

경기도 스마트도시 구축을 위한 도시데이터 센서(SOGO) 위치 선정

팀명 : 쏘고

팀원 : 김도현, 안성재, 유상현

I. 서론

스마트도시란 정보 및 통신 기술을 활용하여 인프라, 서비스, 모빌리티, 에너지 등 다양한 도시기능을 효과적으로 관리하고 최적화하는 혁신 도시이다. 자원 재활용, 에너지 효율 개선, 환경보호등 다양한 측면에서 많은 이점을 제공함으로써 지속가능한 발전과 삶의 질 향상을 도모할 수 있다. 도시데이터 센서인 SOGO를 통해 다양한 데이터를 수집하고 활용하여 경기도의 지속가능한 성장과 시민의 삶의 질 향상에 기여하고자 한다. 특히 지속가능한 성장을 위해 스마트시티에 전 세계가 많은 관심을 가지는 만큼 스마트센서 설치를 통해 이를 실현하고자 한다.



그림 1 스마트시티 시장 규모의 급격한 상승

II. 문제 정의

본 연구의 목적은 유동인구 데이터와 대기오염 데이터 등을 활용해 최적의 도시데이터 센서 입지 선정을 통해 기존 스마트시티의 단점을 보완한 스마트시티를 구축하고자 한다. 도시데이터 센서를 활용한 데이터 플랫폼을 통해 개인, 공공기관, 기업에 도시 데이터 센서를 제공함으로써 데이터 기반의 효과적인 의사결정을 통한 도시의 지속가능한 성장과 도시 서비스 효율화를 통한 삶의 질 향상에 이바지하는 것을 목적으로 한다.

III. 데이터 정의

1. 데이터 수집

센서 입지 데이터는 경기데이터드림의 CCTV 현황 데이터와 기상청의 방재기상관측소 지점 데이터를, 대기오염 데이터는 환경 빅데이터 플랫폼의 대기오염 측정정보와 경기데이터드림의

폐기물처리업체현황 데이터를, 소음 및 유동인구에는 경기데이터드림의 경기도 지하철 승하차 데이터, 국가철도공단의 수도권 호선 역위치 데이터, 경기도 교통정보센터의 버스 승하차 데이터를 활용하고, 마지막으로 건강 취약 계층은 공공데이터 포털의 전국초중등학교 위치표준 데이터를 활용한다.

2. 데이터 전처리

1) CCTV와의 거리계산이 필요한 데이터(지하철, 학교, 관측소)에 대해서는 QGIS 프로그램을 사용해서 새로운 컬럼으로 추가하고, 그 중에서 지하철과 학교의 경우는 거리가 가까울수록 큰 값을 갖도록 하기 위해 역수함수를 적용한다. 그러나 지하철역이나 학교 내에 설치되어 있는 CCTV의 경우 거리가 0이거나 0에 근사하여 편차가 심해지기 때문에 10m를 epsilon 값으로 더해준다(로그변환에서 $\log(1+값)$ 을 해주는 이유와 같다).

$$\text{지하철역 중요도 지수화} = \frac{\text{승하차인원}}{(\text{distance} + \text{epsilon})}$$

그림 2 역수 함수를 통한 지하철역 중요도 지수화

2) 버스정류장과 폐기물 처리시설에 대해서도 마찬가지로 QGIS 프로그램을 사용해서 반경 300m 내에 승하차 인원의 총합과 폐기물 처리시설 개수의 합계를 각 CCTV 데이터에 매핑한다. 특히 소음의 경우 일정 거리를 벗어나면 크게 감소하기 때문에 유의미하게 작용하는 거리를 정하는 것이 중요하게 작용하였다. 버스정류장의 반경 300m기준은 주택건설기준 제 9조를 참조, 자동차의 소음이 최대 300m이내까지 영향을 준다고 판단하였음. 폐기물 처리업체의 반경 300m 기준은 폐기물 처리시설 설치 촉진 및 주변지역지원등에 관한 법률을 참조, 명시된 간접영향권 300m로 설정하였음.

