

한국대학생 15th “산업공학” 프로젝트 경진대회

트위터 데이터 기반 텍스트 마이닝을 통한 글로벌 공급망 리스크 시각화

팀 명 : SCM Blue-bird
팀 원 : 손 정 습

01

글로벌 공급망 위험관리 (Global Supply Chain Risk Management: GSCRM)

- 초국가적인 차원에 초점이 맞추어져 있어 광대한 영역의 공급망을 유지 및 위험원인규명이 어려움
- 글로벌 공급망 위험관리에서 다양한 공급망 위험인자로부터 생기는 불확실한 위험들을 피하거나 그 위험으로 인한 데미지를 최소화 시키는 것이 중요함[1]
- 기존의 공급망 위험관리 연구는 실증적인 데이터들에 치중되어 있어 글로벌 공급망 상의 긴급한 상황 발생 시 해당 위험을 반영하기 어려움

02

03

04

05

트위터(소셜미디어 플랫폼) 선택

- 트위터 API서비스를 통해 매 순간 변화하는 방대한 정보수집이 가능 → 트위터는 인기있는 소셜 미디어 서비스[2]
- 글로벌 위험 데이터의 수집이 쉬움 → 부정적인 성향의 데이터는 트위터 상에서 확산성이 좋음 [3]
- 트위터 데이터는 국가적 분류 용이 → 트위터 데이터는 지리적 정보 인자를 가지고 있음



따라서, 본 연구는

효과적인 **글로벌 공급망 위험관리 (GSCRM)** 를 위해 **트위터 데이터 및 이점** 을 활용하여
글로벌 비즈니스 환경 속 **실시간으로 변화하는 리스크를 포착 및 시각화** 하는 것을 제시하며,
이는 유용한 정보를 제공하여 긴급상황 발생 시, **빠른 대응 및 해결** 을 위한다.

Keywords - Global Supply Chain Risk Management (GSCRM); Social Network Service(SNS); Big data; Visualization; Risk analysis; Text-mining;

