

제7회 교육 공공데이터 분석·활용 대회

지역별 사교육 현황 분석과 학습자원 최적화: SHAP 기반 인사이트와 입지 최적화 모델 적용

SATIMA

활용 데이터 목록

데이터 명	제공기관	출처
학원교습소정보	교육부, 경기도교육청	https://open.neis.go.kr/portal/data/service/selectServicePage.do?page=1&rows=10&sortColumn=&sortDirection=&infId=OPEN19220231012134453534385&infSeq=1
학교기본정보(초)_경기도교육청	한국교육학술정보원	https://www.schoolinfo.go.kr/ng/go/pnnggo_a01_l2.do
학교기본정보(중)_경기도교육청	한국교육학술정보원	https://www.schoolinfo.go.kr/ng/go/pnnggo_a01_l2.do
학교기본정보(고)_경기도교육청	한국교육학술정보원	https://www.schoolinfo.go.kr/ng/go/pnnggo_a01_l2.do
2024년도_학교 현황(초)_경기도교육청	한국교육학술정보원	https://www.schoolinfo.go.kr/ng/go/pnnggo_a01_l2.do
2024년도_학교 현황(중)_경기도교육청	한국교육학술정보원	https://www.schoolinfo.go.kr/ng/go/pnnggo_a01_l2.do
2024년도_학교 현황(고)_경기도교육청	한국교육학술정보원	https://www.schoolinfo.go.kr/ng/go/pnnggo_a01_l2.do
2024년도_성별 학생수(초)_경기도교육청	한국교육학술정보원	https://www.schoolinfo.go.kr/ng/go/pnnggo_a01_l2.do
2024년도_성별 학생수(중)_경기도교육청	한국교육학술정보원	https://www.schoolinfo.go.kr/ng/go/pnnggo_a01_l2.do
2024년도_성별 학생수(고)_경기도교육청	한국교육학술정보원	https://www.schoolinfo.go.kr/ng/go/pnnggo_a01_l2.do
시도 시군구별 졸업자 진학자 진학률	한국교육개발원	https://www.data.go.kr/data/15053808/fileData.do?recommendDataYn=Y#tab-layer-file
지역별(행정동) 성별 연령별 주민등록인구수	행정안전부	https://www.data.go.kr/data/15097972/fileData.do
읍면동 하부행정기관 현황	행정안전부	https://www.data.go.kr/data/15059715/fileData.do
분석시스템_카드매출_성연령별	경기데이터드림	https://data.gg.go.kr/portal/data/service/selectServicePage.do?page=1&sortColumn=&sortDirection=&infId=3BB8GUKK0NIBW955BF2V36585929&infSeq=1&searchWord=카드매출
경기도 공공시설 개방 현황	경기데이터드림	https://data.gg.go.kr/portal/data/service/selectServicePage.do?infId=84E26Y8566L4VKXC5Z5012608120&infSeq=1
경기도_경제활동별_지역내총생산	경기통계	https://stat.gg.go.kr/statgg/kr/dataMng/PublicationForm.html?pub_seq=563
2023년 공공도서관 통계조사 결과표	국가도서관통계	https://www.libsta.go.kr/board/statref/detail/10721

활용 데이터 목록

데이터 명	제공기관	출처
소상공인시장진흥공단_상가(상권)정보	소상공인시장진흥공단	https://www.data.go.kr/data/15083033/fileData.do
전국 시군구 단위 평균 소득	문화 빅데이터 플랫폼	https://www.bigdata-culture.kr/bigdata/user/data_market/detail.do?id=75ab3e79-6f9b-4d80-934b-07746d384096
가구의 월평균 소득별 사교육 참여율	국가통계포털	https://kosis.kr/statHtml/statHtml.do?sso=ok&returnurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHtml%2FstatHtml.do%3Fconn_path%3DDMT_ZTITLE%26list_id%3DH1_10_006%26obj_var_id%3D%26seqNo%3D%26tblId%3DDT_1PE309%26vw_cd%3DDMT_ZTITLE%26itm_id%3D%26language%3Dkor%26lang_mode%3Dko%26orgId%3D101%26
학교급 및 특성별 학생 1인당 사교육 참여시간(주당 평균)	국가통계포털	https://kosis.kr/statHtml/statHtml.do?sso=ok&returnurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHtml%2FstatHtml.do%3Flist_id%3D%26obj_var_id%3D%26seqNo%3D%26docId%3D0223034013%26tblId%3DDT_1PE103%26vw_cd%3DDMT_ZTITLE%26orgId%3D101%26conn_path%3DE1%26markType%3DS%26itm_id%3D%26lang_mode%3Dko%26scrId%3D%26itmNm%3D전국%26
스트레스 정도	국가통계포털	https://kosis.kr/statHtml/statHtml.do?sso=ok&returnurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHtml%2FstatHtml.do%3Fconn_path%3DDMT_ZTITLE%26list_id%3DD2_437001_004%26obj_var_id%3D%26seqNo%3D%26tblId%3DDT_43701N_159%26vw_cd%3DDMT_ZTITLE%26itm_id%3D%26language%3Dkor%26lang_mode%3Dko%26orgId%3D437%26
어제의 행복 정도	국가통계포털	https://kosis.kr/statHtml/statHtml.do?sso=ok&returnurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHtml%2FstatHtml.do%3Fconn_path%3DDMT_ZTITLE%26list_id%3D210_21004_005%26obj_var_id%3D%26seqNo%3D%26tblId%3DDT_20114_2024026_3%26vw_cd%3DDMT_ZTITLE%26itm_id%3D%26language%3Dkor%26lang_mode%3Dko%26orgId%3D210%26
시군구별 스트레스 인지율	국가통계포털	https://kosis.kr/statHtml/statHtml.do?sso=ok&returnurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHtml%2FstatHtml.do%3Fconn_path%3DDMT_ZTITLE%26list_id%3DF_001_001_007%26obj_var_id%3D%26seqNo%3D%26tblId%3DDT_H_MENTAL_STRESS%26vw_cd%3DDMT_ZTITLE%26itm_id%3D%26language%3Dkor%26lang_mode%3Dko%26orgId%3D177%26
교원 1인당 학생수 시도 시군구	국가통계포털	https://kosis.kr/statHtml/statHtml.do?sso=ok&returnurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHtml%2FstatHtml.do%3Fconn_path%3DDMT_GTITLE01%26list_id%3D104%26obj_var_id%3D%26seqNo%3D%26tblId%3DDT_1YL21171%26vw_cd%3DDMT_GTITLE01%26itm_id%3D%26language%3Dkor%26lang_mode%3Dko%26orgId%3D101%26
늘봄학교 및 방과후학교 프로그램 학생 1인당 참여시간 주당 평균	국가통계포털	https://kosis.kr/statHtml/statHtml.do?sso=ok&returnurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHtml%2FstatHtml.do%3Fconn_path%3DDMT_ZTITLE%26list_id%3DH1_10_008%26obj_var_id%3D%26seqNo%3D%26tblId%3DDT_1PE510%26vw_cd%3DDMT_ZTITLE%26itm_id%3D%26language%3Dkor%26lang_mode%3Dko%26orgId%3D101%26

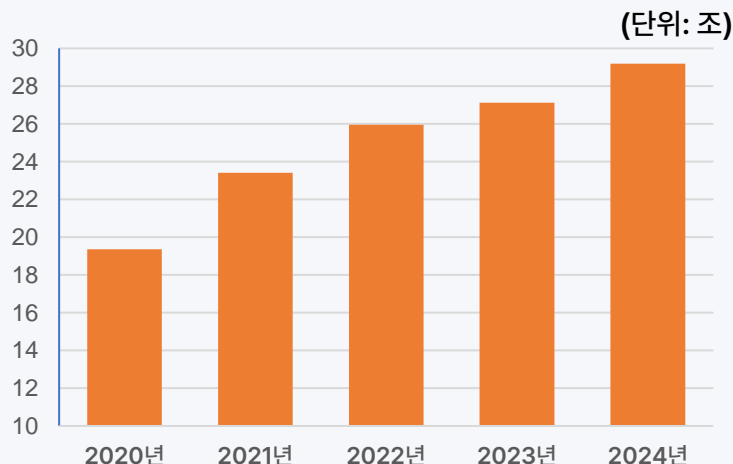
목차

- 01 문제 상황**
문제 상황 및 선행연구 / 분석 진행 과정
- 02 데이터 수집 / 전처리**
데이터 수집 및 전처리 / EDA
- 03 군집화 및 SHAP**
PCA / PCA 후 군집화 / SHAP 분석
/ 군집화 해석 및 공교육 인프라 필요성
- 04 입지 최적화**
자기주도학습센터 신규 입지 최적화
/ P-Median과 MCLP / 입지 최적화 적용 결과
- 05 연속적 시스템 및 시사점**
연속적 시스템 구축 / 시사점 및 활용방안

문제 상황 및 선행 연구

□ 사교육 현황

교육부와 통계청의 「초중고사교육비조사」에 따르면 초중고교생의 사교육비가 4년 연속 역대 최대치를 갱신하였다. 2020년 19조 3,532억 원에서 2024년 29조 1,919억 원으로 4년동안 약 51% 증가하였다. 동 기간 학생 수는 오히려 줄어들어, 1인당 월평균 사교육비 부담은 47만 4000원으로 가중되고 있다.



