

수도권 쓰레기 매립지 입지선정

서희나, 윤경선, 이지윤



목차

01 연구 배경

연구 배경 및 목적
연구의 범위 및 방법
선행연구
연구의 필요성

02 분석 방법

전처리
분석 과정

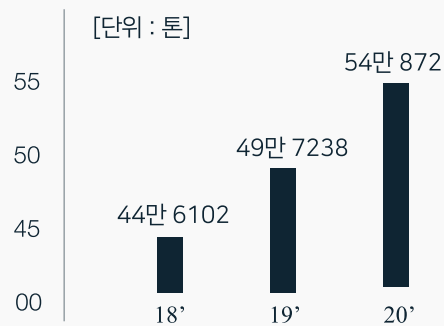
03 분석 활용전략

최종입지선정
기대효과

연구 배경 및 목적

쓰레기 매립지에 관한 연구는 얼마나 진행되고 있는가?

쓰레기 매립 양



코로나 19로 인해 쓰레기 배출량이 증가
쓰레기 양은 늘어나고 있지만 쓰레기를
매립할 장소는 없어지는 상황

수도권 매립지 사용 종료



인천시 2025년 수도권 매립지 사용 종료 선언

쓰레기 매립지 입지 선정에 관해 수도권 지
역은 팽팽한 신경전을 겨루고 있다

공모 내용 정리

구분	내용
토지	1) 부지면적 220만 m^2 이상 (실제 매립면적 170만 m^2 이상) 2) 인구 밀집지역 및 공공기관, 학교, 연구시설, 의료시설, 종교 시설 등과 근접하면 안됨 3) 풍향을 고려, 배수 고려, 시민의 보건위생에 위해를 끼칠 우려가 없는 지역 4) 용수, 동력의 확보 용이, 차량 접근 쉬운 지역
주민	1) 시설 설치비의 20% 이내에서 주민편익시설을 설치하고, 폐기물 반입수수료의 20% 이내에서 주민지원기금을 제공 2) 시설 설치비의 20% 이내에서 주민편익시설을 설치하고, 폐기물 반입수수료의 20% 이내에서 주민지원기금을 제공
기타	1) 매립장 외에 생활폐기물 예비처리시설과 건설 폐기물 분리, 선별 시설 건설 2) 대기 및 수질오염 등 각종 환경오염문제를 고려하여 결정, 은폐 시설 필요

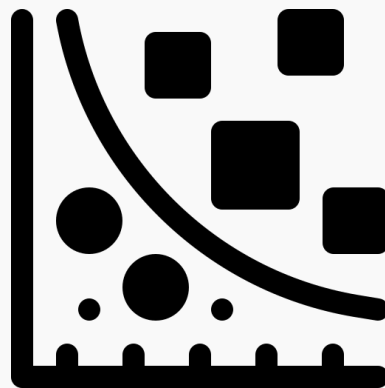
연구의 필요성

쓰레기 매립지 사용 종료에 대비해야 하는 것과 더불어
본 연구에 다양한 필요성이 존재한다.



선행연구 미비

선행연구의 경우 단순한 공간분석으로
인문 사회적인 요소들을 거의 반영하지
못함



통계적 방법 활용

쓰레기 매립지가 혐오 시설인 것을 고려
하여 통계적 방법론을 사용하는 것이 합
리적이라 판단



쓰레기 매립지에 대한 시민의식 부족

쓰레기 매립지를 혐오시설로서만 취급,
경제적인 가치에 대한 사회적 인식 부족

주제

쓰레기 매립지를 위치할 수 있는 입지 선정

쓰레기 매립지란?

쓰레기 매립지를 위치할 수 있는 입지 선정



매립지는 혐오성을 띄는 혐오시설이다.

혐오시설: 국가적으로 지역적으로 매우 긴요하고 필수적인 시설이긴 하지만 그것이 그 지역에는 부정적인 효과로 인하여 지역주민들이 기피하는 공공시설

님비 현상 (NIMBY, Not In My Back Yard): 원하지 않는 시설의 입지에 대한 주민이나 지역에서의 반대

즉, 사회에는 필요하지만 자신의 지역에는 원하지 않는 시설

선행연구

선행연구

국내에는 쓰레기 매립지 관련 연구가 많이 이루어지지 않아
공공시설의 입지 선정에 관한 선행 연구도 함께 참고함

환용, 정일훈, 김철중(2010)

파주시의 사례를 중심으로 도시
공공시설에 대한 우선순위 분석

김병철, 오상영, 류근호 (2006)

DBSCAN-1을 이용하여
영향력을 고려한 적정 입지 선정
모델 연구

배민기, 장병문 (1998)

입지 인자들의 상대적 중요도를
고려하여 경상북도 경산시를
대상으로 폐기물 매립지 입지 선정

오재식, 최준호 (2009)

폐기물처리시설 설치로 인한
입지 갈등에 관한 이론적 고찰과
갈등 해소 방안 탐색

본 연구의 차별성

분석 결과

1. 객관성과 합리성을 높이기 위해 다양한 분야 고려
2. 주민들의 관심과 환경적인 요소와 밀접하게 관련 있음을 확인
3. 공간 분석 및 우선순위 분석 활용
4. 환경 관련 데이터 사용



1. 환경 대신 사회적 항목에 초점
2. 군집화, 이상치 탐지와 같은 다양한 통계적 분석 시도
3. 경제적 기대효과 분석

선행연구

선행연구

국내에는 쓰레기 매립지 관련 연구가 많이 이루어지지 않아
공공시설의 입지 선정에 관한 선행 연구도 함께 참고함

한용, 정일훈, 김철중(2010)

파주시의 사례를 중심으로 도시
공공시설에 대한 우선순위 분석

김병철, 오상영, 류근호 (2006)

DBSCAN-1을 이용하여
영향력을 고려한 적정 입지 선정
모델 연구

배민기, 장병문 (1998)

입지 인자들의 상대적 중요도를
고려하여 경상북도 경산시를
대상으로 폐기물 매립지 입지 선정

오재식, 최준호 (2009)

폐기물처리시설 설치로 인한
입지 갈등에 관한 이론적 고찰과
갈등 해소 방안 탐색

본 연구의 차별성

분석 결과

1. 객관성과 합리성을 높이기 위해 다양한 분야 고려
2. 주민들의 관심과 환경적인 요소와 밀접하게 관련 있음을 확인
3. 공간 분석 및 우선순위 분석 활용
4. 환경 관련 데이터 사용



1. 환경 대신 사회적 항목에 초점
2. 군집화, 이상치 탐지와 같은 다양한 통계적 분석 시도
3. 경제적 기대효과 분석

연구의 범위 및 방법

쓰레기 매립지에 관한 연구는 얼마나 진행되고 있는가?

