Optimal Location Analysis for Solar Power Plants Using AI

Hanyang Univ. Sunwoong Shin
Prof. Jaeweon Jeong

Contents


Chaeryŏng-ŭp
Kosong
Sokch'o
연천군
Ongjin
Gapyeong
Hongch'ŏn
Seoul
Wabu
Inchon
Yangp'yŏng
Pubal
Kwangju
Ansŏng
Sŏsan
Chinch'ŏn
Yŏngju
Asan
Mungyeong
Kongju
Sangju
옥천군
Poryŏng
Daegu
Yŏngch'ŏn
Wanju
Hwawŏn
Changnyeong
Koch'ang
Namwŏn
Ungsang
Naesŏ
Hwasun
Pusan
Kwangyang
Sinhyŏn
Sinan
Kŏje
Korea Strait

1. INTRODUCTION

(1) Background

Limitation of Previous Research

Research Site	There is no way to bring vast amounts of data, so most are limited to one province areas.
Factors	Some study excludes important factors that affect actual solar power generation.
Correlation	Rather than factors independently affecting the amount of power generation, close relationships between factors affect the amount of power generation. However, almost all studies have overlooked this point.

1. INTRODUCTION

(2) Research scope and method

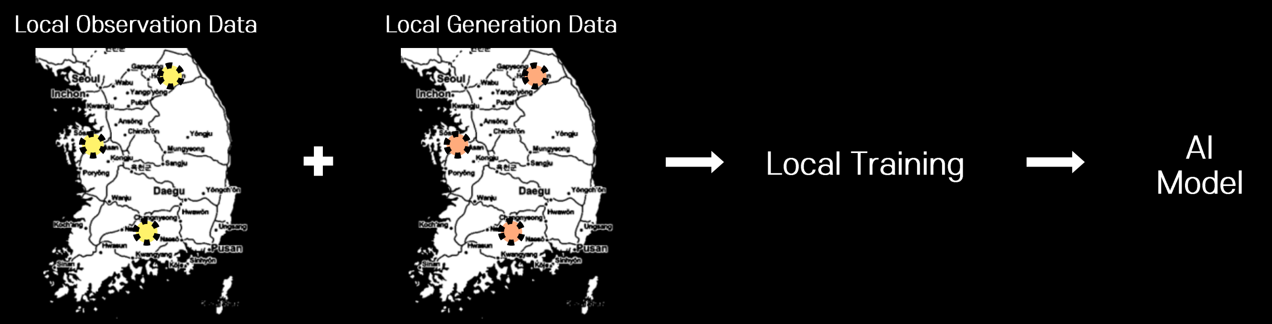
Local Area

In this study, “Local” represents certain regions where we can get not only observation data (features) but also solar power generation data (target). Through machine learning techniques, we can get AI model which will be used to predict global (unknown) solar power generation data.

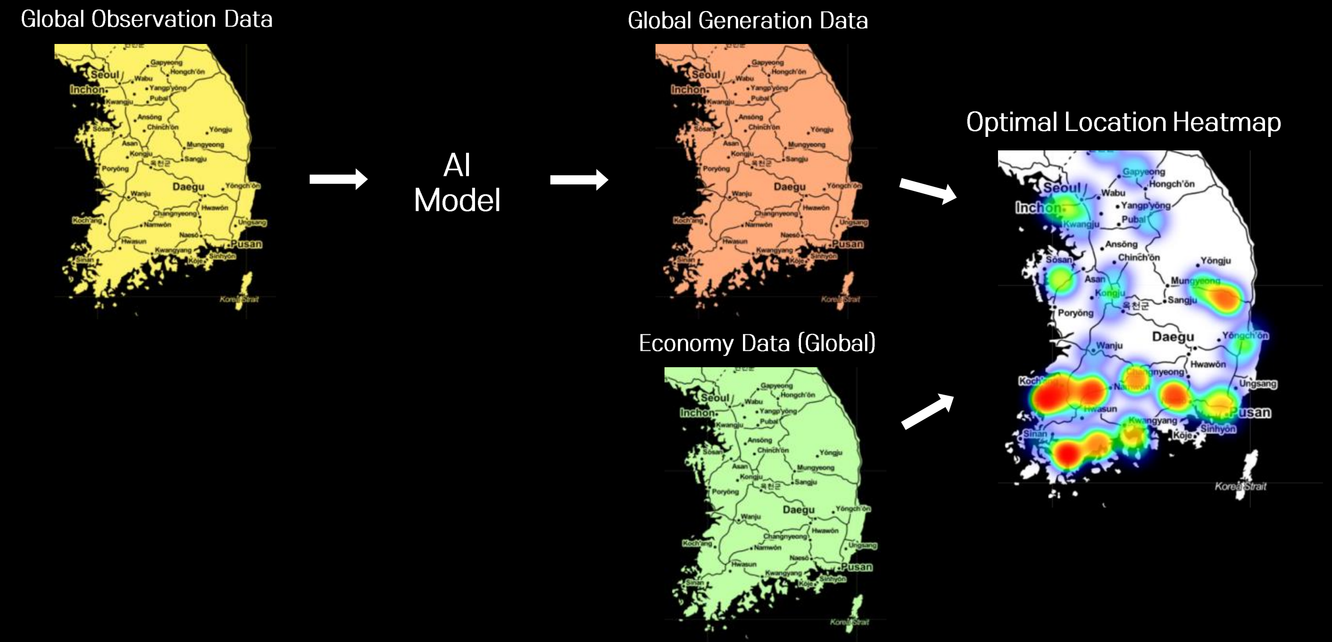
Global Area

In this study, “Global” represents unknown regions where we can get just observation data (features). By using AI model that we made in Local Training, we could predict solar power generation data which we never knew.

Local Area

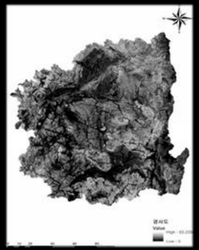


Global Area



2. RELATED WORK

(1) Geographic Conditions



(a) Slope

When installing a solar power plant, it should be installed in an area with a slope of less than 10 degrees (Kwon et al. 2008).

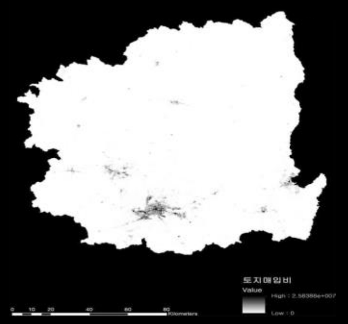
(b) Aspect

The results of the in-depth analysis results, the south, southeast, southwest, and flat terrain is good for sunlight.

(c) Hillshade

In the case of the shading corridor, Azimuth was applied at 230° and Altitude at 10° as of the winter solstice day.

(2) Economic Conditions



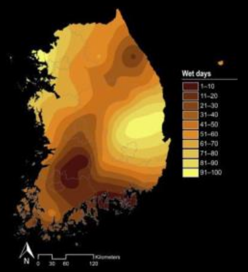
(a) Economic factors of solar power plants

Economic factors can also be important factors in selecting the optimal location for solar power plants.

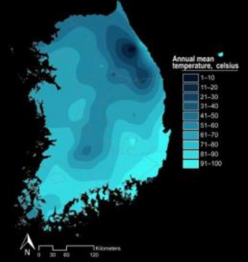
In previous study, land purchase costs were expressed using the Kriging interpolation technique using Biz-gis' GIS DB branch data (as of 2011) based on the official land price standarad of the Ministry of Land, Infrastructure and Transport.

However, since this paper was made only in Daegu, the land costs were all similar, so economy couldn't an effective factor in most previous research.

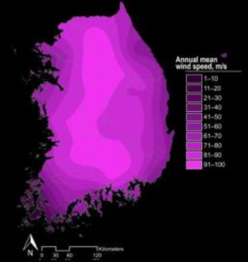
(3) Weather Conditions



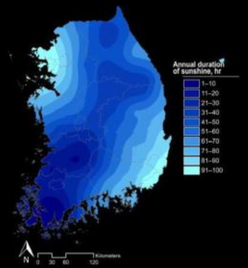
(a) Wet days



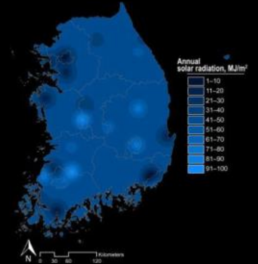
(b) Annual mean temperature



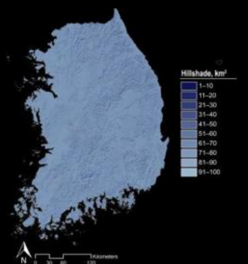
(c) Annual mean wind speed



(d) Annual duration of Sunshine



(e) Annual solar radiation



(f) Hillshade

The most important thing in solar power plants is climate.

The factors commonly used in previous studies are solar radiation, mean temperature, wind speed, duration of sunshine, humidity.

We could see how these factors correlate with solar power generation, but we did not know the close relationship between each factor.

