

산업통상자원부 공공데이터 활용 비즈니스아이디어 공모전 분석 결과 제출 양식

1 명칭

피어슨 상관계수를 이용한 집합자원 모델 개발

2 제안배경

현재 탄소중립은 전세계적으로 중요한 사안으로 인식되고 있으며, 재생에너지는 이를 해결할 수 있는 핵심 요소이다. 재생에너지를 위한 연구와 투자가 활발히 이루어지고 있지만, 재생에너지의 간헐성으로 인해 발전량 예측과 관련된 문제가 발생하고 있다. 일사량, 습도, 설비요소의 문제 등 인간이 관여할 수 없는 외부 요인에 따라 자원을 활용한 전기 생산이 좌우되어 발전소 운영과 에너지 거래에 어려움을 발생하며 수요와 공급의 균형을 맞추기 어려워진다.

현재의 발전량 예측방법은 기상 예보 데이터와 통계 모델을 기반으로 꾸준히 발전하고 있으나, 개별 발전소는 앞서 말한 재생에너지의 간헐성으로 인해 예측이 빗나갈 가능성이 다분히 존재한다. 이러한 문제를 해결하기 위해 산업통상자원부와 한국전력거래소는 2021년부터 재생에너지 발전량 예측제도를 시행하며 재생에너지발전량을 예측 오차율에 따라 (1kw당 6%이하 4원, 8%이하 3원)을 지급하고 있다.

재생에너지의 간헐성문제를 해결하는 방안으로 최근에는 재생에너지 발전량 예측 방법으로 가상발전소(VPP)가 주목받고 있다. 가상발전소는 소규모 신재생 에너지 발전설비와 에너지 저장장치 등 분산형 에너지 자원을 클라우드 기반 소프트웨어로 통합해 하나의 발전소처럼 관리하는 시스템이다. 이를 통해 여러 분산자원을 조합하여 하나의 가상발전소로 운영함으로써, 재생에너지의 간헐성으로 인한 문제를 보완하고 이를 효율적으로 관리할 수 있다.

따라서 빅데이터 분석을 통해 예측 오차율을 최소화하는 가상발전소(VPP) 조합모형을 구성하는 것을 제안한다. 빅데이터를 활용하면, 예측과 실제 발전량 사이의 오차를 분석하여 어떤 자원과 변수들이 예측 정확도에 영향을 주는지 파악할 수 있다. 이를 통해 특정 날짜에 상관없이 안정적으로 예측 정확도를 보장하는 조합모형을 구성하는 데 기여할 수 있다.

3 분석 내용 및 분석 결과

데이터 셋 전처리

- 데이터 분석에 들어가기 전에 데이터의 세부 사항을 파악하고 결측치나 이상치에 대한 처리를 하는 것이 더욱 정확한 분석 결과를 이끌어 낸다. 추후에 시행할 분석의 용이성을 위해 다음과 같은 작업들을 수행하였다.

우선 데이터 셋의 구성 요소부터 파악했다.

	일자	구분1	발전소코드	발전소용량 (KW)	0:00	1:00	2:00	3:00	4:00	5:00	...	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	합계
0	2023-03-01	실측	G1053	249.48	0.00	0.00	0	0	0	0.00	...	0.10	65.80	33.60	133.70	161.1	150.5	0	0	0	640.90
1	2023-03-01	예측	G1053	249.48	0.95	0.97	0.94	0.97	0.96	0.95	...	20.08	93.61	35.62	161.20	161.2	46.95	0.89	0.98	0.98	739.26
2	2023-03-01	실측	G1018	288.00	0.00	0.00	0	0	0	0.00	...	101.26	53.50	29.09	2.21	0.0	0	0	0	0	787.25
3	2023-03-01	예측	G1018	288.00	0.00	0.00	0	0	0	0.00	...	93.33	30.84	27.05	1.43	0.0	0	0	0	0	618.75
4	2023-03-01	실측	G1034	298.08	12.40	0.00	0	0	0	0.00	...	18.60	63.80	28.80	0.60	0.0	164.2	135.6	0	0	552.40

5 rows x 29 columns

제11회 산업통상자원부 공공데이터활용 BI공모전_빅데이터 분석 과제 5_데이터.csv

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10476 entries, 0 to 10477
Data columns (total 29 columns):
#   Column      Non-Null Count  Dtype
---  --
0   일자        10476 non-null object
1   구분1       10476 non-null object
2   발전소코드  10476 non-null object
3   발전소용량(KW)  10476 non-null float64
4   0:00        10476 non-null float64
5   1:00        10476 non-null float64
6   2:00        10476 non-null object
7   3:00        10476 non-null object
8   4:00        10476 non-null object
9   5:00        10476 non-null float64
10  6:00        10476 non-null float64
11  7:00        10476 non-null float64
12  8:00        10476 non-null float64
13  9:00        10476 non-null float64
14  10:00       10476 non-null float64
15  11:00       10476 non-null float64
16  12:00       10476 non-null float64
17  13:00       10476 non-null float64
18  14:00       10476 non-null float64
19  15:00       10476 non-null float64
20  16:00       10476 non-null float64
21  17:00       10476 non-null float64
22  18:00       10476 non-null float64
23  19:00       10476 non-null float64
24  20:00       10476 non-null object
25  21:00       10476 non-null object
26  22:00       10476 non-null object
27  23:00       10476 non-null object
28  합계        10476 non-null float64
dtypes: float64(19), object(10)
memory usage: 2.3+ MB
```

- 데이터의 시각화 모듈 사용 시 한글은 오류가 나는 경우가 있기 때문에 다음과 같이 영어 명칭으로 수정해주었다.

전	일자	구분1	발전소코드	발전소용량(KW)	합계
후	date	type1	power_plant_code	power_plant_capacity	sum

- date열의 데이터 타입을 date_time으로 변환.

- 결측값 '-'으로 인해 object 타입이 된 시간대열도 숫자형으로 변환했다.
- type1 열의 실측과 예측 여부를 활용하기 위해 실측은 1, 예측은 0으로 값을 바꾸는 작업을 실시했다.
- 발전소의 총 개수는 169개인 것을 확인했다.

