

데이터 분석·활용 결과보고서

(새만금 빅데이터 분석·활용 부문)

1. 데이터 분석·활용 개요

1-가. 데이터 분석·활용 명

신재생 에너지 잉여 전력을 이용한 녹색 인프라 구축

1-나. 적용 분야

전기차 충전소 확충, 수질 오염 개선, 그린뉴딜 실천

1-다. 데이터 분석·활용 요약

현재 새만금 지역에서는 막대한 양의 신재생 에너지가 생산되고 있다. 그러나, 에너지 보급량이 생산량에 비해 턱없이 부족하다는 사실을 데이터 분석을 통해 알게 되었고, **새만금 지역에 수많은 잉여 전력이 발생하고 있음을 확인하였다.** 우리는 이러한 잉여 전력을 새만금의 '그린뉴딜'을 실현하기 위한 **녹색 인프라를 구축하는 데 활용하고자 한다.** 잉여 전력을 이용하여 **전기차 충전소를 확충하고, 수질 오염을 개선하는 것이다.** 전기차 충전소 확충을 위해 새만금에 필요한 전기차 충전소 개수를 계산해보았다. 2027년까지 새만금 내에 전기차가 몇 대나 증가할지 예측하고, 이러한 전기차 증가 예측에 따라 필요한 이상적인 전기차 충전소의 개수를 계산했다. 한편, 새만금 하천별 수질 지표 데이터를 이용해 각 하천의 오염도를 분석하고 하천끼리의 유사도에 따라 전체 새만금 하천을 4개의 그룹으로 나누었다. 하천 그룹별 수질 오염 개선 우선 순위를 매겨 개선이 시급한 하천 그룹을 찾았고, 해당 하천 그룹의 수질을 개선하는 데 대략 얼마만큼의 전기를 더 투자할 수 있을지까지 알아보았다. 이를 바탕으로 수질 오염을 개선하는데 더욱 적절한 양의 전기를 투자할 수 있게 된다면 그린뉴딜을 실현하는데 한층 가까워질 것이다.

***그린뉴딜 : 환경 문제 개선과 경제 발전을 동시에 이루기 위한 종합적인 정책 방향을 의미한다. 새만금이 동북아 경제 허브로 나아가기 위해서는 필수적으로 가져야 하는 정책 목표이다.**

2. 데이터 분석·활용의 적절성

2-1. 활용 공공빅데이터

- 지역별 신재생 에너지 생산량, 보급량 / 한국에너지공단 / https://www.knrec.or.kr/biz/statistics/supply/supply01_02_list.do
- Global EV Outlook 2020 / IEA / <https://www.iea.org/reports/global-ev-outlook-2020>
- 지역별 전기차 현황정보 / 한국전력공사 / <https://www.data.go.kr/data/15039554/fileData.do?recommendDataYn=Y>
- 지역별 전기차 충전소 현황정보 / 한국전력공사 /

<https://www.data.go.kr/data/15039765/fileData.do>

- 전기차 및 충전기 보급·이용 현황 분석 보고서 / 전력거래소 /

https://new.kpx.or.kr/board.es?mid=a10502000000&bid=0045&act=view&list_no=51836&tag=&nPage=1