3. DATA ANALYSIS

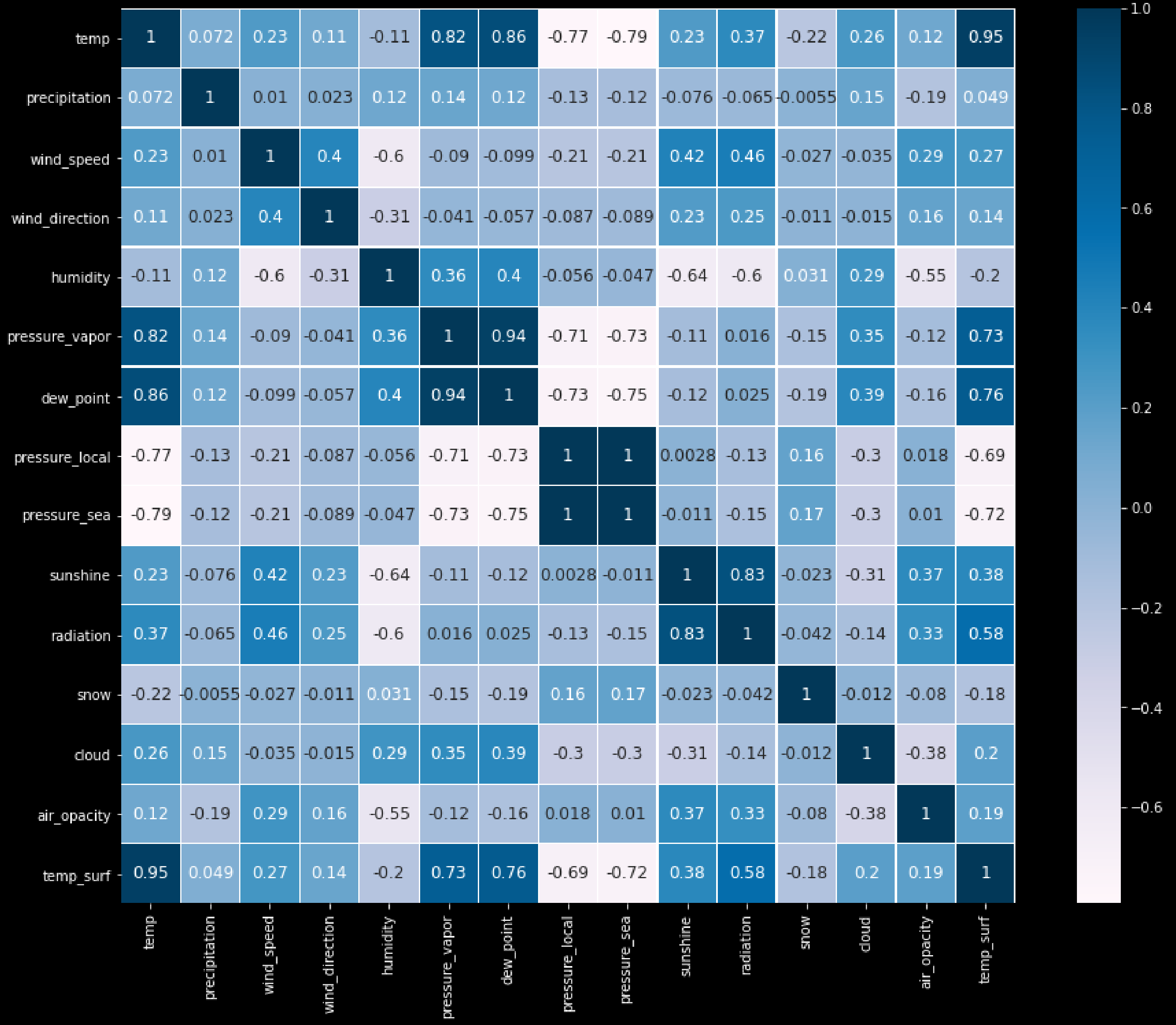
(1) Weather Observation Data

Correlation of weather observation features

Correlation Analysis is a technique for analyzing whether there is an alignment relationship between two variables. The correlation coefficient resulting from the correlation analysis is a number with a range of ”-1 to +1”, representing the linearity of the two variables.

A large correlation coefficient means that the linearity between the two variables is very high and a low correlation coefficient means that the linearity between the two variables is weak.

Correlation analysis of the variables confirms that air pressure and sea pressure, vapor pressure and dew point, air temperature and surface temperature come close to 1, which means that they are almost identical and therefore do not need to be regressed between them.



3. DATA ANALYSIS

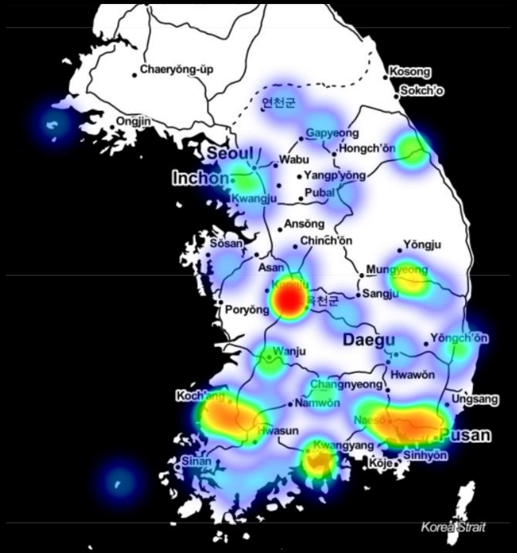
(1) Weather Observation Data

Heatmap using sampled variables

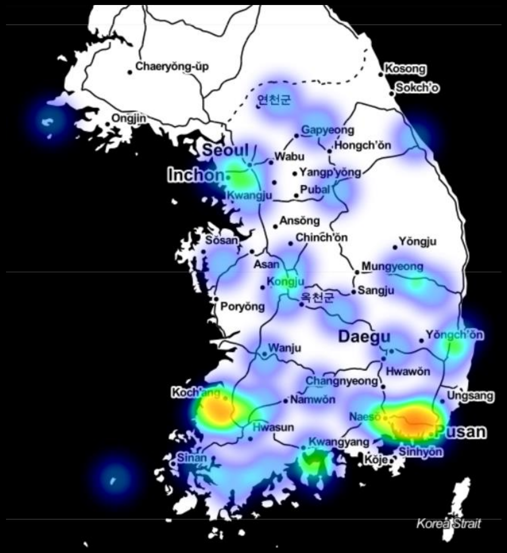
Heatmap is a representation of data in the form of a map or diagram in which data values are represented as colors.

We made a heatmap by sampling variables that are more likely to affect power generation relatively much. When analyzing the solar radiation, temperature, and visibility heatmap, it is predicted that the power generation in the south will be greater than in the north.

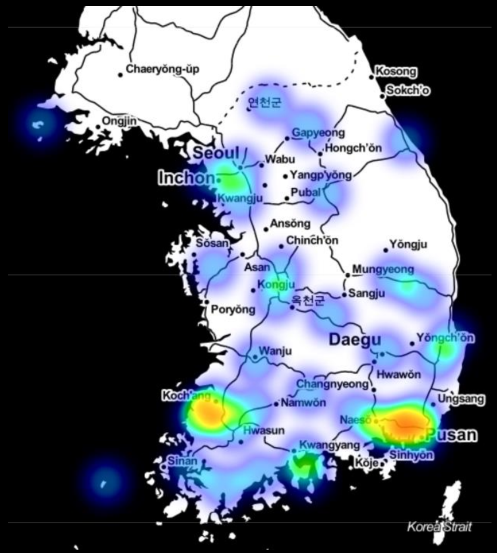
As the correlation analysis suggests, the heatmap of air temperature heatmap and surface temperature heatmap is almost identical. Since the region with zero solar radiation data was removed and visualized, there are many areas with no data at all.



