



텍스트 마이닝, 인공지능 분석 기술을 이용한 경상북도 의료접근성 분석

Jun 10, team project proposal presentation, IMEN-321 Data Mining

2017250005 전재현, 2018250042 정현재

Korea University, Department of Biomedical Engineering

Korea University, Department of Artificial Intelligence

INTRODUCTION

주제 선정 및 배경



- 보건의료 기술을 필요로 하는 이들이 그 혜택을 누릴수 있는지 확인하는 지표
- 삶의 질을 평가하는 가장 직접적인 평가 요소
- 의료접근성의 평등은 모두가 공평히 누려야하는 권리
- 원격의료와 같은 시스템 도입을 가장 경제적이고 효율적으로 진행하기 위해 의료접근성에 대한 조사는 필수적

“의료 시설 또는 의료 행위에 접근할 수 있는 가능성”

METHODS & EXPERIMENTS

분석 – Text Mining

텍스트 데이터를 활용한 의료 접근성 중요 Feature 추출 및 데이터 수집

1. 빈번하게 등장하는 병원 및 지역을 통해 기사 텍스트 데이터를 통해 지역별 의료접근성을 분석
 2. 수업시간에 이론적으로 학습한 토픽 모델링 방법인 **LDA 방식**을 적용하여 최근 의료접근성에 대한 논문들이 어떠한 요소들을 위주로 논문이 작성되고 있었는지를 확인하여 의료접근성에 중요한 요소들을 확인
- K값을 4로 지정하여 가장 많이 언급되고 LDA로 얻어낸 새로운 키워드를 지정하여 이를 **의료접근성을 위해 분석해야할 중요한 요소로 선정**

1. ['농촌간', '의료수요', '의료시설', '탄력성', '선택형태', '의사', '의료', '자원', '보건소', '생산성', '정부', '선택', '보건지소', '양적', '비용', '증가시킬', '약품', '투자']
2. ['지역', '환자', '의료서비스', '서비스', '임금', '연구', '도움', '확보', '의료자원', '공공의료', '기회비용', '국민건강조사', '다양', '보건진료소', '보건의료원', '소득', '저가', '기회', '교육수준', '선택행태']
3. ['분포', '보건', '상반', '소비자', '포함', '기여', '공공', '한계대체', '국민', '공공의료기관', '정책', '기초', '의료원', '불균형', '이용', '수요', '중증도', '의사인력', '교육']
4. ['도시', '시설', '투자모형', '분석', '인력', '장비등', '의료접근', '구매', '농촌지역', '수입하에', '시간', '시간비용',



- 1.의사 수 및 지리적 데이터
2. 성별 장애인 고령인구 등의 지역별 인구 특성
- 3.행정구역 공공의료기관 등의 분포 데이터
4. 도시 및 시설의 인프라

METHODS & EXPERIMENTS

활용 데이터



좌표 데이터

- 경상북도와 대구광역시 내 병원, 보건의료기관, 응급의료기관, 약국의 소재지 및 좌표 데이터
- 국토교통부 국가공간정보포털 오픈마켓 (제공 - (주) bisGIS)
 - 공공데이터포털
 - 국립중앙의료원
 - 응급의료기관 좌표 데이터 : 소재지 데이터 기반으로 직접 변환 (WGS84 좌표계 기반)



통계 데이터

- 전국 시도별 및 경상북도 시군구별 인구, 고령인구, 장애인구, 활동제약인구, 의료기관, 약국 수, 병상수, 의료인력, 인구 천명당 의료기관 종사 의사수, 면적, 도시면적 비율, 백신 예방접종률
- KOSIS
 - 보건의료빅데이터개방시스템
 - 읍면동 단위 행정구역별 면적 : 각 행정구역별 홈페이지



Infrastructure 이미지

- 경상북도 및 대구의 VIIRS Day-Night Band Satellite Image
- The Earth Observations Group (EOG) at NOAA National Centers for Environmental Information, day night band nighttime 사진