```
합력 '2:00'에 1개의 NaN 값이 있습니다.
해당 데이터의 발전소 코드, 일자, 실측예측:
2974 2023-03-13 85024 1
합력 '3:00'에 2개의 NaN 값이 있습니다.
해당 데이터의 발전소 코드, 일자, 실측예측:
224 2023-03-02 85024 1
5224 2023-03-22 85024 1
합력 '4:00'에 1개의 NaN 값이 있습니다.
해당 데이터의 발전소 코드, 일자, 실측예측:
6974 2023-03-29 85024 1
합력 '20:00'에 1개의 NaN 값이 있습니다.
해당 데이터의 발전소 코드, 일자, 실측예측:
4474 2023-03-19 85024 1
합력 '21:00'에 2개의 NaN 값이 있습니다.
해당 데이터의 발전소 코드, 일자, 실측예측:
1224 2023-03-06 85024 1
2474 2023-03-13 85024 1
합력 '22:00'에 1개의 NaN 값이 있습니다.
해당 데이터의 발전소 코드, 일자, 실측예측:
3974 2023-03-17 85024 1
합력 '23:00'에 1개의 NaN 값이 있습니다.
해당 데이터의 발전소 코드, 일자, 실측예측:
6224 2023-03-26 85024 1
```

결측치가 존재하는 데이터 확인

- S5024 발전소의 실측값에 결측치가 몰려있는 것을 확인 할 수 있었다.
- 결측치 발생 시간은 발전이 없을 새벽과 저녁 시간대에만 발생했기 때문에 0으로 값을 대체하였다.

데이터 파일 분할

- 데이터의 활용 용이성을 위해 사본 파일을 각 발전소 별 date 기준 정렬 후 169개의 파일로 분할 (code1.csv ~ code169.csv)

	type1	power plant capacity	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	...	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	sum
count	62.000000	6.200000e+01	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	...	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000
mean	0.500000	2.494800e+02	0.458548	0.462258	0.465223	0.469032	0.465000	0.461129	0.474163	0.433968	...	93.317419	71.337742	30.643226	146.580484	153.913710	143.812097	47.343710	0.402258	0.458387	1000.326129
std	0.504082	5.730745e-14	0.475280	0.478253	0.479620	0.481161	0.479223	0.468801	0.363286	5.842793	...	54.883174	25.233915	10.271082	14.360687	32.531974	44.979197	22.425556	0.419521	0.471309	306.876282
min	0.000000	2.494800e+02	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	9.100000	3.500000	120.990000	1.330000	0.000000	0.000000	0.000000	0.000000	215.810000
25%	0.000000	2.494800e+02	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	46.750000	58.105000	24.725000	133.325000	161.100000	161.000000	46.675000	0.000000	0.000000	794.415000
50%	0.500000	2.494800e+02	0.200000	0.105000	0.140000	0.115000	0.110000	0.270000	0.950000	1.980000	...	118.165000	78.705000	31.375000	134.050000	161.200000	161.100000	51.870000	0.130000	0.215000	1091.095000
75%	1.000000	2.494800e+02	0.967500	0.960000	0.960000	0.970000	0.960000	0.940000	1.040000	14.897500	...	130.832500	90.402000	39.075000	161.200000	161.200000	161.200000	63.975000	0.867500	0.955000	1218.285000
max	1.000000	2.494800e+02	1.010000	1.030000	1.020000	1.030000	1.030000	1.030000	2.460000	22.800000	...	155.450000	107.200000	47.490000	161.200000	161.200000	161.200000	69.400000	1.000000	1.010000	1517.140000

code1.csv를 discribe로 확인

- 오차 데이터의 활용을 위해 “실측-예측”값을 담은 169개의 code(번호)_diff.csv 파일 생성

	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	...	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
count	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	...	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000
mean	0.517097	0.924516	0.930645	0.938065	0.930000	0.922258	0.947742	1.910645	6.638065	5.812258	...	2.045806	10.854194	8.660032	4.493955	26.644839	4.379032	14.246774	15.015806	0.804516	0.916774
std	0.157357	0.153185	0.142113	0.140288	0.141174	0.086128	0.117484	7.680577	15.285392	75.487645	...	66.879212	53.094927	22.906081	8.873475	7.329810	24.042467	47.978116	24.488642	0.153141	0.131995
min	1.010000	1.030000	1.020000	1.030000	1.030000	1.030000	1.160000	9.750000	49.270000	53.370000	...	148.370000	122.790000	49.600000	23.200000	29.700000	0.100000	96.200000	47.380000	1.000000	1.010000
25%	-0.980000	-0.970000	-0.980000	-0.990000	-0.970000	-0.950000	-1.010000	-1.865000	-14.070000	-18.750000	...	-75.219000	-75.470000	-17.410000	-8.690000	-78.100000	-0.100000	-0.700000	11.945000	-0.880000	-0.980000
50%	0.970000	0.960000	0.960000	0.970000	0.960000	0.940000	0.960000	2.120000	9.140000	11.130000	...	6.940000	12.860000	11.750000	4.490000	27.800000	0.100000	0.100000	16.170000	0.870000	0.960000
75%	-0.940000	-0.940000	-0.940000	-0.950000	-0.940000	-0.910000	-0.950000	-0.105000	3.115000	-2.540000	...	10.640000	0.250000	-3.805000	-0.800000	-7.500000	0.000000	-0.100000	17.090000	-0.750000	-0.915000
max	0.400000	0.210000	0.280000	0.250000	0.220000	0.540000	0.520000	3.270000	26.530000	62.920000	...	151.900000	129.100000	58.970000	19.400000	12.710000	133.880000	149.660000	66.440000	0.260000	0.430000

code1_diff.csv를 discribe로 확인

- 오차를 비교를 위해 “(실측-예측)/power_plant_capacity” 값을 담은 169개의 code(번호)_diff_divCapa.csv 파일로 생성

	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	...	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
count	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	...	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000	31.000000
mean	-0.003676	-0.003706	-0.003730	-0.003760	-0.003728	-0.003697	-0.003799	-0.007659	-0.026608	-0.023297	...	-0.008200	-0.043507	-0.034748	-0.018035	-0.106802	0.017553	0.057106	0.060188	-0.003225	-0.003675
std	0.000631	0.000614	0.000570	0.000594	0.000568	0.000345	0.000551	0.010745	0.061309	0.102163	...	0.268074	0.212822	0.091815	0.035568	0.029376	0.096378	0.192313	0.098151	0.000614	0.000529
min	-0.004048	-0.004129	-0.004089	-0.004129	-0.004129	-0.004129	-0.004650	-0.039081	-0.197491	-0.213925	...	-0.594717	-0.492184	-0.198814	-0.093234	-0.119048	-0.000401	-0.385602	-0.189915	-0.004008	-0.004048
25%	-0.003948	-0.003888	-0.003928	-0.003968	-0.003888	-0.003828	-0.004129	-0.014691	-0.059404	-0.075156	...	-0.101150	-0.102092	-0.069785	-0.034832	-0.112634	-0.000401	-0.000802	0.055896	-0.003567	-0.003928
50%	-0.003888	-0.003848	-0.003848	-0.003888	-0.003848	-0.003768	-0.003848	-0.009498	-0.036636	-0.044613	...	-0.027818	-0.051547	-0.047098	-0.017997	-0.111432	-0.000401	0.000401	0.064815	-0.003487	-0.003848
75%	-0.003768	-0.003768	-0.003768	-0.003808	-0.003768	-0.003648	-0.003808	-0.000421	0.012486	0.010181	...	0.042649	0.001002	-0.015252	-0.003207	-0.110229	0.000000	0.000401	0.068502	-0.003006	-0.003668
max	-0.007163	-0.000842	-0.001122	-0.001002	-0.000882	-0.002165	-0.002084	0.013107	0.106341	0.252205	...	0.608866	0.517476	0.236372	0.077762	0.050946	0.536636	0.599888	0.266314	-0.001042	-0.007724

code1_diff_divCapa.csv를 discribe로 확인

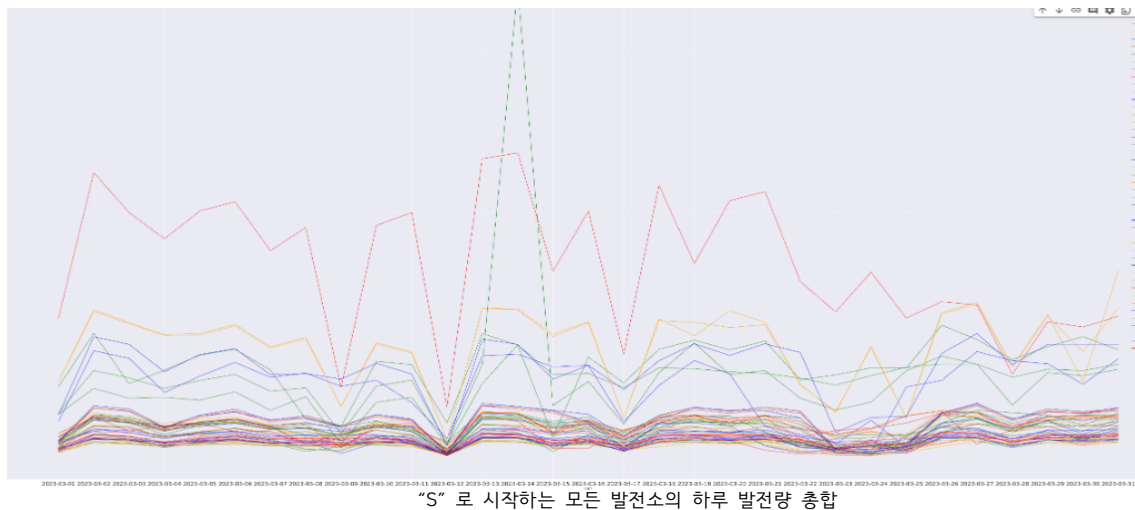
외부 데이터의 사용

발전소의 개수는 총 169개, 코드는 모두 S, G, B, V, N, M으로 시작하는 것을 확인할 수 있었다. 날씨 데이터와 같은 외부 데이터의 활용을 하기 위해서는 발전소의 지역 분포의 특징이 선행되어야 한다고 판단하였다.

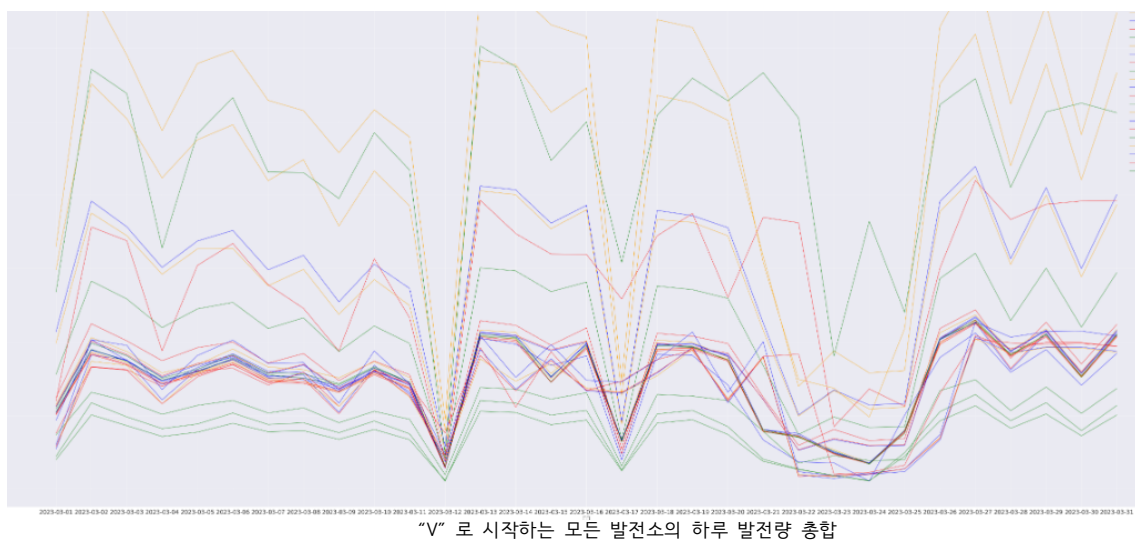
알파벳 별 발전소 코드 개수:

```
S    50
G    38
B    37
V    27
N     9
M     8
dtype: int64
```

– 발전소 코드가 S로 시작하는 발전소의 일별 총 발전량을 살펴 보았다.



– 마찬가지로 발전소 코드가 G, B, V, N, M 으로 시작하는 발전소의 일별 총 발전량을 살펴 보았다.



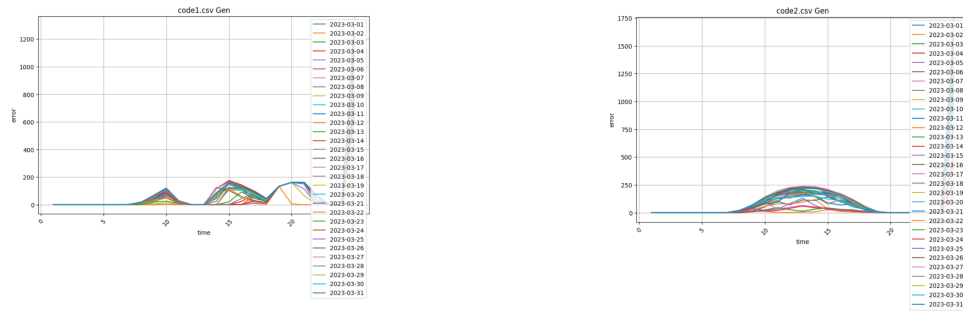
– 발전량의 총합이 모든 발전소 공통적으로 12일, 17일, 24일에 급격하게 줄어드는 것을 확인할 수 있었다.

일요일	월요일	화요일	수요일	목요일	금요일	토요일
			03/01 구름조금 ☀️ 최저기온 : 2.9℃ 최고기온 : 9℃ 일강수량 : 0mm 일적설량 : 0cm	03/02 맑음 ☀️ 최저기온 : -1.8℃ 최고기온 : 6.6℃ 일강수량 : 0mm 일적설량 : 0cm	03/03 구름조금 ☀️ 최저기온 : -1.9℃ 최고기온 : 10.4℃ 일강수량 : 0mm 일적설량 : 0cm	03/04 구름조금 ☀️ 최저기온 : 3.7℃ 최고기온 : 14.1℃ 일강수량 : 0mm 일적설량 : 0cm
03/05 맑음 ☀️ 최저기온 : 2.5℃ 최고기온 : 15.4℃ 일강수량 : 0mm 일적설량 : 0cm	03/06 맑음 ☀️ 최저기온 : 2.2℃ 최고기온 : 16.5℃ 일강수량 : 0mm 일적설량 : 0cm	03/07 맑음 ☀️ 최저기온 : 6.3℃ 최고기온 : 19.4℃ 일강수량 : 0mm 일적설량 : 0cm	03/08 비 ☔️ 최저기온 : 7.4℃ 최고기온 : 17.7℃ 일강수량 : 1.2mm 일적설량 : 0cm	03/09 비 ☔️ 최저기온 : 7.4℃ 최고기온 : 13.1℃ 일강수량 : 3.7mm 일적설량 : 0cm	03/10 맑음 ☀️ 최저기온 : 6.2℃ 최고기온 : 22.2℃ 일강수량 : 0mm 일적설량 : 0cm	03/11 구름많음 ☁️ 최저기온 : 10.2℃ 최고기온 : 22.8℃ 일강수량 : 0mm 일적설량 : 0cm
03/12 비 ☔️ 최저기온 : 0.6℃ 최고기온 : 12.6℃ 일강수량 : 5.6mm 일적설량 : 0cm	03/13 구름조금 ☀️ 최저기온 : -1.9℃ 최고기온 : 7.6℃ 일강수량 : 0mm 일적설량 : 0cm	03/14 구름조금 ☀️ 최저기온 : 3.3℃ 최고기온 : 15.4℃ 일강수량 : 0mm 일적설량 : 0cm	03/15 구름많음 ☁️ 최저기온 : 3.5℃ 최고기온 : 14.4℃ 일강수량 : 0mm 일적설량 : 0cm	03/16 구름조금 ☀️ 최저기온 : 0.4℃ 최고기온 : 9℃ 일강수량 : 0mm 일적설량 : 0cm	03/17 흐림 ☁️ 최저기온 : 0.7℃ 최고기온 : 12.7℃ 일강수량 : 0mm 일적설량 : 0cm	03/18 맑음 ☀️ 최저기온 : 2.5℃ 최고기온 : 14℃ 일강수량 : 0mm 일적설량 : 0cm
03/19 맑음 ☀️ 최저기온 : 2.3℃ 최고기온 : 15.2℃ 일강수량 : 0mm 일적설량 : 0cm	03/20 구름조금 ☀️ 최저기온 : 2.6℃ 최고기온 : 17.1℃ 일강수량 : 0mm 일적설량 : 0cm	03/21 맑음 ☀️ 최저기온 : 5℃ 최고기온 : 22.2℃ 일강수량 : 0mm 일적설량 : 0cm	03/22 구름많음 ☁️ 최저기온 : 9.2℃ 최고기온 : 25.1℃ 일강수량 : 0mm 일적설량 : 0cm	03/23 구름많음 ☁️ 최저기온 : 12.3℃ 최고기온 : 23.3℃ 일강수량 : 0mm 일적설량 : 0cm	03/24 흐림 ☁️ 최저기온 : 7.6℃ 최고기온 : 18.3℃ 일강수량 : 0mm 일적설량 : 0cm	03/25 구름조금 ☀️ 최저기온 : 8℃ 최고기온 : 18.7℃ 일강수량 : 0mm 일적설량 : 0cm
03/26 맑음 ☀️ 최저기온 : 4.7℃ 최고기온 : 14℃ 일강수량 : 0mm 일적설량 : 0cm	03/27 맑음 ☀️ 최저기온 : 1.9℃ 최고기온 : 14.5℃ 일강수량 : 0mm 일적설량 : 0cm	03/28 구름조금 ☀️ 최저기온 : 2.8℃ 최고기온 : 15.7℃ 일강수량 : 0mm 일적설량 : 0cm	03/29 맑음 ☀️ 최저기온 : 5.1℃ 최고기온 : 19.7℃ 일강수량 : 0mm 일적설량 : 0cm	03/30 맑음 ☀️ 최저기온 : 6.4℃ 최고기온 : 21.2℃ 일강수량 : 0mm 일적설량 : 0cm	03/31 맑음 ☀️ 최저기온 : 7.5℃ 최고기온 : 23.9℃ 일강수량 : 0mm 일적설량 : 0cm	

“2023년 3월 전국날씨[기상청 날씨누리-자료]”

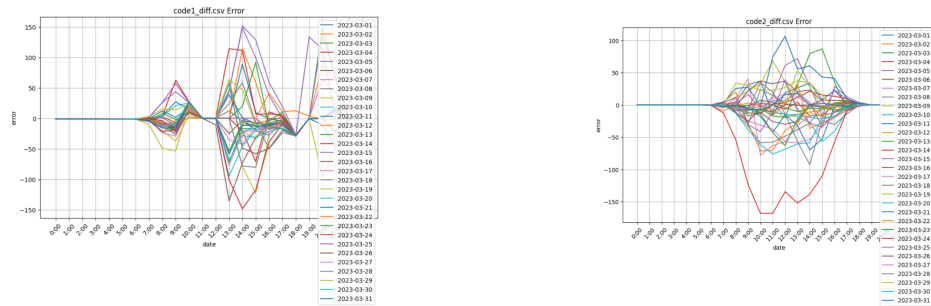
- 날씨 데이터와 급격히 발전량이 줄어드는 날을 비교해보니 전국적으로 흐린 날과 비가오는 날에는 강수량이 줄어드는 전형적인 태양광 발전의 특성을 보여주었다.
- 하지만 원본 데이터에서 비식별화한 코드의 지역을 알아낼 수 없다고 판단, 외부 데이터의 사용은 무의미할 것으로 결론지었다.

● 각 발전소 코드 별 1일부터 31일까지의 발전량 분포를 살펴보았다.

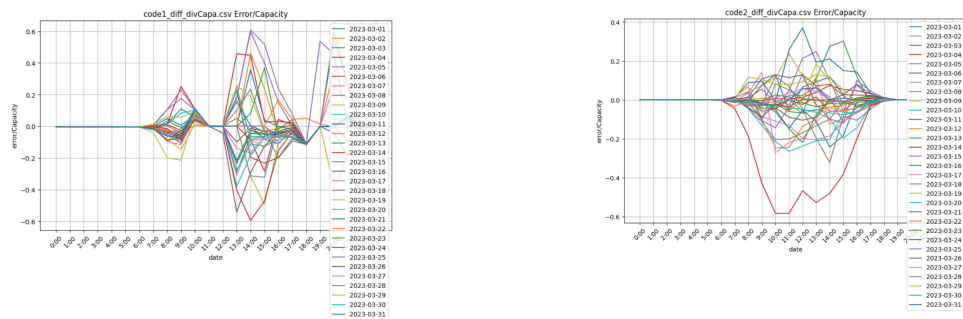


“code1, code2 발전소의 3월1일부터 31일까지의 실제 발전량”

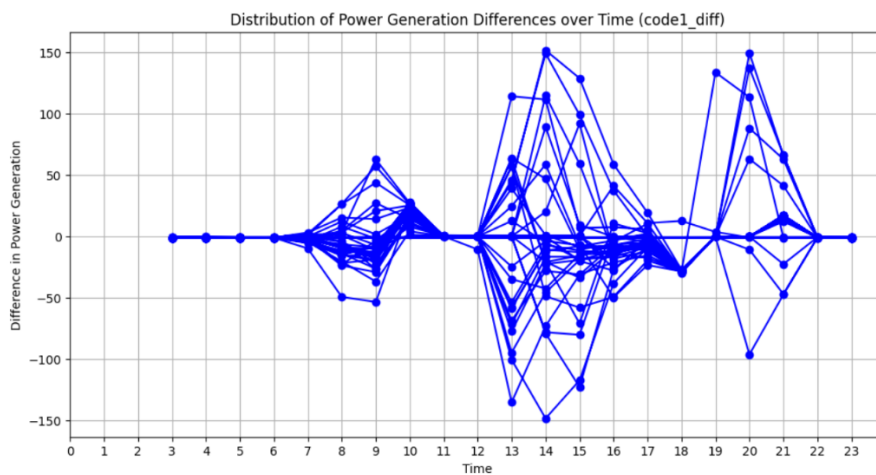
— 각 발전소 코드 별로 1일부터 31일까지의 오차(실측-예측) 데이터를 살펴보았다



“code1, code2 발전소의 3월1일부터 31일까지의 오차”



“code1, code2 발전소의 3월1일부터 31일까지의 오차율”



“code1의 3월1일부터 31일까지의 시간 별 오차 분포

- 각 발전소 별로 모든 날짜의 오차분포를 살펴보았다. 같은 발전소라도 날짜별로 오차의 변동이 매우 큰 것을 확인할 수 있었다.