□ 자기주도학습센터란?

사교육 현황에 대한 교육부의 대책으로, 학교 안팎 학습환경을 개선하고 사교육 부담 경감을 위하여 자기주도학습센터 사업을 추진한다.

이 시설은 학교 밖 학습 여건이 열악한 지역 등의 중고등학생을 대상으로 스스로 공부할 수 있는 공간과 프로그램을 제공한다.

□ 사업 시행 계획

2025년에 50개 내외의 시설을 선정하겠다고 공고하였고, 지역별 여건에 따라 세 가지 유형으로 구분한다.

유형 1: 교육부 사업 연계형 → 교육발전특구, RISE 등과 연계

유형 2: 지역 주도형 → 지역별로 시설 자체 확보

(예: 도서관, 문화회관, 행정복지센터 등)

유형 3: 사업 재구조화형 → 지역별 기 운영 중인 시설에 지원

□ 선행 연구

이지원 외(2019)는 공공데이터를 활용하여 초등학생 돌봄시설의 최적 입지를 P-Median을 활용하여 선정하였다.

조성아와 김성연(2021)은 서울시의 특성화 고등학교 공동실습소 입지를 P-Median과 MCLP방법을 통해 수행하여 분석하였다.

김정근(2020)은 정부의 사교육 정책 방향 및 인식 변화가 필요하고 방과후 학교 프로그램을 사교육 수요 흡수에서 공교육 보완 수단으로 활용해 나가야 하며, 교육 수요자의 선택권을 늘리는 교육 정책을 제안하였다.

심재권(2025)는 클러스터링과 SHAP를 통해 대구시 학업 성취도에 대해 분석하여 학업부진을 예측하는 모델을 연구하였다.

(붉은 색은 본 프로젝트와 관련된 내용)

늘어나는 사교육비 부담 경감과, 공교육 인프라 및 여건 문제는 개선이 시급하다.

여러 지역 중 교육발전특구가 있고, 학생 수와 학구열이 높은 경기도를 본 프로젝트의 분석 대상으로 선정하였다.

이에 지역별 교육 관련 특성을 데이터로 군집화하고, SHAP 기법으로 각 군집의 교육 특성을 명확히 규명한다.

그 후, 학습 자원이 필요한 지역을 선정하여 P-Median, MCLP 방법으로 자기주도학습센터와 같은 공적 학습 자원을 최적화하여 배치하고, 추후에도 성과와 실적을 지속적으로 모니터링하고 정보를 제공할 수 있는 시스템을 개발하고자 한다.

분석 진행 과정

데이터 수집 및 전처리

사교육 관련 데이터

소득기준 사교육 참여율
학생 1인당 학원 카드매출액
입시 검정 보습학원 수
⋮

공교육 관련 데이터

교원 1인당 학생수
공립 학교 수
공공 도서관 수
⋮

학생 관련 데이터

스트레스 인지율
학교생활 스트레스 정도
대학 진학율
⋮

목적에 맞는 데이터 수집 및 전처리

다양한 관점의 데이터를
수집하여 가공, 선별, 전처리를 진행

군집화 및 SHAP 분석

PCA 차원축소

K-means 군집화

XGBoost 기반 지도학습으로 변환

군집 별 SHAP 분석

지역별 분석 및 문제점 도출

현황 시각화

기존 해석을 넘어서는 SHAP 분석

단일 지표, 모델 기반 해석이 아닌
군집화를 통해 교육에 대한 지역별 특성을
유형화한 뒤, 각 유형별 SHAP 분석을
통해 개별 지역의 교육 특성에 영향을 미치는
요인과 특성을 심층 분석

지역 유형별 최적화



수요 집약 유형

MCLP

비용 우선 유형

P-median

지역 특성에 맞는 학습자원 최적화

지역 특성 맞춤형 입지 최적화
군집화와 SHAP 분석 결과를 기반으로,
교육부 현안인 자기주도학습센터를 예시로
각 지역의 유형과 특성을 고려하여
맞춤화된 입지 최적화 알고리즘 적용
이를 통한 합리적 학습자원 최적화 실현

지속 가능한 인사이트

데이터 파이프라인 구축

군집화 및
SHAP 분석

교육 유형별
최적화

교육 특성 현황
시각화 리더보드

지역별 학습 자원
배치 제안 시스템

지속적인 의사결정 지원
시스템 구축

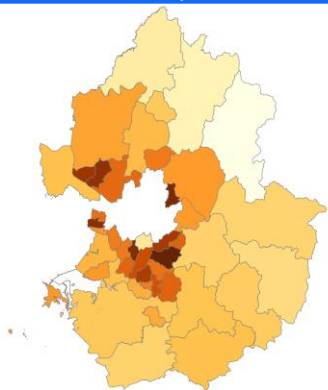
앞선 분석을 통합한 시스템
SHAP 분석과 최적화 알고리즘 기반의
리더보드와 시설 배치 시스템 구축,
이를 통해 실무자의 의사결정을 위한
지속가능 시스템 제안

EDA (탐색적 데이터 분석)

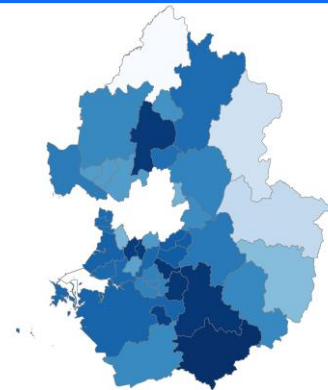
전처리를 거쳐 도출된 19개의 변수를 시각화하고 인사이트를 도출하였으며
아래는 사교육, 공교육, 학생 특성의 관점에서 지역 간 교육 현황을 대표적으로 보여주는 6가지 변수이다.

학원교육 총매출액과 학습 공간인 독서실, 스터디 카페 수를 보았을 때, 경기도의 남부와 북부의 교육 격차를 확인할 수 있다. 사교육과 학습 공간인 프라가 경기 남부, 특히 서울 인근 도시에 집중되어 있다.

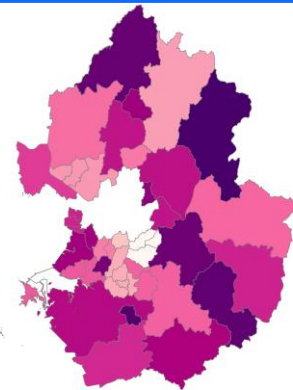
학생1000명당 독서실, 스터디 카페 수



교원1인당 학생수



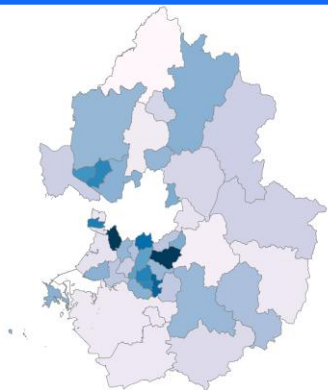
진학률



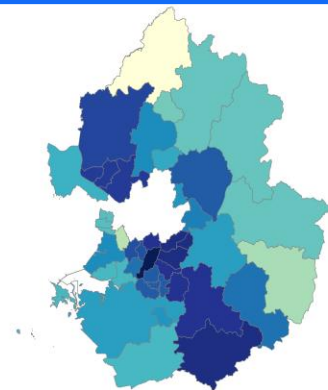
사교육이 심한 도시의 경우 진학율이 낮는데, 이는 역설적으로 높은 학구열과 사교육으로 인해 대학입시를 여러 번 도전하는 경향 때문이다. 스트레스도 마찬가지로 사교육이 심한 지역인 경기 남부가 더 높은 경향을 보인다.

