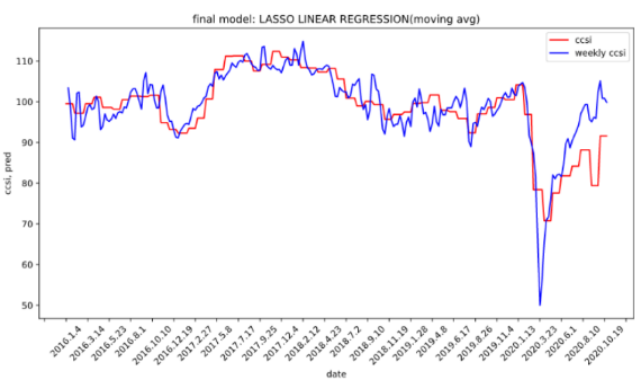


14주차 회의록

회의일시	2020년 11월 30일 (월요일) 13:00 ~ 13:30	장소	zoom
팀원	박채린, 이혜승		
회의안건	<ul style="list-style-type: none"> • 각 전처리+모델에서 사용하는 train, test set이 다른데, test rmse 및 test mae로 비교해도 되는지 • 생성한 주간지표의 cci의 선행 여부 		

논의 및 결정사항	내용
	<ul style="list-style-type: none"> • 상관계수 test set 으로만 하기 • cci 선행여부는 ccsi 조작 cci 에 선행을 보이지 않으므로 어쩔 수 없음. • 유효성을 보이기 위해서는 data 들을 shuffle 하면 안됨. -> shuffle false 을 보면 2020 년 예측이 좋지는 않지만 어느정도 추세를 따르는 것은 알 수 있음. 또한, 라쏘 모델이므로 어느정도 해석이 가능함. 코로나로 인해 사람들의 심리는 살아나지 않았는데 검색은 많이 했을 수 있음 등등으로 예측 성능 자체의 좋고 나쁨 보다 왜 좋지 않았는지 변수들을 추가적으로 살펴보기(사후 분석) • 11 월 데이터들 추가하지 않기: google 과 naver 가 비율이므로 기존 11 월 데이터 추가하지 않은 data set 과 같지 않고 전체적으로 변경됨 -> 변인 통제된 것이 더 좋기 때문에 기존 데이터들만 이용하기 • t-test 에서 p-value 해석 조심하기: 10 개가 독립적이지 않으므로 실제적인 신뢰구간이 더 길어져 p-value 가 더 커질 수 있음. 따라서 p-value 기각할 때 조심하고 오차범위 안에서 거의 차이가 없다는 전제조건 추가하기. <p>(최종 모델)</p> <ul style="list-style-type: none"> • 전처리 방법: mean • 모델: lasso(alpha = 0.001)  <p>(할 일)</p> <ul style="list-style-type: none"> • 1. 파이널 모델 다시 pkl로 저장하기: 혜승 • 2. 파이널 모델에 전체 주간 데이터 다 넣어서, 시각화하기 + 라쏘의 모델 해석: 혜승 • 3. 조금 예측이 잘 안되는 2020년에 대해 사후 분석하기: 채린 • 4. 유효성 입증부분에서, test셋으로만 상관계수 확인하기해서, step 5, 6까지 해오기: 채린
공지사항	<ul style="list-style-type: none"> - 다음 회의: 12/2 수요일 14:00 - 교수님 미팅 마지막: 12/7 월요일 15:20

13 주차(11/23 ~ 29)	<ul style="list-style-type: none"> • 전처리 및 분할방법에 따른 best model 총 6 개 생성 후, test 결과 시각화 및 모델 평가(RMSE 비교)
-------------------	--

14 주차(11/30 ~ 12/6)	11/30 교수님 미팅 <ul style="list-style-type: none"> ● 주간 지표의 유효성 입증 ● 발표 준비 및 영상 제작 ● 보고서 작성
15 주차(12/7 ~12/13)	12/7 교수님 미팅 12/11 프로젝트 발표 영상 제출