- 모든 발전소가 낮 시간대의 발전량이 가장 큰 것을 확인했다.
- 모든 시간대 중 14:00가 가장 발전량이 많았으며, 13시, 15시 등 14시 기준으로 피크 타임이 형성되었다.
- 각 발전소 코드 별로 1일부터 31일까지의 오차(실측-예측) 데이터를 살펴 보니 오차가 가장 큰 시간도 낮 시간대인 것을 확인했다.
- 마찬가지로 오차를 또한 낮 시간대에 특히 14:00시에 가장 높은 것을 확인했다.

분석정리

	date	type	power_plant_code	power_plant_capacity	0:00	1:00	2:00	3:00	4:00	5:00	...	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	sum
0	2023-03-01	1	G1053	249.48	0.00	0.00	0.00	0.00	0.00	0.00	...	0.10	65.80	33.60	133.70	161.10	150.50	0.00	0.00	0.00	640.90
1	2023-03-01	0	G1053	249.48	0.95	0.97	0.94	0.97	0.96	0.95	...	20.08	93.61	35.62	161.20	161.20	161.20	46.95	0.89	0.98	739.26
2	2023-03-01	1	G1018	288.00	0.00	0.00	0.00	0.00	0.00	0.00	...	101.26	53.50	29.09	2.21	0.00	0.00	0.00	0.00	0.00	787.25
3	2023-03-01	0	G1018	288.00	0.00	0.00	0.00	0.00	0.00	0.00	...	93.33	30.84	27.05	1.43	0.00	0.00	0.00	0.00	0.00	618.75
4	2023-03-01	1	G1034	298.08	12.40	0.00	0.00	0.00	0.00	0.00	...	18.60	63.80	28.80	0.60	0.00	164.20	135.60	0.00	0.00	552.40
...
10473	2023-03-26	0	V9010	2019.60	8.15	7.04	6.94	6.92	7.06	7.11	...	863.60	427.73	95.22	37.43	1137.23	1206.20	1195.79	522.96	7.92	7300.76
10474	2023-03-30	0	V9010	2019.60	8.18	7.49	7.43	7.12	6.68	6.93	...	825.32	430.31	100.68	44.82	1129.87	1202.89	1195.94	521.72	7.63	9713.25
10475	2023-03-31	0	V9010	2019.60	8.22	7.78	7.51	6.94	6.91	6.84	...	818.10	415.38	106.57	46.59	1140.02	1207.97	1196.89	522.96	8.82	9738.10
10476	2023-03-12	0	V9010	2019.60	8.77	9.07	7.77	7.25	6.99	7.20	...	28.80	71.61	40.10	25.25	1123.83	202.02	15.10	11.59	8.38	2339.98
10477	2023-03-23	0	V9010	2019.60	11.26	9.42	8.88	9.23	7.88	7.79	...	28.80	168.21	64.03	34.52	1130.57	336.29	16.80	10.51	7.69	2418.82

10478 rows × 29 columns

“설비용량을 초과하는 발전량을 가진 데이터의 처리 후”

- 원본 데이터에서 설비 용량을 초과하는 발전량을 보이는 이상치가 존재하였다. 해당 이상치들을 각 발전소의 발전량의 최대인 설비용량으로 대체하였다.

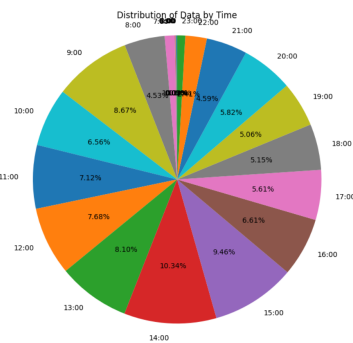
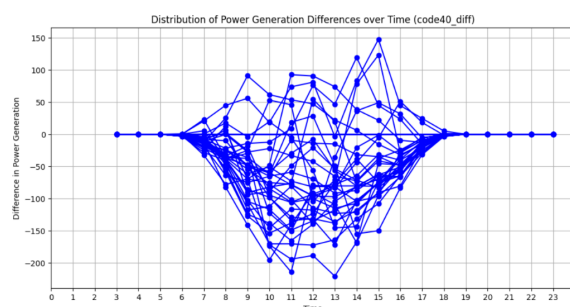
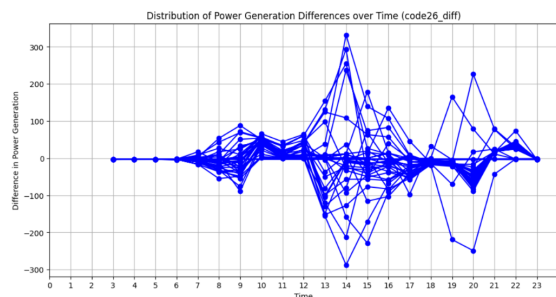


그림 1 시간대별 발전량 비율

- 모든 시간대 중 14:00가 가장 발전량이 많으며 가장 오차율도 높기 때문에 피크타임의 오차를 줄이는 것이 가장 효과적인 정산금 높이는 방법이 될 것이라고 판단하였다.
- 해당 모델에서는 분산형 재생에너지원의 최적 조합 방법에 관한 논문¹⁾을 참고하여 낮은 상관관계의 발전소끼리 조합하여 오차를 줄이는 방법을 채택하였다.
- 가장 역의 상관관계를 갖는 발전소끼리 합치는 것이 오차를 줄일 수 있는 방안이라는 판단하에 서로 높은 역의 상관관계를 갖는 발전소끼리 합친 후 오차의 분포를 보았다.



최적조합 탐색을 위한 알고리즘

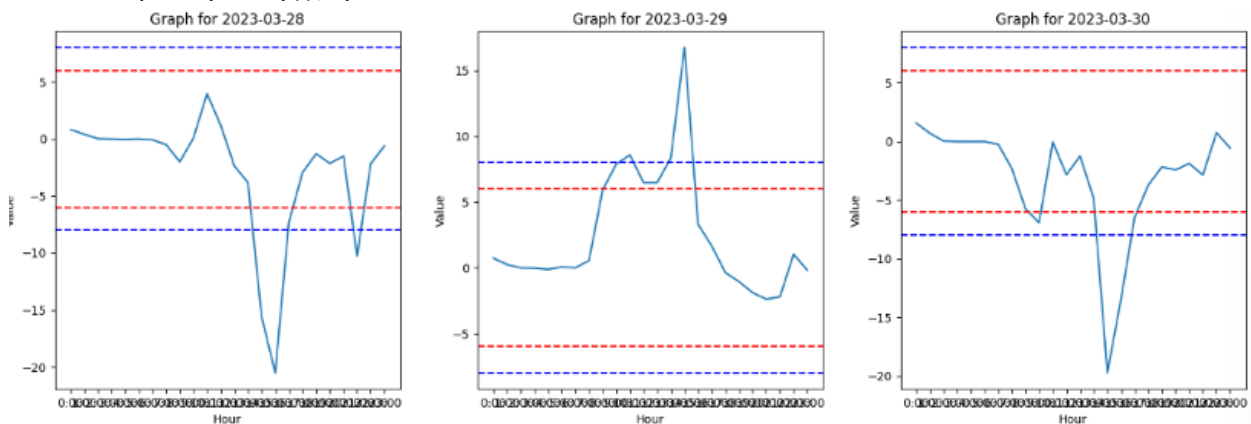
정산금을 최적화하기 위해 예측 오차율이 크게 발생하는 특정 시간대를 추출하고, 발전소별 예측 오차 값 데이터프레임을 재구조화 하였다. 이후 발전설비 용량이 큰 발전소들을 대상으로 K-means 클러스터링을 수행하여 발전소들을 유사한 특성을 갖는 그룹으로 묶었다. 서로 다른 클러스터 간에는 차이가 있다는 점을 기반으로 각 클러스터들에 포함된 발전소들 중 가장 큰 설비용량을 갖는 발전소를 각 조합의 대표 발전소로 설정하였다. 그 다음 피어슨 상관계수를 이용하여 각 조합들과 가장 작은 상관계수를 갖는 발전소들을 각 조합에 병합하여 예측 오차를 보완하는 6개의 최종 조합을 구성하였다. 그 결과 조합에 포함된 발전소들을 병합하여 예측 오차율을 계산하고, 정산금을 구하는 과정을 거쳐 총 정산금을 계산하여 이를 기반으로 분할 시행 개수를 조정하여 가장 높은 총 정산금을 갖는 최종 조합을 도출하였다.

주요 시간대 추출

데이터 탐색 과정에서 전체 발전소들이 발전량의 실측 값과 예측 오차 값이 공통적으로 특정 시간대에 크고 변동성 또한 크다는 것을 발견하였다.