02. Literature Review

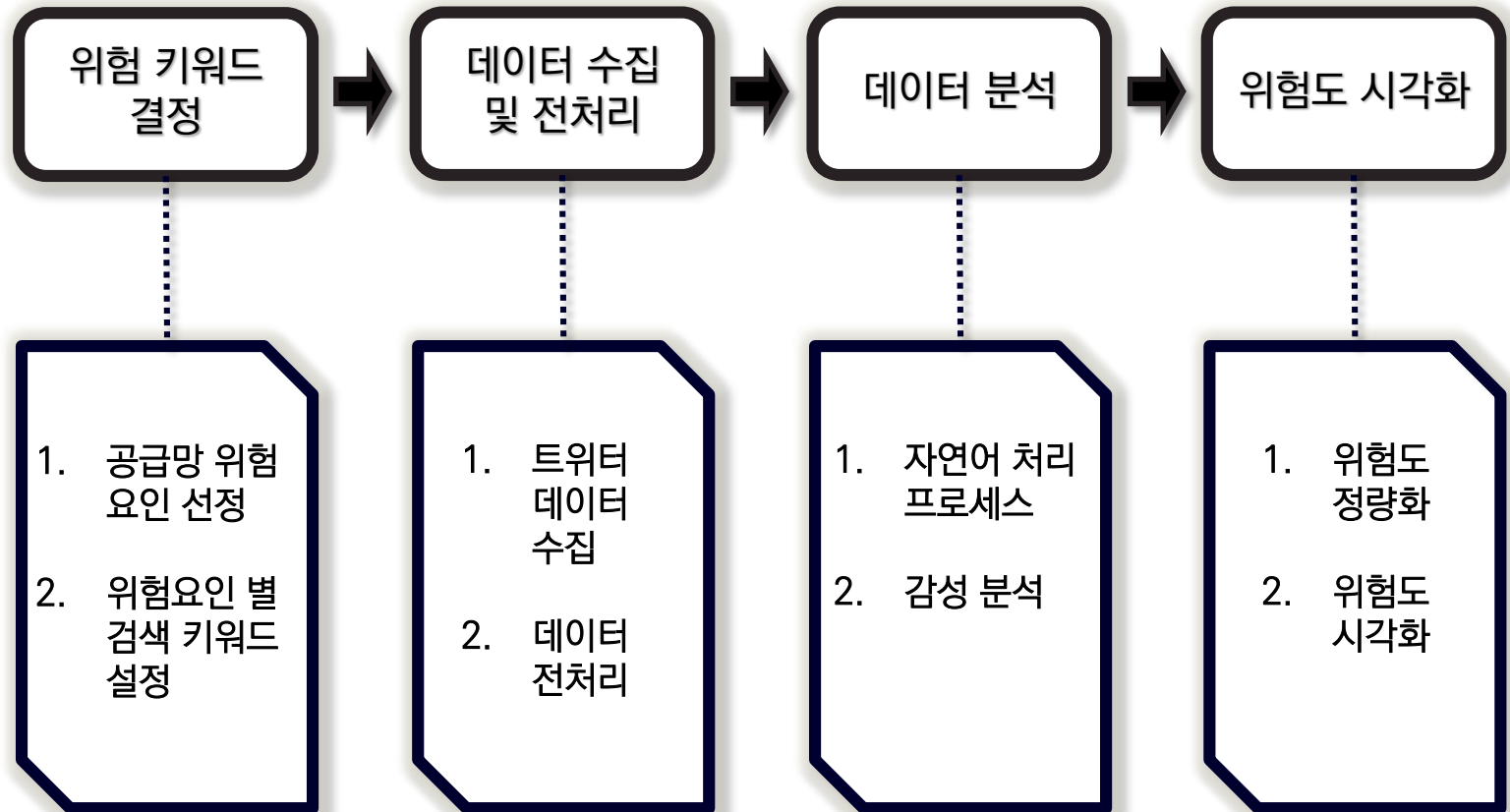
Reference	Approach	Methodology	Data type
Giannakis et al. [2016]	In term of sustainability, Assessing the selected risk's importance and potential causes about environmental, social and econmic	Failure mode and effect analysis technique	interview data
Ho et al. [2015] literature rivew	Literature Review with supply chain risk management research and analysis of potential gaps about papers	Determine and apply criteria for inclusion and exclusion	journal article
Baryannis et al. [2018]	In use of AI methodology, Providing to identify missing and unexplored risk areas	Mathematical parameter, Scenario ,Optimisation, Uncertain parameters(Fuzzy)	x
Behzadi et al. [2018]	Identifying robustness and resilience in agribusiness supply chain risk management , as using mathematical model	quantitative decision modeling	production statistical data
Heckmann et al. [2015]	Reviewing quantitative supply chain risk management focused on definition of supply chain risk	Category analysis Methematical program	literture.
Fahimnia et al. [2017]	developing blood supply network using a coupling function of e-constraint and lagrangian relaxation	hybrid mathematical moeding. (ϵ -constraint method)	data sets
Thun et al. [2011]	Classification of internal and external risk factors and analysis of supply chain risk using probability-impact-matrix	Redesign table	company's empirical data (statistical data)
Basole et al. [2017]	Visualization of Innovation Supply Chain-Network according to layout	Selecting a layout --> mapping data to visual encoding	company's patents
Qazi et al. [2018]	Establishing supply chain risk network using fault tree analysis and expected utility theory	mathematical meodeling	Article

관련 연구의 한계

- 잠재적인 위험원인과 규모에 대해서만 연구 됨
- 위험인자를 규명하고 분류에 집중함
- 공급망 상의 위험을 정량적으로 직관화 시키려는 노력이 진행 중
- 대부분의 데이터 타입은 정적인 데이터로 구성 됨

