

빅데이터 분석을 통한 데이터 3법 인식에 관한 연구*

오 정 주*, 이 환 수**

요 약

산업의 디지털 전환을 촉진하고 혁신을 가속화하고자 우리나라는 디지털 뉴딜 정책을 추진하고 있다. 그러나 엄격한 기존의 데이터 관련 법제 하에서는 디지털 뉴딜정책을 위한 산업계의 데이터 활용에 여전히 제약이 있는 상황이다. 이러한 문제를 해결하기 위해서 데이터 3법 개정안이 발의되었으나 실제로 산업계의 데이터 이용 활성화에 어떠한 영향을 미칠지에 대한 논의는 아직 부족한 상황이다. 이에 본 연구에서는 데이터 3법에 대한 여론의 인식을 분석하여 데이터 3법 개정안의 시사점을 분석하고자 한다. 이를 위하여 데이터 3법 개정안과 관련 연구동향을 분석하고, 빅데이터 분석 기법을 이용하여 데이터 3법에 대한 인식을 분석하였다. 분석결과에 따르면 데이터 3법은 개정 취지에 부합하게 데이터 산업 활성화를 촉진하는 반면에 특정 산업 분야에 치중될 우려가 있는 것으로 나타났다. 본 연구의 결과는 빅데이터 분석을 통해 시행 초기인 데이터 3법의 산업 영향에 대한 온라인 인식을 분석함으로써 향후 개선방안에 대한 시사점을 제공한다라는 점에서 의의가 있다.

A Study on the Perception of Data 3 Act through Big Data Analysis

Jungjoo Oh*, Hwansoo Lee**

ABSTRACT

Korea is promoting a digital new deal policy for the digital transformation and innovation accelerating of the industry. However, because of the strict existing data-related laws, there are still restrictions on the industry's use of data for the digital new deal policy. In order to solve this issue, a revised bill of the Data 3 Act has been proposed, but there is still insufficient discussion on how it will actually affect the activation of data use in the industry. Therefore, this study aims to analyze the perception of public opinion on the Data 3 Act and the implications of the revision of the Data 3 Act. To this end, the revision of the Data 3 Act and related research trends were analyzed, and the perception of the Data 3 Act was analyzed using a big data analysis technique. According to the analysis results, while promoting the vitalization of the data industry in line with the purpose of the revision, the Data 3 Act has a concern that it focuses on specific industries. The results of this study are meaningful in providing implications for future improvement plans by analyzing online perceptions of the industrial impact of the Data 3 Act in the early stages of implementation through big data analysis.

Key words : Data 3 Act, Perception, Big Data, Text-mining

접수일(2021년 04월 29일), 수정일(2021년 06월 22일), 게재 확정일(2021년 06월 28일)

★ 이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행되었음(NRF-2018S1A5A8027174), 2021년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(P0008703, 2021년 산업혁신인재성장지원사업)

* 단국대학교 IT법학협동과정 박사과정(주저자)

** 단국대학교 융합보안학과 조교수(교신저자)

1. 서 론

데이터 경제(Data Economy)란 데이터가 혁신적인 비즈니스와 서비스 창출에 활용되어 산업 발전에 촉매 역할을 하는 경제를 말한다[1]. 데이터 경제는 2011년 데이비드 뉴먼(David Newman)의 가트너(Gartner) 보고서에서 처음 등장한 개념이며, 2014년부터 유럽 집행위원회가 데이터 경제와 함께 데이터 주도 경제(Data-driven Economy) 개념을 도입하면서 주목 받기 시작하였다. 이처럼 데이터의 가치가 높아짐에 따라 주요 선진국들은 데이터 경제로의 전환 계획을 수립하며 디지털 혁신을 거듭해 나가고 있다. 미국은 2016년 빅데이터 R&D 전략 계획을 발표하였고, EU는 2017년 데이터 경제 육성 전략을 수립하여 새로운 데이터 비즈니스를 창출하고자 하였다. 같은 해 중국도 빅데이터 산업 발전 계획을 수립하여 빅데이터 선도 기업을 육성하기 시작했다[2]. 그러나 EU의 경우에는 유럽개인정보보호법(General Data Protection Regulation; GDPR)에서 정보주체에게 데이터 통제권을 부여함으로써 데이터 경제를 활성화하면서도 정보주체의 권리보장도 같이 고려하는 방향으로 추진하고 있다는 점에서 차이가 있다[3].

우리나라는 2020년 7월 14일 ‘한국판 뉴딜 종합계획’을 발표하며 전 산업 분야의 디지털 전환을 촉진하고 산업 혁신을 가속화하고자 하였다. 디지털 뉴딜 정책은 데이터(Data)를 바탕으로 네트워크(Network), 인공지능(AI) 기술 등을 활용하여 데이터 인프라를 구축·개발·활용하고 새로운 혁신산업을 창출하는 것을 강조한다[4]. 디지털 뉴딜 정책에는 데이터 댐(Data Dam), 전 산업 5G·AI 융합 확산, 지능형 정부, 스마트 의료, 교육 인프라의 디지털 전환 등 다양한 추진과제가 포함되어 있다. 그러나 기존의 엄격한 데이터 관련 법제 하에서는 산업계가 데이터를 활용하기 어려운 상황이었다. 이에 법적 규제를 완화함으로써 데이터 이용을 활성화하여 신산업을 육성하고자 데이터 3법 개정안이 발의되었다. 데이터 3법 개정안은 지난해 1월 국회 본회의를 통과하여 8월부터 시행됨으로써 디지털 뉴딜 정책의 법적 기반이 마련되었다.