[그림1]을 통해 9시부터 발전량이 크게 증가하며 오후 17:00이후로 다시 감소하는 것을 확인할 수 있다. 이를 통해 각각의 시간대가 차지하는 비율이 6% 이상이 모두 낮시간대(9:00 ~16:00)임을 파악하고, 이를 발전량이 높은 시간대로 산정하였다.

우리의 목표는 각 조합의 모든 시간대의 예측 오차율이 6%이내로 유지하는 것이므로, 이를 벗어나는 특정 시간대를 추출하여 분석을 수행하였다. 날짜 별로 전체 발전소들의 평균 예측 오차율을 계산하고, 임계 값 $\pm 6\%$ 를 초과하는 시간대를 계산하고 그 횟수를 세어 임계 값을 4회 이상 초과한 시간대를 추출하였다. .



[그림 1] 03/28 ~ 03/30 일자의 전체 발전소의 평균 예측 오차율

수행 결과 이를 통해 임계 값을 초과하는 시간대 중 4회 이상 발생한 시간들은 ['14:00', '15:00', '9:00', '13:00', '10:00', '16:00', '11:00', '12:00']로 실제 발전이 활발히 일어나는 낮 시간대에 예측 오차율이 크다는 것을 확인할 수 있었다.

1) 최어진 외 3명, 「최소 출력변동성을 갖는 가상발전소 구성을 위한 분산형 재생에너지원의 최적 조합 방법」, 『전기학회논문지』, 69, 대한전기학회, 2020, pp.1866~1872.

K-mean Clustering을 통한 조합의 개수 찾기.

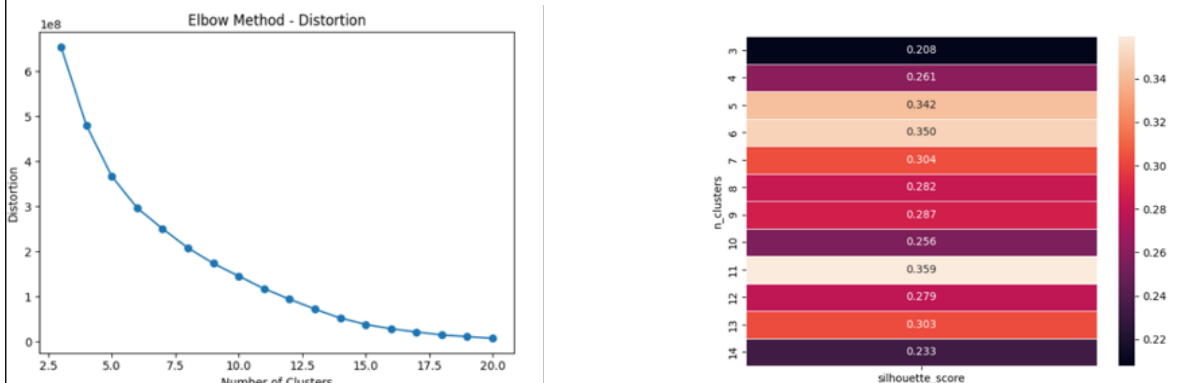
1) 데이터프레임 재구조화

각발전소별 예측 오차 값 데이터프레임(code1_diff부터code169_diff)에서 특정 시간대('14:00', '15:00', '9:00', '13:00', '10:00', '16:00', '11:00', '12:00')를 대상으로 분석을 집중하기 위해 해당 시간대 이외의 열을 제거하였다. 그 후melt() 함수를 사용하여 각 발전소의 예측 오차 값 데이터프레임을 낱 짜 열은 유지하고 시간대 열을 행으로 변환하여 통합하는 작업을 수행하였다.이를 통해diff_melt_df 데이터프레임을 생성하였다. 이후 발전소의 용량이 클수록 조합의 예측 오차율에 큰 영향을 미칠 수 있다는 가정에 따라 안정성을 확보하기 위해, 발전소 용량이2500 이하인 발전소들을 제거하여 cleaned_df 데이터프레임을 생성하였다. 최종적으로cleaned_df 데이터프레임은26개의 발전소와248개의 날짜별 시간별 값으로 구성된다. 이를 통해 특정 시간대에 대한 예측 오차율에 대한 클러스터링과 피어슨 상관관계 분석을 효율적으로 수행할 수 있다.

2) K-mean Clustering

K-means클러스터링은데이터를 비슷한 특성을 가진 그룹으로 묶는 알고리즘으로 클러스터링을 수행하면 각 클러스터는 데이터 간의 유사성에 기초하여 형성되며, 서로 다른 클러스터 간에는 차이가 있다. 이를 기반으로 재구조화한 데이터 프레임(cleaned_df)을 클러스터링 하여 특정시간대의 설비용량이 큰 발전소들의 예측 오차 값이 비슷한 발전소들이 함께 묶였다. 이를 통해 서로 다른 클러스터에 속하는 발전소들 간에 예측 오차 패턴을 비교할 수 있으며 서로 다른 클러스터 간에 차이를 이용하여 각 클러스터에서 가장 큰 설비 용량을 가진 발전소를 각각의 조합의 첫 대표 값으로 설정함으로써 첫 조합들이 서로 다른 큰 설비 용량의 발전소로 구성되도록 하였다.

3) 클러스터링 개수 선정

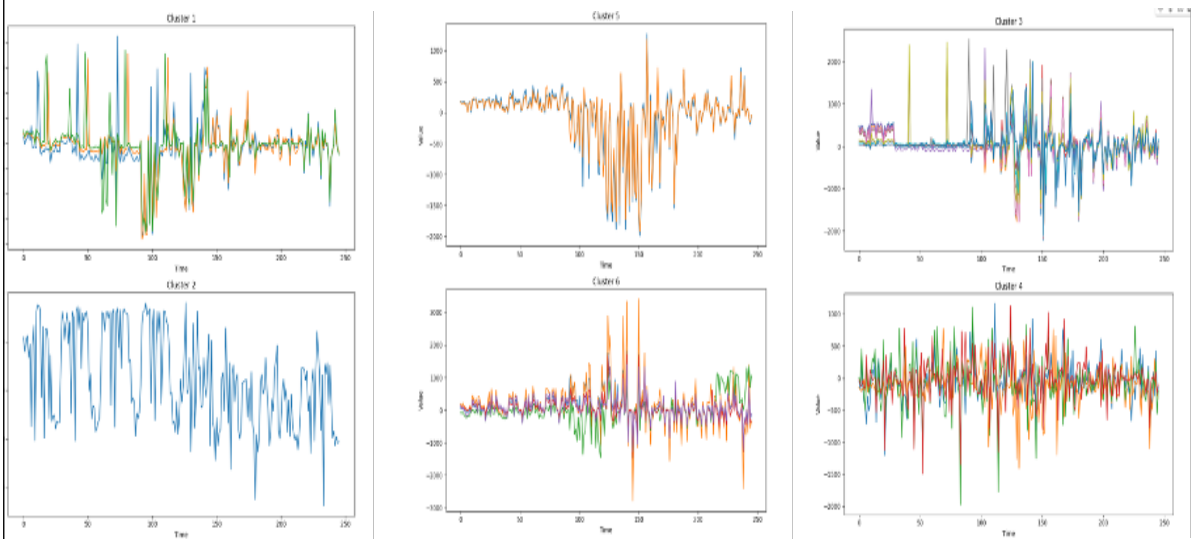


[그림 2] 엘보우 메소드 & 실루엣 점수

K-means 클러스터링을 적용하기 전에적절한 클러스터 개수를 결정하기 위해 엘보우 메서드와 실루엣 점수를 이용하였다. 엘보우 메서드는 클러스터 개수에 따른 응집도를 시각화하여 보여주며 응집도의 변화가 감소하는 속도가 급격히 줄어드는 ‘팔꿈치’인 지점을 적절한 클러스터 개수로 선택한다. 실루엣 점수는 개별 데이터 포인트의 클러스터 내 응집도와 다른 클러스터와의 분리도를 고려하여 -1~1사이의 값을 가지며 1에 가까울수록 좋은 클러스터링 결과를 의미한다.

cleaned_df 데이터프레임을 대상으로 엘보우 메소드와 실루엣 점수를 모두 고려하여클러스터 개수를 6개로 선정하였다. 이때, 실루엣 점수가 6개 보다 11개가 더 높지만, 각 조합의 발전 설비 용량이 20MW가 넘어야 한다는 조건으로 인해 6개를 선정하였다.

[그림 3] 클러스터링 결과



[그림 3] 클러스터링 결과

Cluster 1	['code26', 'code27', 'code48']
Cluster 2	['code86']
Cluster 3	['code19', 'code21', 'code22', 'code24', 'code28', 'code30', 'code31', 'code43', 'code44', 'code46', 'code47']
Cluster 4	['code5', 'code45', 'code71', 'code72']
Cluster 5	['code70', 'code84']
Cluster 6	['code34', 'code73', 'code75', 'code76', 'code85']

[표1] 각 클러스터에 속한 발전소 code

클러스터링 결과를 기반으로 각 클러스터에서 발전소 용량이 가장 큰 발전소를 6개의 조합 리스트의 첫 번째 대표 값으로 선정한 조합 리스트는 다음과 같다 :

Set_list:{1: 'code26', 2: 'code86', 3: 'code19', 4: 'code45', 5: 'code70', 6: 'code73'}

무작위로 조합을 뽑아 비교군 만들기.

상관계수를 이용한 방법의 실효성을 검증하고자 비교군이 필요했다. 이를 위해, 설비용량 20MW를 넘기며, 6개의 조합을 만드는 알고리즘을 개발하였고 이를 1000번 실행한 결과 최대 정산금은 86178811.55원이었다. 아래는 최대 정산금을 보이는 무작위 조합이다.