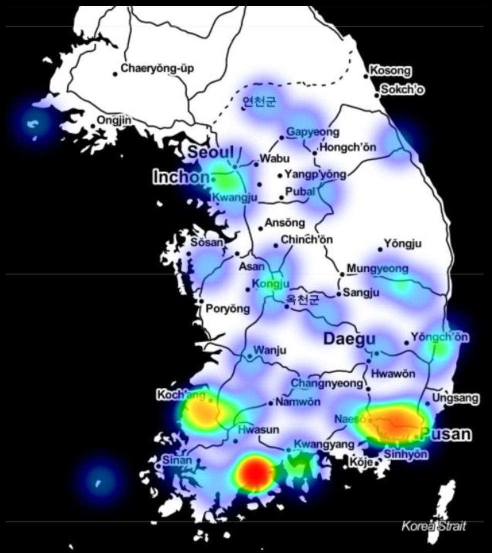
(a) radiation



(b) temperature



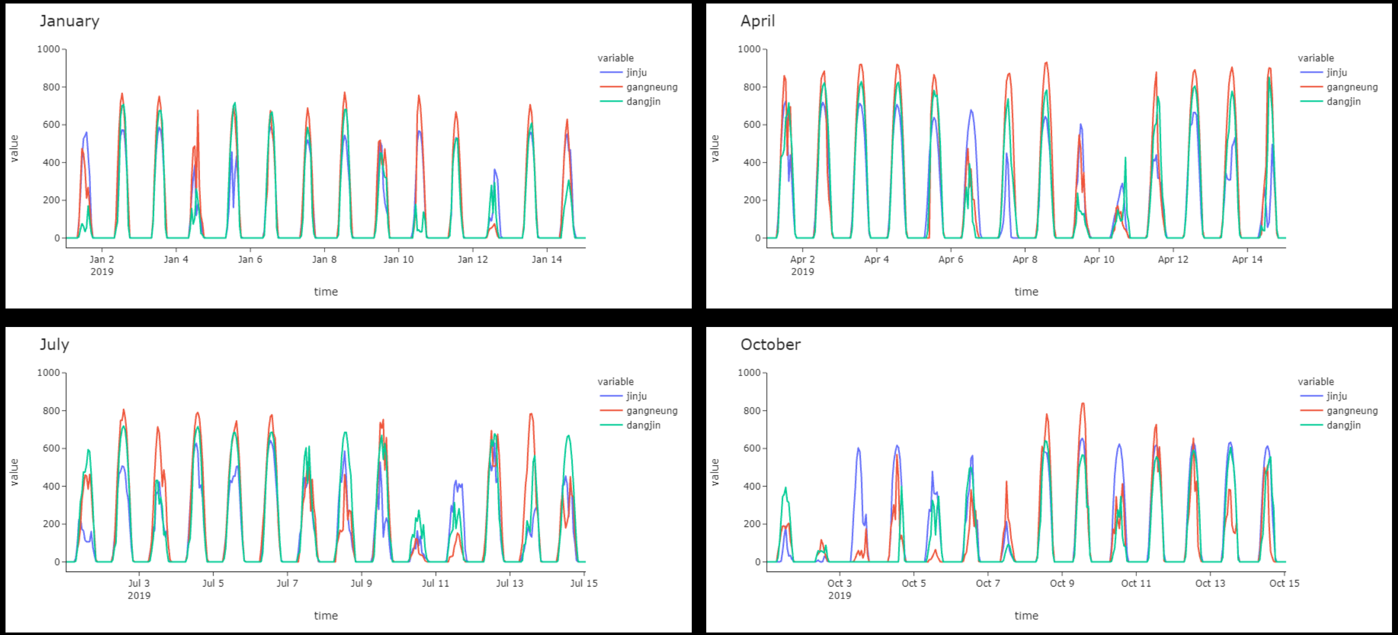
(c) surface temperature



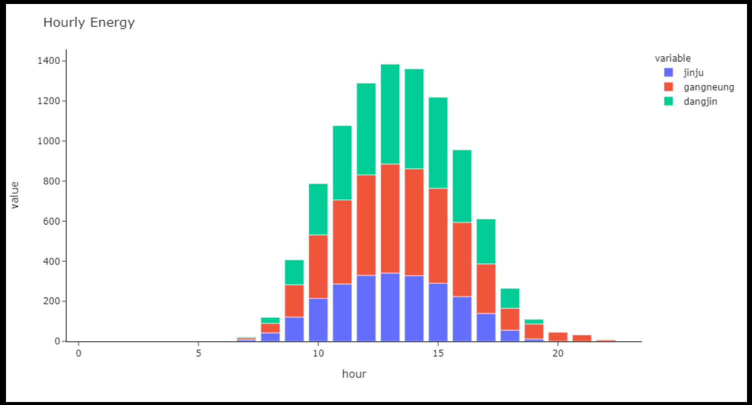
(d) air opacity

3. DATA ANALYSIS

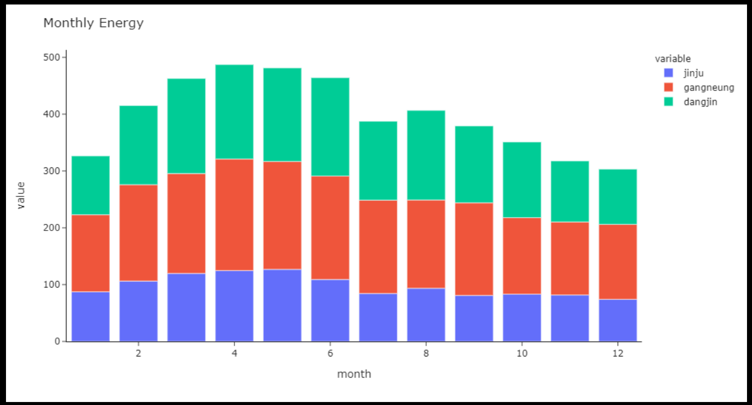
(2) Solar Power Generation Data



Target values are needed for model learning, and solar power generation data could be easily received from public data portals. However, there were constraints: there should not be many missing values, and the weather data provided by the Korea Weather Station should match the region. Power generation data exploratory data analysis was essential, and insight was sought through various visualizations. By visualizing amount of power generation per month, data trend of amount of power generation analyzed. It was possible to confirm that regional power generation was independent, and different aspects could be seen every day.



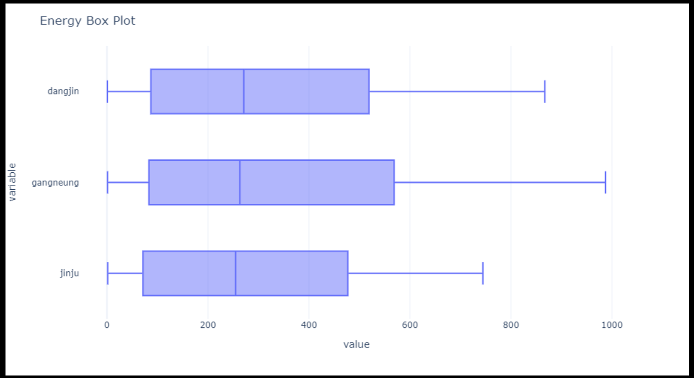
Hourly power generation was highest between 1:00 and 2:00, and before 7:00 a.m. and after 8:00 p.m., power generation converged to zero.

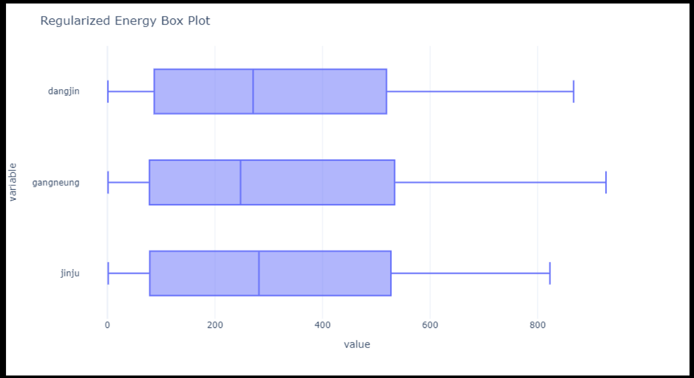


After analyzing the monthly power generation, we expected that there would be more power generation in July and August, when solar radiation was the highest, but 11 we could see that power generation was higher in April and May.

3. DATA ANALYSIS

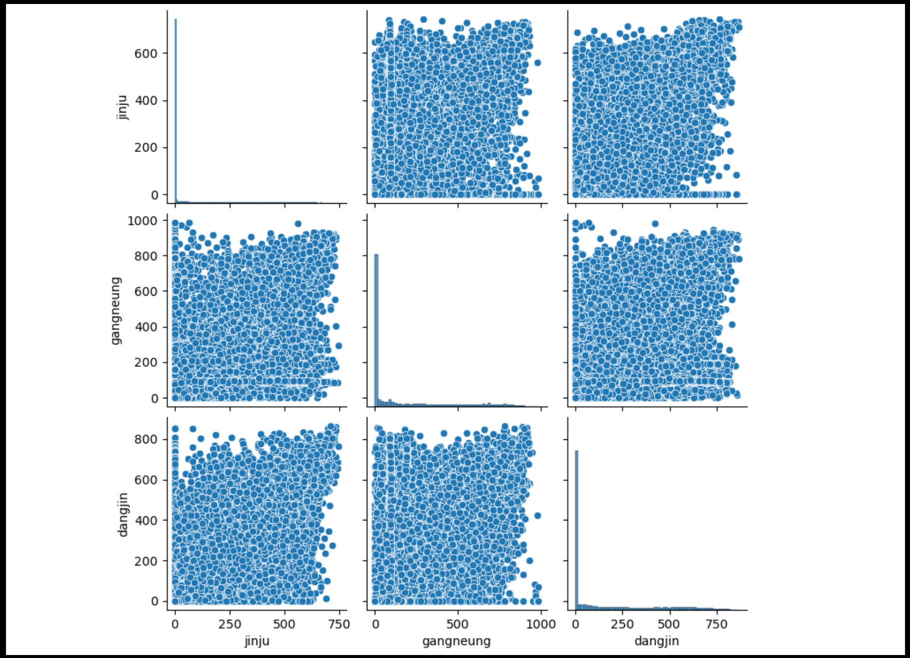
(2) Solar Power Generation Data





We drew a box plot of solar power generation in each region. If you see the left image, the mean and variance were different. But we can see surprising result from using “Normalization”.

After normalizing to the maximum solar power generation (Jinju: 905MW, Gangneung: 1065MW, Dangjin: 1000MW), the mean and variance of each region were almost identical. Judging from this, we could see that the maximum power generation of each solar power plant is linearly proportional.



Pairplot shows us grid of Axes such that each numeric variable in data will by shared across the y-axes across a single row and the x-axes across a single column.

The diagonal plots are treated differently: a univariate distribution plot is drawn to show the marginal distribution of the data in each column. By analyzing pairplot, each generation has regional independence, so we could conclude that it’s suitable for model training.

3. DATA ANALYSIS

(3) Economic Data

	PNU	BASE_YEAR	STDMT	PNILP	PJJI_YN	PANN_YMD
0	1111010600100070038	2021	1	7460000	1	20210531
1	1111010600100070040	2021	1	2008000	0	20210531
2	1111010600100070045	2021	1	4535000	0	20210531
3	1111010600100080000	2021	1	4916000	0	20210531
4	1111010600100090001	2021	1	4820000	0	20210531
...	...	...	...	...	...	...
909146	1174010800104370032	2021	1	4375000	0	20210531
909147	1174010800104370033	2021	1	4375000	0	20210531
909148	1174010800104370034	2021	1	1364000	0	20210531
909149	1174010800104370036	2021	1	5748000	0	20210531
909150	1174010800104380008	2021	1	5244000	0	20210531

909151 rows × 8 columns

	법정동코드	시도명	시군구명	읍면동명	동리명	생성일자	말소일자
0	1100000000	서울특별시	NaN	NaN	NaN	19880423	NaN
1	1111000000	서울특별시	종로구	NaN	NaN	19880423	NaN
2	1111010100	서울특별시	종로구	청운동	NaN	19880423	NaN
3	1111010200	서울특별시	종로구	신교동	NaN	19880423	NaN
4	1111010300	서울특별시	종로구	궁정동	NaN	19880423	NaN
...	...	...	...	...	...	...	...
20558	5013032022	제주특별자치도	서귀포시	표선면	하천리	20060701	NaN
20559	5013032023	제주특별자치도	서귀포시	표선면	성읍리	20060701	NaN
20560	5013032024	제주특별자치도	서귀포시	표선면	가시리	20060701	NaN
20561	5013032025	제주특별자치도	서귀포시	표선면	세화리	20060701	NaN
20562	5013032026	제주특별자치도	서귀포시	표선면	토산리	20060701	NaN

20563 rows × 7 columns

Official Land Price Data

The publicly announced land price is a system in which the Ministry of Land, Infrastructure and Transport of the Republic of Korea investigates and appraises the price of land and discloses it.

