

천하제일 머신러닝 경진대회

따릉이 대여량 예측

머신러닝

임경태

서울시 따릉이 대여량 예측 경진대회

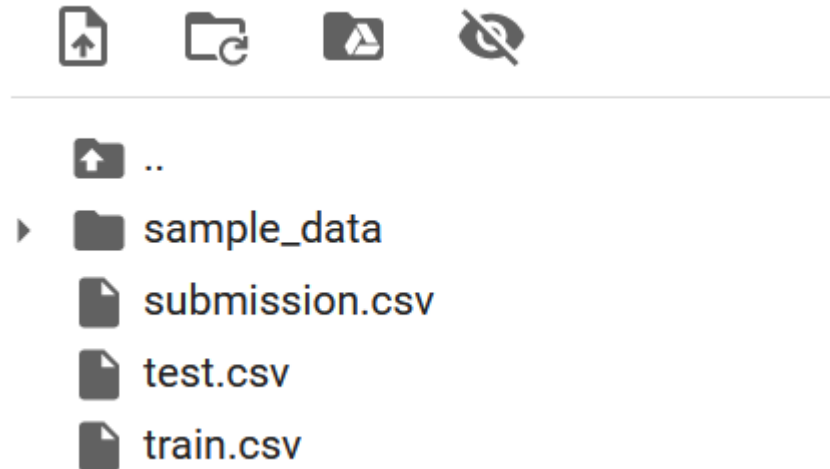
1. 데이터 다운로드 링크 <https://www.dacon.io/competitions/open/235576/data>

대회 참가 후 데이터 다운 가능

2. 참고 코드

<https://drive.google.com/file/d/1Lxl7dhxnYeE81uGauWWNtobYpbmJ88j5/view?usp=sharing>

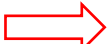
<https://www.dacon.io/competitions/open/235576/codeshare/1535?page=1&dtype=recent>



데이터 다운로드 후 그림처럼 위치시키면
참고 코드 실행 가능

서울시 따릉이 대여량 예측 경진대회 - Data description

	id	hour	hour_bef_temperature	hour_bef_precipitation	hour_bef_windspeed	hour_bef_humidity	hour_bef_visibility	hour_bef_ozone	hour_bef_pm10	hour_bef_pm2.5	count
0	3	20	16.3	1.0	1.5	89.0	576.0	0.027	76.0	33.0	49.0
1	6	13	20.1	0.0	1.4	48.0	916.0	0.042	73.0	40.0	159.0
2	7	6	13.9	0.0	0.7	79.0	1382.0	0.033	32.0	19.0	26.0
3	8	23	8.1	0.0	2.7	54.0	946.0	0.040	75.0	64.0	57.0
4	9	18	29.5	0.0	4.8	7.0	2000.0	0.057	27.0	11.0	431.0
...
1454	2174	4	16.8	0.0	1.6	53.0	2000.0	0.031	37.0	27.0	21.0
1455	2175	3	10.8	0.0	3.8	45.0	2000.0	0.039	34.0	19.0	20.0
1456	2176	5	18.3	0.0	1.9	54.0	2000.0	0.009	30.0	21.0	22.0
1457	2178	21	20.7	0.0	3.7	37.0	1395.0	0.082	71.0	36.0	216.0
1458	2179	17	21.1	0.0	3.1	47.0	1973.0	0.046	38.0	17.0	170.0

- id : 날짜와 시간별 id
- hour_bef_temperature : 1시간 전 기온
- hour_bef_precipitation : 1시간 전 비 정보, 비가 오지 않았으면 0, 비가 오면 1
- hour_bef_windspeed : 1시간 전 풍속(평균)
- hour_bef_humidity : 1시간 전 습도
- hour_bef_visibility : 1시간 전 시정(視程), 시계(視界)(특정 기상 상태에 따른 가시성을 의미)
- hour_bef_ozone : 1시간 전 오존
- hour_bef_pm10 : 1시간 전 미세먼지(머리카락 굵기의 1/5에서 1/7 크기의 미세먼지)
- hour_bef_pm2.5 : 1시간 전 미세먼지(머리카락 굵기의 1/20에서 1/30 크기의 미세먼지)
- count : 시간에 따른 따릉이 대여 수  Target

