

제3회 한국가스공사 (KOGAS)

빅데이터·인공지능 스타트업 경진대회

가스공급량 수요예측 모델 개발

가스 공급량 예측을 위한 새로운 접근 방법

SSUDA

- 팀장 : 김성연
- 팀원 : 이동현, 조경덕, 손형락, 박우빈



한국가스공사



1. 배경 및 개요(천연가스란?)



LNG(Liquefied Natural Gas)



천연가스를 -162°C 로 냉각시켜 부피 1/600로
축소된 액화가스

PNG(Pipeline Natural Gas)



가스전에서 채취한 천연가스(NG)를 기체 상태로
소비지까지 배관을 통해 공급하는 가스

1. 배경 및 개요(LNG 현 상황)

제3회 한국가스공사 (KOGAS)

빅데이터·인공지능
스타트업 경진대회

가스공급량 수요예측 모델 개발



LNG '그린플레이션'...비수기인데 가격 44
0% 폭등

중앙일보 | 입력 2021.10.08 00:04

김남준 기자

예경연 "국제유가LNG가격, 내년 봄 이후 안정"



LNG가격 반년새 3배로 치솟았다

강경민 기자

입력 2021.09.09 17:49 | 수정 2021.09.10 00:08 | 지면 A1

겨울철 앞두고 재고 비축 수요 ↑
LPG 이어 가스요금도 인상 압박

세계 LNG 확보 대란, 한국 뒤흔친다

유럽·중국이 물량 쏠려가자 국제 투기수요까지 가세
한국 구매가격 하루 40% 급등...중소 비용부담 우려

이윤재, 이재철, 김덕식, 김재관, 박상경 기자 | 입력 : 2021.10.07 17:56:00 수정 : 2021.10.07 21:18:17

3

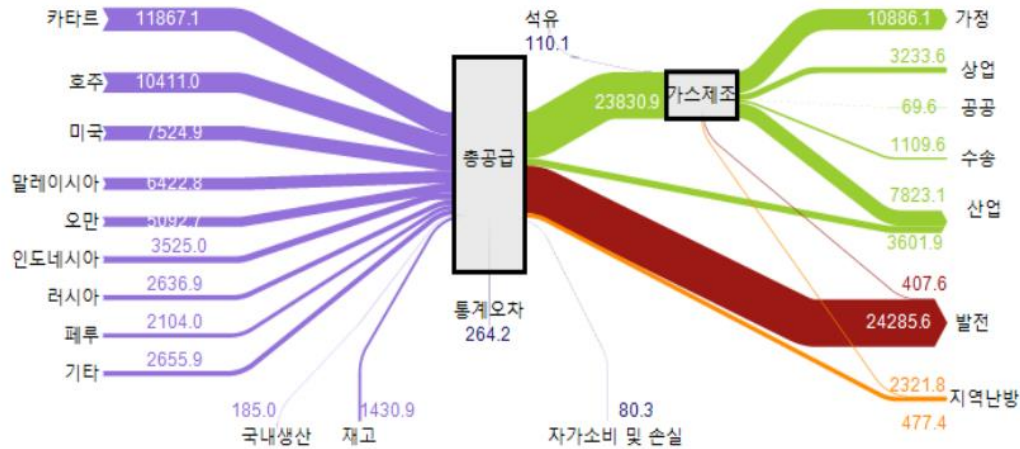
◆ 도미노 에너지 대란 ◆

- 국내 가스 수요량 증가
- 천연가스 가격 증가

1. 배경 및 개요(수입)



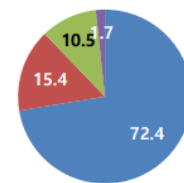
2020년 천연가스 수입 및 공급 흐름 (천 TOE)



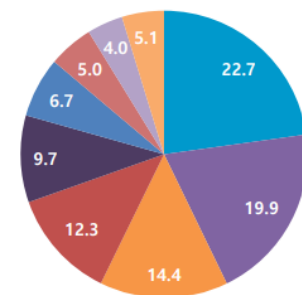
주: 기타 국가는 앙골라, 나이지리아, 브루나이, 파푸아 뉴기니, 아랍에미리트 연합 등
자료: 에너지경제연구원

- 주로 호주, 카타르에서 가스를 수입하여 사용.
- 수입하는 가스는 여러 부분에서 사용되고 있음.