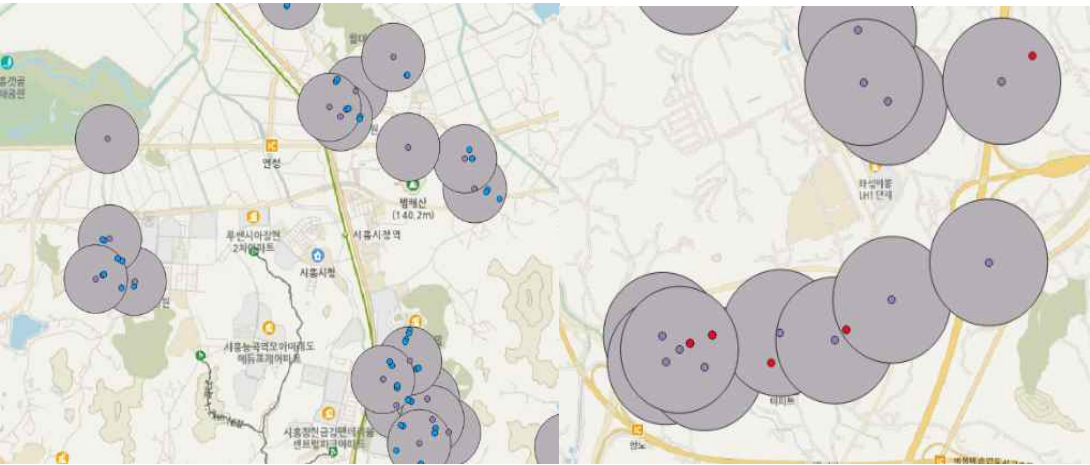


그림 3 CCTV 300m이내 버스정류장

그림 4 CCTV 300m이내 폐기물 처리시설

3) 대기오염 데이터 보간

대기오염 데이터의 경우 측정소 별 데이터를 각각의 CCTV에 매핑하기위해 측정소의 위경도 데이터를 활용하여 IDW 보간을 통해 CCTV위치의 대기오염 농도를 보간한다. 이때 측정소 간의 거리에 따라 가중치가 부여된다. IDW 보간 결과를 사용하여 대기오염 지도를 생성하고

생성된 대기오염지도와 CCTV를 매핑한다.

3. 결측치 처리

대기오염 데이터는 관측지점 별로 결측치 비율을 확인한 결과, 측정소 별로 일부 기간에 결측치가 존재함을 확인. 대기오염 데이터에 있던 결측치를 측정소별로 스플라인 보간법을 활용해 결측치를 대체한다.

4. 이상치 처리

지하철 역과 CCTV간의 거리, 학교와 CCTV간의 거리를 역수 함수를 취하는 과정에서 CCTV가 지하철 역이나 학교와 인접한 경우 기하급수적으로 커진 지표에 대한 이상치를 확인 및 처리한다. 데이터가 정규분포를 이루고 있지 않은점을 고려하여 데이터의 분포에 덜 민감하며 비정규 분포에서도 효과적으로 이상치를 판단하는 IQR(Interquartile range) Method를 통한 이상치를 판단하였다. $Q1 - 1.5 * IQR$ 보다 작거나 $Q3 + 1.5 * IQR$ 보다 큰 값을 이상치로 정의한다. 역수를 취하는 과정에서 기하급수적으로 높은 값을 가지게 된 이상치의 경우 평균이나 중앙값으로 대체할시 데이터의 왜곡의 우려가 있어 KNN을 통해 이상치를 대체하였다. 비슷한 패턴이나 특성을 가진 데이터를 통해 이상치를 대체해 적합하다고 판단하였다.

