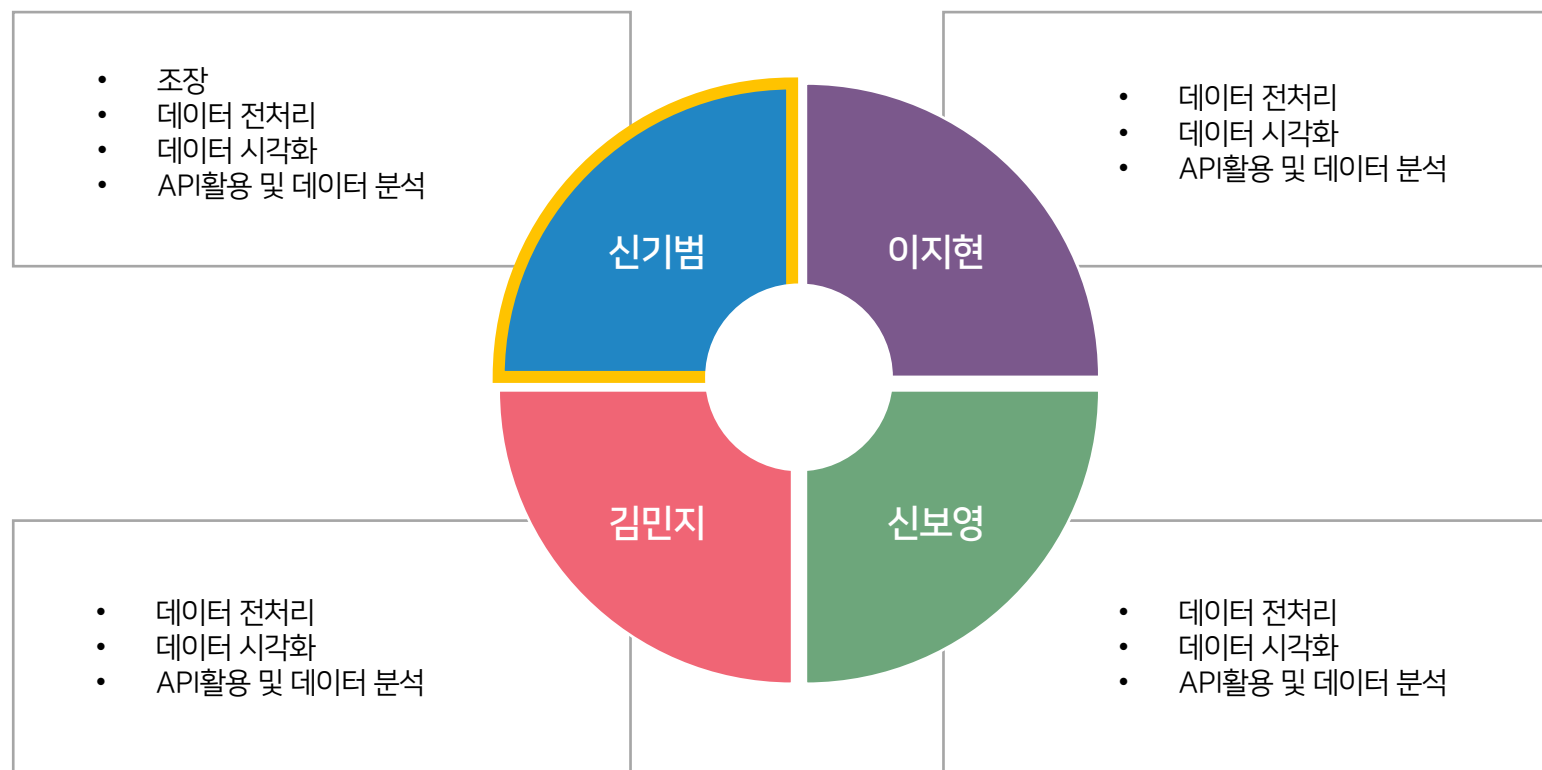
17시 50분을 의미하며, 수업을 마치는 시간입니다.
하루 중 가장 행복한 시간이기도 하고, 모두에게 기억되기
쉽겠다고 생각하여 정했습니다.

