https://doi.org/10.15813/kmr.2024.25.1.001

# 온라인 정보 보호: 소셜 미디어 내 정보 유출 반응 분석[1)]

## Online Privacy Protection: An Analysis of Social Media Reactions to Data Breaches

서승우 (Seungwoo Seo) 가톨릭대학교 경영학과[2)]
고영준 (Youngjoon Go) 연세대학교 정보대학원[3)]
이홍주 (Hong Joo Lee) 가톨릭대학교 경영학과[4)]

### 〈 국문초록 〉

최근 개인 정보 유출 사건이 빈번히 발생하고 빈도가 갈수록 증가하는 추세이지만, 개인 정보 유출 사건에 대한 사회나 정보주체인 시민들의 반응은 크게 대두되고 있지 않다. 또한, 개인 정보 유출 사건들에 대한 정보 주체의 반응을 여러 해 기간동안의 데이터에 기반하여 비교하는 연구는 많이 수행되어 있지 않다. 따라서, 본 연구는 2014년 1월부터 2022년 10월까지 국내에서 발생한 주요 개인정보 유출 사건들에 대한 정보주체의 소셜미디어 반응 변화를 분석하였다. 각 사건들이 발생한 직후 일주일간의 기간 동안 네이버 블로그에 작성된 총 1,317건의 포스팅을 수집하였다. 이 포스팅들에 대해 LDA 토픽 모델링 기법을 적용하여 주제를 분석한 결과, 개인정보 유출, 해킹, 정보기술 등 5개의 주요 토픽이 도출되었다. 토픽 분포의 시간변화를 분석한 결과, 개인정보 유출 사건 직후에는 해당 사건에 대한 직접적인 언급 토픽의 비중이 가장 높았으나, 시간이 지나면서 개인정보 유출과 간접적으로 관련된 토픽의 언급 비중이 증가하는 것을 확인하였다. 이는 개인정보 유출 사건 발생 후 정보주체의 관심이 시간이 지남에 따라 해당 사건에서 벗어나 관련 토픽으로 옮겨지고, 개인정보 보호에 대한 관심 또한 줄어든다는 것을 의미한다. 본 연구 결과는 향후 개인정보 유출 사건 이후 정보주체의 프라이버시 인식 변화에 대한 연구의 필요성을 시사한다.

주제어: 개인정보 유출, 소셜 미디어, 텍스트 마이닝, 토픽 모델링, 지식경영

---

1) 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017S1A5A2A01025690)
2) 제1저자, swseo@catholic.ac.kr
3) 제2저자, wnsldud@naver.com
4) 제3저자, 교신저자, hongjoo@catholic.ac.kr

논문접수일 2023년 11월 11일 | 1차 수정 2024년 01월 08일 | 2차 수정 2024년 02월 07일 | 게재확정일 2024년 02월 16일

## 1. 서론 및 연구필요성

데이터 기반 비즈니스의 증가와 데이터 분석을 통한 고객 경험 개선을 위한 기업들의 노력으로 많은 데이터들이 기업에서 활용되고 있다(박민정, 채상미, 2017). 이를 통해 고객의 디지털 경험을 개선하고 있으며 고객의 이용 패턴 분석을 통한 지속적인 업데이트를 수행하고 있다(문태희, 2018). 반면, 기업의 데이터 보유 증가로 인해 기업의 정보 유출 사건 또한 증가하고 있으며 유출되는 데이터도 계속 증대되고 있다(ITRC, 2023). 정보 유출, 특히 민감한 개인정보 유출은 매출액에 비례한 과징금뿐만 아니라 행정소송과 고객들의 민사소송의 대상이 되어 기업에게 큰 비용을 야기하게 되며(백승준, 이홍주, 2021; Spanos & Angelis, 2016), 기업의 브랜드 가치와 평판에 나쁜 영향을 미치고 있다(Gwebu et al., 2018; Makridis, 2021). 그렇기에 기업의 위험 관리 측면에서 정보보호는 이제 중요한 영역이 되었고(권영옥, 김병도, 2007; 김기현 등, 2020), 개인정보의 위협과 인식은 높아지고 있다(이기혁, 윤재동, 2008).

정보 유출 사건 또는 정보보안 사고가 기업에 미치는 영향에 대해서는 다양한 연구들이 수행되어 왔다. 권영옥, 김병도(2007), 홍일유 등(2015), 황해수, 이희상(2015)은 국내 기업의 개인정보유출 사고를 포함한 정보보안 사고가 기업의 주가에 미치는 영향을 사건연구 방법론(Event study methodology)을 통해 분석하였다. 세 연구 모두 사건 발생 후에 단기적인 주가하락이 있음을 확인하였다. 정보 유출 사건이 기업의 브랜드 평판과 소비자들의 사용행태에도 영향을 미친다는 연구들이 수행되었다(Gwebu et al., 2018; Makridis, 2021; Turjeman & Feinberg, 2023). 브랜드 평판에 미치는 효과는 소비자와 직접 상대하는 기업에 많은 영향을 미치며 가장 큰 규모의 유출을 일으킨 기업에서 가장 부정적인 효과를 보였으며, 유출된 사이트 이용자들의 이용행태에도 단기적으로 부정적인 변화를 야기하는 것으로 분석되었다.

기업의 정보 유출에 대한 소셜 미디어 반응도 분석되었다(Sinanaj et al., 2015). 데이터 유출이 알려진 이후에 가장 부정적인 메시지가 증가하였고, 사건이 알려진 이후 5일까지도 부정적인 감정이 지속되는 것을 확인하였다. 이러한 부정 감정의 지속은 주식시장에서 기업에 미치는 부정적인 영향보다 더 오래 지속되는 것으로 분석되었다(Sinanaj & Zafar, 2016). Syed(2019)은 2014년 Home Depot의 정보 유출 사건에 대한 소셜 미디어 데이터를 분석하여, 시간의 흐름에 따라 어떠한 관점의 언급이 이루어지는 지를 분석하였다.

소셜 미디어는 개인의 일상을 공유하는 것에서부터 다양한 제품이나 서비스 경험 공유, 정부 정책이나 사회 현상에 대한 의견 개진까지 다양한 목적으로 활용되고 있다(Ghose et al., 2012; Kaplan & Mazurek, 2018; Lee et al., 2021). 특히, 자연 재해나 긴급 상황에서 소셜 미디어가 의사 소통 도구로 활용되기도 하며(Anderson, 2021; Oh et al., 2013), 긴급 상황의 인식과 의사결정에 도움을 줄 수 있다(Yin et al., 2012).

기업의 정보 유출에 대한 소셜 미디어 반응이 미치는 파급효과에 대한 연구가 많이 수행되었으나(Rosati et al., 2019; Sinanaj et al., 2015; Sinanaj & Zafar, 2016; Vemprala & Dietrich, 2019; Wang et al., 2013), 정보유출 사건에 대한 소셜 미디어 반응이 어떠한 내용을 주로 언급하고 있는지에 대한 분석은 많이 수행되지 않았다(Syed & Dhillon, 2015; Syed, 2019). Syed and Dhillon(2015)과 Syed(2019)는 Home Depot의 정보 유출 사건에 대한 소셜 미디어 반응을 분석하였으며, 위기 단계 이론(Crisis stage theory)(Fink, 1986)을 적용하여 내용을 단계별로 구분하여 단계에 가장 많이 포함되는 내용과 감정을 분석하였다.

기업이 데이터에 기반한 비즈니스 활동을 펼치고 있기에 기업정보 유출사고가 이전에 비해 빈번하게 발생하고 있다. 그 중에서도 기술의 유출보다 개인정보 유출 사고

가 가장 자주 발생하고 있고 기업의 주가에도 영향을 주고 있다(김태환 등, 2014). 소셜미디어가 개인의 일상뿐만 아니라 사회 현상, 그리고 다양한 의견 개진 등에 활용되고 있기에(Vosoughi et al., 2022), 소셜미디어 내의 언급내용을 분석하여 정보주체의 반응을 연구할 수 있다. 따라서, 개인정보 유출사건에 대한 반응 분석을 통해 개인정보에 대한 인식과 시간의 흐름에 따른 변화를 알아보고자 한다.

본 연구는 국내의 정보 유출 사건이 발생한 후 사건에 대한 소셜 미디어 반응을 수집하여 콘텐츠 분석을 수행하고자 한다. 정보 유출 사건 발생이나 정보유출에 따른 과징금 부가가 신문기사를 통해 알려진 국내 기업들의 정보 유출 사건을 파악하였으며, 유출사건 발생일 혹은 과징금 부과가 알려진 시점으로부터 각 기업에 대한 일주일간의 네이버 블로그 포스팅을 수집하였다. 시간의 흐름에 따라 소셜 미디어 반응의 차이를 분석하기 위해서 다년간의 정보 유출 사건에 대한 소셜 미디어 반응을 수집하여 반응의 변화에 대해서 분석을 수행하였다. 토픽 모델링 등의 자연어 처리 방안을 적용하여 포함된 토픽을 찾고 토픽의 분포 변화에 대해서 분석하였다. 또한, 포스팅에 포함된 감성에 대해서도 분석하여 시간의 흐름에 따른 변화를 파악하였다.