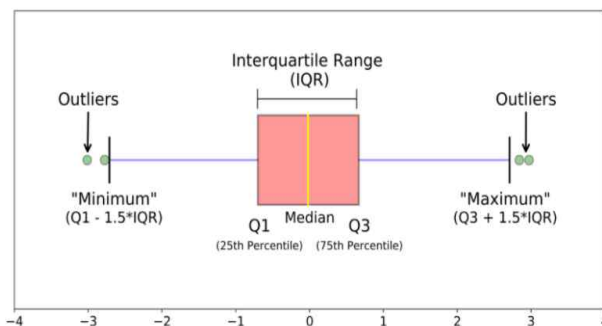


그림 5 IQR을 통한 이상치 판단

IV. 모델정의

1) 클러스터링을 피쳐 엔지니어링 전처리가 완료된 데이터에 대하여 StandardScaler, MinMaxScaler, RobustScaler를 적용한 데이터 분포 비교와 각 스케일러의 특징을 고려하여 StandardScaler를 적용한다. 비정규분포가 가지는 이상치가 모델에 주는 영향을 줄이고 모델의 안정성을 향상시키기 위해 사용.

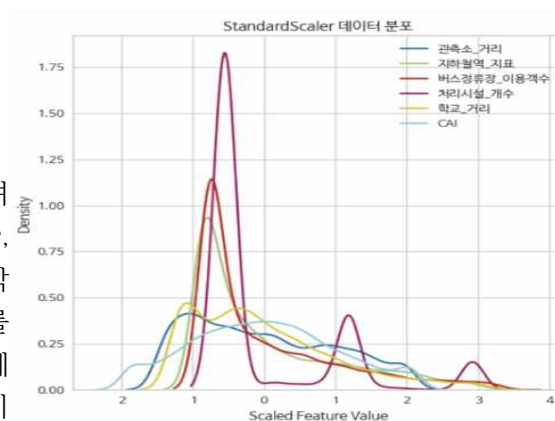


그림 6 standard 스케일러를 통한 분포확인

2) 군집의 중심 좌표를 활용한 군집화 스케일러를 적용한 데이터에 대하여 클러스터링 모델인 K-means, K-medoids,

GMM(Gaussian Mixture Model)을 사용한다. 이는 10,000개의 CCTV 위치 중 다양한 측면에서 데이터 간의 패턴을 식별하고, 각 클러스터별 중심 좌표의 특성을 비교하여 공통되는 CCTV 지점을 도출함으로써 서로 유사한 특성을 가진 지역을 군집화하여 센서 후보지를 선정하기 위함이다. 군집의 개수는 Silhouette Score를 활용해 각 클러스터 별로 최적의 군집 개수를 활용하였다.

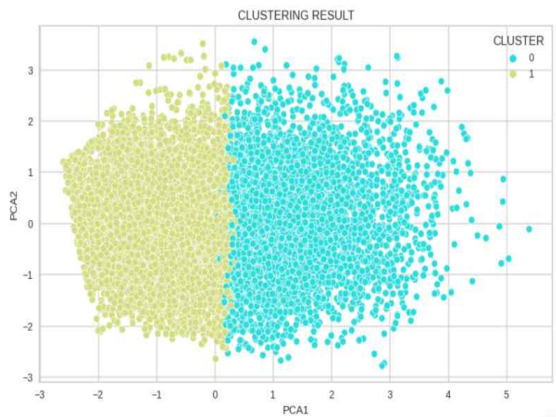


그림 7 K-means 클러스터링

	관측소_거리	지하철역_지표	버스정류장_이용객수	처리시설_개수
CLUSTER 0	-0.404028	0.784483	0.765935	0.440554
CLUSTER 1	0.265669	-0.515837	-0.503640	-0.289686
	학교_거리	CAI	SUM	X
CLUSTER 0	0.616369	-0.184571	2.018741	
CLUSTER 1	-0.405293	0.121365	-1.327423	

그림 8 K-means 군집의 중심 좌표 특성



그림 9 K-medoids 클러스터링

	관측소_거리	지하철역_지표	버스정류장_이용객수	처리시설_개수
CLUSTER 0	-0.322568	0.353609	0.458415	-0.548615
CLUSTER 1	0.204743	-0.673708	-0.737605	-0.548615
	학교_거리	CAI	SUM	X
CLUSTER 0	0.621869	-0.037169	0.525541	
CLUSTER 1	-0.763181	0.062680	-2.455686	

