

연세 DX 아카데미: 파이썬 데이터분석 부트캠프

데이터분석 경진대회 기획안

팀명	9 조			
신청자 (대표자)	성명	현시은	학번	20191422214
	소속	공과대학	전화번호	010-8484-9619
참여자	성명	학번	소속	전화번호
	강우진	2023148017	인공지능융합대학	010-4674-9948
	이수향	2022162008	생명시스템대학	010-3380-1269
데 이 터 분 석 기 획 안				
주제	내용적 속성에 따라 분류된 인터넷 기사들을 사용자에게 맞춤 제공하는 데이터 분석 알고리즘			
활용 목적	<p>[정보 문해력 (Data Literacy) 향상 및 가이드라인 제공]</p> <ol style="list-style-type: none"> 1. 현대 사회로 오면서 정보 문해력의 필요성은 더욱 강조되었다. 첫째, 정보에 대한 접근성이 매우 낮아짐에 따라 정보 문해력의 요소 중 하나인 비판적 사고력 역시 점점 퇴화되었다. 또한, 정보의 양이 급격하게 증가하여 이른바 빅데이터의 시대가 도래함에 따라 현대 시민들은 필연적으로 방대한 양의 데이터를 마주하게 되었다. 이에 따라 정보 처리의 중요성이 입증되었다. 2. 따라서 이러한 사회적 배경을 가진 우리에게 정보 문해력은 꼭 필요한 능력이자 소양이다. 그러나 급격하게 변한 사회에서 우리는 데이터를 통제하지 못하고 정보의 바다에서 표류하곤 한다. 3. 이 알고리즘은 사용자들이 정보 문해력을 연습하고 체험할 수 있는 기회를 제공한다는 점에서 활용 가능성이 높다. 검색 결과 수집된 뉴스를 제목 및 헤드라인을 내용 속성을 기준으로 분류하면, 사용자는 원하는 의도의 기사 검색 결과를 선택할 수 있다. 정보의 의도를 선택하고 인식하는 것만으로도 정보 문해력을 향상시킬 수 있다. <p>[사회적 이슈의 여론 파악 기능]</p> <ol style="list-style-type: none"> 1. 뉴스 소비 편의성을 제공한다. 기존의 뉴스 검색 결과 화면은 일반적으로 관련도 순으로 정렬되어 나타난다. 그러므로 보통 어떤 이슈에 대해 검색하는 경우, 키워드 위주로 검색하는 경향이 있으므로 기존 검색 결과 화면에서 사용자가 원하는 기사를 찾기는 쉽지 않다. 그러나, 원하는 속성의 기사를 검색할 수 있다면, 사용자 입장에서 보다 			

효율적으로 기사를 찾을 수 있을 것이다.

2. 사회적 이슈에 대한 다양한 여론을 파악하여 사회 이해를 증진할 수 있다.

[원하는 속성의 기사만 선택적으로 확인 가능]

1. 개인의 관점에서 예를 들어, 스포브리그에서 응원하는 선수의 긍정적 평가와 부정적 평가를 나눠보고 싶은 경우, 혹은 주식을 투자할 때 투자 종목의 전망 및 가능성을 알아보기 위해 회사의 긍정적 소식만 따로 알아보고 싶은 경우 활용할 수 있다.
2. 영리 단체의 입장에서는, 키워드에 대한 긍/부정 비율을 통해 해당 검색어(주력 상품, 회사 이미지, 마케팅/홍보 효과 등)의 여론을 파악할 수 있다. 예를 들어, A 회사의 검색 결과 속 키워드를 제공하여 홍보/마케팅 방향을 정할 때 용이하다.
3. 더욱 나아가 긍/부정 뿐만 아닌 7개에 달하는 감정으로 분류하는 경우 소비자들의 니즈와 반응을 구체적으로 파악할 수 있다는 장점이 있다.

[언론을 활용한 홍보/마케팅]

1. 예를 들어, 슬로건을 정하거나 새로운 회사 이미지를 만들려고 할 때, 긍정적인 키워드만을 모아 볼 수 있다. 예컨대 보통 가격대에서도 고급스러운 옷과 악세서리를 판매하고 있던 일반 쇼핑몰이 프리미엄 마케팅으로 가격을 올려 상품을 홍보하면 큰 이득을 얻지 못할 수도 있다. 그러나 해당 회사에 대한 긍정적 기사를 모아 보면, 회사의 장점을 살리는 효과적인 마케팅을 진행할 수 있을 것이다. 또한 부정적 기사들로부터 보완해야 할 점에 대한 힌트를 얻을 수 있다.