개인정보 유출 사고가 발생하고 시간이 지나면 소셜미디어상에서 언급이 줄어들고 부정적인 감성이 나타나는 것은 당연하지만, 소셜미디어 언급량 데이터를 바탕으로 실증 분석한 연구가 많지 않다. 또한, 토픽 모델링을 통해 시간의 변화에 따라서 어떻게 변화하는지, 긍정이나 부정적인 감성 외에 어떤 토픽이 주로 언급되고 사건과 관련해서 논의되는지를 파악하여 정보주체 관심사의 변화, 기업이 개인정보 유출 사고가 발생했을 때 대처할 수 있는 방안을 모색하고자 한다.

2장에 관련연구를 정리하였으며, 3장에서 본 논문이 활용한 자료에 대해 기술하였다. 4장에서 분석방안과 결과를 제시하며, 5장에서 본 연구의 결론과 시사점을 논의하였다.

## 2. 관련 연구

### 2.1. 개인정보 유출

프라이버시는 국립국어원 표준국어대사전에 따라 개인의 사생활과 그것을 보호받는 권리로 정의된다. 정보보안 분야에서는 이를 개인정보의 통제 권한으로 해석하며, 프라이버시에 대한 우려(Privacy Concern)는 이러한 통제권이 침해될 수 있다는 불안으로 본다(김상희, 김종기, 2017).

정보 보안에 대한 초기 연구는 프라이버시에 대한 우려를 중심으로 진행되었으며, 프라이버시 염려에 대한 이론과 차원구분, 측정을 위한 연구가 많이 이루어졌다. 대표적인 측정모델은 Smith et al.(1996)의 CFIP(concern for information privacy) 이며, 이에 기반한 많은 측정 모델들이 제안되었다(Stewart & Segars, 2002). Li(2011)은 프라이버시 분야의 실증 연구를 대상으로 문헌적 고찰 후 일반적인 프라이버시 염려와 조직 중심으로 통합적 프레임워크를 제시하였다.

프라이버시에 대한 관심이 높지만, 사용자들은 종종 이익과 위험을 비교하여 정보를 제공하는 결정을 내린다. 이러한 접근 방식을 설명한 이론이 프라이버시 계산이론(Privacy Calculus Theory)이다(Laufer & Wolfe, 1977). 실제로, 많은 사용자들이 프라이버시에 대한 높은 염려에도 불구하고 보상이나 서비스 이용의 혜택을 위해 개인 정보를 제공하는 현상이 있으며, 이는 프라이버시 역설(Privacy Paradox)로 알려져 있다(Belanger & Crossler, 2011; Norberg et al., 2007; Potzsch, 2009). 이 역설적인 현상은 Acquisiti and Grossklags(2003)가 정보 노출의 문

제로 지적하였고, Barnes(2006)는 소셜 네트워크에서 청소년들의 정보 공유에 대한 연구에서 이러한 역설적 행동을 프라이버시 역설로 정의하였다.

또한, 일부 웹사이트의 유용성을 중요하게 생각하는 사용자들은 개인정보 유출사고에도 불구하고 해당 웹사이트를 계속 이용하는 경향이 있으며, 이는 프라이버시에 대한 관심과는 상반된 행동을 보이는 것으로 분석되었다(임명성, 2013). 또한 황용석 등(2020)의 연구에서도 사용자들이 프라이버시에 대한 염려와 편의성 간의 균형을 맞추려고 하는 모순된 행동을 보임으로써 프라이버시 역설을 실증적으로 확인하였다.

김택영 등(2020)은 Google에서 수집된 개인정보 유출사고 데이터를 분석하였다. 이 연구에서는 기업 규모가 개인정보 유출에 큰 영향을 주지 않았으며, 대다수의 대응은 외부 위험에 중점을 두고 있었다는 것을 확인하였다. 한편, 엄재하, 김민정(2016)은 주식시장에서의 개인정보 유출사고가 기업의 가치뿐만 아니라 투자자의 투자성과에도 영향을 미친다는 것을 확인하였다.

임명성(2013)의 연구 결과, 웹사이트의 유용성에 중점을 둔 사용자들은 개인정보 유출이 있더라도 그렇게 걱정하지 않았다. Turjeman and Feinberg(2023)의 연구에서는 불륜 관련 매칭 사이트가 개인정보 유출 후 큰 타격을 받았으나 약 3주 만에 활동이 원래대로 회복되었다는 사실을 발견하였다. 권영옥, 김병도(2017)는 정보보안 사고 발생 후 그 영향이 단기적으로만 제한되는 것을 확인하였고, 이는 정보보안에 대한 일반적인 인식의 부족을 나타낸다.

소병기, 정종수(2021)는 주요 개인정보 유출 사례를 분석하여 중소기업의 사이버보안 강화방안을 제안하였다. 전웅렬(2018)은 개인 정보를 보관하고 관리하는 Personal Data Storage(PDS)에 대한 보안 기준을 제시하였다. 임동성, 이상준(2018)은 수탁사의 정보보안 관리 수준이 조직의 성과에 미치는 영향을 분석하였다. 황윤희, 유진호(2016)는 연간 대형 개인정보 유출 사고가 평균 12회 발생한다는 결과를 제시하였다.

Makridis(2021)의 연구에서는 2002년부터 2018년까지의 정보유출 사건에 따른 기업 브랜드 및 평판 자산의 변화를 분석하였다. 유출이 발생한 기업 중 소비자에게 잘 알려지지 않은 기업은 브랜드 인지도를 높일 수 있는 기회였으나, 큰 사고를 일으킨 기업의 평판은 하락하였다.

## 2.2. 소셜 미디어 텍스트 분석

Cohen and Levinthal(1990)은 새로운 외부 정보의 가치를 인지하고 이해하여 상업적인 목적에 적용시키는 조직의 능력은 조직의 혁신적인 역량으로써 대단히 중요하다고 주장하며 이러한 능력을 흡수역량(absorptive capacity)이라고 정의하였다. Mowery and Oxley(1995)는 전이되는 지식의 암묵적 요소를 다루는 데 필요한 기술과 습득한 지식을 변형하고자 하는 욕구의 집합이라고 흡수역량을 정의하였으며 Kim(1997)은 문제를 해결하고 배우는 역량이라는 정의를 제시하였다. Cohen and Levinthal(1990)에 의하면 외부의 정보를 적절히 활용하는 능력은 혁신 역량의 핵심 요소이며 외부의 정보를 평가하고 활용하는 능력은 관련된 사전 지식의 수준에서 이루어진다. 가장 기본적인 수준에서 사전 지식은 특정 기술과 공유된 언어뿐만 아니라 가장 최근의 과학적、기술적 발전을 포함하고 있으므로 관련된 사전 지식은 새로운 정보의 가치를 인지하고 흡수하며 상업적인 목적에 적용시킬 수 있는 능력을 생성하며, 이러한 능력을 바탕으로 조직의 흡수역량이 구성된다. 소셜 미디어 텍스트 분석을 통해 소비자나 대중의 반응을 분석하여 기업의 마케팅 활동이나 정부의 정책에 반영하려는 연구들이 많이 수행되었다(Ghose et al., 2012; Lee & Bradlow, 2011; Lee et al., 2021). 고객리뷰, 소셜미디어, 구전을 통한 고객간 커뮤니케이션 연구가 생성자이자 수용자인 고객 관점에서 많

이 수행된 연구이며, 기업 관점의 제품 및 브랜드 인식 연구, 투자자와 사회 관점 연구 등이 있다(Berger et al., 2020). 소비자가 인식하는 제품 속성에 대한 연구는 리뷰로부터 속성 추출, 속성에 대한 감성분석, 제품평가로의 영향연구 등이 진행되었다. 속성발굴은 클러스터링, 소셜네트워크분석, 토픽모델링 등 다양한 텍스트마이닝 기법이 적용되었다(Lee et al., 2021). 속성감성 측정에는 감성사전, 딥러닝 모델, BERT 등이 활용되었다(박현정, 신경식, 2020). 시장구조 연구는 제품 언급데이터에서 시장구조를 도출하거나, 속성평가를 활용한 차원축소로 시장구조 맵을 구성하는 연구들이 있었다(Lee & Bradlow, 2011; Netzer et al., 2012).

대부분의 정보 서비스가 개인정보의 수집·저장 및 가공을 기본으로 하는 개인화된 서비스로 발전하고 있으며, 또한 정보주체들이 다양한 서비스를 이용하기 위하여 자발적으로 개인정보를 제공하고 온라인 상에 적극적으로 공유하고 있고(Levmore & Nussbaum, 2010), 소셜미디어 역시 그 중의 하나이다. 소셜미디어에서의 글쓰기 행위나 '좋아요' 등과 같은 표시행위나 추천과 같은 활동은 프라이버시 염려와 무관하지 않을 뿐 아니라 프라이버시 염려에 영향을 미칠 수 있다(서이종, 손준우, 2011)고 보는 것이 타당하다.