그림 10 K-medoids 군집의 중심 좌표 특성

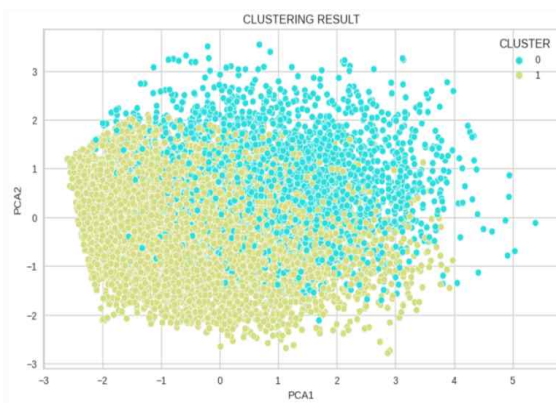


그림 11 Gaussian Mixture 클러스터링

	관측소_거리	지하철역_지표	버스정류장_이용객수	처리시설_개수
CLUSTER 0	-0.220296	0.325197	0.329317	1.460023
CLUSTER 1	0.082778	-0.122195	-0.123743	-0.548615
	학교_거리	CAI	SUM	X
CLUSTER 0	0.194120	0.000239	2.088600	
CLUSTER 1	-0.072942	-0.000090	-0.784808	

그림 12 Gaussian Mixture 군집의 좌표 특성

군집화 후 각 군집의 중심 좌표 특성을 고려해 모두 동일한 가중치를 부여 SUM의 결과값을 통해 더 큰 값을 가지는 클러스터를 선정. 3가지 클러스터링 모두 CLUSTER 0을 선택함. 각 클러스터에서 3958개, 2781개, 5164개의 지점이 선정됨. 최종적으로 세 클러스터에서 모두 0번 군집으로 분류된 1802개의 CCTV 지점을 도시데이터 센서의 후보지로 선정.

3) 군집화 결과 해석

군집화 결과 해석을 위해 최종선정된 1802개 지점에 대한 평균과 그 외 지점들의 평균을 통해 최종 선택된 군집의 특성을 파악하고자 하였음.

최종 선정된 군집의 평균		나머지 point의 평균	
관측소_거리	4027.481341	관측소_거리	6053.050183
버스정류장_이용객수	893.152497	버스정류장_이용객수	359.443376
지하철역_지표	5998.416198	지하철역_지표	2708.631119
학교_거리	0.003836	학교_거리	0.002571
처리시설_개수	1.175805	처리시설_개수	0.127275
CAI	58.677792	CAI	58.866536

그림 13 군집의 특성별 평균 비교

최종 후보지로 선정된 군집은 비교군보다 버스정류장 이용객수, 지하철역 지표, 폐기물 처리 시설 개수에서상대적으로 높았으며, 대기오염 지수인 CAI 지수에서는 큰 차이를 보이지 않았 다. 관측소 까지의 거리는 상대적으로 가까운 것으로 나타났다.

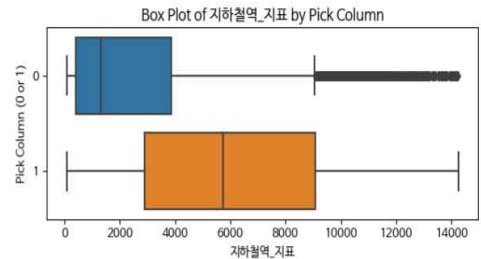


그림 14 지하철역 지표에 대한 군집 분포비교

4) MCLP 알고리즘을 활용한 입지 최적화

MCLP(Maximum Coverage Location Problem) 알고리즘은 최대 커버리지 위치 문제를 해결하기 위해 설계된 알고리즘으로 제한된 수의 위치에 설비(센서, 시설 등)를 배치하여 전체 영역을 최대한 커버하고자 할 때 사용한다. 10000개의 CCTV 대상 후보지에서 군집화를 통해 1802개의 후보지를 선택하였고 해당 후보지들 중에서 최대한 넓은 범위를 커버할 수 있는 후보지를 선정해 제한된 예산안에서 최대한의 효율을 낼 수 있는 센서 후보지를 선정하고자 MCLP 알고리즘을 사용하였음.