After adding all the prices included in the same region, the average land price in the region was calculated by dividing it by the number of each data.

Legal Code Data

Legal dong means dong determined by law, as stated in the administrative district unit designated by law.

This data includes PNU code and region name. Thus, we can merge this data with official land price data. By doing that, we can have “Economy Data” which is equal to the right image.

Economy Data

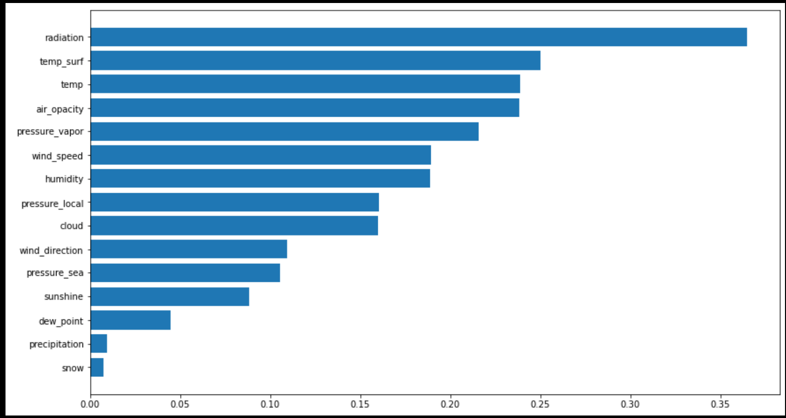
	PNU	price	loc_name
0	1111	5789.0	Jongno
1	1117	6553.0	Yongsan
2	1120	4353.0	Seongdong
3	1121	4157.0	Gwangjin
4	1123	3587.0	Dongdaemun
...	...	...	...
256	4889	20.0	Gyeongsangnamdo
257	4889	20.0	Hapcheon
258	5011	223.0	Jeju
259	5013	137.0	Jeju
260	5013	137.0	Seogwipo

261 rows × 3 columns

4. MODELING

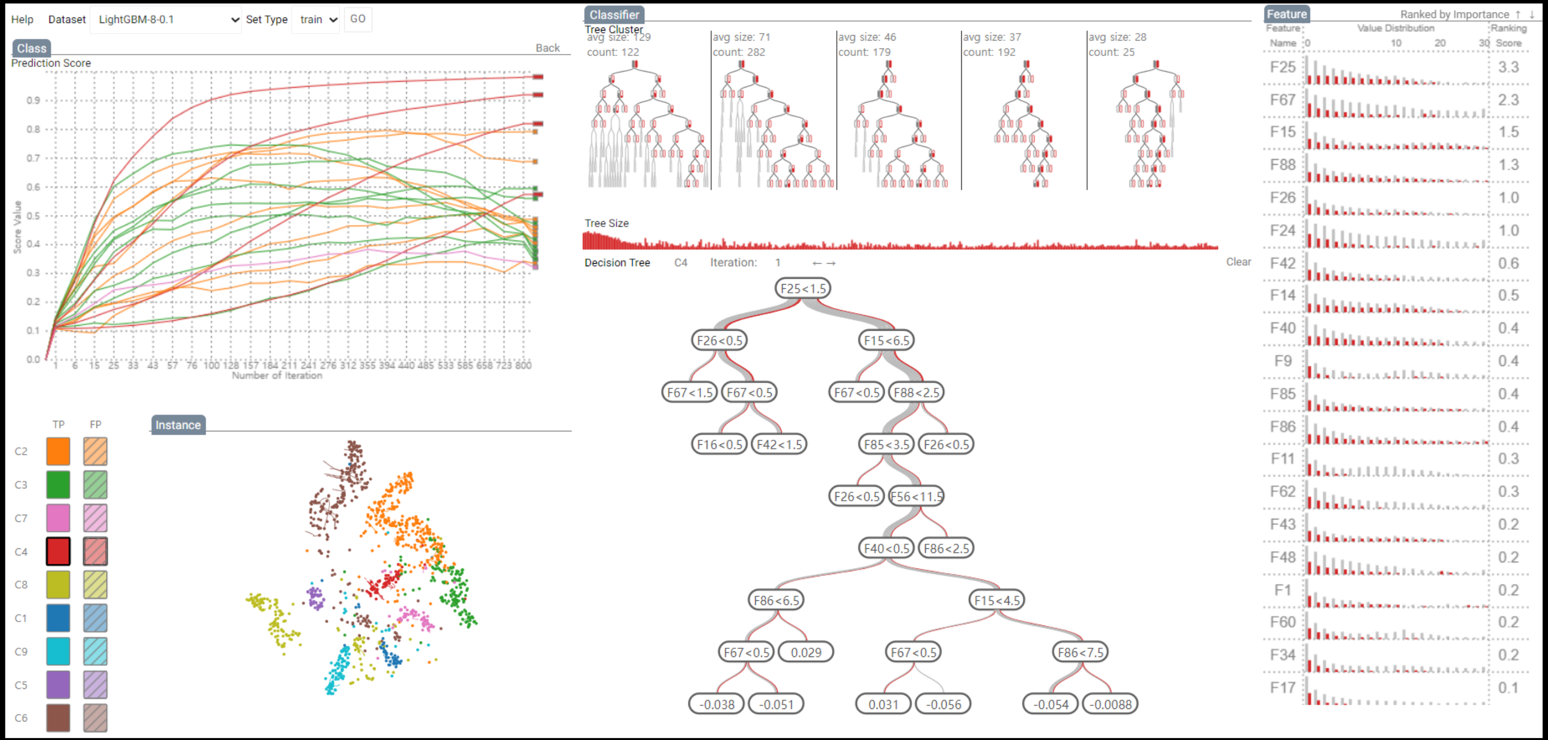
(1) LightGBM

```python
def train_lgb(obs_dict, energy_df):
    feature_importance = pd.DataFrame()
    test_pred_y = {}
    for key, obs_df in obs_dict.items():
        location = obs_df.loc_name[0]; code = obs_df.loc_num[0]
        if location not in energy_df.columns:
            continue

        train_x, val_x, train_y, val_y, test_x = preprocess_local(obs_df, energy_df, location)
        train_dataset = lgb.Dataset(train_x, train_y.values)
        val_dataset = lgb.Dataset(val_x, val_y.values)

        sentence = ' ' + f'{location.capitalize()} Training Start '
        print(f'{sentence:=^110}')
        params = {'learning_rate': 0.01, 'objective': 'regression',
                  'metric':'logloss', 'seed':42, 'force_col_wise': True}
        model = lgb.train(params, train_dataset, 10000, val_dataset, feval=nmae_10,
                          verbose_eval=500, early_stopping_rounds=200)

        sentence = ' ' + f'{location.capitalize()} Training Finish '
        print(f'{sentence:=^110}₩n₩n')
        test_y = model.predict(test_x)
        test_pred_y[location] = np.where(test_y<50, 0, test_y)

        joblib.dump(model, f'../model/{location}_model.pkl')

        feature_importance['feature'] = train_x.columns
        feature_importance[location] = model.feature_importance()

        plot_importance(model, figsize=(14, 8), title=f'{location}_feature_importances');

    return feature_importance, test_pred_y
```



Advantages of using LightGBM

- Faster training speed and higher efficiency
- Lower memory usage
- Better accuracy
- Support of parallel and GPU learning
- Capable of handling large-scale data



We have too much data, three years' worth of observation data, to run machine learning. So we have no choice but to choose LGBM Model which is faster and more accurate than other models.

4. MODELING

(2) Local Training

For more efficient model learning, custom mattresses have been created to prevent actual solar power generation from being reflected in learning when it is less than 10% of maximum power generation. This is because overfitting can occur if you include very little power generation in your learning.

To see how well the model trained during the training, we set the verbose-eval parameter to 500. The final loss value is distributed about 5-6 for each region, so we could see that the training was done well.

Gangneung Data (Sample)

Features (Local Observation Data)

Target (Local Generation Data)