개인 정보 유출과 같은 사건에 대한 소셜 미디어 반응과 이의 영향도 분석되었다(Rosati et al., 2019; Sinanaj et al., 2015; Sinanaj & Zafar, 2016; Vemprala & Dietrich, 2019; Wang et al., 2013). Rosati et al.(2019)은 2011년부터 2014년까지 발생한 73개 기업의 데이터 유출사건 87건에 대해 분석하였으며, 데이터 유출에 대한 소셜 미디어 노출이 주가에 미치는 영향을 분석하였다. 데이터 유출 시의 소셜 미디어 노출은 대체로 주가 하락을 악화시키지만, 회사의 미디어 가시성이 낮을 경우에는 이러한 부정적 영향이 적었다. 이 연구는 소셜 미디어 텍스트 내용에 대한 분석을 수행하지는 않았으며, 데이터 유출 사건의 소셜 미디어 노출 여부만 활용하였다. Sinanaj et al. (2015)은 2010년 1월 1일부터 2021년 11월 1일까지 발생한 데이터 유출 사건과 관련한 소셜 미디어 포스트를 블로그, 트위터, 페이스북, 위키 등에서 수집하였다. 이벤트와 주가 간의 관계를 측정하는데 사용되는 abnormal returns(AR), cumulative abnormal returns(CAR)와 유사한 개념의 abnormal sentiment(AS), cumulative abnormal returns(CAS)를 측정하였다. 데이터 유출이 알려진 이후에 가장 부정적인 메시지가 증가하였고, 사건이 알려진 이후 5일까지도 부정적인 감정이 지속되는 것을 확인하였다. 감성 값은 사전 기반 방안을 활용하여 감성 극성(sentiment polarity)을 측정하여 사용하였다. Sinanaj and Zafar(2016)은 2011년 11월 1일부터 2013년 12월 31일까지의 데이터 유출 사건에 대한 주가 데이터와 소셜 미디어 포스트를 수집하였다. 데이터 유출사건이 주가에 미친 영향을 AR, CAR로 측정하고 소셜 미디어 감성에 미친 영향은 AS, CAS로 측정하였다. 데이터 유출 사건이 알려진 날에만 주가에 부정적인 영향을 미쳤고 그 다음부터는 부정적인 영향이 없었으나, 소셜 미디어 포스트에서는 데이터 유출 사건이 알려진 날부터 10일 후 까지도 부정적인 영향이 있음을 파악하였다. 데이터 유출이 대중이 인식하는 명성에 더 부정적인 영향을 미치는 것을 알 수 있다. Vemprala and Dietrich(2019)은 2018년 5월 1일부터 21일까지 데이터 유출과 관련된 키워드로 검색된 트위터 메시지 73,036건을 수집하여 분석하였다. 트윗의 리트윗 숫자로 측정한 메시지의 전달 속도에서 보안의 기술적인 내용 및 위협과 관련된 메시지가 더욱 빨리 퍼져나갔으며, 기술 관련 그룹이나 보안에 관련된 사람들에게 더욱 넓게 퍼져나갔다. Syed(2019)는 2014년에 발생한 Home Depot의 정보 유출 사건에 대한 소셜 미디어 데이터를 통해 기업에 대한 평판 위협(reputation threat)에 대해 분석하였다. 2014년 9월 8일부터 30일까지의 39,416 건의 트윗을 분석하였으며, 토픽 모델링을

통해 소셜 미디어에 포함된 토픽을 추출하였고 유사한 토픽들을 군집으로 묶은 후에 나타난 평판 위협 프레임을 확인하였다. 기업이 고의로 민감정보를 유출했다는 고의(Intentional) 프레임, 유출이 우연하게 발생했다는 우연(Accidental) 프레임, 기업이 오히려 피해자라는 피해자(Victim) 프레임의 세 가지의 위협 프레임이 분석을 통해 확인되었다. 시기에 따른 위협 프레임의 변화에 대해서도 분석하였으며, 초기(buildup)에는 우연 프레임이 많이 언급되었으나 확산 시기(breakout)에는 고의 프레임이 가장 많이 언급되었다. 완화(chronic) 시기에도 여전히 고의 프레임이 우세하였으며, 종료 시기(termination)에는 우연 프레임이 주요한 주제였다.

자연재해나 범죄와 같은 사건(Event)에 대한 대중들의 반응을 소셜 미디어 텍스트를 통해 분석하는 연구들도 많이 수행되었다(Anderson, 2021; Mayr & Statham, 2021). Anderson(2021)은 2013년 9월에 발생한 콜로라도 지역의 홍수 사건에 대한 트위터 메시지를 분석하였다. 사건 전인 2013년 9월 1일부터 9일까지의 메시지와 홍수가 진행된 9월 10일부터 18일까지의 메시지, 그리고 사건 후인 9월 19일부터 10월 31일까지의 메시지를 비교하였다. 텍스트 내용에 대한 비교는 LIWC 도구를 활용하여 긍·부정, 지칭 대상(사회, 가정, 직장)과 '나'와 '우리' 대명사의 빈도를 분석하였다. 홍수가 진행 중일때는 부정과 긍정 감정이 모두 치솟았으나, 홍수 발생 일주일후부터는 긍정 감정이 더 높게 지속되었다. 사회, 가정, 직장에 대한 사회적 연결 단어들이 사건 발생 후부터 증가하였고, '나'에 대한 언급이 홍수 발생시기에는 높아졌으나 '우리'에 대한 언급이 발생 이후에 증가하였다. 집단에 대한 언급은 사건 이후에 공유 자원을 나누고 함께 활용하는데 기회를 제공하였으며 회복에 영향을 미친 것으로 판단하였다. Mayr and Statham(2021)은 2019년 발생한 39명의 이민자가 트레일러에서 사망한 사건의 트럭 운전자 구속 사건에 대한 페이스북 포스트와 코멘트를 대상으로 질적 분석 연구를 수행하였다. 'Free Mo Robinson'이라는 페이스북 페이지가 대상이었으며, 형벌 포퓰리즘과 불법이민, 인신매매와 관련된 대중의 인식을 강화하는 내용이 많이 언급되었다.

## 3. 연구방법 및 연구결과

시간의 흐름에 따른 정보 유출 사건에 대한 소셜 미디어 반응을 수집하기 위해 자료는 개인정보 유출사고가 발생한 기업 중에서 소셜미디어상에서 가장 언급이 많았던 상위기업 23개를 선정하였다. 각 기업의 정보 유출 사건에 대한 일주일간의 네이버 블로그 포스팅을 수집하였으며, 수집한 기업들의 목록과 수집한 네이버 블로그 포스팅에 대한 기술적인 설명은 각각 <표 1>, <표 2>와 같다.

연구방법으로 활용된 토픽 모델링을 수행하기 위해서는 어느 정도 긴 글이 필요하여 짧은 글 위주인 트위터보다는 페이스북이나 블로그가 분석대상으로 적합하였다. 페이스북은 중첩구조로 웹크롤링을 통한 데이터 수집이 용이하지 않은데 반해 네이버 블로그는 상대적으로 데이터 수집이 용이하였으며, 한국에서는 페이스북과 네이버 블로그의 이용도가 유사하며[6] 네이버 블로그의 20~30대 이용자가 전체 이용자의 57%로 대다수를 차지하였다[7].

사건발생일은 개인정보 유출사고가 발생하고 그 사건을 정보주체가 뉴스나, 인터넷 등을 통해 대중적으로 인지한 일자를 말한다. 사건인지일은 개인정보보호위원회(개보위)나 방송통신위원회(방통위), 행정안전부(행안부) 등 규제기관이 개인정보 유출사고나 개인정보 실태점검 시 위반 사항을 발견하여 과태료나 과징금 등 행정처분을 처분하고 매체에 보도한 날짜를 의미한다. 정보주체는 보도자료나 뉴스, 인터넷 등을 통해서 처분된 후 인지

6) 한국언론진흥재단, "'2021 소셜미디어 이용자 조사", 2021.12.21.
7) 한국경제, "블로그는 옛날 SNS?...20대가 블로그 앱 가장 많이 쓴다", 2022.08.04.