해당 토픽들을 기준으로 데이터를 다시한번 수집하여 구체적으로 분석

METHODS & EXPERIMENTS

분석 - Geographic data

지역별 기초 정보 시각화 - 좌표 데이터

JSON, Shape 파일과 Python 기반의 Feature Engineering을 통해 경상북도의 지도를 표시

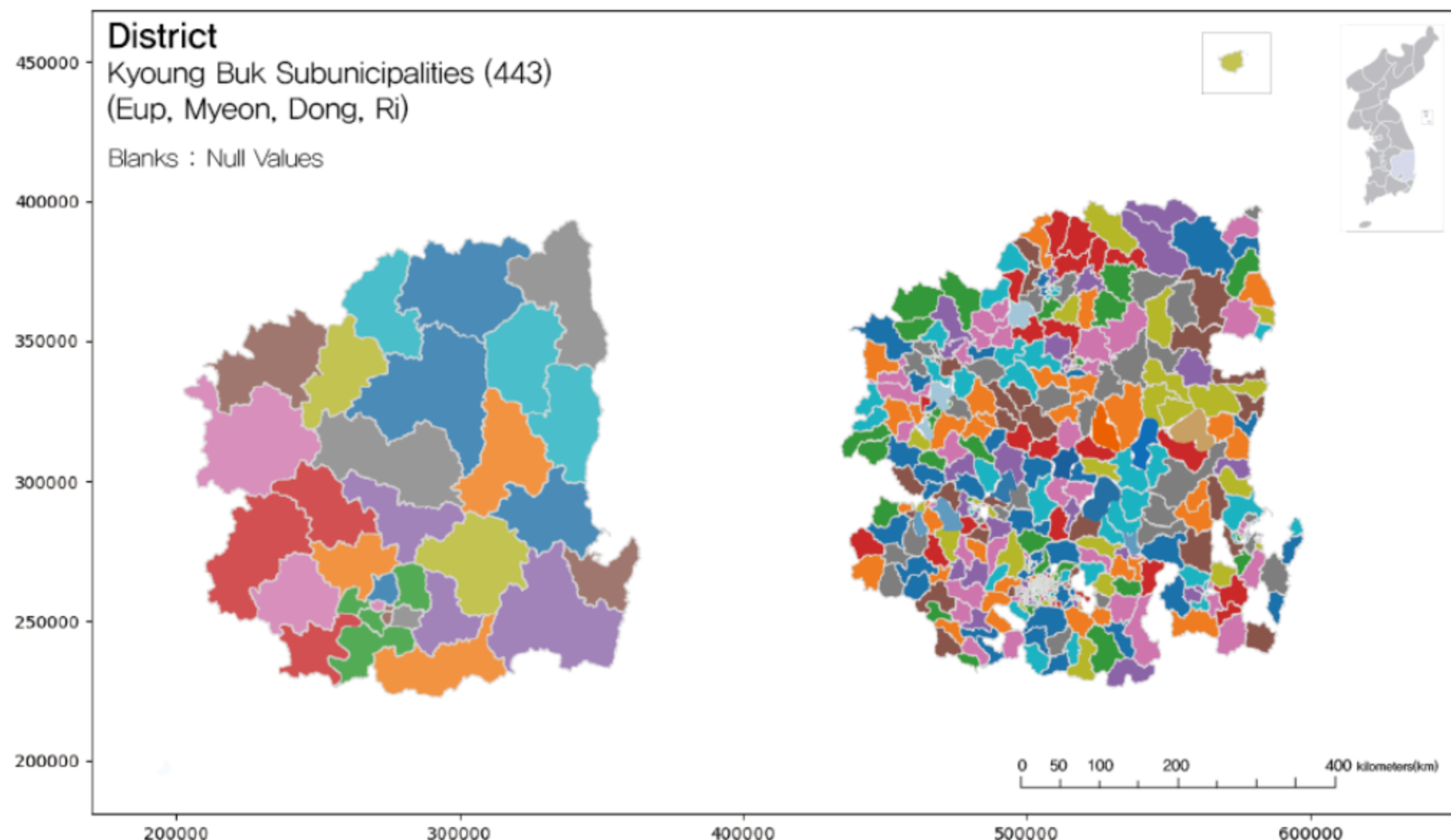
전처리 과정



전처리 이후 좌표 데이터

순서	지역명	코드	Geometry (위도, 경도 다중 점)
1	경주시	11010	MULTIPOLYGON (126.97468 37.629 ...)
2	김천시	11020	MULTIPOLYGON (126.02339 37.571 ...)
3	안동시	11030	MULTIPOLYGON (127.04324 37.558 ...)
4	구미시	11040	MULTIPOLYGON (127.10385 37.572 ...)
5	영주시	11050	MULTIPOLYGON (126.97059 37.550 ...)
6	영천시	11060	MULTIPOLYGON (127.02475 37.222 ...)
7	상주시	11070	MULTIPOLYGON (127.88475 37.332 ...)
...			
30	울진군	11300	MULTIPOLYGON (127.960 36.882 ...)
31	울릉군	11310	MULTIPOLYGON (127.698 37.012 ...)
32	포항시 남구	11320	MULTIPOLYGON (127.552 37.466 ...)
33	포항시 북구	11330	MULTIPOLYGON (127.501 37.277 ...)

Kyoung Buk Municipalities & Districts



METHODS & EXPERIMENTS

분석 - Geographic data

경상북도 내에 존재하는 모든 관내 의료시설

인구와 인구밀도 수치를 통해 실제로 의료 시스템을 활용하는 단편적인 수요와 공급을 정량적으로 분석
좌표 (WGS 80) 통일, 무의미한 Column 제거 등의 공통적인 전처리를 진행함



Fig1. Pharmacy



Fig2. Medical Center



Fig3. Emergency Room



〈선정 기준〉

- Figure 1 - 대학 약국 및 지역 약국
- Figure 2 - 보건소, 사립 병원, 대학 병원
- Figure 3 - 사망률에 직접적으로 연관된 관내 응급실