사용 데이터

쓰레기 매립지 후보 지역 선정

기관명	데이터명	이용과정
환경공간정보서비스	토지 피복도(2020)	토지 관련 공간 분석
국가공간정보포털	토지소유정보(2019)	

인구 관련 데이터

기관명	데이터명	이용과정
통계청	행정동단위주민등록인구 (2022)	밀집도 계산
서울열린데이터광장	서울시생활인구(2022)	
카카오API	병원, 학교, 학원, 문화시설, 관광명소, 교회 개수 (2022)	생활 인구 예측

사회/ 경제/ 환경 관련 데이터

기관명	데이터명	이용과정
국토 교통부 실거래 가 공개시스템	토지 매매가(2022)	경제 관련 지표
국민 재난 안전 포털 자연 재난 상황 통계	10년치 강수 피해액 (2020-2022)	쓰레기 매립지 조건상 비 피해가 많은 지역은 기피해야 함 (환경관련 지표)
한국 전력공사 공공데이터	전력소모량(2020)	쓰레기 매립지 조성 시 전력생산이 가능

* 2022년 기준 데이터 / 활용 가능한 최신 데이터
단, 데이터가 없는 경우 가장 최근 데이터 사용

연구의 범위 및 방법

본 연구는 다음과 같은 흐름으로 진행되었다

분석 방법

1



첫 번째 순서

쓰레기 매립지의
입지 선정에 위한 기준 정립

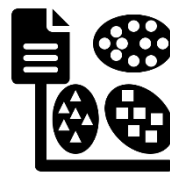
2



두 번째 순서

데이터 수집 및 전처리

3



세 번째 순서

클러스터링을 통한
후보 지역 선정 및 추출

4



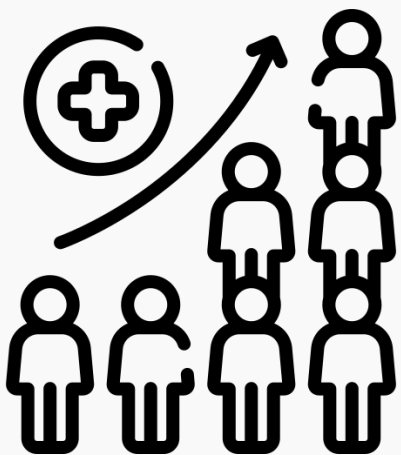
네 번째 순서

최종 입지 선정
및
결과 분석

전처리

생활인구

특정 시점에 개인이 위치한 지역을 집계한 '현주인구'



시간대에 따라 변화하는 인구의 규모로
지역간 특성을 추측해 가능
→ 유동인구 반영 가능

데이터가 서울만 존재

카카오 API



카카오맵 기준 1) 반경내 2) 지역내
시설물 위치 및 개수 제공



시설물이 많을수록 생활 인구가 늘어날 것이라고 판단
∴ 생활 인구를 예측하는데 사용

추출 방법

1. 각 지역 내 읍,면,동 별 중심을 구한다
2. 해당 읍면동의 반지름에 해당하는 거리에 있는 카카오 api가 제공하는 모든 생활 시설 추출 ('병원', '학교', '학원', '문화시설', '관광명소', '대형마트', '편의점', '지하철', '음식점', '카페')



가장 단순한 모델부터 시작해서
생활 인구 예측 시작

전처리

회귀 모델을 이용한 생활인구 예측

- 경기·인천 지역의 생활인구 데이터 부재 → 회귀 모델링을 통해 예측을 진행
- 반응변수(Y) : 수도권 지역 내 읍·면·동 별 생활 인구
- 설명변수(X) : 카카오 API를 이용해 크롤링한 생활시설 데이터, '2022년 06월 총 인구수', '토지 매매가'

예측에 이용한 회귀 모델

다양한 모델을 통해 예측을 진행함



과적합 방지를 위해 단순한 모델 → 복잡한 모델 적합하는 방식으로 예측 진행

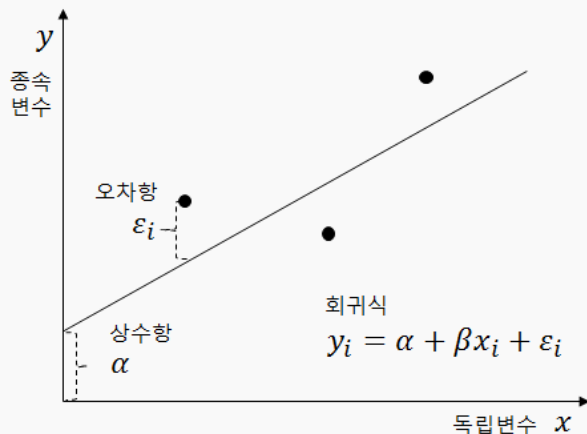
전처리

다중 선형 회귀를 이용한 생활 인구 예측

다중선형 회귀

설명 변수들의 선형 결합 이용해 반응 변수를 예측
→ 최소제곱법을 통해 모수 추정

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

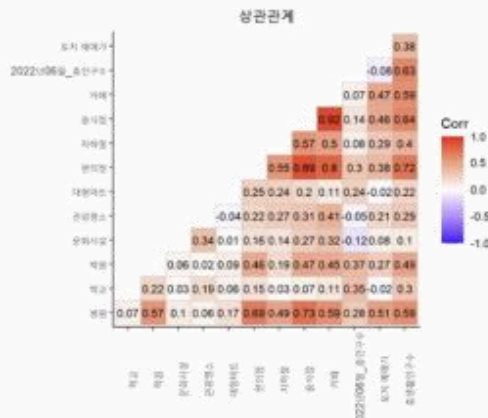


기본 가정:

선형성, 오차항의 독립성·등분산성·정규성,
다중공선성

선형 회귀 모델의 변수 선택

- R의 lmtest 패키지 활용하여 모델 적합
→ F 검정 통과 / 일부 변수 T 검정 통과 X
- 상관관계 플롯을 통해 변수들 간 높은 다중 공선성을 확인 → 변수 선택 진행 (AIC)



'편의점', '2022년 06월 총 인구수', '토지 매매가',
'관광명소', '카페' 설명 변수로 채택

선형 회귀 모델의 가본 가정 충족 확인

- R의 gvlma 패키지와 lmtest 패키지 활용
→ 모델이 선형성·정규성·등분산성 X
- Min-Max 스케일링, 반응변수의 로그 변환, 제곱근 변환 → 선형성과 정규성 X

라이브러리	선형회귀의 기본 가정		가정 충족 / 불충족
gvlma	Global Stat	선형성	불충족
	Skewness	왜도	충족
	Kurtosis	첨도	충족
	Heteroscedasticity	등분산성	충족
lmtest	shapiro.test	정규성	충족
	ncvTest	등분산성	불충족
	dwtest	오차항의 독립성	불충족