〈표 1〉 수집한 포스팅 데이터

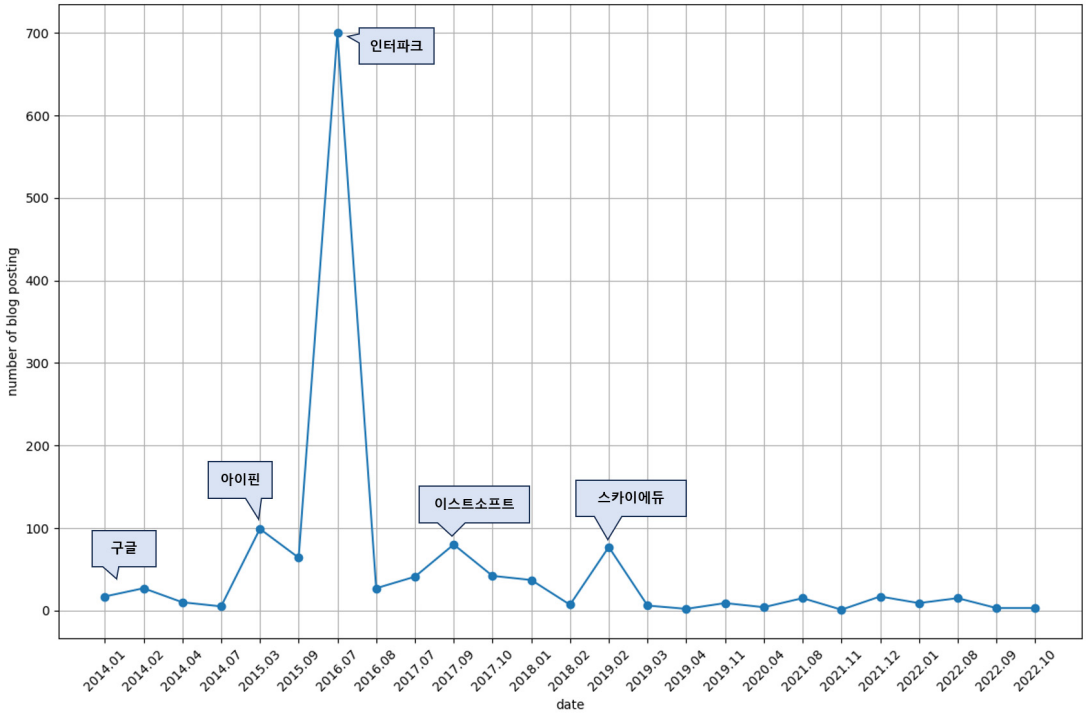
| 구분 | 내용 |
|---|---|
| 수집한 블로그 포스팅 수 | 1,317 |
| 수집된 기간 | 2014년 1월 28일 ~ 2022년 10월 5일 |
| 정보 유출 사건 수 | 23 |
| 포스팅에 포함된 글자 수 평균 | 185.6012 |

〈표 2〉 수집 대상 기업 목록

| No. | 유출 기업 | 연도 | 사건발생일 | 사건인지일 |
|---|---|---|---|---|
| 1 | 구글 | 2014 | | 0128 |
| 2 | 스킨푸드 | 2014 | 0417 | |
| 3 | 아프리카TV | 2014 | 0702 | |
| 4 | 아이핀 | 2015 | 0305 | |
| 5 | 뽐뿌 | 2015 | 0912 | |
| 6 | 한국철도공사(코레일) | 2015 | 0916 | |
| 7 | 인터파크 | 2016 | 0725 | |
| 8 | 빗썸 | 2017 | 0703 | |
| 9 | 이스트소프트 | 2017 | 0905 | |
| 10 | 하나투어 | 2017 | 1017 | |
| 11 | ㈜코빗 | 2018 | | 0124 |
| 12 | ㈜코인원 | 2018 | | 0124 |
| 13 | ㈜이스트소프트 | 2018 | | 0328 |
| 14 | 스카이에듀 | 2019 | 0213 | |
| 15 | 야놀자 | 2019 | 0329 | |
| 16 | ㈜위메프 | 2019 | | 1122 |
| 17 | 스타일쉐어 | 2020 | 0406 | |
| 18 | 페이스북 | 2021 | 0404 | |
| 19 | 샤넬코리아 | 2021 | 0808 | |
| 20 | ㈜브랜디 | 2021 | 1130 | |
| 21 | 로젠㈜ 로젠택배 | 2022 | 0122 | |
| 22 | 스타벅스 | 2022 | 0810 | |
| 23 | ㈜컴투스 | 2022 | | 0928 |

하게 되는데, 개인정보 유출사고가 발생한 기업의 경우 사건발생일과 사건인지일이 모두 나타날 수 있지만 소셜 미디어 데이터 분석을 위해 최초로 인지된 시점을 기준으로 데이터를 수집하였다.

<그림 1>은 수집한 네이버 블로그 포스팅에 대한 월별 분포를 보여준다. <표 2>의 수집대상 기업의 유출사건이 수집기간 동안 매월마다 발생하지는 않았지만, 데이터를 수집하기 시작한 2014년 01월 이후 정보 유출 사건에 대한 블로그 포스팅이 증가하다 2016년 7월에 가장 많은 포스팅이 작성되었고 그 이후 블로그 포스팅이 감소, 유지되는 모습을 보인다. 인터파크의 정보 유출 사건은 2016년 7월에 발생했으며 1,000만 건의 개인정보가



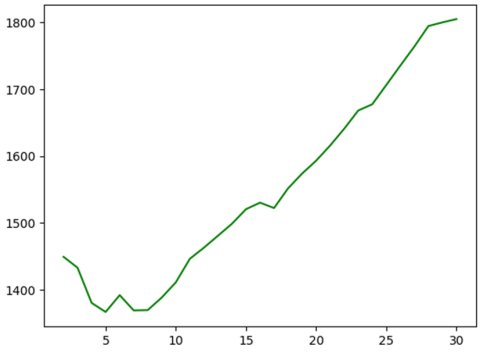

〈그림 1〉 월별 네이버 블로그 포스팅 개수

유출되어 알려진 개인정보 유출 건수 중 단일 기업으로 가장 많은 개인정보가 유출된 사건이기 때문에, 다른 기간에 비해 많은 네이버 블로그 포스팅을 수집할 수 있었던 것으로 판단되며 758개의 포스팅이 수집되었다.

## 4. 분석결과

### 4.1. 토픽 변화 분석

정보 유출 사건에 대한 소셜 미디어 반응 변화를 분석하기 위해 LDA 기법을 활용하여 수집한 문서에 존재하는 하위 토픽을 파악하였다. LDA 분석을 적용하기 위해 Konlpy의 OKT를 활용하여 형태소 분석을 실시하였으며, 한글 불용어[8]를 제거하였다. LDA 기법을 적용하기 앞서 문서의 내용을 분석하기 알맞은 하위 토픽 수를 결정하기 위해 사이킷런(Scikit-learn) 패키지의 혼잡도(perplexity) 계산 기능을 통해 토픽의 수를 증가시켜가며, 지표 값들이 변화하는 추세를 파악하였다. <그림 2>는 토픽의 수에 따른 혼잡도 값을 보여주며, 본 연구에서는 혼잡도 지표 값이 가장 낮게 기록된 토픽의 수인 5개를 선택하여 토픽 분석을 수행하였다.

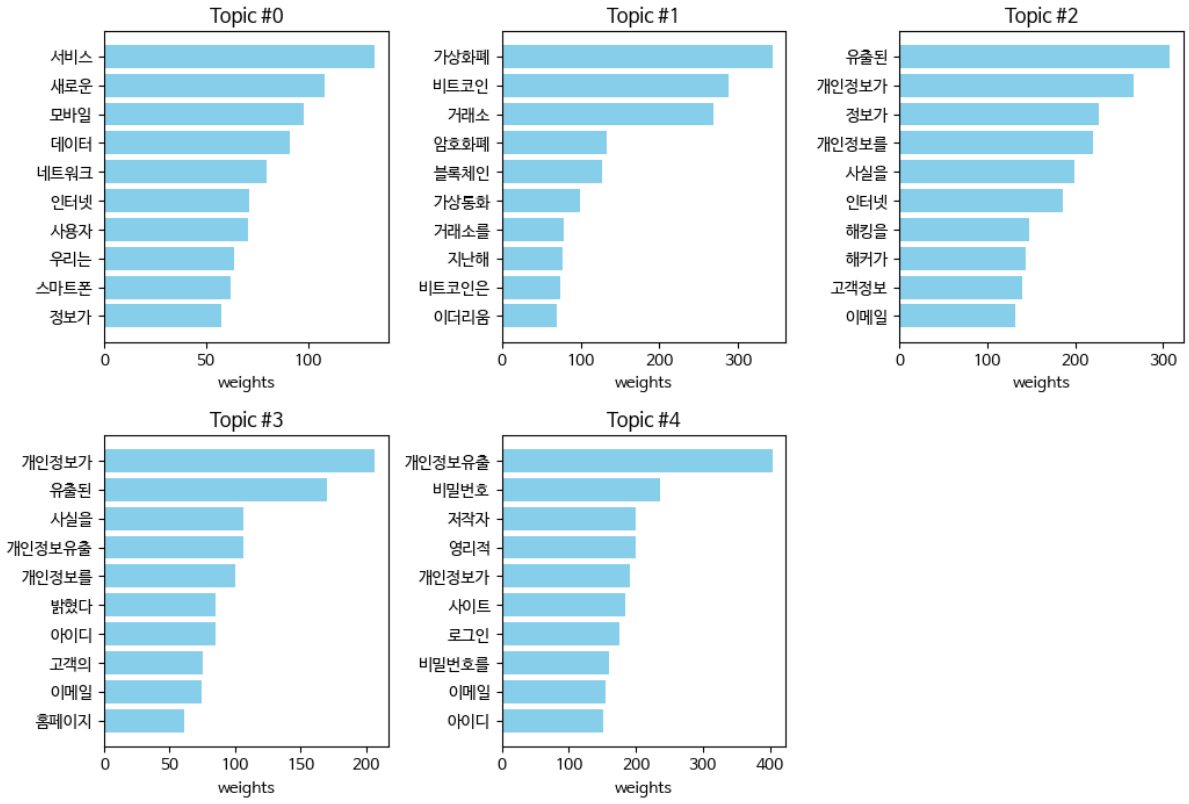



〈그림 2〉 토픽 수에 따른 혼잡도(Perplexity)

8) https://gist.github.com/spikeekips/40eea22ef4a89f629abd87ed535ac6a에 정의된 불용어를 제거하였다.