1995년과 2020년 우리나라의 천연가스 수입선별 점유율 (%)



1995



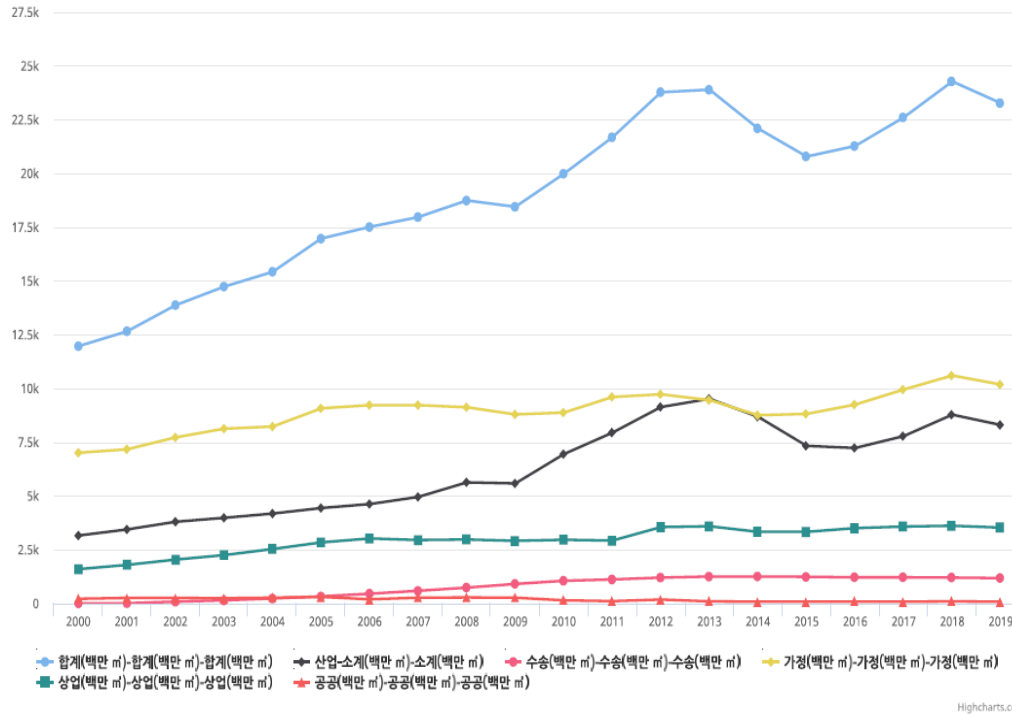
2020p

자료: 무역협회

1. 배경 및 개요(소비량 및 배관망)



<부문별 연간 도시 가스 소비량(2000~2019)>



- 전반적인 분야에서 가스 소비량의 점진적 증가.
- 전국적으로 가스 배관망이 연결되어 있음.

