

빅데이터 기술 교육 세미나 빅데이터와 여론조사

목표 5

빅데이터와 AI 기반 여론 분석 연구와 교육 사례:
썸트렌드 vs. 코딩

이경전 (경희대 교수), 박아름 (용인예술과학대 교수)

빅데이터와 AI 기반 여론·정책 분석 연구와 교육 사례: 썸트렌드 vs 코딩

2022.5.19

이경전*, 박아름**

*경희대학교 경영대학 & 빅데이터응용학과 교수

**용인예술과학대학교 빅데이터경영과 교수



새로운 미디어의 등장과 정책: 빅데이터·AI 활용 분야 및 활용성 확대

- 2002년 12월 대선
 - 오마이뉴스의 등장(vs. 종이 신문)
- 2004년 12월 다음 아고라 서비스 시작
- 2007년 12월 대선
- 2011년 10월 서울시장 보궐선거
 - 2011년 4월 팟캐스트 '나는 꼼수다' 첫방송
- 2011년 12월 종편방송 개국 (vs. 공중파TV)
- 2012년 12월 대선
 - 국정원 댓글 부대 사건, '나는 꼼수다' 종영 (12.18)
- 2016년 가을 종편 방송 '최순실 사건' 적극 보도
- 2016년 9월
 - TBS 김여준의 뉴스 '공장' 첫 방송
- 2017년 5월 대선
 - 드루킹 사건
 - 여론조사 발표의 영향력 및 오남용(예: 역선택)
- 2017년 8월 청와대 국민청원 신설
- 2018년 12월 아고라 폐지

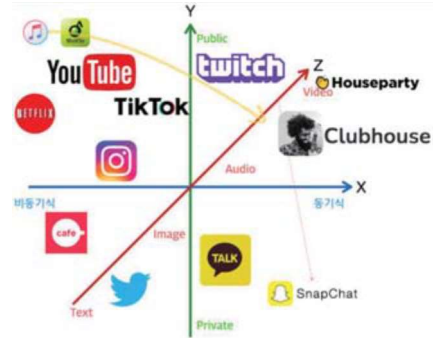
- 2019년 10월 네이버 악성 댓글 차단 클린봇 서비스 시작
- 2020년 4월 총선
 - 빅데이터 선거: 민주연구원
 - 유튜브 선거
- 2020년 5월: 415총선 부정선거 논란 유튜브 미디어에 의해 촉발
- 2020년 6월: 빅데이터·AI 전문가 여의도 연구원장 영입 추진
- 2020년 7월 & 8월: 포털3사 연예 & 스포츠 댓글 폐지
- 2020년 12월, 2021년1월: 미국 대선 부정선거 논란 및 의사당 점거 폭력 사태
- 2021년 1월, 2월: 클럽하우스 열풍
- 2021년 4월: 서울, 부산 시장 보궐선거
- 2022년 3월 대선
 - AI 선거: 크라켄, AI 윤석열, "매타"버스
 - 뽐뿌, 에팸코리아 등 자체 커뮤니티의 영향력
- 2022년 6월 지선: ???
- 2024년 총선: ????
- 2027년 대선: ????

미디어 현황과 빅데이터·AI 활용 정책 연구의 필요성

- 여론이 소셜미디어와 댓글로 표현되어 정책수요/결정자에 영향 미치며, 데이터로 축적, 분석 가능해져, AI/빅데이터 활용 분석 능력이 정책 능력에 영향을 미치는 상황.
- 온라인 국민청원, 인플루언서의 콘텐츠 기사화 등 다양한 의제 설정 채널 등장
- 세대간 미디어 사용의 분절로 인식의 격차가 커지고 있고, 소셜 미디어 알고리즘이 사용자의 정보 편식, 쓸림 현상을 초래하고 있으며, 카톡방, 텔레그램방 등을 활용한 폐쇄적 정보 공유로 탈진실의 증폭이 이루어지면서, 정치 성향의 양극화, 정책의 비합리성, 포퓰리즘의 위험이 커져 합리적인 정책합의가 더욱 어려워질 위험이 커지고 있음
- 신속 정확한 여론 분석 능력을 갖추고, 새로운 의제 설정 기회를 활용하며, 세대간, 집단간 분절, 쓸림, 양극화 현상에 대한 대안을 제시하는 정책 연구 방법론 필요



넷플릭스 최대 경쟁자는 여기서, Z세대가 열광하는 '메타버스', 중앙일보 2021.2.18, mnews.joins.com/amparticle/23994539



<https://brunch.co.kr/@ioojoo/99>

빅데이터·AI기반 정책 연구

- 의제 설정, 정책 결정, 집행, 평가 등 전과정에 빅데이터·AI방법론을 활용
 - 국민의 의견을 신속, 정확히 반영하여 정책을 수립하는 과학적 방법론으로서의 AI
 - 정책 현안과 수요 발굴 및 정책 반응 탐지를 감이나, 여론조사, 전문가에 의존했다면, 비대면 미디어와 AI·빅데이터 등의 도구 활용하여 정책 현안 발굴하고 및 정책 수요자와의 커뮤니케이션을 통해서 수요자의 반응을 고려하여 정책을 튜닝, 미세조정 하여, 정책 수요자에게 적합하고 실질적 도움되는 효과를 기대할 수 있는 정책을 만드는 방법 필요
 - 왜 정책 수요자의 선호도가 바뀌었고, 선호를 바꾸기 위해서 어떤 정책이 필요한가?
 - 정책 성공을 위해 파라미터를 어떻게 조정하여, 맞춤형 정책들을 만들어낼 것인가?
- 참고: 정치나 경영에 AI를 접목한다면?, 전자신문, 2021.2.9. etnews.com/20210209000191
 - “이해관계 대립에 따른 사회적 비용이나 갈등이 증가되는데 여전히 위원회, 청와대 국민청원 등의 불완전한 방법에 의존한다”
 - 사회문제에 대해 미디어를 통해 표출되는 집단 간 의견 수집해 합의 가능 지점 찾는 AI
 - 여론조사, 온라인 청원 등 갈등과 이견 해소 위한 기존 노력 넘어 다양한 의견 간 차별성과 유사성 분석하고, 상호 동의 가능 부분 포착해 집단 간 토론과 의견 공유 지원

빅데이터 · AI 활용 정책 수립 시스템 연구 주제들

- 키워드 분석을 통한 정책 주목도 분석
- 특정 정책 관련 뉴스/댓글 감성 분석을 통한 지지도 분석
- 정책수요자 샘플링과 AI활용을 통한 효율적인 여론 조사
- 댓글러/정책수요자 선호/성향 변화 추이/이유 분석: 사건/사고, 정책, 정치인
- 정책 반응 분석 (뉴스 댓글 반응) 및 피드백을 통한 정책 변화 및 파라미터 조정 연구
- 정책수요자 클러스터링을 통한 맞춤형 정책 설계 및 제시
- 정책 관련 키워드에 대한 빅데이터 분석을 통한 정책 반응 탐지 및 정책 보완 아이디어 도출 방법론
- 특정 정책에 대해서는 도메인 지식에 기반한 모델 및 시뮬레이션, 사례 연구에 의한 정책 대응 방법론
- 정책 관련 미디어 전략의 효과성 평가
- 정책수요자 성향 고려한 감성 분석 고도화: 크롤링, 키워드/유사어 추출, 감성분석, 요약 모델 개발
- 크라우드 소싱(Crowdsourcing)형태의 협업 및 동시 다발적인 데이터 수집과 분석 및 정책 피드백 시스템 구축

빅데이터· AI 기반 여론 조사 연구 사례 및 활용

(사례)

- 특정 키워드 빈도 분석을 통한 주목도 분석
- 특정 정책 관련 뉴스와 댓글의 감성 분석을 통한 지지도 분석 (AI에 의한 감성 분석 필요)
- 특정 정책에 관련한 긍정적 뉴스, 부정적 뉴스에 대한 통계(AI에 의한 감성 분석 필요)
 - 특정 정책에 대한 긍정적 뉴스, 부정적 뉴스의 수
 - 특정 정책에 대한 긍정적 뉴스, 부정적 뉴스에 대한 Like/Dislike 수
 - 특정 정책에 대한 긍정적 댓글 수와 이 댓글에 대한 Like 수, dislike 수
 - 특정 정책에 대한 부정적 댓글 수와 이 댓글에 대한 Like 수, dislike 수

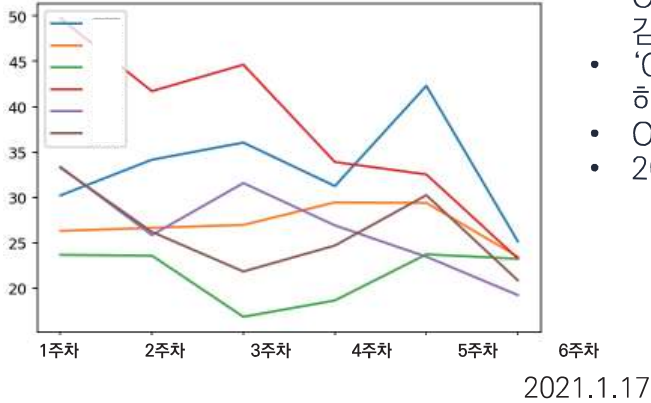
(활용)

- 내부 정보로만 사용
- AI분석 결과 공표에 의한 Signaling Effect에 유의해야 함
 - 왜곡의 유혹, AI 모델 성능의 문제, 'AI'에 대한 대중의 과신
- 정책 커뮤니티에 공유
- 민간 서비스 및 불법적 사용자, 사용자 그룹의 감시
 - 실행 사례: 크라켄

댓글 감성 분석

연구 사례 1: K-정책플랫폼(K-Pol.org) 연구 사례

- 긍정 댓글 vs. 부정 댓글 비율 분석
- KoBERT에 Fine Tuning 과정 거쳐 모델 재학습
- 분석단위 진화: 단어 -> 댓글 -> 댓글러
- 단어 단위 감성 분석은 어느 정도 정확하나 댓글 단위의 감성 분석에 대해서는 더 향상될 여지를 확인하였으며, 댓글러 단위가 오히려 의미가 있음도 확인.