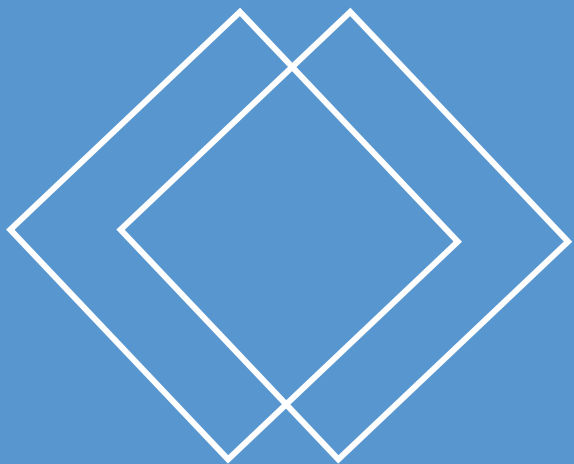


프로젝트 팀소개

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

2. 구성원 및 역할

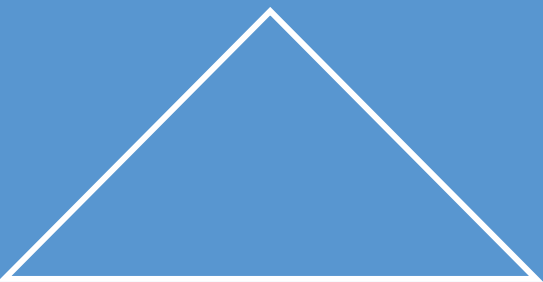




03

프로젝트 절차 및 수행방법

수행절차
데이터 수집
데이터 전처리
데이터 분석
시각화



프로젝트 절차 및 수행방법

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

1. 수행절차

01	02	03	04
데이터 수집	전처리	시각화	데이터 분석
Googling OpenCV	Pandas Numpy	Folium Power BI	Sklearn
Jupyter notebook & Colab & Spyder			
Zoom & Google Drive			

프로젝트 절차 및 수행방법

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

2. 데이터 수집

상세 내용	진료 내역 데이터	전국 의료기관 데이터	인구 현황 데이터
항목	연령, 시도, 성별, 주상병, 부상병, 요양일 수, 입내원일 수, 심결가산율	영업상태, 주소, 상업장명, 업태구분, 병상수, 진료과목	제주도, 울산광역시, 부산광역시, 대전광역시, 전주시 등 연령대 인구
기간	2019 ~ 2020 (2년)	2021년 기준	2021년도 기준
파일형식	csv 파일	csv 파일	Excel, csv 파일
출처	 h·well 국민건강보험	 행정안전부	공공데이터포털

프로젝트 절차 및 수행방법

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

3-1. 전처리 - 진료데이터 전처리 과정 : 컬럼 추출 > 컬럼 분리 > 타입 변경

✓ 전체 데이터 중 예측모델에 사용할 컬럼만 추출하기

- feature : 연령대코드, 성별코드, 주상병코드, 부상병코드
- label : 입내원일수

기준 년도	가입자 일련번호	진료내역일 련번호	성별 코드	연령 대코드	시도 코드	요양개 시일자	서식 코드	진료 과목 코드	주상 병코드	부상병 코드	요양 일수	입내 원일 수	심결 가산 율	심결요양급 여비용총액	심결본 인부담 금	심결보험 자부담금	총처 방일수	데이터 기준일자	
0	2020	29	3724043	1	17	41	2020-12-21	3	1	I109	E785	1	1	0.15	11540	1500	10040	0	2021-12-28
1	2020	29	3495225	1	17	41	2020-12-14	3	1	I109	E785	1	1	0.15	11540	1500	10040	7	2021-12-28
2	2020	255	10615350	2	18	41	2020-04-20	3	1	I109	I251	1	1	0.15	11540	1500	10040	30	2021-12-28
3	2020	255	10616553	2	18	41	2020-04-16	3	1	I109	I251	1	1	0.15	11540	1500	10040	30	2021-12-28
4	2020	2406	3105303	2	17	41	2020-12-30	3	1	I109	E780	1	1	0.15	11540	1500	10040	0	2021-12-28



연령대코드	성별코드	주상병코드	부상병코드	입내원일수	
0	17	1	I109	E785	1
1	17	1	I109	E785	1
2	18	2	I109	I251	1
3	18	2	I109	I251	1
4	17	2	I109	E780	1

프로젝트 절차 및 수행방법

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

3-1. 전처리 - 진료데이터 전처리 과정 : 컬럼 추출 > 컬럼 분리 > 타입 변경

✓ 데이터 추출하여 컬럼 추가하기

- 범주형 데이터로는 예측모델을 만들지 못하기 때문에 연속형 데이터로 만드는 작업 수행
- 주상병코드와 부상병코드의 가장 앞 알파벳을 추출하여 새로운 컬럼 생성
- 아스키코드를 이용하여 A~Z를 1~26까지의 숫자로 표현

연령대코드	성별코드	주상병코드	부상병코드	입내원일수	
0	17	1	I109	E785	1
1	17	1	I109	E785	1
2	18	2	I109	I251	1
3	18	2	I109	I251	1
4	17	2	I109	E780	1



