삼삼사자

데이콘 가스 수요량 예측

박승규, 이성준, 노현곤

목차

사전조사 및 데이터 수집

가정·상업용 도시가스의 경우는 소비의 많은 부분이 기온변화로 설명 될 수 있다. 수요가수가 포화시점에 근접하였으며 대부분 난방용으로 사용되는 가정·상업용 도시가스 소비 증감율은 난방도일1)의 증감

출처 : 산업용 도시가스 수요변화 요인분석(에너지경제연구원)

2018

행정구역별	보급률 (%)	공급권역내 총 가구수 (가구)	도시가스 수요가구수 (가구)
합계	85.0	21,674,404	18,429,378
경기도	88.0	5,306,214	4,669,015
서울특별시	98.2	4,263,868	4,186,336
부산광역시	92.3	1,480,468	1,367,105
인천광역시	92.9	1,115,997	1,201,455
경상남도	75.3	1,360,084	1,023,557
대구광역시	96.4	1,021,266	984,148
경상북도	65.9	1,179,225	776,786
충청남도	70.8	916,667	649,389
광주광역시	99.9	603,107	602,499
대전광역시	94.8	624,965	592,467
전라북도	71.6	806,235	576,948
충청북도	66.2	705,471	466,926
울산광역시	93.9	461,756	433,523

1. 서울특별시 시간별 기온데이터
2. 부산광역시 시간별 기온데이터
3. 인천광역시 시간별 기온데이터

도시가스 수요가구수(가구)가 가장 큰
위 3개의 데이터의 평균을 데이터셋으로 활용

데이터 전처리

데이터 확인 | **전처리**

날씨 데이터에서 **결측치** 다수 발견

Step. 1

결측치에서 가장 **가까운 값들의 평균**으로 처리

ex) 2015-01-01 00:00 (NaN)

(2014-12-31 23:00 기온 + 2015-01-01 01:00 기온) / 2

날씨 데이터와 공급량 데이터의 **시간 표기 차이** 발견

Step. 2

공급량 데이터 기준으로 변경

ex) 2014-01-01 0:00

→ 2013-12-31 | 24

공급량 데이터의 날짜를 year, month, day, weekday로 나눠줌

Step. 3

공급량 데이터의 날짜를

year, month, day, weekday로 나눠줌

ex) 연월일 2013-12-31

Year	month	day	weekday
2013	12	31	2

데이터 전처리

데이터 확인

전처리

구분값이 **문자열임**

Step. 1

구분값을 **숫자열**로 바꿔줌

ex) M → 2

	연월일	시간	일시	year	month	day	hour	weekday	구분	공급량	기온(°C)
122638	2014-12-31	23	2014-12-31 22:00:00	2014	12	31	22	2	6	576.316	-3.633333
122639	2014-12-31	24	2014-12-31 23:00:00	2014	12	31	23	2	6	542.836	-2.183330
122640	2015-01-01	1	2015-01-01 00:00:00	2015	1	1	0	3	0	2228.705	-4.800000
122641	2015-01-01	2	2015-01-01 01:00:00	2015	1	1	1	3	0	2098.593	-5.233333
122642	2015-01-01	3	2015-01-01 02:00:00	2015	1	1	2	3	0	1960.353	-5.666667

모델평가 및 비교

기온 측정모델

	LinearRegression	DecisionTreeRegressor	RandomForestRegressor	GradientBoostingRegressor	xgboost	lightgbm
test_size	9	1	2	6	9	0
train_score	4.093395	100	99.880564	91.492946	98.586207	97.351338
test_score	3.829319	98.704075	99.233752	91.507102	98.151016	97.219483
MAE	706.119927	47.882783	39.485004	178.220452	78.605846	95.461049
MSE	826229.833852	11266.179675	6601.293267	72694.287407	15885.149713	23789.734282
RMSE	908.971855	106.142261	81.248343	269.618782	126.036303	154.239211

