

생활쓰레기 증가 요인 분석 및 개선 방향

기해인 · 김다희 · 김윤경



INDEX

1. 서론

- 문제점 및 진행방향
- 변수 소개

2. 분석

- 분석방법1 : 다변량 회귀분석
- 분석방법2 : 기계학습
 - 1) KNN
 - 2) SVM
 - 3) Random Forest

3. 활용방안

문제점 및 진행방향

인천의 쓰레기 매립지가 2025년 폐쇄 예정인 가운데

생활 쓰레기 처리 문제가 심각해지고 있다.

여러 정책들이 시행되고 있지만 여전히 어려움이 존재하는

상황 속에서 쓰레기 대란의 대책이 절실한 상황이다.

▶가정에서의 일상생활 소비에 의한 쓰레기 배출 문제점을

확인하기 위해 생활 폐기물 분석 자료 채택

▶생활 폐기물에 영향을 주는 요인 파악 및

쓰레기 배출 문제점에 대한 개선 방향 도출

활용데이터	출처	기타
서울시 폐기물 재활용 현황 통계	서울 열린데이터 광장	2010, 2015-2018년도
서울시 연면적별 주택현황 통계	서울 열린데이터 광장	
서울시 1인가구(연령별) 통계	서울 열린데이터 광장	
인구주택총조사	SGIS	2017,2018년도 서울 특별시 20-30대 남녀인구
온라인쇼핑동향조사	KOSIS	2017,2018년도 음/식료품

변수 소개

변수명	계산식	출처 데이터
생활폐기물 변수	= 생활폐기물 배출량	서울시 폐기물 재활용 현황 통계 데이터
재활용 비율 변수	= 생활폐기물 재활용량/생활폐기물 배출량	
주택현황변수 (주택형태 : 아파트, 단독주택, 연립주택, 다세대주택)	= 특정 주택형태 합/모든 주택형태 합	서울시 연면적별 주택현황 통계 데이터
1인가구(연령별) 변수 (연령층 : youth, middle, elderly)	= 특정 연령층 1인가구 수/전체 1인가구 수	서울시 1인가구(연령별) 통계 데이터
배달변수	= 서울시 지역구 20, 30대 인구 비율 × 음식서비스 거래액	인구주택총조사데이터 온라인쇼핑동향조사 데이터

분석방법1 : 다변량 회귀분석

생활폐기물 배출량과 생활폐기물에 영향을 줄 것으로 예측되는 여러 요인들과의 관계를 알아보기 위하여 다변량 회귀분석을 실시

종속변수 : 생활폐기물 배출량

독립변수 : 재활용 비율, 아파트 비율, 단독주택 비율, 연립주택 비율, 다세대주택 비율, 청년층/중년층/장년층 1인가구 비율, 배달 거래액

step1) 종속변수와 독립변수들 간의 상관관계 확인

step2) 종속변수 및 독립변수 log변환

step3) 단계 선택법(stepwise selection)을 통해 최종 변수 선택

⇒재활용 비율, 아파트 비율, 연립주택 비율, 다세대주택 비율,

청년층 1인가구 비율이 최종 변수로 선택

	Estimate	Std.Error	t value
(Intercept)	2.4820	0.8291	2.994 **
재활용 비율(recycle)	0.7243	0.4046	1.790 .
아파트(apart)	2.3682	0.7300	3.244 **
연립주택(town)	5.7115	2.9526	1.934 .
다세대주택(multiplex)	1.9259	0.8109	2.375 *
Youth	1.2840	0.3590	3.577 ***

(*** p<0.001, ** p<0.01, * p<0.05, . p<0.1)

분석방법2 : 기계학습

① WHY?

생활쓰레기 배출량이 비슷한 지역들의 공통적인 특징을 바탕으로 (특정) 지역별 생활쓰레기 배출량을 분류하고, 분류된 범주에 따른 주요 5개 변수들의 분포형태와 현행 정책들을 함께 검토한다면 기존 정책을 보완할 개선점 및 새로운 정책 방안 모색이 가능

② HOW?