그러나 데이터 3법 개정안의 목적이 데이터 이용 활성화를 통한 신산업 육성에 있음에도 불구하고 실제

로 산업계의 데이터 이용 활성화에 어떠한 영향을 미칠지에 대한 논의는 아직 부족하다. 기존 연구들은 데이터 3법 개정안의 의미와 추후 과제에 대한 법적 연구, 특정 산업분야에서의 가명데이터 활용방안, 가명데이터의 보호방안, 보건의료데이터 이용의 쟁점 등에 대하여 질적으로 분석한 연구들이 대부분이다. 이에 본 연구에서는 데이터 3법에 대한 여론의 인식을 분석하여 데이터 3법 개정안의 시사점을 분석하고자 한다. 이를 위하여 데이터 3법 개정안과 관련 연구동향을 살펴보고, 빅데이터 분석 기법을 이용하여 데이터 3법에 대한 인식을 분석한다. 도출된 결과를 바탕으로 데이터 3법의 문제점을 분석하고 개선방향에 대하여 논의한다.

2. 이론적 논의

2.1 데이터 3법

데이터 3법이란 「개인정보보호법」, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률(이하 “정보통신망법”）」, 「신용정보의 이용 및 보호에 관한 법률(이하 “신용정보법”）」을 통칭하여 일컫는 말이다. 4차 산업혁명 시대를 맞이하여 인공지능(AI), 클라우드 컴퓨팅(Cloud Computing), 사물인터넷(IoT) 등 신기술을 이용한 산업들이 등장하면서 그 핵심자원인 데이터의 활용이 중요해졌다. 그러나 개정 전의 데이터 3법은 정보주체의 사전동의를 비롯한 강력한 규제 장치들을 통하여 데이터의 활용을 엄격히 제한해왔다. 헬스케어, 바이오, 금융 등 각종 산업 분야에서는 데이터 3법에 포함된 개인정보 활용 제한 규정들을 완화하여 정보의 보호와 활용을 균형 있게 규율하는 방향으로 개정할 것을 요구하였다[5]. 이에 데이터 이용을 활성화하여 신산업을 육성하기 위한 목적으로 데이터 3법이 개정되었고 2020년 8월 5일자로 시행되었다.

데이터 3법 개정안의 주요 내용으로는 다음과 같다. 우선 개인정보보호법, 정보통신망법, 신용정보법에 산재되어 있던 개인정보 관련 유사·중복 규정을 정비하고 체계를 일원화하였다. 개인정보보호법에서는 가명정보 개념을 도입하여 데이터 활용을 제고하였다. 가명정보란 개인을 알아볼 수 없도록 안전하게 처리한

개인정보를 말한다. 개인정보보호법 개정안에서는 가명정보의 경우 통계작성, 과학적 연구, 공익적 기록 보존 목적으로 정보주체의 동의 없이 처리를 할 수 있도록 하였다. 서로 다른 기업이 보유하고 있는 가명정보를 보안시설을 갖춘 전문기관에서 결합할 수 있도록 하여 상업적 이용가능성을 열어두었다. 또한 동의 없이 처리할 수 있는 개인정보의 범위를 확대하였고 개인정보의 범위를 명확히 하였다. 마지막으로 EU GDPR 적정성 평가의 필수 조건인 감독기구의 독립성을 확보하고자 개인정보보호위원회를 국무총리 소속의 합의제 중앙행정기관으로 격상하였다.

정보통신망법과 신용정보법 개정안에서는 각 법 내 개인정보 보호 관련 사항을 개인정보보호법으로 이관하였다. 신용정보법에서는 일정한 경우에 가명정보를 동의 없이 활용할 수 있도록 하여 금융 분야 빅데이터 분석·이용의 법적 근거를 명확화 하였다. 뿐만 아니라 금융 분야 마이데이터 산업을 도입하여, 다양한 기관에 흩어져있는 개인정보를 통합하여 조회하고 관리할 수 있도록 하였다. 한편 정보활용 동의제도 개선, 프로파일링 대응권 도입, 개인신용정보 이동권 도입, 금융권 정보보호 상시평가제, 개인신용정보 유출에 대한 징벌적 손해배상금 강화 등을 통하여 개인정보 이용 활성화에 따른 불안을 잠재우고자 하였다. 데이터 3법 개정안을 통하여 핀테크, 헬스케어, 바이오 등 데이터 기반의 신산업이 발전하고 다양한 혁신적인 서비스가 등장할 것으로 기대되고 있다[6].

2.2 데이터 3법과 산업 동향

데이터 3법은 금융, 보건의료, 광고, 교육 등 다양한 산업 분야에 영향을 미칠 것으로 예상되고 있다. 그 중 금융 분야는 데이터 3법 개정 후 가장 큰 변화가 일어나고 있는 분야이다. 우선 마이데이터 사업이 본격화되어 개별 금융기관에 분산되어 있던 정보를 하나의 플랫폼에서 정보주체가 통합하여 관리하는 것이 가능해졌다[7]. 마이데이터 사업을 통해 정보주체는 개인정보에 대한 통제권을 강화할 수 있으며, 금융회사는 개인의 소비패턴과 재무현황 등의 정보 분석이 가능해져 개인 금융 및 신용정보 관리, 맞춤형 상품 추천, 금융컨설팅 등의 새로운 서비스를 출시하였다. 또

한 금융데이터거래소가 출범하여 금융 데이터 거래가 활성화되었다. 최근에는 비금융업자도 참여하면서 금융데이터와 결합한 고부가가치 데이터의 생산이 기대되고 있는 상황이다. 한국신용정보원에서는 금융 빅데이터 개방시스템을 개시하여 이용자들에게 금융신용정보 DB와 원격분석 환경을 제공하기 시작하였다[4].