다중선형회귀를 이용한 생활 인구 예측이
적합하지 않다고 판단

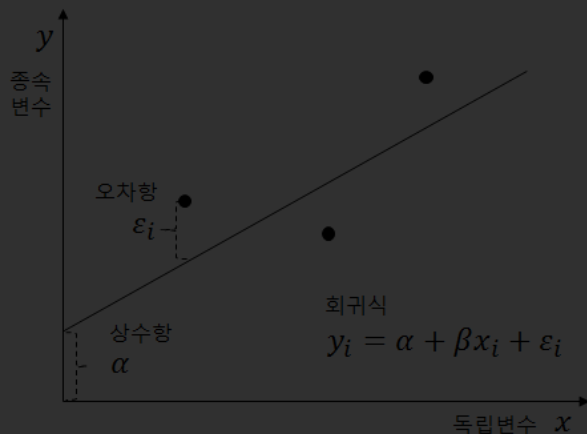
전처리

다중 선형 회귀를 이용한 생활 인구 예측

다중선형 회귀

설명 변수들의 선형 결합 이용해 반응 변수를 예측
→ 최소제곱법을 통해 모수 추정

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$



기본 가정:
선형성, 오차항의 독립성·등분산성·정규성,
다중공선성

선형 회귀 모델의 변수 선택

- R의 lmtest 패키지 활용하여 모델 적합
→ F 검정 통과 / 일부 변수 T 검정 통과 X
- 상관관계 플롯을 통해 변수들 간 높은 다중 공

선형모형으로는 예측하기 어렵다고 판단
따라서 **비선형 모형**을 사용



'편의점', '2022년 06월 총 인구수', '토지 매매가',
'관광명소', '카페' 설명 변수로 채택

선형 회귀 모델의 가본 가정 충족 확인

- 1) R의 gvlma 패키지와 lmtest 패키지 활용
→ 모델이 선형성·정규성·등분산성 X
- 2) Min-Max 스케일링, 반응변수의 로그 변환, 제곱근 변환 → 선형성과 정규성 X

라이브러리	선형회귀의 기본 가정		가정 충족 / 불충족
gvlma	Global Stat	선형성	불충족
	Skewness	왜도	충족
	Kurtosis	첨도	충족
	Heteroscedasticity	등분산성	충족
lmtest	shapiro.test	정규성	충족
	ncvTest	등분산성	불충족
	dwtest	오차항의 독립성	불충족



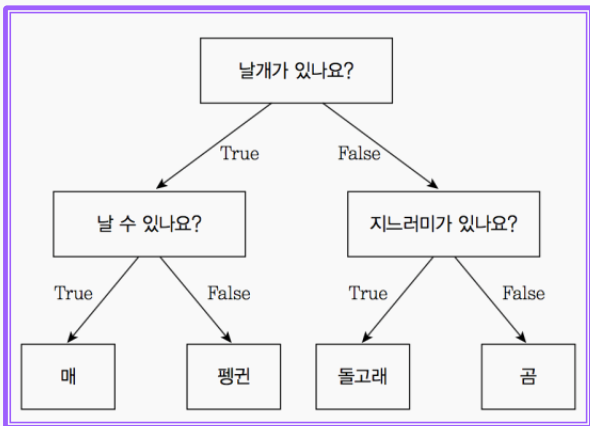
다중선형회귀를 이용한 생활 인구 예측이
적합하지 않다고 판단

전처리

생활 인구 예측에 사용될 모델 설명

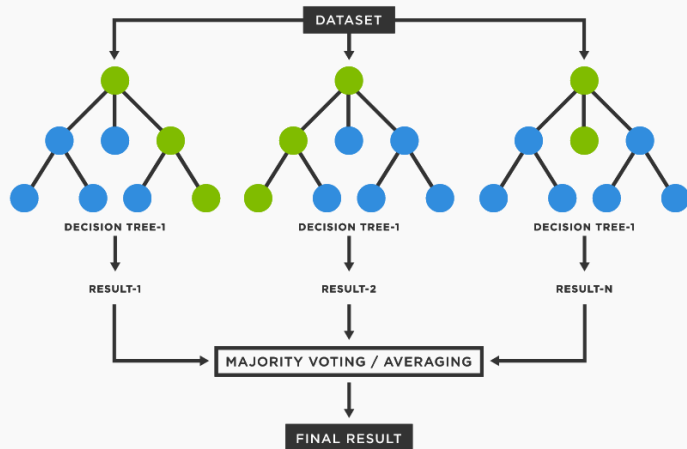
의사 결정 트리 모델

나무 구조의 분류 규칙: 데이터 분류 / 특성 예측



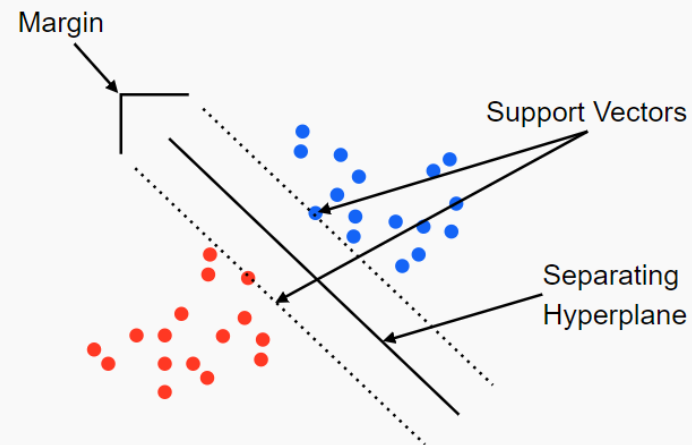
- CART 알고리즘: 이진 분할 방법을 이용해 노드 생성
- 노드의 분할 규칙 : 불순도(지니 지수, 엔트로피 지수, 편차 지수) 최소화

랜덤 포레스트



- 다수의 의사결정 트리를 구성하는 앙상블 학습을 통해 데이터 분류 및 회귀 수행
- 노드를 나타낼 때 무작위로 선택된 설명변수의 집합 중 최적의 결과 내는 방법 이용

SVM(서포트 벡터 머신)



여러 클래스에 속하는 데이터들을 최대한의 margin으로 분리하는 최적의 초평면을 찾는 방식으로 작동

전처리

최적의 모델 선정

데이터 분할

서울시 생활 인구 데이터를 각각 학습 데이터 / 검증 데이터로 8:2 비율로 분할하여 모델의 성능 평가

성능 평가 지표

MSE(평균제곱오차)

생활인구 값이 여덟 자리수로
설명변수의 단위보다 훨씬 큼

MAPE

잔차의 절대값에 대한 평균을 비율로
나타냄 → $100 - \text{MAPE}(\%)$ 값 채택

$$\text{MAPE} = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right|$$

R^2

회귀 모델에서 독립변수가 종속
변수를 얼마만큼 설명해 주는지를
가리키는 지표

전처리

랜덤 포레스트, 의사 결정 트리, SVM의 학습 결과

회귀 모델	100-MAPE(%)	R^2
RandomForestRegressor	79.132158263	0.68718228
DecisionTreeRegressor	73.7436045954	0.3974656714
SVM	59.13980930528	-0.075848099139

전처리

랜덤 포레스트, 의사 결정 트리, SVM의 학습 결과

회귀 모델	100-MAPE(%)	R^2
RandomForestRegressor	79.132158263	0.68718228
DecisionTreeRegressor	73.7436045954	0.3974656714
SVM	59.13980930528	-0.075848099139



평균값으로 예측하는 값보다 결과가 좋지 않음을 의미
일차적으로 **최종 모델 선정에서 배제**

전처리

랜덤 포레스트, 의사 결정 트리, SVM의 학습 결과

회귀 모델	100-MAPE(%)	R^2
RandomForestRegressor	79.132158263	0.68718228
DecisionTreeRegressor	73.7436045954	0.3974656714
SVM	59.13980930528	-0.075848099139