METHODS & EXPERIMENTS

분석 – Geographic data

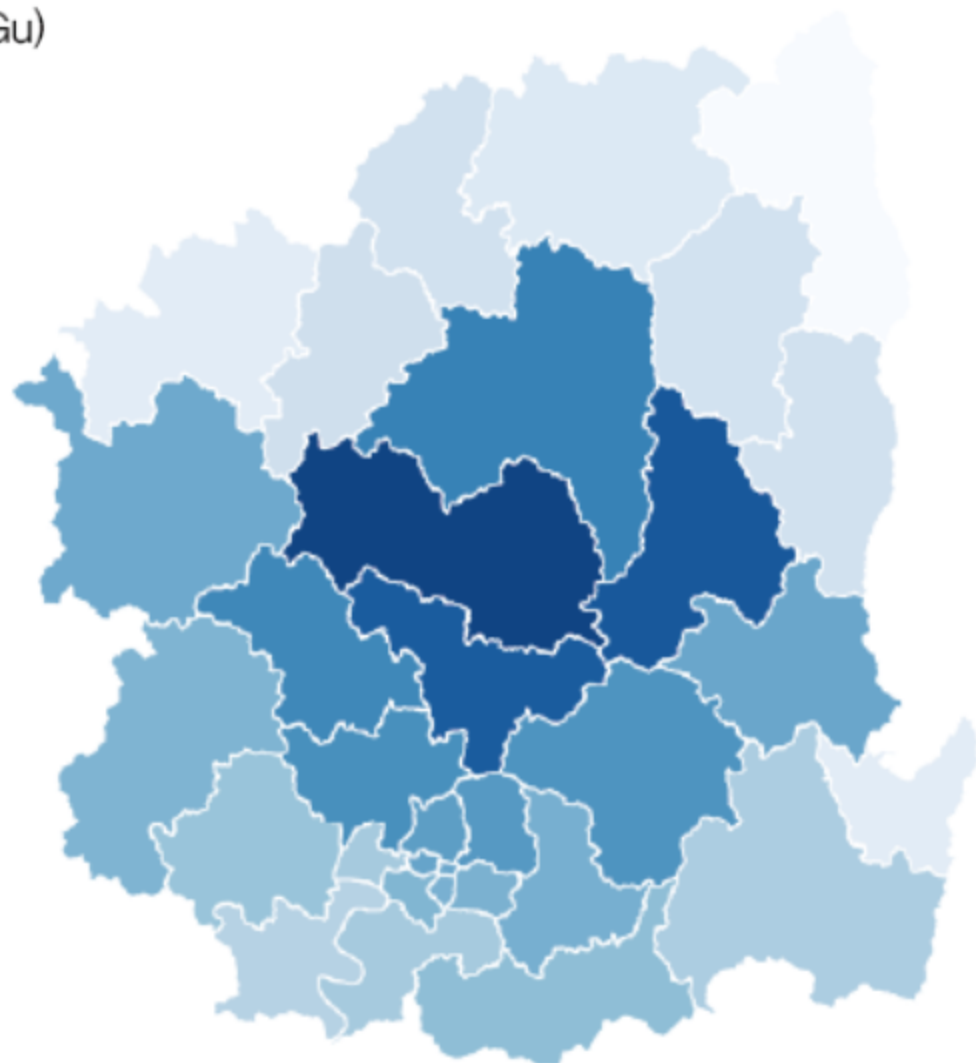
Two-Step Floating Catchment Area

의료접근성 수치 분석 **2SFCA 모델 자체 개발**, 오픈 소스 (GitHub)에 **개발 코드 무료 배포**

수요 및 공급 무게중심 거리 Matrix -> 지역별 인구 요소 곱 연산 -> 2SFCA 중력 모델 평가 -> 의료 접근성 결과 출력

District

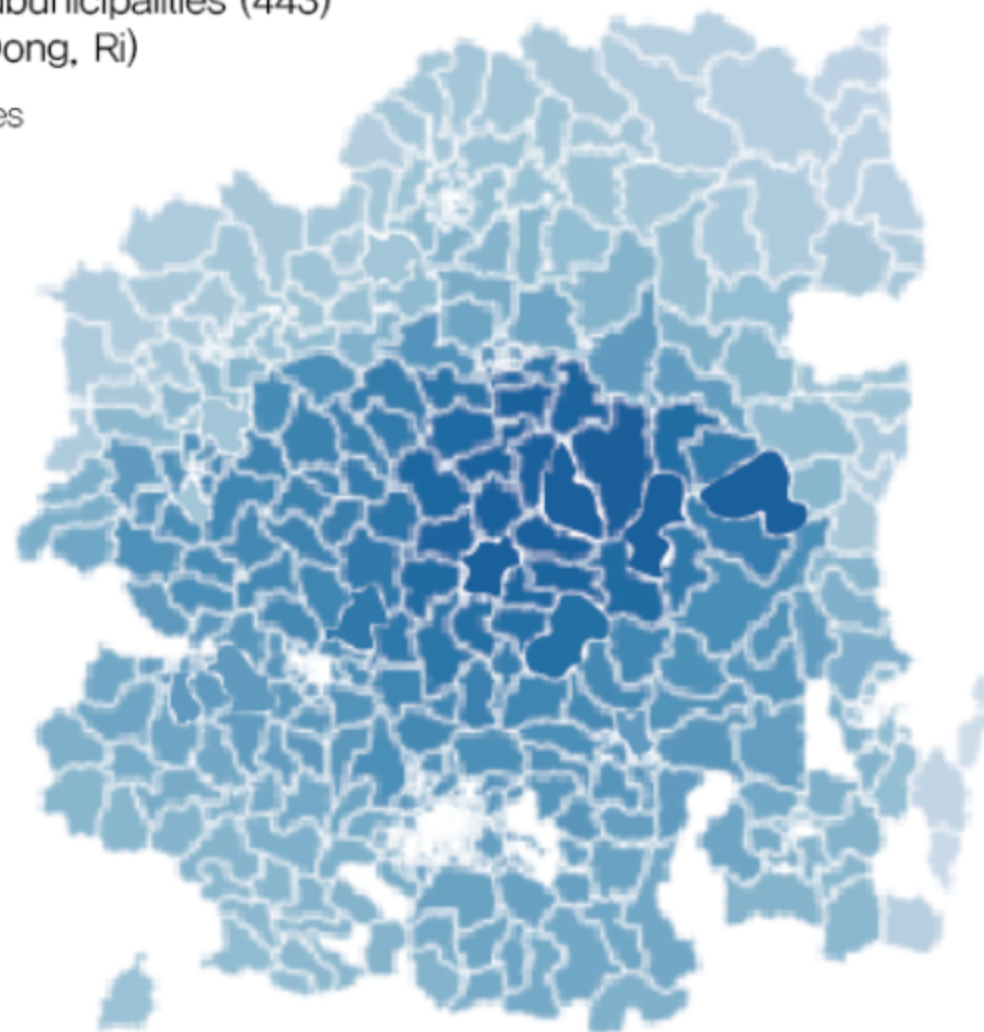
Kyoung Buk Municipalities (23)
(Si, Gun, Gu)