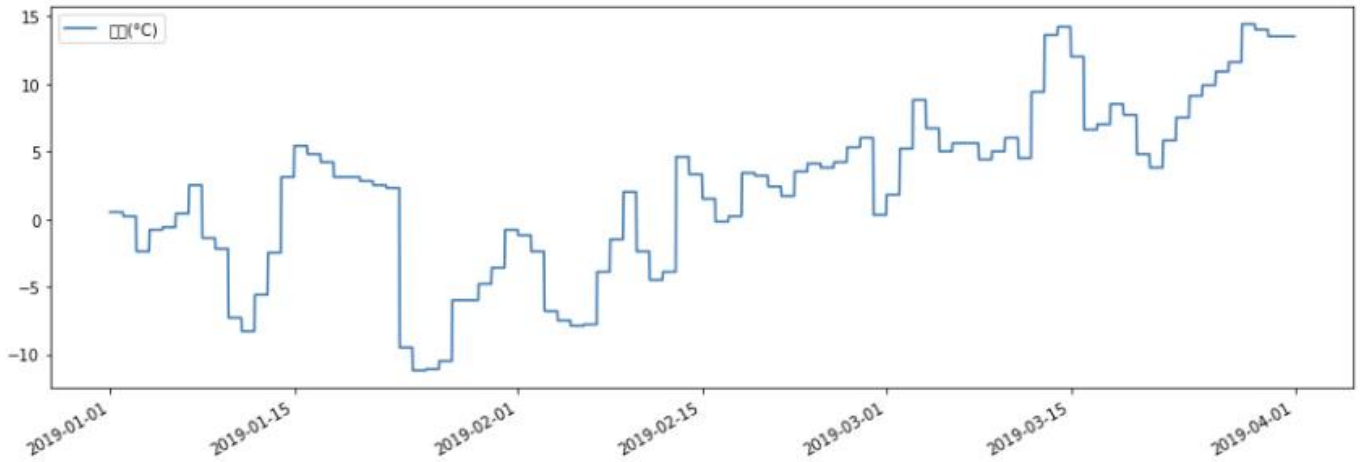
모델평가 및 비교

19년도 기온예측

	일자 시간 구분	일자	시간	구분	기온(° C)
0	2019-01-01 01 A	2019-01-01	1	A	0.5
1	2019-01-01 02 A	2019-01-01	2	A	0.5
2	2019-01-01 03 A	2019-01-01	3	A	0.5
3	2019-01-01 04 A	2019-01-01	4	A	0.5
4	2019-01-01 05 A	2019-01-01	5	A	0.5
5	2019-01-01 06 A	2019-01-01	6	A	0.5
6	2019-01-01 07 A	2019-01-01	7	A	0.5
7	2019-01-01 08 A	2019-01-01	8	A	0.5
8	2019-01-01 09 A	2019-01-01	9	A	0.5
9	2019-01-01 10 A	2019-01-01	10	A	0.5
10	2019-01-01 11 A	2019-01-01	11	A	0.5
11	2019-01-01 12 A	2019-01-01	12	A	0.5
12	2019-01-01 13 A	2019-01-01	13	A	0.5
13	2019-01-01 14 A	2019-01-01	14	A	0.5
14	2019-01-01 15 A	2019-01-01	15	A	0.5
15	2019-01-01 16 A	2019-01-01	16	A	0.5