랜덤포레스트 회귀 모델이 (100-MAPE)(%) 값과 R^2 값이 제일 높음

→ 생활 인구 예측을 위한 최종 모델로 선택

전처리

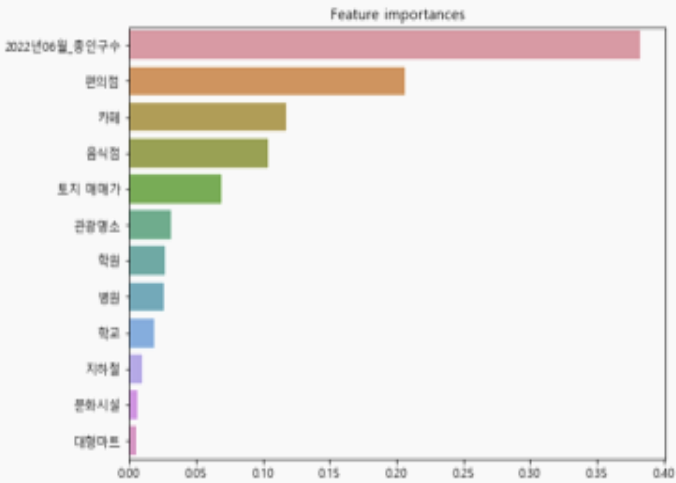
변수 선택 & 하이퍼 파라미터 튜닝을 통한 모델 성능 향상

모델 성능 향상 방법

모델의 예측 값과 실제 값 차이를 줄이는 방향으로 성능을 향상 시킴

변수 선택

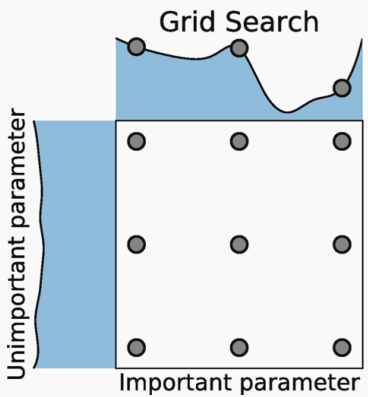
특성 중요도를 시각화 한 결과



상위 5개 변수로
예측 변수 한정

특정 변수를 기점으로 특성 중요도 값이
급격하게 낮아짐을 확인

하이퍼 파라미터 튜닝



그리드 서치 (GridSearch)

하이퍼파라미터를 순차적으로 입력해 학습을 하고 측정을
하면서 가장 좋은 파라미터를 알려주는 방법

GridSearchCV를 이용해 MAPE 값을 최소화하는
최적의 파라미터 값 탐색

	최적의 파라미터 값
n_estimators	300
max_features	3

전처리

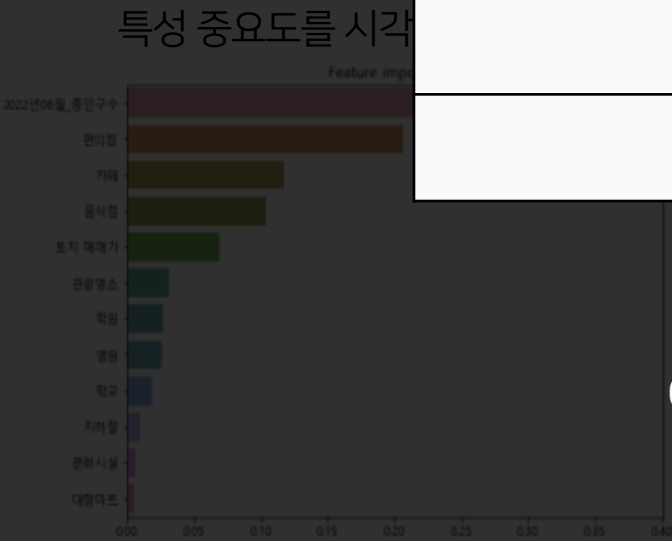
변수 선택 & 하이퍼 파라미터 튜닝을 통한 모델 성능 향상

모델 성능 향상 방법

모델의 예측 값과 실제 값 차이를 줄이는 방향으로 성능을 향상 시키는 방법

모델 튜닝 전후 성능 비교

변수 선택



특정 변수를 기점으로 특성 중요도 값이 급격하게 낮아짐을 확인

변수 추출/ 하이퍼 파라미터 최적화	100-MAPE(%)
이전	79.132158263
이후	80.67996688106766

예측 변수 한정

특성 중요도에 따라 변수를 추출하고
GridSearchCV를 활용하여 하이퍼파라미터 튜닝을 완료했을 때
(100-MAPE)(%) 값이 **상승한 것을 확인** 가능

하면서 가장 좋은 파라미터를 알려주는 방법

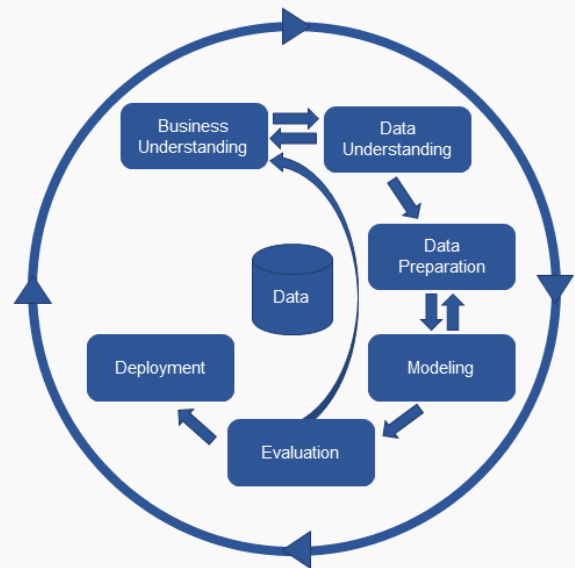
GridSearchCV를 이용해 MAPE 값을 최소화하는
최적의 파라미터 값 탐색

	최적의 파라미터 값
n_estimators	300
max_features	3

전처리

파생변수란?