➤➤ 기존 연구들은 실시간 위험을 탐지하기에 용이하지 않기 때문에,
시시각각 변화하는 글로벌 리스크를 공급망 상에 효과적으로 반영하지 못함

FRAMEWORK



글로벌 공급망 핵심위험인자를 파악하고 해당 위험을 트위터 데이터 감성분석을 통해,
위험을 국가별로 정량화 및 시각화하는 시스템

공급망 위험요인 선정

- 문헌조사를 통해, 글로벌 공급망 상의 영향력 높은 위험요인 선정
 - 자연재해(Natural Disaster), 지리정치학적 문제(Geopolitical Uncertainty)
 - 유류 가격(Fuel Price), 화폐변동(Currency Fluctuation) [4]

위험요인 별 검색 키워드 결정

- 각각의 위험요인과 연관성 있는 검색 키워드 선정
 - 트위터 상에서 검색 시도를 통해 경험적으로 위험탐지에 용이한 키워드 선정
 - 자연재해 키워드 : earthquake, hurricane weather, fire blaze, flood, droughts (5 키워드)
 - 지리정치학적 문제 키워드 수 : trade policy, protest, agreement (3 키워드)
 - 유류 가격 키워드 수 : fuel price, WTI oil, BRENT oil (3 키워드)
 - 화폐 변동 키워드 수 : exchange rate, currency, dollar rate (3 키워드)

Sources of Risk That Affect Global Supply Chain Performance

Risk Factor	Percentage of Supply Chain Affected
Natural disaster	35
Geopolitical uncertainty	20
Volatility of fuel prices	37
Currency fluctuation	29

Source : Supply Chain Management Review (March 2007)

트위터 데이터 수집

- 각각의 위험요인 연관 키워드를 트위터API로 크롤링
 - 사용API : Search API, “tweepy” library in python
 - 데이터 수집기간 : 19/08/26 ~ 19/09/06 (약 10일)

데이터 전처리

- 크롤링 데이터에서 필요한 요소를 추출
 - 데이터 수집 요소 : 데이터 생성날짜, 트윗의 텍스트, 트윗의 나라정보
- 불필요한 정보 제거, CSV파일형식으로 파일을 저장
 - 불필요한 정보 예시) 이모티콘, 특수문자, http정보
 - CSV파일 저장 시 UTF-8 인코딩

자연어 처리 프로세스

- 트위터 텍스트 토큰화와 영어를 제외한 문자 필터링
 - stop_word() 기본셋팅 : English word
 - 사용패키지 : “tidytext” packages in R

감성 분석

- 감성 분석 (AFINN*) : 자연어 처리가 된 토큰(단어)들에 대하여 -5 점에서 부터 +5 점 사이의 감성점수를 매김
 - 감성 기준 : AFINN criterion
 - 사용패키지 : “tidytext” packages in R

*AFINN 기준

R 상에서 사용되는 감성분석 기준 중 하나이다. AFINN 어휘소는 단어들에 -5 에서 +5 사이의 점수를 할당하며 마이너스 점수는 부정적인 감성 그리고 플러스 점수는 긍정적인 감성을 의미한다. 감성 점수의 절대값에 따라 더 그 감성 영향이 큰 것을 의미한다. 해당 AFINN 기준은 라이브러리에 포함된 감성데이터 세트에 이미 테이블링 된 정보를 바탕으로 한다.

위험도 정량화

- 감성점수가 매겨진 토큰(단어)를 나라 별로 그룹화
- 각 나라의 토큰(단어) 감성 점수를 합산
→ 키워드 마다 나라를 기준으로 묶인 감성 점수 분포가 나타남
- 키워드 점수의 평균으로 각 위험요인 채점
- 위험요인의 영향력을 가중치로 두어 전체 위험도 채점