	temp	precipitation	wind_speed	wind_direction	humidity	pressure_vapor	dew_point	pressure_local	pressure_sea	sunshine	radiation	snow	cloud	air_opacity	temp_surf	energy
0	21.203125	0.000000	1.000000	90.0	68.0	17.000000	15.000000	1010.0	1013.0	1.000000	1.240234	0.000000	0.000000	1447.0	22.796875	71.571831
1	18.593750	0.000000	5.699219	230.0	23.0	4.898438	-2.900391	1004.0	1007.0	0.099976	0.330078	0.000000	5.000000	721.0	16.093750	37.138028
2	13.898438	0.000000	0.500000	20.0	89.0	14.101562	12.101562	1024.0	1027.0	0.199951	0.500000	0.000000	9.000000	1097.0	16.593750	67.605634
3	0.899902	0.000000	4.601562	250.0	12.0	0.799805	-25.296875	1025.0	1028.0	1.000000	1.490234	0.000000	1.000000	5008.0	2.500000	532.912676
4	5.898438	0.000000	2.099609	230.0	44.0	4.101562	-5.398438	1020.5	1024.0	0.000000	0.000000	0.000000	0.000000	1628.0	2.000000	0.090141
5	4.500000	0.099976	2.199219	230.0	93.0	7.800781	3.400391	1021.0	1024.0	0.000000	0.140015	0.000000	9.000000	1031.0	5.000000	1.892958
6	21.406250	0.000000	1.500000	360.0	82.0	20.796875	18.093750	1002.0	1005.0	0.000000	0.459961	0.000000	5.000000	1653.0	24.406250	145.667606
7	6.601562	0.000000	5.500000	230.0	19.0	1.799805	-15.398438	1019.5	1022.5	1.000000	2.029297	0.000000	0.000000	5008.0	12.898438	440.247887
8	9.000000	0.700195	2.900391	320.0	91.0	10.398438	7.601562	1011.0	1014.5	0.000000	0.059998	0.000000	10.000000	1120.0	9.898438	0.360563
9	18.593750	0.000000	2.099609	90.0	69.0	14.703125	12.703125	1012.5	1015.5	0.000000	1.809570	0.000000	10.000000	764.0	31.593750	400.315493
10	24.906250	0.000000	1.000000	230.0	58.0	18.203125	16.000000	1004.0	1007.0	0.000000	0.059998	0.000000	6.000000	1418.0	22.593750	9.374648
11	5.500000	0.000000	2.400391	250.0	45.0	4.101562	-5.500000	1029.0	1033.0	1.000000	0.649902	0.000000	0.000000	2296.0	6.199219	131.245070
12	6.500000	0.000000	3.199219	230.0	28.0	2.699219	-10.703125	1019.0	1022.5	1.000000	1.000000	0.000000	2.000000	5000.0	4.101562	394.276056
13	10.500000	0.000000	1.299805	90.0	50.0	6.300781	0.399902	1023.0	1026.0	0.899902	0.830078	0.000000	7.000000	5008.0	9.203125	243.019718
14	6.000000	0.000000	1.599609	250.0	18.0	1.700195	-16.500000	1022.0	1025.0	1.000000	1.500000	0.000000	3.000000	5008.0	4.300781	371.380282
15	19.593750	0.000000	1.799805	250.0	73.0	16.593750	14.601562	1011.0	1014.0	1.000000	0.799805	0.000000	3.000000	5008.0	20.703125	73.194366
16	22.406250	0.000000	2.900391	90.0	66.0	17.796875	15.703125	1005.5	1008.5	1.000000	2.599609	0.000000	4.000000	1161.0	31.203125	647.211268
17	30.000000	0.000000	1.200195	70.0	74.0	31.296875	24.796875	1008.0	1010.5	1.000000	2.650391	0.000000	8.000000	5008.0	34.000000	537.600000
18	16.500000	0.000000	2.500000	250.0	80.0	15.000000	13.000000	1014.5	1017.5	0.000000	0.090027	0.000000	3.000000	4496.0	17.406250	8.292958

Local Training

Local Observation Data

+

Local Generation Data

→ Local Training → AI Model

4. MODELING

(2) Local Training

For more efficient model learning, we invented custom metric to prevent actual solar power generation from being reflected in learning when it is less than 10% of maximum power generation. This is because overfitting can occur if you include very little power generation in your learning.

To see how well the model trained during the training, we set the verbose-eval parameter to 500. The final loss value is distributed about 5-6 for each region, so we could see that the training was done well.

Custom Metric

```python
def nmae_10(y_pred, dataset):
    y_true = dataset.get_label()

    absolute_error = abs(y_true - y_pred)
    absolute_error /= 1000

    target_idx = np.where(y_true>=100)

    nmae = 100 * absolute_error[target_idx].mean()

    return 'score', nmae, False
```

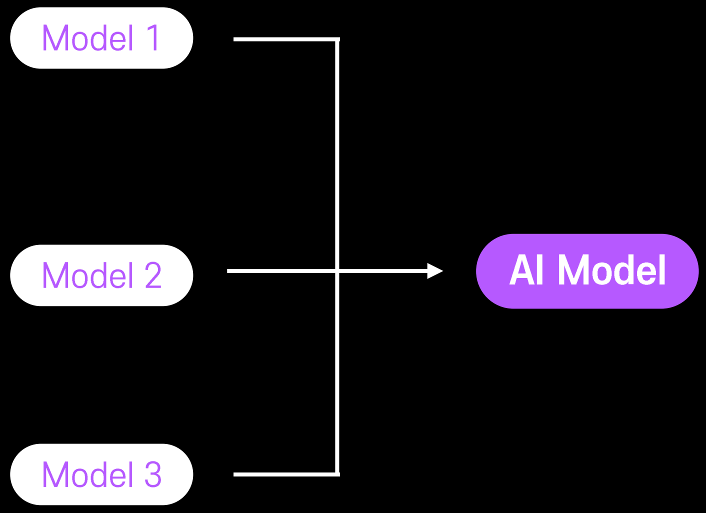
```python
def sola_nmae(answer, pred):
    absolute_error = np.abs(answer - pred)

    absolute_error /= 1000

    target_idx = np.where(answer>=100)

    nmae = 100 * absolute_error[target_idx].mean()

    return nmae
```

Schemetic Model

```
======================================== Dangjin Training Start ========================================
[LightGBM] [Info] Total Bins 2281
[LightGBM] [Info] Number of data points in the train set: 11621, number of used features: 18
[LightGBM] [Info] Start training from score 313.387058
Training until validation scores don't improve for 200 rounds
[500]	valid_0's score: 6.14567
[1000]	valid_0's score: 5.77038
[1500]	valid_0's score: 5.58683
[2000]	valid_0's score: 5.4754
[2500]	valid_0's score: 5.40042
[3000]	valid_0's score: 5.3485
[3500]	valid_0's score: 5.31184
[4000]	valid_0's score: 5.27776
[4500]	valid_0's score: 5.24603
[5000]	valid_0's score: 5.22724
[5500]	valid_0's score: 5.21488
[6000]	valid_0's score: 5.19617
Early stopping, best iteration is:
[6090]	valid_0's score: 5.19242
======================================== Dangjin Training Finish ========================================
```

```
======================================== Gangneung Training Start ========================================
[LightGBM] [Info] Total Bins 1990
[LightGBM] [Info] Number of data points in the train set: 6759, number of used features: 17
[LightGBM] [Info] Start training from score 299.112368
Training until validation scores don't improve for 200 rounds
[500]	valid_0's score: 6.19167
[1000]	valid_0's score: 6.04034
[1500]	valid_0's score: 5.99863
[2000]	valid_0's score: 5.95411
[2500]	valid_0's score: 5.92756
Early stopping, best iteration is:
[2559]	valid_0's score: 5.9218
======================================== Gangneung Training Finish ========================================
```

```
======================================== Jinju Training Start ========================================
[LightGBM] [Info] Total Bins 1983
[LightGBM] [Info] Number of data points in the train set: 6801, number of used features: 17
[LightGBM] [Info] Start training from score 233.619984
Training until validation scores don't improve for 200 rounds
[500]	valid_0's score: 5.87133
[1000]	valid_0's score: 5.46206
[1500]	valid_0's score: 5.25439
[2000]	valid_0's score: 5.09687
[2500]	valid_0's score: 5.04748
Early stopping, best iteration is:
[2362]	valid_0's score: 5.04584
======================================== Jinju Training Finish ========================================
```



4. MODELING

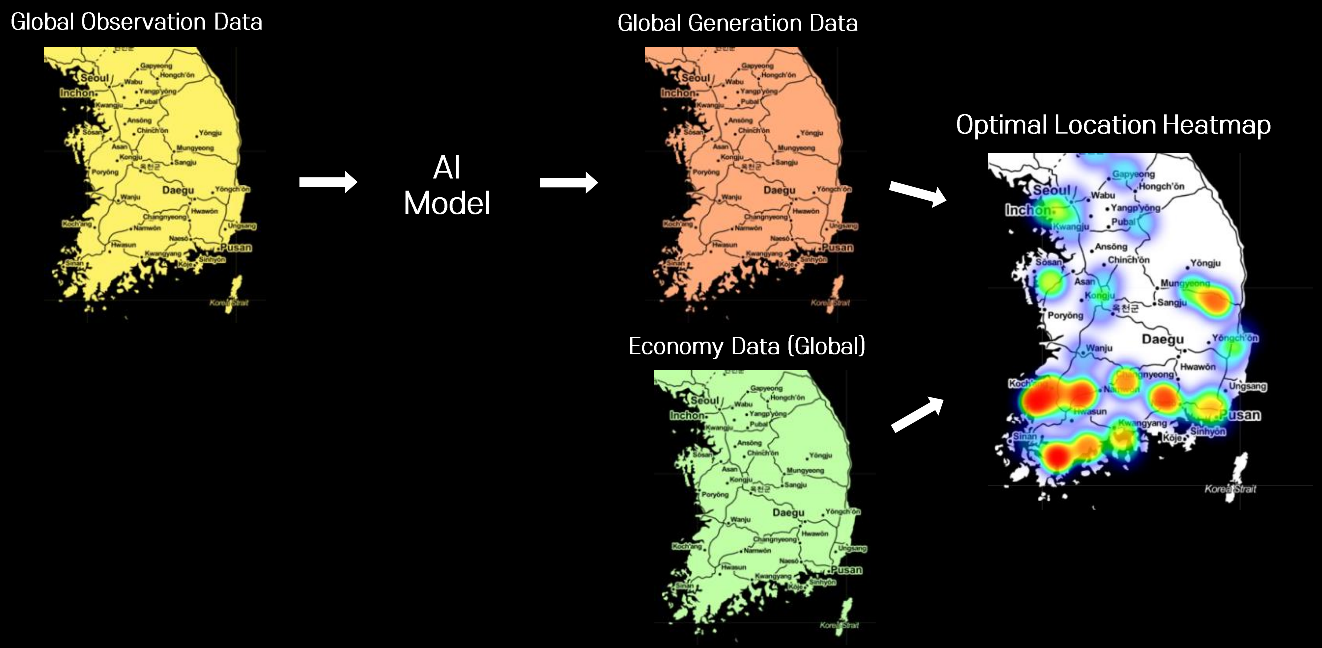
(3) Global Prediction

By using AI model that we just made in Local Training, we could predict solar power generation data. Let's define this as "Observation Score" because it is determined by unknown region's observation data.