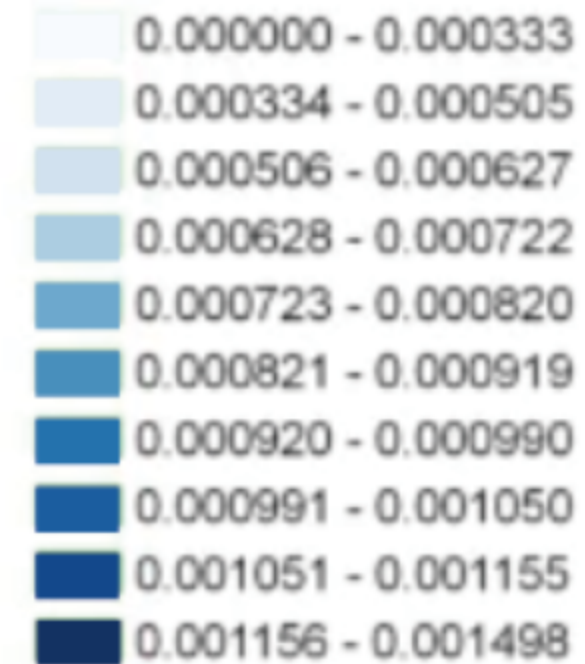
District

Kyoung Buk Subunicipalities (443)
(Eup, Myeon, Dong, Ri)

Blanks : Null Values



Accessibility (Ai score)



Increasing
spatial accessibility
↓

0 50 100 200 400 kilometers(kr)