기존의 변수를 조합하여 새로운 변수를 만들어내는 방법



생활 인구 예측 완료 후 최종적인 군집 분석에 있어서 수집한 데이터를
그대로 사용하기 보다는 지역의
특성을 더 잘 반영할 수 있는 파생 변수를 형성

파생변수

변수	수식
인구	$\frac{\text{주민등록인구}}{(\text{총 생활 인구 수} + \text{주민 등록 인구})}$
토지 매매가	$\frac{\text{토지 매매가}}{\text{면적}}$
산업 발전	$\frac{\text{학교} + \text{문화시설} + \text{관광명소} + \text{음식점}}{\text{면적}}$
전력 소모량	$\frac{\text{전력량}}{(\text{총 생활 인구수} + \text{주민 등록 인구})}$

* 인구: 유동적인 생활인구보다 주민등록인구가 쓰레기 매립지 건설 시 더 많은 영향을 받을 것이라고 판단하여 주민등록인구가 분자로 설정됨

Clustering: K-means

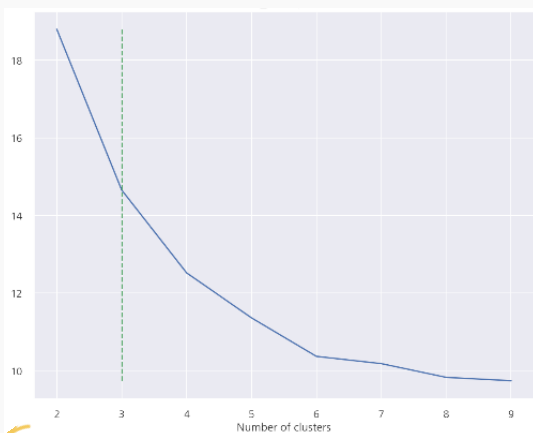
각 패턴과 클러스터의 중심 값과의 거리 차이를 최소화하는 방법

작동 원리

1. 패턴을 k개의 클러스터로 임의로 나눔
2. 클러스터에 포함된 패턴들의 평균으로 클러스터의 중심 값을 계산
3. 이 중심 값을 각 패턴과의 거리를 계산한 후 가장 거리가 가까운 클러스터에 패턴을 포함

K 개수 결정

-엘보우 기법-

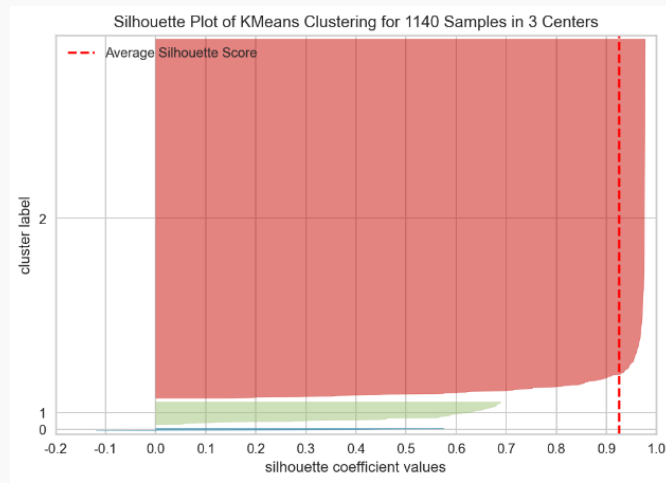


-실루엣 기법-

	silhouette_score
k=2	0.925032445836060
k=3	0.926185416835622
k=4	0.916950352350167
...	
k=8	0.894296261914259
k=9	0.857828060443337
best_n : 3	0.926185416835622

엘보우 기법과 실루엣 기법을 모두 고려하여 k값을 3으로 설정

결과



클래스 불균형이 커 데이터 군집이
잘 이루어지지 않았다고 판단



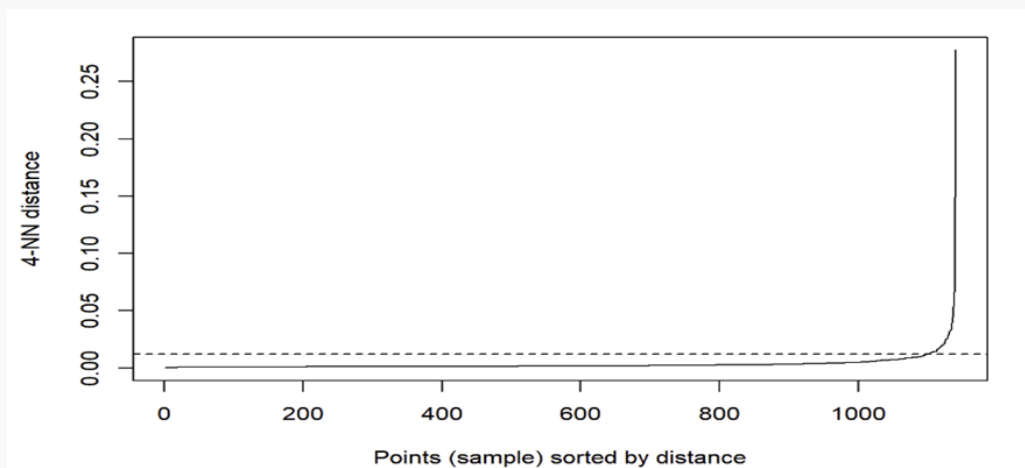
K-means 방법 **기각**

Clustering: DBSCAN (Density-based spatial clustering of applications with noise)

밀도 기반 군집화 방식으로 주어진 데이터 집합에서 클러스터와 클러스터에 속하지 않은 노이즈를 식별하는 방법

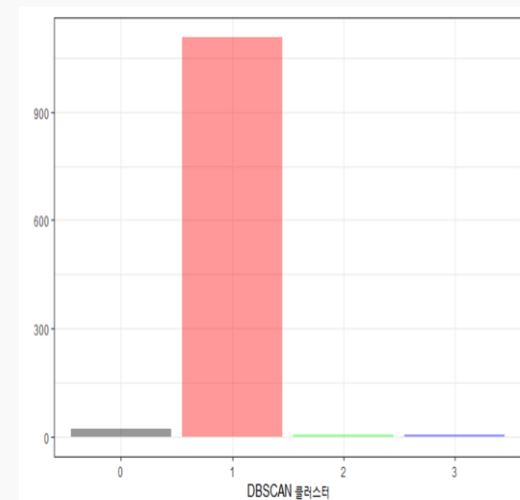
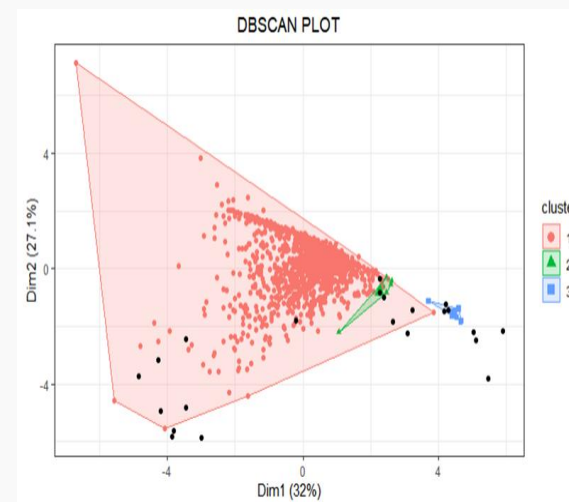
입력 모수 결정

<MinPts=4일 때의 k-dist의 시각화>



Marin Ester, Hans-Peter Kriegel, Jorg Sander, Xiaowei Xu가
제안한 Heuristic 한 방법으로
Ept=0.012 & MinPts=4로 설정 후 클러스터링

