

데이터 셋 구성

이동시간	이동시간_lag1	이동시간_lag2	이동시간_lag3	이동시간_lag4	이동시간_lag5	이동시간_lag6	이동시간_lag7	이동시간_lag24	resid	trend	seasonal
440.066000	281.550000	282.516000	287.216000	273.708000	282.658000	285.588667	274.860667	251.250	87.930493	205.137653	-11.518146
384.191500	312.088333	281.550000	282.516000	273.708000	282.658000	285.588667	251.250	119.125396	209.942219	-16.979282	-19.749380
351.044333	317.525000	312.088333	281.550000	282.516000	287.216000	273.708000	282.658000	251.250	122.527596	214.746785	-20.499776
355.820500	325.441000	317.525000	312.088333	281.550000	282.516000	287.216000	273.708000	258.983	126.389426	219.551351	-20.601312
366.750000	268.016667	325.441000	317.525000	312.088333	281.550000	282.516000	287.216000	288.466	64.262062	224.355917	11.631671
265.266500	238.258000	260.816000	248.955333	257.077667	252.133000	286.858250	256.888667	265.783	-49.706883	276.333212	9.497546
271.666000	241.400000	238.258000	260.816000	248.955333	257.077667	252.133000	286.858250	261.250	-43.404726	275.307181	7.512458
270.850000	244.600000	241.400000	238.258000	260.816000	248.955333	257.077667	252.133000	248.633	-37.426670	274.428899	8.667045
270.016000	244.600000	244.600000	241.400000	238.258000	260.816000	248.955333	257.077667	248.633	-36.553170	273.640712	
262.616000	244.600000	244.600000	244.600000	241.400000	238.258000	260.816000	248.955333	324.066	-37.276861	273.209816	

- 고속도로 공공 데이터 포털에서 도시간 소요시간 데이터 크롤링
- 24이전 시점부터 48 이전시점까지 모두 shift하여 데이터 프레임 구성
- 이 중, 24시간 이전 시점 데이터를 이용하여 시계열 분해를 시행
- 시계열 분해로 얻은 trend, seasonal, residual을 모두 하나의 열로 만들어 병합
- 가설 : 24시간~48시간 이전 시점 데이터와 trend,seasonal, residual은 다음날 도착 시간과 영향이 있다.

검증 및 하이퍼 파라미터 튜닝

```
# 베이지안 옵티마이저 결과
```

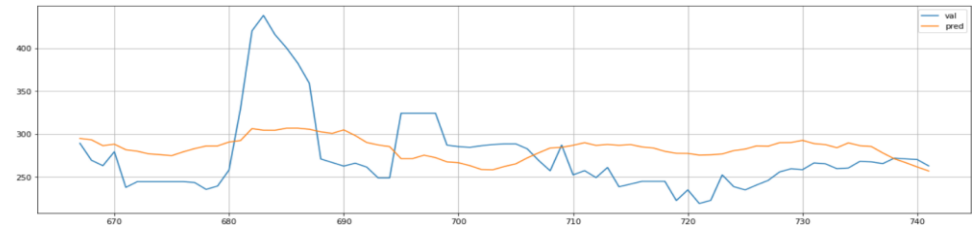
#	iter	target	baggin...	depth	learn...	subsample
#	1	-0.1227	4.753	11.8	0.01011	0.3093
#	2	-0.1353	2.321	2.385	0.1944	0.3521
#	3	-0.1305	4.571	9.082	0.425	0.6884

- Train set을 90%, Test set을 10%로 놓고 검증을 실시
- 시계열 데이터이므로 데이터를 랜덤하게 섞지 않음
- 하이퍼 파라미터 튜닝은 베이지안 옵티마이저를 사용함
- 시간의 여유가 없었기 때문에 베이지안 옵티마이저에서 추출된 3개의 경우의 수만 이용

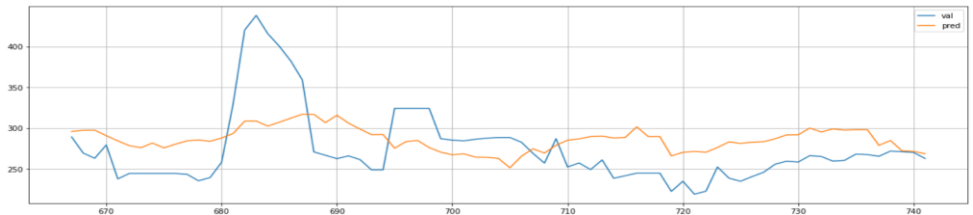
BEST Model과 앙상블

BEST 1 MODEL MAPE	0.13091643358837007
3 MODEL 앙상블	0.13574296587756307

- BEST 1 MODEL



- 3 MODEL 앙상블



- 앙상블을 한 모델보다 성능이 가장 좋은 모델 하나를 선택하는 것이 더 좋게 나왔다.
- 크게 차이는 나지 않지만, 성능이 가장 좋은 모델을 선택하고 예측하였다.

결론

- 예측 결과, 300.5229270130459분이 나왔다.
- 그러므로, 반올림해서 5시간 1분으로 정했다.
- 12시에 출발하면, 17시 1분에 도착한다!