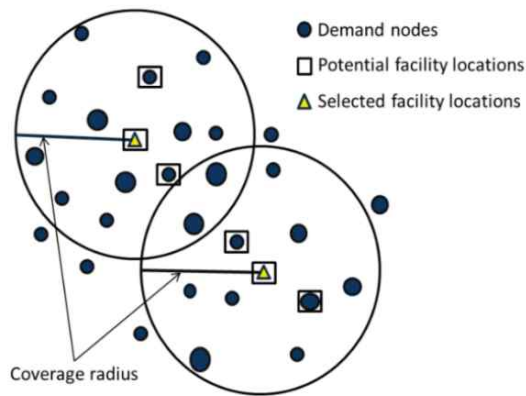


그림 16 MCLP 알고리즘

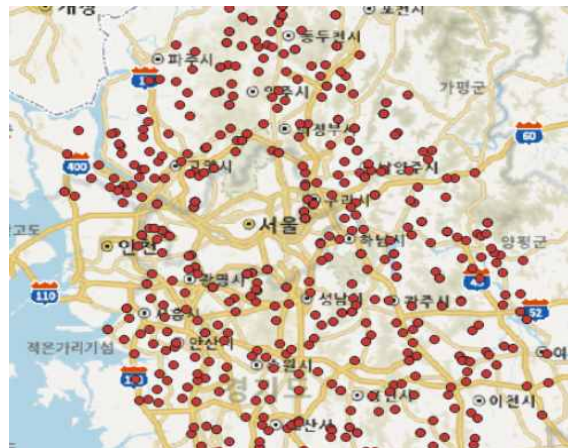


그림 15 MCLP 알고리즘을 통해 선정된 500개 지점

1802개의 후보지에서 최대효율을 내는 500개 지점을 뽑아냄. 이후 500개 지점을 최근접 CCTV 후보지와 매핑해 최종적인 도시데이터 센서 후보 지점을 정함.

V. 결과 해석

최종선정 된 후보 지점들을 프로젝트 계획 단계에서 반영하고자 했던 소음과 유동인구, 건강 취약 계층, 기존 데이터 측정의 불확실성 해소 등의 관점으로 반영된 변수별 상위 n개의 지점을 시각화하여 변수의 특성이 잘 반영되었는지 확인.

