

Covid-19에 따른 글로벌 창업 트렌드 분석: Crunchbase를 중심으로*

김신호 (서울과학기술대학교 데이터사이언스학과 석사과정)**

금영정 (서울과학기술대학교 산업공학과 부교수)***

국 문 요 약

전 세계적으로 유례없는 Covid-19 팬데믹으로 인해 기업의 비즈니스 방식 및 내용이 크게 변화했다. 따라서 Covid-19 이후 산업 전반에서 일어나고 있는 혁신 양상을 면밀히 모니터링하고, 이를 바탕으로 미래 비즈니스를 기획/설계하는 것이 산업현장에서 크게 요구되고 있다. 다수 연구에서 팬데믹 이후 비즈니스 동향 분석이 시도됐으나 특정 산업에 국한된 분석을 수행하는 점, 설문 및 문헌 연구 중심으로 수행되어 객관적 데이터의 활용이 부족한 점에서 한계점이 존재하고 있다. 이에 본 연구에서는 글로벌 스타트업 데이터인 Crunchbase를 활용하여 Covid-19 이후 비즈니스 산업의 트렌드를 분석하고자 한다. 트렌드 분석을 위해 Crunchbase에서 2018년부터 2021년까지 2년 단위로 데이터를 수집 및 전처리를 진행하였다. 산업 변화를 확인하기 위해 네트워크 분석, LDA 기반 토픽 분석, Doc2vec 클러스터링 분석을 결합한 접근 방식을 사용하여 비즈니스 콘텐츠의 변화를 분석하였다. 연구 결과 각 분야에서 비대면/온라인 기술들이 훨씬 전문성을 갖추며 성장하고 있으며, 기술융합에 초점을 두고 많은 산업들이 성장하고 있는 것으로 파악되었다. 본 연구는 Covid-19의 영향으로 기업 투자자 및 예비 창업자들에게 급속도로 변화하는 산업의 흐름을 파악할 수 있게 하고 투자 의사결정에 많은 도움을 줄 것으로 기대된다.

핵심주제어: Covid-19, 비즈니스 트렌드, 트렌드 분석, 창업 동향, 산업 동향, Crunchbase

1. 서론

Covid-19는 전 세계에서 유례없는 팬데믹(Pandemic)으로 국내외 산업 및 생활 전반에 큰 변화를 가져왔다(문명주·고인곤, 2022). 많은 기업에서 오프라인에서 온라인 중심으로 고객 채널을 확장하였을 뿐 아니라 비대면 비즈니스가 크게 성장하는 등 다양한 산업 분야에 Covid-19로 인한 혁신이 일어나고 있다.

이렇듯 Covid-19의 영향이 사회 전반에 크게 작용하는 상황에서 많은 기업이 비즈니스 방식 및 내용을 크게 변화시켜 왔으며, 새로운 비즈니스 접근법에 대한 지속적인 고민이 요구되고 있다(이완형, 2020). 비대면/온라인 산업의 확장, 재택 근무의 확산, 가상세계의 확산 등과 같이 일반적으로 잘 알려진 동향 뿐 아니라 Covid-19 이후 산업 전반에서 일어나고 있는 혁신 양상을 면밀히 모니터링하고, 이를 바탕으로 미래 비즈니스를 기획/설계하는 것이 산업현장에서 크게 요구되고 있다.

이러한 이유로 다수 연구에서 코로나 팬데믹 이후 특정 산업의 사회적 패턴을 포착하기 위한 비즈니스 동향 분석이 활

발하게 수행됐다(송홍규, 2022; 현보람·김경식, 2022; Yang et al., 2022; Persada et al., 2022). 비즈니스 동향 분석은 미래를 예측하거나 과거의 불확실한 현상을 확인하기 위해 사용되며(서은경, 2010), 새로운 변화에 대한 전략을 수립하는데 효과적인 방법으로 알려져 있다(김상겸·장성용, 2016).

Covid-19 전후의 산업 변화를 분석하는 연구들은 주로 특정 산업에 대한 분석이 중심을 이루고 있으며 특히 Covid-19의 영향을 많이 받은 의료(Spoorthy et al., 2020; 최연석 2020; 윤경재 2021), 교육(이용상·신동광, 2020; Daniel, 2020; 윤옥한, 2020; 김경근, 2020; 김은미 외, 2022), 서비스 산업(김정오, 2022; 신용재, 2020; 최상주·조주은, 2021; 김광중, 2021; 문명주·고인곤, 2022; Jung, 2022) 등에서 연구가 활발하게 이루어져 왔다.

이처럼 특정 산업 분야에 대한 Covid-19의 영향을 분석하는 동향 분석 연구는 다수 수행됐으나, 기존 연구들은 다음과 같은 한계점이 있다. 먼저 분석 범위의 측면에서, 대부분 기존 연구들은 특정 산업 분야에 국한되어 Covid-19 전후 분석을 수행한 경우가 대다수이다. 이는 분명 특정 산업의 동향을 면

* 본 연구는 한국연구재단 중견연구 (NRF-2023R1A2C1004752)의 지원을 받아 수행되었으며 2023년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥회의의 지원을 받아 수행되었음 (P0017123, 2023년 산업혁신인재성장지원사업).

** 주저자, 서울과학기술대학교 일반대학원 데이터사이언스학과, shk240@seoultech.ac.kr

*** 교신저자, 서울과학기술대학교 산업공학과 부교수, yjgeum@seoultech.ac.kr

· 투고일: 2023-05-08 · 1차 수정일: 2023-06-10 · 2차 수정일: 2023-06-19 · 게재확정일: 2023-06-28

밀히 분석한다는 점에서 충분한 의의가 있으나 거시적 관점에서 산업의 신규 유망 분야, 쇠퇴 분야 등을 파악하고 새로운 비즈니스의 성장을 추적하기에는 어려움이 존재한다. 따라서 산업 전체를 포괄할 수 있도록 거시적 측면에서의 체계적 분석이 요구된다. 산업 전반을 대상으로 한 거시적 분석은 산업 전체의 흐름 및 혁신 방향을 제시할 뿐 아니라, 개별 기업에도 혁신 방향에 대한 단서를 제공하고 벤치마킹을 할 수 있다는 점에서 매우 중요한 활동이기 때문이다.

다음은 분석 데이터의 측면에서, 선행 연구들의 경우 대부분 설문 조사, 문헌 자료를 통해 연구를 진행한 경우가 많은 비중을 차지하고 있다(정민정·최규완, 2021; 김지선·이형룡, 2022). 설문조사의 경우 응답자 선택에 따른 편차가 있어 분석 결과에 대한 신뢰성을 보장하기 어려우며(이채운, 2007), 비즈니스 산업 전체의 변화를 파악하기 어렵다는 한계점도 있다. 따라서 실질적이면서 Covid-19 전후 동향을 면밀하게 파악할만한 데이터 원천을 발굴하고 이를 활용할 필요가 있으며, 특히 Covid-19 전후로 크게 변화된 기업 현장을 명확하게 반영할 수 있는 기업 수준의 데이터가 요구된다.

마지막으로 분석 과정 측면에서 이전의 선행 연구를 살펴보면 연구자의 주관성을 반영하여 동향을 분석하는 경우가 대다수이며(박준태, 2022), 이는 연구의 객관성 확보에 어려움이 있으며 정확한 분석이 수행되기 힘든 한계점이 있다. 동향 분석에 있어 연구의 객관성을 확보하는 것은 사회적 현상을 쉽게 파악하고 연구의 주요 개념 및 관계성을 통해 연구주체의 방향성과 미래 예측에 도움을 주기 때문에 상당히 중요한 요인으로 작용한다(오상은 외, 2020). 따라서 연구의 객관성을 확보할 수 있으며, 제공된 데이터에서 의미 있는 인사이트를 추출하여 정확한 비즈니스 동향 분석을 수행할 수 있는 연구가 요구된다.

정리하면 현재 비즈니스 동향 분석에 관한 기존 연구의 경우 특정 산업 분야에 대해서만 국한된 분석을 수행하는 점, 분석을 수행하는 데이터에 대한 신뢰성이 부족한 점, 연구의 객관성 확보가 어려운 점들이 한계점으로 보인다. 이에 본 논문에서는 글로벌 스타트업 데이터인 Crunchbase를 활용하여 Covid-19 이후 비즈니스 산업의 변화를 분석하고자 한다. Crunchbase는 전 산업을 대상으로 글로벌 스타트업의 정보가 업데이트되는 대규모 플랫폼으로, 비즈니스 현장의 변화를 추적·관찰이 가능한 이점이 있다. 비즈니스 환경은 기본적으로 비즈니스 및 산업을 구성하는 기업의 생태계이기 때문에 Crunchbase를 통한 기업의 창업 특성 파악은 산업 변화를 이해하는 데 큰 도움이 될 수 있다. 특히 Covid-19 전후로 새로운 패러다임 전환에 적응하면서 이를 새로운 창업의 기회로 인식하는 계기가 관측되는 등 창업 현황은 Covid-19 전후 산업변화를 파악하는 데 도움을 줄 수 있다(박철우·강경란, 2020).

특히 Crunchbase는 각 기업이 관련된 산업 특성을 카테고리 형식으로 파악할 수 있을 뿐 아니라, 기업 설명을 통해 기업의 서비스 및 비즈니스 특성을 파악할 수 있다. 이외에도 창

업 시기, 창업 지역, 임직원 수 등과 같은 부가적 정보를 함께 포함하고 있어 다각적 분석을 하기 적합한 것으로 알려져 있다. 특히 스타트업 창업은 해당 시점에서의 산업의 인기도 및 매력도와 크게 관련되어 있으므로 Covid-19 전후 스타트업 창업 현황을 살펴보면 각 시점에서의 거시적 동향을 효과적으로 파악할 수 있다는 장점이 있다. 이러한 이유로 본 연구에서는 Crunchbase 데이터를 바탕으로 한 창업 트렌드 분석을 통해 산업 전반의 트렌드 변화를 분석하고자 한다.