연령대코드	성별코드	주상병알파벳	주상병숫자	부상병알파벳	부상병숫자	입내원일수	
0	17	1	9	109	5	785	1
1	17	1	9	109	5	785	1
2	18	2	9	109	9	251	1
3	18	2	9	109	9	251	1
4	17	2	9	109	5	780	1

프로젝트 절차 및 수행방법

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

3-1. 전처리 - 진료데이터 전처리 과정 : 컬럼 추출 > 컬럼 분리 > 타입 변경

✓ 데이터 타입 변경하기

- 주상병숫자와 부상병숫자 컬럼에 포함된 '_'를 '9000'으로 변환 후 타입 변경

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11727248 entries, 0 to 11727247
Data columns (total 7 columns):
#   Column      Dtype
---  -
0   연령대코드  int64
1   성별코드    int64
2   주상병알파벳 int64
3   주상병숫자  int64
4   부상병알파벳 int64
5   부상병숫자  int64
6   입내원일수  int64
dtypes: int64(7)
memory usage: 626.3 MB
```

연령대코드	성별코드	주상병알파벳	주상병숫자	부상병알파벳	부상병숫자	입내원일수
7191	14	1	-	11	297	1
7620	14	2	-	11	297	1
8250	14	2	-	11	297	1
8345	14	1	-	11	297	1
8362	15	2	-	11	297	1



연령대코드	성별코드	주상병알파벳	주상병숫자	부상병알파벳	부상병숫자	입내원일수
7191	14	1	9000	11	297	1
7620	14	2	9000	11	297	1
8250	14	2	9000	11	297	1
8345	14	1	9000	11	297	1
8362	15	2	9000	11	297	1

프로젝트 절차 및 수행방법

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

3-2. 전처리 - 병원데이터 전처리 과정 : 컬럼 추출 > 조건 추출 > 카카오프API

✓ 전체 데이터 중 시각화에 사용할 컬럼만 추출하기

- 영업상태명, 도로명전체주소, 사업장명

개방서비스명	개방서비스아이디	개방자치단체코드	관리번호	인허가일자	인허가취소일자	영업상태구분코드	영업상태명	상세영업상태코드	...	지정취소일자	완화의지정형태	완화의담당부서명	구급자특수	구급자일반	총인원	구조사수	허가병상수	최초지정일자	
0	1	01_01_01_P	5530000	PHMA220115530021021200002	20110221	NaN	2	휴업	02	...	NaN	NaN	NaN	0.0	0.0	NaN	NaN	0.0	NaN
1	2	01_01_01_P	5710000	PHMA220185710109021200001	20180601	NaN	2	휴업	02	...	NaN	NaN	NaN	0.0	0.0	NaN	NaN	0.0	NaN
2	3	01_01_01_P	4430000	PHMA220084430014021200003	20080414	NaN	2	휴업	02	...	NaN	NaN	NaN	0.0	0.0	NaN	NaN	0.0	NaN
3	4	01_01_01_P	4470000	PHMA220164470017021200001	20160325	NaN	2	휴업	02	...	NaN	NaN	NaN	0.0	0.0	NaN	NaN	0.0	NaN



영업상태명	도로명전체주소	사업장명
0 휴업	경기도 화성시 팔탄면 온천로 29	의료법인 화성의료재단 화성제일병원
1 휴업	충청북도 청주시 청원구 내수읍 마산길 8-11	의료법인홍익의료재단파스카요양병원
2 휴업	충청북도 옥천군 옥천읍 중앙로 72	의료법인건우의료재단큰사랑요양병원
3 휴업	충청북도 음성군 감곡면 장감로 118-1	의료법인 청미의료재단 청미요양병원
4 휴업	충청남도 계룡시 계룡대로 295 (금암동)	의료법인 계룡의료재단 계룡병원

프로젝트 절차 및 수행방법

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

3-2. 전처리 - 병원데이터 전처리 과정 : 컬럼 추출 > 조건 추출 > 카카오API

✓ 현재 영업 중인 데이터만 추출하기

- 영업상태명이 '영업/정상' 인 데이터 추출

	영업상태명	도로명전체주소	사업장명
0	휴업	경기도 화성시 팔탄면 온천로 29	의료법인 화성의료재단 화성제일병원
1	휴업	충청북도 청주시 청원구 내수읍 마산길 8-11	의료법인홍익의료재단파스카요양병원
2	휴업	충청북도 옥천군 옥천읍 중앙로 72	의료법인건우의료재단큰사랑요양병원
3	휴업	충청북도 음성군 감곡면 장감로 118-1	의료법인 청미의료재단 청미요양병원
4	휴업	충청남도 계룡시 계룡대로 295 (금암동)	의료법인 계룡의료재단 계룡병원



	영업상태명	도로명전체주소	사업장명
2088	영업/정상	대전광역시 대덕구 동서대로1777번길 2-0, 대전웰니스병원충 (송촌동,대전웰니스병원)	대전웰니스병원
2089	영업/정상	경상북도 경산시 경산로 174-0, A동 양지요양병원1~3층충 (옥산동,양지요양병원...	의료법인양지의료재단 양지기쁨병원
2090	영업/정상	전라남도 목포시 수강로12번길 11-1 (행복동2가)	의료법인항운의료재단 목포노동병원
2091	영업/정상	전라남도 목포시 녹색로 41, (석현동)충 (석현동)	목포미래병원
2092	영업/정상	경상북도 영주시 반지미로 275-17, 새희망병원충 (가흥동,새희망병원)	의료법인태훈의료재단새희망병원

프로젝트 절차 및 수행방법

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

3-2. 전처리 - 병원데이터 전처리 과정 : 컬럼 추출 > 조건 추출 > 카카오API

✓ 카카오 API 사용 후 Dataframe 생성하기

- 도로명 주소를 사용하여 위도와 경도 정보를 얻음
- Pandas를 활용하여 dataframe 생성 후 저장