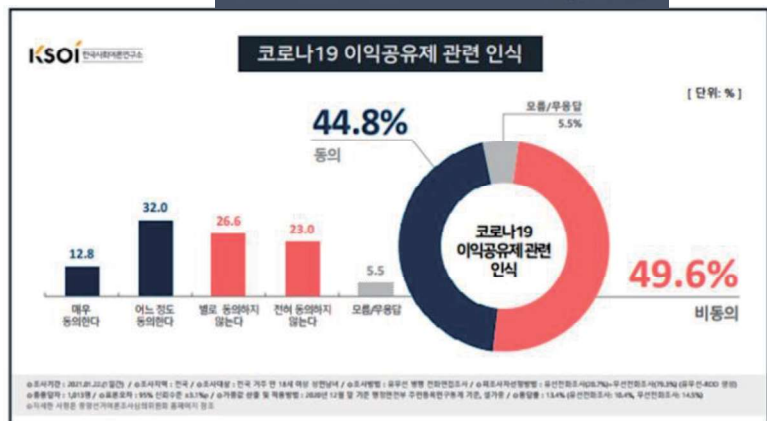
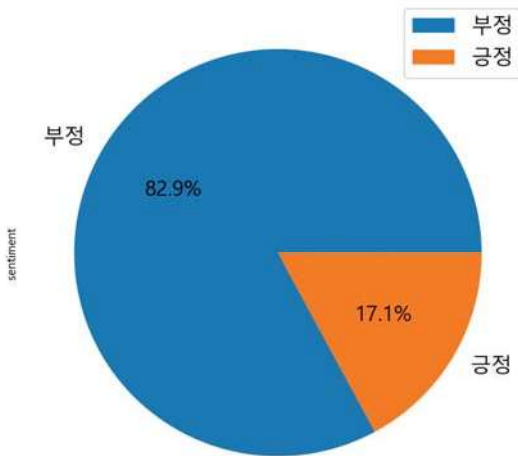


- 감성 분석 결과 전체적으로 각 후보자에 대한 이미지 감소
- 'OOO' 후보자는 긍정 댓글의 비율이 가장 높았지만 점차 감소하여 5주차에 2위로 하락
- 'OOO' 후보자는 긍정 댓글 비율이 3위 였지만 점차 증가하여 5주차 1위 달성
- OOO-OOO 단일화가 필요했던 이유
- 2021.3.23: 야권 단일 후보 선출

정책 여론 분석 사례: 이익 공유제

- 뉴스 댓글 분석 결과 이익 공유제 관련 인식과 한국사회 여론연구소에서 조사한 이익 공유제 관련 인식에서 큰 차이 확인
- 주주대상 전경련 조사와는 상대적으로 차이가 적음

연구 사례 1: K-정책플랫폼(K-Pol.org) 연구 사례





사랑
좋은
기대
좋다
웃음
활약
성공
화려
안벽

범죄
불법
갑질
비판
특혜
이기
논란
부패하다
혐의

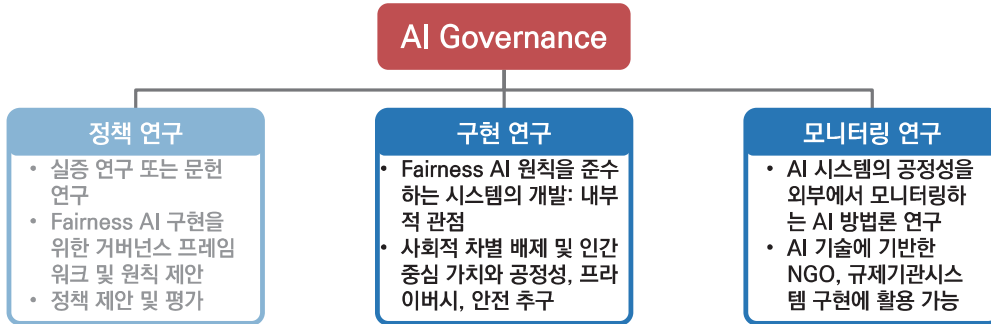
K-POL(K-정책플랫폼) 빅데이터· AI 기반 여론 조사 연구 시사점

- 댓글 감성 분석 고도화 필요 - 이 부분이 정확하기 전에는 이를 자동적인 여론 조사 결과로 공표하는 것은 무리
- 뉴스, 뉴스 댓글, SNS, 기존 여론 조사의 결과가 상충되는 경우 이를 앙상블하고 대시보드화 하는 방법론 필요
- 빅데이터· AI 기반 여론 조사는 자동화하여 발표하는 Signaling Effect를 주기 보다는 내부적 정책 개선 및 전략 수립에 활용하는 것이 바람직
 - 댓글 수준의 감성 분석이 95%~99%의 정확성을 가지게 될 경우는 자동화 발표도 가능할 것임
 - 댓글러의 성향 변화는 댓글 수준의 감성 분석보다 먼저 더 높은 정확성을 가지게 될 것으로 예상되므로, 댓글러 패널 구성에 의한 여론 조사 방법론을 연구 개발하는 것은 의미가 있을 것임
- 다만 기존 여론 조사가 발표되었을 경우 어느 정도 공신력을 갖춘 기구를 통해 빅데이터· AI 기반 여론 조사 결과가 경쟁 및 보완하는 체제로 갈 필요가 있음
- 빅데이터· AI 기반 조사는 기술, 경험, 데이터가 축적되어서 일종의 데이터 효과를 일으키는 분야
 - 데이터 효과 (Data Effect): 제품과 서비스, 비즈니스 모델에 데이터가 쌓임에 따라 경쟁력과 가치가 커지는 효과(coined by 이경전)
 - K-POL(K-정책플랫폼)과 같은 민간 ThinkTank 등 민간과 여의도연구원, 민주연구원 등 정당 연구소에서 빅데이터· AI 기반 정책 연구 및 여론 조사 시스템을 제대로 갖추어 '축적의 시간'을 확보할 필요가 있음.
 - 정책가, 정책, 여론 조사 전문가, AI 빅데이터 전문가의 협업 시너지 확인

연구 사례 2: 인공지능 거버넌스 연구(2020)

네이버 클린봇 평가 및 대안 모델 개발

유지웅, 항보유정, 손동성, 이경전, 악성 댓글 분류 시스템 모니터링 연구: 네이버 클린봇 분석(Malicious Comment Classification System Monitoring Research: Naver Cleanbot Analysis), 2020 한국지능정보시스템학회 춘계학술대회, 2020.
 유지웅, 인공지능 거버넌스 시스템: 모니터링 및 구현 연구: 네이버 클린봇 분석, 경희대학교 빅데이터융합학과 석사학위논문, 2021.
 Yoo, Park, & Lee, Governance of News Comment Filtering AI: Naver Cleanbot Case, Working Paper, 2022.



인공지능을 활용하여 악성 댓글 차단 사례

- 네이버 클린봇: 사용자가 뉴스 기사에 댓글을 작성하면 악성 유무를 판단하여 자동으로 노출 제한
- 뉴욕타임즈 모더레이터
 - 직소의 Perspective API를 활용하여 Moderator 시스템 개발
 - Moderator는 사용자에서 '유해', '스팸', '음란' 세가지 항목에 대해 점수를 할당
 - 뉴욕 타임즈는 Moderator의 점수를 기반으로 최종적으로는 사람이 댓글 노출 여부 판단

네이버 뉴스 댓글 수집 & 레이블링

악성 댓글 분류 모델인 클린봇을 평가하기 위해 네이버 뉴스에서 작성된 댓글을 수집

- 기간: 2019년 11월 12일 ~ 2019년 12월 14일
- 키워드: '문재인', '김건모'
- 수집된 뉴스 건수: 360건
- 수집된 댓글 건수: 91,716건

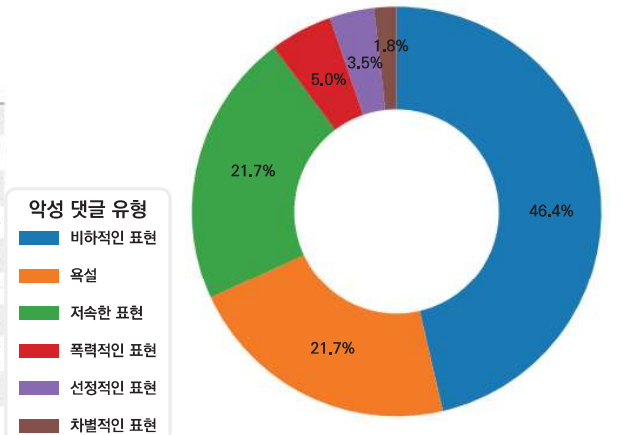
키워드	댓글 개수	차단된 댓글 개수	차단율
문재인	24,726	1,102	4.46%
김건모	66,990	1,809	2.70%
총합	91,716	2,911	3.17%

id	date	keyword	link	nick	type	비고	댓글
3314	2019.12.11.23.55	문재인	https://news.naver.com/main/read.nhn?mode=LSD&...	ljs0****	댓글		자제분 물기가학한다*
11059	2019.12.08.03.24	문재인	https://news.naver.com/main/read.nhn?mode=LSD&...	dcs****	대댓글		국민들은 원자력을 발전시켜 안전한 원전이 국가 산업에 중요한 공간이라 생각합니다 국...
18633	2019.12.06.10.47.45	문재인	https://news.naver.com/main/read.nhn?mode=LSD&...	ye2****	댓글		대한민국의 정의수호와 권력중의 부패, 권력비리적결에 앞선바 임무를 다하고 있는 걸음장...
29905	2019.12.13.15.21.32	김건모	https://news.naver.com/main/read.nhn?mode=LSD&...	ys92****	댓글		이러다 나중엔 다 사실이더라, 얼른 사과하고, 처벌 받아라---
88945	2019.12.09.18.43	김건모	https://news.naver.com/main/read.nhn?mode=LSD&...	onou****	대댓글		현대부 있는 술집에 가온 모든 낯선분 다 온 강간범이라는 놀리
89185	2019.12.11.10.32	김건모	https://news.naver.com/main/read.nhn?mode=LSD&...	yan9****	댓글		관중 대 관중 관대 통병에 한일씩 간계 성숙할 증거는 아니잖아?
44412	2019.12.11.10.08	김건모	https://news.naver.com/main/read.nhn?mode=LSD&...	wan9****	댓글		지금 악동자 잘 생각하십시오. 좋을때는 글방 지나가니...
22686	2019.12.06.09.58	문재인	https://news.naver.com/main/read.nhn?mode=LSD&...	algy****	댓글	클린봇	*거지 쓰레기같은 문재인 청와대는 거짓말을 밥먹듯이 거짓말로 포장하여 전설처럼 이야...
12018	2019.12.07.03.49	문재인	https://news.naver.com/main/read.nhn?mode=LSD&...	lnde****	댓글	클린봇	오후 문재인 똥 1선놈 이빨을 핏물었다. 모친 권리디언들이 전파 실드중 처벌, 이정도...
38451	2019.12.13.11.52	김건모	https://news.naver.com/main/read.nhn?mode=LSD&...	slar****	대댓글	클린봇	ㅈ.. 총합커가 가해자면이라는 ㅈ개소리위나침을관한봐도 A..