19년도 기온 90일 예측

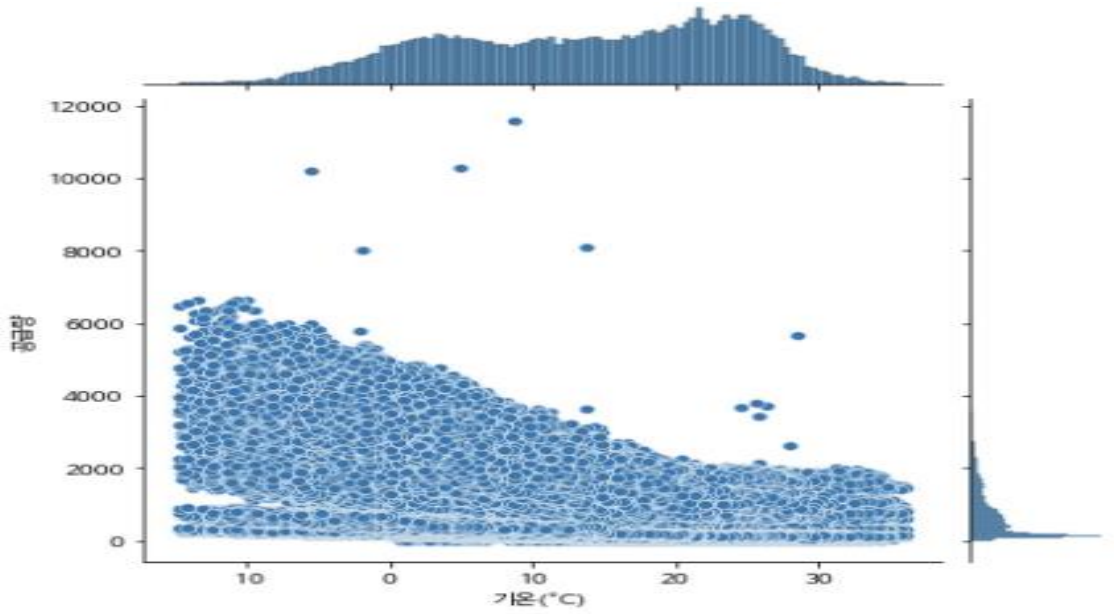
2019-01-01 01:00 ~ 2019-03-31 24:00



모델평가 및 비교

데이터 상관관계

	month	day	weekday	구분	공급량	기온(° C)
month	1.000000	0.011691	-0.000302	-0.112220	-0.335137	0.342848
day	0.011691	1.000000	0.004394	-0.002303	-0.019385	0.017556
weekday	-0.000302	0.004394	1.000000	-0.000443	-0.033323	-0.010292
구분	-0.112220	-0.002303	-0.000443	1.000000	0.237706	0.008419
공급량	-0.335137	-0.019385	-0.033323	0.237706	1.000000	-0.611679
기온(°C)	0.342848	0.017556	-0.010292	0.008419	-0.611679	1.000000



모델평가 및 비교

피쳐엔지니어링 전 모델 평가

	LinearRegression	DecisionTreeRegressor	RandomForestRegressor	GradientBoostingRegressor	xgboost	lightgbm	Catboost
test_size	6	1	1	1	9	2	2
train_score	31.49072	100	99.73431	95.55194	95.354505	98.67351	99.21437
test_score	31.42217	98.238919	98.77315	95.481119	95.305833	98.6827	99.21418
MAE	583.2795	61.08911	51.63444	137.87098	140.21451	267.10956	49.96599
MSE	586986.5	15310.03	10665.71	39285.090	40328.924	11348.63	6769.877
RMSE	766.1505	123.7337	103.2749	198.20466	200.82062	106.5299	82.27926
NMAE	4.262607	0.06954	0.070613	0.9256156	1.0036665	0.535277	0.4634

```
X = train[["year", "month", "day", "hour", "weekday", "기온(° C)", "구분"]]
y = train["공급량"]
```

```
print("정규화, 확장 전 데이터 셋 : ", X.shape, y.shape)
```

```
정규화, 확장 전 데이터 셋 :  (368088, 7) (368088,)
```

```
ex_X = PolynomialFeatures(degree=2, include_bias=False).fit_transform(nor_X)
print( ex_X.shape )
```

```
(368088, 35)
```

```
select = SelectPercentile(score_func=f_regression, percentile=50)
select.fit(X_train, y_train)
```

모델평가 및 비교

피쳐엔지니어링 후 모델 평가

	LinearRegression	DecisionTreeRegressor	RandomForestRegressor	GradientBoostingRegressor	xgboost	Lightgbm	Catboost
test_size	7	1	2	3	9	2	2
train_score	36.59598	100	99.70342	95.90891	95.90874	98.70071	99.3295
test_score	36.58995	98.17498	98.78098	95.82315	95.83771	98.70042	99.3096
MAE	549.3485	62.98522	52.98869	128.0361	132.7087	66.15616	46.94483
MSE	543689.7	15865.89	10501.96	35935.81	37620.58	11195.99	5947.822
RMSE	737.3531	125.9599	102.4791	189.5674	193.9603	105.8111	77.12212
NMAE	4.108543	0.078663	0.091127	0.78111	0.851044	0.542082	0.418261

모델평가 및 비교

최종 모델 선택

피쳐엔지니어링 전
catboost

학습용 : 8, 테스트용 : 2
학습용 데이터 결정계수: 0.992
테스트 데이터 결정계수: 0.992
MAE : 49.69598503900728
MSE : 6769.877180970 41
RMSE : 82.27926337158355
NMAE : 0.46339975093411834