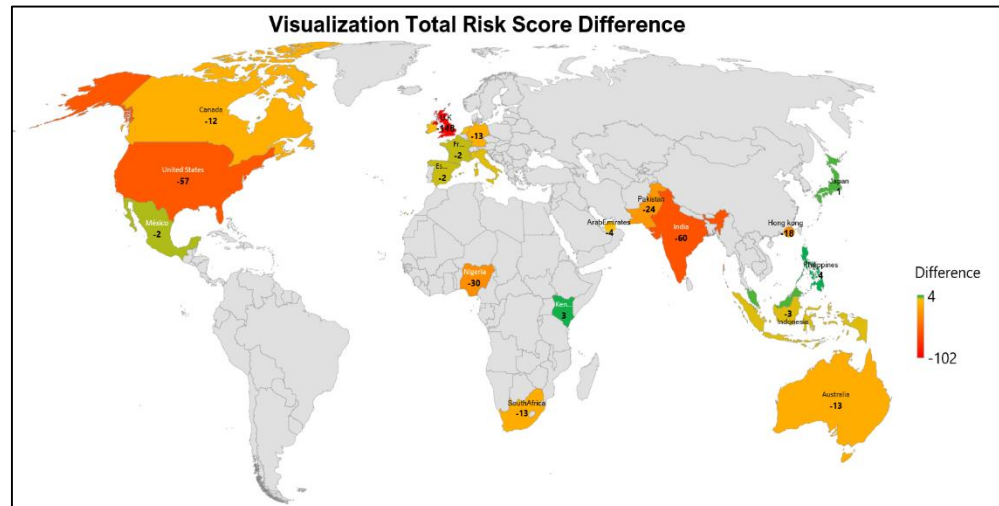
키워드
점수위험요인
점수최종
위험점수

→ 세부 위험점수 계산식은 다음 장을 참고

위험도 시각화

- 전체 위험도 점수를 바탕으로 나라 별 위험점수현황을 그래프화, 위험도를 직관적으로 시각화

→ 예시)



키워드 점수 - 예시) 키워드 : hurricane + weather

부정적인 단어 분포

country	word	value
<fct>	<chr>	<dbl>
1 Antigua and Barbuda	weak	-2
2 Antigua and Barbuda	pressure	-1
3 Australia	bomb	-1
4 Australia	destruction	-1
5 Australia	appalling	-2
6 Australia	pressure	-1
7 Australia	pressure	-1
8 Australia	disaster	-2

나라별 점수 합산

Country	negative score
Antigua and Barbuda	-3
Australia	-10

긍정적인 단어 분포

country	word	value
<fct>	<chr>	<dbl>
1 Antigua and Barbuda	fantastic	4
2 Antigua and Barbuda	curious	1
3 Australia	proud	2
4 Australia	adventures	2
5 Australia	lol	3
6 Australia	lol	3
7 Australia	lol	3
8 Australia	help	2

나라별 점수 합산

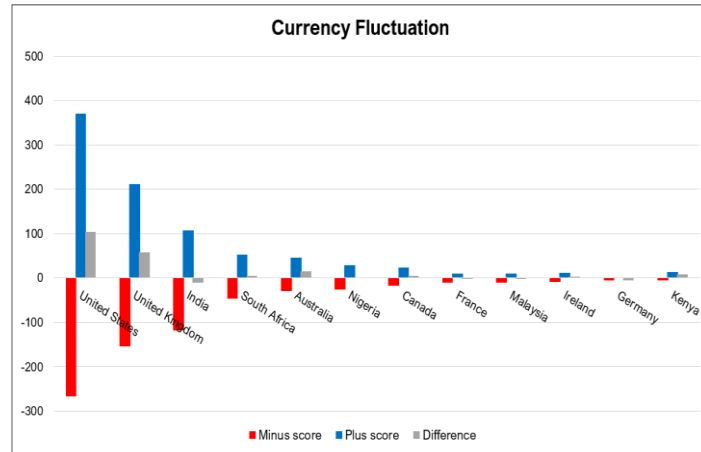
Country	positive score
Antigua and Barbuda	5
Australia	15

