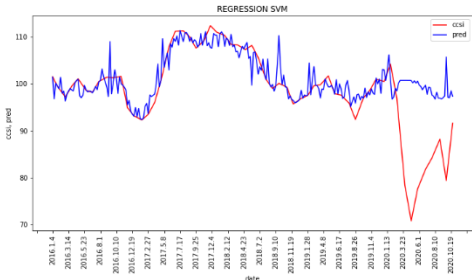
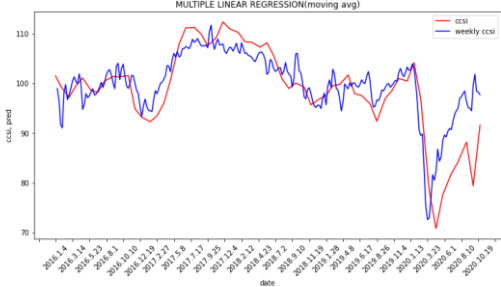
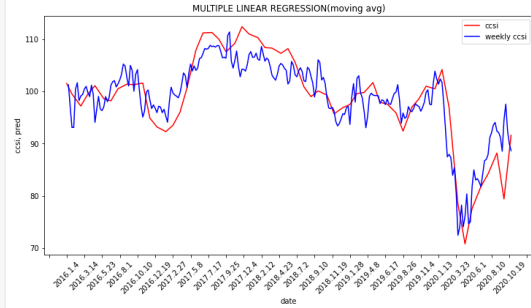


# 12주차 회의록

회의일시	2020년 11월 21일 (토요일) 14:30 ~ 15:00	장소	온라인
팀원	박채린, 이혜승		
회의안건	<ul style="list-style-type: none"> <li>다양한 파라미터를 고려하는 모델들은 어떻게 k-fold cv를 적용하는 지</li> <li>T-test 결과로 귀무가설 채택이 나오는데, 어떤 결론을 낼 지</li> </ul>		

논의 및 결정사항	내용	비고
	<p>(이번주 해온 것)</p> <ul style="list-style-type: none"> <li>전처리 방법 여러가지 적용(채린) [<a href="#">코드</a>]</li> </ul>  <p>[ 전처리 - 내삽, 분할 - 2020년을 test셋으로, 모델: svm regressio]</p> <ul style="list-style-type: none"> <li>2020 년을 test set 으로 떼어놓으니, svm regression 의 경우에는 거의 제대로 추정하지 못함.</li> </ul>  <p>[ 전처리 - 내삽, 분할 - 2020년을 test셋으로, 모델: multi regression]</p> <ul style="list-style-type: none"> <li>2020 년을 test set 으로 떼어놓으니, multiple linear regression 의 경우 2020 년 추세는 따라가지만 실제 값과의 차이 존재. 선행도 존재하는 것으로 보임.</li> </ul>  <p>[ 전처리 - 내삽, 분할 - 랜덤, 모델: multi regression]</p> <ul style="list-style-type: none"> <li>train, test data 를 랜덤하게 뽑은 결과 내삽하기 전 mean 값 적용한 것과 비슷함.</li> </ul>	

	<ul style="list-style-type: none"> <li>● 모델의 성능에 차이가 있는지 확인하는 t-test 진행[코드][참고 1][참고 2] : 전처리 - 평균 기준</li> <li>· Model1: 다중선형회귀모델, Model2: 랜덤포레스트, Model3: GAM, Model4: Lasso</li> <li>· model1 과 model2 검정 결과. H0: 두 모델의 성능이 같다(RMSE)를 높은 p-value 로 채택하게 된다. 모든 검정에서 비슷한 결과가 나오게 된다면, 모델에 대한 이해도, 설명력이 가장 높은 선형회귀모델을 사용하게 될 듯.</li> </ul> <p>(할 일)</p> <ul style="list-style-type: none"> <li>● GAM 모델링 코드</li> <li>● Parameter 서치마다 k-fold cv하는 코드 구현해오기</li> <li>● 주간 지표의 유효성 입증             <ol style="list-style-type: none"> <li>1. Ccsi 의 추세를 따르는지 시각화</li> <li>2. 전체 레코드에 대한 rmse</li> <li>3. GDP, cci 순환변동치를 선행되는지를 확인.</li> </ol> </li> </ul>	
공지사항	<ul style="list-style-type: none"> <li>- 교수님 미팅 날짜: 11/30 월요일 13:00, 12/7 월요일 15:20</li> <li>- 다음 회의: 11/23 월요일 20:00                 <ul style="list-style-type: none"> <li>✓ Parameter 서치마다 k-fold cv하는 코드 구현해오기 (혜승)</li> <li>✓ GAM 모델링 코드 (채린)</li> </ul> </li> </ul>	

	계획
10 주차(11/2 ~ 8)	11/4 교수님 미팅 전 <ul style="list-style-type: none"> <li>● <del>뉴스 기사 수 키워드별 DB table 저장.</del></li> <li>● <del>카카오, 네이버, 구글의 포털트렌드 검색어 비율 변수 유의성 확인(시각화, 상관계수, 회귀분석)</del></li> </ul> 11/4 교수님 미팅 후 <ul style="list-style-type: none"> <li>● <del>모든 변수들을 활용해서 multi regression 모델링</del></li> </ul>
11 주차(11/9 ~ 15)	<ul style="list-style-type: none"> <li>● 모든 변수들을 활용해서 모델링 및 weekly CCSI 예측               <ul style="list-style-type: none"> <li>~ <del>Multiple linear regression</del></li> <li>~ <del>Ridge, Lasso linear regression</del></li> <li>~ <del>Decision tree</del></li> <li>~ <del>Random forest</del></li> <li>~ <del>SVM</del></li> </ul> </li> </ul>
12 주차(11/16 ~ 22)	11/18 교수님 미팅 <ul style="list-style-type: none"> <li>● <del>전처리 방법 여러가지 적용해보기(채린)</del></li> <li>● <del>10 fold CV 코드 다시 구현(혜승)</del></li> <li>● 다양한 모델 더 고려! (Spline(GAM)) (채린)</li> <li>● 모델 test 결과 시각화 및 모델 평가(RMSE 비교)</li> <li>● 주간 지표의 유효성 입증</li> <li>● 보고서 작성 시작!!</li> </ul>
13 주차(11/23 ~ 29)	프로젝트 완성 목표> <ul style="list-style-type: none"> <li>● 유의성 확인(소비자심리지수와 비교)</li> </ul>

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