보건의료 분야도 데이터 활용 비즈니스가 활발한 영역이다. 캐나다 스타트업 '블루닷(Bluedot)'은 전세계 뉴스와 인구이동, 기후변화 등의 데이터를 수집·분석하고 있으며, 2019년 코로나19 바이러스의 발생과 확산을 예측한 바 있다. 메디블록은 블록체인 기술을 이용한 개인건강기록 플랫폼을 개발하였다. 코아제타에서는 건강보험심사평가원의 청구 데이터를 분석하여 제약사에게 제공하여 의약품을 분석하고 시각화하는 서비스를 개발하였다. 중앙정부와 지방자치단체에서도 공공데이터를 효과적으로 사용하기 위한 방안이 모색되고 있다. 서울시는 빅데이터 캠퍼스를 설립하여 공공데이터를 개방하여 다양한 서비스를 개발하고, 경기도에서는 민관 공동 데이터 플랫폼 구축사업을 추진하였다. 또한 국민들이 공공데이터를 활용하여 사회문제를 해결할 수 있도록 지원하고 있다[8].

2.3 관련 연구

데이터 3법이 개정되면서 가장 주목받는 이슈는 가명정보 이슈이다. 그 동안 정부는 비식별정보의 활용을 장려해왔으나 가명정보의 활용은 명확히 언급하지 않았다. 이번 개정안은 가명정보 활용에 대한 명확한 법적 근거를 확보했다는 점에서 의미가 있다고 평가된다[9]. 데이터 3법 개정안 통과로 의료, 금융, 광고 등 다양한 산업 분야에서 데이터 활용을 통한 고객맞춤형 서비스 및 상품 제공, 새로운 서비스 및 상품 개발 등의 기대효과가 있을 것으로 기대하고 있다.

그러나 정일영·박선호·현대원(2020)의 연구에 의하면 데이터 3법이 개정되었음에도 불구하고 추가 개정의 필요성을 강조한다[10]. 데이터 활용에 대한 기준과 방법 등이 모호하여 법 해석상의 혼란이 발생할 수 있기 때문이다[11]. 가령 개인정보보호법 제28조의2 제1항과 신용정보법 제32조 제6항 9의2호 모두 정보주체의 동의 없이 가명정보 처리를 허용하는 목적으로 통

제작성, 과학적 연구, 공익적 기록보존 등을 두고 있다. 그러나 신용정보법의 경우 상업적 목적의 통계작성과 산업적 연구를 포함한다고 명시하고 있다. 이것은 개인정보보호법과 신용정보법의 서로 다른 규정이 규제기관의 해석에 따라 가명정보를 처리가능범위가 달라질 수 있음을 시사한다[12]. 즉 데이터 3법을 통하여 데이터 활용에 대한 법적 근거는 마련되었지만, 법의 모호함으로 인하여 데이터 활용에 여전히 제약이 있다는 것이다. 이와 함께 불명확한 규정으로 인한 가명정보의 오·남용[13], 형사처벌로 이어질 경우 죄형법정주의 중 명확성의 원칙 위배[9] 등의 우려도 제기되고 있다. 이에 데이터 산업이 활성화되기 위해서는 명확한 가명처리 기준과 처리방법, 결합 프로세스, 익명화 기준 등이 필요하다는 의견이 제시되고 있다[14-16].

정보주체의 프라이버시 보호를 위해서 가명정보 활용 방안에 대해서도 활발한 연구가 진행되고 있다. 김동현과 김순석(2020)은 데이터 상황 접근법에 따라 각 상황을 국내의 환경에 따라 분류하고 전반적인 위험도를 측정·진단하고 조치 수준을 결정할 수 있도록 방법론을 제안하였다[17]. 김정선(2020)은 가명정보 활용을 위한 기업의 준비사항을 제시하였으며[15], 차건상(2020)은 가명정보 Life-Cycle별 위험분석을 실시하여 처리단계별 보호조치를 제안하였다[16]. 천지영과 노건태(2020)는 서울시 공공자전거 이용정보를 분석한 결과, 사람들은 서비스의 목적에 따라 프라이버시 침해문제에 대해 다르게 반응하는 것을 발견하였다. 이에 개인정보의 활용 목적에 따라 개인정보 처리의 수준을 다르게 할 것을 제안하였다[6].

데이터 3법은 특히 보건의료 분야에 큰 영향을 미쳐 보건의료 데이터의 활용과 보호에 관한 많은 논의가 진행되었다. 데이터 3법 개정은 보건의료 데이터의 활용으로 신약개발과 원격의료 등 보건의료 산업 발전에 기여할 것으로 기대되었다[18]. 그러나 이와 동시에 가명정보 활용 관련 규정이 모호하여 보건의료 데이터가 오·남용될 것을 우려하는 견해들이 공존하였다[19-20]. 이에 많은 연구자들이 보건의료 데이터의 특수성과 민감성을 고려하여 명확하고 구체적인 기준, 엄격한 관리, 철저한 안전조치가 필요하다고 주장했다[18-19]. 보건복지부에서는 2020년 9월 보건의료 데이

터 활용 가이드라인을 발표하였지만 이기호와 김계현(2020)은 보건의료 데이터 활용 가이드라인에는 정보주체 보호방안 부족, 개인정보 재식별에 대한 가이드 부족 등의 한계가 있다는 점을 제시하였다[19].

이외에도 데이터 3법이 미치는 영향을 분석한 다양한 연구가 진행되었다. 데이터 3법 개정이 금융회사에 미치는 영향[21], 보험업에 대한 영향[22], 창업기업에 미치는 영향[23], 광고산업에 미치는 영향[24] 등의 연구를 통하여 데이터 3법 개정이 산업계에 큰 영향을 미쳤다는 점이 확인되었다. 특히, 데이터 3법 관련 선행연구는 주로 데이터 3법에 대한 법적 분석, 특정 산업분야에서의 가명데이터 활용방안, 가명데이터 보호방안, 보건의료데이터 이용의 쟁점 등에 대하여 질적으로 분석한 연구가 대부분인 것으로 나타났다. 그러나 데이터 3법 개정안이 어떻게 인식되는지 실증적으로 분석한 연구는 부족하다. 데이터 3법에 대한 인식 분석을 통해 데이터 3법이 입법취지에 맞게 작동하는지, 그 현황과 문제점으로는 어떠한 것들이 있는지 검토할 필요가 있다.