Hurricane weather 키워드점수

County	negative score	positive score
Antigua and Barbuda	-3	5
Australia	-10	15

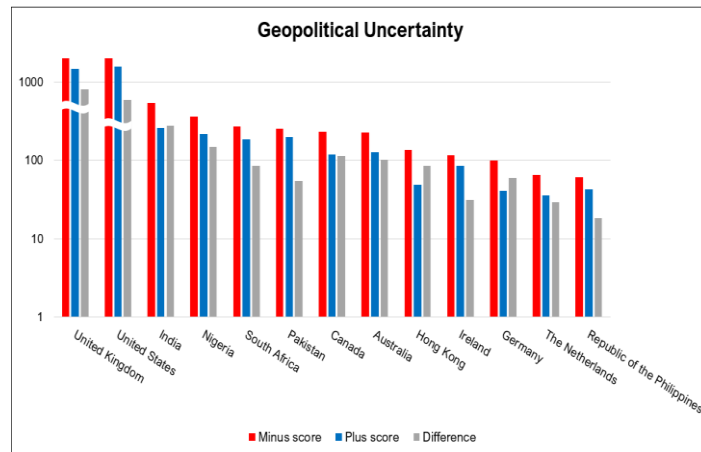
위험요인 및 총 위험점수 계산식

- 위험요인 점수 = 키워드 점수 평균(\sum 키워드 별 긍정(부정) 점수 / 위험요인의 키워드 개수)
 - 최종 위험 점수 = \sum (위험요인 점수 * 위험요인 가중치¹⁾)
- 1) 위험요인 가중치 : 결정된 키워드의 공급망 성과 영향력 (3-1 참고)
 → 자연재해 = 0.35, 지리정치학적 문제 = 0.2, 유류 가격 = 0.37, 화폐변동 = 0.29



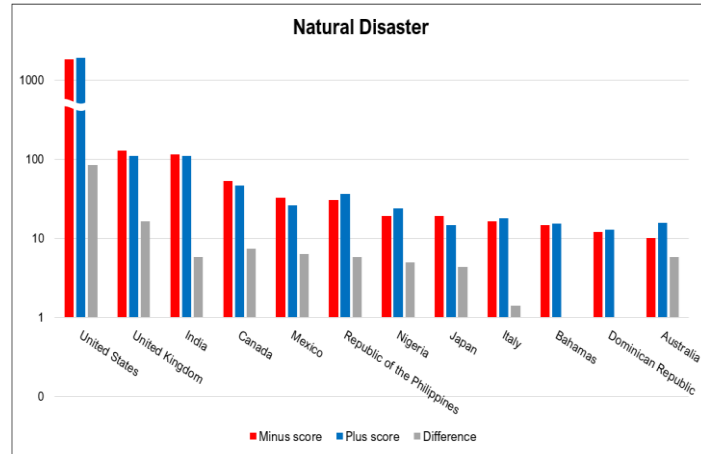
Country	Minus score	Plus score	Difference
United States	-267	371	104
United Kingdom	-155	212	58
India	-119	107	-11
South Africa	-47	52	5
Australia	-30	46	16
Nigeria	-27	28	2
Canada	-18	23	5
France	-11	10	-1
Malaysia	-10	10	0
Ireland	-8	11	3
Germany	-6	1	-5
Kenya	-6	14	8

- 인도와 독일은 화폐변동 측면에서 부정적인 성향임
→ 인도화폐(루피)의 하향하는 추세 [5]