METHODS & EXPERIMENTS

분석 - Geographic data

The VIIR day-night band(DNB) image

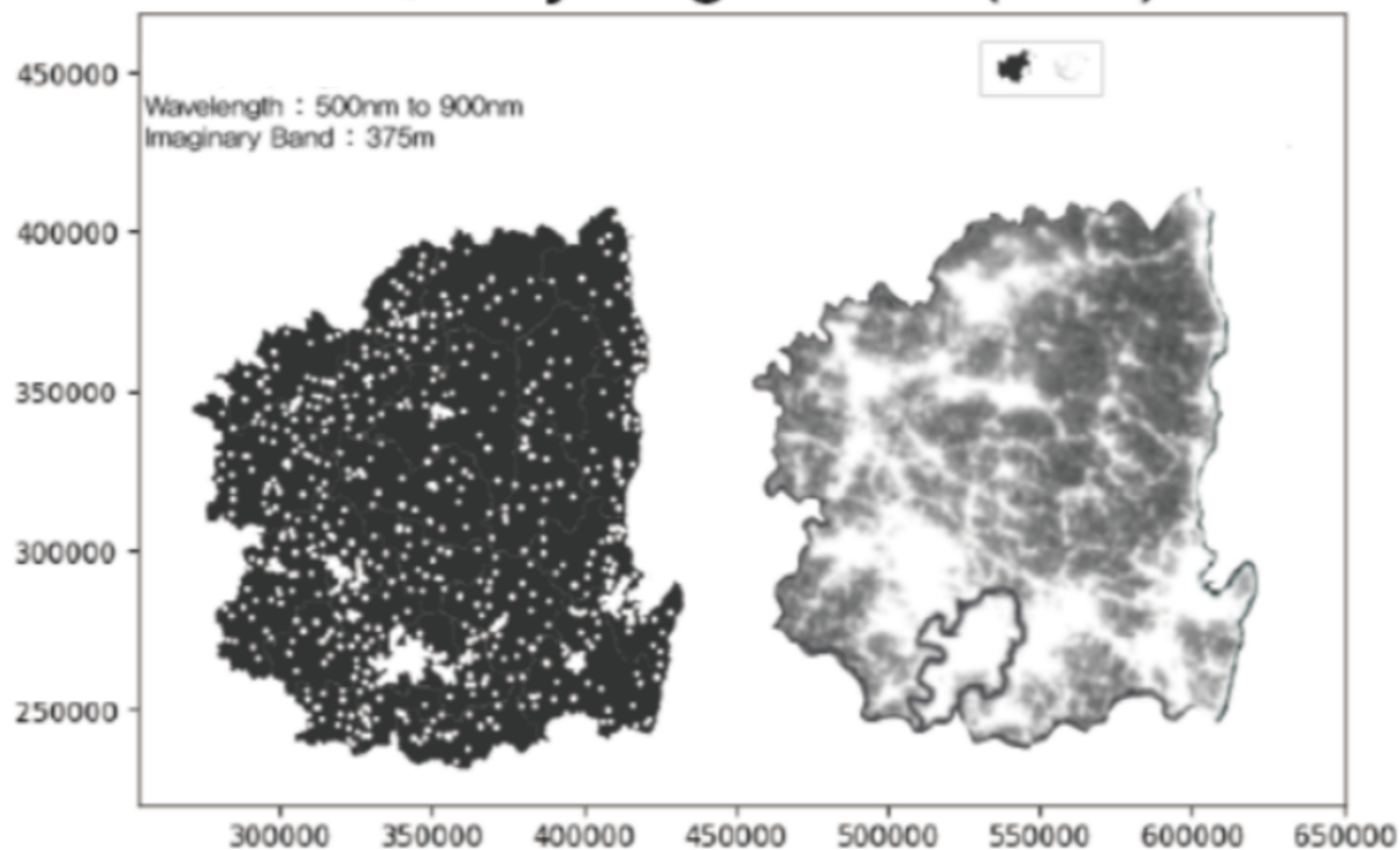
