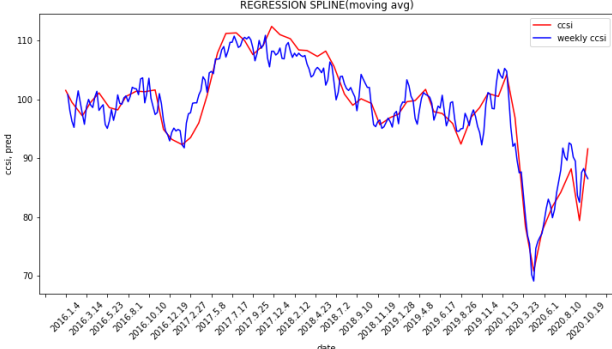


12주차 회의록

| | | | |
|------|---|----|-----|
| 회의일시 | 2020년 11월 26일 (목요일) 20:00 ~ 21:50 | 장소 | 온라인 |
| 팀원 | 박채린, 이혜승 | | |
| 회의안건 | <ul style="list-style-type: none"> 다양한 파라미터를 고려하는 모델들은 어떻게 k-fold cv를 적용하는 지 T-test 결과로 귀무가설 채택이 나오는데, 어떤 결론을 낼 지 | | |

| 논의 및 결정사항 | 내용 |
|-----------|---|
| | <p>(이번주 해온 것)</p> <ul style="list-style-type: none"> GAM 모델 적합 및 예측(채린) [코드]  <p>[전처리 - 내삽, 분할 - shuffle해서 train:test 분할, 모델: GAM]</p> <ul style="list-style-type: none"> 모델의 성능에 차이가 있는지 확인하는 t-test 진행 <ul style="list-style-type: none"> H_0: model A 의 rmse = model B 의 rmse 즉, 두 모델의 성능에 차이가 없다. H_1: 두 모델의 성능에 차이가 있다. <p>(할 일)</p> <ul style="list-style-type: none"> 채린: shuffle = True 혜승: shuffle = False <p>해서, step 5, 6까지 해오기!</p> <p>+) 6개의 best model중 어떤 것을 고를지 평가하는 척도 찾아보기</p> |
| 공지사항 | <ul style="list-style-type: none"> 교수님 미팅 날짜: 11/30 월요일 13:00, 12/7 월요일 15:20 다음 회의: 11/29 일요일 14:00 |

| | 계획 |
|-----------------|---|
| 10 주차(11/2 ~ 8) | <p>11/4 교수님 미팅 전</p> <ul style="list-style-type: none"> 뉴스 기사 수 키워드별 DB table 저장. 카카오, 네이버, 구글의 포털트렌드 검색어 비율 변수 유의성 확인(시각화, 상관계수, 회귀분석) <p>11/4 교수님 미팅 후</p> <ul style="list-style-type: none"> 모든 변수들을 활용해서 multi regression 모델링 |

| | |
|---------------------|---|
| 11 주차(11/9 ~ 15) | <ul style="list-style-type: none"> ● 모든 변수들을 활용해서 모델링 및 weekly CCSI 예측 — Multiple linear regression — Ridge, Lasso linear regression — Decision tree — Random forest — SVM |
| 12 주차(11/16 ~ 22) | 11/18 교수님 미팅 <ul style="list-style-type: none"> ● 전처리 방법 여러가지 적용해보기(채린) ● 10 fold CV 코드 다시 구현(혜승) ● 다양한 모델 더 고려! (Spline(GAM)) (채린) |
| 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) ● 주간 지표의 유효성 입증 ● 보고서 작성 |
| 14 주차(11/30 ~ 12/6) | 11/30 교수님 미팅 <ul style="list-style-type: none"> ● 발표 준비 및 영상 제작 |
| 15 주차(12/7 ~ 12/13) | 12/7 교수님 미팅 12/11 프로젝트 발표 영상 제출 |