II. 이론적 배경

2.1. 산업별 동향 연구 현황

Covid-19가 유행하면서 팬데믹 이후 산업의 변화를 파악하기 위한 많은 동향 분석 연구가 시도됐다. 전 세계적으로 팬데믹의 영향을 가장 크게 받은 산업은 서비스, 의학, 교육 분야가 대표적이며(He et al., 2020b; 김영환 외, 2015) 이들 산업에서 동향 분석을 수행할 때는 팬데믹 이전과 이후로 나누어 각 산업에 대한 분석이 진행됐다. 이를 통해 해당 산업의 회복 가능한 시나리오를 제공하거나(윤문길·장윤희, 2020; 박혜림·이수정, 2022; 이소영 외, 2022; 윤승환 2021), 팬데믹이 산업에 미친 영향을 파악하는 연구가 주로 수행되어 왔다(권영환, 2022; 조대형·김정주, 2020; 최동현 외, 2022; He et al., 2020a). 이처럼 산업별로 동향을 분석하는 연구가 광범위하기에 Covid-19와 관련된 연구 현황을 제대로 파악하기 위해서는 현재 존재하는 선행연구를 정리하고 고찰하는 단계가 필요하다. 따라서 본 절에서는 이전에 언급한 서비스, 의학, 교육산업을 그 예시로 들어 분석 방법과 목적 관점에서 연구 동향을 확인하고자 한다.

먼저 서비스 산업의 경우 송민정 외(2022), 조근위·박현정(2022), 조주은(2022),에서는 주로 설문조사, 문헌자료, 인터뷰 자료 등을 수집 후 ANOVA, 이중차이, 집중지수, 회귀 등을 통한 통계분석과 질적·양적 연구분석을 시행하여 Covid-19 이후 서비스 산업시장의 경제 동향을 파악하여 서비스 산업 종사자들의 경제·심리적 영향에 관해 분석하는 연구가 주로 진행되었다. 서비스 산업은 경제와 가장 밀접한 연관이 있는 산업이기에 수치를 활용한 통계분석이 주를 이루는 것으로 사료된다.

다음은 의료 산업으로 뉴스, 문헌, 설문조사 등의 자료를 수집하여 통계분석, 네트워크 분석, 토픽 모델링, 델파이 기법을 통해 원격의료(최연석, 2020), 백신·예방접종(최보운 외, 2022), 의료기관 종사자 또는 일반인의 심리 동향을 분석(박혜림·이수정, 2022; 전소라·김호영, 2021)하는 연구가 진행됐다. 특히 토픽 모델링과 네트워크 분석을 활용한 연구가 활발하게 수행되고 있다. 네트워크 분석은 보통 키워드 기반의 언어 네트워크 분석으로 단어 간의 맥락 정보를 효과적으로 분석 가능하며(함승경 외, 2021) 토픽 모델링은 토픽에 대한 양상과 쟁

점, 트렌드 등을 비교 분석할 수 있다는 이점이 있다(박민영 외, 2022).

교육산업에서는 주로 문헌 자료, 뉴스 기사를 수집하여 계량서지정보, 네트워크 분석, 토픽 모델링 방법이 활용되었다. 이를 통해 Covid-19가 교육산업에 미친 영향과 Covid-19의 장기화로 인한 교육방식의 변화 등을 관찰 및 고찰하는 연구가 진행되었으며(김건우 외, 2020; 김상미, 2020; 김선희, 2021; 조성범·한송이, 2022) 주로 토픽 모델링, 네트워크 분석 등의 방법들이 동향 연구에 사용되었다. 그리고 교육산업은 동향 분석을 통해 앞으로 나아가야 할 교육 방법과 지점을 제시한다는 점에서 다른 산업과의 차이를 보인다.

앞선 논의처럼 Covid-19 이후 각 산업에 대한 동향 분석 연구는 많은 관점에서 다양한 방법으로 수행됐으며, 특정 산업에 대한 Covid-19 전후 변화를 잘 설명하고 있음을 확인할 수 있다. 그러나 Covid-19 대부분의 동향 연구는 한 가지 산업에 대한 분석은 성공적으로 수행한 연구가 많으나 특정 산업이 아닌 산업 전체의 관점에서 수행된 사례 연구는 거의 수행되지 않고 있다. 이는 산업 전체 동향을 파악하기 위한 데이터 수집의 어려움에 기인한 것으로 보인다. 따라서 특정 산업뿐 아니라 전 산업에 걸친 동향을 파악할 수 있는 객관적이고 공신력 있는 데이터의 수집 및 분석을 통해 Covid-19 전후 산업의 변화를 파악하고 이를 바탕으로 혁신 전략을 수립하는 것이 요구된다.

2.2. Crunchbase기반 동향 연구

본 연구는 산업 동향 분석을 위해 Crunchbase 데이터를 활용하여 분석하였으며, 특히 산업군(Industry group), 산업(Industry), 요약 설명(Description), 상세 설명(Full Description)을 활용하여 분석하였다.

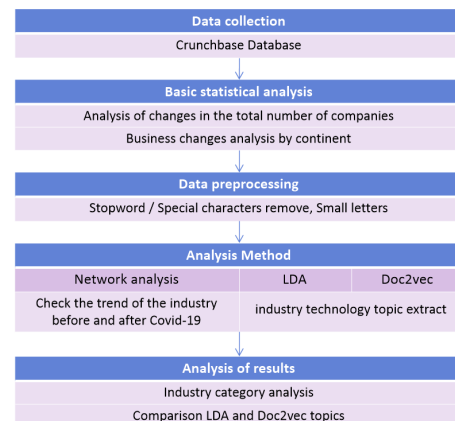
Crunchbase는 전 세계의 스타트업 및 기업 정보를 제공하고 있는 포털로 이를 이용한 다양한 연구가 수행되어 왔다. Dalle et al.(2017)의 연구에서는 경제 및 경영 전반에서의 Crunchbase의 역할에 주목하여 Crunchbase의 주요 콘텐츠를 리뷰하고 이들의 경영학적 활용 가능성에 초점을 맞추었다. Liang & Yuan(2016)의 연구에서는 Crunchbase 내의 투자자-기업 네트워크를 바탕으로 미래의 기업투자를 예측하는 연구가 수행되기도 하였으며, Lee & Geum(2023)의 연구에서는 Crunchbase 내의 M&A 정보를 바탕으로 이를 유형화하는 연구가 수행되었다.

특히 Crunchbase의 산업 및 요약 설명 데이터의 경우 기업들의 혁신 양상을 분석하는 데 활용됐다. Kim et al.(2023)의 연구에서는 Crunchbase의 산업 및 요약 설명 데이터를 바탕으로 산업융합 양상을 분석한 바 있으며, Savin et al.(2023) 역시 Crunchbase를 통한 토픽 분석을 수행하고 이를 산업 카테고리 데이터와 비교 분석을 시행하였다.

요약 설명 속성 역시 다수 연구에서 기업의 혁신 양상을 분

석하는 데 활용됐다. 일부 연구에서는 Crunchbase의 기업을 대상으로 각 요약 설명을 활용하여 Industry Classification을 수행하는 연구가 수행됐다(Batista & Carvalho, 2015; Felgueiras et al., 2020). Savin et al.(2023)의 연구에서도 스타트업 기업의 글로벌 트렌드 분석을 위해 요약 설명 데이터를 바탕으로 토픽 분석을 시행하였다.

III. 연구 방법론



<그림 1> 연구 프레임워크

본 연구의 순서는 <그림 1>을 따른다. 먼저 Crunchbase에서 2년 단위로 18년부터 21년까지 데이터를 수집하고 이를 바탕으로 정량적 분석을 수행한다. 그리고 네트워크 분석, 토픽 모델링, Doc2vec을 통해 산업·기술의 흐름 분석을 진행 후 이를 종합하여 Covid-19이후 전체 산업의 동향을 최종 분석한다.

먼저 거시적 측면에서 산업의 동향을 다각적으로 확인하기 위해 산업군 및 산업 속성을 통해 변화와 흐름을 관찰한다. 산업군 속성은 상위 분류에 관한 데이터를, 산업 속성은 산업군 하위 분류에 관한 데이터를 포함하고 있다. 따라서 산업군 상위, 하위 분류 데이터를 활용하여 네트워크 분석을 시행한다. 네트워크 분석은 각 산업군이 다른 산업군과 어떤 관계를 맺고 있는지 그 관계를 시각화하기 때문에 거시적 측면에서의 산업 동향을 분석하는 데 유용하다.

다음으로 미시적 측면에서 각 창업기업의 특성이 어떻게 변화했는지 파악하기 위해 상세 설명 속성을 활용한다. 상세 설명은 기업에 대한 소개를 상세하게 기술한 속성으로 기업 설명을 비롯한 해당 기업이 소유한 기술적 특성이 반영되어 있어 산업별 기술 콘텐츠를 식별하는데 용이하다. 본 연구에서는 상세 설명을 LDA 토픽 모델링을 통해 분석하며, 이는 앞서 산업 분류체계가 파악하지 못한 미시적 흐름을 토픽 수준에서 파악하기 위해 필요하다.

