구분	질문	답변
크롤링/데이터 수집	데이터 크롤링 목적 및 데이터 분석/시각화 목적	1. 크롤링 목적: 정신건강정책 키워드로 검색한 네이버뉴스/블로그 텍스트 데이터 2. 데이터 분석 목적: 정신건강 정책과 방향 제안
	크롤링 사이트 링크 화면 캡쳐	**Figure 1**
	크롤링 데이터 요소	네이버 블로그: 내용 네이버 뉴스: 날짜, 뉴스 제목, 링크, 내용
	크롤링 데이터 타입	DataFrame을 csv 파일로 저장
	크롤링 데이터 크기	네이버 블로그: 49,408문단 네이버 뉴스: 2,603개
	크롤링 데이터 파일 원본	naver_blog_정신건강정책_startdata_enddate.csv(12개, 1개월 단위) 정신건강정책_startdata_enddate.csv(6개, 2개월단위)
	크롤링 시 사이트 특성 상 어려웠던 점	1. 네이버뉴스 및 블로그가 출력되는 최대 페이지 수의 제한이 있었기 때문에 한 번에 1년치의 데이터를 크롤링하기가 어려웠음. 2. 따라서 네이버 블로그는 1개월 단위로, 네이버 뉴스는 2개월 단위로 크롤링을 진행하였음.
	크롤링 코드 에러 시 해결	1. 네이버 블로그 크롤링을 진행시, 최대 페이지 수를 초과하는 for문을 돌리게 되면, 오류가 발생하였다. 2. 따라서, 최대 페이지만큼 for문을 돌린 후 tqdm 모듈을 사용하여 error가 난 위치를 파악하여 for의 반복횟수를 수정하며 크롤링을 진행하였다.

	크롤링 코드 특징	1. 뉴스데이터는 requests와 BeautifulSoup을 이용한 정적크롤링을 진행하였다. 2. 블로그데이터는 selenium과 chromedriver를 이용한 동적크롤링을 진행하였다.
	크롤링 데이터 파일 전처리 부분	1. blog_preprocessing과 news_preprocessing ipnyb 파일에서 불필요한 데이터 제거, mecab 모듈을 이용한 형태소 태깅과 토큰화를 진행하였다. 이후 모델에 input으로 넣기위한 list 자료형의 데이터를 pickle을 사용하여 따로 저장하였다.
	크롤링 라이브러리, api	requests, BeautifulSoup4, Selenium 등 트위터API의 경우 academic researcher API의 발급을 신청하였으나, Not Aprroved 되어 twitter data는 사용하지 못하였다.(Figure 2)

Figure 1

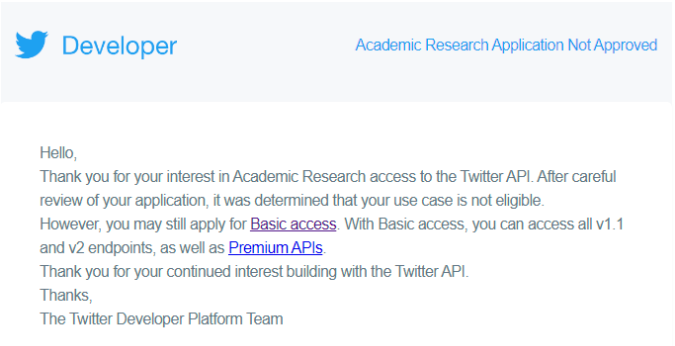

Figure 2