<그림 3>은 토픽의 수 5개로 LDA를 실행하여 얻은 토픽들과 토픽별로 각 토픽을 대표하는 주요 단어 10개, 그리고 단어들의 가중치를 표시했다. 수치적으로 토픽의 수를 5개로 나눌 때 가장 문서 간의 구분이 잘되지만, 주관적 판단으로는 다소 모호한 부분이 있는 것으로 보일 수 있다. 토픽 2, 3, 4에 공통적으로 개인정보라는 단어가 많이 포함되지만, 이는 개인정보 유출에 대한 반응을 분석하는 본 연구의 특성상 수집한 데이터에 빈번히 포함되는 단어이기 때문이다. 따라서 조금 더 자세히 토픽을 구분한다면 토픽 2는 개인정보유출+해킹, 토픽 3은 개인정보유출+공지, 토픽 4는 개인정보유출+비밀번호로 나눌 수 있다. <표 3>은 토픽 2, 3, 4를 세부적으로 나눈 것처럼 5개의 토픽에 대해 명칭을 부여하여 토픽들의 의미를 구별하기 쉽게 하였다.

각 토픽의 의미를 비교해보면 토픽 0번은 서비스, 모바일, 데이터 등 정보 기술에 관한 단어가 속한 토픽으로




〈그림 3〉 토픽 별 상위10개 단어들의 가중치

〈표 3〉 토픽 레이블

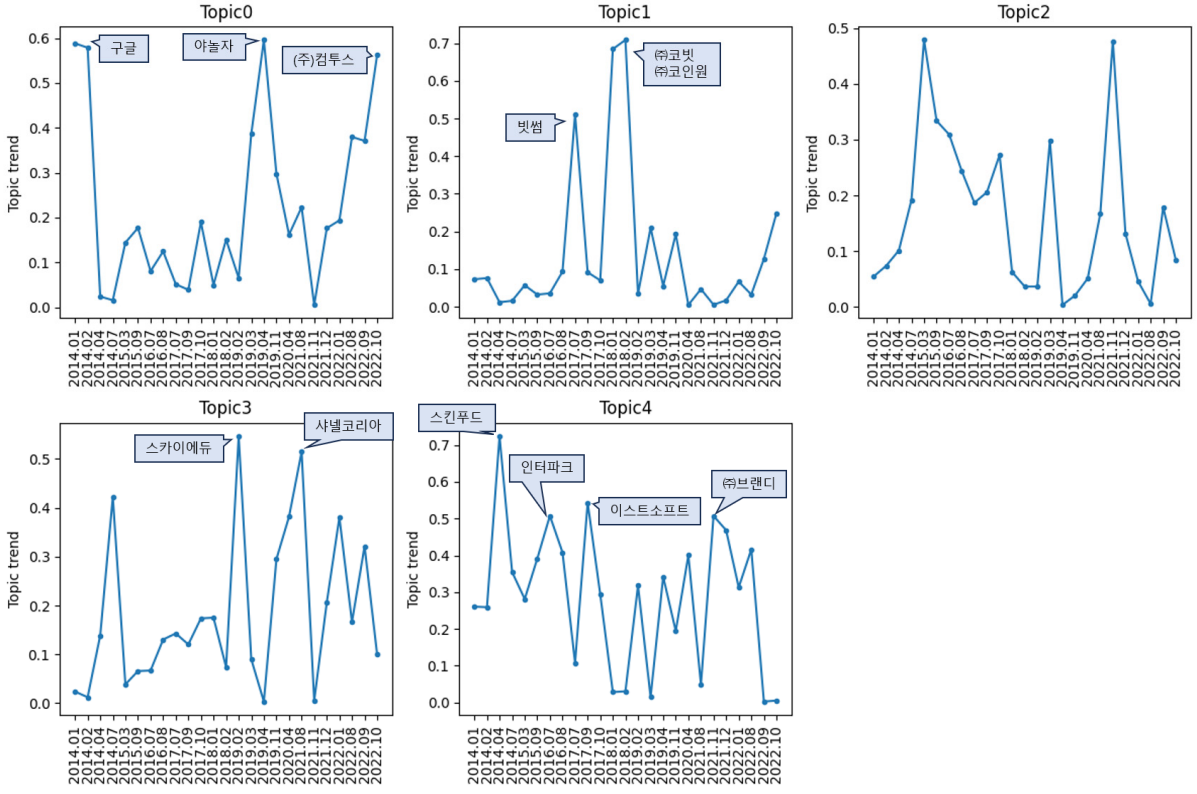
| 토픽 번호 | 레이블 |
|---|---|
| 0 | 정보 기술 |
| 1 | 암호 화폐 |
| 2 | 개인정보유출+해킹 |
| 3 | 개인정보유출+공지 |
| 4 | 개인정보유출+비밀번호 |

개인정보 유출 사건을 통해 증가한 관련 정보 기술에 관한 관심을 나타낸다고 볼 수 있다. 토픽 1번은 가상화폐, 비트코인, 거래소 등 암호화폐와 관련된 단어가 속한 토픽으로 분석을 위해 수집한 데이터 중 암호화폐 거래소를 운영하는 기업의 개인정보 유출 사건에 관한 내용을 담고 있다고 볼 수 있다. 토픽 2번, 3번, 4번은 공통으로 개인정보, 개인정보유출이 속한 토픽들로 세부적으로 분류하면 토픽 2번은 해킹과 관련된 내용, 토픽 3번은 기업의 개인정보 유출 사실 공지에 관련된 내용, 4번은 사이트의 이메일, 비밀번호 유출에 관련된 내용으로 볼 수 있다.

토픽 1번에서 암호화폐와 관련한 주제가 등장한 이유는 첫째, 실제 연도별 개인정보 유출 사고 발생일 혹은 규제기관으로부터 행정처분 등을 받아서 외부에 알려진 사건 인지일을 기준으로 상위에 있는 23개 기업을 대상으로 수집하였고, 분석한 기업 중 3개의 기업이 가상자산 거래소로 데이터의 13%의 비중을 차지하고 있다. 둘째, 대표적인 암호화폐인 비트코인(Bitcoin)이 2017년 말 하락을 시작하여 2018년 1월까지 73%가 하락하였고, 그 시점에 가상자산 거래소 중 가장 상위에 있거나 상위 거래소에 포함되는 기업인 빗썸과 코인원, 코빗이 언급되면서 많은 관심을 받았기 때문이라고 볼 수 있다.

시간의 흐름에 따른 개인정보 유출 사건에 대한 소셜 미디어 반응의 변화를 파악하기 위해 5개 토픽들의 비중이 시간에 따라 어떻게 변화하는 지를 분석해 보고자 한다. 각 문서에는 생성된 날짜를 통해 특정 기간에 만들어진 문서들의 토픽 분포에 대한 평균을 계산하여 그 기간의 토픽 분포를 알아낼 수 있다.

문서들이 작성된 기간을 월별로 분류하고, 각 월의 토픽 분포 평균을 계산함으로써 데이터를 수집한 전체 기간 동안 소셜 미디어에서의 개인정보 유출 사건에 대한




〈그림 4〉 월별 토픽 비중 변화

관심이 어떻게 변화하는지를 알 수 있다. <그림 4>는 데이터 수집 기간 동안 토픽들의 비중이 어떻게 변화하는지 나타낸다.

토픽 0(정보 기술)은 2022년 이전까지는 구글, 야놀자의 개인정보 유출 사건 당시 각각 전체 문서 중 약 59% 비중을 차지하지만, 그 이외의 개인정보 유출 사건 때에는 20% 내외의 적은 비중을 갖는 토픽이였다. 하지만 2022년 8월 이후부터 30% 대의 비중을 보이며, (주)컴투스의 개인정보 유출 사건이 발생했을 때 56%의 비중을 차지한다.

토픽 1(암호 화폐)은 빗썸과 같은 날 사건을 인지한 (주)코빗, (주)코인원의 개인정보 유출 사건이 발생했을 때 각각 51%, 68%의 비중을 차지했다. 하지만 그 이외 기간에는 매우 낮은 비중을 가진다. 이를 통해 토픽 1(암호 화폐)은 암호화폐 거래 플랫폼 기업인 빗썸, (주)코빗, (주)코인원의 개인정보 유출 사건에 대한 문서들에 의해 생성된 토픽이라고 볼 수 있다.