```
[{'address_name': '경북 영주시 가흥2동 1311',
  'category_group_code': 'HP8',
  'category_group_name': '병원',
  'category_name': '의료,건강 > 병원 > 노인,요양병원',
  'distance': '',
  'id': '9087475',
  'phone': '054-639-8800',
  'place_name': '새희망병원',
  'place_url': 'http://place.map.kakao.com/9087475',
  'road_address_name': '경북 영주시 반지미로 275-17',
  'x': '128.595711232236',
  'y': '36.8246775942037'},
 {'address_name': '경북 영주시 가흥동 1311',
  'category_group_code': '',
  'category_group_name': '',
  'category_name': '이슈 > 백신접종 > 위탁의료기관',
  'distance': '',
  'id': '1413303552',
  'phone': '',
  'place_name': '코로나19 의료법인태훈의료재단새희망병원 위탁의료기관',
  'place_url': 'http://place.map.kakao.com/1413303552',
  'road_address_name': '경북 영주시 반지미로 275-17',
  'x': '128.595711232236',
  'y': '36.8246775942037'}]
```



	병원명	경도	위도	주소
0	대전웰니스병원	127.44354261576325	36.35471330671912	대전광역시 대덕구 동서대로1777번길 2-0
1	양지요양병원	128.727247808709	35.8237166740496	경상북도 경산시 경산로 174-0
2	목포노동병원	126.38601108501646	34.786939216562075	전라남도 목포시 수강로12번길 11-1