교원 1인당 학생 수와 고교생당 공립학교 수를 통해, 공교육 자원의 경우 경기 남부보다 북부가 풍부하다고 보여진다. 이는 공급되는 공교육 자원의 양의 차이 뿐 아니라 학생 수에 따른 공교육 집약도 차이를 보이는 자료이다.

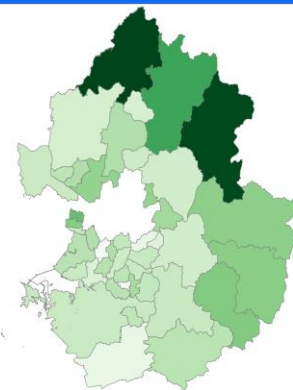
학생1인당 학원교육총매출액



스트레스_많음 이상



고교생1000명당 공립 고등학교 수

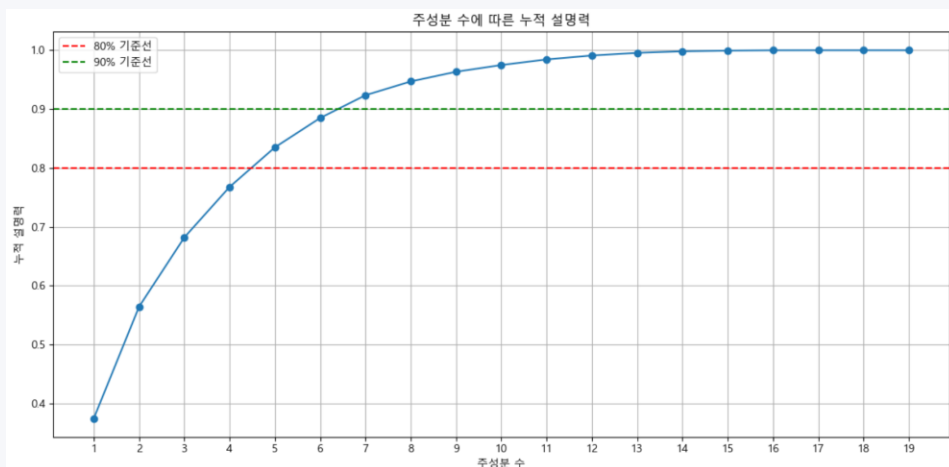


하지만 이러한 해석들은 지도를 통해 간단히 살펴본 경향성일 뿐, 경기 남부와 북부를 넘어 각 시군구별 본질적 특질은 파악할 수 없다. 따라서 본 프로젝트에서는 경기도 시군구별 교육 특성과 그 차이를 더욱 고도화된 분석으로 살펴보고자 한다.

PCA (주성분 분석)

- 군집화에 앞선 차원 축소와 군집 수 설정에서 고려된 핵심 아이디어는 아래와 같다.
1. 원본 데이터의 구조를 최대한 유지하되 복잡하지 않은, 강건한 차원 수를 선정한다.
 2. 군집화가 지역별 유형 분류에 사용될 점을 고려하여, k가 너무 크지 않아야 한다.

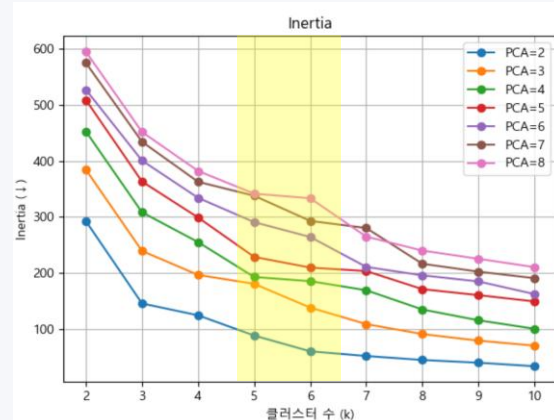
1. 주성분 수 선택



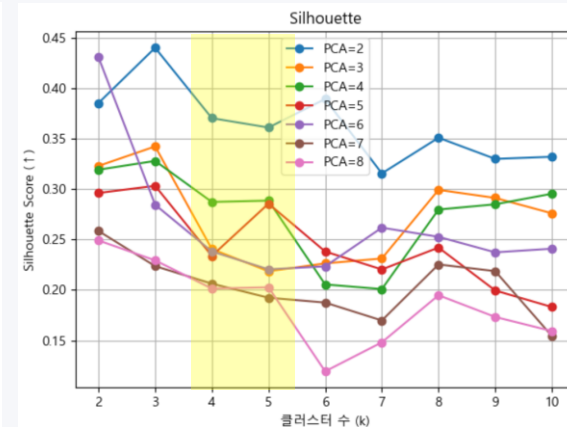
주성분 수에 따른 누적 설명력

교육 관련 변수들은 다차원적이고 서로 상관관계가 있을 수 있기 때문에, 기존의 차원으로 바로 군집화를 시행할 경우 군집 결과가 왜곡될 우려가 있다. 따라서 PCA를 통해 변수 간 중복 정보를 줄여 주요한 설명력을 유지하고 노이즈를 줄여 보다 뚜렷한 군집 구조를 파악하기 위해서, 적절한 주성분 수(차원 수)까지 차원을 축소할 필요가 있다. 이 과정에서 84.42%의 누적 설명력을 유지하는 5개의 주성분(5차원)을 우선적으로 선택하였다.

2. 군집 수 선택



Inertia (by Elbow Method)



Silhouette

5차원일 때 (PCA=5)와 인근 차원들에 따른 Inertia와 Silhouette Score는 위와 같다. Elbow 포인트는, 5차원의 경우 5, 인근 차원들 역시 5-6의 값을 가진다. 이를 통해 군집(유형)을 k=5 인근으로 설정하는 것이 타당하다고 판단하였다. 추가적으로 실행한 Silhouette의 경우에도, k=5일때 5차원이 대부분의 다른 차원들 보다 높은 점수를 가진다. 앞선 단계에서 진행한 누적 설명력을 고려하여, 80%이하의 설명력을 가지는 차원들을 고려하지 않았다. 최종적으로 5차원으로 차원을 축소하고, 5개의 군집을 가지도록 군집화를 수행하는 것이 적합하다고 판단하였다.