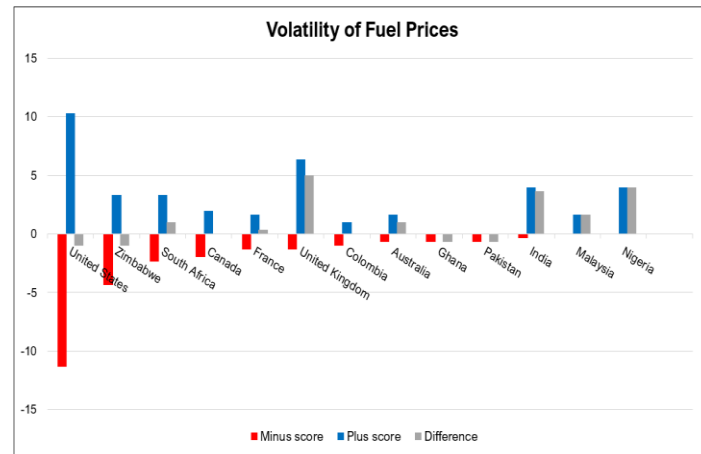
Country	Minus score	Plus score	Difference
United Kingdom	-2276	1474	-802
United States	-2166	1582	-584
India	-539	260	-280
Nigeria	-365	217	-147
South Africa	-269	184	-85
Pakistan	-255	200	-55
Canada	-233	118	-115
Australia	-227	127	-101
Hong Kong	-134	49	-85
Ireland	-117	85	-31
Germany	-100	41	-59
The Netherlands	-65	36	-29
Republic of the Philippines	-61	42	-18

- 지표1위 국, 영국을 포함한 여러 국가에서 지리정치학적 문제는 유효함
→ 영국 저항운동이 지표상에 잘 반영 됨 [6]



Country	Minus score	Plus score	Difference
United States	-1815	1900	85
United Kingdom	-127	111	-16
India	-115	109	-6
Canada	-53	46	-7
Mexico	-33	26	-6
Republic of the Philippines	-31	37	6
Nigeria	-19	24	5
Japan	-19	15	-4
Italy	-16	18	1
Dominican Republic	-15	15	0
Chile	-12	13	1
Bahamas	-10	16	6

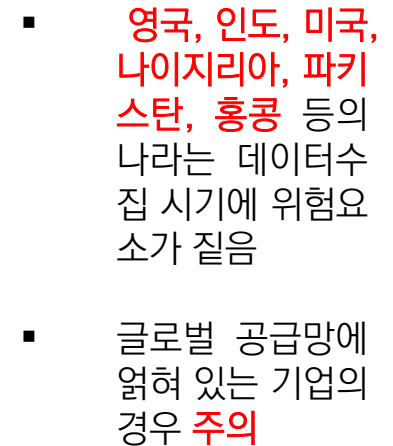
- 영국, 인도, 캐나다, 멕시코 그리고 일본에서 부정적인 성향이 유효하게 나타남
→ 영국은 강한 홍수로 인한 피해 [7]



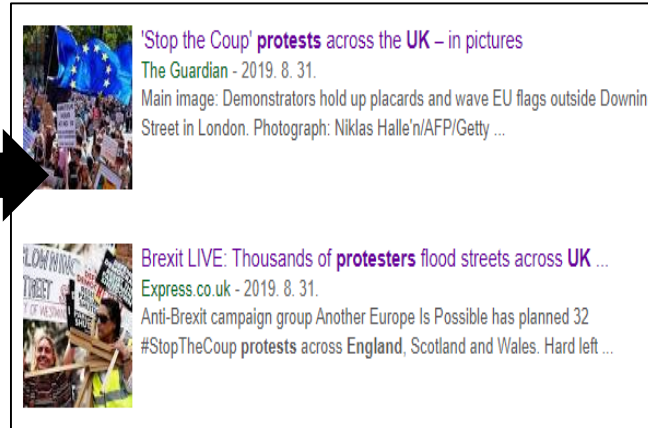
Country	Minus score	Plus score	Difference
United States	-11	10	-1
Zimbabwe	-4	3	-1
South Africa	-2	3	1
Canada	-2	2	0
France	-1	2	0
United Kingdom	-1	6	5
Colombia	-1	1	0
Australia	-1	2	1
Ghana	-1		-1
Pakistan	-1		-1
India	0	4	4
Malaysia		2	2
Nigeria		4	4

