

데이터 분석 계획서

구분	내용		비고			
팀명	이건뫼조					
팀원(역할)	김예진 - 추천 서비스 모델링 및 적용 최윤석 - 뉴스 크롤링 및 데이터 전처리, 분류, 발표 송현근(팀장) - 뉴스 크롤링 및 데이터 전처리, DB 모델링 한유진 - 챗봇 서비스 연동 및 페이지 구축, PPT					
기간	2021.07.22. ~ 2021.07.28. (07.28 비대면 발표)					
제목	뉴스 콘텐츠 큐레이션 서비스 - 뉴스켓					
분석주제	뉴스 추천 서비스 알고리즘 분석					
분석목적	뉴스 모음에 대한 편향을 배제하고 개인 취향에 따라 적합한 뉴스 콘텐츠를 제공할 수 있는 서비스를 개발하고자 함.					
분석배경	네이버의 실시간 검색어가 없어지고 실시간 이슈 및 다수의 사람들이 이 관심 갖는 주제에 대해 직접적으로 노출될 기회가 적어짐. 또한 기존 포털들의 편향적인 뉴스 노출 배열에 대한 문제도 존재. 이에 따라 편향없이 개별 취향에 따라 카톡 챗봇으로 편리하게 맞춤형 뉴스 콘텐츠 큐레이션을 제공할 수 있는 서비스를 개발하고자 함.					
분석대상	온라인 뉴스 기사 콘텐츠					
분석데이터	웹크롤링을 통한 기사 자료 획득 - 한국일보 https://www.hankookilbo.com/ - 중앙일보 https://joongang.joins.com/ - 조선일보 https://www.chosun.com/ - 동아일보 https://www.donga.com/ - 한겨레 http://www.hani.co.kr/ - 경향신문 http://www.khan.co.kr/ - KBS https://news.kbs.co.kr/common/main.html - MBC https://imnews.imbc.com/pc_main.html - SBS https://news.sbs.co.kr/news/newsMain.do - 네이버 뉴스 https://news.naver.com/main/main.naver?mode=LSD&mid=shm&sid1=100					
분석방법	- 뉴스 기사 크롤링 - 크롤링한 뉴스 data 전처리 및 분류 - 콘텐츠 기반 추천 알고리즘 모델 활용 및 적용					
분석결과활용	- Daily 큐레이션 콘텐츠에 대해 지속적인 user 피드백을 수집하여 유저별 모델 고도화 및 초개인화 추천 서비스로 발전					
SWOT 분석	<table><tr><td></td><td>Opportunity[기회] · 실시간 검색어 폐지로 인한 사람들의 뉴스 관심도 하락 · 크롤링한 데이터를 통해 다양한 방향으로 서비스 제공 가능</td><td>Threat[위협] · 언론사마다 다른 성향 · 기자들의 자극적인 단어 선택 · 넉넉하지 않은 프로젝트 기간</td></tr></table>			Opportunity[기회] · 실시간 검색어 폐지로 인한 사람들의 뉴스 관심도 하락 · 크롤링한 데이터를 통해 다양한 방향으로 서비스 제공 가능	Threat[위협] · 언론사마다 다른 성향 · 기자들의 자극적인 단어 선택 · 넉넉하지 않은 프로젝트 기간	
	Opportunity[기회] · 실시간 검색어 폐지로 인한 사람들의 뉴스 관심도 하락 · 크롤링한 데이터를 통해 다양한 방향으로 서비스 제공 가능	Threat[위협] · 언론사마다 다른 성향 · 기자들의 자극적인 단어 선택 · 넉넉하지 않은 프로젝트 기간				

	<p>Strength[강점]</p> <p>다양한 전공으로 인해 각자의 장점을 살리는 역할분담 조원들이 트렌드에 민감한 20대 중반 텍스트마이닝 관련 프로젝트 경험이 있는 인원</p>	<p>SO 전략</p> <p>- 실시간 검색어를 대체할 수 있게 텍스트마이닝을 통해 핫키워드 노출</p>	<p>ST 전략</p> <p>- 길지 않은 프로젝트 기간이지만 각자의 장점을 살려 신속한 프로젝트 진행</p> <p>- 다양한 단어들이 나오지만 관련 프로젝트 경험이 있어 수월한 전처리 가능</p>	
	<p>Weakness[약점]</p> <p>. 각 분야 전문용어 지식 부족 . 크롤링을 통한 텍스트 전처리 기준이 애매함 . 추천 알고리즘 모델 구현 복잡 . 불용어set 직접 생성</p>	<p>WO 전략</p> <p>- 프로젝트 진행중 뉴스 관련 전문용어 지식 습득 - 크롤링한 데이터를 전처리 하기 위한 기준을 조원 간 의논을 통해 확고히 하여 만들어진 데이터는 활용방안이 다양할 것이라 예상됨</p>	<p>WT 전략</p> <p>- 프로젝트 기간내 알고리즘 모델 구현이 불가능할 경우 클러스터링에 집중</p> <p>- 기자들의 사건이나 자극적인 단어들을 불용어set에 추가하여 필터링</p>	