PCA 후 군집화

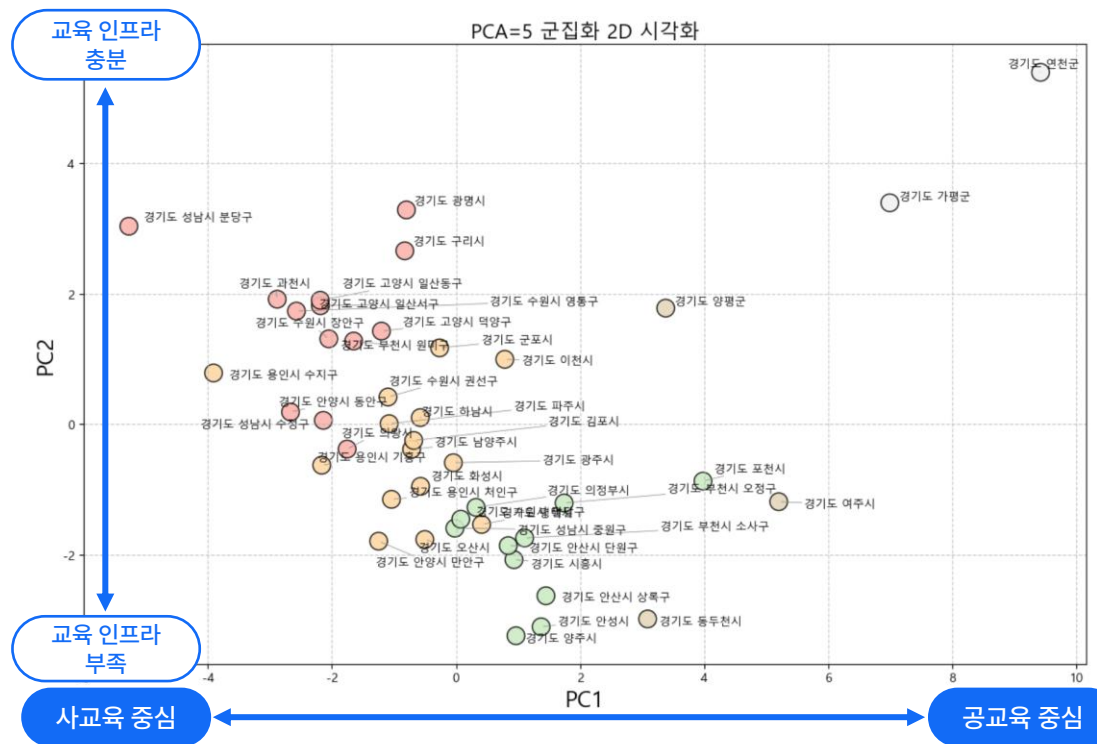
차원별 기여도

PC 1 사교육 ↔ 공교육			PC 2 교육 인프라 제공 정도		
변수명	PC1 기여도		변수명	PC2 기여도	
0	총초등생1000명당_공립_초등학교수	0.326531	0	재적학생100명당_실질_교육서비스업_총생산_22	0.408708
1	총중등생1000명당_공립_중학교수	0.264377	1	재적학생100명당교원수	0.349290
2	재적학생100명당교원수	0.255496	2	총고교생1000명당_공립_고등학교수	0.292874
3	총고교생1000명당_공립_고등학교수	0.245991	3	총학생1인당학원카드매출액_2312	0.244311
4	진학률	0.194446	4	총학생1000명당_입시검정보습학원수	0.241425
5	재적학생100명당_실질_교육서비스업_총생산_22	0.169250	5	소득기준_사교육참여시간	0.237106
6	총중등생1000명당_사립_중학교수	0.157296	6	소득기준_사교육참여율	0.236969
7	총고교생1000명당_사립_고등학교수	0.111289	7	소득기준_일반교과_사교육참여율	0.236626
8	스트레스인지율_24	0.071468	8	총중등생1000명당_공립_중학교수	0.233434
9	총학생1인당교육기관카드매출액_2312	0.055414	9	총학생1인당_학원교육총매출액_2312	0.225250
10	총학생1인당_학원교육총매출액_2312	-0.207960	10	총초등생1000명당_공립_초등학교수	0.156680
11	교원일인당학생수	-0.222522	11	총학생1000명당_독서_스카수	0.103246
12	총학생1인당학원카드매출액_2312	-0.236150	12	진학률	-0.007667
13	총학생1000명당_독서_스카수	-0.261603	13	총학생1인당교육기관카드매출액_2312	-0.034462
14	총학생1000명당_입시검정보습학원수	-0.262101	14	학교생활_스트레스_많음이상_24	-0.062841
15	소득기준_사교육참여시간	-0.270276	15	총중등생1000명당_사립_중학교수	-0.128746
16	소득기준_일반교과_사교육참여율	-0.270498	16	총고교생1000명당_사립_고등학교수	-0.142096
17	소득기준_사교육참여율	-0.270540	17	스트레스인지율_24	-0.181192
18	학교생활_스트레스_많음이상_24	-0.284357	18	교원일인당학생수	-0.351581

위의 절차를 따라 5차원을 통해 5개의 군집으로 군집화를 시행하였지만, 실무적으로 이를 시각화 하기 위하여 2차원(PC1, PC2)에 결과를 투영하였다.

이를 위하여 2차원 축에 대한 기여도를 분석한 결과, 일관된 경향성을 통해 각 PC에 대한 정의를 내릴 수 있었다. **PC1**은 양의 방향일수록 공교육 위주, 음의 방향일수록 사교육 위주의 특성이 많아 해당 지역의 **교육 제공의 주체**에 대한 정보를 담고 있다. **PC2**는 양의 방향일수록 교육 인프라나 매출과 관련된 특성이 크고, 음의 방향일수록 학생들이 겪는 스트레스의 정도나 교원 한 명이 맡는 학생 수가 많은 등 **교육 인프라의 제공의 양과 수준**에 대한 정보를 담고 있다고 해석할 수 있다.

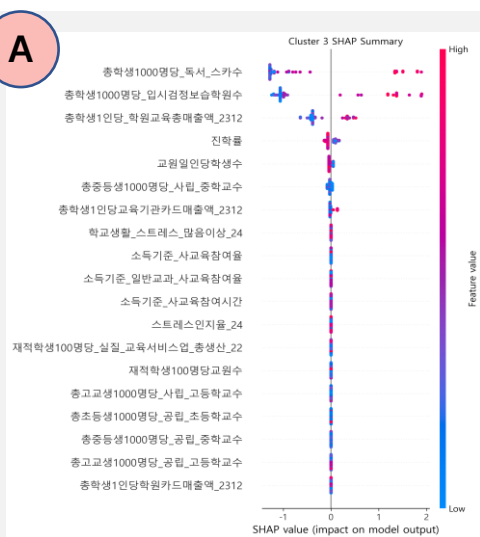
따라서 이후 분석에서는 해당 기준을 통해 해석을 진행할 것이다.



SHAP 분석

군집화를 완료한 후, 각 지역의 어떤 요인이 해당 군집으로 분류 되도록 하는 요인인지를 SHAP 분석을 통해 산출하였다. 비지도 학습 환경을 XGBoost를 활용하여 지도학습 환경으로 변환한 후, 각 클러스터링 레이블에 대한 SHAP를 적용하였다. 각 점들이 좌우로 많이 벌어질수록, 그 데이터는 해당 군집에 속하게 되는 주요 요인이라는 것이며, 빨간색이면 높을수록, 파란색은 낮을 수록 더 속할 확률을 높인다는 의미이다.

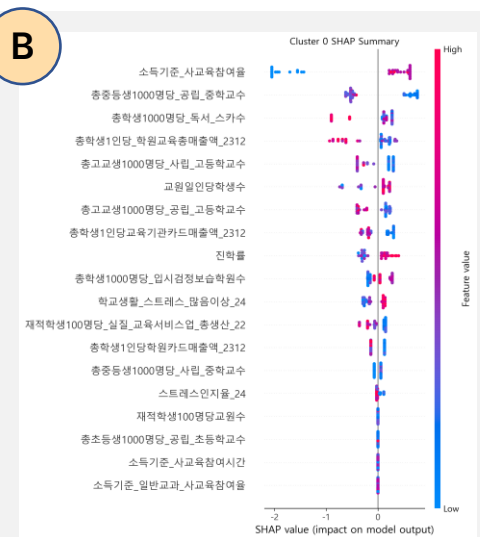
A



사교육 중심 고투자형

학원 매출액이 높거나 학원 수가 많은 수록 해당 군집에 속할 확률이 높다. 학교 생활 중 겪는 스트레스 또한 높은 것으로 나타났다. SHAP 분석만 보더라도 전형적인 학구열이 높은 지역임을 유추할 수 있다.

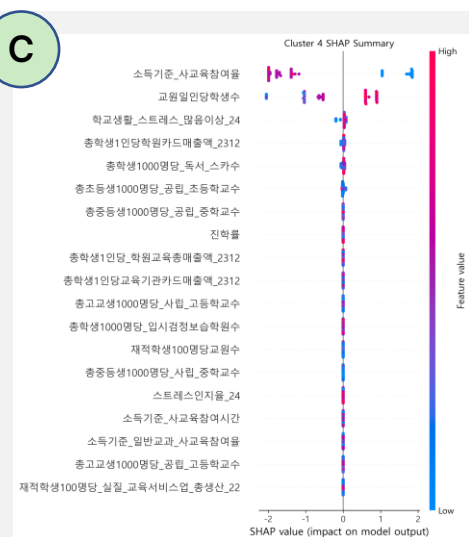
B



사교육 중심 저투자형

사교육 참여율이 높지만, 1인당 학원의 총 매출액은 낮은 지역이다. 사교육 자체에 대한 관심은 높은 지역으로 보이나, 그 투자는 아직까지는 그리 높지 않다.

C



공교육 중심 저투자형

PC1 축에서 양수 방향으로 치우쳐져 있던 것을 고려하면 사교육보다는 공교육 중심의 교육이 이루어지고 있다. 그러나 교원 한 명당 맡는 학생 수가 높은 것으로 나타나, 교육 관련된 투자가 전반적으로 낮다고 판단된다.

D



교육 인프라 개선형

경기도에 존재하는 군 지역 3곳이 모두 포함되어 있으며, 지리적으로는 경기 외곽 지역들이다.

