

[따릉이 대여량 예측 AI 경진대회]

주최: 단국대학교, 용인시
주관: 데이콘
심승현, 박선희

목 록

1. 팀원 및 대회 소개
2. 데이터 확인
3. 모델링
4. 결과 및 아쉬운 점

팀원 소개



심승현

Main: 데이터 분석/ 파생변수 생성
Sub: 모델링/ 테스트



박선희

Main: 모델링/테스트
Sub: 데이터 분석/ 파생변수 생성

대회 소개

목적

서울시에서 운영중인 무인 공공 자전거
대여 서비스 '따릉이' 수요 예측

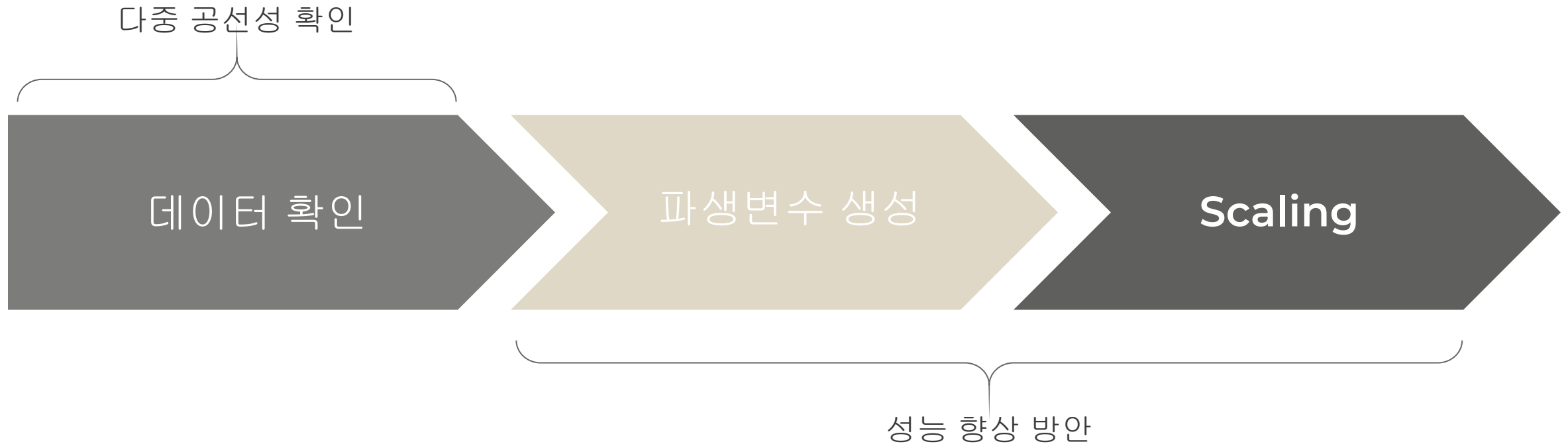
평가
지표

NMAE

$$NMAE = \frac{1}{n} \sum_i^m \frac{|true_i - predict_i|}{true_i}$$

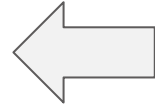
기간

2022.06.13 ~ 2022.07.01

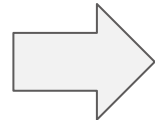


데이터 확인

date	0
precipitation	678
temp_mean	0
temp_highest	0
temp_lowest	0
PM10	67
PM2.5	68
humidity	0
sunshine_sum	5
sunshine_rate	0
wind_mean	0
wind_max	0
rental	0