- 인구밀도(인구주택총조사기준) / 통계청 KOSIS /

https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1B040A3&conn_path=l2

- 지역별 관광 현황 / 한국관광데이터랩 /

<https://datalab.visitkorea.or.kr/datalab/portal/loc/getAreaDataForm.do#>

- 새만금 하천별 수질지표 / 물환경정보시스템 /

http://water.nier.go.kr/web/waterMeasure?pMENU_NO=571

- 공공하수처리시설 현황 / 한국환경공단 /

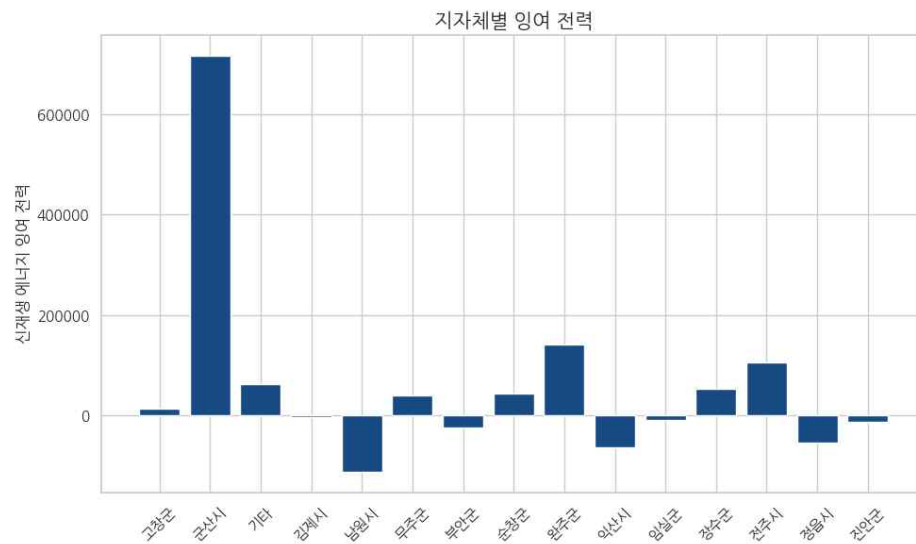
<https://www.data.go.kr/data/3073222/fileData.do?recommendDataYn=Y>

- 전북권통게이트진출입일교통량 / 한국도로공사 /

<https://www.data.go.kr/data/15062155/fileData.do>

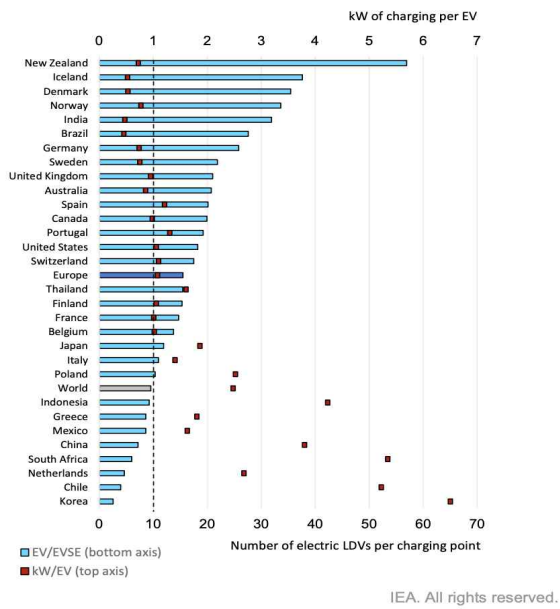
2-2. 데이터 분석·활용 배경 및 활용 적절성

한국에너지공단의 '지역별 신재생 에너지 생산량, 보급량' 데이터를 활용하여 지역별 잉여 전력을 비교해본 결과, 새만금 지역(군산 지역)이 다른 지역 대비 지나치게 많은 양의 잉여 전력이 발생되고 있음을 알게 되었다. 그래서 이러한 잉여 전력을 효율적으로 활용할 수 있는 방안으로 수질 오염 개선과 전기차 충전소 확충을 생각해보았다.



첫 번째로 잉여 전력을 전기차 충전소 확충에 활용하고자 한다. 현재 한국은 전기차 전력 사용량이 많은 것에 비해 전기차 충전 인프라가 매우 부족한 상태이다. 이를 IEA의 Global EV Outlook 2020 데이터를 통해 확인할 수 있었다.

Charging points per EV and kW per electric LDV, 2021



필요한 전기차 충전소 개수를 구하기 위해 2027년까지 증가할 전기차 대수부터 예측하기로 했다. 한국전력공사의 '지역별 전기차 현황정보' 데이터에서 전국의 2022년 1년부터 2023년 3월까지의 전기차 대수 데이터만 뽑아와 ARIMA 모델을 적용해 2027년까지 증가할 전기차 대수를 예측했다. 이렇게 나온 예측 전기차 대수 데이터를 이용해 한국관광데이터랩의 '지역별 관광 현황' 데이터와 연결지어 향후 4년간 전라북도의 방문자수와 목적지 검색건수를 예측했고 전력거래소의 '전기차 및 충전기 보급·이용 현황 분석 보고서' 데이터와도 연결지어 향후 사용될 전력사용량을 예측했다. 이와 같이 예측한 데이터로 이상적인 전기차 충전소 개수를 예측했다. 그리고 한국도로공사의 '전북권톨게이트진출입일교통량' 데이터로 가장 교통량이 많은 5군데의 고속도로에 전기차 충전소가 설치된다면 좋을 것이라 판단해 시각화했다.