단, 분석 결과 자체는 공교육 자원이 충분한 것으로 나오는데, 이는 공교육의 역할이 충분해서가 아니라, 군집화에 사용한 데이터가 단위 학생 당을 기준으로 하였기에 학생 수가 적어 상대적으로 커보이는 착시이다. 전반적인 투자 자체가 적은 인프라 개선이 필요한 지역들이다.

E



군집화 해석 및 공교육 인프라 필요성

사교육 중심 고투자형 (13)

성남시 분당구/수정구, 고양시 일산동구/일산서구/덕양구, 수원시 장안구/영통구, 부천시 원미구, 안양시 동안구, 과천시, 광명시, 구리시, 의왕시

학구열이 높고 학원가가 밀집한 지역으로 평가되는 1기 신도시가 포진되어 있다. 일산, 영통, 평촌, 분당으로 대표되는 지역들이 이 군집에 속한다.

또한 사교육과 공교육 모두 적절히 기능하고 있어 자기주도학습센터를 입지시켜 오랜 기간 이 지역을 이끌어온 사교육과 경쟁하여 대체하려 하기보다는, 시급한 지역에 우선 배치하는 것이 적절할 것이라고 정무적으로 판단했다.

사교육 중심 저투자형 (15)

용인시 처인구/기흥구/수지구, 안양시 만안구, 군포시, 이천시, 파주시, 수원시 권선구, 하남시, 김포시, 광주시, 남양주시, 화성시, 평택시, 오산시

반도체 클러스터 주변인 처인, 기흥, 이천, 화성, 평택, 오산과 최근에 개발되어 젊은 부부와 아이가 늘어나는 김포, 하남, 남양주 등이 속해 있다. 사교육에는 관심이 많지만 사교육이 아직 성숙하게 발달하지는 못한 특징을 보이고 있다.

신도시의 이점을 살려 자기주도학습센터가 빠르게 입지를 선점하여 사교육과 경쟁하기 좋은 구도이다.

공교육 중심 저투자형 (11)

부천시 오정구/소사구, 수원시 팔달구, 안산시 단원구/상록구, 성남시 중원구, 포천시, 의정부시, 시흥시, 안성시, 양주시

사교육이 그리 발달하지 않은 도시들로 구성되어 있다. 다른 지역에 비해 학구열이나 투자가 특별히 높지도 않다. D, E 군집보다는 나은 상황이나, A, B 군집보다는 열악한 환경이라고 판단했다. 이 지역에도 자기주도학습센터를 도입하여 학습 환경을 개선하고 학생들의 열의를 높이는데 공교육이 역할을 할 수 있을 것이다.

인프라 개선형 (2)

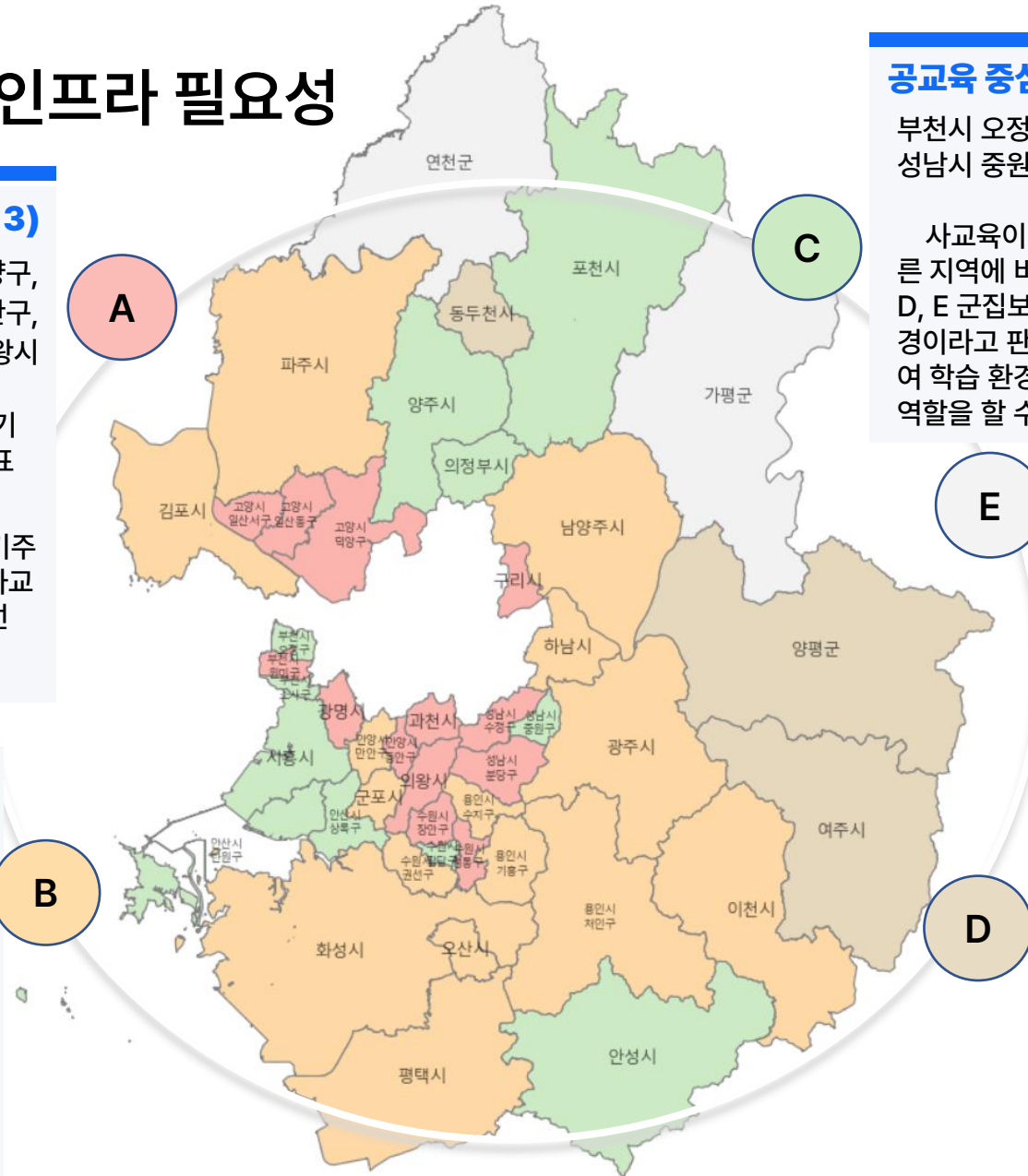
가평군, 연천군

경기 외곽 지역으로, 공교육 및 사교육 인프라가 모두 제한되는 지역이다. 경기도 3개 군 지역 중 2개가 이 군집에 속한다. 자기주도학습센터를 입지시켜 지역 교육 관련 인프라를 개선해야 한다.

인프라 개선형 (3)

여주시, 동두천시, 양평군

마찬가지로 경기 외곽 지역으로, 공교육 및 사교육 인프라가 모두 제한되는 지역이다. 자기주도학습센터를 입지시켜 지역 교육 인프라를 개선해야 한다.



자기주도학습센터 신규 입지 최적화

※ (참고) 교육발전특구 (경기도 7개 지역): 동두천시, 양주시, 고양시, 파주시, 김포시, 포천시, 연천군

군집 A는 이미 사교육과 공교육 모두 충분한 역할을 하고 있기 때문에 제외하였으며, 군집 B, C, D&E에 속한 지역들 중, 교육발전특구로 지정된 지역을 대상으로 각각 적합한 입지 선정 모델을 적용하고자 한다.

최적화 모델

최적화 방식

분석 지역

Capacitated P-Median

Capacitated P-Median은 시설의 용량 제약을 고려하여, 수요지까지의 거리 총합 (수송비용)을 최소화하는 것을 목표로 하는 위치 선정 모형이다.

C

D

E

군집 C는 학구열과 교육에 대한 투자가 다른 지역에 비해 두드러지지 않는 지역으로, 교육 수요가 광범위하게 퍼져 있는 특성을 지닌다. 군집 D와 E는 인구 밀도가 낮아 교육 시설과 수요지가 넓고 드물게 분산되어 있다.

이처럼 세 군집은 **물리적 접근성 확보**가 핵심 과제이므로, 수요지와의 거리 총합을 최소화하는 P-Median 모델을 적용하는 것이 타당하다.

“포천시” “동두천시”

Capacitated MCLP (Maximal Covering Location Problem)

Capacitated MCLP는 시설의 용량 제약을 고려하여, 주어진 거리 범위 내에서 충족하는 수요량을 최대화하는 것을 목표로 하는 위치 선정 모형이다.