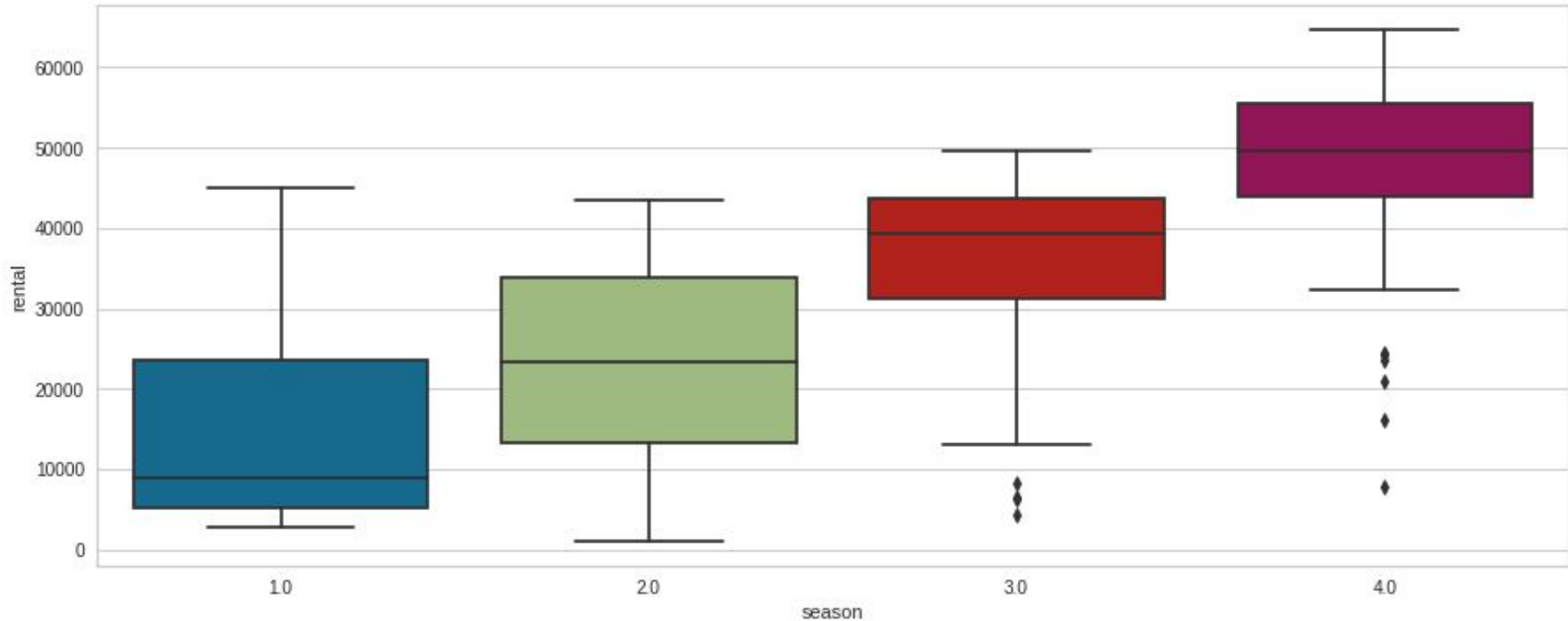


12개의 X(feature)가 존재,
전부 날씨 관련 데이터



다수의 결측치 확인

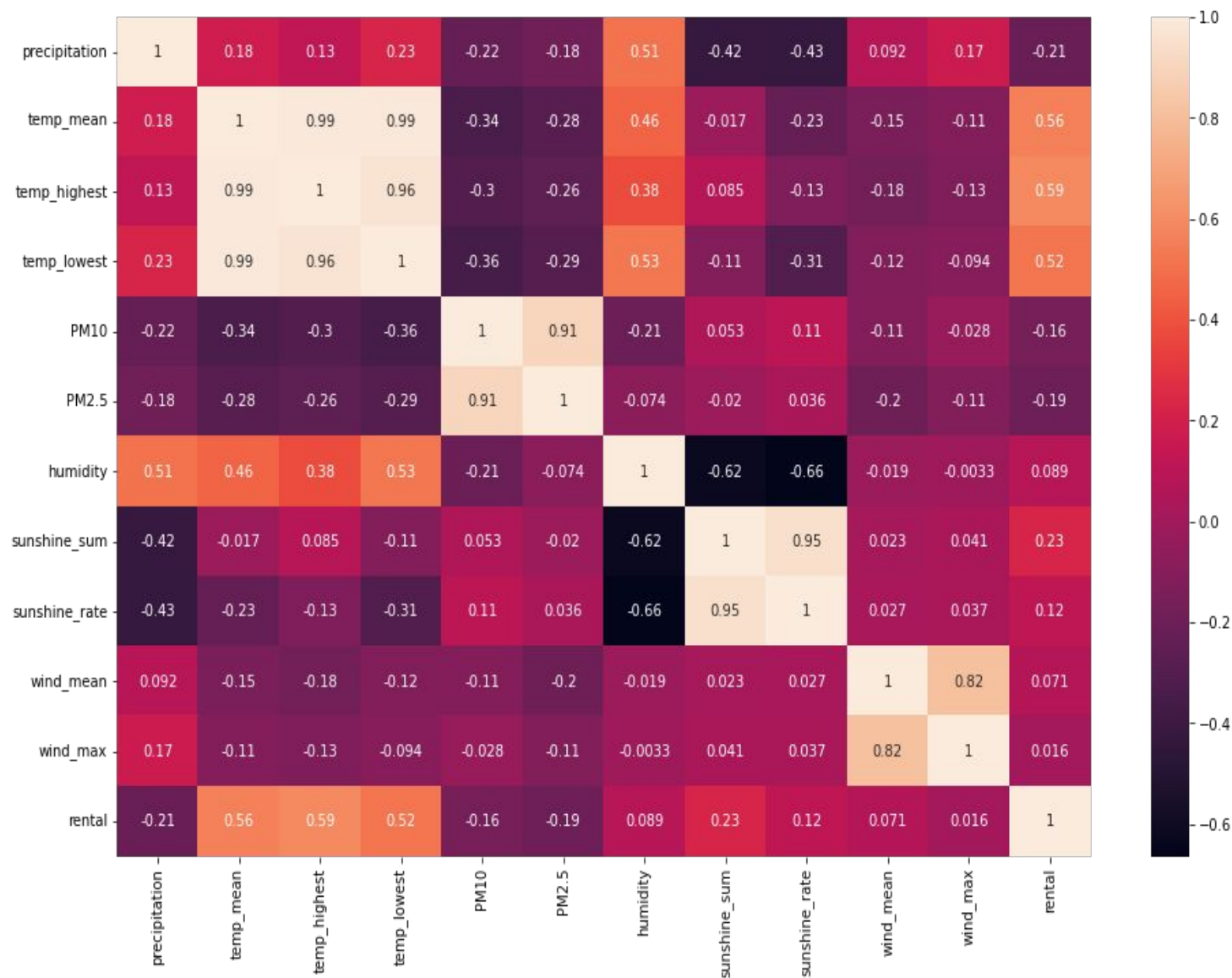
결측치 이전 2일 평균으로 대체
(데이터 누수 방지)



season feature에서 특히 **y**와 밀접한 관계가 존재함

Part 2-1

데이터 확인



서로 영향을 끼치는
(다중공선성)이 존재

> PCA나 파생변수를 만들어 공선성을 제거할 수도 있으나, 딥러닝 모델을 사용할 것이기에 제거 x

파생변수 생성

파생 변수와 구간화를 통해 점수 상승

1. 불쾌지수

공식 : $1.8 * \text{온도} - 0.558 * (1 - \text{습도}) * (1.8 * \text{온도} - 26) + 32$

2. 체감온도

공식: $13.12 + (0.6215 * \text{온도}) - (11.37 * \text{풍속}^{0.16}) + (0.3965 * \text{온도} * \text{풍속}^{0.16})$