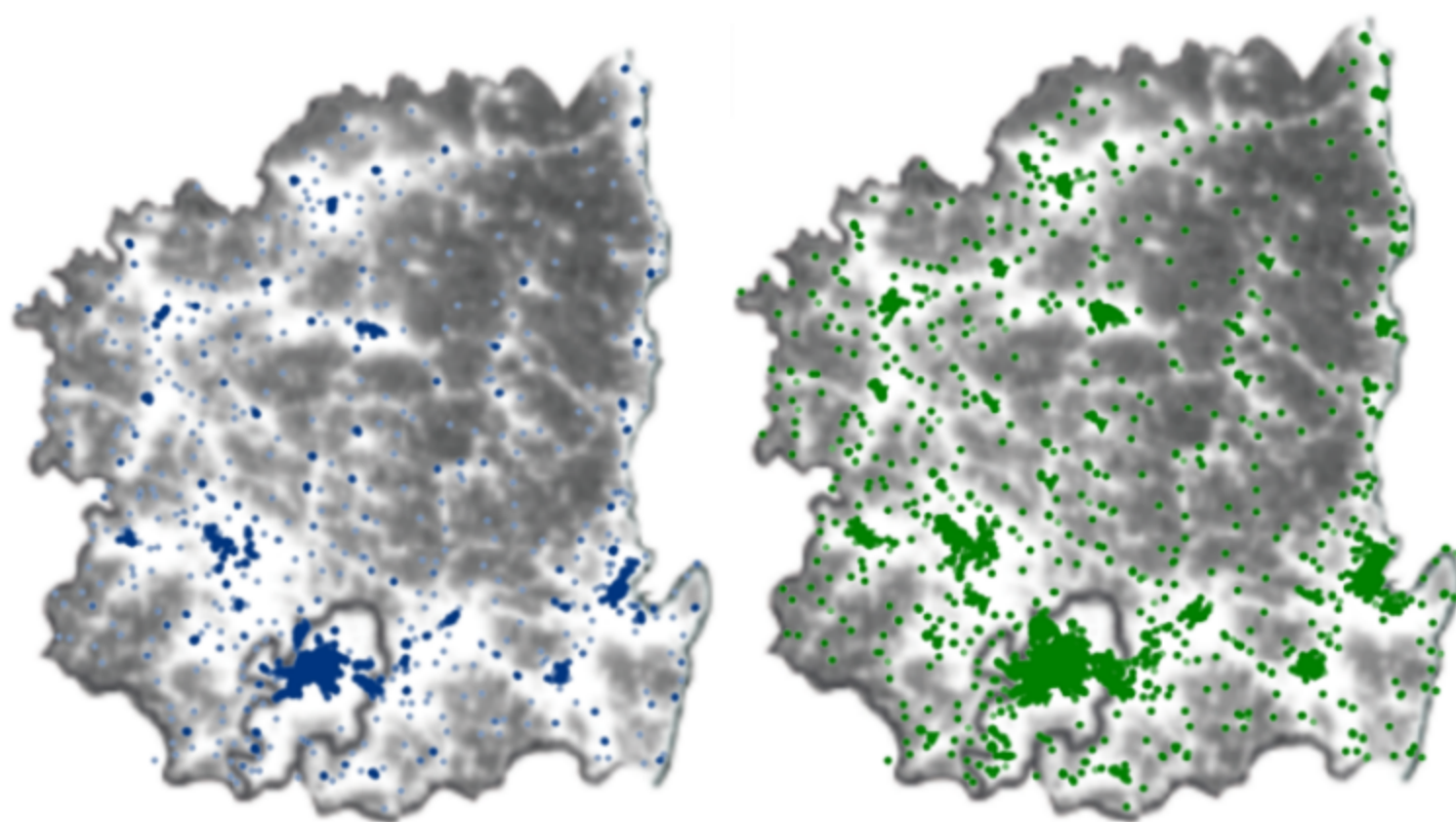
인공광원 복사휘도 정보를 추출하여 야간 에어로졸 광학두께를 추정하는 방법