두 번째로 잉여 전력을 새만금 하천의 수질 오염을 개선하는 데 활용하고자 한다. 물환경정보시스템의 '새만금 하천별 수질지표' 데이터를 '국가수자원관리종합정보시스템'의 하천수 수질환경 기준을 활용해서 하천별 수질 오염도를 분석했다. 그다음 오염도까지 합쳐진 수질지표 데이터에 SOM Clustering 기법을 적용해서 하천끼리의 유사도에 따라 하천 그룹을 4개로 나누었다. 이중 가장 수질 개선이 시급한 하천 그룹을 그룹 내 오염도 수치를 이용해 찾을 수 있었다.

현재 수질 개선이 가장 시급한 하천 그룹에 사용되는 하수처리 전력량이 어느 정도인지, 얼마만큼의 전력을 이 하수처리 그룹에 더 투자할 수 있을지 알아보기 위해 한국환경공단의 '공공하수처리시설 현황' 데이터를 분석해보았다. 새만금 지역의 전체 하수처리시설 전력 사용량 평균을 구하고, 수질 개선 우선 하천 그룹에 속한 하천별 전력 사용량 평균을 구해 비교하였다. 전체 평균인 592731kWh에 비해 수질 개선 우선 하천 그룹에 실제 사용하고 있는 전력량은 월등히 낮았다.

```
data['연간 총 에너지 사용량_총 전력사용량(kWh)'].mean()
```

```
592731.21875
```

```
In [65]: data[data['지류'] == '동진강']['연간 총 에너지 사용량_총 전력사용량(kWh)'].mean()
```

```
Out[65]: 313062.0
```

```
In [66]: data[data['지류'] == '원평천']['연간 총 에너지 사용량_총 전력사용량(kWh)'].mean()
```

```
Out[66]: 226733.75
```

```
In [67]: data[data['지류'] == '용암천']['연간 총 에너지 사용량_총 전력사용량(kWh)'].mean()
```

```
Out[67]: 40618.0
```

```
In [68]: data[data['지류'] == '용호천']['연간 총 에너지 사용량_총 전력사용량(kWh)'].mean()
```

```
Out[68]: nan
```

```
In [69]: data[data['지류'] == '고부천']['연간 총 에너지 사용량_총 전력사용량(kWh)'].mean()
```

```
Out[69]: 95987.125
```

```
In [70]: data[data['지류'] == '마산천']['연간 총 에너지 사용량_총 전력사용량(kWh)'].mean()
```

```
Out[70]: 73908.0
```

3. 데이터 분석·활용의 독창성

- 정규 분포가 아닌 원본 데이터를 정규 분포와 유사하게 변환하기 위해 Box-Cox 변환 방법을 사용했다.
- 필요한 전기차 충전소 개수를 예측하는 데에 있어서 단순히 지역별 전기차 충전소의 '전력 사용량' 데이터만 활용한 것이 아니라 한국 관광 데이터랩의 지역별 '방문자 수'와 '목적지 검색 건수' 데이터도 활용하여 전기차 충전소가 방문자 수, 목적지 검색 건수에 비례해서 많이 필요하다는 인사이트를 새롭게 얻게 되었다.
- 랜덤 포레스트와 그리드 서치를 함께 모델링함으로써 가장 성능이 좋은 하이퍼 파라미터로 최종 모델을 튜닝했다.
- 수질지표와 오염도를 합친 데이터에 일반적인 clustering 기법이 아닌 Ward 연결법과 차원 축소 및 군집화를 동시에 수행하는 SOM clustering 기법을 적용하였다. 이 기법은 단순히 개체 간의 거리에 따라 clustering을 하는 것이 아니라 군집 내의 오차 제곱합에 기초하여 clustering을 수행하므로 개체들을 하나로 그룹화하며 발생하는 정보손실이 최소화된다.
- SOM clustering을 본 모델에 적용하기 전에 map_size, sigma, learning_rate, init_method 등 다양한 매개변수 조합에 대한 실험부터 수행했다. 실험 결과, cluster의 개수가 우리가 사용하는 하천 데이터 개수에 가장 적합하며, quantization_error 값도 가장 적은 매개변수 조합을 이용해 모델링을 하였다.