또한 상세 설명을 바탕으로 Doc2vec을 통해 문서 임베딩을 거친 후 클러스터를 추출한 후 토픽 모델링 결과와 비교함으

로써 토픽 수준이 아닌 키워드 수준에서 Covid-19에 따른 기술 토픽의 변화를 살펴본다.

요약하면 네트워크 분석을 통해서 전체 산업의 변화를, 토픽 모델링과 Doc2vec 과정을 통해 산업들의 기술토픽 변화를 비교 분석한 결과를 종합하여 Covid-19 전후 산업 변화를 비교 분석하고자 한다.

3.1. 데이터 수집

Crunchbase(<https://www.crunchbase.com/>)에서 데이터를 수집하였다. Covid-19 이전은 2018~2019년 설립된 기업 데이터로 총 75,881개, Covid-19 이후는 Covid-19가 일어난 2020년 이후 데이터로 2020~2021년간 총 30,382개의 기업 정보를 수집하였다. 데이터 속성은 기업 이름, 설명, 산업 분류, 본사 위치, 펀딩 수 등 스타트업 설립과 관련된 총 102개의 속성으로 구성된다. Crunchbase 는 투자, 자금조달 및 경영진 정보, 인수합병 등에 관한 정보도 추가로 제공하고 있으나, 본 연구에서는 Covid-19를 거쳐 비즈니스 방식 및 내용의 변화에 초점을 맞추고 있으므로 이를 직접적으로 반영하는 산업, 산업군, 상세 설명, 요약 설명 속성만을 활용한다.

<표 1>은 본 연구에서 활용된 데이터 정보이다. 수집된 기업의 수가 2018년~2019년과 2020년~2021년과 비교해볼 때 그 수가 현저히 감소한 것을 확인할 수 있다. 이는 스타트업 설립에 있어 Covid-19의 영향을 직접적으로 받은 것으로 보인다.

<표 1> Crunchbase 데이터 정보

Division	2018-2019	2020-2021
Path	Crunchbase	
Criteria	Before Covid-19	After Covid-19
Number of data	75,881	30,382
Number of column	102	
Selected column	Description, Full description, Headquarters location, Industries, Industry groups	
Analysis program	Python / Jupyter notebook	
Collection form	.xlsx	

3.2. 데이터 전처리

수집된 Crunchbase 데이터의 산업군, 산업, 요약 설명, 상세 설명, 본사 지역 위치 속성에 대해 결측치를 제거하는 작업을 수행하였다.

분석 시에는 키워드 단위로 나누어 활용되기 때문에 특수문자, 불용어 제거, 소문자 변경 과정을 거친 후 마지막으로 표제어 형태로 키워드를 추출하였다.

3.3. Descriptive analysis

연도별로 수집된 Crunchbase 데이터를 Covid-19 이전과 이후로 창업된 기업 빈도와 산업별 창업된 기업의 빈도 분석을 수행하여 전후 비교를 수행한다. Covid-19의 최초 유행 시기는 2020년 1월 이므로 2018~2019년을 Covid-19 이전 시기로, 2020~2021년을 Covid-19 이후 시기로 지정한다.

이후 Covid-19 전후로 창업연도별 기업 수 변화 분석 및 산업별 기업 수의 변화를 분석하여 Covid-19 가 산업 전반에 미친 영향을 파악한다.

3.4. Network Analysis

네트워크 분석은 개인이나 집단을 점으로, 서로 간의 관계성을 선으로 개념화하고 네트워크를 구성하여 개인과 조직이 주는 영향을 평가하기 위해 수학 및 시각적인 방법을 통해 패턴을 파악하여 분석하는 과정을 말한다(Scott, 2012).

본 연구에서는 키워드를 활용한 키워드 네트워크 분석을 활용하였다. 키워드 네트워크 분석은 특정한 주제 영역의 문헌 집합에서 의미가 있는 키워드를 찾아내고 키워드 쌍의 동시출현 빈도 계산을 통해 키워드 간의 유사도를 계산하여 구성된 네트워크를 분석하는 과정을 의미한다(이성신, 2016). 키워드 네트워크 분석은 방대한 문서에 관해 연구의 객관성을 높일 수 있으며(윤진아 외, 2016), 핵심 단어 간 의미론적 연관성을 파악할 수 있는 이점이 있어 자주 활용되는 방법이다(박지순, 2017).

본 연구에서는 데이터를 동시 발생 행렬(Co-occurrence matrix) 형태로 변형하고 Gephi를 통해 시각화하였으며, 모듈성 분석(Modularity Analysis)을 수행하였다. 모듈성 분석은 연결망을 구성하는 노드들의 군집구조를 식별하는 데 많이 사용되는 방법이다(조수곤·김성범, 2012).

3.5. 토픽 모델링

토픽 모델링(Topic modeling)이란 텍스트를 요약할 수 있는 텍스트 마이닝의 한 방법으로 문서 집단에 잠재된 토픽, 즉 주제들을 도출하기 위해 문서에서 등장하는 단어들의 패턴 확률을 모형화한다(문성현 외, 2018). 토픽 모델링 방법 중에는 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 방법이 현재까지도 가장 자주 사용되는 방법 중 하나다(Blei et al., 2003).

LDA는 토픽의 단어 분포와 문서의 토픽 분포의 결합확률이 최대가 되도록 하여 개별 단어들이 어떤 주제 비율에 따른 구성을 하는지를 파악하고, 문서들에 대한 토픽 비율 및 토픽에 따른 단어 비율을 결정한다.

본 연구에서는 Crunchbase의 데이터에서 기업에 관한 상세 설명을 기술한 상세 설명 속성을 문서로 정의하고 불용어 제

거, 소문자 변경, 특수문자 제거 등의 텍스트 전처리 과정을 거친 후 LDA 토픽 모델링 기법을 통해 산업들이 소유한 기술 토픽을 추출하고 이를 통해 Covid-19 기간 동안 어떠한 기술의 변화가 있는지, 유망·쇠퇴한 기술의 양상이 어떠한지 알아보고 동향 분석을 수행한다

3.6. Doc2Vec&Clustering

Doc2vec은 문장 임베딩을 수행하는 모델로 기존의 BoW(bag-of-words) 방법의 한계점인 단어의 정보 상실과 의미론적 정보를 무시하는 문제점들을 극복하고자 제안되었다(송진주·강승식, 2019). Doc2vec에서 가장 자주 활용되는 예측 방법은 PV-DM(Paragraph vector with distributed memory)과 방법이다(김도우·구명완, 2016). PV-DM 방식은 다음에 오는 단어를 예측할 때 로그 확률 평균을 최대화 하는 과정에서 Paragraph id에 대한 학습 과정을 추가하여 Paragraph에 대한 임베딩을 수행한다(Douzi et al., 2017).

본 연구에서는 기업의 설명 정보를 담고 있는 요약 설명 속성에 대해 임베딩을 수행한다. 이후 K-Means 알고리즘으로 동일한 군집끼리 묶은 후 각 군집의 주요 키워드를 추출하고 이를 기반으로 Covid-19 이전과 이후 기술 키워드의 변화를 살펴본다.

IV. 연구 결과

4.1. Descriptive Analysis

4.1.1. 창립연도별 기업 수 변화 분석

먼저 창립연도별 기업 수의 변화에 대한 분석을 수행하였다. 그 결과가 <표 2>에 나타나 있다. 먼저 1년 단위부터 살펴보면 Covid-19 이전 시기인 2019년의 기업 감소율은 2018년에 설립된 스타트업 수와 비교할 때 전년 대비 -29.9% 감소율을 보인다. 그리고 Covid-19가 본격적으로 유행한 시점인 2020년부터 -31.4%로 2019년보다 -1.5% 더 감소한 것을 확인할 수 있다. Covid-19가 장기화된 시점인 2021년에는 -58.3%로 전년 대비 -26.9% 감소하여 가장 큰 감소세를 보인다. 이러한 감소세는 2년 단위로 묶어 살펴볼 때 더 명확하게 나타난다. 2018~2019년에서 2020년~2021년으로 가면서 감소율이 -60%로 2년 동안 절반 넘게 창업 빈도가 감소한 것을 확인할 수 있다.

<표 2> 연도별 스타트업 빈도 변화

Period		Before Covid-19		After Covid-19	
		2018	2019	2020	2021
Number of startups	1year	44,606	31,275	21,451	8,931
	2year	75,881		30,382	
Startup decline rate	1year	-	-29.9%	-31.4%	-58.3%
	2year	-		-60%	

이를 종합해볼 때 창업 빈도는 Covid-19 시점 이전부터 감소세를 보이고 있기 때문에 스타트업 창업 감소의 직접적 원인이 Covid-19로 판단하기는 힘들다. 하지만 Covid-19가 진행되면서 감소율이 가속화되는 경향을 보였고 전년 대비 꾸준히 높은 수치의 감소율을 기록하고 있기 때문에 Covid-19는 스타트업 설립에 있어 간접적인 영향을 끼친 것으로 판단된다.

4.1.2. 산업별 기업 수 변화 분석

다음은 Covid-19 이전과 이후 산업간의 변화를 확인하기 위해 산업별 기업 통계 분석을 수행하였다. <표 3>, <표 4>에서는 Covid-19 이전과 이후 각각 상위 10개 산업의 기업 수 비율 변화를 보여주고 있다.