실제 도입이 가능한 시설 기반 및 도시 개발률, 건물 설비, 실제 인구 밀집 지역까지 분석 가능.



NOAA National Centers for Environmental Information

VIIRS Day-Night Band(DNB)



METHODS & EXPERIMENTS

분석 – Tabular data

머신러닝 기반 분석

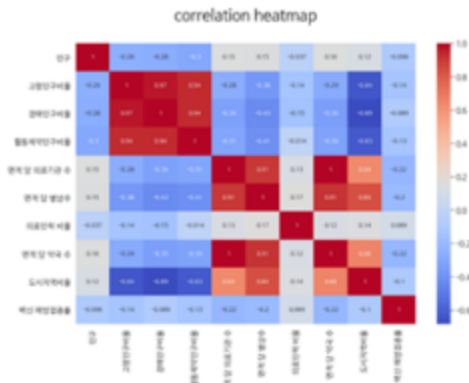
일부 feature값을 비율로 변경

-> 인구, 고령인구비율, 장애인구비율, 활동제약인구비율, 면적 당 의료기관 수, 면적 당 병상 수, 의료인력 비율, 면적 당 약국 수, 도시지역비율, 백신 예방접종률

Healthcare manpower를 8개 class로 구분

Upsampling을 통해 각 class의 데이터 500개로 증가

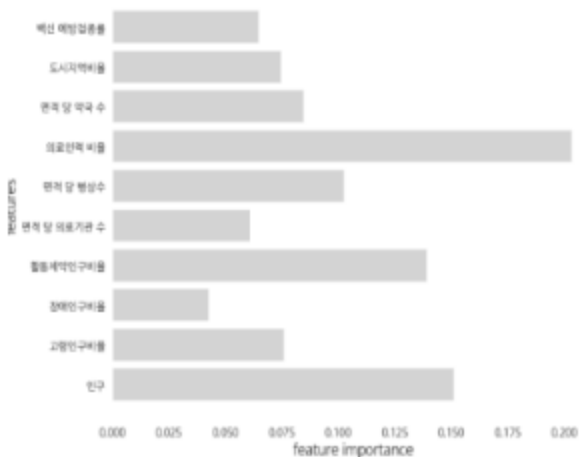
correlation heatmap을 확인하여 feature들 간의 상관 관계를 파악