4. 데이터 분석·활용의 기술성

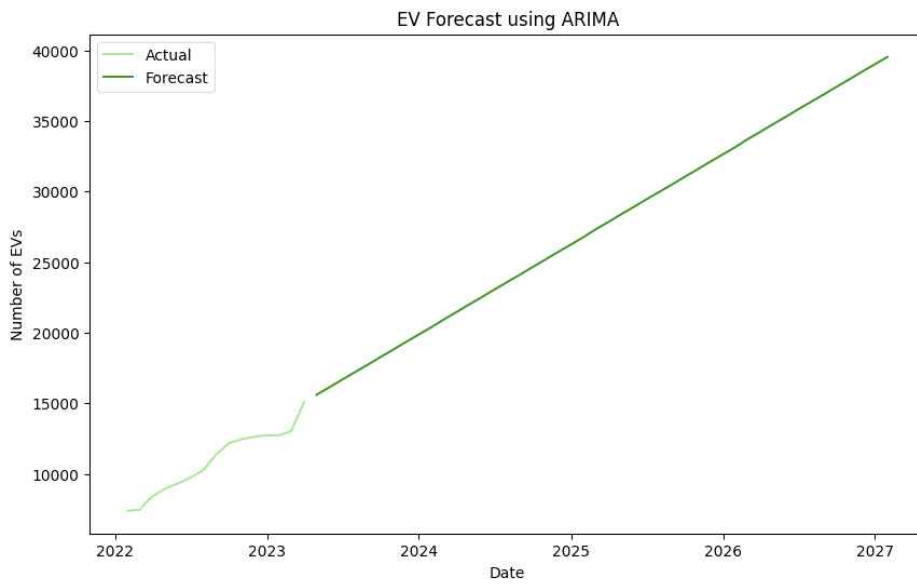
1. 2023년 4월부터 2027년 1월까지 46개월 동안 증가할 전기차 대수 예측

1) ARIMA 모델

- 전북의 2022년 1년부터 2023년 3월까지의 전기차 대수 데이터를 이용함.
- 성능 평가 Mean Absolute Error (MAE): 8417.74

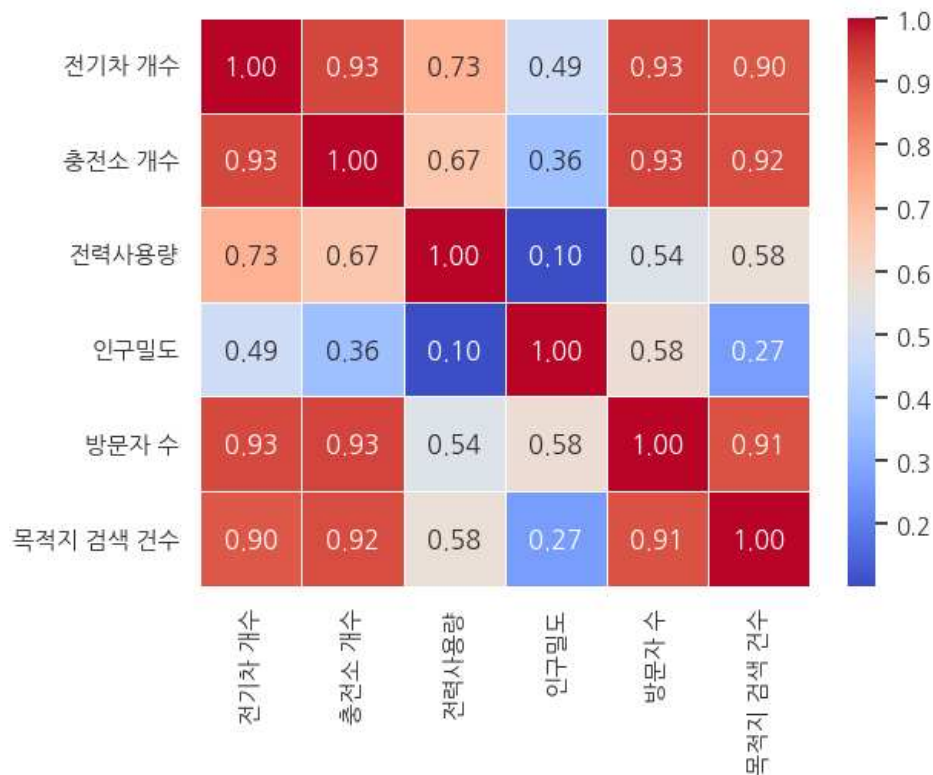
2) 예측 결과 (증가할 전기차 대수)