<표 3> Covid-19 이전 상위 10개 산업의 기업 수 비율 변화

Industry		Portion
1	Health Care	2.86%
2	Financial Services	2.37%
3	Commerce and Shopping	2.11%
4	Software	1.93%
5	Biotechnology, Health Care, Science and Engineering	1.65%
6	Information Technology	1.16%
7	Financial Services, Lending and Investments	1.12%
8	Information Technology, Software	1.07%
9	Artificial Intelligence, Data and Analytics, Science and Engineering, Software	0.98%
10	Advertising, Sales and Marketing	0.97%

<표 4> Covid-19 이후 상위 10개 산업의 기업 수 비율 변화

Industry		Portion
1	Health Care	3.17%
2	Software	3.14%
3	Commerce and Shopping	3.07%
4	Financial Services	2.80%
5	Education, Software	1.56%
6	Biotechnology, Health Care, Science and Engineering	1.48%
7	Financial Services, Lending and Investments	1.28%
8	Advertising, Sales and Marketing	1.10%
9	Information Technology	1.09%
10	Information Technology, Software	1.07%

1순위~5순위 산업을 살펴보면 ‘Health Care’, ‘Software’, ‘Commerce and Shopping’, ‘Financial Services’, ‘Biotechnology’가 대부분을 차지하고 있다. 이를 보면 의료, 소프트웨어, 온라인 쇼핑, 금융 서비스, 생명공학의 경우 고정적으로 창업이 활발한 산업임을 확인할 수 있다. 그리고 Covid-19 이후를 보면 ‘Software’, ‘Commerce and Shopping’, ‘Financial Services’,

‘Education, Software’ 순서로 비대면 기술 및 개개인의 금융과 관련된 산업이 상위 산업을 차지하고 있다. 이는 Covid-19 여파로 인해 일상생활의 대부분이 대면에서 비대면으로 바뀌고 경제적인 부분에서 타격을 받은 결과가 산업에도 그대로 반영되어 나타난 것으로 사료된다.

Covid-19 이후 소프트웨어 기반의 교육산업(Education, Software)이 새로 출현하였다. Covid-19의 장기화로 교육도 점차 비대면으로 진행되는 상황이 많이 발생함에 따라 자연스럽게 온라인에 기반한 교육 산업에도 그 영향을 끼친 것으로 보인다.

4.2. Network Analysis

4.2.1. Industry Groups Network Analysis

전체 산업별 동향을 파악하기 위해 산업군과 산업 속성을 통해 네트워크 분석을 수행하였다. 먼저 상위 산업군 관점에서 전반적인 산업의 변화를 살펴본다. <그림 2>과 <그림 3>에서 네트워크 그래프 형태로 시각화한 결과를 확인할 수 있다. 노드, 레이블은 모두 연결 중심성(Degree Centrality)을 기준으로 크기 조절을 진행하였으며, 이는 산업의 변화를 명확하게 식별할 수 있기 때문에 해당 네트워크 척도로 활용하였다.

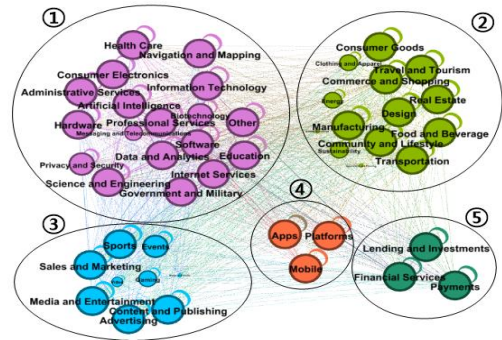
<표 5>에서는 산업군의 네트워크 그래프에 대한 기본 정보를 나타내고 있다. 이를 살펴보면 Covid-19 이전과 이후에서 노드의 변화는 없었으며, 엣지는 1,091에서 1,077(-1.28%), 모듈성 군집의 경우 5에서 4(-20%)로 감소하였다. 데이터의 개수는 Covid-19 이전과 비교했을 때, 절반 이상(60%) 감소한 것을 확인하였다. 결론적으로 네트워크 정보를 종합해보면 일부 컬럼의 수치가 감소하는 경향을 보였으나 Covid-19 이전과 이후의 큰 변화는 감지되지 않았다.

<표 5> Covid-19 전후 산업군 기본 정보

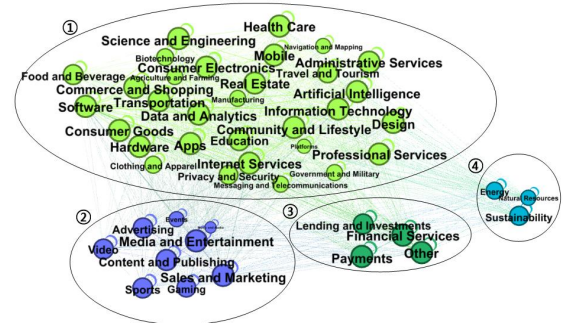
	Before Covid-19	After Covid-19
Column	Industry groups	
Period	2018~2019	2020~2021
Number of data	75,881	30,382(-60%)
Node	47	47
Edge	1,091	1,077(-1.28%)
Modularity	5	4(-20%)

네트워크 분석 결과, Covid-19 이전 상위 산업군에서는 총 5개의 군집이 도출되었다. Covid-19 이전에는 교육, 하드웨어, 소프트웨어, 인공지능, 헬스케어, IT, 보안의 산업이 포함된 IT 기반의 SW·HW 서비스 클러스터가 38.30%로 가장 큰 비율을 차지하고 있으며, 그다음으로 문화·일상 29.79%, 미디어·광고 19.15% 클러스터가 다음 순위를 이루고 있다. 모바일과 금융 서비스는 6.38%로 동일한 비율을 차지하고 있다. 이를 종합해보면 군집마다 많은 산업 키워드가 포함되어 있지만 대부분

비슷한 성격을 가진 산업이 주를 이루고 있다. IT 기반의 SW·HW 서비스 산업이 가장 큰 비중을 차지하고 있고, 그 뒤로 문화·일상, 미디어·광고 산업이 주를 이루고 있다. IT와 의료, 교육이 하나로 묶인 것을 고려할 때 IT와 의료, 교육산업 간의 융합이 발생하여 온라인 교육, 의료 형태로 산업 발전이 이루어진 것으로 보인다.



<그림 2> Covid-19 이전 산업군 네트워크 시각화



<그림 3> Covid-19 이후 산업군 네트워크 시각화

<표 6> 및 <표 7>은 Covid-19 이전과 이후의 산업군에 대한 주요 카테고리 변화 및 각 카테고리의 비중을 나타낸다.

<표 6> Covid-19 이전 산업군 카테고리 분류

Cluster		Industry category	Portion
1	IT-based SW/HW service	Education, Hardware, Software, AI, Health Care, IT, Privacy and Security	38.30%
2	Culture/Daily life	Design, Consumer Goods, Food and Beverage, Travel and Tourism, Transportation, Commerce and Shopping, Energy	29.79%
3	Media/Advertising	Advertising, Sports, Gaming, Media and Entertainment, Sales and Marketing	19.15%
4	Mobile service	Mobile, Apps, Platforms	6.38%
5	Financial service	Lending and Investments, Financial Services, Payments	6.38%

<표 7> Covid-19 이후 산업군 카테고리 분류

Cluster	Industry category	Portion
1	IT/Culture	Education, Consumer, Transportation, IT, Privacy and Security, Hardware, Design, Biotechnology, AI, Real Estate, Travel and Tourism, Food and Beverage, Commerce and Shopping, Health Care
2	Media/Advertising	Sports, Gaming, Sales and Marketing, Media and Entertainment, Advertising
3	Financial service	Lending and Investments, Financial Services, Payments
4	Nature/Renewable energy	Energy, Natural Resources, Sustainability

Covid-19 이후에는 군집 개수가 4개로 감소하였다. 군집 그 자체만 살펴보면 때는 큰 차이가 없는 것처럼 보이지만 실제로 각 클러스터의 비중을 살펴보면 Covid-19 이후의 변화가 뚜렷하게 나타나는 것을 알 수 있다. 먼저 IT가 포함된 그룹이 훨씬 다양한 산업을 포함하면서 그 비율이 전체 네트워크의 절반 이상인 65%를 차지하고 있는데 이와 같은 결과가 나타난 이유는 Covid-19의 장기화에 따라 비대면, 온라인과 관련한 산업이 주목받게 되면서 기존 산업들도 비대면, 온라인 산업의 형태로 변화하거나 융합하기 위해 IT, 소프트웨어, 인공지능과 관련된 산업군에 소속된 것으로 생각된다. 미디어, 광고 산업은 Covid-19 이전과 비교하여 큰 변화 없이 비율을 유지하였고, 금융클러스터 비율은 Covid-19 이후 2.13% 증가하였다. 이는 Covid-19 이후 폭락했던 금융시장이 시간이 지남에 따라 다시 안정을 찾으면서 투자 수요가 높아지고, 이에 따라 자연스럽게 금융과 관련된 산업의 중요도가 높아진 것이 원인으로 보인다. 한 가지 주목할 점은 Covid-19 이전 문화·일상 클러스터에 소속되어 있던 에너지 산업이 Covid-19 이후로 넘어오면서 독립적인 그룹을 형성했다는 점이다. Covid-19 이후 전 세계적으로 천연자원이나 재생 가능한 에너지 산업에 투자하면서 스타트업에도 이러한 변화가 그대로 반영된 것으로 보인다. 이를 종합해볼 때 Covid-19 이후 모바일, 문화와 관련된 산업들은 IT, 소프트웨어 그룹에 포함되면서 온라인이나 비대면의 성격을 띠는 형태로 변화하였고, 경제침체가 진행됨에 따라 금융과 관련한 산업들의 비중은 자연스럽게 증가한 것을 확인하였다.