3. 강수량 / 기온 / 풍속... 구간화

공식적으로 지정한 기점으로 구간화

Scaling

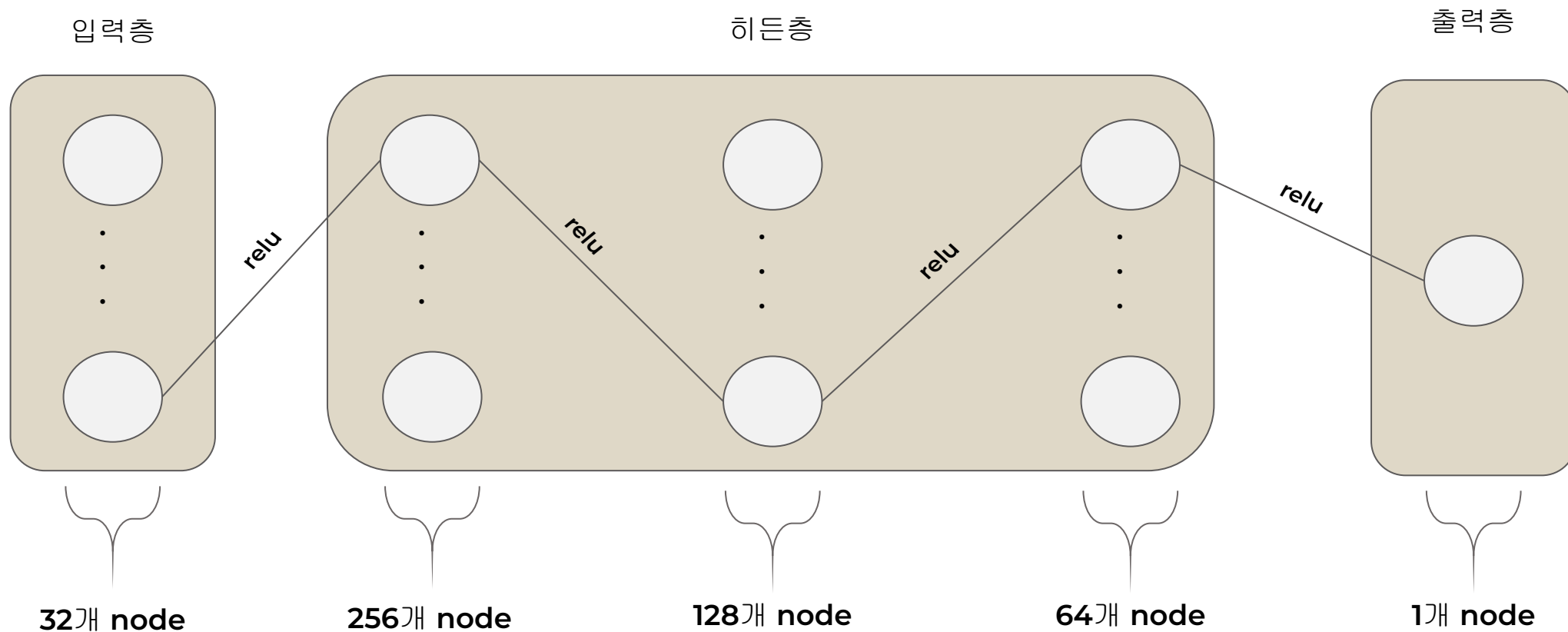
sunshine_sum	sunshine_rate	wind_mean
7.4	61.2	1.6
11.1	79.9	2.2
12.0	85.1	1.5
7.3	72.3	2.7
0.2	1.8	2.0



sunshine_sum	sunshine_rate	wind_mean
0.384267	0.921466	-0.903973
0.283363	0.793470	-0.314688
0.459944	0.988511	0.274596
-0.145477	0.232729	-0.903973
0.359041	0.860516	-0.462010

feature별 단위를 맞춰주기 위해 **Scale** > 성능향상

Modeling



optimizer : adam
k-fold : 3

learning_rate = 0.0001
epoch : 250

Earlystop = 10
batch_size = 1

결과 및 아쉬운 점

결과

- 다른 팀들과는 다르게 딥러닝으로 접근하여 점수를 높임
- **17등 / 94팀** 이라는 준수한 성적 달성

아쉬운 점

1. 여러 모델들을 **양상블** 해보지 못한 점
2. 모델 **최적화**를 진행하지 못한 점
3. 연도별 **증가치**를 **적용**해 보지 못한 점