3. 연구방법

3.1 연구대상

본 연구는 데이터 3법에 대한 여론의 인식을 살펴보고, 데이터 3법 시행 전후의 인식을 비교하는 것을 목적으로 하였다. 이를 위하여 대표적인 포털사이트인 네이버(Naver), 다음(Daum), 구글(Google)에서 제공하는 뉴스 데이터를 수집채널로 선정하였다. 자료분석을 위하여 사용한 키워드로는 ‘데이터 3법’을 사용하였다. 자료의 수집기간은 데이터 3법 시행 전후 6개월인 2020년 2월 1일부터 2021년 1월 31일까지 총 1년으로 설정하였다. 데이터 수집 결과 네이버 10,269개, 다음 9,024개, 구글 2,936개로 총 22,229개의 데이터가 수집되었다.

3.2 데이터 정제방법

데이터 분석에 앞서 텍스트를 통하여 수집된 데이터를 정제하는 과정을 거쳤다. 수집된 키워드에서 대,

등, 것 등 의미 파악이 어려운 단어와 최근, 지난, 이후 등 데이터 3법과 관련이 없는 단어를 삭제하였다. 다음으로 단어를 이루지 못한 불완전한 단어, 동일한 의미이지만 명칭이 다르게 적용된 단어들은 공통된 키워드로 정제하였다. 정제 후에는 가장 많이 등장한 상위 100개의 키워드를 선정하였다.

<표 1> 데이터 정제 예시

데이터 정제	삭제된 단어
빅데이터 (빅데이터활, 빅데이터관련, 빅데이터업무)	대, 등 것, 위, 시, 하, 통 등 의미 파악이 어려운 단어
마이데이터 (마이데이터사업, 마이데이터산업, 마이데이터시장)	최근, 지난, 이후, 올해, 이번, 내년 등 시점과 관련된 일반적 단어
개인정보보호위원회 (개보위, 개인정보위)	관련, 위하, 밝혔, 밝히, 때 문, 에서, 관하, 각종 등 데이터 3법과 관련이 없는 단어

3.3 분석도구 및 절차

텍스트마이닝이란 많은 양의 텍스트에서 정보를 자동으로 추출하여 새롭고 의미 있는 정보를 발견하는 분석기법을 말한다[25-26]. 의미 있는 정보를 추출해냄으로써 텍스트에 대한 효율적인 분석을 가능하게 한다는 장점이 있어 많은 연구에서 활용되고 있다[27]. 본 연구에서는 데이터 수집 및 분석 도구로 빅데이터 분석 솔루션인 텍스트롬(Textom) 서비스를 이용하였다. 텍스트롬을 이용하여 수집된 데이터는 네트워크 분석 도구인 UCINET 6.721을 사용하여 의미연결망 분석(SNA; Semantic Network Analysis)을 수행하였다. 의미연결망 분석이란 단어들 간의 연관성 및 상호작용의 정도를 분석하는 기법으로, 언론의 의제 설정과 담론 형성 구조를 이해하는 데 도움을 준다[28-29]. 연결 중심성(Degree Centrality)은 의미연결망 분석의 분석 지표 중 하나로, 한 노드에 직접적으로 연결되어 있는 노드의 수를 나타낸다. 연결중심성이 높은 단어는 다른 단어와의 교류가 높은 단어이므로 의제와 담론에 있어서 주도적인 역할을 하는 단어로 볼 수 있다. 마지막으로 데이터 시각화 도구인 Netdraw를 사용하여 데이터 3법 관련 단어들의 네트워크를 시각화하였으며, CONCOR 분석을 실시하여 유사성을 가진 단어 군집을 도출하였다.

<표 2> 분석 데이터 정보

구분	내용
수집범위	네이버(뉴스), 다음(뉴스), 구글(뉴스)
수집기간	2020년 2월 1일 ~ 2021년 1월 31일 (총 1년)
키워드	데이터 3법
수집데이터 수	네이버 10,269개, 다음 9,024개, 구글 2,936개 (총 22,229개)
분석도구	Textom, UCINET 6.721, Netdraw

4. 분석결과

4.1 키워드 분석결과

2020년 2월 1일부터 2021년 1월 31일(총 1년) 동안 ‘데이터 3법’을 키워드로 추출된 단어는 총 22,229개였다. 데이터 정제 과정을 거친 후 선정된 100개의 단어를 <표 3>과 같이 정리하였다.

<표 3> 추출된 단어의 빈도

단어	빈도	단어	빈도
데이터	15867	정보	1831
빅데이터	5452	정부	1784
활용	4427	인공지능	1728
마이데이터	3600	추진	1672
시행	3590	플랫폼	1584
개인정보	3519	제공	1544
사업	3462	기술	1520
개정	3302	마련	1467
디지털	3274	뉴딜	1460
기업	3057	결합	1436
통과	3032	보안	1434
산업	2940	정책	1296
서비스	2688	시스템	1289
기반	2677	혁신	1267
법	2589	지원	1264
금융	2534	시장	1262
국회	2451	대표	1192
개인정보 보호법	2321	개최	1169
가명정보	2317	활성	1137
구축	2181	국내	1078
신용정보법	2140	도입	1017
데이터 3법	2018	처리	970
개정안	1980	클라우드	928

정보통신망법	1909	의료	913
분석	1843	진행	906

데이터 3법에 대한 연결중심성 분석을 실시한 결과를 <표 4>와 같이 정리하였다.