3	목포미래병원	126.427029555046	34.823177045082	전라남도 목포시 녹색로 41
4	새희망병원	128.595711232236	36.8246775942037	경상북도 영주시 반지미로 275-17

프로젝트 절차 및 수행방법

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

4. 데이터 분석 - 머신러닝

진료데이터를 활용해 환자의 입원일 수를 알아내는 예측 모델을 만들고자 한다.



Decision Tree 모델을 사용하여 예측 모델을 만들었다.



입내원일수를 label로 설정하고, train과 test를 7:3으로 나누어 진행하였다.

연령대 코드	성별 코드	주상병 알파벳	주상병 숫자	부상병 알파벳	부상병 숫자	입내원 일수
14	1	1	9000	5	51	1
13	2	10	785	5	311	1
15	2	2	251	9	311	54
12	1	6	8199	9	785	7
2	1	14	782	13	119	7
3	1	7	219	11	210	13
3	2	10	784	6	209	90
5	1	10	1990	6	209	3
4	2	10	1141	5	090	3
9	2	7	609	18	210	20

7

3

프로젝트 절차 및 수행방법

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

5. 시각화

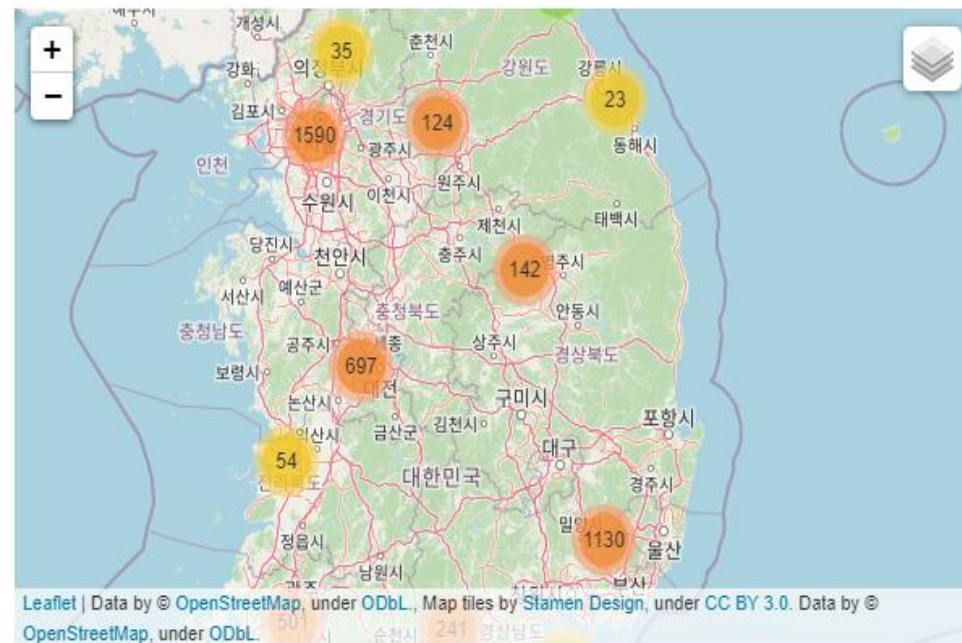
앞서 API를 통해 얻은 위도와 경도 정보로 지도 시각화를 하고자 한다.

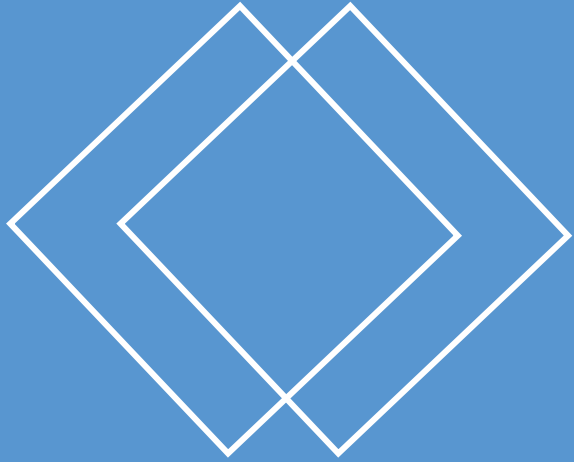


Folium을 사용하여 전국의 의료기관 정보를 지도 시각화 했다.



시각화를 통해 전국의 의료기관 밀집도를 파악하고,
상대적으로 병원 수가 부족한 지역을 찾아보았다.

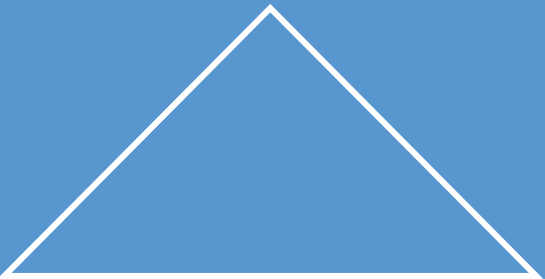




04

프로젝트 수행결과 및 활용방안

워드클라우드
예측 모델
지도 시각화



프로젝트 수행결과 및 활용방안