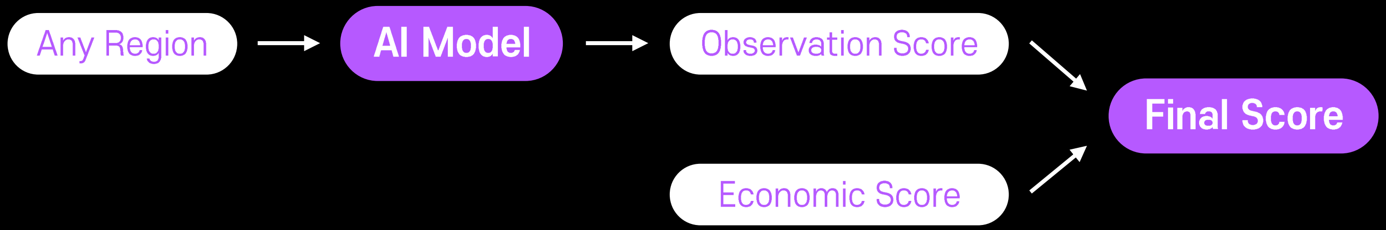
After collecting the data obtained by the Korea Meteorological Administration, around 40 regions have adequate observation data without missing values which are found to be suitable for global regions.

Once we get observation score, we can get final score by merging with economic score. Each region's score is equal to "Observation Score / $\sqrt[3]{\text{Economic Score}}$"

Global Area



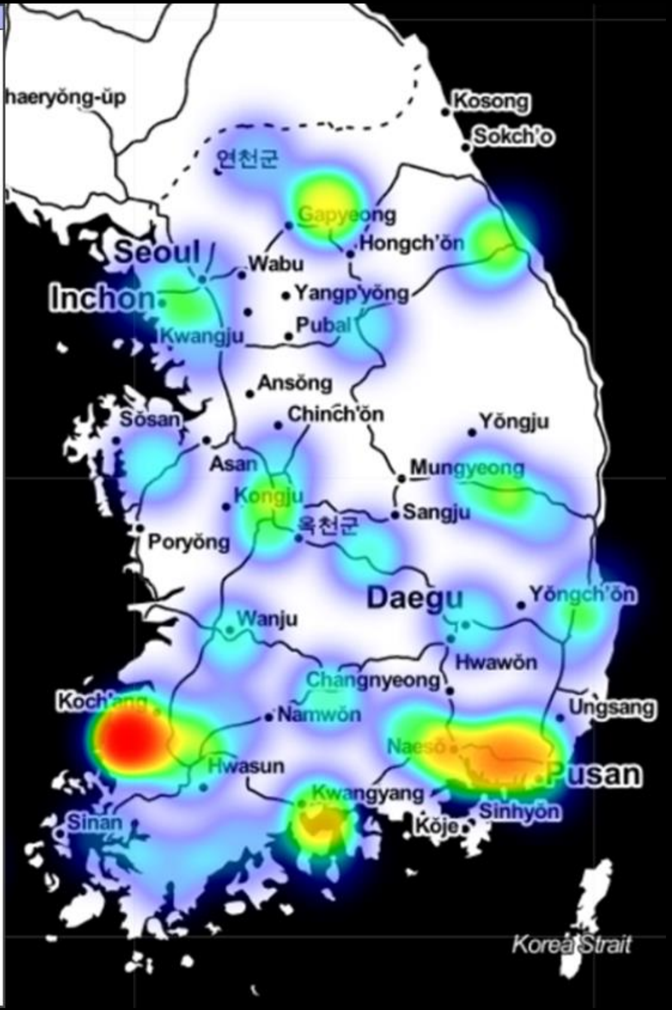
Schemetic Model



5. RESULTS (1) Visualization

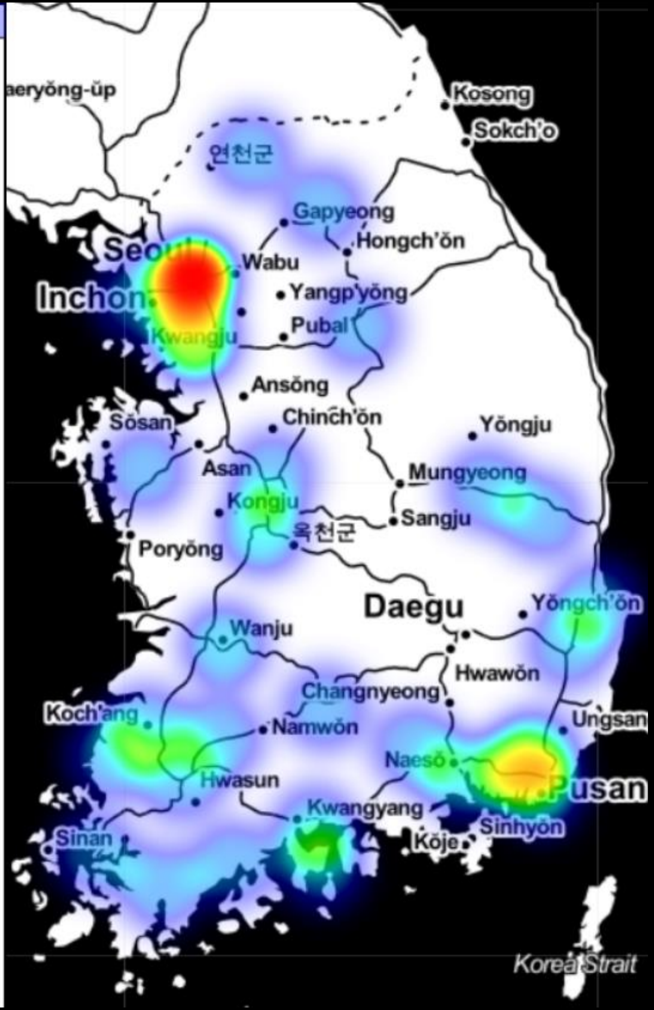
Observation Score

location name	latitude	longitude	score
Gwangyangsi	34.943	127.691	509
Yeonggwanggun	35.284	126.478	503
Uiryeonggun	35.323	128.288	503
Daejeon	36.372	127.372	500
Andong	36.573	128.707	499
Hamyanggun	35.511	127.745	496
Jeonju	35.822	127.155	493
Gwangju	35.173	126.892	493
Chupungnyeong	36.220	127.995	493
Chuncheon	37.903	127.736	491
Hongseong	36.658	126.688	491
Daegu	35.885	128.619	490
Changwon	35.170	128.573	490
Wonju	37.338	127.947	487
Cheongsonggun	36.435	129.040	484
Cheongju	36.639	127.441	483
Daegwallyeong	37.677	128.718	482
Busan	35.105	129.032	479
Mokpo	34.817	126.382	471
Bukchuncheon	37.948	127.755	468
Bukgangneung	37.805	128.855	468
Gangjingun	34.645	126.784	467
Suwon	37.272	126.985	466
Incheon	37.478	126.625	466
Yeosu	34.739	127.741	461
Seoul	37.571	126.966	461
Gimhaesi	35.230	128.891	457
Gyeongjusi	35.817	129.201	452
Sunchanggun	35.371	127.129	451
Pohang	36.032	129.380	449
Jeju	33.514	126.530	446
Heuksando	34.687	125.451	445
Gochanggun	35.427	126.697	441
Bukchangwon	35.227	128.673	441
Gochang	35.348	126.599	436
Gosan	33.294	126.163	433
Boseonggun	34.763	127.212	428
Yangsansi	35.307	129.020	411
Baengnyeongdo	37.974	124.712	372
Ulleungdo	37.481	130.899	364
Cheorwon	38.148	127.304	323



Economic Score

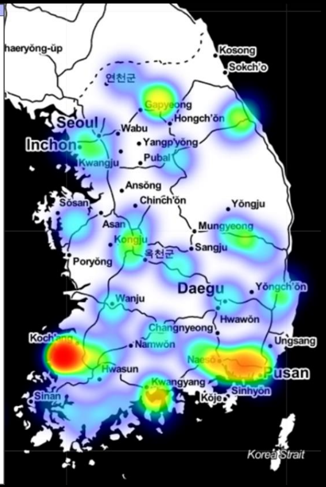
location name	latitude	longitude	price
Gwangyang	34.9434	127.6914	71.0
Yeonggwang	35.2837	126.4778	18.0
Uiryeong	35.3226	128.2881	22.0
Daejeon	36.372	127.3721	464.0
Andong	36.5729	128.7073	62.0
Hamyang	35.5114	127.7454	26.0
Jeonju	35.8215	127.155	368.0
Gwangju	35.1729	126.8916	215.0
Chuncheon	37.9026	127.7357	149.0
Hongseong	36.6576	126.6877	45.0
Changwon	35.1702	128.5728	320.0
Wonju	37.3375	127.9466	146.0
Cheongsong	36.4351	129.0401	21.0
Cheongju	36.6392	127.4407	163.0
Busan	35.1047	129.032	323.0
Mokpo	34.8173	126.3815	251.0
Gangjin	34.6446	126.7841	13.0
Suwon	37.2723	126.9853	1268.0
Incheon	37.4777	126.6249	58.0
Yeosu	34.7393	127.7406	116.0
Seoul	37.5714	126.9658	3934.0
Gimhae	35.2298	128.8908	239.0
Gyeongju	35.8174	129.2009	108.0
Sunchang	35.3713	127.1286	13.0
Pohang	36.032	129.38	168.0
Jeju	33.5141	126.5297	223.0
Gochang	35.4266	126.697	16.0
Boseong	34.7634	127.2123	14.0
Yangsan	35.3074	129.0201	245.0
Cheorwon	38.1479	127.3042	30.0