연구 사례 2: 인공지능 거버넌스 연구(2020)

수집된 데이터를 사람의 눈으로 확인하여 악성 댓글 유무를 판단 또한 네이버에서 정의한 6가지의 악성 댓글 유형을 기반으로 악성 댓글 유형 분류

1. 욕설: 일반적인 욕설, 모욕적인 표현 또는 남을 저주하는 표현
2. 저속한 표현: 타인에게 불쾌감을 주는 속되고, 격이 낮은 표현
3. 선정적 표현: 성적으로 자극적인 표현
4. 폭력적인 표현: 신체적 위협에 대한 표현
5. 차별적인 표현: 지역/인종/국가/종교/ 등에 기반한 차별 표현
6. 비하적인 표현: 상대방에게 모멸감과 수치심을 주는 비하 표현



모델 안정성 평가

연구 사례 2: 인공지능 거버넌스 연구(2020)

네이버의 클린봇이 여러 상황에서 악성 댓글을 안정적으로 분류하는지 모델의 품질 평가

- 중요한 단어나 키워드를 직관적으로 시각화하는 도구인 WordCloud를 사용하여 사용자가 주로 사용하는 욕설을 확인
- “욕설”로 분류된 댓글 데이터의 WordCloud로 사용자는 주로 ‘미친’, ‘지랄’, ‘ㅈㅂ’, ‘ㅂㅈ’의 욕을 사용하는 것으로 확인
- 모델 평가를 위해 ‘지랄’이라는 단어를 얼마나 안정적으로 분류하는지 확인
- 전체 댓글 중 ‘지랄’이 포함된 댓글은 총 527건, 그 중 클린봇에 의해 차단된 댓글은 67건으로 12.71%의 차단율을 확인
- 같은 댓글이지만 서로 다르게 판별한 경우
- 같은 작성자, 유사한 뉴스 기사, 비슷한 작성 시간 등 입력값이 유사한 상황에서 서로 다른 결과를 출력

id	date	keyword	link	nick	type	비고	댓글
1141	2019.12.12. 09:35	문제인	https://news.naver.com/main/read.nhn?mode=LSD&...	acez****	댓글	클린봇	손지는 왜도 지랄 인해도 지랄할 모양이구나?
1980	2019.12.10. 10:16	문제인	https://news.naver.com/main/read.nhn?mode=LSD&...	ghw****	대댓글	클린봇	가도 지랄 안가도 지랄
4189	2019.12.11. 10:27	문제인	https://news.naver.com/main/read.nhn?mode=LSD&...	wse****	댓글	클린봇	지랄보네 ㅋㅋㅋ 잘하길 바라겟지 ㅋㅋㅋㅋ
43570	2019.12.12. 22:23	김건모	https://news.naver.com/main/read.nhn?mode=LSD&...	yoog****	댓글	클린봇	저런 놀다가 왜 지랄하고있네
84686	2019.12.09. 13:00:22	김건모	https://news.naver.com/main/read.nhn?mode=LSD&...	hyun****	댓글	클린봇	이년 운동장기러다 말대로 안되니까 저지랄하는군ㅋㅋ

“지랄”이 포함된 차단 댓글

id	date	keyword	link	nick	type	비고	댓글
20332	2019.11.28. 15:29	문제인	https://news.naver.com/main/read.nhn?mode=LSD&...	fran****	댓글		지랄도 풀넌이다 진짜
4005	2019.12.11. 19:50	문제인	https://news.naver.com/main/read.nhn?mode=LSD&...	chi****	대댓글		좌를 새끼를 꼭 엑트로 조치면 저 지랄을이저
9857	2019.12.07. 20:14:39	문제인	https://news.naver.com/main/read.nhn?mode=LSD&...	log****	대댓글		문재인 대통령님 지랄합니다. 한반도 경제파탄에 예비주셔서 잘 사합니다
61935	2019.12.11. 06:34:27	김건모	https://news.naver.com/main/read.nhn?mode=LSD&...	seoy****	대댓글		지랄 언주말문거지 사 프 오
76139	2019.12.09. 15:12	김건모	https://news.naver.com/main/read.nhn?mode=LSD&...	ch25****	대댓글		참 강릉서 또 지랄한다-

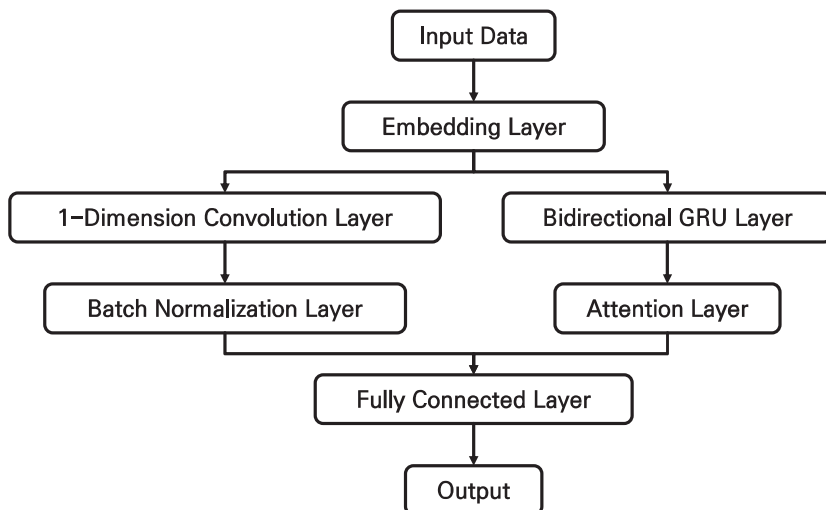
“지랄”이 포함된 비차단 댓글

	클린봇 적용 전	클린봇 적용 후
연합 뉴스	Imia**** 2019.12.09. 15:03 썬레기	Imia**** 2019.12.09. 15:03 · 신고 썬레기
서울 신문	Imia**** 2019.12.10. 18:50 썬레기	Imia**** 2019.12.10. 18:50 ① 클린봇이 부적절한 표현!

대안 모델 구현: CNN, GRU, Attention을 조합

연구 사례 2: 인공지능 거버넌스 연구(2020)

- 빠르고 불규칙적으로 변하는 욕설 및 비속어 댓글은 추상화 과정을 통해 비 선형적인 특징을 찾는 CNN 기법이 적합
- 욕설은 없지만 문맥을 파악하여 악성 여부를 판단해야 할 경우엔 GRU + Attention 기법이 적합
- Attention 메커니즘은 전체 문장을 동일한 비율로 참고하지 않고 분류 문제에 연관 있는 단어에 더 집중하여 학습
- 성능 비교를 위해 기존의 클린봇과 그 외 세가지의 모델을 설계하여 모델의 정확도 측정
- 4개의 모델의 정확도 및 악성 댓글 유형별 평균 차단율을 비교한 결과 CNN + GRU + Attention 모델이 가장 뛰어난 성능으로 확인

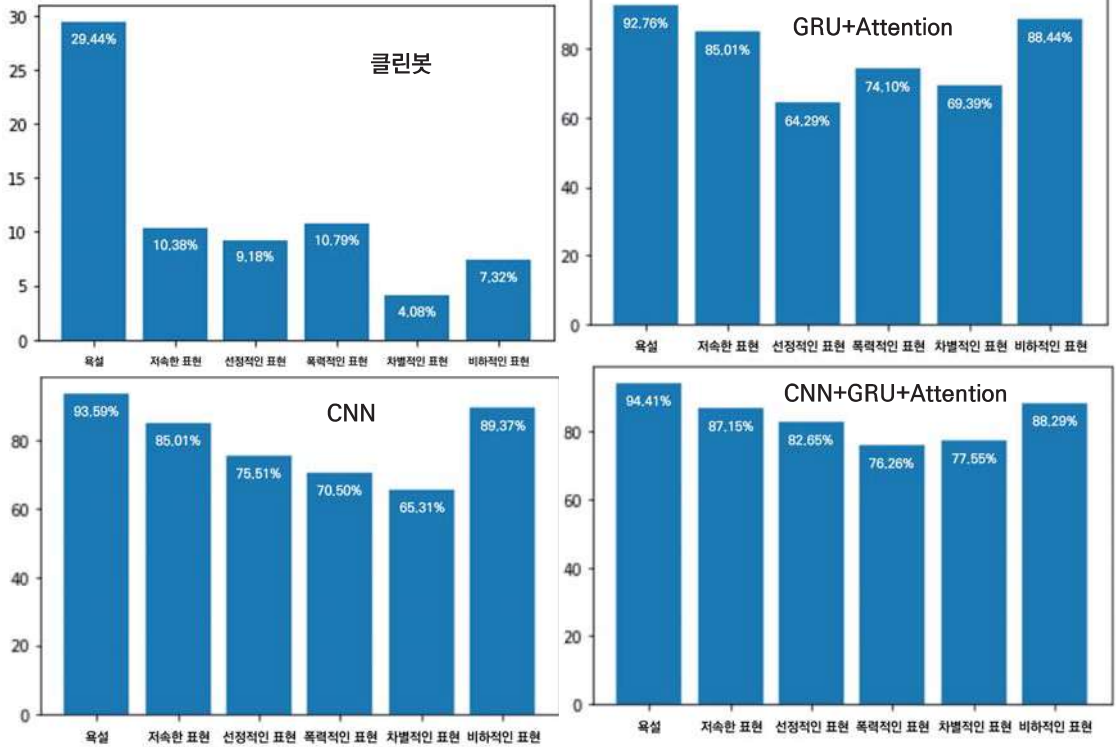


모델	정확도	악성 댓글 유형별 평균 차단율
클린봇	55.87%	11.87%
CNN	87.01%	79.88%
GRU + Attention	87.12%	79.00%
CNN + GRU + Attention	88.21%	84.39%