<표 4> 추출된 단어의 연결중심성

단어	연결중심성	단어	연결중심성
디지털	0.021399038	활용	0.007598957
데이터	0.020397472	핀테크	0.007585399
서비스	0.017226808	마이데이터	0.007527011
국회	0.015536534	안전	0.007372871
개인정보 보호법	0.014793397	분야	0.007356786
구축	0.014282941	확대	0.00730623
신용정보법	0.014134481	시행	0.007270555
개인정보	0.013211417	필요	0.007263552
정보통신망법	0.012899886	플랫폼	0.007230244
추진	0.011846544	제공	0.007194505
마련	0.010539964	기술	0.007143122
빅데이터	0.010194644	네이버	0.007133349
기반	0.009809129	개인	0.007061778
정책	0.009642477	대응	0.007059136
개정	0.008807039	서울	0.006694647
시장	0.008686432	제도	0.006690407
대표	0.008602913	발표	0.006553712
활성	0.008233454	가명정보	0.006305828
국내	0.008204752	수집	0.006197781
통과	0.008169618	데이터 3법	0.005991363
도입	0.008089955	산업	0.005943735
정보	0.007860806	가능	0.005880407
진행	0.007841545	전문	0.005828679
개발	0.007743454	개정안	0.005793704
인공지능	0.007729124	신한카드	0.005771595

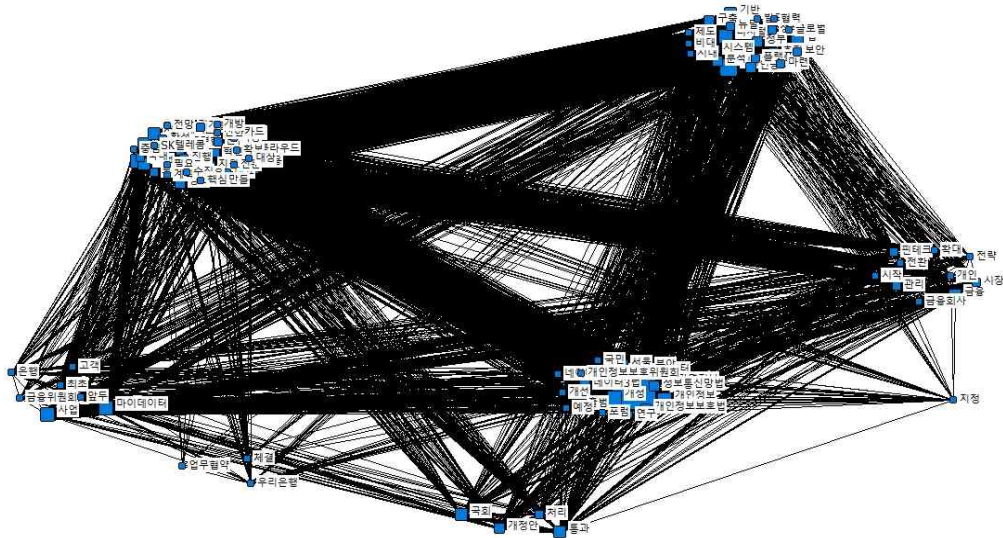
4.2 네트워크 시각화 분석결과

(그림 1)과 같이 네트워크 내 연결 관계와 패턴을 시각화 및 군집화하기 위하여 수집된 단어들을 대상으로 CONCOR 분석을 실시하였다. 시행전후 6개월간의 데이터로 분석한 결과 8개의 그룹이 형성되었다. ①

산업계에서의 데이터 활용과 관련된 단어 (SK텔레콤/ 신한카드/ 의료/ 산업/ 기술/ 수집 등), ② 기술 및 정책 관련 단어 (인공지능/ 빅데이터/ 플랫폼/ 시스템/ 제도/ 정책 등), ③ 데이터 3법 논의 관련 단어 (개인정보보호법/ 정보통신망법/ 신용정보법/ 개인정보보호위원회/ 포럼/ 출범 등), ④ 금융 관련 단어 (금융/ 금융회사/ 핀테크/ 시장 등), ⑤ 마이데이터 관련 단어 (마이데이터/ 금융위원회 / 은행 / 사업 등), ⑥ 입법부 관련 단어 (개정안/ 국회/ 처리/ 통과 등), ⑦ 업무협약 관련 단어 (우리은행/ 업무협약/ 체결 등), ⑧ 지정 등 총 8개의 그룹이 형성되었다.

다음으로 데이터 3법 시행일을 기준으로 CONCOR 분석을 추가로 실시하였다. 데이터 3법 시행 전에는 ① 데이터 활용 및 규제 관련 단어 (데이터/ 개인정보/ 가명정보/ 제공/ 결합/ 규제 등), ② 국가기관 관련 단어 (정부/ 입법/ 정책/ 제도/ 경제/ 공공데이터 등), ③ 데이터 활용 기술 관련 단어 (클라우드/ 빅데이터/ 인공지능/ 의료/ 시스템/ 산업 등), ④ 입법부 관련 단어 (본회의/ 국회/ 개정안/ 통과 등), ⑤ 마이데이터 관련 단어 (마이데이터/ 금융위원회/ 금융보안원/ 은행 등), ⑥ 금융 관련 단어 (금융/ 금융회사/ 우리은행 등), ⑦ 데이터 유통 관련 단어 (유통/ 협력/ 체결 등), ⑧ 법률 관련 단어 (정보통신망법/ 신용정보법 등) 총 8개의 그룹이 형성되었다.

데이터 3법 시행 후에는 ① 데이터산업 관련 단어 (SK텔레콤/ 신한카드/ 클라우드/ 인공지능/ 의료/ 산업 등), ② 금융 관련 단어 (금융/ 핀테크/ 마이데이터/ 기업/ 사업/ 시장 등), ③ 기술 활용 관련 단어 (빅데이터/ 플랫폼/ 디지털/ 시스템 등), ④ 개인정보 관련 단어 (데이터/ 개인정보/ 개인정보보호법/ 개인정보보호위원회 등), ⑤ 가명정보 관련 단어 (가명정보/ 정보통신망법/ 신용정보법/ 관리/ 처리/ 출범 등), ⑥ 입법부 관련 단어 (국회/ 개정안/ 의원/ 제정/ 통과 등), ⑦ 포럼 관련 단어 (포럼/ 서울/ 개최), ⑧ 업무협약 관련 단어 (업무협약/ 핵심/ 지정) 총 8개의 그룹이 형성되었다.