<전국 가스 배관망>



2. 사용 데이터

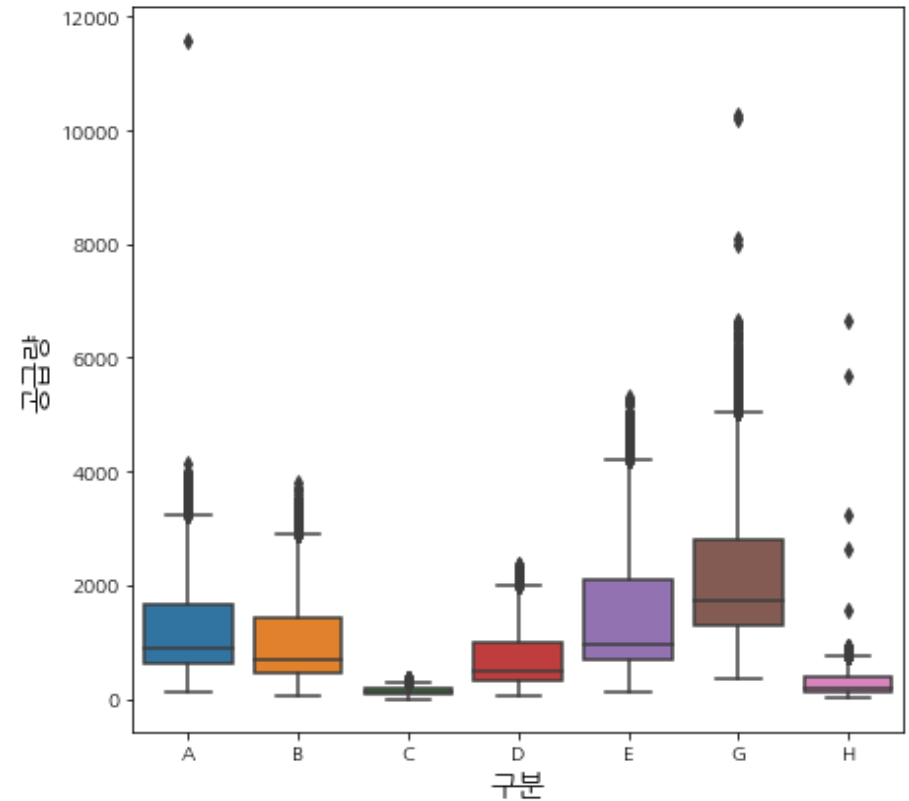


<대회 제공 데이터 셋>

	연월일	시간	구분	공급량
0	2013-01-01	1	A	2497.129
1	2013-01-01	2	A	2363.265
2	2013-01-01	3	A	2258.505
3	2013-01-01	4	A	2243.969
4	2013-01-01	5	A	2344.105
...
368083	2018-12-31	20	H	681.033
368084	2018-12-31	21	H	669.961
368085	2018-12-31	22	H	657.941
368086	2018-12-31	23	H	610.953
368087	2018-12-31	24	H	560.896

- 2013-01-01~2018-12-31 연/월/일/시간 단위로 구분 된 데이터
- 구분은 비식별화 처리된 데이터
- 공급량은 원데이터와 유사한 패턴을 갖도록 처리됨.(가상데이터)

<구분 별 공급량 Boxplot>

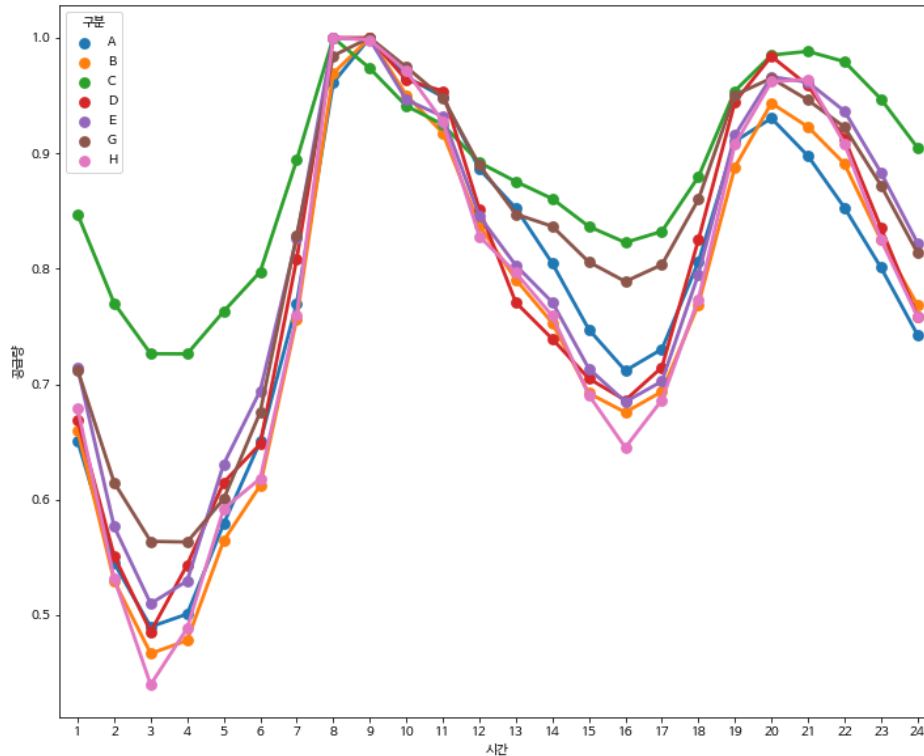


- C와 H에서 전체적인 공급량이 높지 않다.
- G는 공급량이 상대적으로 크기 때문에 수도권지역의 공급량으로 생각해 볼 수 있다.