B

군집 B는 교육에 대한 관심과 수요는 높지만, 이를 뒷받침할 수 있는 교육 인프라가 아직 성숙할 정도로 발달하지 못한 지역이다.

이러한 환경에서는 **가능한 한 많은 수요**를 충족시키는 것이 우선 과제로 부각된다. 따라서 해당 지역에는 일정 반경 내에서 서비스를 제공할 수 있는 수요를 최대화하는 MCLP 모델을 적용하는 것이 적절하다.

“김포시”

해당 프로젝트에서는 현실의 환경을 반영하기 위해 기존의 Capacitated P-Median과 Capacitated MCLP에 여러 제약조건을 추가하여 후보지를 한 곳씩 탐욕적(greedy)으로 확정하는 휴리스틱 기반 반복적 탐색 알고리즘을 구축했다.

P-Median과 MCLP

최적화 방법론

최적화 방식

공식

파라미터 및 결정변수

Capacitated P-Median

1. 각 후보지는 평균 통학거리 이내에 위치하고, 수요가 남은 학교를 거리 순으로 배정해 후보지 시설 용량을 채운다.
2. 배정된 학생 수가 후보지 시설 용량의 80% 미만이면 해당 후보지를 제외한다.
3. '배정된 수요지의 학생 수×수요지까지의 거리'의 합이 최소인 후보지를 최적 시설로 선정하고, 배정된 수요를 차감하여 다음 단계를 진행한다.
4. 모든 수요가 소진되거나 활용률 미달 후보지만 남으면 알고리즘을 종료한다.

Capacitated MCLP (Maximal Covering Location Problem)

1. 각 후보지가 커버할 수 있는 반경에 위치하면서 수요가 남은 학교를 거리 순으로 배정해 후보지 시설 용량을 채운다.
2. 배정된 학생 수가 후보지 시설 용량의 80%미만이면 해당 후보지를 제외한다.
3. '배정된 수요'의 합이 최대인 후보지를 최적 시설로 선정한다. 동률일 경우, 사전에 정의해둔 우선순위 기준으로 선정하고, 배정된 수요를 차감하여 다음 단계를 진행한다.
4. 모든 수요가 소진되거나 활용률 미달 후보지만 남으면 알고리즘을 종료한다.

$$\begin{aligned}
 \min_{x_{ij}, y_j} \quad & \sum_{j \in F} \sum_{i \in D: d_{ij} \leq D_{\max}} d_{ij} x_{ij} \\
 \text{s.t.} \quad & \sum_{j \in F: d_{ij} \leq D_{\max}} x_{ij} \leq w_i \quad \forall i \in D \\
 & \sum_{i \in D: d_{ij} \leq D_{\max}} x_{ij} \leq C_j y_j \quad \forall j \in F \\
 & \sum_{i \in D: d_{ij} \leq D_{\max}} x_{ij} \geq \alpha C_j y_j \quad \forall j \in F \\
 & \sum_{j \in F} y_j = 1 \\
 & y_j \in \{0, 1\}, \quad x_{ij} \geq 0 \quad \forall i \in D, j \in F
 \end{aligned}$$

$$\begin{aligned}
 \max_{x_{ij}, y_j} \quad & \sum_{j \in F} \sum_{i \in D: d_{ij} \leq R_{\max}} x_{ij} \\
 \text{s.t.} \quad & \sum_{j \in F: d_{ij} \leq R_{\max}} x_{ij} \leq w_i \quad \forall i \in D \\
 & \sum_{i \in D: d_{ij} \leq R_{\max}} x_{ij} \leq C_j y_j \quad \forall j \in F \\
 & \sum_{i \in D: d_{ij} \leq R_{\max}} x_{ij} \geq \alpha C_j y_j \quad \forall j \in F \\
 & \sum_{j \in F} y_j = 1 \\
 & y_j \in \{0, 1\}, \quad x_{ij} \geq 0 \quad \forall i \in D, j \in F
 \end{aligned}$$

인덱스

D 수요지(학교) 집합

F 후보지(시설) 집합

후보지는 자기주도학습센터 후보 유형2를 참고하여 공공시설(도서관, 행정복지센터 등) 및 지역 중고등학교로 선정하였다.

파라미터

d_i 학교 i와 시설 j 간의 거리

w_i 학교 i의 수요량(학생 수)

C_j 시설 j의 용량(최대 수용 인원)

α 개설된 시설의 최소 활용률 80%

D_{\max} 최대 배정 거리 17308m

R_{\max} 최대 커버 거리 5000m

학교별 수요량은 학교 간 형평성을 고려하기 위해 시군별 공급 예정 인원 수를 학교별 전체 학생 수로 나누었다. 최대 배정 거리는 평균 버스 통학시간에 평균 버스 속도를 곱하여 산출했다.

결정변수

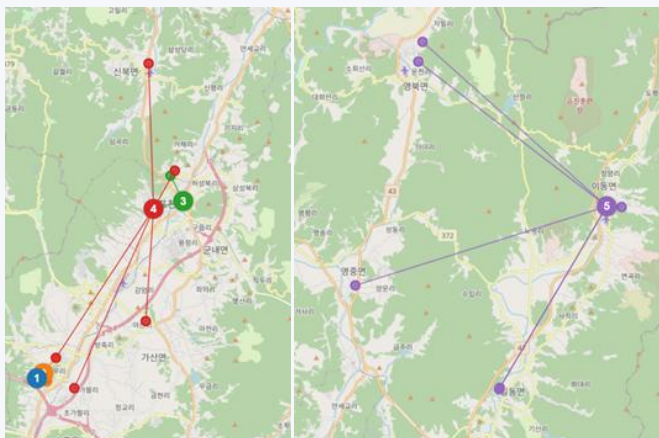
y_j 시설 j의 선정 여부 (선정=1, 미선정=0)

x_{ij} 학교 i의 학생 중 시설 j로 배정되는 인원

입지 최적화 적용 결과

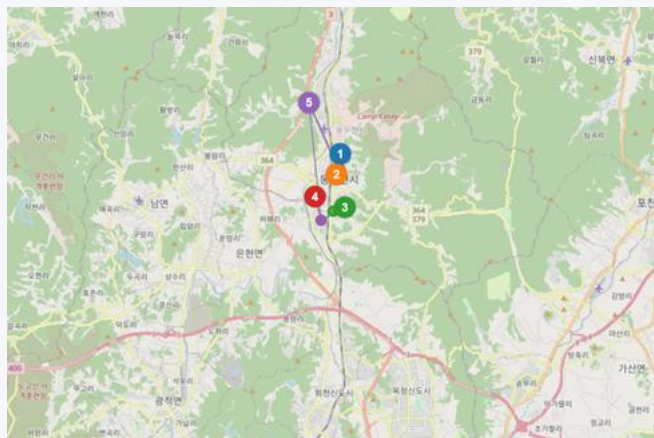
Capacitated P-Median

포천시



	시설	학교
1	동남 고등학교	동남고등학교(33명), 동남중학교(18명), 대경중학교(9명)
2	송우 고등학교	송우고등학교(35명), 대경중학교(8명), 송우중학교(17명)
3	포천일 고등학교	포천일고등학교(17명), 포천여자중학교(19명), 포천고등학교(12명)
4	포천 고등학교	포천고등학교(18명), 포천중학교(22명), 경북중학교(3명), 삼성중학교(1명), 송우중학교(9명), 갈월중학교(7명)
5	이동면 행정복지 센터	이동중학교(4명), 일동중학교(14명), 일동고등학교(14명), 영북고등학교(8명), 영중중학교(4명), 영북중학교(5명)