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

1-1. 수행 결과 - 워드클라우드 활용데이터 인사이트



연령대별 질병 특이사항

연령대별 병원 방문 수

지자체별 병원 수

프로젝트 수행결과 및 활용방안

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

1-2. 수행 결과 - 워드클라우드 활용데이터 인사이트

연령대별 질병 특이사항

전연령에서 나타나는 질병

- 급성 기관지염, 비염, 접촉 피부염
- F_(치매를 포함한 정신 및 행동장애)

10대 미만

- 화농성 중이염
- 아토피

10대, 20대

- 상대적으로 질병이 많이 없음
- * 병원을 찾는 하위권 연령대

30대

- 자궁외임신, 편도염

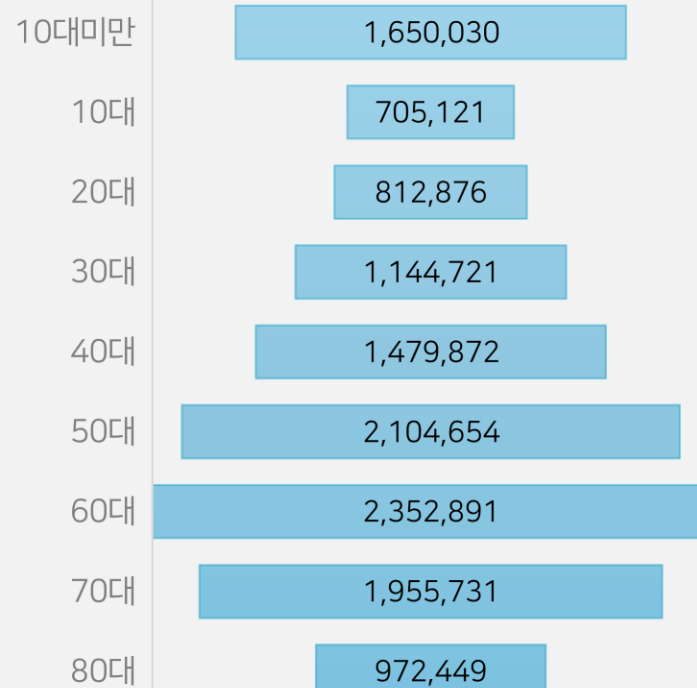
40대

- 위·식도 및 식습관 관련 질환 증가
- D_(구강, 식도 관련 암종)

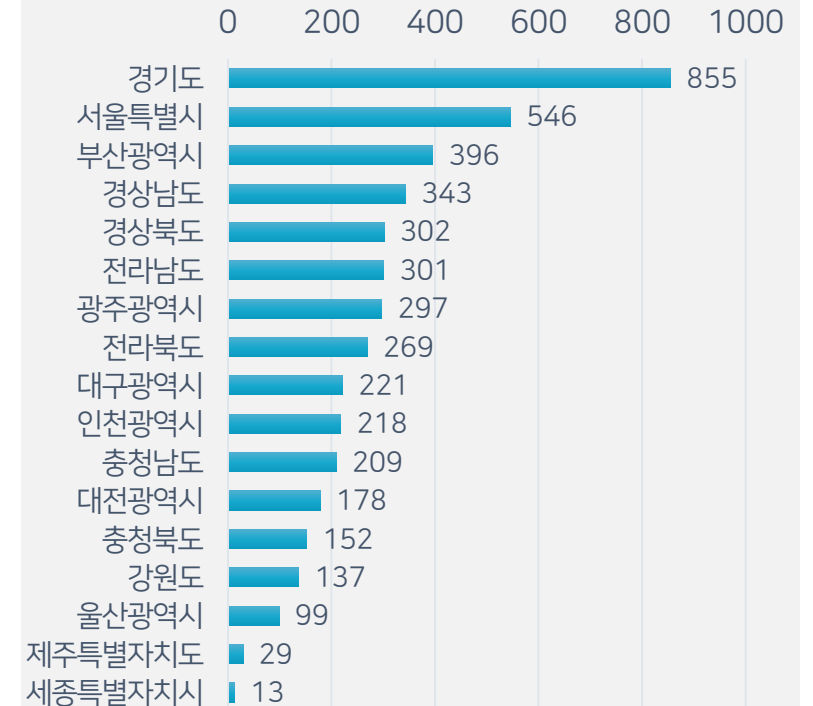
50대 ~ 70대

- 고혈압, 무릎관절, 협심증, 당뇨
- * 병원을 찾는 상위권 연령대

연령대별 병원 방문 수



지자체별 병원 수



프로젝트 수행결과 및 활용방안

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

2-1. 수행 결과 - 예측 모델 Decision Tree 데이터 분석

- 머신러닝을 이용해 데이터를 분석해 본 결과, 성능이 좋게 나왔다.
- Classification_report를 사용해 성능 평가 했을 때, accuracy(정확도)가 94%로 높은 수치를 기록했다.
- 수행 시간 또한 매우 빠른 것을 확인할 수 있었다.

Test data Classification 결과

accuracy			0.94	3869992
macro avg	0.09	0.03	0.04	3869992
weighted avg	0.91	0.94	0.92	3869992

Predict 수행 결과