4.2.2. Industries network analysis

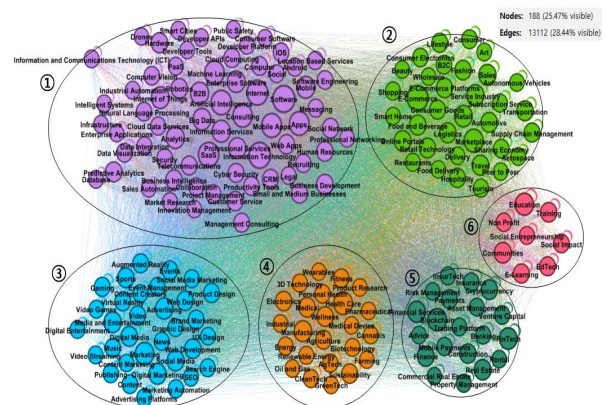
다음은 산업 속성에 대한 네트워크 분석을 수행하였다. 이는 산업군에 비해 훨씬 세부적인 산업의 변화를 보기 위함이다. <표 8>를 살펴보면 Covid-19 이후 군집의 개수가 1개 증가하였는데 이는 특정 산업의 성장으로 인해 독립적으로 새로운 군집이 생성된 것으로 보인다. 그리고 옛지의 경우 데이터의 개수를 고려하더라도 Covid-19 이전보다는 감소가 있다고 판단된다. 이를 통해 전체적으로 Covid-19 이후 스타트업이 감소하면서 자연스럽게 산업 간의 관계성에도 그 영향을 끼친 것으로 보인다. 그리고 개별 군집이 추가로 생성된 것을 통해

특정 산업에 대한 독보적인 성장이 있음을 시사하고 있다.

<표 8> Covid-19 전후 산업 기본 정보

	Before Covid-19	After Covid-19
Column	Industries	
Period	2018-2019	2020-2021
Number of data	75,881	30,382(-60%)
Node	738	721(-2.3%)
Edge	46,098	28,244(-38.73%)
Modularity	6	7(+17%)

<그림 4>은 Covid-19 이전 산업 속성에 대한 네트워크 그래프를 시각화한 결과를 나타내고 있다. 현재 보이는 노드들은 산업 키워드들을 한눈에 파악하기 위해 Cuf-off 값을 활용해 노드의 연결 정도가 100 이상인 것만 표시하였다. 노드는 전체 738개 중 188개로 25.74%, 엣지는 46,098중 13,112개로 28.44%로 나타난다. Covid-19 이전에는 6개의 군집을 확인할 수 있는데 대부분 동일한 성격을 가지는 산업끼리 그룹화가 되어있다. 그중에서는 IT 기반의 SW·HW 서비스 산업 비중이 31.98%로 가장 큰 비중을 차지하고 있으며 다음으로 문화·일상, 첨단기술/미디어 기반 마케팅·광고 산업이 뒤를 따르고 있다. <표 9>는 Covid-19 이전 네트워크 요약을 나타낸다.



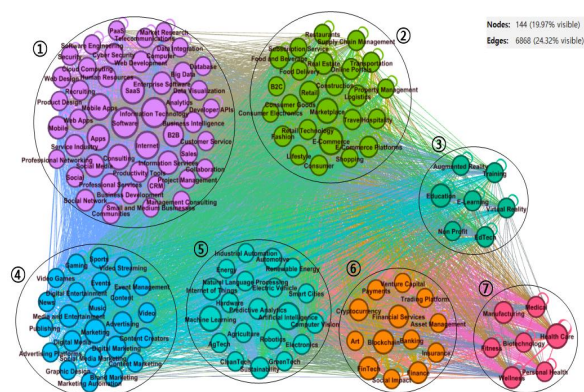
<그림 4> Covid-19 이전 산업 네트워크 시각화

<표 9> Covid-19 이전 산업 카테고리 분류

Cluster	Industry category	Portion
1	IT-based SW/HW service	Software, IT, AI, Security, Hardware, Robotics, Mobile, Apps, Social Network, Machine Learning
2	Culture/Daily life	Retail, Fashion, Consumer Goods, Logistics, Automotive, Travel, Food Delivery, Food and Beverage, Shopping, Restaurants, E-Commerce
3	Advanced technology /Media-based Marketing&Advertising	Digital Media, Marketing, Digital Marketing, Brand Marketing, Product Design, Virtual Reality, Sports, Gaming, Augmented Reality, Media and Entertainment, Advertising
4	Medical/life/pharmaceutical	Health Care, Agriculture, Energy, Biotechnology, Pharmaceutical

5	Financial service	Insurance, Financial Services, Real Estate, Cryptocurrency, Asset Management	9.89%
6	Education/Non-profit	Education, Not Profit, E-Learning, EdTech, Social Impact, Social Entrepreneurship	9.35%

다음으로 <그림 5>는 Covid-19 이후 네트워크 시각화를 나타내고 있다. Cut-off에 대한 기준은 이전과 동일하게 적용되었으며, 노드는 721개 중 144개로 19.97%, 엣지는 28,244개 중 24.32%를 나타내고 있다. <표 10>은 Covid-19 이후 나타난 클러스터의 특성 및 비중을 요약하여 나타내고 있다. 먼저 IT 산업에 소속되어 있던 머신러닝, 인공지능, 하드웨어, 컴퓨터 비전, 로봇, IoT 산업이 독립적인 새 그룹을 이루어 등장하였다. 이는 분명 인공지능, 머신러닝과 같은 AI 산업의 발전에 의한 것으로 보인다. AI/ML 기반 첨단기술 클러스터의 비율은 11.51%로 적지 않은 수치를 기록하였고, 농업, 산업 자동화 같은 다른 산업과의 융합 가능성을 고려할 때, Covid-19 이후에도 다른 산업과 융합하여 AI가 적용된 새로운 형태의 신산업 등장을 기대할 수 있다. 교육산업에도 많은 변화가 일어났는데 Covid-19 이전에는 교육의 비율이 9.35%로 군집중에서는 가장 적은 비중을 차지하고 있었는데 Covid-19 이후 9.35%→15.95%로 증가한 것을 확인할 수 있다. 7개의 군집중에서도 3순위를 차지할 만큼 중요도가 높아졌다고 볼 수 있으며, 이는 곧 대면으로 진행되어야 할 교육산업에 많은 변화가 일어나고 있음을 짐작할 수 있다. 특히 가상현실, 증강현실 산업과 융합되는 양상을 보여 기존 대면 교육의 사·공간 제약에 대한 한계점을 극복하기 위한 기업들의 혁신적인 시도가 지속되고 있음을 알 수 있다.



<그림 5> Covid-19 이후 산업 네트워크 시각화

<표 10> Covid-19 이후 산업 카테고리 분류

Cluster	Industry category	Portion
1	IT-based online service	23.72%
2	Culture/Daily life	16.64%
3	Non-contact/Advanced technology based education service	15.95%
4	Online/Media based advertising/marketing	13.45%
5	AI/ML based advanced technology	11.51%
6	Financial service	9.71%
7	Health care/Biotechnology/Personal Health	8.6%

4.3. Topic modeling

네트워크 분석에 이어 산업들이 소유하고 있는 기술 동향을 살펴보기 위해 LDA 토픽 모델링을 수행한다. 앞서 수집한 기간별 데이터를 통해 불용어 제거, 특수문자 제거 등의 텍스트 전처리 과정을 거쳐 분석을 수행하였다. 토픽 모델링은 효과적인 토픽 추출을 위해 Coherence score와 Perplexity score가 존재한다. Coherence score는 토픽이 얼마나 의미론적으로 일관성이 있는지를 판단하는 척도이며 값이 클수록 일관성이 높다고 판단한다. Perplexity score는 확률 모델이 결과를 얼마나 정확하게 예측하는지를 판단하는 척도이며 Coherence score와는 반대로 낮을수록 정확하게 해석한다. 따라서 해당 연구에서는 Coherence score가 높으면서 Perplexity score가 낮은 지점을 찾아 각각 최적의 토픽 수를 산정하였다. 그 결과 <표 11> 및 <표 12>와 같이 Covid-19 이전은 7개, Covid-19 이후는 7개로 도출되었다. 그리고 각 토픽은 단어와 해당 토픽에 기여 점수가 포함된 값을 기준으로 상위 20개의 데이터를 보여준다.