4개 모델의 악성 댓글 유형별 차단율

악성 댓글을 6가지 항목에 따라 분류한 정보를 기반으로 각 모델의 유형별 차단율 확인

- CNN 모델은 ‘욕설’, ‘비하적인 표현’, ‘선정적인 표현’에서는 높은 성능
- 문맥을 파악해야 하는 ‘폭력적인 표현’, ‘차별적인 표현’에서는 GRU+Attention 우수



연구 사례 2: 인공지능 거버넌스 연구(2020)

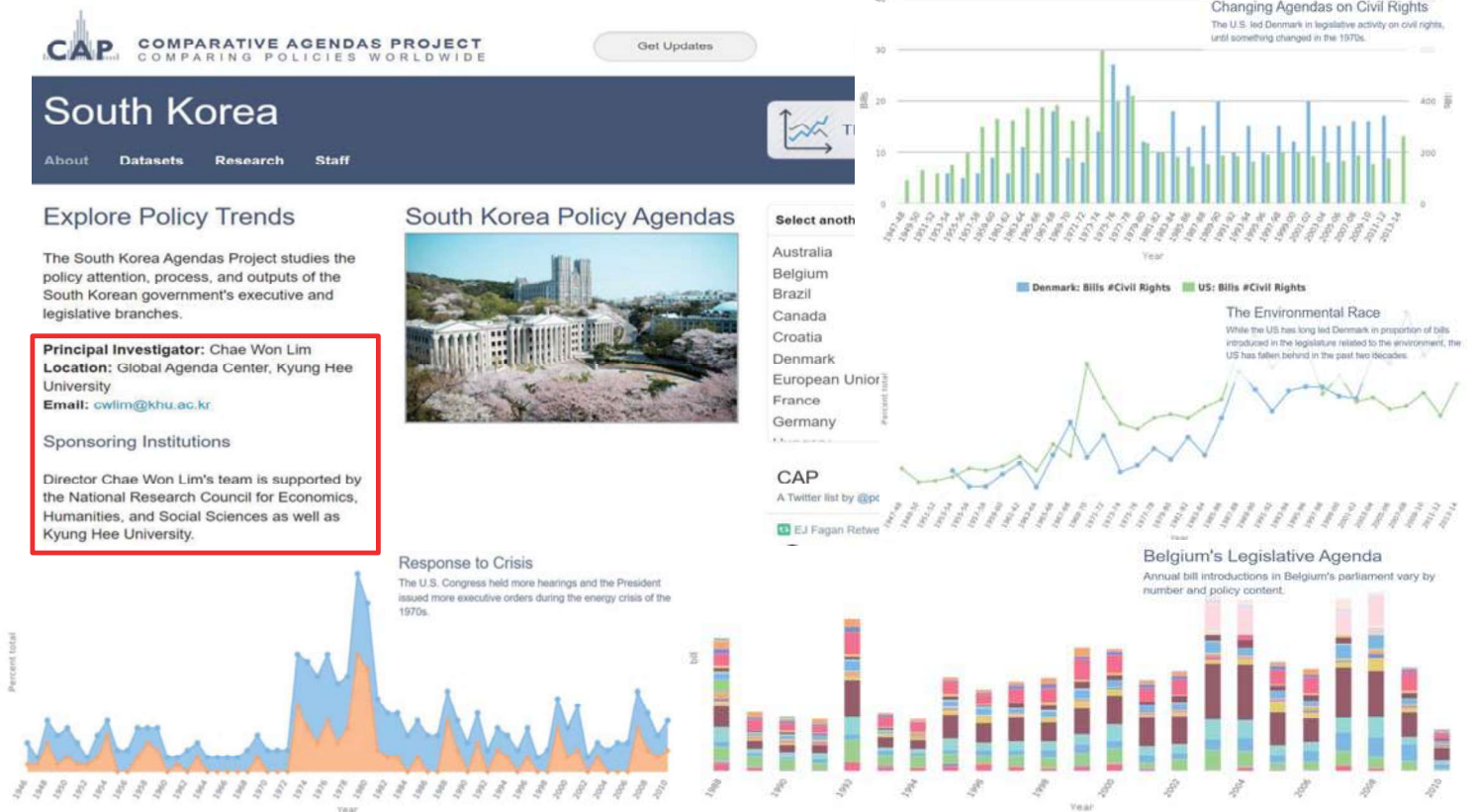
네이버 클린봇 평가 및 대안 모델 연구 시사점 및 결론

- 미디어(임원/실무진, 비밀창구) 알고리즘의 불공정한 운용에 대한 사회적 의심 검증
 - 네이버 '추미애' 검색 오류 사건
- 미디어 알고리즘의 오류로 발생하는 문제의 해결
- 미디어 사용자 관점의 오남용 및 불법적 사용의 감시
 - 예) 김경수/드루킹
 - 자동화된 댓글은 적법한가? K대 B교수 vs. 영업 방해, 선거법 위반
 - 자동화된 댓글에 대한 감시 필요
- 포털 및 미디어의 자정 노력은 한계가 있고, 자칫 여론 공론장의 폐쇄와 규제로 이어질 수 있음(예: 스포츠, 연예 댓글)
 - 빅데이터와 AI에 기반한 민간의 노력이 표현의 자유와 건전한 공론장이라는 두마리 토끼를 잡을 수 있음

불공정한 운용, 오류의 발견 및 시정, 정치권/사용자의 오남용 및 불법적 사용의 감시와 대응을 위한 시 기반 자동 모니터링 시스템과 이른바 AI NGO의 역할 필요

연구 사례 3: 정책 분류 자동화를 위한 AI 모형 개발에 관한 연구

- 비교 아젠다 프로젝트(Comparative Agendas Project, CAP) – 에딘버러대학 주도
 - 정책변동 및 정책 아젠다에 대한 글로벌 연구로서 비교 아젠다 프로젝트가 진행되어 왔음
 - 정책 공통기준으로 분류, 공공의제 변동 특성, 글로벌 공공정책 비교를 통해 협력체계 구축
 - 참여국은 거시 경제, 교육, 보건 등 21의제로 범주화(213 Subtopics): 질적, 계량분석 활용
 - 구축된 데이터 통해 정책변동 특성과 미래 글로벌 의제의 등장 예상하고 정책적 시사점 도출
 - 미, 영 등 국가는 1900년부터 데이터 수집: 정책 전문가들이 모두 수작업으로 정책을 분류
 - 한국: 1987-2019 미디어, 입법, 행정부 의제 분석과 구축을 정책 분석가의 수작업으로 진행
- 수작업 정책 분류는 정책 전문가에 따라 다르게 해석될 수 있으며, 상황에 따라 휴먼에러 발생 할 수 있기에 정책 분류작업의 일관적 분류 및 생산성 높이는 AI 알고리즘 필요
- 수작업 구축 DB 기반으로 AI 알고리즘 활용하여 정책 분류 자동화 모델 세계 최초 개발
- 정책 분류 자동화를 위한 인공지능 모형을 개발하여 알고리즘을 구현한 Policy Perceptron 을 제안: 정책 분류의 첫 알고리즘으로서의 의미
- 이경전, 황보유정, 정백, 유지웅, 배성원, 임채원, *Policy Perceptron: 정책 분류 자동화를 위한 인공지능 모형 개발, 2022 한국행정학회 춘계학술대회, 여수.*



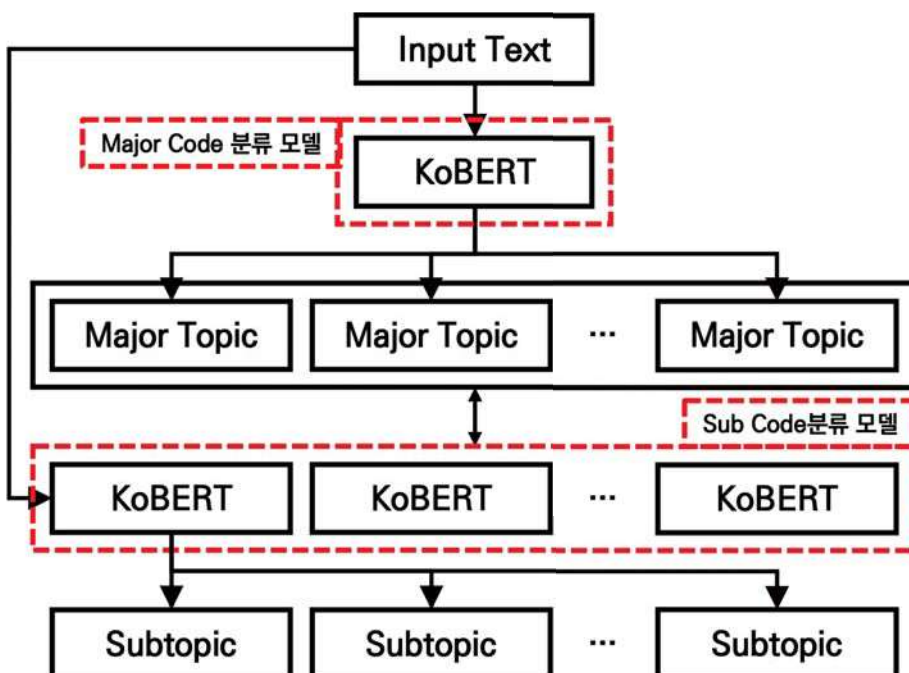
한국판 코딩북 예시: 한국판 정책범주 및 하위범주 코드

Major Topic		Subtopic
1	거시 경제	100. 국내 거시 경제 관련 정책
		101. 인플레이션, 가격, 이자율
		103. 실업률
		104. 통화정책, 한국은행
		105. 국가예산 및 국가부채
		107. 조세, 조세정책, 조세개혁
		108. 산업정책
		110. 물가관리 및 안정
		119. 그 외
		2
201. 사회적 소수자에 대한 차별		
202. 성차별, 성차에 따른 차별		
204. 연령 차별		
205. 장애인 및 특정 질병 보유자에 대한 차별		
206. 선거권		
207. 표현, 종교, 결사의 자유		
208. 사생활에 대한 권리 및 정부 정보에 대한 접근		
209. 반정부활동		
210. 종교단체에 대한 규제		
230. 이민 및 난민		
299. 그 외		