결과



하나의 클러스터와 소수의 노이즈로 인식되는
심각한 클래스 불균형이 발생



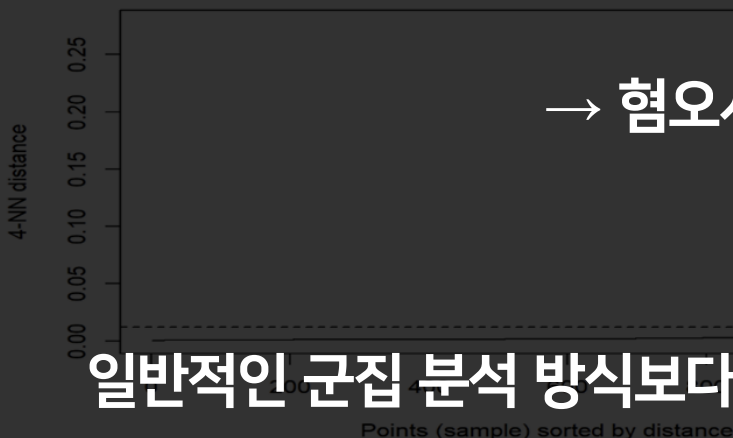
DBSCAN 방법 기각

Clustering: DBSCAN (Density-based spatial clustering of applications with noise)

밀도 기반 군집화 방식으로 주어진 데이터 집합에서 클러스터와 클러스터에 속하지 않은 노이즈를 식별하는 방법

입력 모수 결정

<MinPts=4일 때의 k-dist의 시각화>



심각한 클래스 불균형 발생

→ 혐오시설이라는 매립지의 시설적 특징을 고려



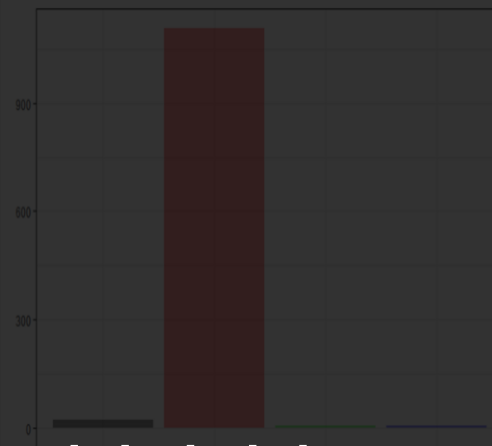
일반적인 군집 분석 방식보다는 이상치 탐지로 대체 매립지 후보 지역을 찾는 것이 더 타당

Marin Ester, Hans-Peter Kriegel, Jorg Sander, Xiaowei Xu가

제안한 Heuristic 한 방법으로

Ept=0.012 & MinPts=4로 설정 후 클러스터링

결과



하나의 클러스터와 소수의 노이즈로 인식되는
심각한 클래스 불균형이 발생



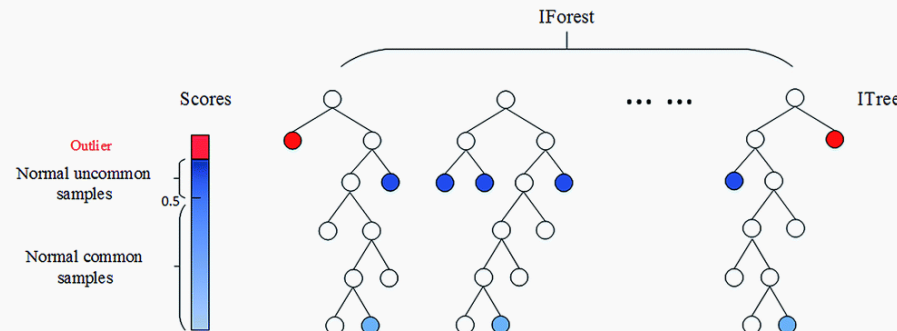
DBSCAN 방법 기각

이상치 탐지: Isolation Forest

이상치의 고립이 정상치를 고립시키는 것보다 쉽다는 개념에서 착안한 이상치 분류 모델

알고리즘

- 트리 기반 비지도 이상치 탐지 모델- 무작위로 데이터를 분할하여 관측치를 고립
- 이상치의 경우 더 적은 횟수의 분할로 고립시킬 수 있다는 개념 하에 모델이 작동
- 이상치는 고립에 취약 → 트리모델의 뿌리에 가깝게 고립될 가능성이 높다 (상단부)
- 정상치는 트리모델의 하단부에 깊게 위치



수도권 데이터를 Isolation Forest에 적합

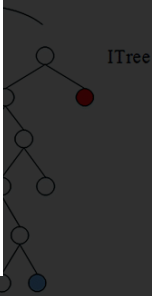
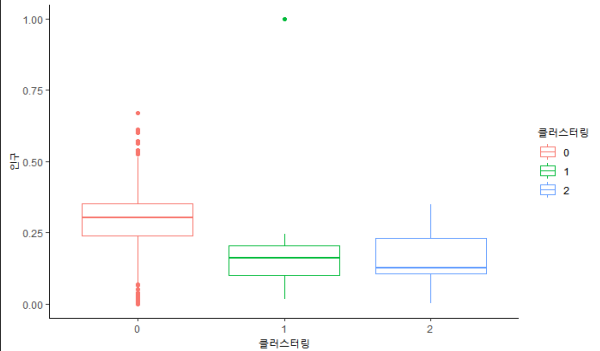
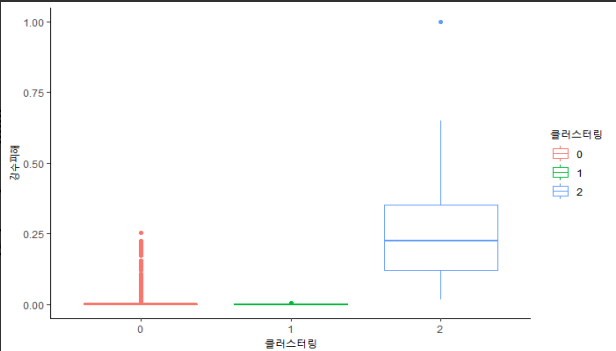
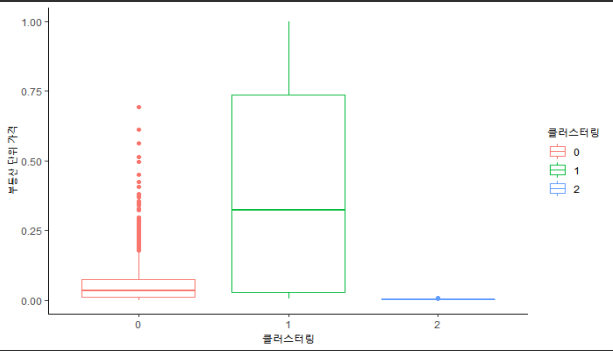
1. Min-Max 스케일링 진행 → R의 isotree 패키지를 활용 → Isolation Forest에 적합
2. 이상치에 해당하는 지역의 개수는 26개로 나타남
3. 매립지 선정에 있어서 실매립 면적인 100만 평방미터를 수용할 수 없는 지역을 후보에서 제외
4. 이상치로 분류된 지역 : (주민 등록 인구+ 생활 인구) 값을 기준으로 내림차순 정렬
→ 상위 50% 와 하위 50%으로 나누어 분포를 확인