2. 사용 데이터(시간대별 추세)



<시간에 따른 공급량 차이>

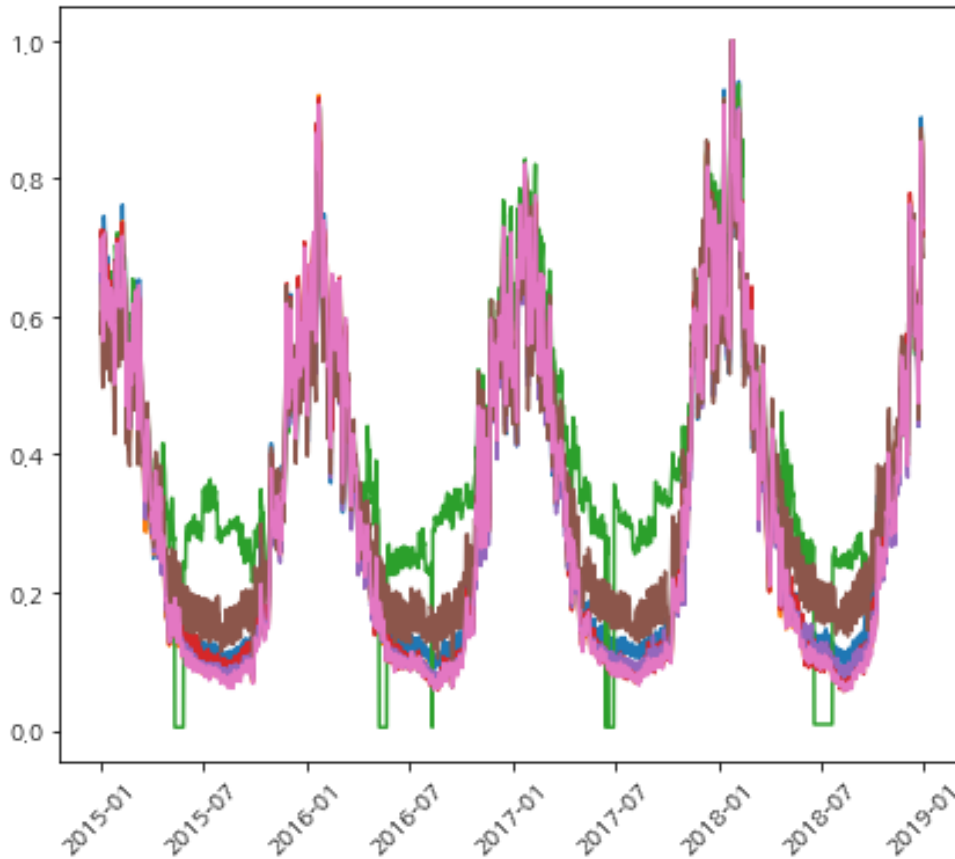


- 6년치의 시간별 공급량을 평균 내어 나타냄. 동일한 축상에서 보기 위해 최댓값으로 나눠 줌
- 시간에 따른 공급량의 차이를 보이며 특히 아침과 저녁 시간에 구분 별로 공급량의 사용량이 많음.

2. 사용 데이터(연간 추세)



<4년간 일별 공급량 추이>

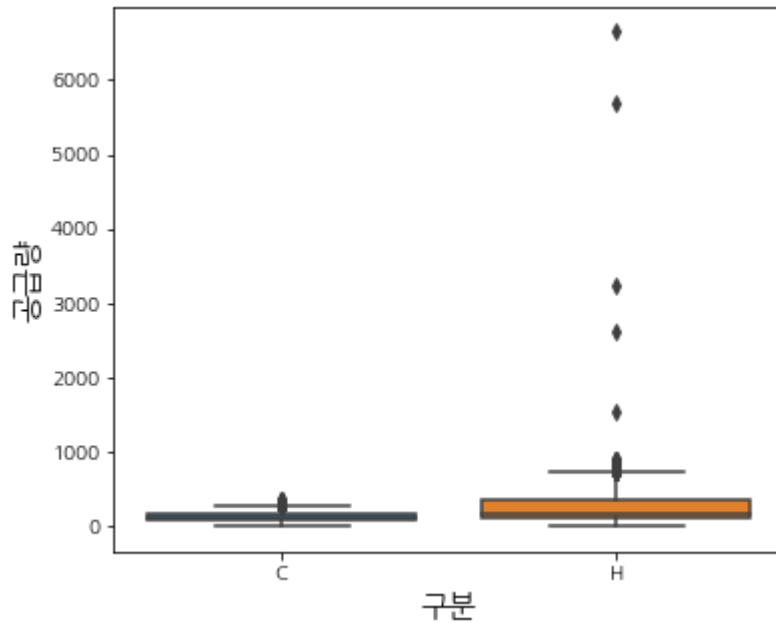


- 공급량이 계절성을 갖는 것으로 보임.
- 겨울철에 공급량이 많은 것으로 보임.
- 구분별로 모두 같은 계절성을 갖는 것으로 보임.