Vpp1	['S5005', 'V9025', 'B1741', 'S5019', 'S5034', 'B1784', 'B1729', 'S5001', 'M1201', 'B1788', 'M1205', 'B1789', 'B1778', 'S5010', 'B1786', 'S5048', 'S5035', 'G1020', 'B1776', 'S5018', 'B1738', 'S5014', 'B1780', 'M1208', 'N6002', 'B1732', 'N6005', 'S5007', 'V9007', 'V9006', 'G1056', 'V9012', 'B1733', 'G1033', 'N6007', 'B1782', 'V9020', 'N6009', 'G1031', 'S5039']
Vpp2	['N6003', 'V9016', 'S5028', 'V9008', 'B1728', 'G1066', 'V9010', 'G1057', 'V9004', 'S5032', 'S5036', 'G1062', 'S5017', 'G1022', 'G1030', 'G1019', 'S5040', 'N6001', 'V9013', 'S5037', 'G1028']
Vpp3	['S5046', 'G1051', 'S5041', 'V9003', 'V9001', 'V9026', 'S5031', 'G1029', 'G1034', 'G1061', 'V9027', 'S5033', 'G1058', 'S5027', 'S5025', 'M1202', 'V9017', 'S5009', 'G1055', 'G1017', 'G1023', 'S5006', 'V9022', 'G1025', 'S5021', 'B1740', 'S5045', 'M1204', 'N6004']
Vpp4	['B1794', 'G1052', 'B1777', 'V9023', 'B1795', 'V9018', 'S5044', 'B1792', 'G1050', 'V9005', 'S5013', 'B1783', 'V9019', 'S5002', 'S5004', 'B1781', 'G1053', 'M1206', 'G1067', 'B1790', 'V9002', 'G1068', 'G1059', 'G1065', 'S5029', 'S5016', 'V9011', 'G1039', 'B1743', 'G1064', 'V9009', 'B1785', 'G1018', 'B1793', 'S5030', 'S5050', 'S5011', 'B1727', 'N6006', 'B1787', 'S5015', 'V9021']
Vpp5	['S5023', 'N6008', 'B1731', 'G1024', 'S5003', 'G1063', 'B1734', 'S5047', 'G1060', 'B1730', 'S5020', 'G1021', 'G1026', 'S5038', 'B1736', 'B1742', 'B1737', 'S5012', 'S5026', 'S5008', 'G1027', 'B1779']
Vpp6	['B1735', 'B1728', 'S5022', 'G1059', 'S5042', 'S5028', 'N6001', 'B1779', 'S5043', 'S5004', 'S5007', 'S5016', 'G1030', 'G1065', 'S5044', 'B1792', 'S5030', 'S5010', 'S5002', 'S5049', 'S5021', 'S5026', 'S5033', 'M1206', 'G1023', 'G1017', 'G1019'] ₁₂ -

켄달타우 상관계수를이용한 조합 구성.

Kendal Tau 상관계수는 두 변수들 간의 순위를 비교하여 연관성을 계산한다. C는 각 변수의 비교대상의 상하관계가 같은 것들의 수, D는 상하관계가 다른 것들의 수이다.

$$p = \frac{C - D}{C + D}$$

그림 1kendal tau 상관계수 공식

예를 들어 아래 표와 같을 때 C는 3개, D는 2개 이므로 계수는 0.2이다.

A 사람	B 사람	결과
1	3	C
7	-1	D
-2	-3	C
-1	2	C
-3	9	D

해당 데이터에서는 같은 날짜, 같은 시간에 두 발전소의 각각의 오차가 양수인지 음수인지를 따져 계산한다. 해당 알고리즘을 기반으로 정산금을 구한 결과 87176225.9원이 나왔다. 랜덤하게 구한 정산금보다 높았으나 해당 알고리즘은 두 데이터의 양수, 음수만 판별하기 때문에 값의 크기는 반영하기 어렵다는 문제가 있었다.

Vpp1	['B1735', 'M1207', 'N6003', 'S5012', 'S5013', 'G1065', 'S5019', 'S5029', 'S5025', 'V9027']
Vpp2	['B1736', 'S5044', 'B1782', 'N6005', 'V9005', 'B1785', 'B1784', 'S5031', 'S5049', 'G1050', 'G1056', 'G1057', 'V9016', 'V9002', 'G1068', 'G1067', 'S5001', 'G1054', 'G1051', 'S5050', 'S5046', 'V9009', 'S5015', 'G1021', 'S5020', 'B1786', 'B1739']
Vpp3	['M1208', 'N6004', 'B1731', 'B1733', 'B1783', 'N6006', 'B1728', 'V9003', 'M1202', 'B1793', 'B1743', 'V9015', 'G1055', 'G1061', 'B1778', 'V9025', 'G1023', 'S5039', 'B1730', 'S5041', 'V9007', 'S5010', 'S5016', 'S5033', 'G1017', 'S5030', 'S5026', 'B1790', 'G1066', 'S5038', 'S5028', 'S5032', 'S5027', 'S5007', 'S5043', 'B1788', 'B1741', 'B1738', 'B1742']