<표 11> Covid-19 이전 토픽별 상위 10개 키워드

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7
11.44%	25.05%	15.74%	13.04%	15.64%	8.48%	10.61%
Marketing (Web, Media, Brand, Digital)	Service, Software	Financial service, Capital investment, Startup	E-Commerce	AI, ML, Customer service	Online, Education, Game, Sports	Health care, Treatment development
marketing	service	financial	online	learning	students	health
digital	software	digital	home	business	online	medical
development	management	business	food	experience	headquartered	care
media	app	companies	app	intelligence	games	technology
design	users	technology	want	team	game	healthcare
content	solution	market	every	work	education	patients
business	customers	global	share	technology	world	patient
web	cloud	investment	community	customer	sports	developing
social	business	capital	fashion	software	gaming	treatment
website	mobile	world	world	provide	united	research

<표 12> Covid-19 이후 토픽별 상위 10개 키워드

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7
7.97%	21.15%	19.55%	13.63%	16.73%	10.41%	10.56%
Cryptocurrency, trading	Delivery service, energy, Construction/Logistics/Manufacturing	Marketing, Software	Content, Entertainment	Capital investment, Startup, Venture	Health care service	Financial. Online app
online	food	marketing	content	technology	health	financial
crypto	delivery	digital	news	investment	service	online
market	energy	business	video	capital	care	business
payment	sustainable	development	music	companies	technology	learning
business	supply	design	social	intelligence	security	world
customers	local	management	nft	world	software	create
marketplace	technology	software	world	investors	healthcare	users
shopping	carbon	media	games	technologies	provide	community
blockchain	customers	web	media	artificial	solution	students
sell	waste	team	game	startup	business	experience
buy	chain	sales	entertainment	stage	medical	app

토픽 분석은 크게 변화가 포착된 산업들 기준으로 파악하고자 한다. 먼저 미디어 산업의 경우 Covid-19 이후의 토픽 4에서 nft와 게임 키워드와 묶여 주제로 출현이 되었고, nft의 경우 Covid-19 이후 토픽 1의 가상화폐와 독립적으로 구분되어 토픽 4에 출현이 되었는데 이는 Covid-19 이후 유튜브와 같은 1인 미디어 플랫폼이 활성화되면서 미디어 기업들이 자신들의 콘텐츠 IP로 캐릭터 nft를 만들어내면서 미디어, nft, 게임과 같은 키워드와 토픽 4의 주로 나타나고 있음을 확인할 수 있다. 의료 산업 분야에서는 Covid-19 이전의 토픽 7과 이후 토픽 6에 각각 꾸준히 출현 되고 있다. 다만 Covid-19 이전의 의료 산업 분야의 경우 주로 특정 질병·질환에 대한 치료 및 치료제 개발 자체를 목적으로 하는 스타트업들이 대부분이었다면, Covid-19 이후 치료 자체의 목적보다는 의료 원격 서비스와 같이 환자화 비대면 서비스를 제공하는 스타트업들이 대거 등장하였음을 시사한다.

Covid-19 이후에는 배달 대행이라는 산업이 활발해지기 시

작했는데 Covid-19 이전에도 음식을 의미하는 food 키워드는 online 키워드와 결합하면서 Covid-19 이전 토픽 4에도 출현되었다. 식료품을 온라인으로 주문하는 서비스를 제공하는 스타트업들이 존재하였으나, Covid-19 장기화가 진행됨에 따라 식료품뿐만 아니라 외식이 줄어들면서 집에서 배달할 수 있는 배달 대행 관련 서비스 산업이 활발해지면서 배달 대행 산업에 가속화가 붙기 시작하였다.

교육산업에도 변화가 보였는데 Covid-19 이전 토픽 6과 Covid-19 이후 토픽 7에서 각각 교육산업이 등장하였으나 Covid-19 이후에는 주로 온라인 금융, 온라인 앱과 관련한 키워드와 묶여 토픽 7에 출현하고 있다. 이를 Covid-19 이전 시기와 비교할 때, 주로 온라인, 모바일, 앱과 같은 키워드와 묶여 온라인 교육 및 앱을 이용한 교육 방법은 동일하나 교육을 받는 대상과 범위가 확장되면서 금융 및 투자와 관련한 금융 교육 서비스를 제공하는 스타트업들이 등장하기 시작한 것을 확인할 수 있다.

4.4. Doc2Vec&Clustering

LDA 토픽 모델링과 같이 산업별 기술 콘텐츠를 다각적으로 분석하기 위해 Doc2vec을 통해서 요약 설명 문서에 임베딩 과정을 학습하고 대표적인 군집 알고리즘인 K-Means 알고리즘을 적용하였다. K-Means 알고리즘의 최적의 클러스터 개수를 선정하기 위해 토픽 모델링에서 최적의 토픽 개수 8개를 기준으로 클러스터링을 진행하였다. 토픽 모델링과 출력 결과는 비슷하지만, 토픽 모델링의 경우에는 단어와 해당 단어가 토픽에 기여한 점수가 출력되고 Doc2vec&Clustering에서는 키워드(단어)와 해당 단어의 빈도 및 비율이 출력된다는 점에서 근소한 차이를 보인다. <표 13>과 <표 14>는 Covid-19 이전과 이후의 각 군집별 상위 키워드를 나타낸다.

<표 13> Covid-19 이전 군집별 대표되는 상위 10개 키워드

Cluster	Keyword
1 Marketing, Web	digital marketing, web design, web development, social media marketing, brand marketing, graphic design, management consulting, content marketing, business development, mobile
2 Food and Beverage, Tourism	food and beverage, retail, fashion, biotechnology, travel, digital marketing, gaming, mobile, wellness, automotive
3 Security, Software	enterprise software, cyber security, b2b, security, productivity tools, cloud computing, business intelligence, cloud data services, developer tools, computer
4 Health care	medical, biotechnology, wellness, medical device, pharmaceutical, therapeutics, life science, personal health, fitness, hospital
5 Financial	cryptocurrency, finance, banking, payments, venture capital, bitcoin, trading platform, financial exchanges, real estate investment, personal finance
6 Energy, Biotechnology, Food and Beverage, Robotics, Flight	biotechnology, agriculture, energy, internet of things, robotics, agtech, electronics, food and beverage, automotive, renewable energy
7 Education, Sports	edtech, e-learning, training, recruiting, communities, higher education, sports, wellness, fitness, non profit
8 Tourism, Fashion, Food and Beverage	travel, fashion, food and beverage, retail, lifestyle, social network, mobile, hospitality, shopping, tourism

Doc2vec&Clustering을 적용한 결과 각각 식품, 의료, 엔터테인먼트 산업에서 Covid-19 이후 주목할만한 변화를 관찰하였다. 먼저 식품(food and beverage)의 경우 Covid-19 이후 cluster 3의 배달 대행, 패션 키워드와 함께 묶여 주제로 출현하였고, cluster 5에서는 생명공학, 디지털마케팅, 웹 개발 키워드와 묶여 주제로 출현 되었다. 즉 식품 키워드는 Covid-19 이후 2개의 cluster에서 서로 다른 키워드들과 묶여 있는데 cluster 3 주제인 식품, 배달 대행, 패션, 농업에서의 식품 키워드는 주로 사람들의 여가·편의 서비스와 관련된 주제와 묶여 온라인으로

식품·음식·의료 등을 주문·배달할 수 있는 서비스를 제공하는 스타트업들이 다수 등장한다는 점, 그리고 cluster 5의 주제인 식품, 생명공학, 디지털마케팅, 웹 개발에서는 식품의 키워드가 식품 자체를 제조하는 스타트업들이 등장한다는 점에서 서로 다른 차이를 보인다.

의료 산업에서는 Covid-19 이전의 경우 특정 질환·질병에 대해 1차원적 치료를 목적으로 하는 스타트업들이 대부분 설립되었으나, Covid-19 이후 DTx(Digital Therapeutics)와 같이 치료를 보조하는 역할을 수행하는 웨어러블 디바이스 및 의료진·환자들에게 의료서비스(의료 스케줄)를 제공하는 스타트업들이 등장함을 확인하였다.

마지막으로 엔터테인먼트로 정의되는 음악(music), 게임(game), 비디오(video)와 같은 키워드의 경우 Covid-19 이전에는 주로 여가·편의 서비스 키워드와 묶여 cluster 2와 cluster 7에 등장하였다.

<표 14> Covid-19 이후 군집별 대표되는 상위 10개 키워드

Cluster	Keyword
1 Security, Software	cyber security, enterprise software, b2b, productivity tools, business intelligence, cloud computing, internet of things, data integration, predictive analytics, developer tools
2 Financial	finance, cryptocurrency, banking, venture capital, payments, personal finance, trading platform, bitcoin, mobile payments, insurance
3 Food and Beverage, Food delivery, Fashion	retail, food and beverage, fashion, food delivery, consumer goods, logistics, e-commerce platforms, retail technology, shopping, delivery
4 Marketing, Web, Graphic	digital marketing, web development, web design, social media marketing, professional services, brand marketing, graphic design, content marketing, management consulting, business development
5 Food and Beverage, Biotechnology, Digital marketing, Web	food and beverage, biotechnology, digital marketing, web development, retail, web design, wellness, cryptocurrency, gaming
6 Education, Financial	edtech, e-learning, training, higher education, recruiting, wellness, tutoring, corporate training, secondary education, stem education
7 Health care	medical, biotechnology, wellness, medical device, personal health, therapeutics, life science, health diagnostics, pharmaceutical, mhealth
8 Entertainment, Social Network	music, social network, gaming, news, mobile, digital media, digital entertainment, sports, travel, cryptocurrency

마지막으로 엔터테인먼트로 정의되는 음악(music), 게임(game), 비디오(video)와 같은 키워드의 경우 Covid-19 이전에는 주로 여가·편의 서비스 키워드와 묶여 cluster 2와 cluster 7에 등장하였다. Covid-19 이후 장기화가 진행됨에 따라 디지털 활동이 활발해지면서 음악(music), 게임(game), 비디오

(video) 등과 같은 엔터테인먼트 산업이 소셜네트워크(social network) 키워드와 함께 묶여 cluster 8에 출현하였다. 이를 해석하면 음악, 게임, 비디오 키워드는 Covid-19 이전에는 단순히 여가·편의를 위한 산업 일부였으나 디지털 활동이 활발해지는 Covid-19 이후 시기부터 음악, 게임, 비디오 등을 포괄하는 엔터테인먼트 산업이 소셜네트워크 키워드와 함께 묶이면서 특히 VR(Virtual Reality), AR(Augmented Reality), Metaverse와 같이 시·공간 제약이 없는 온라인 플랫폼 산업의 활성화를 의미하고 있다.