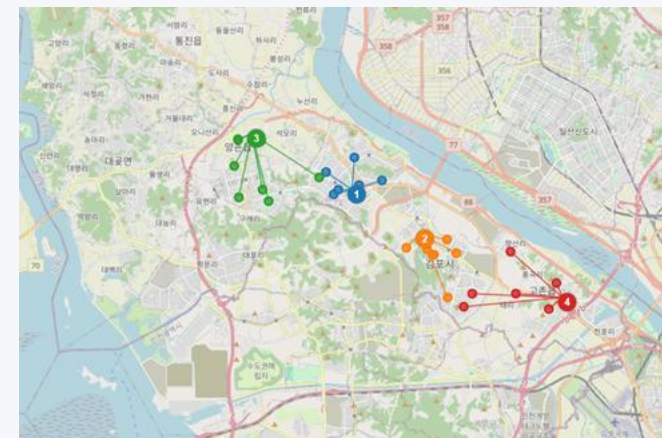
동두천시



	시설	학교
1	한빛누리 고등학교	한빛누리고등학교(32명), 한빛누리중학교(23명), 동두천여자중학교(5명)
2	동두천 고등학교	동두천중학교(16명), 동두천고등학교(24명), 한국문화영상고등학교(20명)
3	동두천외국어 고등학교	동두천외국어고등학교(30명), 동두천중앙고등학교(29명), 송내중학교(1명)
4	생연 중학교	생연중학교(34명), 송내중앙중학교(26명)
5	신흥 중학교	신흥중학교(20명), 신흥고등학교(19명), 동두천여자중학교(15명), 한국문화영상고등학교(4명), 송내중앙중학교(1명)

Capacitated MCLP

김포시



	시설	학교
1	김포시 장기도서관	운양중학교(12명), 고창중학교(10명), 운양고등학교(11명), 김포제일고등학교(12명), 장기중학교(13명), 장기고등학교(2명)
2	김포시 중봉도서관	김포중학교(3명), 김포과학기술고등학교(5명), 감정중학교(8명), 김포여자중학교(2명), 김포고등학교(11명), 금파중학교(11명), 사우고등학교(12명), 풍무고등학교(8명)
3	김포시 양곡도서관	양곡고등학교(8명), 양곡중학교(4명), 신양중학교(3명), 김포한가람중학교(9명), 솔터고등학교(10명), 나래중학교(4명), 은여울중학교(10명), 운유고등학교(5명)
4	김포시 고촌도서관	고촌중학교(5명), 김포신곡중학교(9명), 고촌고등학교(10명), 향산중학교(4명), 양도중학교(8명), 풍무중학교(12명)

입지 최적화 결과는 다음 지도에 나타나 있다. 포천시는 두 후보지가 서로 멀리 떨어져 있어, 지도를 2개로 나누어 게재하였다. 지도에는 시설이 들어설 위치와 각 위치의 입지 (자기주도학습센터 설치 우선순위)를 함께 표기하였다. 또한, 표에서 [학교]는 각 후보지에 할당된 학교와 학생 수를 나타내었다.

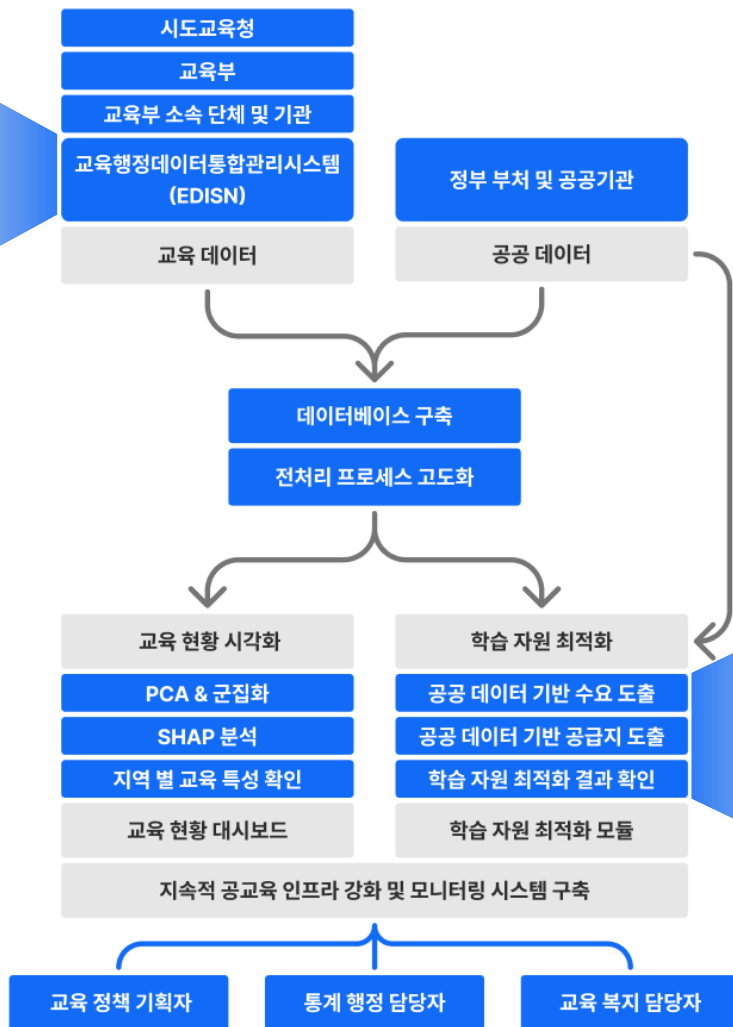
연속적 시스템 구축 - 파이프라인

지금까지 진행한 군집화 및 입지 최적화 방법을 자동화하여 지속적이고 연속적으로 분석이 가능한 시스템을 구축하기 위한 파이프라인이다.

산발적으로 분포되어 있는 교육 데이터들을 통합, 관리하는 시스템을 활용하여 더욱 고 품질의 데이터를 지속적으로 받아와 데이터 파이프라인을 안정화 한다.
이를 위하여 향후 구축될 교육부의 EDISN을 사용하는 방안을 제안한다.

1. 범용성

본 파이프라인을 통하여 양질의 데이터를 기반으로, 발전 된 지역 유형화와 학습 자원 최적화를 이루어 낼 수 있다. 이는 경기도의 지역별 교육 특성을 확인하는 것을 넘어, 데이터 파이프라인 구축이 가능한 다른 지역에도 적용이 가능하다. 또한 학습 자원 최적화의 경우 현안인 자기주 도학습센터 외에도 공교육 인프라를 강화할 수 있는 학습 자원을 분배하는 문제에 적용이 가능하다.