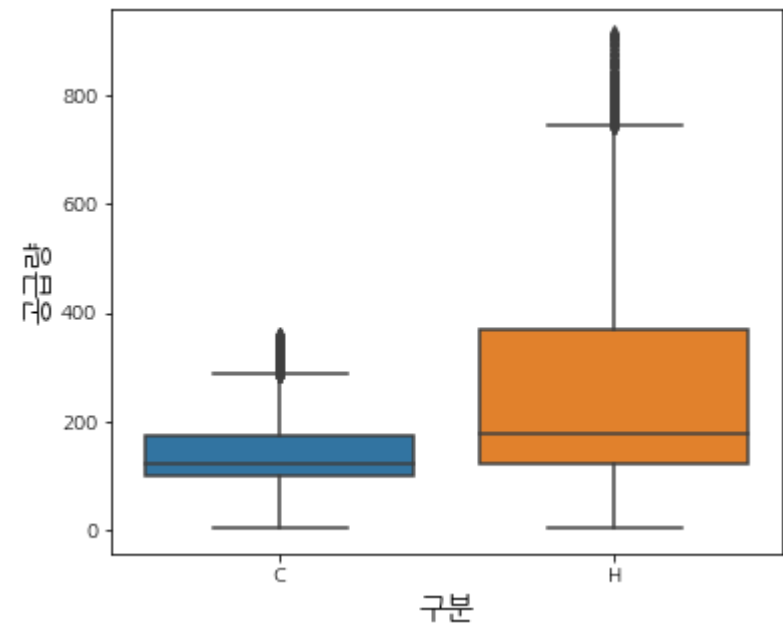
2. 사용 데이터(이상치 제거)



<범위를 크게 벗어난 이상치를 제거하기 전 Boxplot>



<범위를 크게 벗어난 이상치를 제거한 Boxplot>



```
1 train = train[(train['구분'] != 'H') | (train['공급량'] < 1500)]  
2 train = train[(train['구분'] != 'C') | (train['공급량'] > 3)]
```

2. 사용 데이터(변수 추가)



<파생변수 생성 후 Data Frame>

	연월일	시간	구분	공급량	Month	Year	Weekday	하루 가스공급량	시간당 비율	group_day	holiday
0	2013-01-01	1	A	2497.129	1	2013	1		0.041532	60125.336	True
1	2013-01-01	2	A	2363.265	1	2013	1		0.039306	60125.336	True
2	2013-01-01	3	A	2258.505	1	2013	1		0.037563	60125.336	True
3	2013-01-01	4	A	2243.969	1	2013	1		0.037322	60125.336	True
4	2013-01-01	5	A	2344.105	1	2013	1		0.038987	60125.336	True
...
365323	2018-12-31	20	H	681.033	12	2018	0		0.047997	14189.118	False
365324	2018-12-31	21	H	669.961	12	2018	0		0.047217	14189.118	False
365325	2018-12-31	22	H	657.941	12	2018	0		0.046369	14189.118	False
365326	2018-12-31	23	H	610.953	12	2018	0		0.043058	14189.118	False
365327	2018-12-31	24	H	560.896	12	2018	0		0.039530	14189.118	False

- 파생변수 Month 생성
- 파생변수 Year 생성
- 파생변수 Weekday 생성 (주중 여부)
- 파생변수 holiday 생성 (공휴일 여부)
- 파생변수 일별 시간당 가스 공급량 비율 생성

3. 사용 모델기술



가스 공급량 예측을 위한 시도해본 모델

< ARIMA, SARIMA >

- 정상성을 확인하고 차분을 통한 시계열 방법을 접근.
- 시간 별 예측에 있어 모델을 적합하는데 시간이 오래 걸리는 단점.
- 공급량의 추세를 잘 따라가는 경향을 보이지 못함.

<LGBM, XGB>

- 머신러닝 방법론 접근.
- 과적합되는 경향을 보임.