▶기계학습 활용자료

종속변수 : 3구간 (low,middle,high)으로 범주화시킨 생활폐기물 배출량

독립변수 : 다변량 회귀분석 단계를 통해 도출된 5개의 변수

(재활용 비율, 아파트 비율, 연립주택 비율, 다세대주택 비율, 청년층 1인가구 비율)

▶분석순서

step1) 5-fold cross validation을 50~100번 반복해 최적 모수 추정

step2) 자료를 train set과 test set으로 나누어 추정 모수를 이용한 모형(분류모델:KNN,SVM,RF) 훈련 및 정확도 확인

step3) 분류된 test set을 바탕으로 생활 폐기물 범주별 독립변수들의 boxplot 확인

분석방법2 : K-Nearest Neighbor

▶ KNN : 분류하고자 하는 데이터 주변의 k개 데이터가 무엇인지에 따라 분류하는 기계학습 방법

모델	k	Train.accuracy	Test.accuracy
KNN	5	0.5043	0.6667

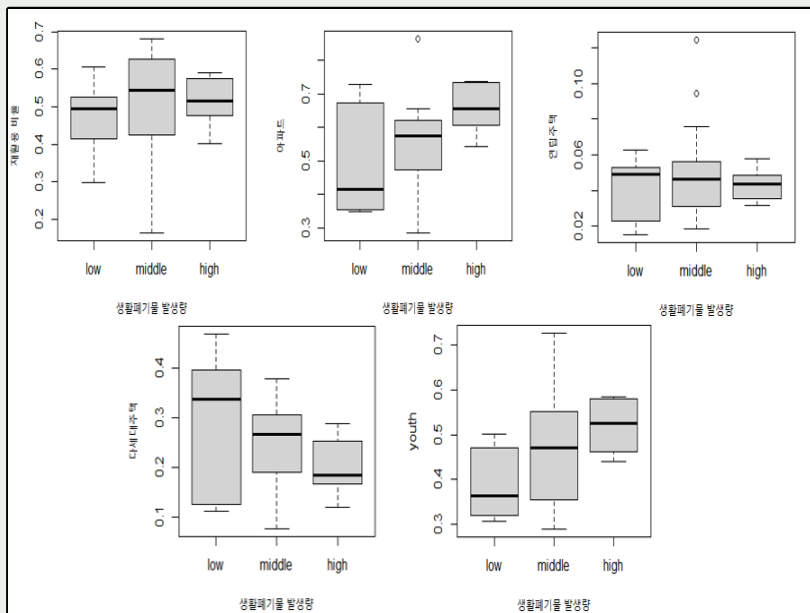
⇒최적의 모수 k=5로 최종 분류한 결과 정확도는 0.6을 넘는 것으로 확인

boxplot

⇒생활폐기물 배출량이 많은 지역에서 아파트와 청년층의 비율이 높게 나타남

⇒다세대주택의 경우 생활폐기물 배출량이 적은 지역들의 범위가 비교적 크지만 middle과 high에 비해서는 다세대 주택 비율이 비교적 높게 나타남

⇒생활폐기물 배출량에 따라 재활용 비율, 연립주택 비율은 크게 차이 나지 않음



분석방법2 : Support Vector Machine

▶ SVM 모델 : 퍼셉트론을 토대로 최적의 초평면을 경계로 하여 데이터 집합을 나누는 선형분류기법

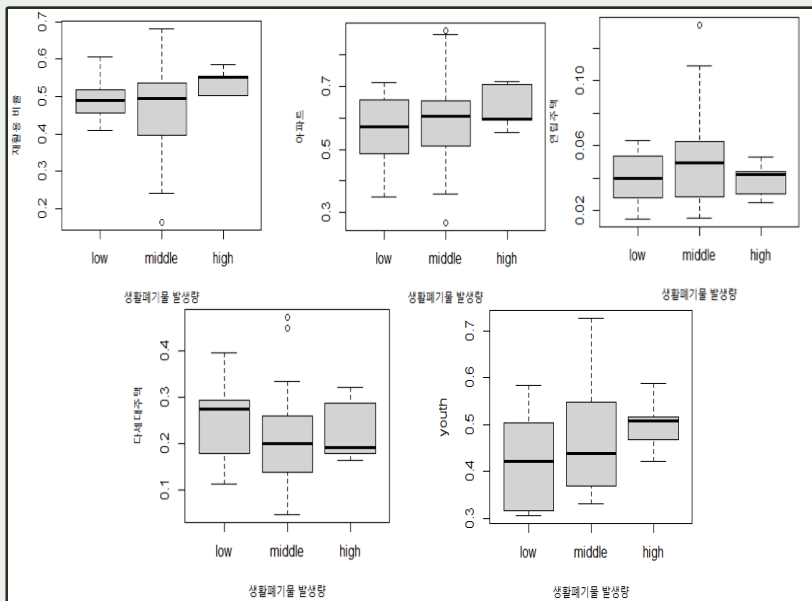
모델	gamma	Cost	Train.accuracy	Test.accuracy
SVM	10	10	1	0.75