Major Topic	Subtopic	
3	보건	300. 보건 일반
		301. 보건 의료 체계 개혁
		302. 국민 건강 보험
		321. 제약, 의료 기기, 의원에 대한 규제
		322. 의료 관련 시설의 건설, 규제 및 보조
		323. 의료인 및 보험 단체
		324. 의료 사고, 의료인의 일탈 행위 및 사기 행위
		325. 의료인의 육성 및 관리
		331. 질병의 예방 및 관리와 건강 검진
		332. 영유아
4	농업	400. 농업 일반
		401. 농수산물 교역
		402. 농가 및 어가에 대한 보조금 지급 및 재해 피해 대책
		403. 농수산물 안전 및 검역
		404. 농수산물 판촉
		405. 가축 및 농작물의 질병 및 해충 관리
		406. 가축의 건강과 돌봄
		407. 농수산물 생산과 관련된 환경 문제
		408. 어업
		498. 농업 관련 연구 개발(R&D)

연구 사례 3: 정책 분류 자동화 AI

Policy Perceptron 구조도



한국 정책 아젠다 데이터베이스 구축

- 한국 비교아젠다연구팀은 2018년도 부터 비교 아젠다 프로젝트에 참여 23 Major Topic 및 213개의 Subtopic 지정하여 각 기준에 기반한 입법, 행정, 미디어에 대한 데이터베이스 구축
- **행정 데이터:** 1988-2018년 대통령 연설문을 문장 단위로 분해하여 구축. 행정 데이터는 대통령의 연설문을 문장 단위로 쪼개어 각 Major Topic 및 Subtopic의 코드로 레이블링 하였으며 만약 해당 문장이 둘 이상의 정책 내용을 포함하면 추가 분해 작업을 통해 세분화하였다. 또한 문장이 정책 내용을 포함하지 않을 경우 Major Topic 및 Subtopic을 9999(그 외)로 레이블링.
- **입법 데이터:** 한국의 주요 통과 법안에 대한 내용을 수집한 데이터베이스로 1987년부터 2018년 까지의 데이터가 수집. 각 법안의 발의 날짜와 발의한 상임위원회, 법안명 등의 데이터 수집.
- **미디어 데이터:** 미디어 데이터는 각 시기에 미디어의 제 1면에 기재된 뉴스 기사들을 수집한 데이터베이스. 조선일보 신문사의 기사를 수집하였으며 1988년부터 2020년까지의 기사 데이터베이스를 구축. 해당 기사의 발행 날짜, 기사 제목, 기사 요약 문장 수집.
- **최종 데이터베이스:** 데이터의 수가 10개 미만인 Subtopic은 제외 => Major Topic은 23개, Subtopic은 184개로 확정.

데이터 추가 수집, 전처리 & 모델 설계

- 입법 데이터: 국회 의안정보시스템 홈페이지에서 웹 크롤링 과정을 통해 해당 법률에 대한 세부 내용을 수집. 불필요한 텍스트 또는 특수문자 제거 및 한자->한글 변환.
- 미디어 데이터: 조선일보 신문 제 1면에 보도된 뉴스 기사를 기반으로 구축. 조선 뉴스 라이브러리 100 홈페이지에서 웹 크롤링 과정을 통해 뉴스의 본문 내용을 수집. 한자 -> 한글 변환.
- 데이터가 상대적으로 긴 입법과 미디어 데이터를 각기 다른 방식으로 데이터 압축
 - TF-IDF(Term Frequency-Inverse Document Frequency)를 데이터 전부에 적용 불용어 제거.
 - 카카오 브레인 Pororo(Platform of Neural Models for Natural Language Processing)를 입법/미디어 데이터에 적용, 각기 다른 두 개 데이터를 재생산하여 공통된 딥러닝 모델에 적용 성능 비교 통해 최종 기법 선택.
- Major Topic 및 Subtopic 코드는 각각 23개, 184개: 184개 중 하나 맞추는 문제
- 2중 분류 모델 설계: 먼저 Major Topic을 분류하고 다음으로 Subtopic을 분류
- 1차 딥러닝 모델은 전체 데이터에 Major Topic을 레이블링으로 학습하고 2차 딥러닝 모델은 전체 데이터를 Major Topic에 따라 분류하여 각각 학습
- Major Topic 분류 모델 1 & 각 Major Topic의 Subtopic 분류 모델 23: 총 24개 딥러닝 모델
- 정답 범주를 예측할 경우 하나의 범주를 예측하여 맞추는 경우(Top-1)와 3개의 범주를 예측하여 정답을 맞추는 경우(Top-3)를 비교하여 성능을 확인.

Policy Perceptron 모형 성능

연구 사례 3: 정책 분류 자동화 AI

- Major Topic 분류 모형
 - 문장 입력 시 Major Topic 23개의 범주를 분류하는 모형
 - Top-1일 경우 23개 중 1개를 예측하여 정확도를 확인한 결과 69.2%
 - Top-3에서는 23개 중 3개를 예측하였을때 87.6%의 정확도
- Subtopic 분류 모형
 - Subtopic 분류 모형의 성능을 나타낸 것으로, 각 Major Topic에 해당하는 모형의 성능과 평균을 Top-1과 Top-3의 정확도를 확인
 - 정확도는 각 23개 Subtopic 모델의 정확도를 평균낸 값으로 Top-1의 경우 74.2%, Top-3의 경우 91.4%의 정확도
- Policy Perceptron 최종 분류 모형
 - Major Topic 분류 모형과 Subtopic 분류 모형을 결합한 Policy Perceptron 모형의 최종 모형 성능
 - Top-3의 경우 Major Topic 분류 모형에서 도출되는 3개의 결과 확률과 이에 해당하는 Subtopic 분류 모형의 9개 도출 결과 확률을 곱하여 가장 높은 확률을 가지는 3개의 결과를 토대로 정확도를 출력
 - Top-1의 경우 정확도가 62.4%로 나타났으며, Top-3의 경우 71.6%의 정확도

연구 사례 3: 정책 분류 자동화 AI

정책적 활용 내용 & 기대효과

- 비교 아젠다 정책 분류를 인공지능 모형 Policy Perceptron이 정책 전문가를 보조하게 되는 경우, 정책 전문가의 생산성이 향상되어 다양한 영역에서의 정책 연구를 진행 가능
- 중앙 정부 정책 외에도 지방자치 정책 등 많은 수의 정책을 분류하고 비교·평가함으로써 국정운영의 방향을 모색하는데 기여.
- 정책 분류 자동화의 인공지능 모형 개발은 동일한 기준으로 정책을 분류 가능
- Policy Perceptron 모델은 향후 사람의 정책 분류 의사결정을 돕는 DSS로 활용 가능
 - AI와 빅데이터 분석, 인간의 시너지 창출: AI는 자동 분류하고 빅데이터 기법은 시사점 도출 및 시각화 - 인간은 해석 및 AI의 오류 검증
- 향후 연구: Policy Perceptron과 사람이 함께 최적의 선택을 할 수 있도록, 사람과 기계의 지속적 상호 작용을 다룬 휴먼인더루프(Human In The Loop) 시스템을 구현하여 Policy Perceptron의 잘못된 판단을 수정하여 점점 더 정확도를 높이는 시스템을 구현
- 본 연구에서 세계 최초로 개발된 정책 분류 자동화 인공지능 모형은 한국어를 사용하여 개발하였지만, 향후 글로벌 정책을 비교를 위하여 영어 등 다른 언어로 개발하게 된다면 글로벌 협력체계를 구축하는 비교 아젠다 프로젝트에 큰 기여를 하게 될 것으로 기대
- 정책 전문가와 AI 전문팀의 원활한 협조로 신속한 Pilot개발
 - 데이터 검증 및 모델 검증

연구 사례 4: 썬트렌드 비즈를 활용한 빅데이터 활용 교육 사례

발표: 박아름 교수 (용인예술과학대학교 빅데이터경영과)



빅데이터경영과의 커리큘럼 무엇이 문제인가

K대, D대, Y대, 한국데이터산업진흥원 빅데이터인재양성 프로그램 등 학부생들을 대상으로 코딩교육을 한 결과, 학생들은 어렵고, 재미없으나 해야만 하는 과목정도로 인식

The Jobs Landscape in 2022

emerging roles, global change by 2022

133 Million

Top 10 Emerging

1. Data Analysts and Scientists
2. AI and Machine Learning Specialists
3. General and Operations Managers
4. Software and Applications Developers and Analysts
5. Sales and Marketing Professionals
6. Big Data Specialists
7. Digital Transformation Specialists
8. New Technology Specialists
9. Organisational Development Specialists
10. Information Technology Services

declining roles, global change by 2022

75 Million

Top 10 Declining

1. Data Entry Clerks
2. Accounting, Bookkeeping and Payroll Clerks
3. Administrative and Executive Secretaries
4. Assembly and Factory Workers
5. Client Information and Customer Service Workers
6. Business Services and Administration Managers
7. Accountants and Auditors
8. Material-Recording and Stock-Keeping Clerks
9. General and Operations Managers
10. Postal Service Clerks