2024년 1월	20388대
2025년 1월	26769대
2026년 1월	33147대
2027년 1월	39524대



2. 예측된 전기차 대수에 따른 이상적인 충전소 개수 예측

- 1) 지역별 전기차 대수, 충전소 개수, 인구밀도, 전력 사용량, 방문자 수, 목적지 검색 건수로 이루어진 데이터 프레임 생성
- 2) 히트맵을 확인한 결과 가장 correlation이 적은 '인구밀도' 열 삭제



3) PowerTransformer(method = 'box-cox')로 데이터 스케일링 & PCA로 차원축소

- 전기차 대수, 충전소 개수에 비해 전력사용량, 방문자 수, 목적지 검색 건수의 값이 훨씬 크기 때문에 저 세 열만 각각 데이터 스케일링 후 차원 축소

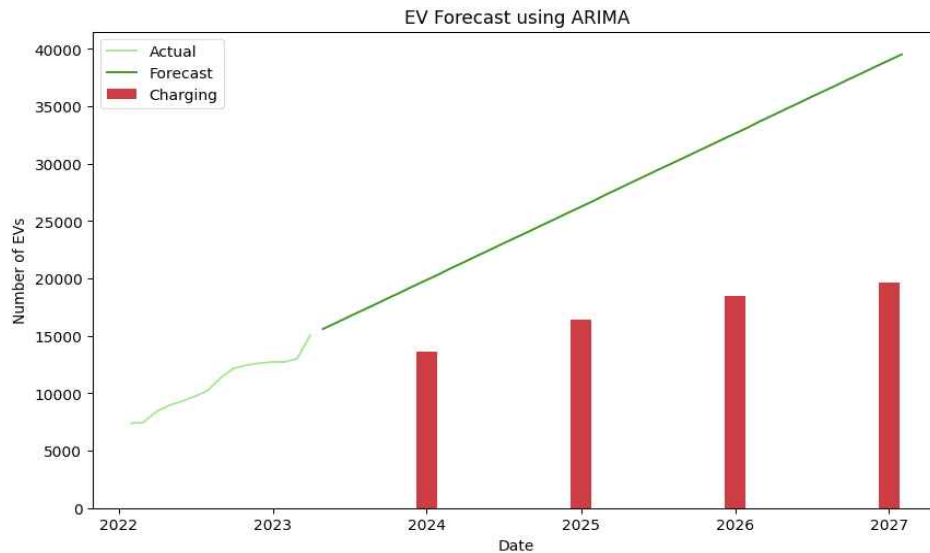
4) 증가하는 전기차 대수에 따라 전력사용량, 방문자 수, 목적지 검색 건수가 영향을 받을 것이라고 보고 선형 회귀 모델을 이용하여 각각의 값 예측

5) 랜덤 포레스트 & 그리드 서치로 모델 생성

- 독립변수 : 전기차 대수, 전력 사용량, 방문자 수, 목적지 검색 건수
- 종속변수 : 충전소 개수
- 전체 데이터의 20%를 test 데이터로 사용
- 성능 평가 R2 score : 0.84

6) 예측 결과 (이상적인 전기차 충전소의 개수)

2024년 1월	273개
2025년 1월	329개
2026년 1월	370개
2027년 1월	393개



3. 전라북도 전기차 충전소의 적절한 위치 파악

전기차 충전소의 설치 장소는 한국도로공사 데이터를 이용해 전라북도에서 총교통량이 많은 고속도로 5개 (전주IC, 익산IC, 동군산IC, 논산IC, 완주IC)로 지정해 대시보드에 시각화 했다. (남대전은 전라북도과 많이 떨어져 있다고 판단해 제외했고 서전주는 전주IC와 가깝기 때문에 제외했다.)