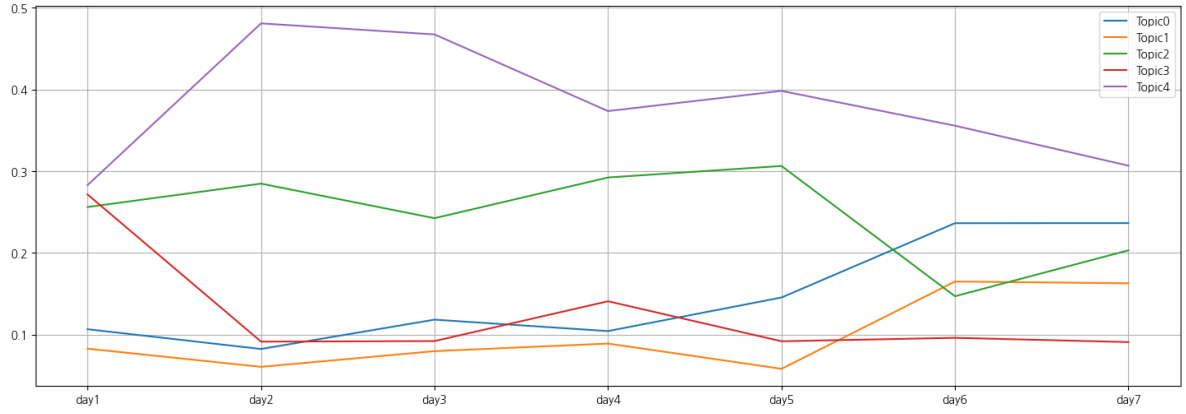
토픽 2(개인정보유출 + 해킹)은 공공 아이핀 시스템이 해킹 공격을 받은 2015년 3월에 48%의 비중을 차지한 후, 2015년 7월까지 30%대의 비중을 유지하고 있었으나 그 이후로 비중이 계속해서 하락했다. (주)브랜디 이용자 66만명의 개인정보가 해커에 의해 판매되는 정황이 확인된 2021년 11월에 48%까지 비중이 상승했지만 그 이후 비중이 급격히 하락했다.

토픽 3(개인정보유출 + 공지)은 아프리카TV의 개인정보 유출 사건이 발생한 2014년 7월에 42%의 비중을 차지한 것을 제외하고 10% 대의 비중을 가지고 있었다. 하지만 스카이에듀의 개인정보 유출 사건과 샤넬코리아의 개인정보 유출 사건이 발생했을 때 각각 55%, 51%의 비중을 차지했다.

토픽 4(개인정보 유출 + 비밀번호)은 스킨푸드의 개인정보 유출 사건이 발생한 2014년 4월에 72%로 전체 기간 중 가장 높은 비중을 갖는 토픽이었다. 이후에도 인터파크, 이스트소프트, (주)브랜디의 개인정보 유출 사건이 발생했을 때 각각 50%, 54%, 50%로 지속적으로 높은 비중을 갖는 토픽이다.

사건 발생 후 일주일간의 반응 변화를 분석하기 위해 각 기업의 정보 유출 사건에 대해 수집한 기간을 day 1, day 2, …, day 7로 분류하고, 각 일의 토픽 분포 평균을 계산함으로써 수집 기간 동안 하위 토픽의 비중 변화를 파악하였다. <그림 5>는 토픽들의 비중이 일주일간 어떻게 변화하는지를 나타낸다.

관련 정보 기술에 관한 내용을 담고 있는 토픽 0(정보 기술)은 day 1에서 약 10% 정도의 낮은 비중을 차지하고 있지만 day 4부터 비중이 상승해 최종적으로 전체 문서 중 약 25%의 비중을 차지한다. 암호 화폐에 관련된 내용




〈그림 5〉 일주일간의 토픽 분포 변화

을 담고 있는 토픽 1(암호 화폐)은 day 1에서 10% 미만의 비중을 차지하고 있지만 day 5에서 비중이 증가하여 최종적으로 10% 이상의 비중을 차지한다. 토픽 0, 1의 비중의 변화를 통해 개인정보 유출 사건 발생 직후에 관련 정보 기술, 암호 화폐 등 개인정보 유출 사건과 간접적으로 연관된 주제에 대한 언급은 매우 적지만, 사건 발생 후 시간이 흐름에 따라 관련 언급에 대한 비중이 증가하고 있음을 알 수 있다.

직접적으로 개인정보 유출에 관한 내용을 담고 있는 토픽 2, 3, 4는 사건 발생 직후인 Day 1에서 각각 약 30% 정도로 높은 비중을 차지하고 있다. 토픽 2(개인정보유출 + 해킹)는 비중이 큰 변화 없이 유지 되다, day 5에서 하락한 뒤 다시 반등하여 최종적으로 약 20%의 비중을 차지한다. 토픽 3(개인정보유출+공지)은 day 2에서 비중이 급격히 하락하여 최종적으로 약 10% 정도의 매우 낮은 비중을 차지한다. 토픽 4(개인정보유출+비밀번호)는 day 2에서 비중이 약 20% 정도 상승하지만, day 4와 day 5에서 비중이 감소해 최종적으로 day 1과 비슷한 수준의 비중을 유지하는 것을 알 수 있다. 토픽 2, 3, 4의 비중의 변화를 통해 개인정보 유출 사건 발생 직후에는 개인정보 유출 사건에 대한 직접적인 언급의 비중이 약 90%로 대다수를 이루지만, 사건 발생 후 시간이 경과함에 따라 그 언급이 점점 줄어들면서 관련된 다른 주제로 관심이 변화한다고 볼 수 있다.

## 4.2. 감성 분석

개인정보유출 사건에 대한 소셜 미디어의 반응을 좀 더 심도있게 분석하기 위해 50,000개의 댓글에 대해 44개의 정서로 레이블링된 KOTE(Korean Online That-gul Emotions) 데이터 셋으로 미세 조정된 트랜스포머(Transformer) 계열 한글 언어 모델인 KcELECTRA를 사용하여 수집한 데이터에 대한 감성 분석을 진행하였다.

KOTE 데이터 셋으로 미세 조정한 KcELECTRA 모델은 입력된 텍스트 데이터에 대해 각각 44개의 감성 레이블과 해당 레이블에 속하는 점수를 출력한다. 이를 통해 각 문서를 문장 단위로 분리하여 감성 분석을 진행하였다. 각 문장 별로 출력된 감성 레이블과 점수의 평균을 통해 개별 문서에 대한 감성 레이블과 점수를 부여하였다. 전체 문서에 대해서도 이와 같은 방법으로 상위 6개의 감성 레이블과 점수를 부여하였다.

<표 4>는 전체 문서의 상위 6개의 감성 레이블과 평균 점수이다. '없음' 레이블은 어떤 레이블에도 포함되지 않는 중립 감성으로 블로그 포스팅을 시작하는 인사말이나 끝 맺음말 등 본문 내용과는 관련없는 형식적인 문장 또는 개인정보 유출 사건에 대한 주관적 의견 이전에 사건에 대해 소개하는 문장에서 비롯된 감성 레이블이라고 볼 수 있다. 그 외의 5개의 감성 레이블은 '불평/불만', '짜증', '화남/분노', '안타까움/실망', '어이없음'으로 모두 부정적인 감성임을 알 수 있다.

<표 5>는 개인정보 유출 사건 발생 이후 소셜 미디어

〈표 4〉 문서 전체의 Top 6 감성 레이블과 평균 점수

| 순위 | 감성 레이블 | 평균 점수 |
|---|---|---|
| 1 | 없음 | 0.8407 |
| 2 | 불평/불만 | 0.7754 |
| 3 | 짜증 | 0.7628 |
| 4 | 화남/분노 | 0.7567 |
| 5 | 안타까움/실망 | 0.7469 |
| 6 | 어이없음 | 0.7352 |

〈표 5〉 사건 발생 일주일 간의 빈도 수 상위 4개 감성 레이블과 평균 점수

| 일자 | 감성 레이블 | 개수 | 평균 점수 |
|---|---|---|---|
| Day 1 | 없음 | 84 | 0.8786 |
| | 불평/불만 | 14 | 0.9090 |
| | 짜증 | 11 | 0.8907 |
| | 안타까움/실망 | 9 | 0.8635 |
| Day 2 | 없음 | 196 | 0.8709 |
| | 불평/불만 | 72 | 0.8964 |
| | 화남/분노 | 36 | 0.9051 |
| | 짜증 | 34 | 0.9036 |
| Day 3 | 없음 | 79 | 0.8686 |
| | 불평/불만 | 23 | 0.8659 |
| | 화남/분노 | 17 | 0.9224 |
| | 짜증 | 14 | 0.909 |
| Day 4 | 없음 | 59 | 0.8714 |
| | 불평/불만 | 13 | 0.8684 |
| | 감동/감탄 | 5 | 0.8847 |
| | 고마움 | 5 | 0.8945 |
| Day 5 | 없음 | 39 | 0.8683 |
| | 불평/불만 | 9 | 0.9319 |
| | 짜증 | 6 | 0.9180 |
| | 안타까움/실망 | 5 | 0.8649 |
| Day 6 | 없음 | 28 | 0.8769 |
| | 불평/불만 | 5 | 0.8849 |
| | 화남/분노 | 4 | 0.9102 |
| | 깨달음 | 3 | 0.8629 |
| Day 7 | 없음 | 65 | 0.8624 |
| | 불평/불만 | 7 | 0.8408 |
| | 감동/감탄 | 5 | 0.8683 |
| | 짜증 | 5 | 0.8789 |

의 감성 변화를 알기 위해 사건 발생 이후 일주일간 각 일의 해당 문서 수 상위 4개의 감성 레이블과 그 레이블에 해당하는 네이버 블로그 포스팅의 개수, 각 레이블의 평균 감성 점수이다. 개인정보 유출 사건 발생 이후 일주일 간 모든 기간에 대해 '없음'과 '불평/불만' 이 공통적으로 포함되어 있으며, 4일차와 7일차에는 '감동/감탄' 감성이 3번째이고, 나머지 다른 기간에는 '짜증'과 '화남/분노'가 3번째 감성 라벨이다.