Source: Future of Jobs Report 2018, World Economic Forum



경영직군을 위한 빅데이터분석 교육커리큘럼

<K대 커리큘럼>

<국내외 데이터과학 커리큘럼>

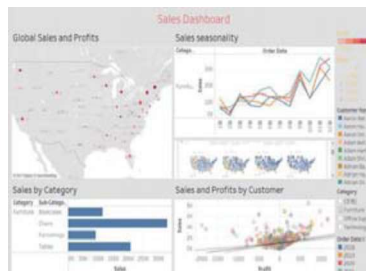
Phase	데이터베이스 입문	데이터처리운영	운영체제 Internal
Phase01 빅데이터 분석을 위한 기초학습	<ul style="list-style-type: none"> Big DATA 이해 DB 운영 관리 	<ul style="list-style-type: none"> DB성능관리 SQL tuning 고 가용성을 위한 서버 구축 	<ul style="list-style-type: none"> Unix / Linux Storage 관리 기법 OS shell programming
Phase02 빅데이터 분석도구 학습	R빅데이터 기본분석 <ul style="list-style-type: none"> R Studio 기본실습 R 과 선형대수 데이터 시각화 및 탐색분석 기술통계 및 통계분석 	파이썬 기본분석 <ul style="list-style-type: none"> 파이썬 개발환경 구축 및 기본분석 Numpy / pandas 학습 데이터 수집 및 가공 시각화/시계열분석 	통계분석 실습 <ul style="list-style-type: none"> 공공데이터 활용한 실습 빅데이터 분석 결과 시각화 pandas + matplotlib 활용 SciPy
Phase03 머신러닝+딥러닝 인공지능강령	통계기반 심화학습 <ul style="list-style-type: none"> 지도학습(삼진, 분류분석) 비지도학습(군집분석, 연관분석) 시계열 분석(회귀도식화) 기술통계분석 	머신러닝 (Machine learning) <ul style="list-style-type: none"> Clustering (고객 세그먼트 분석, 고객 분류) Association Rule (유통바구니 분석, 상품 추천) Decision tree(결정 트리) 회귀분석 - 연관성 분석 	Deep Learning + 인공지능강령 <ul style="list-style-type: none"> 엠티다부스트 알고리즘 Tensorflow 구현 Tensorflow/REMI NN/DNN CNN - RNN
Phase04 빅데이터 프로젝트	빅데이터 설계 분석 프로젝트 <ul style="list-style-type: none"> Data 분석 Data 분석 설계 데이터마이닝 기법 적용 시각화 구현 	분석 도구 활용 <ul style="list-style-type: none"> R Python 분석 알고리즘 	프로젝트 시연 <ul style="list-style-type: none"> 모바일용 정보자료 및 기술서 작성 발표 및 시연

역량 구분	역량 강화를 위한 표준 커리큘럼	교육 기관별 표준 커리큘럼 커버리지 비율				
		해외 A 대학	해외 B 사실교육	국내 A 대학	국내 B 사실교육	국내 C 사실교육
기반 역량 (Foundation)	산업 별 빅데이터 활용 사례, 빅데이터와 Creative Thinking, 빅데이터 보안 분석, 데이터 과학자의 역할 등	40 %	0 %	25 %	20 %	20 %
기술 역량 (Platform Technique)	하둡 Core 및 Eco System의 이해, HDFS와 MapReduce의 활용, NoSQL(Mongo DB, Cassandra 등)	100 %	16.7 %	50 %	83.5 %	16.7 %
분석 역량 (Analysis Technique)	분석 모형의 이해, R분석 및 Visualization, 상용 Tool 활용법, 데이터 마이닝 프로세스, 텍스트 마이닝, Social Network Analysis 등	100 %	100 %	12 %	20 %	100 %
사업 역량 (Business Analytics)	산업 별(제조, 유통, 통신, 금융, 공공, 소매 등) 핵심 업무의 이해, 산업/업무 별(Risk, Social, CRM 등) Analytics 방법 및 적용 등	62.5 %	37.5 %	13 %	12.5 %	12.5 %

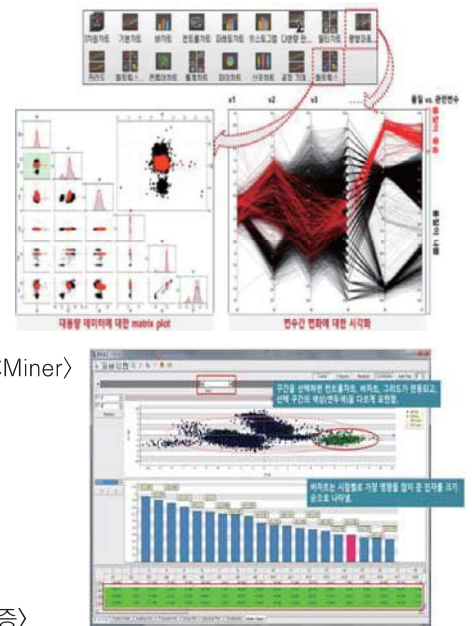
→ 경영직군의 데이터 분석가를 양성하는데 있어, 사업역량과 분석역량의 비중이 더 중요하며, 분석역량을 위해 R, Python 중심의 교육은 학생들이 흥미를 잃을 수 있고 포기할 수 있음

경영직군을 위한 빅데이터분석 교육커리큘럼

후보별 공약 키워드별 감성분석-이준희
(Sometrend)



<ECMiner>



<Tableau와 자격증>

→저학년 학생 대상으로 데이터분석과 시각화를 쉽게 할 수 있는 Sometrend, EC miner, Tableau, 엑셀 등을 활용한 교육진행 결과 학생들의 참여도가 높아짐
→ 특히, Z세대에게 롤플레이 기반의 교육이 효과적이며 이러한 툴 활용시 롤플레이 강의가 원활하게 진행될 수 있음

섬트렌드 비즈를 활용한 빅데이터 활용 교육: 지방선거 결과 예측

- 과목: 디지털마케팅 전략, 마케팅관리
- 프로젝트 주제: '빅데이터 분석에 기반한 선거결과 예측 및 정책별 감성분석을 통한 선거전략 제안' 을 각 팀별 인스타그램계정에 공유
- 기간: 4.25~6.15
- 주차별 팀프로젝트
 - 1주차: 후보자별 추이분석과 연관어 분석
 - 2주차~4주차: 각 후보자의 공약별 감성분석과 타겟분석 이슈 도출(5.19일 발표)
 - 5주차: 최종 결과 예측
 - 6주차: 예측과 결과 비교 및 토론

섬트렌드 비즈를 활용한 빅데이터 활용 교육_1주차 과제사례

1주차: 후보자별 추이분석과 연관어 분석

- 세종시장: 이춘희 vs 최민호
- 서울시장: 오세훈 vs 송영길
- 부산시장: 박형준 vs 변성완
- 경기도지사: 김동연 vs 김은혜
- 용인시장: 백군기 vs 이상일
- 강원도지사: 이광재 vs 김진태
- 인천시장: 박남춘 vs 유정복

섬트렌드 비즈를 활용한 빅데이터 활용 교육_1주차 이슈도출 과제사례

후보자별 추이분석과 연관어 분석 경기도지사 예시: 단순 추이분석과 연관어 분석

김은혜 후보 채널별 추이분석

김은혜 후보 채널별 추이

김은혜 후보는 5월 2일 윤석열 당선인과 오찬으로 인해 언급량이 많았고 5월 4일에 교육 공약 기자회견 발표로 인해 뉴스에 언급량이 많았습니다.

지방선거 여론 분석

김은혜 후보 연관어 분석

김은혜 후보의 주요 연관어로는 경기, 김동연, 경기도지사, 선거, 경기도, 국민의 힘, 후보, 윤석열, 민주당, 여론, 경기지사 등 연관어가 분석되고 있습니다.

youbigdata • 활동일

youbigdata 경기도지사 선거가 얼마 남지 않은 차를, 가장 최근에 작성된 댓글을 읽어 경기도가 되면, 김은혜 후보가 가장 많은 검색량을 보였습니다.

4월 28일부터 5월 5일까지 2주간 각종 SNS, 커뮤니티와 뉴스에서 김은혜 후보의 언급량이 꾸준히 상승세를 기록했습니다.

#윤석열과하태고 #비박이대안과 #2022년지방선거 #경기도지사후보 #youbigdata

섬트렌드 비즈를 활용한 빅데이터 활용 교육_1주차 이슈도출 과제사례

후보자별 추이분석과 연관어 분석 인천시장 예시: 공약과 추이분석을 통해 주요 정책 이슈 도출

- 01 인천 어르신은 날은이가 제일 잘 모셔요**
 - 1.65세 이상 어르신은 무료 대상료 면 혜택제공
 - 2.취약계층 어르신 식사배달,어르신 돌봄방산 (푸드뱅크 연계)
- 02 혜택모아 청년지원: 청년자산 형성 지원 드림 5종 패키지**
 - 1.드림체크: 구직형년 최대 300만원
 - 2.드림포인트: 첫해당 월년 최대 120만원
 - 3.드림 for 청년통장: 연 120만원 저축시 3년후 1000만원 수령
 - 4.드림학비 대출이자지원 2배 확대
 - 5.드림 카드 (청년 E음 카드)
- 03 학교 졸업해도 체육 시간이 있어요**
 - 1.교에서 15분 내 운동할 수 있는 체육강사명 제정
 - 2.시민 누구나 1회 신청해 체육 지도를 받을수 있는 '체육 배우기' 도입
- 04 변경동별이랑 인천으로 이사가요**
 - 1.시민운동보안선 순행
 - 2.변경 공원,동반체육시설 확충 및 변경 문화숙제 개
 - 3.광고량어 및 미장권 활성화(TNR) 지원
- 05 버스야 택시야? 버스야 여객선이야?**
 - 1.인천시민 누구나 버스요금으로 연안여객선 이용
 - 2.부르면 달려오는 I-MOD 버스 확대
 - 3.맞춤형 E음버스 확대로 20분 버스 기다리기 없는 도시로
 - 4.도시지역 100명 택시 확대
 - 5.장애인콜택시 서비스 개선

인천시장 후보 박남춘 채널별 추이분석

채널별 추이

2022 인천시장 후보 박남춘은 트위터와 뉴스에서 압도적으로 많은 검색량을 보이고 있으며 뒤이어 커뮤니티와 블로그에서도 많은 검색량을 보여주고있다. 특히 5월5일 가장 많은 검색량을 기록하였다.