(그림 1) 데이터 3법 시행전후 CONCOR 분석결과

5. 논의 및 결론

본 연구는 빅데이터 분석 기법을 활용하여 데이터 3법에 대한 인식을 확인하고, 데이터 3법의 시사점과 향후 개선방향에 대하여 살펴보고자 하였다. 분석결과에 의하면 ‘데이터’, ‘빅데이터’, ‘활용’, ‘마이데이터’, ‘시행’ 등의 단어가 많이 추출되었다. 데이터 이용을 활성화함으로써 신산업을 육성하겠다는 데이터 3법 개정안의 취지에 부합하는 결과이다. 동시에 ‘보안’과 ‘안전’, ‘관리’ 등의 키워드들이 추출되었다. 이는 데이터 3법이 데이터의 활용에 중점을 두고 있지만 동시에 데이터 보호 차원도 고려하고 있음을 시사한다. 기업들은 데이터 3법이 개정되면서 안전한 데이터 활용을 위한 기반이 마련되었으나 아직은 제도적으로 불확실성이 높다고 호소하고 있다[28].

한편 추출된 키워드의 상당수가 금융 관련 키워드로 나타나 데이터 활용이 금융 분야에 편중되어 있다는 점을 확인하였다. 손정연(2021)의 연구에서도 금융위원회, 금융감독원, 금융보안원 등 금융 관련 기관의 의제형성과 정책 조정에 타 산업분야 기관에 비해 높은 빈도로 언론기사에 출현한 것으로 나타났다. 이는 데이터 3법과 관련하여 의제가 형성되고 정책이 조

정되는 시기에 금융 관련 기관이 활발하게 참여했기 때문인 것으로 보인다[29].

데이터 3법 시행 전과 후 6개월의 키워드에 대하여 CONCOR 분석을 실시한 결과, 산업계에서의 데이터 활용과 관련된 단어들이 주요 군집을 이루고 있음을 확인할 수 있었다. 제도적 인프라(법, 정책, 시스템 등)와 기술적 인프라(빅데이터, 플랫폼, 인공지능, 보안 등)가 하나의 군집을 이루었다. 해당 군집은 데이터 3법 관련 사업인 ‘데이터 댐’ 또는 ‘마이데이터’ 등의 시행에 필요한 인프라에 대한 것으로 판단된다. 정부에서 디지털뉴딜의 핵심과제 중 하나로 제시된 ‘데이터 댐’은 국가 주요 인프라에 인공지능과 빅데이터 등의 기술을 접목하여 스마트 관리체계를 구축하는 사업이다. 한편 마이데이터 사업에서 데이터와 플랫폼, 인공지능 경쟁력은 마이데이터 산업에서의 성과를 좌우하는 요소로 여겨진다[30]. 한국데이터산업진흥원에 의하면 2019년 9월부터 2020년 9월까지의 뉴스데이터를 분석한 결과 기업들의 데이터센터 확보 경쟁이 심화되었다[31]. 데이터산업의 핵심원료인 방대한 양의 데이터를 보관하기 위해 인프라를 구축하고 있음을 확인할 수 있다. 기업의 데이터 활용 계획 관련 군집의 등장은 데이터 3법 개정에 따라 기업의 데이터 활용 계획에 대한 보도가 증가하였음을 시사한다. 산업 및 기업

명(SK텔레콤, 신한카드, 의료 등)과 데이터의 활용 관련 단어(수집, 결합, 제공 등), 기업의 계획 관련 단어(계획, 도입, 활성화 등)가 함께 군집화되었다. 다음으로 마이데이터 사업과 금융 관련 군집이 각각 등장하였다. 마이데이터 사업 관련 군집에서 특징적인 것은 금융위원회와 은행과 같은 금융 관련 키워드가 함께 군집화되었다는 점이다. 마이데이터 산업은 금융 분야에서 먼저 활성화되고 있음을 시사한다. 이와 관련하여 마이데이터 산업은 금융, 통신, 의료 등 다양한 분야에서 시행될 수 있으나 금융 산업에 가장 먼저 변화를 일으킬 것으로 기대된다는 견해가 있다[30]. 한편 데이터 3법에 해당하는 법률들과 ‘개인정보보호위원회’, ‘포럼’, ‘출범’, ‘국민’, ‘개정’ 등의 단어들이 군집을 이루었다. 데이터 3법의 개정에 따라 개인정보보호위원회가 출범하였고 관련 논의를 위하여 국민과 전문가들의 의견을 수렴하는 과정이 진행되었음을 확인할 수 있다.

다음으로 데이터 3법의 시행 전과 시행 후의 군집을 비교해 본 결과 몇 가지 시사점이 확인되었다. 우선 데이터 3법 개정의 영향을 가장 많이 받은 산업분야는 바로 금융 분야라는 점이다. 데이터 3법 시행 전에는 마이데이터 사업 관련 키워드(마이데이터, 은행, 금융위원회, 금융보안원 등)와 금융회사 관련 키워드(금융회사, 금융, 우리은행 등)가 각각의 군집을 이루었다. 시행 후에는 ‘핀테크’, ‘마이데이터’, ‘금융’ 등의 금융데이터 사업 관련 키워드가 군집화되었다. 시행 전에는 2개, 시행 후에는 1개의 별도 군집이 형성된 것으로 보아 데이터 3법 시행으로 금융기업의 마이데이터 사업이 활발하게 진행되고 있음을 시사한다. 그러나 금융 이외의 분야는 별도의 군집을 형성하지 않아 데이터 활용이 다소 미진한 것으로 보인다.