이상치 탐지: Isolation Forest

이상치의 고립이 정상치를 고립시키는 것보다 쉽다는 개념에서 착안한 이상치 분류 모델

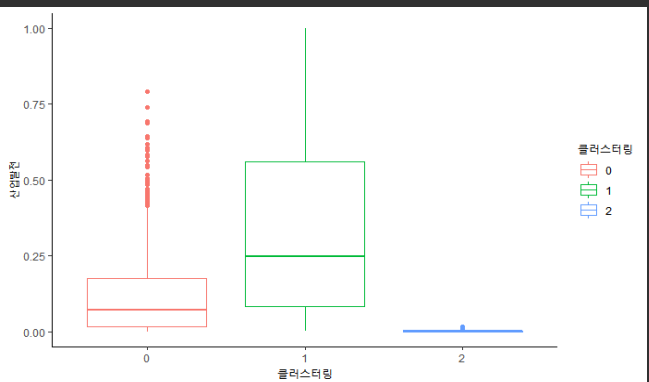
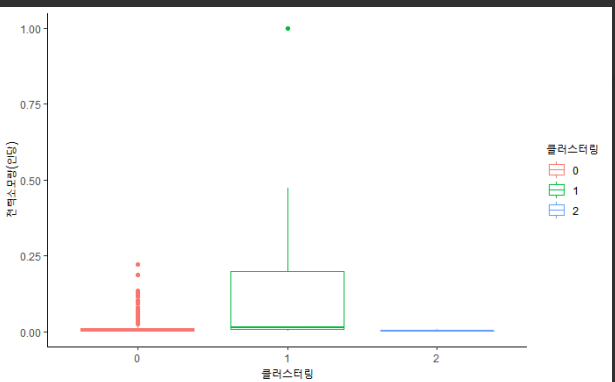
알고리즘

- 트리 기반
- 이상치 탐지
- 이상치 분류
- 정상치 분류



수도권 데이터를 Isolation Forest

1. Min-Max 스케일링 진행 →
2. 이상치에 해당하는 지역의 가
3. 매립지 선정에 있어서 실매립 면적인 100만 평방미터를 수용할 수 없는 지역을 후보에서 제외
4. 이상치로 분류된 지역 : (주민 등록 인구 + 생활 인구) 값을 기준으로 내림차순 정렬



[전력 소모량 (인당)]

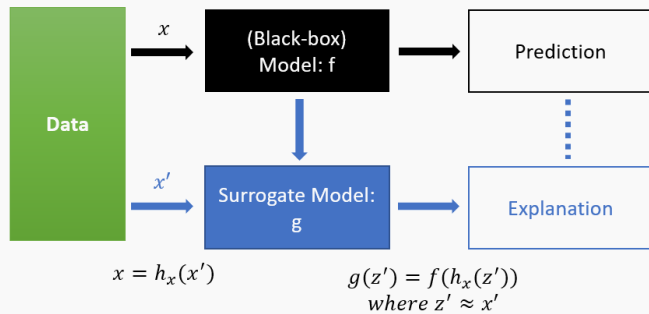
[산업발전]

부동산 단위 가격, 전력 소모량, 산업발전, 강수피해에서 뚜렷한 분포 차이를 확인 가능

SHAP

인공지능, 머신 러닝 모델 위에 설명성을 부여해 블랙박스를 들여다 볼 수 있게 해주는 방법

알고리즘 및 작동원리



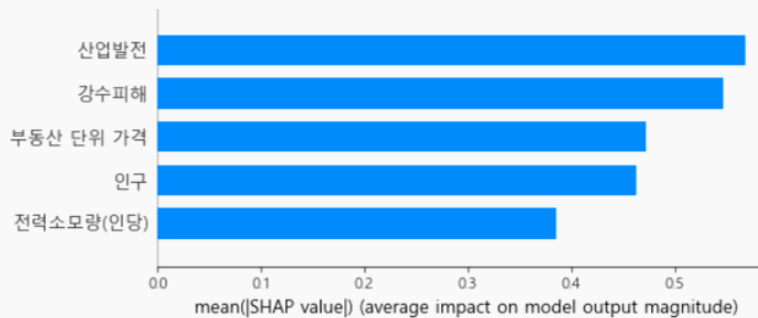
예측 모델의 결과에 대한 변수 기여도를
모든 가능한 변수 조합에 대하여 종합적으로 구한다
(변수 간 의존성까지 고려)