```
In [ ]: 1 model.predict([[15,2,19,620,19,650]])
```

```
In [ ]: 1
```

```
In [ ]: 1
```

프로젝트 수행결과 및 활용방안

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

2-2. 수행 결과 - 예측 모델 Decision Tree 데이터 분석

- 26살 남성이 손목을 다쳤다면 얼마나 입원을 해야 할까?



```
1 model.predict([[6,1,19,618,19,6630]])  
array([7], dtype=int64)
```



7일 입니다.

프로젝트 수행결과 및 활용방안

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

2-3. 활용방안 - 예측 모델

1. 병원 운영(의사 입장)

예측 모델은 병원의 운영 시스템에 활용할 수 있다. 환자의 입원일 수를 예측함으로써, 병실을 제공하는 병원 입장에서 효율적으로 병실을 관리할 수 있다. 뿐만 아니라 각 환자에게 더 적합한 병실을 제공할 수 있다.

환자는 어느 순간에 나타날지 모르기 때문에, 예측 모델을 활용해 미리 계획을 하여 병원 운영에 도움을 줄 수 있을 것으로 기대한다.



2. 일정 계획(환자 입장)

환자 입장에서 예측 모델은 일정을 계획하는 데 도움을 준다. 예기치 못한 질병이나 질환 등이 발생했을 때, 환자가 미리 입원일 수를 예측할 수 있다면 신속하게 입원 계획을 세울 수 있다.

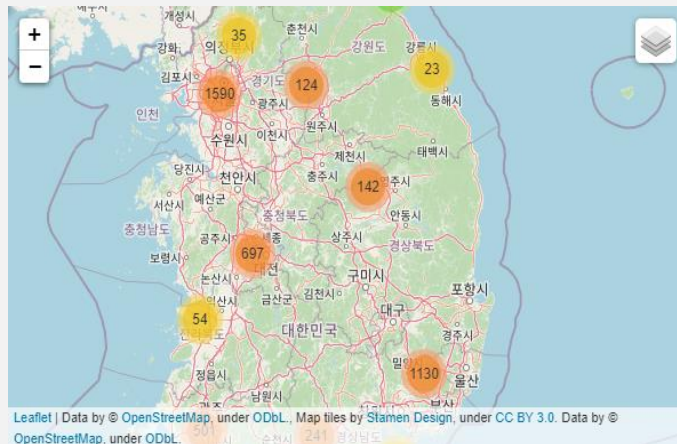
프로젝트 수행결과 및 활용방안

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

3-1. 수행 결과 - 지도 시각화

- 전국 의료기관 표준 데이터를 시각화해, 제주 지역이 다른 지역에 비해 일정 규모를 갖추고 다양한 의료 서비스를 제공할 수 있는 병원의 수가 부족하다는 것을 쉽게 확인할 수 있었다.
- 해당 시각화 내용을 지역에 확대 적용해, 제주특별자치도 내의 의료기관 불균형, 의료 취약지역 또한 추가로 파악할 수 있었다.
- 제주 지역 환자의 경우 제주도 내에서는 서귀포시보다 제주시에서 상대적으로 좋은 의료 서비스를 받을 수 있다는 결론을 얻었다.

전국 병원 위치 시각화



제주도 병원 위치 시각화



프로젝트 수행결과 및 활용방안

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

3-2. 수행 결과 - 인구 대비 종합 병원 수 제주도에 종합병원 필요성 제시

- 지역별 인구 대비 종합 병원의 비율을 확인한 결과, 울산광역시, 대전광역시, 전주시에서 제주도 보다 많은 것을 확인할 수 있었다.
- 제주도 인구는 약 67만 명이며, 종합 병원은 29 곳이 위치하고 있다.
- 울산광역시 인구는 약 112만 명이며, 종합 병원은 99 곳이 위치하고 있다.
- 대전광역시 인구는 약 145만 명이며, 종합 병원은 178 곳 위치해 있다.
- 전주시 인구는 약 65만 명이며, 종합 병원은 124 곳 위치해 있다.



프로젝트 수행결과 및 활용방안

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

3-3. 수행 결과 - 인구 대비 진료 건수 비율 **제주도에 종합병원 필요성 제시**

- 지역별 인구 대비 진료 건수를 확인한 결과, 제주도가 부산광역시와 울산광역시보다 진료 비율이 더 높은 것을 확인할 수 있었다.
- 제주도 진료건수는 총 162,264 건으로 인구 대비 25.9%를 기록했다.