## 5. 결론

기술적 발전과 함께 인터넷과 소셜 미디어의 활용이 증가하면서 개인정보의 보호와 관련된 이슈가 중요해지고 있다. 소셜 미디어는 개인정보를 실시간으로 생성, 통합, 활용하는 플랫폼으로서 정보 유출에 대한 반응이 크게 나타나곤 한다.

본 연구는 개인정보 유출 사건에 대한 소셜 미디어의

반응 변화를 시계열적으로 분석하였다. 특히, 정보 주체의 프라이버시에 대한 염려와 인식의 차이를 중점적으로 살펴보았다. 소셜 미디어 반응을 수집·분석하기위해 연도별 개인정보 유출사건이 발생한 이후 소셜 미디어 반응이 높았던 유출사건을 기준으로 네이버 블로그 포스팅을 수집하고, LDA 토픽 모델링 기법을 활용하여 주제를 분석하였다.

본 연구의 차별성은 최근 9년 동안 발생한 개인정보 유출 사건에 대한 소셜 미디어 데이터를 기반으로 실증적인 분석을 수행한 점에 있다. 분석 결과, 토픽의 수를 5개로 구분하였을 때 각 토픽이 정보기술, 암호화폐, 개인정보 유출 등의 주제를 포함하고 있음을 확인하였다. 특히, 개인정보 유출 사건 발생 직후에는 해당 사건에 대한 언급이 90%에 이르는 등 대다수를 차지하였으나, 시간이 경과함에 따라 그 언급은 점차 감소하였다. 또한, 수집된 블로그 포스트의 감성분석 결과 개인정보 유출사건을 소개하는 문구를 제외하고는 대부분 불평/불만, 짜증, 화남/분노, 안타까움/실망 등을 나타내는 부정적인 감성임을 확인했다. 다만, 사건 발생 후 5일째부터는 불평/불만에 해당하는 포스트의 건수가 10건 미만으로 줄어들어 발생 초기의 매우 부정적인 평가는 급하게 줄어드는 패턴을 보이고 있다. 암호 화폐 및 그와 관련된 토픽은 초기에는 데이터 분석 전에 예측하기 어려웠던 주제 중 하나였다. 그러나 2010년 후반에 암호화폐 투자에 대한 과도한 열기와 이와 동시에 발생한 암호화폐 거래소의 개인 정보 유출 사건은 많은 관심을 끌었다. 이 주제는 특히 빗썸, 코빗, 코인원과 관련된 유출 사건이 발생한 시기에 큰 관심을 받았으며, 다른 기간에는 상대적으로 더 적은 주목을 받았다.

본 연구의 시사점은 다음과 같다. 첫째, 개인정보 유출 사고에 대한 소셜 미디어 반응은 주로 개인정보 유출, 비밀번호 등 정보 기술과 관련된 주제로 집중되었다. 이는 사건의 원인과 예방 방안에 대한 정보주체들의 관심을 반영하는 것으로 해석된다. 개인정보 유출 사건에 대한 대중의 높은 관심과 개인정보 보호와 관련된 주제에 집중되었다는 점을 고려할 때, 기업은 최신 보안 기술과 프로토콜을 도입하여 데이터 보호를 강화하는 방안을 도입할 필요가 있다. 또한, 연구에서 사용된 시계열적 분석 방법과 유사하게 소셜 미디어 상의 대화를 실시간으로 모니터링하여 유출 사건에 대한 대중의 반응을 파악하고 적절히 대응할 수 있는 체계를 구축해야 한다. 둘째, 짧은 일주일간의 데이터분석이지만, 시간의 흐름에 따라 개인정보 유출에 대한 관심이 점차 감소하였다. 또, <그림 1>에서 알 수 있듯이 개인 정보 유출 사건 발생 이후의 블로그 포스팅 수가 시간이 흐름에 따라 감소하는 패턴이 보인다. 이는 임명성(2013)의 연구와 일맥상통하며, 프라이버시에 대한 인식이 점차 감소하고 있다는 것을 시사한다. 따라서, 기업은 유출 사건 발생 시 즉각적이고 투명한 커뮤니케이션 전략을 수립하여 대중의 불안과 불만을 최소화하는 것에 중점을 둘 필요가 있다.

본 연구는 이러한 실증적 함의에도 불구하고 다음과 같은 한계점이 존재한다. 첫째, 9년간 개인정보 유출 사고 시 소셜 미디어 언급량이 많았던 23개의 기업을 대상으로 소셜 미디어 반응을 분석하여 대략적인 추세를 예측하였기에, 결과를 일반화하기에는 주의가 필요하다. 둘째, 개인정보 유출 사건 발생 후 시간의 경과에 따라 프라이버시 인식이 감소함은 확인하였으나 일시적인 현상인지 전반적인 추세인지는 연도별로 소셜 언급량 변화를 분석하는 연구가 더 필요하다. 셋째, 소셜 미디어 플랫폼마다 사용자의 인구통계 특성과 의사소통 스타일의 차이에서 오는 특이성이 존재하기에 다른 소셜 미디어 플랫폼에서도 비슷한 결과가 나오지 않을 수 있다. 사용자들의 차이로 인한 데이터의 잠재적 편견이 존재할 수 있으며, 특정 측면이 배제될 수 있는 한계가 있다. 따라서 추후 연구에는 많은 플랫폼 데이터를 중장기적인 관점에서 실증분석하면 더 의미있는 연구결과를 기대할 수 있다.

## 〈참고문헌〉

### [국내 문헌]


1. 권영옥, 김병도 (2007). 정보보안 사고와 사고방지 관련 투자가 기업가치에 미치는 영향. **Information Systems Review, 9**(1), 105-120.
2. 김기현, 조혜진, 임소희 (2020). 4차산업혁명 핵심기술 도입 및 정보보호조직에 관한 탐색적 연구: 성과측면에서의 비교분석. **지식경영연구, 21**(1), 41-59.
3. 김동현, 김순석 (2020). 개인정보 비식별 조치를 위한 데이터 상황 기반의 위험도 측정에 관한 새로운 방법. **정보보호학회논문지, 30**(4), 719-734.
4. 김영일, 이재훈 (2013). 개인정보보호투자의 성과측정방안에 관한 연구. **디지털융복합연구, 11**(1), 99-106.
5. 김정규, 이경호 (2017). FAIR를 통한 개인정보 유출에 따른 기업의 손해금액 산출에 대한 연구. **정보보호학회논문지, 27**(1), 129-145.
6. 김채현, 장은조, 류혜원, 이기용 (2022). 정보량을 활용한 개인정보 노출 위험도 측정 기법. **2022년 한국정보과학회 학술발표논문집**, 926-928.
7. 김태환, 이해니, 유진호 (2014). 개인정보 유출사고 이후 기업의 주가변동 패턴에 대한 고찰. **2014년 한국경영정보학회 춘계공동학술대회 논문집**, 89-92.
8. 김택영, 김태성, 전효정 (2020). 기업의 특성이 개인정보 유출 사고에 미치는 영향. **한국IT서비스학회지, 19**(4), 13-30.
9. 김환희 (2019). 학생 개인정보보호의 인식수준에 관한 연구. **2019년 한국엔터테인먼트산업학회 학술대회 논문집**, 155-166.
10. 문태희 (2018). 빅데이터를 활용한 고객 경험 품질 관리를 통한 디지털 경영 혁신. **2018년 한국경영과학회 추계학술대회 논문집**, 322-328.
11. 박민정, 채상미 (2017). 빅데이터 환경 형성에 따른 데이터 감시 위협과 온라인 프라이버시 보호 활동의 관계에 대한 연구. **지식경영연구, 18**(3), 65-82.
12. 박현정, 신경식 (2020). BERT를 활용한 속성기반 감성분석: 속성카테고리 감성분류 모델 개발. **지능정보연구, 26**(4), 1-25.
13. 백승준, 이홍주 (2021). 정보보호 공시제도의 운영실태와 효과성 분석. **지식경영연구, 22**(1), 309-330.
14. 서이종, 손준우 (2011). "신상털기" 현상과 배태된 프라이버시. **사이버커뮤니케이션학보, 28**(4), 49-87.
15. 소병기, 정종수 (2021). 개인정보유출 사고 방지를 위한 중소기업의 사이버 위험관리. **한국재난정보학회논문집, 17**(2), 375-390.
16. 손영수, 황선호, 안다미, 변지연, 김종성 (2011). 국내 개인정보 유출 사고의 사례와 실태. **한국정보처리학회 학술대회논문집, 18**(2), 975-976.
17. 신영진 (2023). 메타버스 서비스에서의 개인정보 침해요인 도출 및 개인정보보호 개선방안. **한국범죄정보연구, 9**(1), 31-57.
18. 엄재하, 김민정 (2016). 정보보안사고가 투자주체별 투자성과에 미치는 영향: 개인정보유출사고 중심으로. **정보보호학회논문지, 26**(2), 463-474.
19. 이기혁, 윤재동 (2008). 민간 기업의 개인정보 유출 위험에 대한 측정 방법과 그 사례에 대한 연구. **정보보호학회지, 18**(3), 92-100.
20. 임동성, 이상준 (2018). 수탁사 개인정보 보호 관리 수준 점검 활동과 정보보안 성과간의 실증 연구. **정보화연구(구 정보기술아키텍처연구), 15**(1), 31-422.
21. 임명성 (2013). 개인정보 유출 사고 후 웹 사이트 가입 지속 및 프라이버시 무관심에 영향을 미치는 요인에 관한 연구. **디지털융복합연구, 11**(1), 107-119.
22. 전응렬 (2018). 개인정보 관리를 위한 PDS의 구조 및 보안기능 연구. **정보화연구(구 정보기술아키텍처연구), 15**(3), 345-356.
23. 천명호, 최종석, 신용태 (2013). SNS에서 개인정보유출방지를 위한 개인정보 유출위험도 측정 방법. **정보보호학회논문지, 23**(6), 1199-1206.
24. 홍일유, 이재훈, 강성민 (2015). 정보보안 사고에 대한 공시가 시장에서 기업의 주식가치에 미치는 영향. **Entrue Journal of Information Technology, 14**(2), 33-56.
25. 황용석, 김기태, 이현주, 이원태 (2020). 인공지능(Artificial Intelligence: AI) 스피커 이용자의 인지지도와 잠재된 프라이버시 인식. **사이버커뮤니케이션학보, 37**(3), 195-231.
26. 황윤희, 유진호 (2016). 개인정보유출 사고의 분포 추정에 관한 연구. **정보보호학회논문지, 26**(3), 799-808.
27. 황해수, 이희상 (2015). 정보보안 사고가 기업가치에 미치는 영향 분석: 한국 상장기업 중심으로. **정보보호학회논문지, 25**(3), 649-664.