- 미국, 짐바브웨, 가나, 파키스탄에서 부정적인 성향이 미미함

01
02
03
04
05

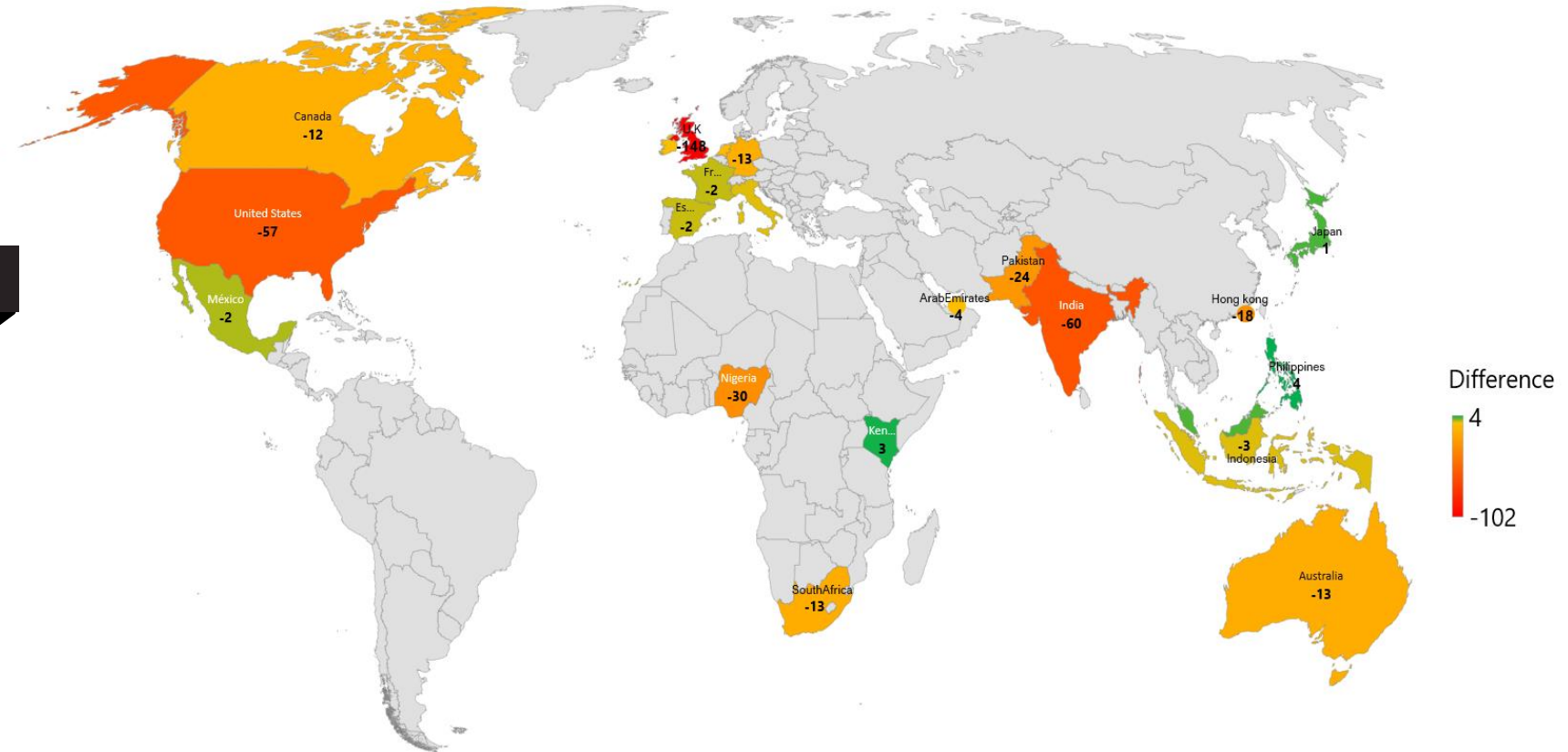


PROTEST 관련 영국 기사



- 데이터 수집 당시, **영국**의 EU 탈출 문제로 인한 **시민들의 저항운동**이 있었음
- 해당 위험 요소가 **가장 유효하게 작용**

Visualization Total Risk Score Difference



- 위험도를 정량화 시켜, **직관적으로 글로벌 시장의 위험도**를 파악
 → 기업의 시장 진출 시, 진입판단의 지표로 사용 가능
 → 실시간으로 나라범위의 시장의 위험성 파악 용이
- 특정 기업을 위한 **특화 위험도** 모델 제시
 → 기업마다 중요시 되는 위험인자를 더 높은 가중치를 두어 반영하여, 특정 기업에 알맞은 위험지표 제공

결론의 시사점

- 트위터 데이터를 기반으로 한 글로벌 공급망 위험관리는 실시간 데이터 반영에 효과적
→ 트위터 사용자 개개인의 실시간 트윗들을 통해 현재 위험요인의 상태 및 위험국가를 보임
- 데이터 수집 기간동안, 가장 유효하게 관찰되는 단어는 “**PROTEST**”
- 데이터 수집 기간동안, **지리정치학적 문제**는 전세계적으로 부정적인 성향이 우세한 위험요인
- **영국**은 공급망 위험관리 관점에서, 가장 **위험**하다고 판단
→ 해당 위험의 원인으로는 “**No - Deal - Brexit**” 라고 불리는 저항운동이며, 이에 따라 부정적인 영향이 야기됨
→ 관련 기사에서는 영국시장에서 자금이 유출되고 있는 위험한 상태라 보도[8]

향후 연구

- 세계 공용어 영어를 기반으로 했으나, 모국어가 영어인 나라에서 상대적으로 데이터가 우세
→ 향후, 번역기능 및 다양한 언어의 검색을 통해, 더 나은 모델링 구축가능
- 위험요인을 대변하는 키워드가 한정적
→ 글로벌 공급망 상의 존재하는 위험요인은 더 많으며, 해당 위험요인을 잘 보여줄 수 있는 다양하고 정확한 키워드들로 재구성한다면 위험을 더 정확하게 측량가능

한국대학생 15th “산업공학” 프로젝트 경진대회

참고자료

- [1] : (Chopra, S., Sodhi,M.S. 2004. Managing risk to avoid supply-chain breakdown. MIT Sloan Management Review, 46, 53-62.)
- [2] : ((2019). Top 15 Most Popular Social Networking Sites & APPs.
