안녕하세요, 데이터 분석 팀프로젝트 1조의 발표를 맡은 이현주입니다.

저희는 SNS 댓글 및 텍스트 데이터를 통한 백신 접종자 후기 감성 분석을 진행하여 사람들이 갖고 있는 백신의

인식에 대해 조사하고, 이후 팬데믹 상황이 재발할 시에 대해 어떤 대처를 할 수 있을지에 대해 조사해보고자

해당 주제를 선택하게 되었습니다.

--(2번째 슬라이드: 목차)

우선 개요를 소개하겠습니다. 주요 목차는 다음과 같습니다. 첫번째로 주제를 선정하게 된 이유, 그리고 이 주제를 통해 분석하고자 하는 목적과 가설 설정, 그리고 해당 가설을 검증하기 위해 수집한 데이터와 전처리, 그리고 데이터를 분석하고 내린 결론에 대해 소개하겠습니다.

--(3번째 슬라이드: 첫번째 목차)

첫번째로, 주제 선정의 이유입니다. 본격적으로 저희의 데이터 분석 방법과 그 결과를 설명드리기에 앞서, 저희가 어떤 토론을 통해 이 주제를 선택했는지에 대해 말씀드리고 시작하겠습니다.

--(4번째 슬라이드: )

2019년 11월, 중국 우한에서 첫번째 신종 코로나바이러스 환자가 발생하였습니다. 이후 2020년 1월, 세계보건기구(WHO)는 신종 코로나바이러스 감염증에 대해 **국제적 공중보건** 비상사태를 선언하였습니다. 한국에서도 1월 20일에 신종 코로나 바이러스의 최초 감염자가 등장한 이후, () 빠른 속도로 확산되었습니다.

같은 해 3월, WHO는 신종 코로나바이러스 감염증에 대해 범유행전염병, 팬데믹을 선언하였습니다.

--(5번째 슬라이드: )

이 그래프는 코로나 팬데믹이 선언된 이후 1년 동안 한국의 코로나 환자를 나타내고 있습니다. 하루에도 수많은확진자가 발생하고, 병동이 없어 코로나 뿐만 아니라 다른 중환자를 수용하지 못하여 문제가 발생하였습니다. 그것은 그 가족들 뿐만 아니라, 사회 전반에 큰 상처를 안겨주었습니다.

또한, 경제 역시 심각한 타격을 입었습니다. 많은 국가에서 사회적 거리두기와 이동 제한 조치를 시행하면서 여행, 호텔, 레스토랑 등의 서비스 업계는 마비되었고, 제조업과 소매업도 강한 타격을 입었습니다. 많은 사람들이 일자리를 잃고 기업들은 경영난에 처했습니다. 이로 인해 경기 침체와 불황의 그늘이 사회와 경제에 가로막힌 상황이었습니다.

이러한 상황에서 백신의 필요성이 대두되었습니다. 백신은 코로나바이러스와의 싸움에서 가장 효과적인 무기 중 하나입니다. 백신은 개인의 면역력을 향상시키고 바이러스에 대항할 수 있는 항체를 형성하는 데 도움을 줍니다. 이를 통해 중증 증상 및 사망률을 크게 줄일 수 있습니다. 또한, 백신의 대규모 접종은 사회적 거리두기와 규제 조치를 완화할 수 있는 가능성을 제공하며, 사람들이 일상 생활로 돌아갈 수 있는 환경을 조성합니다. 이는 경제의 회복과 사회적인 균형을 촉진합니다

이러한 이유들로 인해, 코로나 백신은 기존의 임상체계를 생략한 채 유례를 찾기 힘들 정도로 빠른 시일 내에 개발되었습니다.

중요한 점은 이러한 불안을 해소하기 위해서는 왜

과학적인 근거와 신뢰할 수 있는 정보를 제공하는 것입니다.

저희는 이런 코로나 백신에 대한 사람들의 반응이 어떠한지, 불안은 어디서 오는지 살펴보도록 하겠습니다.

저희 조의 목적은 코로나 백신에 대한 반응과 첫 번째 팬데믹 시기의 혼란을 조사하고, 팬데믹 상황이 재발할 경우 어떤 혼란이 발생할 수 있는지 이해하고, 그 혼란을 어떻게 줄일 수 있는 방안을 찾는 것입니다

코로나 백신에 대한 불안이 어디서 오는지, 백신에 대한 개선점이나 고찰을 제시할 수 있지 않을까에 대해 생각이 들어 이 주제를 선택하게 되었습니다.

--

두번째로, 목적과 가설 설정입니다.

저희는 첫번째 팬데믹 시의 혼란스러움을 조사하고, 이후 팬데믹 상황이 재발할 시 일어날 수 있는 혼란과, 그 혼란을 줄이기 위한 방안을 찾아보고자 하였습니다.

첫 번째 가설로는 "코로나 백신에 대하여 언론이 사람들의 반응에 영향을 미쳤을 것이다"입니다. 우리는 이 가설을 분석하고 시각화하여 검증해보았습니다. 분석 결과를 보여드리겠습니다.