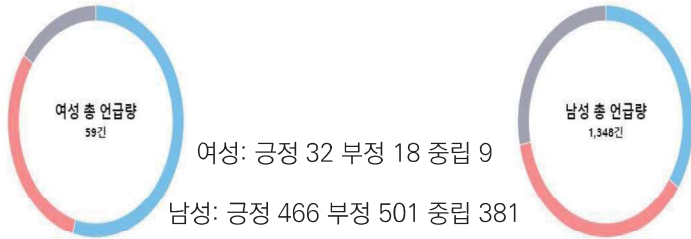
트위터 > 뉴스 > 커뮤니티 > 블로그 > 인스타그램

썸트렌드 비즈를 활용한 빅데이터 활용 교육_2주차 공약별 감성분석 과제

각 후보자의 공약별 감성분석과 타겟분석을 통한 여론분석 경기도지사 예시

감성 언급량 비교 여성 vs 남성 분석 기간 2022/05/09 ~ 2022/05/17

역설 다운로드 이미지 다운로드



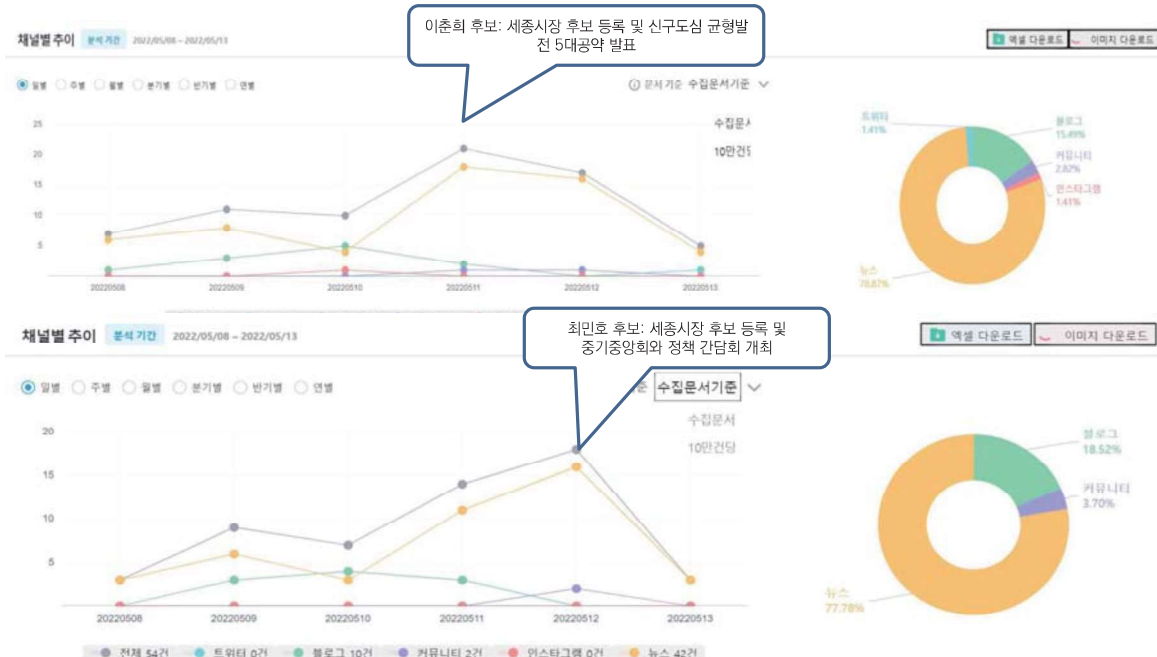
[김은혜 공약 육아 키워드 감성분석 결과]

[육아 공약에 대한 긍정적 의견]

- 김은혜 국민의힘 경기도지사 후보가 '여성의 건강한 성장과 각종 폭력으로부터 안전한 경기도'를 주요 내용으로 하는 여성 공약을 발표하고 본격적인 정책 선거에 돌입
- '여성안심귀갓길 1000개소 구축'을 통해 주거침입이나 성범죄 등 범죄가 빈번하고 1인 가구가 밀집

썸트렌드 비즈를 활용한 빅데이터 활용 교육_2주차 공약별 감성분석 과제

각 후보자의 공약별 감성분석과 이슈 도출 세종시장 예시



섬트렌드 비즈를 활용한 빅데이터 활용 교육_2주차 공약별 감성분석 과제

후보자의 공약별 감성분석과 이슈 도출 - 부산시장 예
연령대별 공약에 대한 상반된 의견(박형준 후보)

워드클라우드를 통한 감성분석 (어반루프에 관하여)



타겟 감성분석 (어반루프에 관하여)



섬트렌드 비즈를 활용한 빅데이터 활용 교육_2주차 공약별 감성분석 과제

각 후보자의 공약별 감성분석과 이슈 도출 부산시장 예시: 부정적 감성에 대한 이슈 도출

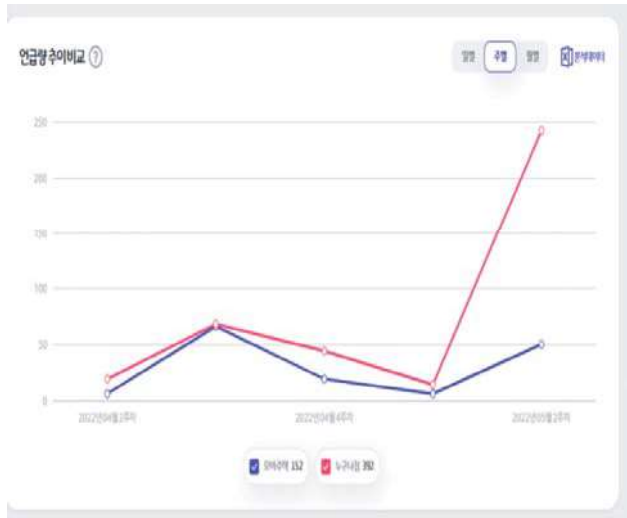
박형준 후보

공약 감성분석에 따른 시사점

- "어반루프 2030 완공" 이라는 공약은 초고속 교통시설을 통해 미래 기술 도시, 국제적인 물류 허브공항 등 지역 이미지를 탈바꿈하고 지역 인식을 고양시킨다는 점에서 긍정적인 감성을 불러일으키고 지역 주민들의 박형준후보의 지지를 늘릴 것으로 보인다.
- 다만, 안전성과 상용화 가능성에 대한 문제, 예산 등 다양한 문제점에서 부정적인 감성을 불러일으키는 것을 알 수 있다.

썸트렌드 비즈를 활용한 빅데이터 활용 교육_2주차 공약별 감성분석 과제

각 후보자의 공약별 감성분석과 이슈 도출 서울시장 예시



오세훈 후보, 송영길 후보 공약별 긍부정 키워드 추이



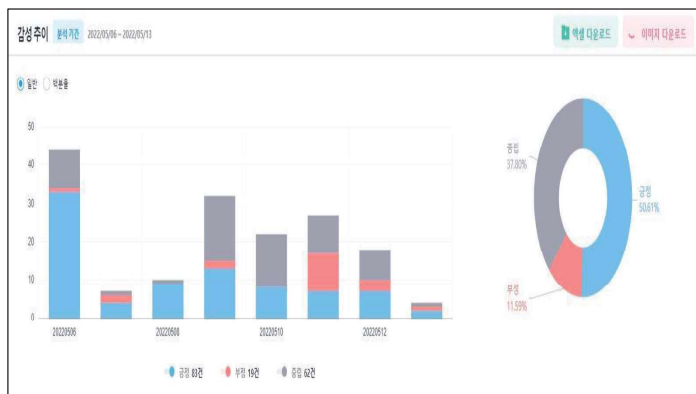
오세훈 후보 모아 주택 공약



송영길 후보 누구나 집 공약

썸트렌드 비즈를 활용한 빅데이터 활용 교육_2주차 공약별 감성분석 과제

각 후보자의 공약별 감성분석과 이슈 도출 용인시장 예시



백군기 후보 주요 공약 : 개발이익 시민환원

- SK하이닉스 유치로 매년 늘어나는 세수 1조 5000억원 중 10%를 시민기금으로 적립하고, 운영수익 전액을 시민 제안사업에 투자 하는 것
- 이 공약에서 유의해야 할 점은 원금상환 이외에 민간사업자들이 추가로 얻는 수익을 어떻게 지역사회로 환원해야 할 것인지, 세수가 감소하는 경우에는 어떻게 적정선을 정해야 할 것 인지에 대한 이슈가 있음

섬트렌드 비즈를 활용한 빅데이터 활용 교육_2주차 공약별 감성분석 과제

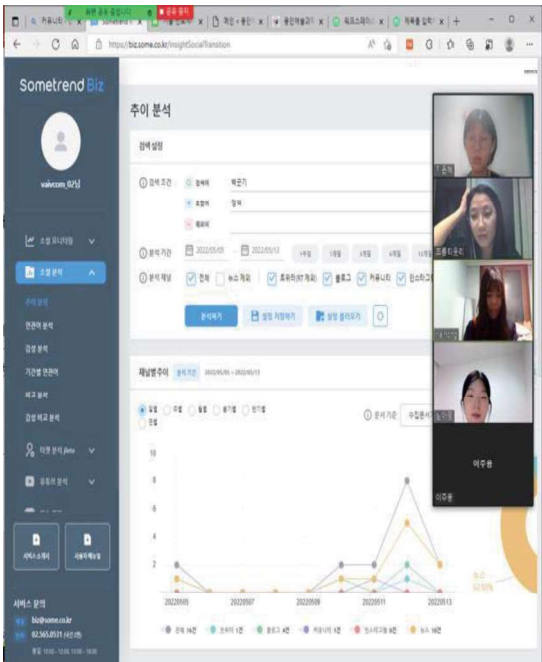
각 후보자의 공약별 감성분석과 이슈 도출 용인시장 예시



이상일 후보 주요 공약 : 재산세 감면

- 과세표준 3억원(공시가격 5억원)이하 1가구 1주택의 재산세 100% 감면해준다는 것
- 조세법률주의에 의거한 국회 입법사항으로 도지사나 특례시장권한 밖의 사항이며 포퓰리즘 공약이라는 것. 이를 활용하기 위해선 국회와 지방정부, 지방의회 차원에서 구체적이고 종합적인 대책마련을 하여 공평과세의 원칙을 철저히 지켜야 해당 공약을 활용할 수 있을 것이라는 이슈가 있음