	영업소명	총교통량
28	전주	1756859
7	남대전	1520104
20	서전주	1414088
26	익산	1318649
25	완주	1279094
9	논산	1058542
12	동군산	1004098

4. 새만금 하천별 오염도 측정

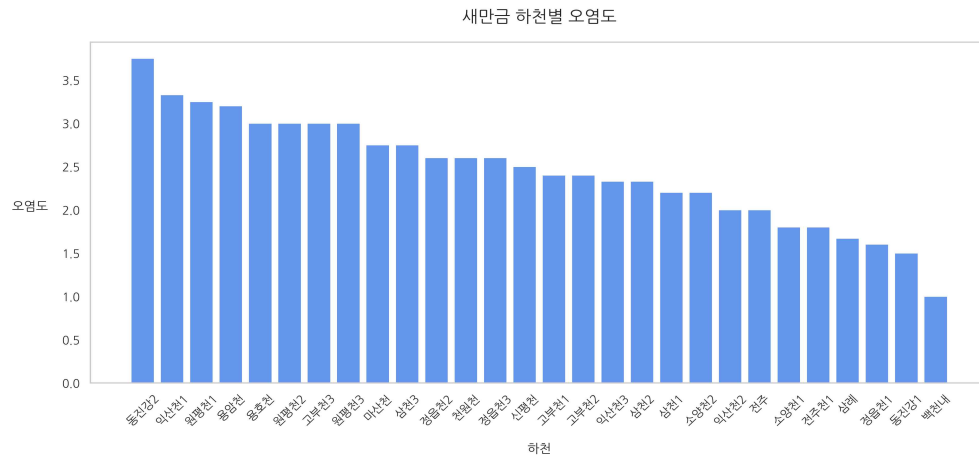
1) pandas 라이브러리를 이용해서 새만금 하천별 수질지표 데이터 전처리.

- 하천간 분별력이 없는 pH(수소이온농도), DO(용존산소량) 수질 지표 제거.

2) 수질 지표 별로 국가수자원관리종합정보시스템의 하천수 수질환경 기준에 따라 오염도 측정 후 하천 별로 평균 냄.

- 매우 좋음 : 1, 좋음 : 2, 약간 좋음 : 3, 보통 : 4, 약간 나쁨 : 5, 나쁨 : 6, 매우나쁨 : 7

3) 오염도 분석 결과 :



5. 하천 그룹 생성 (SOM clustering 적용)

1) 기존 전처리한 수질지표 데이터에 '오염도' 열이 추가된 새로운 데이터 프레임 구성.

2) 속성 별로 수치 값의 범위가 천차만별이므로 표준화 진행

- 표준화 공식 : $(\text{속성값} - \text{평균}) / \text{표준편차}$

3) SOM의 4개 매개변수(map_size, sigma, learning_rate, init_method) 조합에 대한 실험 진행.

- quantization_error 값이 가장 적은 매개변수 조합 선택

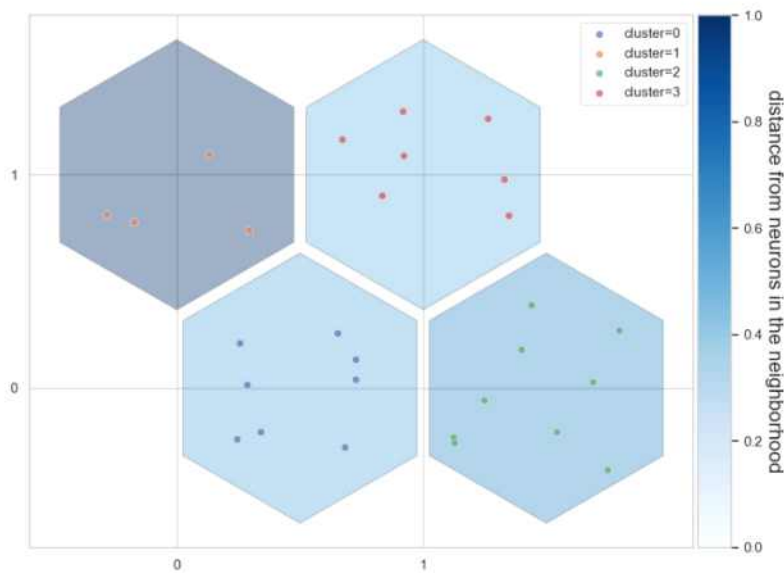
- 적당한 n_cluster 개수 선택

	map_size	sigma	learning_rate	init_method	qe	n_cluster
81	2x2	0.9	0.5	random_init	1.169116	4
82	2x2	0.9	0.4	random_init	1.169709	4
83	2x2	0.7	0.5	random_init	1.169744	4
84	2x2	0.8	0.5	random_init	1.169914	4
85	2x2	0.8	0.4	random_init	1.170042	4
...
157	2x2	0.3	0.4	random_init	1.322381	4
158	2x2	0.4	0.4	random_init	1.322381	4
159	2x2	0.1	0.1	random_init	1.329264	4
160	2x2	0.2	0.1	random_init	1.329264	4
161	2x2	0.3	0.1	random_init	1.329524	4