Vpp4	['S5034', 'M1201', 'B1792', 'N6002', 'N6001', 'V9004', 'V9023', 'V9008', 'B1794', 'B1727', 'B1776', 'V9021', 'G1033', 'M1206', 'V9018', 'V9017', 'V9014', 'V9013', 'V9012', 'G1064', 'B1791', 'G1027', 'G1028', 'G1031', 'V9022', 'V9024', 'V9026', 'B1732', 'B1777', 'S5011', 'S5040', 'S5047', 'G1020', 'S5006', 'G1030', 'G1026', 'S5018', 'S5009', 'S5002', 'G1019', 'S5014', 'G1022', 'S5004', 'G1053', 'B1787', 'B1740', 'B1737']
Vpp5	['N6007', 'N6008', 'S5036', 'M1203', 'M1204', 'B1729', 'S5037', 'V9010', 'S5035', 'V9006', 'S5048', 'G1052', 'V9019', 'B1789', 'G1062', 'S5042', 'G1024', 'G1063', 'S5024', 'S5017', 'G1039', 'G1034', 'G1018']
Vpp6	['S5045', 'N6009', 'M1205', 'V9020', 'G1058', 'V9001', 'B1795', 'G1029', 'S5023', 'S5003', 'S5008', 'S5022', 'G1059', 'B1734', 'S5005', 'G1060', 'G1025', 'V9011', 'B1780', 'B1779', 'B1781', 'S5021', 'G1032']

피어슨 상관계수를 이용한 조합 구성

피어슨 상관관계는 두 변수간의 선형 관계의 강도와 방향을 나타내는 통계적 지표로, 주어진 데이터에서 두 변수가 얼마나 밀접하게 관련되어 있는지를 평가하는데 사용할 수 있다. 피어슨 상관계수는 -1~1의 범위를 가지며 -1에 가까운 값을 가질수록 강한 음의 성향 관계가 있음을 나타낸다. 이에 근거하여 각 조합들과 발전소들의 상관계수를 구하고, 각 조합과 가장 음의 상관관계를 가지는 발전소들을 병합시키며 조합 내의 발전소들이 상호 보완적인 역할을 수행하여 예측 오차를 최소화하도록 유도하였다.

먼저, 모든 발전소의 특정시간대의 예측 오차 값의 데이터 프레임인 `diff_melt_df`에서 각 조합(set)들과 가장 작은 피어슨 상관계수를 갖는 발전소를 추출하고, 해당 발전소들을 조합에 병합하는 과정을 반복하였다. 이때, 모든 조합들의 발전소 용량 총합이 20MW를 초과해야 하므로 이에 대한 안정성을 확보하기 위해, 각 조합에 대해 가장 작은 피어슨 상관계수를 갖는 발전소를 동시에 찾고 6개의 조합에 각각 1개씩 6개의 발전소를 병합하는 과정을 19~25번 반복하였다. 이 결과 6개의 조합은 모두 동일한 개수의 발전소들이 병합된 상태이다. 이후 보다 아직 병합되지 않은 남은 발전소들은 세밀 조정을 위해 전체 피어슨 상관계수 매트릭스에서 조합들(set1~set6)과 발전소 중 가장 작은 값을 추출하여 해당 조합에 하나씩 병합하였다.

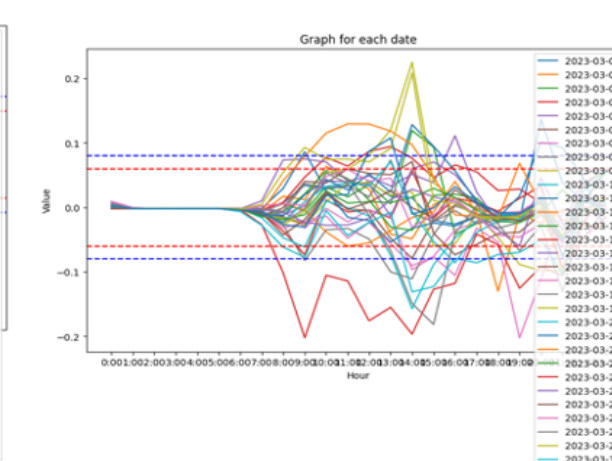
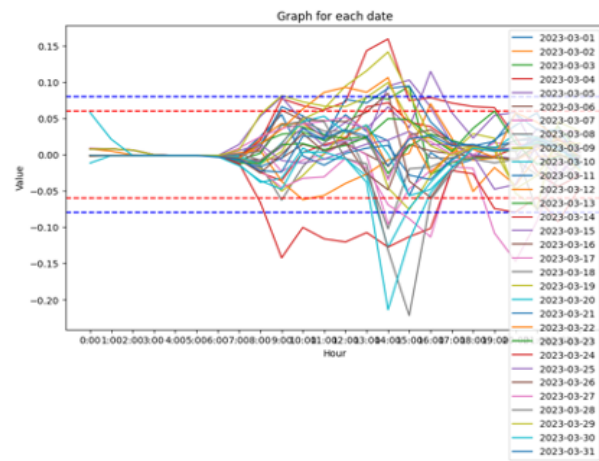
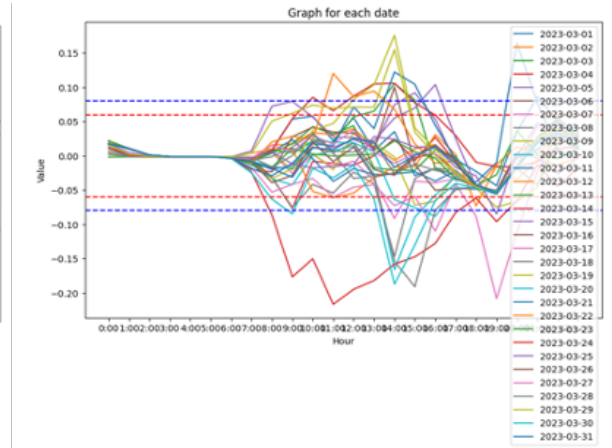
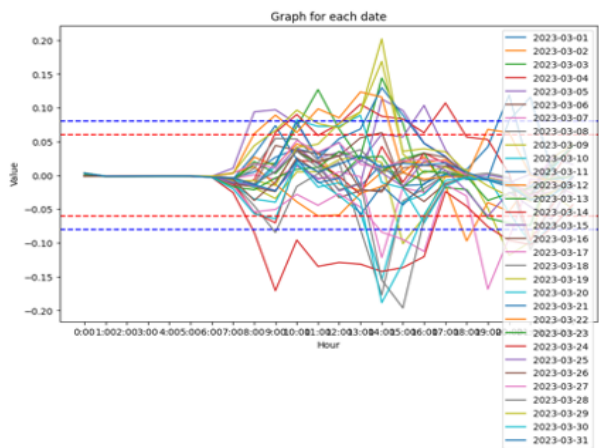
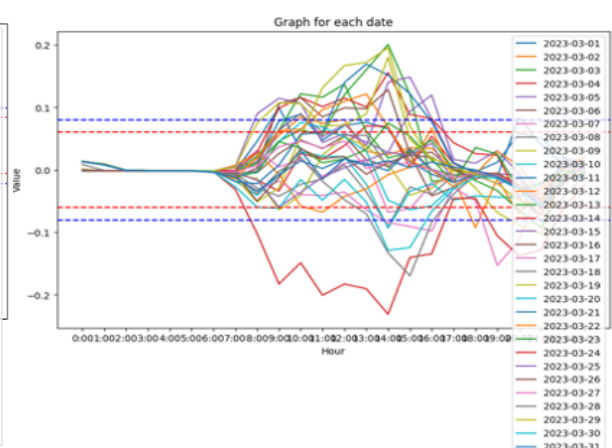
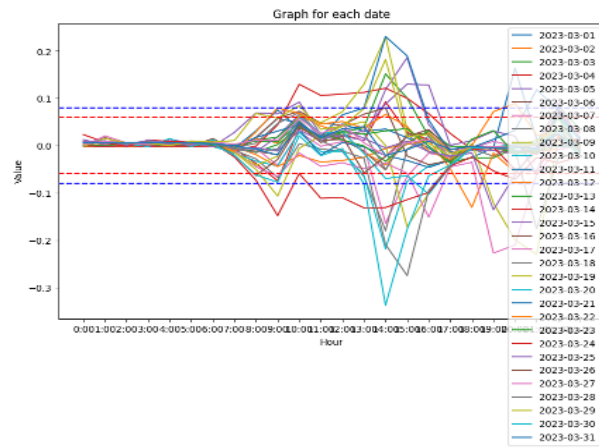
6. 정산금 계산

구성된 조합에 속하는 발전소들을병합하여 날짜별, 시간대별 예측 오차율을 계산하였다. 정산금은 예측 오차율이6% 이하인 경우에는 실제 발전량에4원을 곱하고, 6% 초과8% 이하인 경우에는 실제 발전량에3원을 곱하여 계산하였다.이를통해 일별 정산금을 계산하고, 31일치의 일별 정산금을 합산하여 총 정산금을 구하는 함수 code_money를 통해 조합들의 결과를 비교하였다.

분할 시행 개수 (동시비동시)	총 정산금
19 / 43	86705176.31
20/37	87170257.91
21/31	87198754.83
22/25	87868972.89
23/19	87446090.76
24/13	87376446.62
25/7	87401370.27

[표2] 분할 시행 개수에 따른 총 정산금

분석 결과 각각의 조합들에 대해 가장 음의 상관계수를 갖는 발전소들을동시에 병합해주는 방식을22번 진행하고, 가장 낮은 상관계수를 가지는 조합과 발전소를 찾아 해당 조합에 하나씩 병합하는 방식을 25번 실행했을 때, 가장 높은 총 정산금을 얻었다.



각조합의 날짜 별 시간대의 오차율 그래프를 통해 조합의 성능을 시각화 하였다. 이를 통해 각 조합에 속한 발전소들을 하나의 가상 발전소로 묶음으로써 오차 값이 서로 상호 보완되어 낮아진 것을 확인할 수 있었다.

이를기준으로 최종적으로 도출된 조합은 다음과 같다:

[용량(kw): 소수점 두 자리 수 올림]

발전그룹 code	발전소 코드	발전원	용량(kw)
Vpp1	['N6007', 'V9026', 'S5035', 'G1067', 'N6008', 'B1783', 'V9009', 'G1066', 'B1736', 'G1052', 'G1033', 'S5032', 'B1738', 'B1737', 'B1788', 'S5027', 'B1787', 'V9007', 'B1730', 'S5005', 'S5018', 'B1734', 'V9011', 'G1039', 'S5019', 'B1727', 'B1732', 'M1205', 'M1204', 'M1203']	SOLAR	35421.76