< LSTM, RNN, Prophet >

- 전통적 시계열 접근 방법에서 딥러닝 기반 시계열 접근법 사용.
- 지난 며칠간의 공급량을 참고하여 2019년 공급량 예측.
- 추세는 따라가나 지난 일을 참고하기 때문에 급격한 변동에는 유연하게 예측하지 못하는 경향이 있음.

<TABNET>

- 정형데이터를 위한 딥러닝 모델로, 머신러닝 접근 모델들을 앙상블 모델로 적합시키기 위해 메타 모델로 사용.

3. 사용 모델기술



< Random Forest – 일별 공급량 >

- 일별 공급량과 시간대 별 비율을 예측하기 위해서 데이터 셋을 다르게 설정.
- 공급량과 일정 상관관계를 갖을 것으로 보이는 Weekday, Month, Holiday를 포함하고, 구분 별로 값에 유의한 차이가 있다고 판단하여 구분 변수도 이용.

<일별 공급량 예측 부분>

	Month	Weekday	holiday	구분_A	구분_B	구분_C	구분_D	구분_E	구분_G	구분_H
0	1	1	True	1	0	0	0	0	0	0
1	1	2	False	1	0	0	0	0	0	0
2	1	3	False	1	0	0	0	0	0	0
3	1	4	False	1	0	0	0	0	0	0
4	1	5	False	1	0	0	0	0	0	0
...
15217	12	3	False	0	0	0	0	0	0	1
15218	12	4	False	0	0	0	0	0	0	1
15219	12	5	False	0	0	0	0	0	0	1
15220	12	6	False	0	0	0	0	0	0	1
15221	12	0	False	0	0	0	0	0	0	1

- group_day는 당일 총 공급량을 의미함.
- 일별 총 공급량은 group_day 변수에 반복적으로 포함 되어 있으므로 하나씩만 추출하여 사용.
- 구분 변수는 One-Hot Encoding을 이용.
- 예측할 때 공급량 로그 변환.

3. 사용 모델기술



< Random Forest – 시간별 공급량 비율 >

- 공급량과 일정 상관관계를 갖을 것으로 보이는 시간, Weekday, Month, Holiday를 포함하고, 구분 별로 값에 유의한 차이가 있다고 판단하여 구분 변수도 이용.

<시간별 공급량 비율 예측 부분>

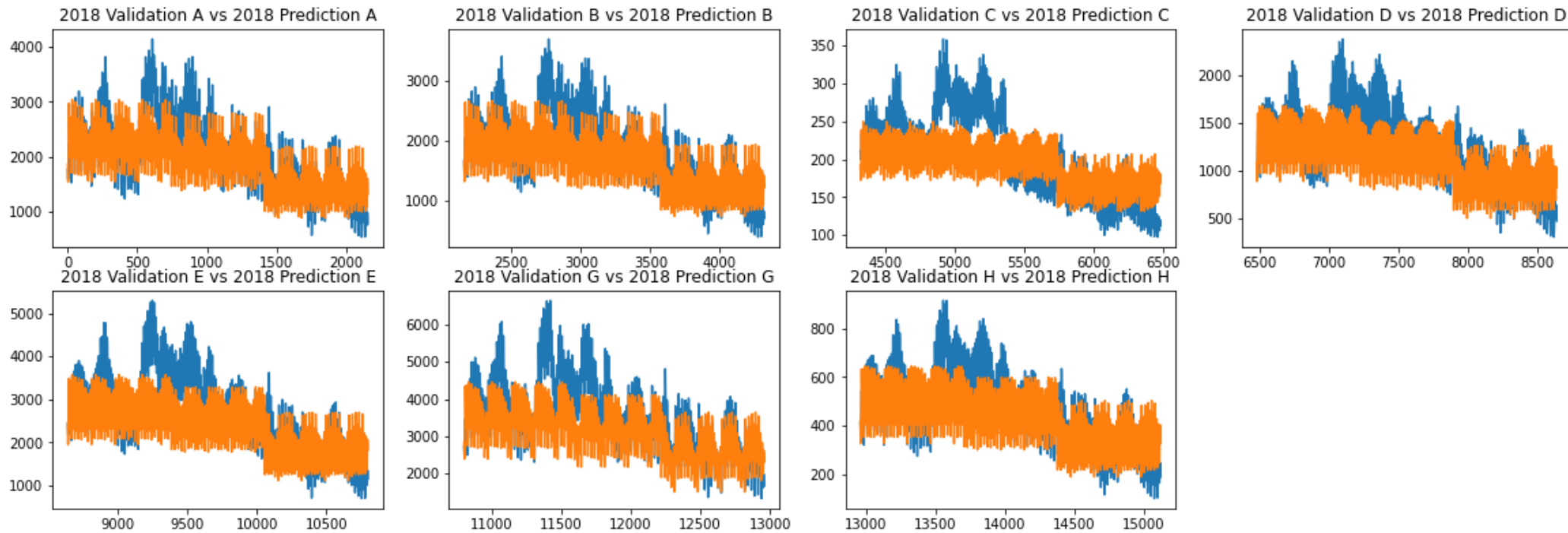
	시간	Month	Weekday	holiday	구분_A	구분_B	구분_C	구분_D	구분_E	구분_G	구분_H
0	1	1	1	True	1	0	0	0	0	0	0
1	2	1	1	True	1	0	0	0	0	0	0
2	3	1	1	True	1	0	0	0	0	0	0
3	4	1	1	True	1	0	0	0	0	0	0
4	5	1	1	True	1	0	0	0	0	0	0
...
365323	20	12	0	False	0	0	0	0	0	0	1
365324	21	12	0	False	0	0	0	0	0	0	1
365325	22	12	0	False	0	0	0	0	0	0	1
365326	23	12	0	False	0	0	0	0	0	0	1
365327	24	12	0	False	0	0	0	0	0	0	1