5. RESULTS (1) Visualization

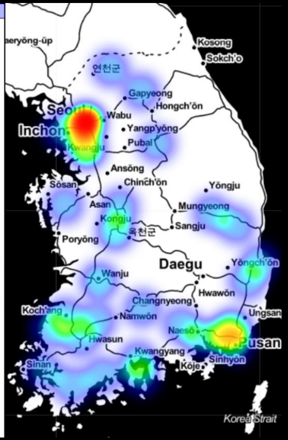
Observation Score

location name	latitude	longitude	score
Gwangyangsi	34.943	127.691	509
Yeonggwanggun	35.284	126.478	503
Uiryeonggun	35.323	128.288	503
Daejeon	36.372	127.372	500
Andong	36.573	128.707	499
Hamyanggun	35.511	127.745	496
Jeonju	35.822	127.155	493
Gwangju	35.173	126.892	493
Chupungnyeong	36.220	127.995	493
Chuncheon	37.903	127.736	491
Hongseong	36.658	126.688	491
Daegu	35.885	128.619	490
Changwon	35.170	128.573	490
Wonju	37.338	127.947	487
Cheongsonggun	36.435	129.040	484
Cheongju	36.639	127.441	483
Daegwallyeong	37.677	128.718	482
Busan	35.105	129.032	479
Mokpo	34.817	126.382	471
Bukchuncheon	37.948	127.755	468
Bukgangneung	37.805	128.855	468
Gangjingun	34.645	126.784	467
Suwon	37.272	126.985	466
Incheon	37.478	126.625	466
Yeosu	34.739	127.741	461
Seoul	37.571	126.966	461
Gimhaesi	35.230	128.891	457
Gyeongjusi	35.817	129.201	452
Sunchanggun	35.371	127.129	451
Pohang	36.032	129.380	449
Jeju	33.514	126.530	446
Heuksando	34.687	125.451	445
Gochanggun	35.427	126.697	441
Bukchangwon	35.227	128.673	441
Gochang	35.348	126.599	436
Gosan	33.294	126.163	433
Boseonggun	34.763	127.212	428
Yangsansi	35.307	129.020	411
Baengnyeongdo	37.974	124.712	372
Ulleungdo	37.481	130.899	364
Cheorwon	38.148	127.304	323



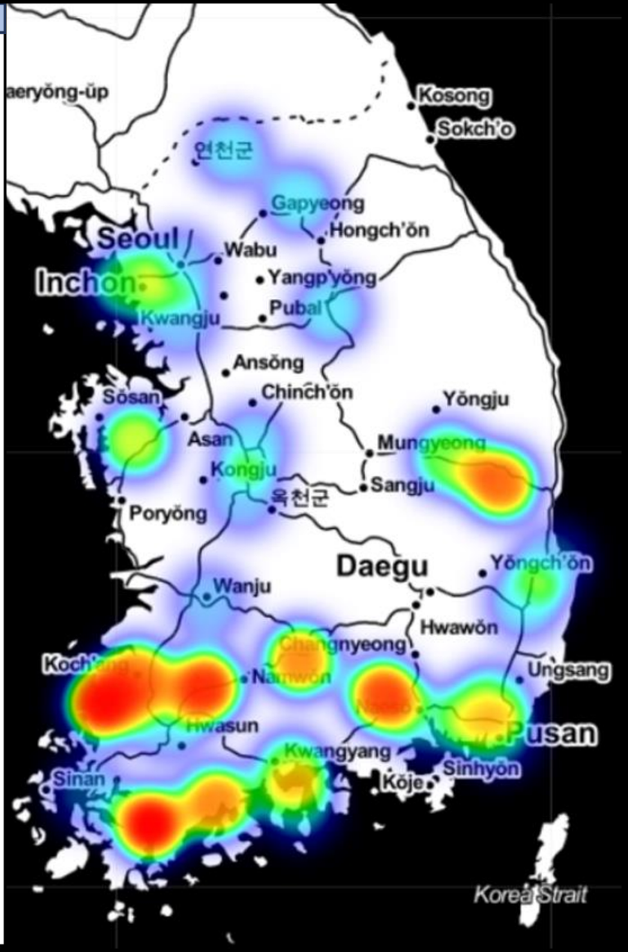
Economic Score

location name	latitude	longitude	price
Gwangyang	34.9434	127.6914	71.0
Yeonggwang	35.2837	126.4778	18.0
Uiryeong	35.3226	128.2881	22.0
Daejeon	36.372	127.3721	464.0
Andong	36.5729	128.7073	62.0
Hamyang	35.5114	127.7454	26.0
Jeonju	35.8215	127.155	368.0
Gwangju	35.1729	126.8916	215.0
Chuncheon	37.9026	127.7357	149.0
Hongseong	36.6576	126.6877	45.0
Changwon	35.1702	128.5728	320.0
Wonju	37.3375	127.9466	146.0
Cheongsong	36.4351	129.0401	21.0
Cheongju	36.6392	127.4407	163.0
Busan	35.1047	129.032	323.0
Mokpo	34.8173	126.3815	251.0
Gangjin	34.6446	126.7841	13.0
Suwon	37.2723	126.9853	1268.0
Incheon	37.4777	126.6249	58.0
Yeosu	34.7393	127.7406	116.0
Seoul	37.5714	126.9658	3934.0
Gimhae	35.2298	128.8908	239.0
Gyeongju	35.8174	129.2009	108.0
Sunchang	35.3713	127.1286	13.0
Pohang	36.032	129.38	168.0
Jeju	33.5141	126.5297	223.0
Gochang	35.4266	126.697	16.0
Boseong	34.7634	127.2123	14.0
Yangsan	35.3074	129.0201	245.0
Cheorwon	38.1479	127.3042	30.0



Final Score

location	latitude	longitude	score	price	final score
Gangjin	34.6446	126.7841	467.0	13.0	107.77
Sunchang	35.3713	127.1286	451.0	13.0	104.08
Boseong	34.7634	127.2123	428.0	14.0	91.71
Yeonggwang	35.2837	126.4778	503.0	18.0	83.83
Gochang	35.4266	126.697	441.0	16.0	82.69
Cheongsong	36.4351	129.0401	484.0	21.0	69.14
Uiryeong	35.3226	128.2881	503.0	22.0	68.59
Hamyang	35.5114	127.7454	496.0	26.0	57.23
Hongseong	36.6576	126.6877	491.0	45.0	32.73
Cheorwon	38.1479	127.3042	323.0	30.0	32.3
Andong	36.5729	128.7073	499.0	62.0	24.15
Incheon	37.4777	126.6249	466.0	58.0	24.1
Gwangyang	34.9434	127.6914	509.0	71.0	21.51
Gyeongju	35.8174	129.2009	452.0	108.0	12.56
Yeosu	34.7393	127.7406	461.0	116.0	11.92
Wonju	37.3375	127.9466	487.0	146.0	10.01
Chuncheon	37.9026	127.7357	491.0	149.0	9.89
Jeju	33.5141	126.5297	446.0	137.0	9.77
Cheongju	36.6392	127.4407	483.0	163.0	8.89
Pohang	36.032	129.38	449.0	168.0	8.02
Gwangju	35.1729	126.8916	493.0	215.0	6.88
Jeju	33.5141	126.5297	446.0	223.0	6.0
Gimhae	35.2298	128.8908	457.0	239.0	5.74
Mokpo	34.8173	126.3815	471.0	251.0	5.63
Yangsan	35.3074	129.0201	411.0	245.0	5.03
Changwon	35.1702	128.5728	490.0	320.0	4.59
Gwangju	35.1729	126.8916	493.0	326.0	4.54
Busan	35.1047	129.032	479.0	323.0	4.45
Jeonju	35.8215	127.155	493.0	368.0	4.02
Daejeon	36.372	127.3721	500.0	464.0	3.23
Suwon	37.2723	126.9853	466.0	1268.0	1.1
Seoul	37.5714	126.9658	461.0	3934.0	0.35



5. RESULTS

(2) Optimal Locations

강진군
GANGJIN CITY

순창군
SUNCHANG

보성군

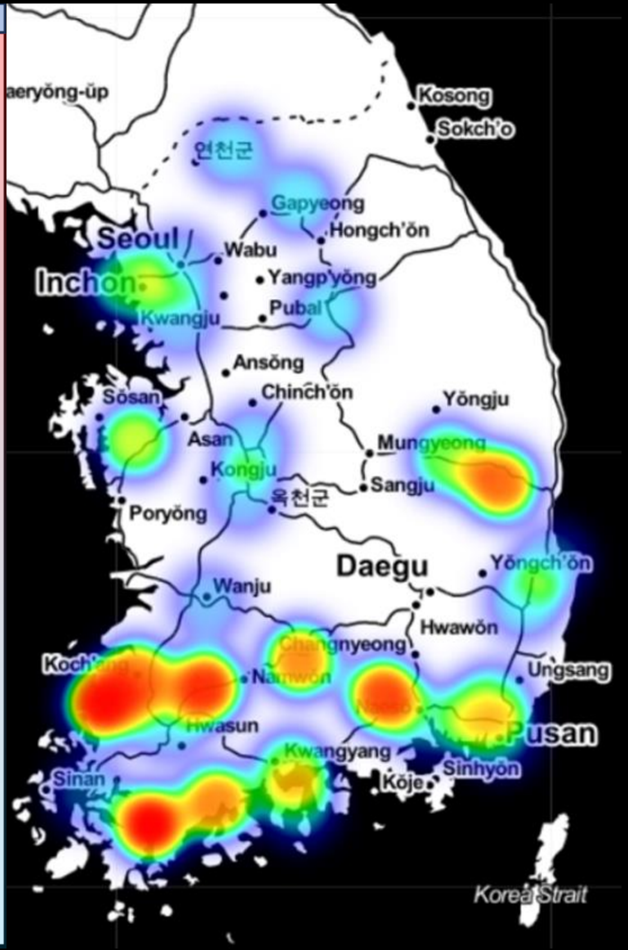
영광군
YEONGGWANG-GUN

Final optimal regions considering observation data and economic data were derived.