### [국외 문헌]


28. Acquisti, A., & Grossklags, J. (2004). Privacy attitudes and privacy behavior – losses, gains, and hyperbolic discounting: An experimental approach to information

security attitudes and behavior. In: Camp, L.J., Lewis, S. (eds) ***Economics of Information Security. Advances in Information Security***, vol 12. Springer, Boston, MA.

29. Anderson, A. A. (2021). Expressions of resilience: Social media responses to a flooding event. ***Risk Analysis, 41***(9), 1600-1613.
30. Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. ***First Monday, 11***(9).
31. Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. ***Journal of Marketing, 84***(1), 1-25.
32. Fink, S. (1986). ***Crisis management: Planning for the inevitable***. American Management Association.
33. Ghose, A., Ipeirotis P. G., Li B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. ***Marketing Science, 31***(3), 493-520.
34. Gwebu, K. L., Wang, J., & Wang, L. (2018). The role of corporate reputation and crisis response strategies in data breach management. ***Journal of Management Information Systems, 35***(2), 683-714.
35. Kaplan, A., & Mazurek, G. (2018). Social media. ***Handbook of Media Management and Economics***, 273-286.
36. Lee, H. J., Lee, M., Lee, H., & Cruz, R. A. (2021). Mining service quality feedback from social media: A computational analytics method. ***Government Information Quarterly, 38***(2), 101571.
37. Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. ***Journal of Marketing Research, 48***(5), 881-894.
38. Levmore, S., & Nussbaum, M. C. (eds.) (2010). ***The Offensive Internet: Speech, Privacy, and Reputation***. Harvard University Press.
39. Li, Y. (2011). Empirical studies on online information privacy concerns: Literature review and an integrative framework. ***Communications of the Association for Information Systems, 28***.
40. Mayr, A., & Statham, S. (2021). Free Mo Robinson: Citizen engagement in response to a crime event on social media. ***Social Semiotics, 31***(3), 365-382.
41. Makridis, C. A. (2021). Do data breaches damage reputation? Evidence from 45 companies between 2002 and 2018. ***Journal of Cybersecurity, 7***(1), 1-8.
42. Oh, O., Agrawal, M., & Rao, H. R. (2013). Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. ***MIS Quarterly, 37***(2), 407-426.
43. Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. ***Marketing Science, 31***(3) 521-543.
44. Rosati, P., Deeney, P., Cummins, M., Van der Werff, L., & Lynn, T. (2019). Social media and stock price reaction to data breach announcements: Evidence from US listed companies. ***Research in International Business and Finance, 47***, 458-469.
45. Sinanaj, G., Muntermann, J., & Cziesla, T. (2015). How data breaches ruin firm reputation on social media!-insights from a sentiment-based event study. ***In Proceedings of Wirtschaftsinformatik***, Osnabrueck, Germany, 61.
46. Sinanaj, G., & Zafar, H. (2016). Who wins in a data breach? - A comparative study on the intangible costs of data breach incidents. ***In proceedings of PACIS 2016***, Chiayi, Taiwan.
47. Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy: Measuring individuals' concerns about organizational practices. ***MIS Quarterly, 20***(2), 167-196.
48. Spanos, G., & Angelis, L. (2016). The impact of information security events to the stock market: A systematic literature review. ***Computers & Security, 58***, 216-229.
49. Stewart, K. A., & Segars, A. H. (2002). An empirical examination of the concern for information privacy instrument. ***Information Systems Research, 13***(1), 36-49.
50. Syed, R., & Dhillon, G. (2015). Dynamics of data breaches in online social networks: Understanding threats to organizational information security reputation. ***In proceedings of ICIS 2015***, Fort Worth, Texas, USA.
51. Syed, R. (2019). Enterprise reputation threats on social media: A case of data breach framing. ***The Journal of***

*Strategic Information Systems, 28*(3), 257-274.

52. Turjeman, D., & Feinberg, F. M. (2023). When the data are out: measuring behavioral changes following a data breach. *Marketing Science, 0*(0), forthcoming.
53. Vemprala, N., & Dietrich, G. (2019). A Social Network Analysis (SNA) study on data breach concerns over social media. *In proceedings of the 52nd Hawaii International Conference on System Sciences(HICSS)*, Hawaii, USA, 7186-7193.
54. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151.
55. Wang, T., Ulmer, J. R., & Kannan, K. (2013). The textual contents of media reports of information security breaches and profitable short-term investment opportunities. *Journal of Organizational Computing and Electronic Commerce, 23*(3), 200-223
56. Wong, N., Yan, J., & Chua, H. N. (2020). A path analysis model to identify the effects of social media, news media and data breach on data protection regulation awareness. *In proceedings of 2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 1-6.
57. Yin, J., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2012). Using social media to enhance emergency situation awareness. *IEEE intelligent systems, 27*(6), 52-59.

## [URL]

58. ITRC (2023). H1 2023 Data Breach Analysis: 2023 Data Compromises Are on a Blistering Pace to Set a New Record, https://www.idtheftcenter.org/publication/h1-2023-data-breach-analysis/

## 저 자 소 개

### 서 승 우 (Seungwoo Seo)

현재 한국정보통신진흥협회 부설 정보통신인증센터에 재직중이다. 동의대학교 컴퓨터공학과를 졸업하고, 서울시립대 경영대학원에서 경영학석사를 취득하였으며, 가톨릭대학교 경영학전공 박사 과정에 재학중이다. 주요 관심분야는 정보보안, 개인정보보호, 정보보호 및 개인정보보호 제도, 정보보안 Automation 등이다.

### 고 영 준 (Youngjoon Go)

현재 연세대학교 정보대학원 석사 과정에 재학 중이다. 가톨릭대학교 경영학과와 빅데이터인문경영융복합전공을 졸업했다. 주요 관심분야는 데이터 분석, 텍스트 마이닝, 딥러닝 등이다.

### 이 홍 주 (Hong Joo Lee)

현재 가톨릭대학교 경영학전공 교수로 재직 중이다. KAIST 산업경영학과를 졸업하고 KAIST 테크노경영대학원에서 석사 및 박사학위를 취득하였다. 주요 관심분야는 데이터 분석, 지능형 정보시스템, 온라인 사용자들의 상호작용 등이다. 지금까지 International Journal of Electronic Commerce, Expert Systems with Applications, Journal of Electronic Commerce Research, Government Information Quarterly 등 주요 학술지에 논문을 발표하였다.

## 〈 Abstract 〉

# Online Privacy Protection: An Analysis of Social Media Reactions to Data Breaches

Seungwoo Seo*, Youngjoon Go**, Hong Joo Lee***

This study analyzed the changes in social media reactions of data subjects to major personal data breach incidents in South Korea from January 2014 to October 2022. We collected a total of 1,317 posts written on Naver Blogs within a week immediately following each incident. Applying the LDA topic modeling technique to these posts, five main topics were identified: personal data breaches, hacking, information technology, etc. Analyzing the temporal changes in topic distribution, we found that immediately after a data breach incident, the proportion of topics directly mentioning the incident was the highest. However, as time passed, the proportion of mentions related indirectly to the personal data breach increased. This suggests that the attention of data subjects shifts from the specific incident to related topics over time, and interest in personal data protection also decreases. The findings of this study imply a future need for research on the changes in privacy awareness of data subjects following personal data breach incidents.

Key words: Data breach, Personal information, Social media, Text mining, Knowledge management

* swseo73@catholic.ac.kr
** wnsldud@naver.com
*** hongjoo@catholic.ac.kr