- 하루 가스 공급량 시간당 비율은 당일 공급량을 총 공급량으로 나눈 값임.
- 구분 변수는 One-Hot Encoding을 이용.
- 예측할 때 공급량 로그 변환.

4. 모델링 결과



<구분별 2018년 실제 값과 예측 값 비교>

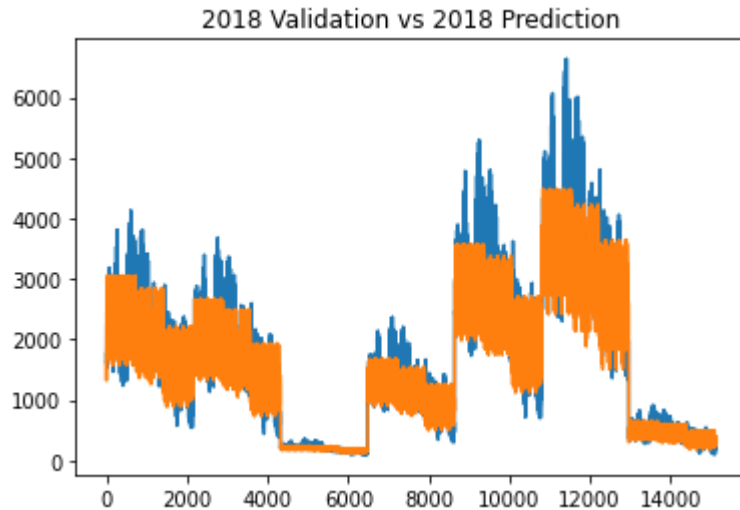


- 구분 별로 실제 값의 추세를 따라가는 것으로 보임.