섬트렌드 비즈를 활용한 빅데이터 활용 교육: 팀 프로젝트 사진 및 소감



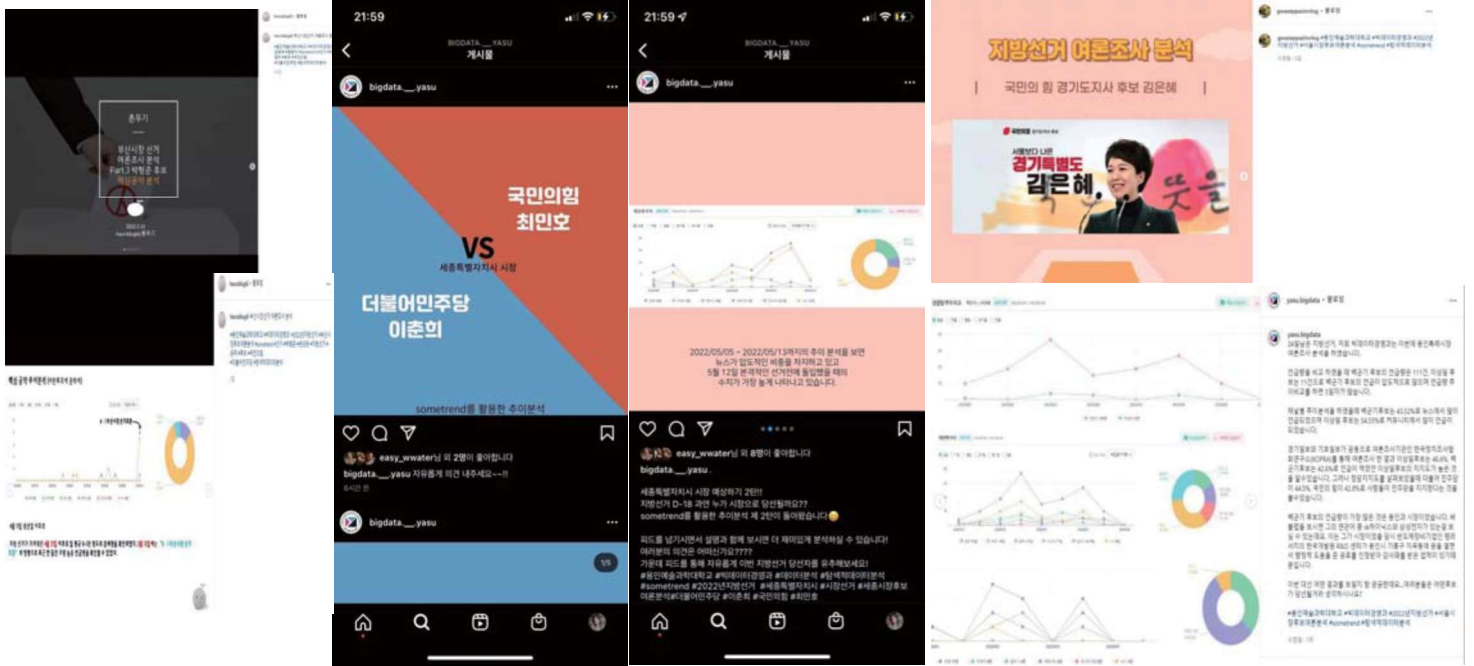
<이지수 학생>



<서주영 학생>

<섬트렌드를 활용함 팀플 줌회의>

썬트렌드 비즈를 활용한 빅데이터 활용 교육: 감성분석 인스타그램 공유



- 팀별 인스타그램에 아래 태그와 함께 빅데이터분석 결과 공유
 #용인예술과학대학교 #빅데이터경영과 #2022년지방선거 #sometrend #탐색적데이터분석

썬트렌드 비즈를 활용한 빅데이터 활용 교육: 썬트렌드의 기능별 의견

1. 추이분석/연관어분석:

- 인터넷에 단어가 언급되는 게시글 원본을 확인하기 쉽고 시각화가 잘 되어 있어 그래프를 이해하기 쉽다
- 트위터, 블로그, 뉴스 등 주요 채널별 원문 찾기에 쉽다.
- 알아보고 싶은 키워드의 일별 언급량 및 채널의 비율을 한눈에 볼 수 있는게 좋았고 전체적으로 정확한 량이 채널별로 잘 나뉘어져 있다.
- 관심있는 단어를 검색하면 각 미디어별로 언급량을 한 눈에 확인 할 수 있고, 어떤 미디어에서 가장 많이 언급이 되었는지도 알 수 있다.
- 일별, 주별, 월별, 분기별, 반기별, 연별로 간편하게 볼 수 있다.
- 원하는 기간 내에 있던 정보를 알아서 분석 할 수 있다.
- 간단하게 많은 양의 데이터를 확인하고 활용할 수 있어 사용하는데 용이하며 분석 단어가 얼마나 많이 언급되고 있는지 파악할 수 있고 특정 단어의 화제성 또는 이슈성을 확인하고 싶을 때 사용할 수 있다.

썸트렌드 비즈를 활용한 빅데이터 활용 교육: 썸트렌드의 기능별 의견

2. 감성분석:

- 속성별 기간별 긍정 부정 단어 변화 추이를 분석하여 분석 단어에 대한 인식과 평판을 파악할 수 있다.
- 다양한 색깔들로 표시가 되어 보기 편했고 여러가지 연관어들도 편리하게 볼 수 있다.
- 키워드에 대한 사람들의 긍정, 부정, 중립적인 면이 나오니까, 분석하기 편하다.

3. 타겟분석:

- 각 연령별, 키워드별 단어를 빠르게 선택하여 분석하는 시간이 단축되었고, 일일이 검색하지 않아도 언급량과 연령별 등 필요한 자료들을 쉽게 얻을 수 있어서 좋았다.
- 관심있는 단어와 그 단어를 언급한 타겟을 직접 정해서 그에 대해 긍정, 부정, 중립인지의 의견을 확인할 수 있고, 그 타겟의 관심사를 유추할 수 있다. 날짜를 지정했을 때 어느 날에 가장 높았는지도 알 수 있다.
- 타겟을 정하기 쉽게 되어있고 분석을 하였을 때 자세히 분석이 되어 조사할 때 좋았다.
- 키워드에 대한 연령대별로 타겟을 분석해서 보기에 편하고, 어떤 연령대가 더 검색을 많이 하는지 빠르게 보기 편하다.

썸트렌드 비즈를 활용한 빅데이터 활용 교육: 썸트렌드에 대한 제안

1. 추이분석:

- 해당 인물을 검색하였을 때 유명한 연예인들이 함께 언급되어 수정하는데 불편함이 있었다.
- 검색 하고싶은 사람이 동일인물 일 때 조건을 추가 하지 않으면 내가 원하는 사람을 검색하기 어려웠다.
- 여러 채널의 데이터를 제공하나, 채널별 주 사용연령층 등 특징을 같이 제공한다면 더 좋을 것 같다.

2. 타겟분석:

- 제시된 키워드의 양이 적어서 검색하는데 약간의 불편함이 있다. 다른 분석들처럼 필수 키워드 입력란이 있으면 좀 더 정확하게 타겟분석이 가능할 것 같다.
- 타겟과 날짜를 지정해서 검색했을 때, 가장 높거나 낮게 언급이 된 이유를 확인할 수 없다.
- 타겟을 검색 했을 때 아무것도 뜨지 않는 경우가 있다.
- 타겟을 특징으로 정하고 하기 때문에 특정 타겟만 분석할 수 있다.
- 설정한 타겟에 대한 분석 결과가 아주 적거나 아예 없을 수도 있어 분석하기 전 조금 더 꼼꼼하게 조사 할 필요가 있다.
- 여성, 남성이 아닌 혼성으로 분석 되어 타겟을 확실하게 알고싶은데 그렇지못해 아쉽다.(타겟이 불분명해 진다)
- 온라인 매체를 거의 사용하지 않는 고령 사용자들의 타겟분석은 정보와 신뢰성이 떨어지는 것 같다.

썸트렌드 비즈를 활용한 빅데이터 활용 교육 효과

6.1 지방선거 결과 예측을 위한 썸트렌드 활용결과

- 썸트렌드를 활용하여 탐색적 데이터 분석을 비교적 쉽게 이행

1) 추이분석과 연관어 분석 기능

추이분석을 통한 이벤트 시점과 이슈 발견-> 연관어 분석을 통해 키워드 도출->공약과 키워드 비교/주차별 여론 추이 확인

2) 감성분석 기능: 후보의 정책별 대략적 여론 파악가능

: R이나 python을 활용한 감성분석은 코딩을 하는데 익숙치 않은 저학년 학부생들이 흥미를 잃을 수 있음. 썸트렌드를 활용함으로써 감성분석 결과를 빠르게 도출하고 원문으로의 쉬운 접근으로 감성분석 결과에 대한 검증이 가능

3) 타겟분석 기능: 정책에 대한 조건별 선호도 파악 가능

: 연령대별/성별/관심사별 정책에 대한 감성분석 결과로 향후 선거방향에 제언 가능

결론 및 향후 필요 연구: 썸트렌드 비즈를 활용한 빅데이터 활용 교육효과

- 빅데이터활용 교육을 위한 썸트렌드 활용 의의
 - 데이터분석가로서 분석역량은 상당히 중요. 그러나 R이나Python을 기반으로 한 코딩부터 접근할 경우, 일부 학생을 제외한 대부분은 흥미를 잃고 포기
 - 따라서, 흥미 유지와 문제 및 이슈를 탐색하기 위한 단계로서 분석결과를 빠르게 도출할 수 있는 툴을 활용하는 것이 탐색적 사고를 향상시키는데 효율적
 - 탐색적 데이터 분석 과정을 빠르게 경험하고 분석결과를 기반으로 한 시사점을 도출하는 과정을 통해 비판적 사고와 논리적 사고를 기를 수 있으며 학생들의 참여도를 높일 수 있음
 - 향후, 롤플레이 방식 수업을 진행시 유용한 툴임
- '빅데이터 분석 도구를 활용 교육이 학습역량을 높이는데 미치는 영향'에 대한 후속 연구 진행
 - ✓ Sometrend 활용 학생들을 대상으로 서베이 진행 예정
 - ✓ 가설
 - 빅데이터 분석 도구를 통한 교육이 직업에 대한 태도에 긍정적인 영향을 미친다.
 - 빅데이터 분석 도구를 통한 교육이 학습태도 및 학습효과에 긍정적인 영향을 미친다.
 - 빅데이터 분석 도구를 통한 교육이 자기효능감에 긍정적인 영향을 미친다.
 - 빅데이터 분석 도구를 통한 교육이 분석 분야에 대한 이해와 태도에 긍정적인 영향을 미친다: 빅데이터 분석 교육 뿐만 아니라 분석 대상 분야 교육(예: 정치)에서 빅데이터 분석 도구 활용 교육의 필요성 인식

빅데이터 기술 교류 세미나 빅데이터와 여론조사



한국경영정보학회

The Korea Society of Management Information Systems
서울시 용산구 한강대로 115, 702호 | office@kmis.or.kr