[언론사 분석]

1. 언론사에 따른 긍/부정 비율을 조사하여 언론사가 지향하는 관점을 엿볼 수 있다. 이 기능을 활용하면, 업무 관점에서 타깃 키워드에 대한 언론사의 반응을 미리 살펴, 광고를 내거나 혹은 엠바고를 거는 등의 언론사와 회사의 컨택이 중요한 일에 연결할 언론사를 선택하는 일에 도움을 받을 수 있다.

<p>분석 대상 (데이터)</p>	<p>[데이터 분석 방법]</p> <ol style="list-style-type: none"> 1. 네이버 뉴스 기사 크롤링 <ol style="list-style-type: none"> A. 사용자가 원하는 기사 제목과 게시 날짜 및 기간을 선택하면, 해당 데이터를 수집. 2. 수집한 데이터를 기사 제목, 게시 날짜, 원본 링크, 헤드라인으로 정리 및 전처리. 3. 전처리한 데이터 파일을 기반으로 AI를 활용하여 기사 제목 및 헤드라인을 내용 속성에 따라 긍정, 부정으로 분류
<p>발전가능성 /전망</p>	<ol style="list-style-type: none"> 1. 이미 여러 다른 필터링 알고리즘이 존재한다. 이 알고리즘이 다른 서비스와 차별점을 가지기 위해서는 emotion 별 필터링의 장점을 극대화해야 한다. 제목 2. 언론사 및 파워블로거 등과 제휴를 맺어 댓글이나 블로그 글 등으로 수집하는 데이터 범위가 확장될 수 있다면, 언론뿐만 아니라 일반인의 여론을 확인할 수 있다. 더불어 일상언어는 인공지능이 주로 학습하는 언어 데이터와 유사하므로 인공지능 엔지니어링 관점에서 서비스의 신뢰도가 향상될 수 있을 것이다. 그러나, 기사뿐만 아니라 개인의 생각도 데이터에 수집되므로 정보적 검색어의 경우, 신뢰도를 잃을 수 있다. 따라서 이를 해결하기 위해서는 private data와 official data를 구분하여 수집한 후, private data에는 언론사처럼 출처를 태그하고 다른 카테고리를 마련하여 사적 관심도 정보를 희망하는 사용자에게만 제공해야 한다. 또한 개인정보가 포함된 데이터를 삭제하는 과정이 필수적이다. 3. 마지막으로, 사용자의 연령에 따른 기사 검색 결과를 선택하게 된다면, 청소년 및 유아에게 유해한 검색결과를 보이지 않게 만들 수 있다. 4. 인공지능을 활용하므로 뉴스 기사를 분류하는 과정을 완전히 신뢰할 수 없을 것이라는 우려가 있을 수 있다. 그러나 기사의 내용 속성을 분류하는 일을 사람이 하게 된다면, 주관적 평가가 개입되므로 오히려 인공지능의 자동 분류 시스템이 획일적이며, 객관적일 수 있다. 그러나 분류 오류를 수정하기 위해서는 꾸준한 데이터 전처리와 알고리즘 수정/업데이트가 필요할 것이다.

참고문헌

Narr, B. (2022, August 8). 8 Simple Ways to Enhance Your Data Literacy Skills. Forbes.
<https://www.forbes.com/sites/bernardmarr/2022/08/08/8-simple-ways-to-enhance-your-data-literacy-skills/?sh=224f3af32eb6> (2023-07-26)

Luci Pangrazio and Julian Sefton Green. (2019). The social utility of 'data literacy'. Learning, Media, and Technology.

개인정보의 수집 및 이용 동의

학술문화처 도서관은 <데이터분석 경진대회> 신청 접수를 위하여 다음과 같이 귀하의 개인정보를 수집 및 이용하고자 합니다.

개인정보 수집 · 이용 동의

1. 개인정보 수집·이용 목적: 데이터분석 경진대회 신청 접수
2. 개인정보 수집 항목: 필수 - 학과(소속), 성명, 학번, 핸드폰번호
3. 개인정보 보유기간: 대회 종료 후 1년
4. 동의하지 않을 권리 및 그에 따른 불이익

귀하는 개인정보의 수집 및 이용에 동의하지 않으실 수 있으나, 필수정보 수집 동의 거부시 경진대회 참가 신청이 불가합니다.

☒ 동의함

☐ 동의하지않음

위 기재된 내용이 사실과 일치함을 확인하며 경진대회 참가를 신청하고자 합니다.

2023년 7월 28일

신청자 : 현시은 (서명)

신청자 : 강우진 (서명)

신청자 : 이수향 (서명)

Jonathan
Seh