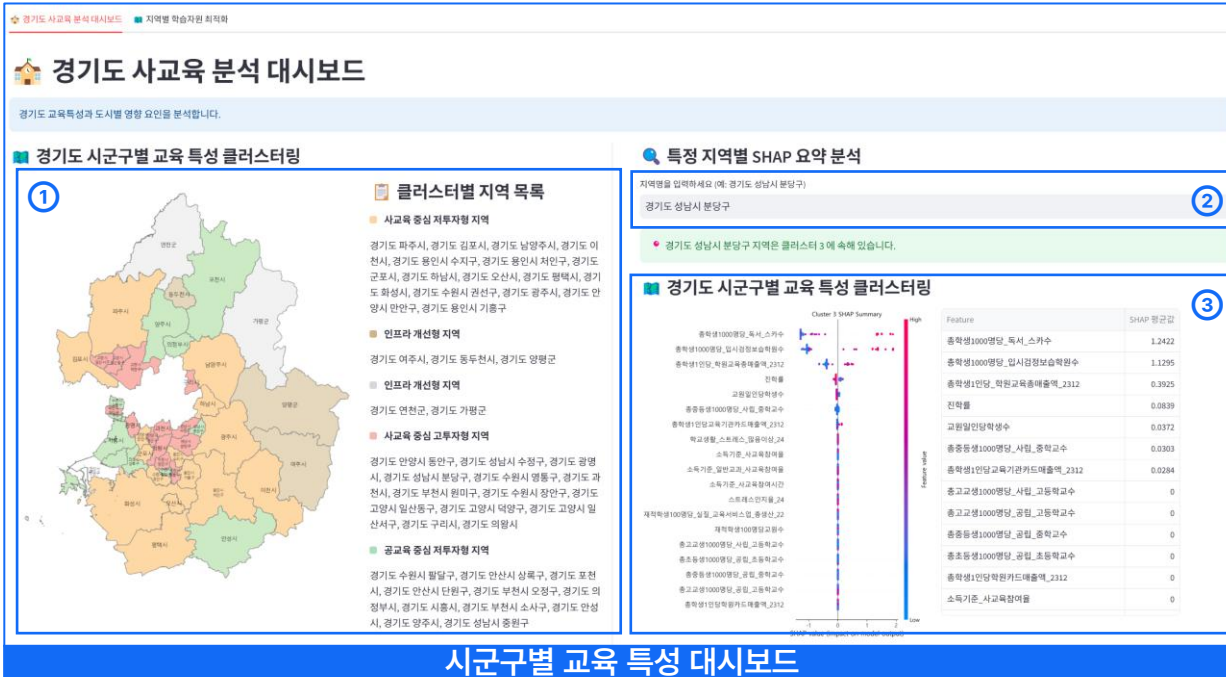


2. 실무적 활용 방안

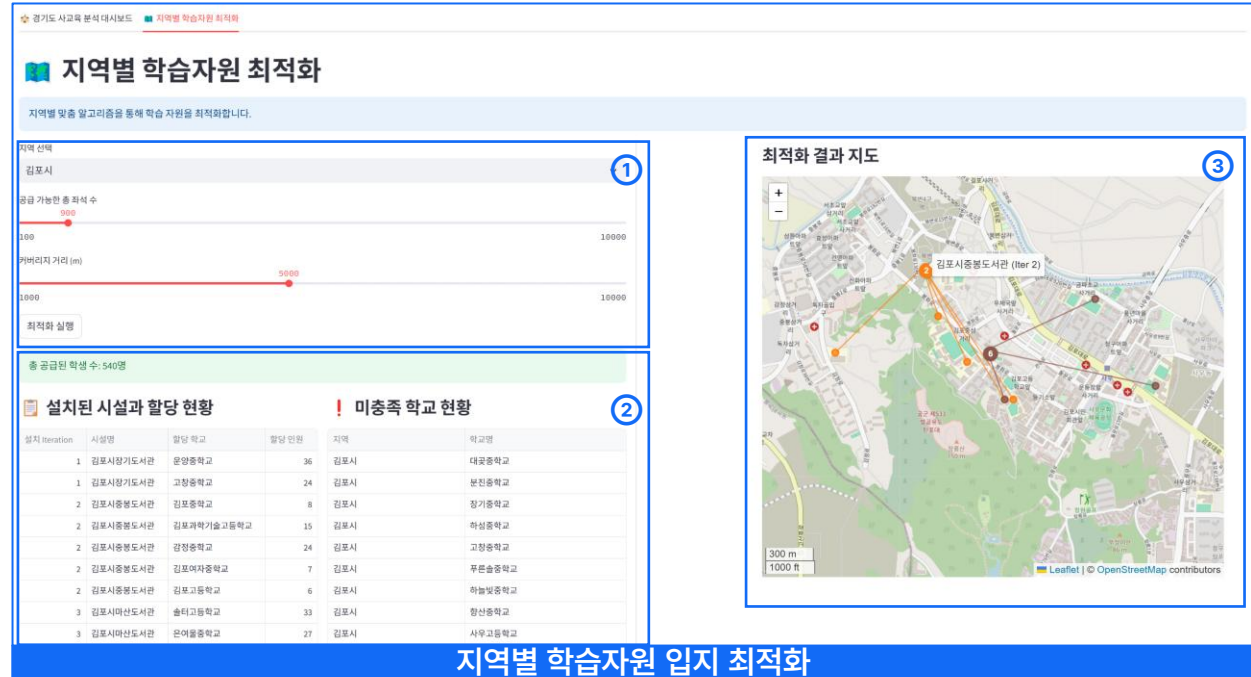
연속적인 시스템을 구축하여 교육부와 시도교육청의 정책 및 복지 담당자는 지속적으로 교육 상황을 모니터링하고, 교육 인프라가 부족하고 투자가 필요한 지역을 식별 가능하다. 더불어, 데이터 기반의 정량적 통계와 분석을 통해서 실무적 관점과 큰 조화를 이루는 본질적인 인사이트를 도출 하는데 활용할 수 있을 것이다.

일부 공공데이터를 이용해 구축한 현재 수요-공급 관계를 넘어서서, 다양한 공공 데이터를 활용하여 사회적 / 물리적 환경을 함께 고려하는 실효성 있는 수요-공급과 정밀한 최적화 알고리즘을 구현한다.

연속적 시스템 구축 - 의사결정 대시보드



- ① 분기 혹은 특정 시기마다 변하는 데이터들을 반영하여 사용자가 확인하는 시점의 지역별 교육 특성과 지역별 유형을 나타낸다.
- ② 사용자가 세부 요인 분석을 원하는 지역을 입력 받을 수 있다.
- ③ 입력한 지역의 교육 영향 변수들을 SHAP 분석을 기반으로 시각화하고 수치를 나타낸다.



지역별 학습자원 입지 최적화

- ① 학습 자원 분배가 계획된 도시를 입력하고, 공급량과 자원별 가용 거리를 입력한다.
- ② 입력한 수용량과 제약을 고려하여 최적의 세부 입지와 수용 가능한 인원 수를 도출하고, 미충족 수요를 확인한다.
- ③ 반응형 지도를 통해 수요지와 공급지 주변 환경과 시설을 확인할 수 있다.

Streamlit을 활용하여 실무자의 의사결정을 지원할 수 있는 실제 시스템을 구현하였다. 해당 시스템을 통해 지속적인 현황 파악과 지역별 공급-수요에 따른 유동적 최적화가 가능하다.

시사점 및 활용방안



시사점

- 최근 늘어나는 사교육비 부담을 경감하고, 부족한 교육 인프라를 채우기 위해 교육부의 신규 사업인 자기주도학습센터에 대한 분석이다.
- 군집화와 SHAP를 통해 지역별 교육 특성을 확인할 수 있었고, 공교육과 사교육에 대한 인사이트를 도출하였다.
- 입지의 우선순위와 지역별로 다른 최적화 기법을 적용해 자기주도학습센터의 자원 배치의 효율성을 높일 수 있었다.



한계점 및 보완방안

- 공공데이터에서만 수집하여 지역별 성과와 같은 민감한 정보까지는 접근하지 못하였다. 더 많은 공공기관과 협업, 파이프라인 확대를 통하여 고품질의 데이터를 활용하면 더 완성도 높은 분석이 가능할 것이다.
- 본 프로젝트는 경기도에 한정된 분석이었고, 입지 최적화는 포천시, 동두천시, 김포시에 한정하였다. 하지만, 자동화 파이프라인을 구축하여 데이터를 본 프로젝트의 시스템에만 연결할 수 있으면 어느 지역이든 실시간 인사이트를 수집할 수 있을 것이다.



추후 활용방안

- 다른 데이터들과 연계하여 지속적인 인사이트를 얻는 파이프라인을 구축할 수 있다.
- 교육 데이터 뿐 아니라 타 기관과의 통합적인 데이터 시스템을 구축하여 더 실효성 있는 활용이 가능할 것이다
- 연속적 시스템을 통해 얻은 결과로 학습 자원을 배분하고, 지역별 차별화된 정책을 수립할 수 있을 것이다.

분석 도구

데이터 수집 및 분석



군집화 및 시각화



최적화 모델링



의사결정 시스템



참고 문헌

참고 문헌

[1] 이지원, 김지영, 유기윤, & 양성철. (2019). 공공데이터를 활용한 초등 학생 돌봄시설의 최적입지 선정. *지적과 국토정보*, 49(2), 109-122.

[2] 조성아, & 김성연. (2021). 서울시의 특성화고등학교 공동실습소 입지 분석. *한국콘텐츠학회논문지*, 21(4), 393-403.

[3] 김정근. (2020). 사교육비 경감대책으로서 공교육 정상화 정책 논증 분석. *정책분석평가학회보*, 30(3), 139-179.

[4] 심재권. (2025). 학업부진 예측모델 개발 연구: 대구교육종단연구의 국어와 수학을 중심으로. *정보교육학회논문지*, 29(1), 53-61.

[5] 교육부. (2024. 5. 28). 제5차 사회관계장관회의 개최 [보도자료].

[6] 교육부. (2024. 5. 28). 교육데이터 개방·활용 확대방안 [보도자료].

뉴스 기사

[1] 이승은. (2025, 3월 13일). 학생 수 주는데 사교육비 29.2조 4년 연속 역대 최대. *YTN*.

https://www.ytn.co.kr/_ln/0102_202503131249235706

[2] 정인지. (2025, 4월 9일). 교육부, 사교육비 절감 위해 '자기주도 학습센터' 세운다. *머니투데이*.

<https://news.mt.co.kr/mtview.php?no=2025040910223832329>

[3] 이상미. (2025, 4월 10일). "공부 습관까지 맞춤 관리"...자기주도 학습센터 50곳 지정, *EBSNEWS*.

<https://news.ebs.co.kr/ebsnews/allView/60588841/N>