⇒ 교차 검증으로 최적 모수가 gamma=10, cost=10로 추정.
이 때의 정확도는 0.75로 확인

boxplot

⇒ 분석결과 청년층과 아파트에 비례하여 영향이 크고, 연립주택의 관련성은 적은 것으로 나타남

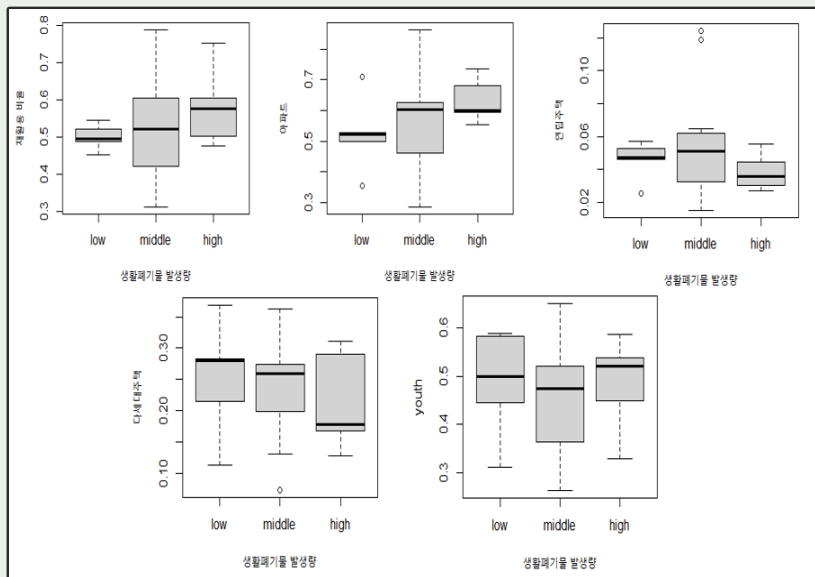
⇒ 다세대주택, 재활용 비율은 연립주택보다는 영향력이 있지만 생활폐기물 발생량 범주에 따른 차이는 적음



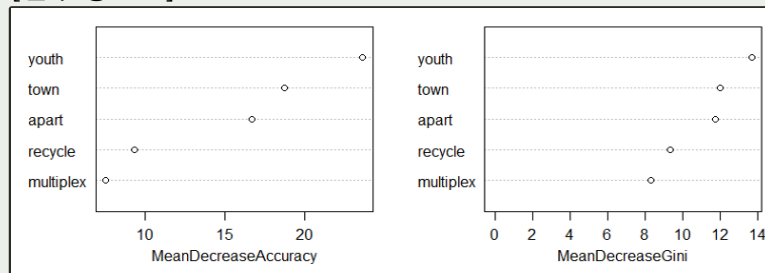
분석방법2 : Random Forest

▶ 랜덤포레스트 : 다중의 무작위 의사결정 나무(decision tree)들이 분류한 결과들을 종합하여 최종 분류하는 기계학습 방법

모델	Mtry	Ntree	Train.accuracy	Test.accuracy
Random forest	2	300	0.7416	0.7222



[변수 중요도]



⇒ 최적 모수 mtry=2, ntree=300를 이용한 최종 분류 정확도는 0.7이 넘는 것으로 확인

⇒ 변수의 중요도를 나타내는 2개의 지수 모두 청년층 비율 변수가 가장 큰 값을 가짐(데이터 분류에 가장 큰 기여)

boxplot

⇒ 앞의 두 모형과 달리 생활폐기물 발생량이 낮은 지역에서 청년층 비율이 비교적 높게 나타남

⇒ 생활폐기물 발생량에 따른 나머지 변수들의 분포는 앞의 두 모형의 결과와 유사

활용방안

▶다중 회귀분석을 통해 도출된 5개의 변수에 초점을 맞추어 개선방향 소개

▶재활용 비율-재활용에 대한 인식 개선과 적극적인 참여 필요

① 올바른 분리배출 방법의 적극적 홍보

② 재활용 품목 간소화

▶아파트/연립주택/다세대주택 비율-혼합배출의 문제점 해결

① 지자체 거점 수거 시설 설치 및 수거 체계 확립

② 지역화폐 및 포인트 형태 등의 보상 제공으로 주민 참여 활성화

▶청년층 1인가구 비율-일회용품 줄이기에 대한 경각심 필요

① 분리수거 참여 또는 일회용품 사용 지양에 대한 보상 제공

▶궁극적인 재활용 방안

① 생산자 책임 재활용 제도(EPR) 확대

② 자원순환형 사회 - 폐기물의 자원화

Thank you!