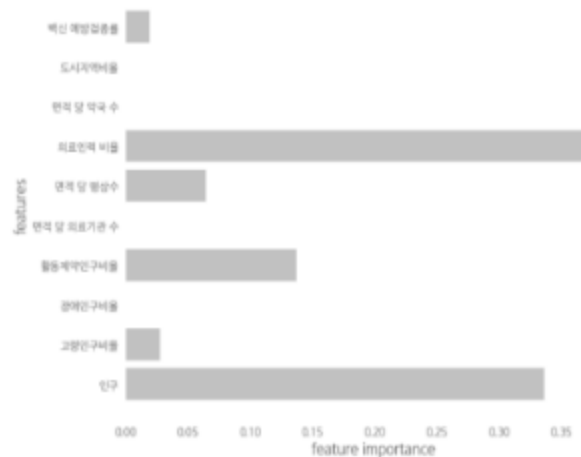
Feature Engineering

Correlation Heatmap

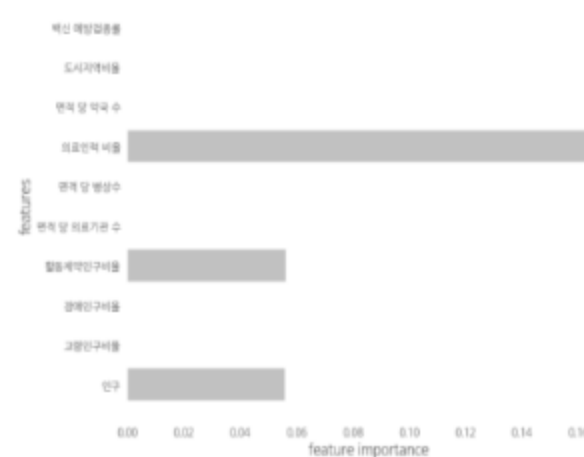
random forest MDI feature importance



Xgboost permutation feature importance



random forest permutation feature importance



차원 축소 기반 분석

5차원

의료 접근성과 관련된 중요한 feature

인구

활동제약인구비율

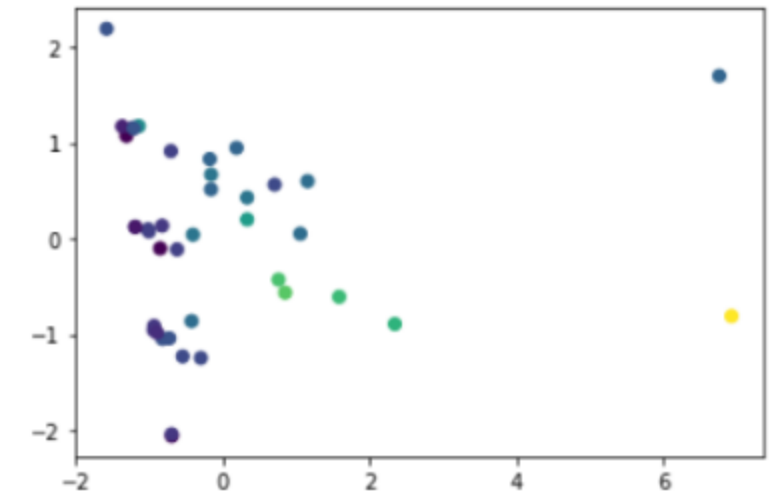
면적 당 의료기관 수

의료인력 비율

백신 예방접종률

PCA

2차원



각 점 -> 전국 시/도(17개)와 경상북도의 시/군/구(23개)를 의미

CONCLUSION

분석 결론

- 경상북도 전역의 의료 시설을 Clustering 해본 결과, 광역시 및 대도시 부근에서 많은 서비스가 있는 것을 확인 할 수 있었고, **동북부 지역의 의료시설이 상대적으로 적은 것**을 확인할 수 있었다.
- **응급실이 존재하지 않는 시,도의 존재를 분석**하였고, 응급실 도입성에 있어서 이러한 특정 지역(포항시 북부,봉화군 동부) 등에서 우선도입의 필요성을 느꼈다.
- 상위 미충원 비율 지역에 포함되는 지역 중에서도, **높은 인구비율을 차지하는 지역을 확인**하였고, 이러한 지역의 **우선적인 도입**이 필요해보였다.
- 인구, 수요와 공급, 면적등의 요소를 모두 고려하는 의료접근성에서는 대도시 및 광역시부근의 의료접근성이 그렇게 크지 않았다. **대도시에 의료접근성이 꼭 우수한것은 아니**라는 것을 분석을 통해 결론지을 수 있었다.
- 많은 인프라에도 불구하고 의료 시설이 도입되지 않은 곳도 확인할 수 있어, 우선 도입 지역을 확인해볼 수 있었다.



“

객관적인 분석을 통해, 경상북도 의료 불균형 및 지역별 특성을 효율적으로 파악

”