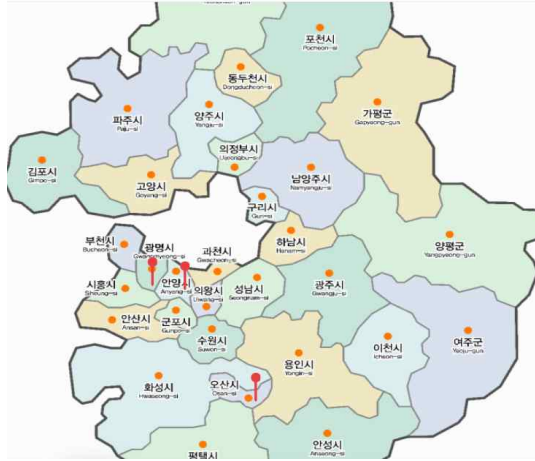


그림 17 소음(버스 승하차 인원, 지하철 지표)기준 상위 3개 지점

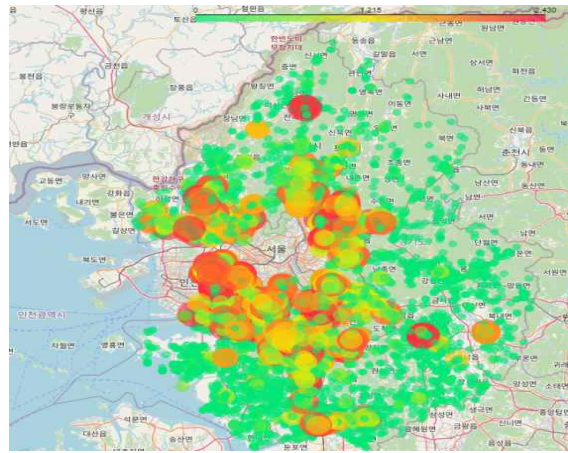


그림 18 지하철역 지표 시각화



그림 19 기존 AWS 데이터 불확실성 기준의 상위 3개 지점

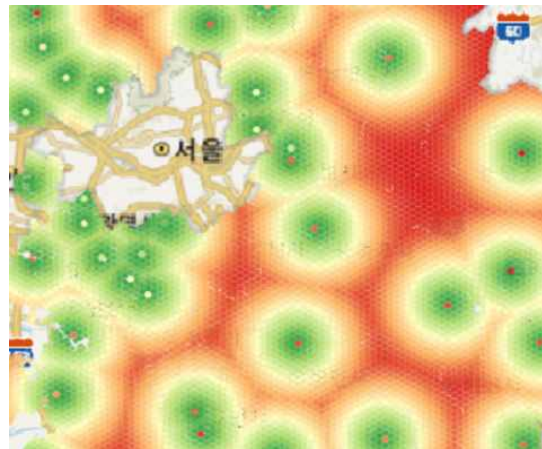


그림 20 AWS- 관측소 거리 지표 시각화

버스 승하차 인원, 지하철 지표 데이터를 통해 분류한 소음 상위 3개 지점이 지하철역 지표 시각화와 일치, 기존의 AWS의 데이터 불확실성을 해소하기 위한 상위 3개 지점이 시각화 자료와 일치함을 확인. 군집의 좌표 특성과 시각화 결과와의 일치성을 고려하였을 때 군집화 후 MCLP 알고리즘을 통해 선정된 후보 지점이 데이터의 특성을 잘 반영하고 있음을 확인.

VI. 활용 방안 및 기대효과

1. 기대 효과

본 프로젝트를 통해 센서를 설치함으로써 결과적으로 미세먼지, 소음 등 다양한 환경 정보를 지역별로 더 세분화하여 제공할 수 있게 되며, 이 서비스의 적용대상은 크게 개인, 공공기관, 기업으로 나눌 수 있다.

시민들에게 도시의 환경정보를 제공함으로써 삶의 질을 향상시킬 수 있고, 건강 취약 계층에

게 미세먼지나 폭염, 한파 등의 정보를 제공해 위협으로부터 사전에 보호할 수 있다.

공공기관에서는 도시 데이터를 활용해서 새로운 정책을 개발하거나 도시 시스템 효율화와 도시 시설 운영 등에 필요한 도움을 받을 수 있다.

그리고 기업에서는 새로운 사업 모델을 발굴 및 제품, 시스템 개발을 통해 더 많은 가치를 창출할 수 있을 것으로 기대한다.

비용적 측면에서는 세계경제포럼에 따르면 국가 디지털화의 10% 증가는 국내총생산을 0.75% 증가시키는 결과를 가져올 수 있다고 발표했다. 또한, 세계 각국의 스마트시티 사례를 보면 스마트시티를 구축함으로써 각종 도시 비용을 30% 이상 절감하고, 생산성을 20% 이상 높이는 효과가 나타났다.

개인	[도시데이터를 통한 삶의 질 향상] 도시의 발전으로 삶의 질 향상과 에너지 비용 절약 가능 . 쾌적한 교통과 편리한 삶을 살 수 있음 .
공공 기관	[도시 설계 및 정책 질적 향상] 도시 데이터를 활용하여 도시 내의 에너지 사용 절약과 탄소 배출량 감축 . 교통 최적화 및 치안 강화 효과 기대 .
기업	[다양한 데이터 확보] 도시 데이터 센서로부터 확보한 다양한 데이터를 통해 기업의 플랫폼 개발과 데이터로부터 새로운 가치 창출을 기대 .