이 결과를 통해 언론의 영향은 백신에 대한 여론에 영향을 주는 주요 요인이 아니었음을 확인할 수 있었습니다.

두 번째 가설로는 "코로나 백신의 종류 별로 사람들의 여론에 유의미한 차이가 있을 것이다."입니다.

이 가설을 검증하기 위해 우리는 코로나 백신의 반응을 조사하고 비교했습니다. 긍정적인 여론이 높은 백신을 확인하고, 해당 백신을 주로 사용하여 혼란을 줄일 수 있는 방안을 모색하고자 합니다.

우리는 이러한 조사를 통해 첫 번째 팬데믹 시의 혼란을 분석하고, 이후 팬데믹 상황이 재발할 경우 어떤 혼란이 발생할 수 있는지 이해하고자 합니다. 또한, 혼란을 줄이기 위한 방안을 찾아보고자 합니다.

저희는 혼란을 줄이기 위해 다양한 방안을 고려할 것입니다. 예를 들어, 정보의 정확성과 신뢰성을 보장하기 위해 공식적인 정부 기관 및 건강 당국의 지침과 의견을 홍보하는 등의 노력을 할 수 있습니다. 또한, 의사소통과 교육의 중요성을 강조하여 대중들이 적절한 정보를 받을 수 있도록 돕는 것이 필요합니다. 또한, 긍정적인 여론을 받은 백신을 주로 사용함으로써 대중들의 혼란을 최소화하는 방안도 고려할 수 있습니다.

이러한 조사를 통해 우리는 팬데믹 상황에서의 혼란을 이해하고, 혼란을 줄이기 위한 방안을 모색하고자 합니다. 이는 미래의 팬데믹 상황에서 사회적인 혼란을 최소화하고 대처할 수 있는 중요한 토대가 될 것입니다.

--

세번째로, 데이터 수집 및 전처리입니다.

데이터 수집

데이터 수집은 셀레니움과 뷰티풀수프를 이용하여 크롤링을 진행했습니다.

(크롤링 데이터 중 백신에 대한 관심이 제일 많았을 2020년 데이터부터 2022년 데이터까지만 추려냄)

제목에 각 백신의 상표명이 들어간 영상의 URL을 관련도 순으로 300개씩 추려서 해당 영상의 덧글을 추출했습니다.

(csv를 통해 df로 만듦)

(전처리 코드 보여주면서 한글 외의 모든 단어 제거했음)

(konlpy 라이브러리의 코모란 메소드 사용.)

(명사만 추출하여 데이터 저장)

--

마지막으로, 분석과 그 결과입니다.

우리는 감성 사전을 활용하여 감성 분석을 수행하였습니다. 이를 통해 어떤 백신이 가장 긍정적인 반응을 받았는지를 알아보고, 각 백신에서 어떤 명사가 가장 많이 언급되었는지 분석해보았습니다.

우선, 우리는 감성 사전을 활용하여 감성 분석을 수행했습니다. 분석을 위해 KNU에서 제공되는 감성 사전을 사용했습니다. 추출한 명사들을 이 감성 사전의 데이터와 비교하여 감성 점수를 확인했습니다. 감성 점수는 -2부터 +2까지의 범위로 나타나며, 부정적인 점수에서 긍정적인 점수까지의 범위를 통해 백신의 긍부정 비율을 확인할 수 있었습니다.

저희는 화이자, 모더나, AZ, 얀센 순서로 각 백신의 긍부정 점수를 체크했습니다. 그리고 이 점수를 백분율화하여 각 백신의 긍정도를 확인했습니다. 특히, 얀센 백신이 가장 긍정적인 반응을 받았음을 확인할 수 있었습니다.

또한, 저희는 각 백신에서 어떤 명사가 가장 많이 언급되었는지 분석해보았습니다. 이를 통해 백신 별 반응의 키워드를 확인할 수 있었습니다. 어떤 백신에서 어떤 명사가 가장 많이 언급되었는지를 분석함으로써 대중들이 어떤 측면에서 백신에 관심을 가지고 있는지 알 수 있었습니다.

마지막으로, 각 백신 단어 별 연관도를 분석한 결과입니다.

(이 연관도를 통해 어떤 단어가 특정 백신과 가장 연관성이 높은지를 확인 가능. 이를 통해 백신과 관련된 특정 단어들의 주요 키워드를 도출.)

이러한 데이터 분석을 통해 우리는 각 백신의 긍부정 점수와 반응의 명사 분석, 그리고 단어 간의 연관도를 확인할 수 있었습니다. 이는 우리에게 백신의 인식과 대중의 반응을 파악하는 데 중요한 정보를 제공해줄 것입니다.

(높은 이유에 대해 확인)

(코로나 백신의 개발사 별로 나눠서 키워드 별로 형태소 분석하여 나온 결과 보여줌)

(백신의 인식과 언론의 영향성 분석)

(2023년 5월 5일, (2개월 전이죠?) WHO는 3년만에 국제적 공중보건 비상사태가 해제되었음을 공식 선언했습니다.)