피쳐엔지니어링 후
catboost

학습용 : 8, 테스트용 : 2
학습용 데이터 결정계수: 0.993
테스트 데이터 결정계수: 0.993
MAE : 46.944827180546 82
MSE : 5947.821538745806
RMSE : 77.12212094299407
NMAE : 0.41826113989121205

학습용 결정계수	테스트 결정계수	MAE	MSE	RMSE	NMAE
+0.001	+0.001	-2.751158	-6,692.75506	-5.157143	-0.04513857

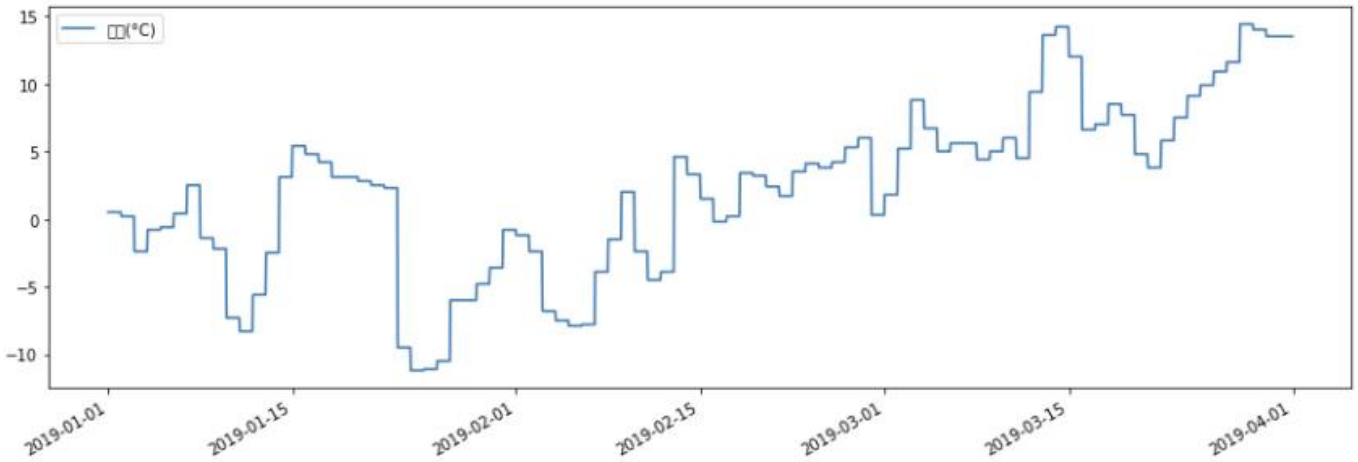
모델평가 및 비교

19년도 공급량측정

	일자|시간|구분	일자	시간	구분	구분_int	기온(° C)
0	2019-01-01 01 A	2019-01-01	1	A	0	0.5
1	2019-01-01 02 A	2019-01-01	2	A	0	0.5
2	2019-01-01 03 A	2019-01-01	3	A	0	0.5
3	2019-01-01 04 A	2019-01-01	4	A	0	0.5
4	2019-01-01 05 A	2019-01-01	5	A	0	0.5
5	2019-01-01 06 A	2019-01-01	6	A	0	0.5
6	2019-01-01 07 A	2019-01-01	7	A	0	0.5
7	2019-01-01 08 A	2019-01-01	8	A	0	0.5
8	2019-01-01 09 A	2019-01-01	9	A	0	0.5
9	2019-01-01 10 A	2019-01-01	10	A	0	0.5
10	2019-01-01 11 A	2019-01-01	11	A	0	0.5
11	2019-01-01 12 A	2019-01-01	12	A	0	0.5
12	2019-01-01 13 A	2019-01-01	13	A	0	0.5
13	2019-01-01 14 A	2019-01-01	14	A	0	0.5
14	2019-01-01 15 A	2019-01-01	15	A	0	0.5
15	2019-01-01 16 A	2019-01-01	16	A	0	0.5

19년도 공급량 90일 예측

2019-01-01 01:00 ~ 2019-03-31 24:00



최종결과 제출

162	벵튜2	벵튜	0.22919	2	17일 전
163	BluBerry		0.2466	5	1분 전
164	TonyStank		0.33589	4	12일 전