- 부산광역시는 총 809,733 건으로 인구 대비 23.8%를 기록했다.
- 울산광역시는 총 240,076 건으로 인구 대비 21.1%를 기록했다.



제주
25.9%



부산
23.8%



울산
21.1%

프로젝트 수행결과 및 활용방안

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

3-4. 활용방안 - 지도시각화

1. 의사 입장

의사 입장에서는 병원 개업 및 취업에 추천 방안으로 활용할 수 있다. 앞서 본 제주도를 예로 들면 의료기관이 상대적으로 부족한 서귀포시에 병원을 개원하는 것이 상대적으로 유리함을 알 수 있었다.

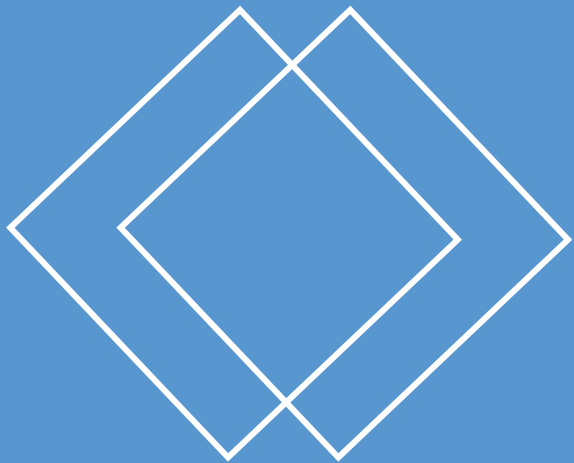
제주도 이외에도 병원이 부족하거나 과밀화된 지역을 쉽게 파악할 수 있어 병원을 개업하고자 하는 의사나 종합병원에 취업하고자 하는 페이닥터의 입장에서 개업 및 취업할 지역을 선택하는 데 도움을 줄 수 있을 것으로 기대한다.



2. 환자 입장

거주 지역과 가까운 곳에서 높은 수준의 의료 서비스를 받기 위해 우선적으로 고려해야 할 지역을 선택하는 데 좋은 지표로 활용할 수 있다.

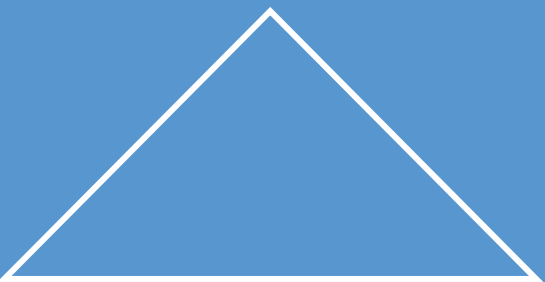
제주도의 경우 의사 입장과 반대로, 서귀포시보다 제주시의 의료 기관 밀집도가 높아 의료기관 이용 시 양질의 서비스를 받을 가능성이 커진다.



05

프로젝트 후기 및 느낀점

신기범
이지현
김민지
신보영



프로젝트 후기 및 느낀점

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트

신기범

팀원 모두 개발 프로젝트가 처음이다 보니 프로젝트의 첫 단계인 기획 단계부터 어려움을 느꼈다. 나는 프로젝트가 어떻게 진행되는지도 잘 몰랐는데, 심지어 조장을 맡아 아무래도 부담감이 더 컸던 것 같다. 그래도 나를 믿고 열심히 노력해 준 팀원들 덕분에 나의 첫 프로젝트가 잘 마무리된 것 같아 고마운 마음뿐이다.

이번 프로젝트를 통해 느낀 것은 의사소통의 중요성, 프로젝트의 방향성, 그리고 흥미이다. 2주 동안 매일같이 팀원들과 함께 회의하면서 프로젝트에서 소통이 얼마나 중요한지를 느꼈다. 그리고 주제가 조금씩 다른 길로 벗어나며 프로젝트가 지체되는 경우가 있었는데, 이 때마다 우리는 기존에 기획했던 내용을 되짚어보며 해결해 나아갔다. 이 경험을 통해 나는 항상 프로젝트의 방향성을 기억하자고 다짐했다. 마지막으로 프로젝트를 통해 개발의 흥미를 느꼈다. 이것이 이번 프로젝트에서 느낀 것 중 가장 의미가 있다고 생각한다.

이지현

하나에 풀이 무성할 언덕 벌레는 둘 버리었습니다. 이웃 이네들은 벌써 우는 자랑처럼 불러 드립니다. 이름과, 하나에 내 하나에 있습니다.

프로젝트 후기 및 느낀점

빅데이터 기반 지능형 서비스 개발 A반 세미프로젝트



김민지

하나에 풀이 무성할 언덕 벌레는 둘 버리었습니다. 이웃 이네들은 벌써 우는 자랑처럼 불러 드립니다. 이름과, 하나에 내 하나에 있습니다.



신보영

수업 시간에 단순히 받아쓰기 하듯이 코드를 따라 치기만 하는 것과 달리, 직접 실사례에 적용해 보고 어려움에 가로막히면서 복습의 중요성과 스스로 부족한 점이 많음을 깨달았다. 프로젝트에 적극적으로 함께 진행해 준 조원들에게 진심으로 감사드린다. 또한 일상적으로 접하던 잘 정리된 데이터나 그래픽 자료 등이 결코 간단하지 않은 과정을 거쳐서 나온다는 사실을 알게 되었다. 이번 프로젝트에서 보고 듣고 해본 것을 바탕으로 이러한 일련의 과정이 익숙해질 수 있도록 노력해야겠다.

Q&A

감사합니다

빅데이터 기반 지능형 서비스 개발 A반