

7주차 회의록

회의일시	2020년 10월 18일 (일요일) 14:00 – 15:00	장소	Zoom
팀원	박채린, 이혜승		
회의안건	<p><문제></p> <ul style="list-style-type: none"> - 키워드 5개를 어떻게 활용할 것인가? -> 같은 키워드들의 대푯값(ex. 평균)을 이용하여 모델링 변수로 이용한다. - 기간이 다른 것은 어떻게 할 것인가? -> 질문 - 인터넷 뉴스를 많이 읽고 제공하는 시기 이전의 자료들에 가중치를 주고싶는데 어떻게 줄 것인가? -> 네이버에 있는 전체 뉴스 건수를 크롤링 하여 비율을 이용한다. 뉴스 전체 기사 수를 알 수 없어 모든 뉴스 기사에 들어가는 단어인 '기자'를 검색하여 크롤링한다. <p><목표></p> <ul style="list-style-type: none"> - 단기 지표 생성(주간 지표) : 변수로 뉴스 카운팅, 포털 검색 트렌드 비율 사용 → 소비자심리지수(월별) 예측(하고 싶은데, 실제로는 예측 변수가 더 필요할 것 같아서, 같은 추세를 보이는 지 정도를 확인) - <중간보고서 이전> 사람들이 경제에 관심을 갖는다(포털 트렌드 이용) = 경제 불황이다. 소비자 심리지수와 관계 파악(이건 이미 불러옴) or 다른 경제 지표(경기동행지수) openAPI로 불러오기. <p><한 일></p> <ul style="list-style-type: none"> - 구글 csv 파일 다운로드하여 DB 생성[코드] - 카카오 csv 파일 다운로드하여 DB 생성[코드] - 네이버 api 이용하여 크롤링 및 DB 생성[코드] - 한국은행 open api 이용하여 소비자심리지수 DB 생성[코드] <p><할 일></p> <ul style="list-style-type: none"> - 뉴스 전체 기사 수 읽기 -> 채린 - DB 정리(year, month, day 외래키 설정) -> 채린 - 소비자심리지수, 경기동행지수 api 이용해서 자료 가져와 DB 만들기 ->혜승 - 시간이 되면 포털 트렌드 '경제' 키워드 ratio 와 경기동행지수의 상관관계 확인 및 시각화 (일변량 회귀분석) 		

논의 및 결정사항	내용	비고
	<ul style="list-style-type: none"> - 각 키워드 활용법 : 키워드별 평균 / 대푯값 산출 : 포털마다 제공하는 검색어 트렌드 기간 다름 질문 - 뉴스 카운팅 가중치 처리(어떤 기준으로 해야 할까) : 연도별 뉴스 데이터 수 자체를 더해서 weight주면 어떨까 → 뉴스 기사 수 스크래핑('기자'로 카운트) - 경제 불황임을 파악하는데 사용할 변수 :경기동행지수 	
공지사항	<ul style="list-style-type: none"> - 중간 보고서 작성 완성일: 토요일 - 포스터 제작: 일요일 	

(목표) 단기 지표 생성(주간 지표)

: 변수로 뉴스 카운팅, 포털 검색 트렌드 비율 사용 → 소비자심리지수(월별) 예측(하고 싶은데, 실제로는 예측 변수가 더 필요할 것 같아서, 같은 추세를 보이는 지 정도를 확인)

Process

키워드에 초점을 맞추자. 키워드별 평균 / 대푯값 산출 → **기간이 다른 걸 어떻게 해결?**.

1. 증명

<중간보고서 이전> 사람들이 경제에 관심을 갖는다(포털 트렌드 이용) = 경제 불황이다.

소비자 심리지수와 관계 파악(이건 이미 불러옴) or 다른 경제 지표(경기동행지수) openAPI 로 불러오기.

2. 단기지표를 만들겠다.

→ 주 단위 지표를 만들어서, 매월 말(25~29 일정도)에 나오는 소비자심리지수를 예측할 수 있을 것 같음. 혹은 비슷한 추세를 갖는지 확인 정도. (예측하기위해서는 다른 정보들이 더 필요할 것 같다.)

주간 지표가 필요한 이유:

- 단기적인 경제상황을 알고 싶을 때.
- 경제정책 대안 제안 혹은 시행을 했을 때, 피드백을 빨리 받아 보고 싶을 수 있음. 이런 상황에서 탄력적 대응이 가능함.