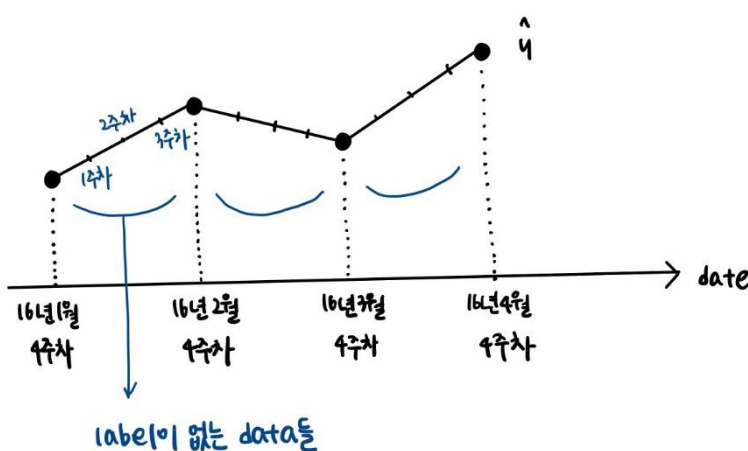


10주차 회의록

회의일시	2020년 11월 4일 (수요일)	장소	학교(교수님 면담)
팀원	박채린, 이혜승		
회의안건	<ul style="list-style-type: none"> 어떻게 월간지표인 CCSI를 예측하고, 주차별 값을 낼 것인지 예측 방법 고려 		

논의 및 결정사항	내용	비고
	<ul style="list-style-type: none"> ● 침체 금융위기 불황 외환위기 폭락을 뉴스 키워드로 변경해서 db table에 저장 (키워드 선택기준: recession(침체) 단어 기준으로 상관계수가 높은 것들) [코드] ● 주제를 예측문제로!! \hat{y}(주간 데이터 레코드로 CCSI를 예측한 값) y(월말에 제공되는 CCSI소비자심리지수) → 월의 가장 마지막 주차의 레코드들만 가지고 train, val, test를 진행한다. (예측에 날짜가 사용되지 않기 때문에, 랜덤으로 섞어서 train, val, test 분할과 CV가 가능함.) → 이때 마지막 주차의 변수값들 말고, 해당 월의 주차들의 평균값 사용해보자 (-) 모델링과 평가에 활용할 수 있는 데이터 수가 현저히 적음. challenging으로 남길 수 밖에.. ● 예측 방법으로 먼저 multi regression을 시도하고, CCSI를 예측해보자. 예측 방법 <ol style="list-style-type: none"> multi regression(먼저 시도!) decision tree(회귀) random forest 내삽: 월의 마지막 주차의 심리지수를 예측하여, 점 사이를 직선으로 연결한 후 4등분해서, 차례로 1, 2, 3주차의 심리지수로 예측한다.  v. Spline: 월의 마지막 주차의 심리지수를 예측하여, smooth하게 연결. 	

	<ul style="list-style-type: none"> ● 포털 트렌드 변수들의 다중공산성은 고려하지 않기로!(네이버, 구글 변수 활용) (예측 오차를 줄일 수 있는 것을 최우선으로 고려) 	
공지사항	<p>할 일)</p> <ol style="list-style-type: none"> 1. 모든 변수들을 활용해서 모델링 및 단기지표생성 : 구글, 네이버, 키워드1 ~ 5 2. P-value로 포털 트렌드와 뉴스 키워드 변수들 중 유의한 것 골라서 모델링 및 단기지표생성 	

	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 및 CCSI 예측
11 주차(11/9 ~ 15)	<ul style="list-style-type: none"> ● 경제 단기 지표 생성
12 주차(11/16 ~ 22)	<p>11/18 교수님 미팅</p> <ul style="list-style-type: none"> ● 경제 단기 지표 생성
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>