4.5. LDA, Doc2Vec&Clustering 비교 분석

다음으로 이전 단계에서 진행한 LDA와 Doc2vec의 결과 비교 분석을 수행한다. <표 15>을 통해 확인해보면 우선 LDA와 Doc2vec에서 의료, 배달, 교육산업에서 공통점을 보인다. 먼저 사람을 대면해야 하는 서비스 산업들이 Covid-19 이후 비대면 서비스로 전환되는 산업 분야에서 공통점을 보인다. 그 예로 배달 산업은 Covid-19 장기화의 여파로 외식이 줄어들면서 집에서 음식을 배달하는 형태의 비대면 서비스로 제공된다. 의료 산업도 비슷한 성격을 가지는데 Covid-19 이전 환자를 대면하는 1차원적 치료 서비스를 제공하는 형태에서 원격 서비스와 같은 비대면 서비스의 활성화가 진행되어 환자가 병원을 방문하지 않고도 원격으로 치료가 가능한 산업의 형태로 변화하였다. 두 번째로는 기존 산업이 더 크게 확장되는 형태로 교육산업의 경우 원래 단순히 학생들의 교육 자체에 초점을 맞춘 산업이 주를 이루었다면, Covid-19 이후 금융 산업 등과 연계하면서 교육의 범위와 받는 대상에 대한 다양화가 진행되는 관점에서 공통점을 보인다.

<표 15> Covid-19 변화에 대한 분석 방법별 차이점 및 공통점

	LDA	Doc2vec
Commonality	Health care, Food delivery, Education	
	Activation of Non-contact service Expansion of existing industries	
Difference	Entertainment	
	Differences between groups with specific industry keywords	

차이점으로는 특정 산업 키워드가 속한 그룹 간의 차이가 발생한다는 점이다. 엔터테인먼트 산업의 경우 토픽 모델링에서는 콘텐츠(미디어), nft 키워드와 묶이면서 콘텐츠(미디어)와 엔터테인먼트를 통한 nft 산업의 확장을 의미하고 있고, Doc2vec 방법에서는 엔터테인먼트 산업이 소셜네트워크 산업과 묶이면서 디지털 산업의 확장을 의미한다는 점에서 다소 차이를 보인다.

VI. 결론

본 연구는 Covid-19 이후 산업을 거시적 관점에서 동향 변화를 살펴보고 체계적으로 분석하기 위해 Crunchbase를 통해 산업 수준의 동향 분석을 수행하였다. 이를 위해 네트워크 분석을 통해서 전체적인 산업 변화의 흐름을 관찰하였으며, 토픽 모델링과 Doc2vec 방법을 통해 해당 산업들이 보유한 기술 콘텐츠를 분석하는 과정을 거쳤다.

분석 결과 Covid-19 이후 소프트웨어, 전자상거래, 금융 서비스, 교육 서비스가 크게 성장하였으며, 특히 소프트웨어 기반 교육산업이 새로 출현하여 크게 성장한 것으로 보인다. 에너지, IT, 교육산업 분야에서도 많은 변화가 나타났으며, 에너지 산업은 Covid-19 이후 독립적인 개별 그룹을 형성하였고, IT 산업의 경우 IT 산업 내에 존재하던 AI 관련 산업이 독립적인 그룹을 형성하였다. 교육산업은 특히 가상현실, 증강현실 등과 같이 융합되면서 비대면 시대에 특화된 새로운 교육 비즈니스 혁신이 일어나고 있는 것으로 보인다.

의료, 배달, 교육산업에서는 비대면 서비스 전환으로 인해 기존 산업이 확장되는 양상이 뚜렷히 관측되었으며 엔터테인먼트 산업 역시 콘텐츠 측면에서 변화를 보이며 새로운 혁신을 주도해가는 것으로 보인다. 이를 정리하면 Covid-19는 기존 산업의 확장·다양화(식품, 교육), 디지털화(엔터테인먼트, 미디어), 비대면 서비스화(의료, 배달 대행) 등 여러 가지 산업의 형태를 변화시켜 왔고 이를 통해 기존 산업의 장벽을 뛰어넘는 새로운 기회를 가져다준 것으로 파악되었다.

결론적으로 Covid-19 이전과 이후 산업의 동향 변화는 단순 시간 변화에 따른 산업 동향 분석과는 많은 차이를 보인다. 특히 비대면과 관련한 기술·산업에서 뚜렷한 성장을 나타내고 있으며 그 외에 의료 및 개인의 생활과 관련된 산업이 활발한 창업이 이루어지고 있는 것을 확인할 수 있다. 특히 비대면, 온라인 기술들이 이전보다 더 전문성을 갖추고 있으며 단일 기술의 발전이 아닌 융합 기술에 초점을 두고 많은 산업이 성장하고 있는 것으로 확인되었다.

본 연구는 창업자, 벤처투자자, 예비창업자 등 다양한 이해관계자들이 산업의 흐름을 파악할 수 있게 하고 기업 투자의 사결정에 많은 도움을 줄 수 있는 자료를 제공한다는 점에서 그 의미가 있다. 주로 Covid-19 전후로 특정 산업 내에서의 변화 분석에 초점을 맞춘 기존 연구와는 달리(Spoorthy et al., 2020; 최연석 2020; 윤경재 2021; 이용상·신동광, 2020; Daniel, 2020; 윤옥한, 2020; 김경근, 2020; 김은미 외, 2022; 김정오, 2022; 신용재, 2020; 김광중, 2021; Jung, 2022), 산업 전체를 대상으로 Covid-19 전후를 분석함으로써 거시적 관점에서 산업 변화를 분석하고, 추후 혁신을 위한 단서를 제공했다는 점 역시 본 연구의 의의로 보인다.

본 연구는 프로세스, 분석, 데이터 측면에서 한계점을 가지고 있다. 먼저 프로세스 측면의 경우 다양한 분석 방법론을 통해 전문가 관점의 분석보다는 더 객관적 분석을 시도하고

자 하였다. 그러나 토픽 모델링과 Doc2vec 등 일부 방법론의 경우 토픽의 네이밍 과정 등에서 연구자의 주관성이 일부 반영될 수 있다는 한계가 있어, 완전히 객관적인 방법을 활용했다고 보기는 어렵기 때문에 추후 연구에서는 객관성을 더 높일 수 있는 접근을 고려해 볼 수 있다. 그리고 네트워크 분석의 경우 산업의 범주가 다양하므로 세부적인 동향까지 식별하기 어렵다는 한계점이 있다. 분석 측면의 경우 연구가 동향 분석이라는 점에서 세부 산업들의 명확한 분석이 수행되지 못하였다. 연구에서는 단순 산업들의 시간적 변화를 파악하여 유망·쇠퇴한 산업 및 산업 간의 융합 정도를 파악하였으나, 실제 이러한 결과가 나온 것에 대한 원인을 설명하기 어렵다. 마지막으로 연구에 사용된 Crunchbase 데이터의 경우 결측치 비율이 매우 높아 분석에 활용할 수 있는 속성이 한정되는 문제가 있다. 실제 Crunchbase 데이터에는 펀딩 수, 페이지 방문 수 등 다른 관점에서 기업을 분석하는 방법들이 존재하지만, 결측치 문제 때문에 다양한 관점의 분석 수행을 하지 못하였다. 따라서 이러한 한계점들을 극복한다면 현재보다 더 정확하고 많은 관점에서 산업 동향을 파악할 수 있을 것으로 생각한다.