표 1 간접적 기대효과

공공기관 정책 효율화	[도시 설계 및 정책 질적 향상] 도시 데이터를 활용하여 정책의 효율화와 신사업 발굴 , 시설 설치시 입지 선정에 사용할 수 있음.
취약 계층 보호	[건강 취약 계층 보호] 청소년 ,임산부 , 노인통과 같은 건강 취약 계층에 대한 위협을 사전에 감지하고 보호할 수 있다.
도시의 생산성 증가	[도시의 효율화] 교통 , 에너지 사용 , 자원 재활용 등의 효율성 증가로 도시의 효율화를 통해 도시의 생산량 증가를 기대할 수 있다.

표 2 직접적 기대효과

2. 기존의 서비스와 차별점.

기존 국내에서 진행한 U(ubiquitous)-city 프로젝트는 기존에 존재하던 구도시가 아닌 새로 개발되는 신도시에 적용된 프로젝트로 CCTV 카메라 중심의 통합관제 수준으로 일반적인 공공 정보와 감시 수준에 머물렀으며 신도시에 거주하는 인구가 적어 체감도와 활용성이 적었다는 한계점이 존재.

서울시에 설치된 사물인터넷 IOT 센서는 설치 기준이 불분명하고 특정 지역에만 다수 설치되는 문제점이 존재함.

SOGO 도시데이터 유동인구 데이터를 활용하여 시민들의 체감도를 상승시키고자 하였고 다양한 데이터를 활용한 입지선정 과정을 통해 미세먼지 안전구역, 폭염 쉼터 등 다양한 정책 및 도시서비스 개발에 활용할 수 있다. 또한 MCLP 알고리즘을 활용한 입지선정 과정으로 센서가 특정지역에만 설치되는 것이 아닌, 경기도 전체에 넓게 설치된다.

VII. 결론

도시 데이터 센서 입지 선정에 있어서, 대기오염 데이터, 버스와 지하철 승하차 인원, 유동인구 데이터, 그리고 쓰레기 처리시설 데이터를 종합적으로 고려하여 수행하였음. 다양한 도시

요인을 고려함으로써 종합적이고 심층적인 분석을 통해 효과적인 센서 설치 지점을 결정. 후보군을 클러스터링하면서 각 군집의 중심 좌표를 고려하여 대기오염 지수, 유동인구 데이터 등이 높은 군집을 선정하고, 이러한 과정에서 데이터의 다양한 특성을 반영하고자 대기오염 데이터 지수화 등 데이터의 차원을 줄이고 스케일링을 통해 모델의 분류 성능을 향상시키고자 하였다. 클러스터링 결과를 해석함으로써 최종 군집을 선정. 이 과정에서 데이터들의 특성이 잘 반영되었음을 확인하였음.

센서 설치를 통해 도시는 에너지를 효율적으로 사용할 수 있게 되며, 자원 재사용성이 증가하고 교통 정책의 효율화에 기여할 것으로 예상. 이는 도시의 지속 가능한 성장과 삶의 질 향상에 긍정적인 영향을 미칠 것으로 예측되며, 따라서 도시 데이터 센서 설치의 미래를 대비한 전략으로, 지속 가능한 발전과 시민들의 삶의 질 향상을 위한 중요한 요소로 작용할 것으로 여겨짐.

참고문헌

장환영 외 3명, 사업구조적 관점에서 바라본 국내 스마트시티 현황과 개선 이슈(2022), 한국 도시지리학회지

정회훈, 국내 스마트시티 현황과 시사점(2021), KDB미래전략연구소 산업기술리서치센터

이우진, 미국의 스마트시티에 관한 연구개발 동향과 시사점(2018), 한국토지법학회

김기봉 외 2명, 4차 산업혁명시대의 스마트시티 현황과 전망(2018), 한국융합학회논문지

강민영 외 2명, 내일의 도시: 또 한번의 진화를 앞둔 스마트도시(2023), 삼성KPMG 경제연구원