4) 3번에서 선택된 매개변수 조합으로 모델 생성 후 clustering 진행

cluster	이름	개수
0		8
1		4
2		9
3		7

5) Clustering 결과 :



- Cluster 0 : 삼천1, 소양천2, 소양천1, 전주천1, 삼레, 정읍천1, 동진강1, 백천내
(오염도 1.00 ~ 2.20)
- Cluster 1 : 익산천1, 익산천3, 삼천2, 익산천2 (오염도 2.00 ~ 3.33)
- Cluster 2 : 원평천2, 삼천3, 정읍천2, 천원천, 정읍천3, 신평천, 고부천1, 고부천2, 전주
(오염도 2.00 ~ 3.00)
- Cluster 3 : 동진강2, 원평천1, 용암천, 용호천, 고부천3, 원평천3, 마산천
(오염도 2.75 ~ 3.75)

5. 활용방안

5-1. 활용방안

1. 데이터를 분석해서 얻은 전기차 충전소 필요 개수를 바탕으로 구체적 전기차 충전소 확충 계획을 짤 수 있다. 미래에 부족할 필수 인프라를 현재에서부터 미리 미리 채워나가며 새만금은 경제적으로 빠르게 한 단계 성장할 수 있고, 잉여 전력도 효율적으로 활용함으로써 새만금 내 재생 에너지 순환이 원활히 돌아가도록 할 수 있다.

2. 데이터 분석을 통해 얻은 새만금 내 하천별 오염도와 하천 그룹을 이용하여, 하천 그룹별 수질 개선 우선 순위를 매기고 수질 오염 개선 계획을 짤 수 있다. 현재 나온 4개의 cluster 중 cluster 3이 오염도가 2.75 ~ 3.75로 다른 cluster 대비 가장 높다. 따라서 해당 하천 그룹에 새만금은 더 많은 잉여 전력을 투자하여 수질 오염 개선에 힘쓸 수 있고, 신재생 에너지 발전으로 생성된 잉여 전력도 묵혀두는 것이 아니라 효율적으로 사용할 수 있게 되는 것이다.

3. 오염도가 높은 하천 그룹 순서로 잉여 전력을 좀 더 투자할 때 일반 하수처리시설에 전력을 더 투자할 수도 있겠지만, 전기 침전과 같이 전기 자체를 활용해 수질을 개선하는 방법에 전력을 투자해볼 수도 있을 것이다. 새만금 내에는 아직 전기만을 이용해 수질을 개선하는 하수처리시설이 거의 없다.

5-2. 사회적 가치 창출

넘쳐나는 신재생 에너지 전력을 새만금이 아직 충분히 갖추지 못한 녹색 인프라를 구축하는 데 사용하자는 아이디어는 경제와 환경이 더욱 동적으로 순환하는 데 박차를 가할 것이다. 새만금을 향한 주목과 관심, 지원이 커짐에 따라 현재 다양한 신생 산업들이 새롭게 자리를 잡고 있다. 특히 신재생 에너지 발전 산업은 그 수가 어마무시하다. 이렇게나 많은 산업들이 새만금 내에 새롭게 자리 잡을 수 있도록 신경을 쓰는 것도 새만금의 발전을 위해 당연히 중요하지만, 다양한 산업과 정책들이 마치 하나의 유기체처럼 조화를 이루며 서로 긍정적인 영향을 주고받을 수 있도록 힘쓰는 것도 새만금의 성장을 길게 볼 때 빼놓을 수 없을 정도로 중요하다. 새만금이 그린뉴딜이라는 하나의 큰 목표 아래, 여러 다채로운 산업들이 서로 영향을 주고받으며 성장할 수 있도록 많은 관심을 기울인다면 새만금은 눈부신 자연과 경제적 활력을 갖는 동북아 경제허브로 성장할 수밖에 없을 것이다.

6. 기타