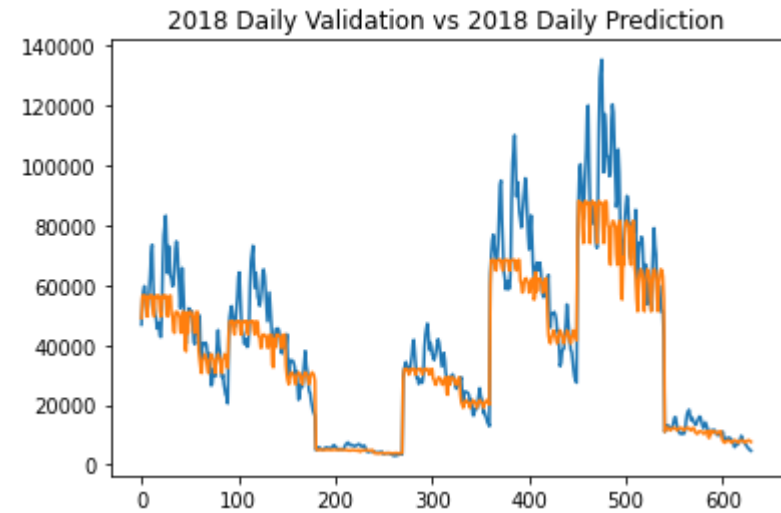
4. 모델링 결과



<2018년의 공급량 예측 후 실제 값과 비교>



<2018년 일별 예측 후 실제 값과 비교>



- 2018년을 검증 데이터로 일별 및 전체 공급량 예측을 진행 함.
- 일별, 전체 공급량 모두 추세를 따라가는 것을 보임.
- 일별 예측 검증 NMAE 점수 : 0.1552
- 시간별 예측 검증 NMAE 점수 : 0.1643
- 데이콘 제출 시 NMAE 점수 : 0.09956

5. 특징점(차별성/우수성)



	연월일	하루	가스공급량	시간당	비율	group_day
0	2013-01-01				0.041532	60125.336
1	2013-01-01				0.039306	60125.336
2	2013-01-01				0.037563	60125.336
3	2013-01-01				0.037322	60125.336
4	2013-01-01				0.038987	60125.336
...
365323	2018-12-31				0.047997	14189.118
365324	2018-12-31				0.047217	14189.118
365325	2018-12-31				0.046369	14189.118
365326	2018-12-31				0.043058	14189.118
365327	2018-12-31				0.039530	14189.118

- 시간별, 일별에 따른 공급량의 차이가 있다고 생각하여 시계열 모델 접근.
- LSTM , tabnet, RNN 등을 사용하여 예측
- 시계열 접근에서 예측 값의 추세는 따르지만 좋은 성능을 갖지 못함.
- 특히 시계열 모델의 사용에서 예측되는 기간이 길어질수록 후반부에서는 제대로 된 예측을 하지 못하는 경향이 있었음.
- 이를 해결하기 위해 시계열 접근을 최소화 하려는 기법을 고민함.



- 특정 시간과 특정 주기에 따라 공급량의 차이가 있을 것이라고 생각.
- 각 1~24시에 해당하는 공급량의 비율과 계절성을 반영한 일자 별 공급량의 합의 값을 곱하여 계절성을 함께 고려한 공급량을 예측.

6. 발전(사업화)계획



- 가스 공급량은 기온에 대하여 민감하게 반응하고 가스를 공급하는 배관 역시 기온에 직접적으로 영향을 받음.
- 따라서 이를 잘 대처하여 안정적인 공급을 할 수 있도록 보장해야 함.
- 따라서 계절성의 의미를 가지고 있는 시간당 공급량의 비율, 일별 공급량의 비율 등을 함께 고려한 변수를 만들어 모델에 반영함.
- 또한 기온 외부 데이터를 사용하여 하루 당 기온이 어느정도 비율로 상승하거나 하락할 때 공급량의 증감률을 고려하여 기온의 급격한 변화에도 공급량을 잘 예측할 수 있도록 발전시킬 수 있음.

-

7. 결론(정리)



- 국내에서 가스를 사용량의 증가와 더불어서 국제 LNG 가격도 상승하고 있음.
- 따라서 가스를 안정적으로 공급하는 것이 더욱 중요해짐.
- 특히 겨울철에는 기온의 급격한 변화로 가스 공급을 할 때 불안정성이 있을 수 있음.
- 이러한 가스 공급에 있어서 안정성을 확보하기 위해 일별, 시간별 변수를 사용하여 기온에 잘 대처할 수 있는 모델을 제안함.
- 실제 가스 공급량의 추세를 어느정도 따라감.
- 이를 통해 급격한 기온 변화에도 안정성 있는 가스 공급을 제공할 수 있는 방법을 마련할 수 있음

8. 팀원 역할 및 참여도



이름	역할
김성연	팀원 일정 조율, 팀 베이스 라인 만들기, 랜덤 포레스트 모델 제안
박우빈	시계열 모델 시도, ppt/포스터 및 발표 준비
손형락	시계열 모델 시도, ppt/포스터 및 발표 준비
이동현	양상블 모델 시도, 시계열 모델 시도, 1차 코드 제출
조경덕	시간별/일별 분할 모델 제안, 최종 모델 구축



제3회 한국가스공사 (KOGAS)

빅데이터·인공지능 스타트업 경진대회

가스공급량 수요예측 모델 개발

Q&A

제3회 한국가스공사 (KOGAS)

**빅데이터·인공지능
스타트업 경진대회**

감사합니다