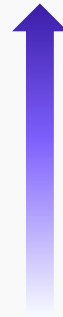
결과에 미치는 영향력의 방향성과 크기도 확인 가능

변수 중요도

- # 각 변수의 평균 영향력을 나타냄
- # Isolation Forest에 SHAP을 적용



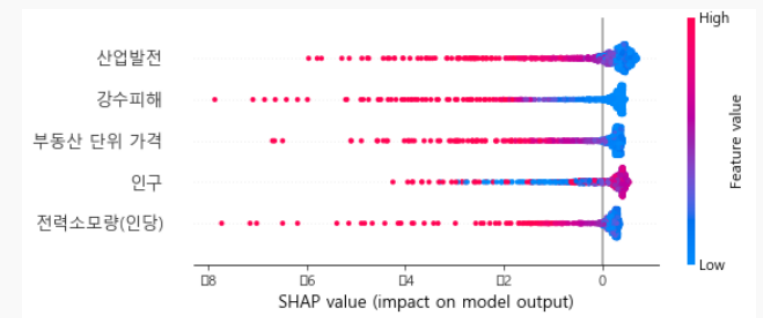
산업발전
강수 피해액
부동산 단위 가격
인구
인당 전력소모



영향력이 커짐

값의 분포도 및 영향도

- # SHAP value의 절대값이 높은 순으로 표시
- # 중앙선 기준으로 왼쪽은 예측 값이 낮아지는데 기여



산업발전이 낮을수록
강수피해액이 낮을수록
부동산 단위 가격이 낮을수록
인구가 높을수록
인당 전력소모가 낮을수록



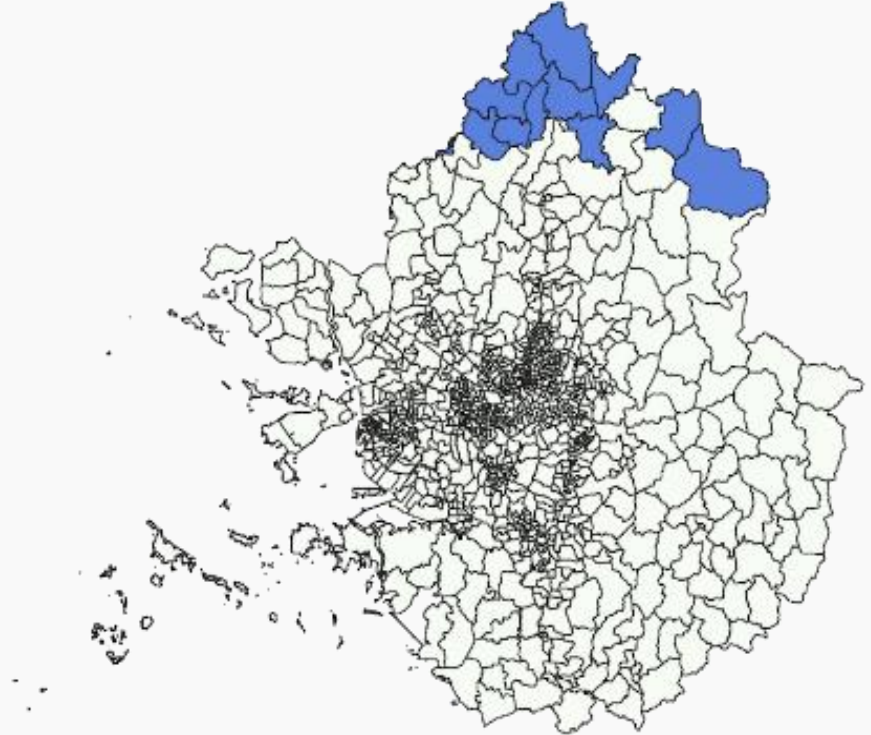
선정될 확률이 높음

인구: 주민등록인구 / (주민등록인구 + 생활 인구)

결과

한 줄 설명

결과



최종 결과
□ 안됨
■ 됨

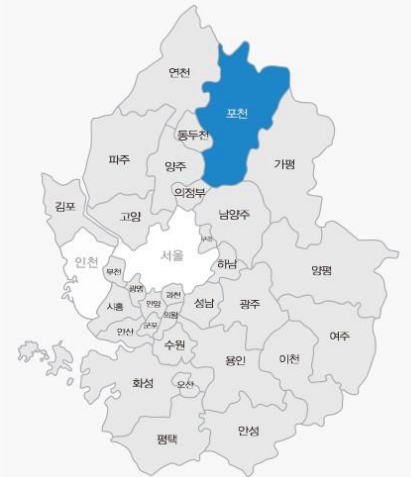
● Isolation Forest (이상치 탐지) 결과
7개의 읍면동이 선정됨

7개의 읍면동
포천시 창수면, 포천시 이동면, 포천시 관인면,
연천군 연천읍, 연천군 관남면, 연천군 백학면, 연천군 미산면

● 쓰레기 매립지는 큰 면적을 요구함으로 포천시 / 연천군으로 선정

연천군은 군사기밀 지역이기 때문에 제외

● 최종 선정 지역: 포천시



갈등 해소 및 경제적 효과

혐오시설로 여겨지는 것이 아닌 하나의 문화공간이자
경제적 효과를 볼 수 있다



주민 친화적 & 환경 친화적

- 남양주 에코-랜드의 경우 부지 면적 일부분을 실내 체육 시설 등으로 활용
- 매립지 외곽으로는 산책로 조성
- 환경 교실 운영으로 이미지 개선
- 수질 및 가스 확산 여부 등 정기조사



새로운 관광지로 탈바꿈

- 해외 사례의 경우 소각장 주변을 공원화하고 스포츠 시설로 조성함으로써 관광지가 됨
- 지역의 경제와 토지 이용의 능률 소생



매립지 폐쇄 후 생태 복원 및 인근 환경 개선 효과

난지도, 친환경적인 공원으로 탈바꿈

- 많은 시민들이 찾는 공원
- 다양한 동식물이 살고 있음
- 매립가스 회수를 통해 자원 절약
- 미개발 지역 택지 개발

결론

이상치 탐지를 이용한 쓰레기 매립지 입지선정 및 향후 토지 이용 방향

분석 흐름

데이터를 확보하는 데 있어 생활인구 관련 데이터가 부족하여 랜덤포레스트와 같이 다양한 방법론을 통해 생활인구를 예측

→ K-means 클러스터링과 DBSCAN 방법 결과 클래스 불균형이 심하다는 것을 확인

→ 이상치 탐지를 통해 최종 후보 지역을 선정

→ 해당 결과를 토대로 SHAP을 통해 변수 중요도를 구하여 각 변수가 의사결정에 얼마만큼의 효과가 있었는지 확인

→ 정부가 내세운 쓰레기 매립지 입지 기준에 맞추어 최종 지역을 선정

의의

- 공간 분석과 환경 데이터 위주로 분석했던 기존의 연구들과 달리 사회적인 관점에서 접근
- 공공데이터의 활용 및 데이터 기반의 의사결정의 필요성과 접근 방법을 보여줌

한계점

- 구하고자 하는 시설의 특성상 다양한 모델을 사용하는 것이 힘들어 Isolation Forest만 최종적으로 사용
- 사용된 데이터의 수집시점이 모두 다름

참고문헌

- 김병철, 오상영, 류근호, 2006, 영향력을 고려한 적정입지선정 모델 연구, 한국산학기술학회논문지, p895-900
- 김판준, 2019, 랜덤포레스트를 이용한 국내 학술지 논문의 자동분류에 관한 연구, 정보관리학회지, p59
- 김현일, 이연수, 김병현, 2021, 랜덤포레스트 회귀모형을 적용한 도시지역에서의 실시간 침수 예측, 한국수자원학회 논문집, p1121
- 김현정, 2021, 소하천 재난관리체계 마련을 위한 k-means 기반의 군집화 평가, 부산대학교, p 11
- 민준영, 1995, 신경망 클러스터리의 성능평가: GLVQ와 k-means알고리즘을 중심으로, 성균관대학교 대학원, p25-26
- 박성우, 노윤아, 정승민, 황인준, 2021, LSTM을 사용한 SHAP 기반의 설명 가능한 태양광 발전량 예측 기법, ACK 2021 학술대회 논문집, p845-848
- 박순호, 1997, 농촌지역 쓰레기 매립장 입지 선정에 관한 연구-경상북도 영양군을 사례로, 한국지역지리학회지, p66-80
- 박환용, 정일훈, 김철중, 2010, 도시공공시설의 적정입지 선정에 관한 연구: 파주시를 중심으로, 국토연구 제 66권, p148-168
- 배민기, 장병문, 1998, 지리정보체계를 이용한 일반폐기물 매립후보지의 입지선정에 관한 연구, 한국지리정보학회지 1권 2호, p14-25
- 송용정, 김현덕, 2012, 의사결정나무분석을 이용한 컨테이너 수출입 물동량 예측, 한국항만경제학회지, p198
- 오재식, 최준호, 2009, 매립장 입지선정의 갈등과 해소방안 - 대구시 쓰레기 매립장 사례를 중심으로-, 영남대학교, p1-112
- 정상용, 2001, 비위생 생활폐기물 매립지의 문제점과 대책, 한국농공학회지 제 43권 제6호, p.44-53
- 진승종, 유상철, 김남기, 하윤우, 왕지남, 2022, AutoEncoder/Isolation Forest 알고리즘을 이용한 용접 공정 시계열 데이터 이상탐지, 한국경영과학회 학술대회논문집, p3
- Martin Ester, Hans-Peter Kriegel, Jiirg Sander, Xiaowei Xu, 1996 ,A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Institute for Computer Science, University of Munich,p230
- 안재현. 『XAI 설명가능한 인공지능, 인공지능을 해부하다』. 위키북스, 2020