

12주차 회의록

회의일시	2020년 11월 18일 (수요일) 13:30 ~ 15:40	장소	온라인
팀원	박채린, 이해승		
회의안건	<ul style="list-style-type: none"> 모델링 방법 고려 전처리 방법 고려 		

논의 및 결정사항	내용	비고
	<ul style="list-style-type: none"> Issue <ul style="list-style-type: none"> : random state 조정의 의미가 없는 듯(두가지 상황의 반복) - training set에 CCSI가 낮은 2020년 데이터가 들어가면, 그 부분이 잘 적합되는 대신에, test set에 쉬운 데이터들이 몰려서 unknown fit이 생김. - 반대의 경우에는, 갑자기 불황이 찾아온 순간에 예측이 잘 안됨. → 2016 ~ 2019년을 training set으로, 2020년을 test set으로 활용해보자. (2020년이 경험하지 못한 이벤트인데 잘 예측된다면 좋고, 제대로 되지않더라도 2020년이 갑작스런 상황을 누구나 이해할 수 있을 듯!!) 더 고려해 볼 방법 <ul style="list-style-type: none"> 전처리 방법(주간과 월간을 맞추기 위한 전처리) : <u>채린</u> <ul style="list-style-type: none"> 마지막 주차만 활용 평균(현재 사용하고 있는 방안) 최대/최소 중앙값 기하/조화평균 분위 수 내삽(이건 제일 마지막에 !!) CV(gridsearchCV말고, 다시 구현!) : <u>혜승</u> <ul style="list-style-type: none"> 10-fold CV(10-fold CV의 분포를 찍어보자.) → t - test. 비등비등하다면 linear로 가는게 좋을 듯! (지표의 해석력을 위해) 더 고려할 모형 <ul style="list-style-type: none"> Polynomial Spline(GAM) PCR Quantile regression 보고서 내용 <ul style="list-style-type: none"> Introduction <ul style="list-style-type: none"> : '빅데이터를 이용한 경기판단지표 개발' 저널의 연구 배경 참고 Future work <ul style="list-style-type: none"> : 경기 침체가 있었던 이전 데이터(2008년부터)들을 얻을 수 있다면, 좀 더 잘 적합하고, 경기 침체 상황에서의 경기선행지수를 잘 내줄 것임. 	

	<ul style="list-style-type: none"> · 질의응답? <p>: 인사이트, 상관성을 분석해보고자 하는 것이 아니고 좋은 지표를 개발하는 것이 목표!</p>	
공지사항	<ul style="list-style-type: none"> - 교수님 미팅 날짜: 11/30 12/7 월요일 15:20 - 다음 회의: 11/21 토요일 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)	<p>11/18 교수님 미팅</p> <ul style="list-style-type: none"> ● 전처리 방법 여러가지 적용해보기(채린) ● 10-fold CV 코드 다시 구현(혜승) ● 다양한 모델 더 고려! (Polynomial, Spline(GAM), PCR, Quantile regression) ● 모델 test 결과 시각화 및 모델 평가(RMSE 비교) ● 주간 지표의 유효성 입증 <p>● 보고서 작성 시작!!</p>
13 주차(11/23 ~ 29)	<p>프로젝트 완성 목표></p> <ul style="list-style-type: none"> ● 유의성 확인(소비자심리지수와 비교)
14 주차(11/30 ~ 12/6)	<p>12/2 교수님 미팅</p> <ul style="list-style-type: none"> ● 발표 준비 및 영상 제작 <p>12/11 프로젝트 발표 영상 제출</p>