두 번째로 데이터 3법에 따라 각광을 받고 있는 연관 기술에 대하여 산업계의 활용 계획을 추측할 수 있다. 데이터 3법 시행 전에는 ‘빅데이터’, ‘클라우드’, ‘인공지능’, ‘플랫폼’ 등의 기술 관련 키워드와 ‘디지털’, ‘혁신’, ‘성장’, ‘뉴딜’, ‘전환’ 등의 키워드가 군집화된 것으로 보아, 데이터 3법 시행을 앞두고 데이터 기술 활용에 대한 기대감을 확인할 수 있었다. 데이터 3법 시행 후에는 구체적인 기업명과 제도적·기술적 인프라 관련 키워드, 연관 기술, 향후 계획 관련 키워드가 군

집을 이루어 산업계의 데이터 활용 계획을 추정할 수 있다.

마지막으로 데이터 3법 시행 전후로 정부 중심에서 민간으로 확대되어 논의되고 있다는 점이다. 데이터 3법의 시행 전에는 ‘정부’, ‘공공데이터’, ‘정책’, ‘제도’, ‘입법’ 등의 단어들이 하나의 군집을 이루었다. 정부는 공공데이터 개방을 확대함으로써 데이터산업의 활성화를 촉진하고 있다. 2018년까지 2.8만개, 2019년 4.9만개, 2020년 9.8만개의 공공데이터가 개방되었고 2021년에는 14.2만개가 개방될 예정이다[28]. 데이터 3법 시행 후에는 정부 관련 군집이 사라지고 금융기관과 데이터 산업 관련 키워드 군집이 등장하였다. 수요자가 원하는 데이터를 제공함으로써 민간의 참여를 확대하겠다는 정부의 방향과 부합하는 결과이다[28]. 한편 ‘활성’, ‘수집’, ‘개방’, ‘안전’, ‘지원’ 등의 키워드가 함께 군집을 이루는 것은 공공데이터 개방을 통한 데이터 활용 활성화와 함께 안전한 처리를 고려하고 있음을 시사한다.

본 연구는 빅데이터 분석을 통하여 데이터 3법에 대한 인식을 확인하였다. 연구 결과 데이터 3법은 개정 취지에 부합하게 데이터 산업 활성화를 촉진하는 것으로 나타났다. 하지만 특정 산업 분야에 치중될 우려가 있음이 분석 결과에서 나타났다. 디지털 뉴딜의 과제가 스마트 의료 및 돌봄, 스마트 물류체계, 디지털 기반 교육 등 다양한 산업 분야를 대상으로 하고 있다는 점을 고려하면 아쉬움이 남는 결과이다. 이는 데이터 활용에 대한 기준과 방법이 아직 모호하여 다른 산업 분야에서는 데이터 활용을 주저하는 것으로 판단된다. 한편 보건의료 분야는 관련 연구가 다수 수행된 분야이지만 별도의 군집을 형성하지는 않았다. 국내 보건의료 분야에서 데이터 이용이 활성화될 경우 성장 잠재력이 높지만 그에 따른 우려도 많으므로 신중히 검토하기 때문인 것으로 판단된다. 다수의 연구에서 보건의료 데이터의 민감성과 특수성을 고려하여 명확한 활용 기준 및 안전장치가 필요하다는 주장이 제기되고 있기 때문이다. 만약 보건의료 데이터 활용의 명확한 기준과 안전장치가 마련된다면, 높은 품질의 공공데이터와 IT 의료 인프라를 바탕으로 보건의료 산업 강국이 될 것으로 보인다. 한편 이번 데이터 3법 개정이 이용자의 개인정보 활용에 대한 규제를 완화하

는 법안입에도 불구하고 이용자 관점의 논의는 부족하다는 한계가 있다. 법률체계상의 개인정보 유형과 이용자가 실제로 인식하는 개인정보 유형 간에 분명한 차이가 존재하는 만큼 이용자의 검토와 논의가 이루어질 필요가 있다.

본 연구는 데이터 3법에 대한 문헌 연구와 실증 분석을 통하여 데이터 3법에 대한 인식과 시사점, 한계점을 발견하였다는 점에서 학술적 의의가 있다. 데이터 3법이 입법 취지에 부합하게 작용하기 위해서는 산업분야별 특성을 고려한 명확한 데이터 활용 기준과 방법이 보완되어야 한다는 점을 실증적으로 확인하였다는 점에서도 의미가 있다. 이는 데이터 3법 재개정 시 한계점 보완에 도움이 될 것으로 기대한다. 그러나 본 연구는 데이터 3법의 한계점에 대하여 구체적인 개선방안은 제시하지 못했다는 점에서 한계가 있다. 본 연구는 뉴스 빅데이터를 바탕으로 데이터 3법에 대한 인식 분석에 그쳐 전문가들의 의견을 포함하지 못했다는 점에서도 한계가 있다. 추후 연구에서 실제 전문가들의 견해와 여론의 차이를 분석하여, 이를 기반으로 한 개선방안 연구도 필요할 것으로 사료된다.