<http://www.ebizmba.com/articles/social-networking-websites>)
- [3] :(Naveed,N., Gottron,T., Kunegis,M., Alhadi,A.C. 2011. Bad news travel fast : a content-based analysis of interestingness on Twitter, Proceedings of the 3rd International Web Science Conference : 1-7.)
- [4] : (Jaume Ferre, Johann Karlberg, and Jamie Hintlian. 2007. Integration: The Key to Global Success, Supply Chain Management Review : 24-30.)
- [5] : (Saikat Das. (2019). Rupee fall aids valuation gains in external debt: RBI report.
<https://economictimes.indiatimes.com/markets/forex/rupee-fall-aids-valuation-gains-in-external-debt-rbi-report/articleshow/70899458.cms>)
- [6] : (Boris Johnson. (2019). 'Stop the Coup' protests across the UK - in pictures.
<https://www.theguardian.com/politics/gallery/2019/aug/31/stop-the-coup-protests-across-the-uk-in-pictures>)
- [7] : (Tom Bawden. (2019). Flooding in UK over the past 50 years is worst in Europe as climate change 'increases the amount of rainfall.
<https://inews.co.uk/news/environment/uk-flooding-news-latest-worst-europe-climate-change-494940>)
- [8] : (Don Pittis. (2019). More Brexit chaos for markets after U.K. PM moves to suspend Parliament: Don Pittis. <https://www.cbc.ca/news/business/brexit-economy-confusion-1.5253359>)