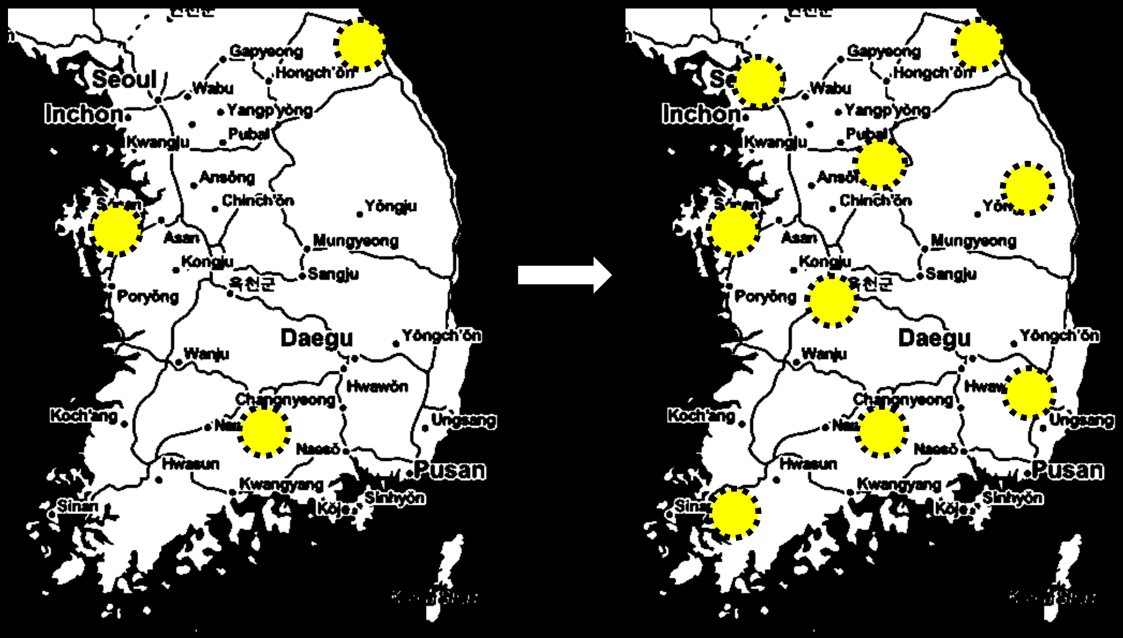
Gangjin, Sunchang, Boseong, Yeongguang is the optimal location for solar power plant in order. Overall, it is distributed at the bottom of the map.

Final Score

location	latitude	longitude	score	price	final score
Gangjin	34.6446	126.7841	467.0	13.0	107.77
Sunchang	35.3713	127.1286	451.0	13.0	104.08
Boseong	34.7634	127.2123	428.0	14.0	91.71
Yeonggwang	35.2837	126.4778	503.0	18.0	83.83
Gochang	35.4266	126.697	441.0	16.0	82.69
Cheongsong	36.4351	129.0401	484.0	21.0	69.14
Uiryeong	35.3226	128.2881	503.0	22.0	68.59
Hamyang	35.5114	127.7454	496.0	26.0	57.23
Hongseong	36.6576	126.6877	491.0	45.0	32.73
Cheorwon	38.1479	127.3042	323.0	30.0	32.3
Andong	36.5729	128.7073	499.0	62.0	24.15
Incheon	37.4777	126.6249	466.0	58.0	24.1
Gwangyang	34.9434	127.6914	509.0	71.0	21.51
Gyeongju	35.8174	129.2009	452.0	108.0	12.56
Yeosu	34.7393	127.7406	461.0	116.0	11.92
Wonju	37.3375	127.9466	487.0	146.0	10.01
Chuncheon	37.9026	127.7357	491.0	149.0	9.89
Jeju	33.5141	126.5297	446.0	137.0	9.77
Cheongju	36.6392	127.4407	483.0	163.0	8.89
Pohang	36.032	129.38	449.0	168.0	8.02
Gwangju	35.1729	126.8916	493.0	215.0	6.88
Jeju	33.5141	126.5297	446.0	223.0	6.0
Gimhae	35.2298	128.8908	457.0	239.0	5.74
Mokpo	34.8173	126.3815	471.0	251.0	5.63
Yangsan	35.3074	129.0201	411.0	245.0	5.03
Changwon	35.1702	128.5728	490.0	320.0	4.59
Gwangju	35.1729	126.8916	493.0	326.0	4.54
Busan	35.1047	129.032	479.0	323.0	4.45
Jeonju	35.8215	127.155	493.0	368.0	4.02
Daejeon	36.372	127.3721	500.0	464.0	3.23
Suwon	37.2723	126.9853	466.0	1268.0	1.1
Seoul	37.5714	126.9658	461.0	3934.0	0.35



6. FUTURE WORK



Increase the amount of Local Region

In this paper, we made AI model using just three local areas (gangneung, jinju, dangjin) because it's difficult to get solar power generation data (target data).

In the future, we'll increase the local area as much as possible. By doing that, we can get a more accurate AI model and a more accurate observation score.

Apply AI model to all over the world

The the way we get the global data in Korea, we could predict other countries like USA, England, China…

We could suggest the optimal location for solar power plants for them just by getting observation data of their regions.

Reference

[1] Lars Buitinck et al. “API design for machine learning software: experiences from the scikit-learn project”. In: arXiv preprint arXiv:1309.0238 (2013).

[2] Nicholas Cox. “PAIRPLOT: Stata module for plots of paired observations”. In: (2007).

[3] Md Ziaul Hassan et al. “Forecasting day-ahead solar radiation using machine learning approach”. In: 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE). IEEE. 2017, pp. 252–258.

[4] Ju-Hee Jang et al. “A preliminary research of the bifacial PV system under installation conditions”. In: Journal of the Korean Solar Energy Society 38.6 (2018), pp. 51–63.

[5] Guolin Ke et al. “Lightgbm: A highly efficient gradient boosting decision tree”. In: Advances in neural information processing systems 30 (2017), pp. 3146– 3154.

[6] Chang Ki Kim et al. “Derivation of typical meteorological year of Daejeon from satellite-based solar irradiance”. In: Journal of the Korean Solar Energy Society 38.6 (2018), pp. 27–36.

[7] YH Kwon, JY Kim, and MJ Lee. “Environmental considerations in the siting of solar and wind power plants”. In: Korea Environment Institute (2008).

[8] J Lee, HM Chung, and SS Lee. “Analysis on the location of the sunray energy power plants”. In: Korea Knowledge Information Technology Society 3.3 (2008), pp. 31–37.

[9] Kirim Lee and L. Hee. “Solar Power Plant Location Analysis Using GIS and Analytic Hierarchy Process”. In: 2015.

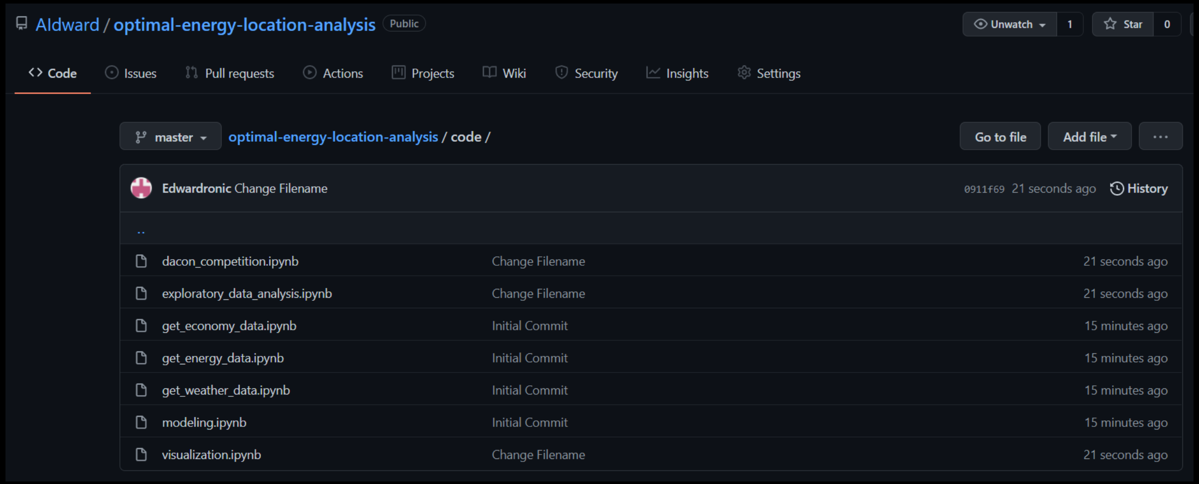
[10] Sung-Hun Lee et al. “Economic evaluation method for photovoltaic system development using insolation data analysis”. In: Journal of the Korean Institute of Illuminating and Electrical Installation Engineers 25.10 (2011), pp. 38–46.

[11] JI Park, MH Park, and SY Choi. “A study on GIS based suitability analysis of solar photovoltaic power generation using correlation analysis”. In: The Korean Society of Cadastre 28 (2012), pp. 91–107.

[12] Wim CM Van Beers and Jack PC Kleijnen. “Kriging interpolation in simulation: a survey”. In: Proceedings of the 2004 Winter Simulation Conference, 2004. Vol. 1. IEEE. 2004.

[13] Sung-Wook Yun et al. “A Study for Planning Optimal Location of Solar Photovoltaic Facilities using GIS”. In: (2019).

Code Link (Jupyter Notebook)



Get Weather Data Code

https://nbviewer.org/github/AIdward/optimal-energy-location-analysis/blob/master/code/get_weather_data.ipynb

Get Economy Code

https://nbviewer.org/github/AIdward/optimal-energy-location-analysis/blob/master/code/get_economy_data.ipynb

Get Solar Power Generation Code

https://nbviewer.org/github/AIdward/optimal-energy-location-analysis/blob/master/code/get_energy_data.ipynb

Visualization Code

https://nbviewer.org/github/AIdward/optimal-energy-location-analysis/blob/master/code/visualization.ipynb

Modeling Code

https://nbviewer.org/github/AIdward/optimal-energy-location-analysis/blob/master/code/modeling.ipynb