Vpp2	['S5034', 'B1742', 'N6003', 'B1794', 'B1777', 'B1731', 'V9025', 'G1060', 'G1051', 'S5006', 'S5038', 'S5014', 'G1028', 'M1202', 'S5039', 'B1740', 'V9012', 'S5046', 'G1064', 'S5012', 'M1201', 'V9016', 'S5013', 'G1062', 'B1784', 'V9005', 'B1729', 'V9004', 'G1058', 'V9002', 'V9001', 'G1022', 'G1018', 'G1020', 'V9003']	SOLAR	42343.41
Vpp3	['M1208', 'G1027', 'B1795', 'V9010', 'S5003', 'S5023', 'N6002', 'V9021', 'B1780', 'V9020', 'B1793', 'G1029', 'G1034', 'B1776', 'S5048', 'G1053', 'S5009', 'S5029', 'G1031', 'S5020', 'S5036', 'V9019', 'G1063', 'V9018']	SOLAR	28603.32
Vpp4	['S5045', 'B1741', 'S5024', 'B1739', 'V9022', 'S5008', 'S5037', 'B1778', 'N6005', 'S5025', 'G1032', 'B1743', 'V9014', 'N6004', 'V9023', 'B1789', 'G1026', 'S5041', 'S5011', 'V9013', 'G1025', 'B1733', 'S5047', 'M1207', 'B1791', 'G1056', 'G1061', 'G1057', 'V9015']	SOLAR	35110.73
Vpp5	['B1782', 'G1055', 'V9024', 'S5031', 'B1781', 'G1024', 'G1054', 'G1068', 'B1790', 'S5015', 'G1050', 'G1021', 'N6006', 'B1786', 'V9006', 'S5040', 'S5001', 'B1785', 'V9027', 'S5050', 'V9008', 'N6009', 'V9017', 'S5017']	SOLAR	27033.03

Vpp6	['B1735', 'B1728', 'S5022', 'G1059', 'S5042', 'S5028', 'N6001', 'B1779', 'S5043', 'S5004', 'S5007', 'S5016', 'G1030', 'G1065', 'S5044', 'B1792', 'S5030', 'S5010', 'S5002', 'S5049', 'S5021', 'S5026', 'S5033', 'M1206', 'G1023', 'G1017', 'G1019']	SOLAR	29287.38
------	---	-------	----------

4. 결론

무작위로 뽑은 조합, 켄달 타우 상관계수, 피어슨 상관계수를 이용한 결과 피어슨 상관계수를 이용했을 때 가장 큰 정산금이 나왔다. 즉, 피어슨 상관계수를 이용한 방법을 통해 조합을 구할 수 있다.

정산금(원)	랜덤	켄달 타우	피어슨
	86231246.77	87176225.91	87868972.89

- 1) 최어진 외 3명, 「최소 출력변동성을 갖는 가상발전소 구성을 위한 분산형 재생에너지원의 최적 조합 방법」, 『전기학회논문지』, 69, 대한전기학회, 2020, pp.1866~1872.

4 활용데이터

분석 과제 데이터 : 2023-03-01 ~ 2023-03-31 의 169개의 31일치 실제 발전량과 예측량이며 발전소 코드는 비식별화 코드이며, 태양광 에너지 발전량 및 예측량에 대한 데이터

5 사업화방안 및 기대효과

분석을 통해 도출된 특정 날짜에 관계없이 안정적으로 낮은 예측 오차율을 제공하는 조합 모델(가상발전소)은 전력 거래의 안정성과 효율성을 향상시킬 수 있다.

소규모 집합 자원을 생산하는 소규모 집합 자원들을 모아 가상발전소를 구성하여 거래하는 중개 사업자는 정확한 예측을 통해 예측 정산금을 높일 수 있으며, 전력거래소는 다수의 소규모 발전사업자와 거래하는 대신 중개 사업자를 통해 분산된 전력 자원을 효율적으로 관리할 수 있게 되며 국가가 안정적인 전력망 운영을 보장할 수 있도록 돕는다.

도출한 가상발전소 조합 모델 알고리즘을 기반으로 전력 거래 중개 사업자 및 거래 플랫폼을 구축하여 사업화 할 수 있다. 소규모 발전사업자들과 전력거래소 간의 중개 역할을 하여 정확한 예측을 통해 수익을 얻을 수 있다. 더 나아가 알고리즘을 확장해 정부와 파트너십을 구축함으로써 우리나라의 전체 재생에너지 발전소들의 예측 오차를 조사하고, 이들 간의 상관관계를 분석하여 국가 차원에서 지속 가능한 에너지로의 전환을 위한 환경을 조성하는데 기여할 수 있다.

이러한 가상발전소 조합 모델은 재생에너지의 발전 효율을 향상시키고 에너지 생산 비용을 줄이는 데 기여하며, 재생에너지 생산을 확대할 수 있는 잠재력을 가지고 있다. 이를 통해 기존의 화석 연료 발전을 대체함으로써 온실가스 감축과 환경 보호에 기여할 수 있으며, 탄소중립을 실현하는데 중요한 역할을 할 것으로 기대된다. 빅데이터 분석을 활용한 가상발전소 조합 모델은 에너지 산업의 혁신과 탄소중립을 위한 핵심적인 기술로써 매우 중요하며, 전력 시장의 안정성과 지속가능성을 향상시키는 데 큰 기여를 할 것이다.

※ 분량제한은 없으며, 공모요강에 적시된 평가항목을 참고하여 작성하여 주시기 바람 (상세 설명을 위해 도표, 스케치 등 별도파일 추가 가능)