- 리더보드

- a. 평가 산식 : RMSE

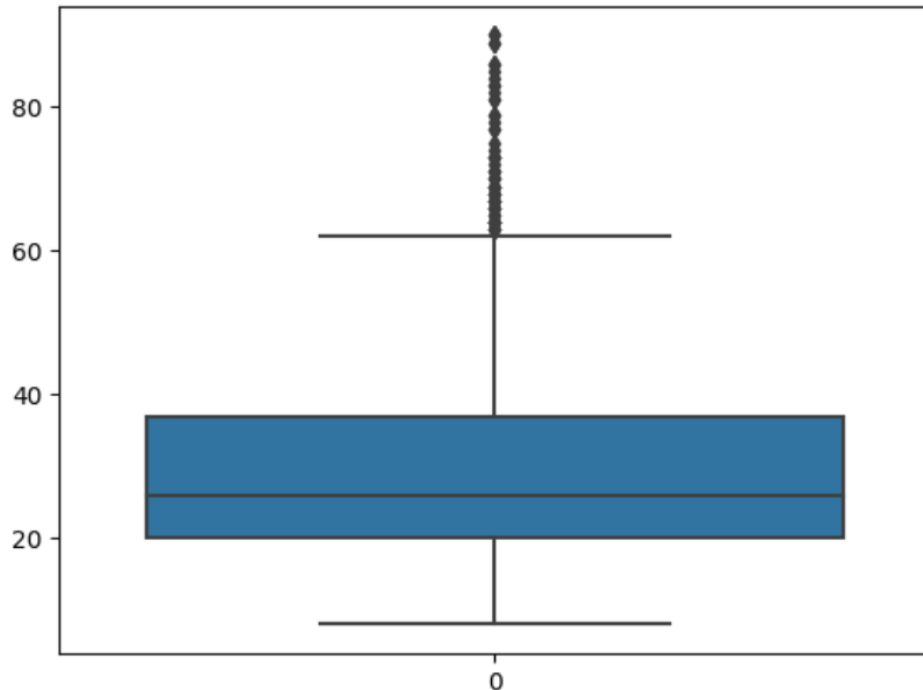
- RMSE (Root Mean Square Error)

$$RMSE(\theta) = \sqrt{MSE(\theta)} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

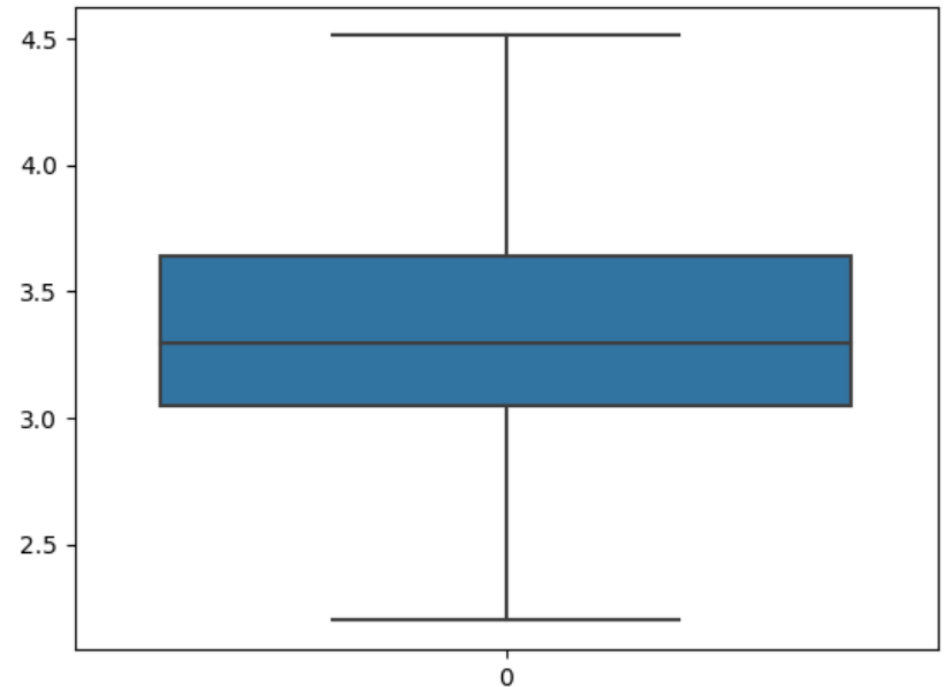
서울시 따릉이 대여량 예측 경진대회

Hint

- 결측값 확인 : `data.isna().sum()`
- 결측값 처리 : `dropna()`, `fillna()`
- 이상치 확인 : `data.boxplot(column_name)`
- 이상치 처리 : 직접 제거, `np.sqrt()`, `np.log1p()` 등



Log 변환



서울시 따릉이 대여량 예측 경진대회

Hint

- Scaler 사용
- data.corr()을 통해 correlation 확인
- train[['hour', 'count']].groupby('hour').mean().plot() – 시간별 대여량 그래프
- validation set 만들기
- Kfold
- Ensemble



감사합니다