참고문헌

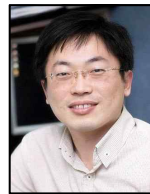
- [1] 한국정보화진흥원, “데이터 경제의 부상과 사회경제적 영향”, IT & Future Strategy, 제7호, 2018.
- [2] 관계부처 합동, “데이터 산업 활성화 전략 -I-Korea 4.0 데이터 분야 계획, I-DATA-”, 4차산업혁명위원회, 2018.
- [3] 한국데이터산업진흥원, “IMF 데이터 경제학의 주요 내용 및 시사점”, 데이터산업 동향 이슈 브리프, 제1호, 2020.
- [4] 한국데이터산업진흥원, “데이터산업 백서”, 통권 제23호, 2020.
- [5] 정현승, 박선호, 현대원, “빅데이터 산업 활성화를 위한 정책추진 과정의 우선순위 분석 : AHP를 활용한 정책 중요도 분석을 중심으로”, 한국방송학보, 제35권, 제1호, pp.283-313, 2021.
- [6] 천지영, 노건태, “데이터 3법 시대의 익명화된 데이터 활용에 대한 제언”, 정보보호학회논문지, 제30권 제3호, pp.503-512, 2020.
- [7] 백한중, “금융분야에서 마이데이터 서비스의 수용의도에 영향을 미치는 요인에 관한 연구”, 박사학위, 숭실대학교 대학원, 2019.
- [8] 배영임, 신혜리, “데이터 3법, 데이터경제의 시작”, 이슈&진단, 제405호, pp.1-26, 2020.
- [9] 김송옥, “가명정보의 안전한 처리와 합리적 이용을 위한 균형점 -데이터 3법에 대한 헌법적 평가를 곁다하여-”, 공법연구, 제49권, 제2호, pp.371-407, 2020.
- [10] 정일영, 이관호, 목은지, “데이터 산업의 이해관계자 기반 규제 이슈 분석 및 대응방안”, STEPI Insight, 제248호, 2020.
- [11] 김영국, “개정 데이터 3법과 보험업의 과제 - 디지털 헬스케어서비스 활성화를 중심으로 -”, 보험법연구, 제14권, 제1호, pp.495-524, 2020.
- [12] 김정수, “지능정보사회의 입법과 기본권 보호의 상관관계에 대한 소고 -‘데이터 3법’ 관련 개인정보 보호 논의를 중심으로-”, 공법학연구, 제21권, 제4호, pp.339-367, 2020.
- [13] 김혜진, “데이터 3법이 치매 조기 예측을 위한 의료 빅데이터 활용에 미치는 영향 연구”, 디지털융복합연구, 제18권, 제5호, pp.305-315, 2020.
- [14] 김서안, “데이터 3법 개정의 의미와 추후 과제”, 융합보안논문지, 제20권, 제2호, pp.59-68, 2020.
- [15] 김정선, “가명 데이터 활용연구 - 기술적 처리방법 및 기업의 활용방향을 중심으로 -”, 정보보호학회 논문지, 제30권, 제2호, pp.253-261, 2020.
- [16] 차건상, “가명정보 Life-Cycle에 대한 위험 분석을 통한 관리적/기술적 보호조치 방안에 대한 연구”, 융합보안논문지, 제20권, 제5호, pp.53-63, 2020.
- [17] 김동현, 김순석, “개인정보 비식별 조치를 위한 데이터 상황 기반의 위험도 측정에 관한 새로운 방법”, 정보보호학회논문지, 제30권, 제4호, pp.719-734, 2020.
- [18] 최경진 “보건의료 공공데이터의 안전한 활용 및 정보주체 보호를 위한 과제”, 정책동향, 제14권, 제3호, pp.21-28, 2020.
- [19] 이기호, 김계현, “보건의료 데이터 활용 가이드라인의 내용과 한계에 관한 연구”, 과학기술법연구, 제26권, 제4호, pp.89-118, 2020.

- [20] 이양복, “데이터 3법의 분석과 향후과제”, 비교사법 제27권, 제2호, pp.423-463, 2020.
- [21] 이성남, “데이터(DATA) 3법의 개정이 금융실무에 미치는 영향과 법적 개선방안에 관한 연구”, 비교사법, 제28권, 제1호, pp.81-133, 2021.
- [22] 유주선, “데이터 3법 개정과 보험업에 관한 법적 연구”, 상사판례연구, 제33권, 제3호, pp.157-189, 2020.
- [23] 송인방, 나병민, “데이터 3법 개정 시행에 따른 창업 환경의 변화”, 법학연구, 제20권, 제4호, pp.331-357, 2020.
- [24] 김동성, 정해원, “데이터 3법 개정에 의한 광고계의 변화와 준비 : 전문가 심층인터뷰를 중심으로”, 한국소통학보, 제20권, 제1호, pp.123-159, 2021.
- [25] I. H. Witten, “Text Mining”, 2004.
- [26] V. Gupta and G. S. Lehal, “A Survey of Text Mining Techniques and Applications”, Journal of emerging technologies in web intelligence, Vol. 1, No. 1, pp.60-76, 2009.
- [27] JISC, “The Value and Benefits of Text Mining”, 2012.
- [28] 4차산업혁명위원회, “4차산업혁명위원회 제21차 회의”, 2021.2.17.
- [29] 손정연, “개인정보보호 관련 법안 개정과정 분석: 정책네트워크 관점에서”, 석사학위, 서울대학교 행정대학원, 2021.
- [30] 김상윤, “마이데이터 시대의 개막과 기업의 경쟁력”, 월간 SW 중심사회, 통권 제83호, 2021.
- [31] 한국데이터산업진흥원, “Data economy global news trends in korea”, 제1권, 제7호, 2020.
- [32] 황서이, 심지원, “국내 언론에 나타난 ‘스마트시티’ 이슈의 의미연결망분석”, 한국디지털콘텐츠학회 논문지, 제21권, 제5호, pp.941-950, 2020.
- [33] 이송이, 김성원, 이환수, “빅데이터 분석을 통한 프라이버시 인식에 관한 연구”, 디지털융복합연구, 제17권, 제10호, pp.49-58, 2019.

[저 자 소 개]



오 정 주 (Jungioo Oh)
2017년 2월 단국대학교 법학과 학사
2019년 2월 단국대학교 IT법학협동과정 석사
2020년 ~ 단국대학교 IT법학협동과정 박사과정
email : winteroot@gmail.com



이 환 수 (Hwansoo Lee)
2005년 2월 연세대학교 산업정보시스템 공학과 석사
2014년 2월 KAIST 기술경영학과 박사
2017년 ~ 단국대학교 융합보안학과 교수
email : hanslee992@gmail.com