REFERENCE

- 김건우·한서희·이재석·김나영·옥지원(2020). COVID-19 이후 국내 교육 동향 분석: 비대면 교육 중심으로. *차세대컨버전스정보서비스기술논문지*, 9(4), 377-390.
- 김광중(2021). *코로나시대 패밀리레스토랑 배달 서비스 품질 요인이 브랜드 만족도, 브랜드 신뢰도, 브랜드 충성도에 미치는 영향*. 석사학위논문, 세종대학교 관광대학원.
- 김경근(2020). 변화의 시대를 맞이한 한국 고등교육의 주요 과제 및 대안. *교육사회학연구*, 30(4), 1-32.
- 김도우·구명완(2016). Doc2Vec 을 활용한 CNN 기반 한국어 신문 기사 분류에 관한 연구. *한국어정보학회 학술대회*, 67-71.
- 김상미(2020). 코로나19 관련 온라인 교육에 관한 국내 언론보도기사 분석. *디지털콘텐츠학회논문지*, 21(6), 1091-1100.
- 김상겸·장성용(2016). 토픽 모델링을 이용한 국내 산업경영공학 연구동향 분석. *한국경영공학회지*, 21(3), 71-95.
- 김선희(2021). COVID-19 시대의 교육 관련 글로벌 연구 동향 분석: 아시아태평양 지역을 중심으로. *아시아리뷰*, 11(2), 187-215.
- 김영환·정주훈·강두봉·박경민·김상미(2015). 『평생학습사회』의 연구주제 변화 동향 분석: 네트워크 텍스트 분석방법의 적용. *평생학습사회*, 11(1), 291-315.
- 김은미·Shi Bowen·민윤경(2022). COVID-19 상황 속 사회적경제의 조직 레질리언스에 대한 질적연구: 발달장애인 교육 서비스 협동조합과 사회적협동조합을 중심으로. *사회적가치와 기업연구*, 15(1), 165-260.
- 김정오(2022). COVID-19 팬데믹 상황에서 서비스산업종사자들의 직무만족이 삶의 만족에 미치는 영향: 가족지지의 매개효과를 중심으로. *관광연구저널*, 36(7), 97-108.
- 김지선·이형룡(2022). 소셜미디어 빅데이터를 활용한 코로나 19 전·후 호텔고객의 변화된 인식비교: 텍스트마이닝과 의미연결망 분석을 중심으로. *호텔경영학연구*, 31(4), 1-15.
- 권영환(2022). 코로나바이러스감염증-19 후유증과 항공안전. *항공우주의학지*, 32(3), 80-84.
- 문성현·정세환·지석호(2018). Latent Dirichlet Allocation 기법을 활용한 해외건설시장 뉴스기사의 토픽 모델링(Topic Modeling). *대한토목학회논문집*, 38(4), 595-599.
- 문명주·고인곤(2022). 포스트 코로나 시대의 항공물류 서비스기업의 경영성과에 대한 연구: 신뢰 및 호혜 개념을 중심으로. *벤처창업연구*, 17(3), 135-145.
- 박지순(2017). 키워드 네트워크 분석을 통한 장애인 교육 관련 연구 동향 분석. *특수교육재활과학연구*, 56(2), 107-126.
- 박혜림·이수정(2022). COVID-19 팬데믹 상황에서 의료기관 종사자의 불안, 스트레스, 대처유형 분석. *한국산학기술학회논문지*, 23(2), 57-66.
- 박민영·정성화·김희선·이은지(2022). COVID-19 발생 전·후 언론 보도에 나타난 간호사 이미지에 대한 텍스트 네트워크 분석 및 토픽 모델링. *Journal of Korean Academy of Nursing*, 52(3), 291-307.
- 박준태(2022). 텍스트 마이닝을 활용한 코로나 19 이후 장애인 체육 연구동향 분석. *한국스포츠학회*, 20(3), 241-249.
- 박철우·강경란(2020). 앙트러프러너십과 변혁적 리더십이 취·창업 인식에 미치는 영향 연구: 코로나 19 사회적 이슈의 매개효과. *벤처창업연구*, 15(5), 51-63.
- 신용재(2020). 항공운송서비스 산업의 경제적 역할과 한국 경제에 미치는 영향: COVID-19 상황을 대처하기 위해. *한국항공경영학회지*, 18(3), 115-136.
- 송민정·이슬자·조장희(2022). 코로나19 로 인한 대구·경북지역 관광 산업 피해와 파급효과: 산업연관분석을 중심으로. *한국경영컨설팅학회*, 22(6), 239-277.
- 송진주·강승식(2019). Doc2Vec을 이용한 특허 문서 자동 분류. *한국정보처리학회 학술대회논문집*, 26(1), 239-241.
- 송홍규(2022). 빅데이터를 활용한 밀키트 산업의 동향 연구: 코로나19 발생 전후 시점 비교를 중심으로. *외식경영연구*, 25(5), 443-457.
- 서은경(2010). 『정보관리학회지』 연구의 동향분석. *정보관리학회지*, 27(4), 7-32.
- 이완형(2020). COVID-19 팬데믹으로 인한 유통물류산업의 비즈니스 환경변화와 리스크 극복 전략. *유통경영학회지*, 23(3), 81-95.
- 이소영·윤소원·박지수·이지영·김나희·문세연·김남영(2022). 코로나바 이러스 감염증-19 에 대한 지식과 두려움이 간호대학생의 취업불안에 미치는 영향. *한국응용과학기술학회지(구 오일 및 응용과학 학회지)*, 39(3), 407-416.
- 이성신(2016). 키워드 네트워크 분석을 통한 도서관마케팅 연구 경향 분석. *한국문헌정보학회지*, 50(3), 383-402.
- 이채윤(2007). 직무스트레스와 상사의 리더십이 이직에 미치는 영향에 관한 연구. 석사학위논문, 한양대학교 경영대학원.
- 이용상·신동광(2020). 코로나19로 인한 언택트 시대의 온라인 교육 실태 연구. *교육과정평가연구*, 23(4), 39-57.
- 오상은·박재현·최창환(2020). 체육학에서 노인 관련 연구의 동향 탐색: 고령화 및 고령사회에 따른 키워드 네트워크 분석. *한국체육측정평가학회지*, 22(4), 111-121.
- 윤문길·장윤혜(2020). COVID-19 이후 우리나라 항공시장 회복과 정 연구. *한국항공경영학회지*, 18(4), 89-106.
- 윤승환(2021). 코로나19(COVID-19)의 동북아시아 3국에 대한 파급효과: GDP와 서비스 산업을 중심으로. *경상논총*, 39(2), 97-113.
- 윤옥한(2020). 코로나19 이후 교양 교육 방향 탐색. *교양교육연구*, 14(4), 25-34.
- 윤경재(2021). 코로나19 팬데믹 상황에서 의료관광 산업의 발전 방

- 안에 관한 탐색적 연구. *The Journal of the Convergence on Culture Technology (JCCT)*, 7(4), 577-582.
- 윤진아·김수진·서혜애(2016). 2010~2015년 사회네트워크분석 (SNA) 방법 활용 국내외 영재교육 연구동향 분석. *영재교육연구*, 26(2), 347-363.
- 정민경·최규환(2021). 텍스트 마이닝 기법을 활용한 COVID-19 발생 이전·이후의 배달주문플랫폼 서비스에 대한 주요 토픽 분석. *지역산업연구*, 44(3), 283-305.
- 전소라·김호영(2021). 장기화된 COVID-19 에서 경험한 스트레스가 심리적 고통에 미치는 영향: 사회적 지지의 조절효과. *stress*, 29(4), 207-219.
- 조성범·한송이(2022). 네트워크 텍스트 분석을 활용한 COVID-19 이후 대학교육 연구 경향. *열린교육연구*, 30(6), 281-305.
- 조수곤·김성범(2012). 텍스트마이닝을 활용한 산업공학 학술지의 논문 주제어간 연관관계 연구. *대한산업공학회지*, 38(1), 67-73.
- 조대형·김정주(2020). 코로나-19 사태가 국내경제에 미치는 영향과 향후 과제. *인문사회*, 21, 11(6), 2597-2612.
- 조군위·박현정(2022). COVID-19 불안감과 식생활 라이프스타일에 따른 드론 음식배달 서비스에 대한 소비자 태도. *디지털융복합연구*, 20(2), 175-184.
- 조주은(2022). Covid-19 상황에서 항공서비스 전공 대학생들의 진로준비이행과 취업 스트레스에 대한 경험 탐색. *호텔관광연구*, 24(3), 115-128.
- 최보윤·윤희수·성원준·조금준·나성훈·정영미·고현선·박종신(2022). 임신 중 COVID-19 백신 예방접종에 대한 의료인의 인지도, 수용도 및 권장 요인 연구. *한국모자보건학회지*, 26(4), 245-253.
- 최상수·조주은(2021). Covid-19상황에서 서비스산업 종사원의 고용 불안이 조직시민 행동에 미치는 영향: 그릿의 조절효과를 중심으로. *관광레저연구*, 33(8), 311-331.
- 최연석(2020). 원격의료의 도입에 관한 연구: 코로나바이러스감염증 19 전염병과 원격의료 도입의 필요성. *국제법무*, 12(1), 113-137.
- 최동현·송보미·박다현·이성우(2022). 텍스트마이닝을 활용한 Covid-19 기간 동안의 항공산업 관련 키워드 트렌드 분석. *한국산업정보학회논문지*, 27(2), 115-128.
- 함승경·김혜정·김영옥(2021). 코로나19 언론보도 경향에 대한 빅데이터 분석: 이슈 주기 및 언론사 정치적 지향에 따른 주제 분석과 언어 네트워크 분석 적용. *한국언론학보*, 65(1), 148-189.
- 현보람·김경식(2022). 사회연결망 분석 및 토픽 모델링을 활용한 한국 무용학의 연구동향 규명. *Journal of Convergence and Consilience*, 5(3), 283-299.
- Batista, F., & Carvalho, J. P.(2015). Text based classification of companies in CrunchBase. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-7). IEEE.
- Blei, D. M., Ng, A. Y., & Jordan, M. I.(2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Choi, D. H., Song, B. M., Park, D. H., & Lee, S. W.(2022). Keyword trends analysis related to the aviation industry during the Covid-19 period using text mining. *Journal of the Korea Industrial Information Systems Research*, 27(2), 115-128.
- Choi, S. S., & Cho, J. E.(2021). Effect of Job Insecurity of Employees in Service Industry on the Organizational Citizen Behavior under Covid-19 Situation: Focusing on the Moderating Effect of Grit. *Journal of Tourism and Leisure Research*, 33(8), 311-331.
- Choi, Y. S.(2020). A Study on the Introduction of Telemedicine: Coronavirus Disease 2019 and the Need for the Introduction of Telemedicine. *International Law Review*, 12(1), 113-137.
- Cho, D. H., & Kim J. J.(2020). The Impact of the COVID-19 Pandemic on the Domestic Economy and Future Challenges. *The Journal of Humanities and Social science*, 11(6), 2597-2612.
- Cho, S. G., & Kim, S. B.(2012). Finding meaningful pattern of key words in IIE transactions using text mining. *Journal of Korean Institute of Industrial Engineers*, 38(1), 67-73.
- Cho, S. B., & Han, S. I.(2022). Academic Research Trends on College Education After the COVID-19 Pandemic. *The Journal of Yeolin Education*, 30(6), 281-305.
- Cho, J. E.(2022). An exploratory research on the implementation of career preparation and experience of job-seeking stress of the university students of majoring in aviation service during Covid-19. *Journal of Hospitality and Tourism Studies*, 24(3), 115-128.
- Choi, B. Y., Yoon, H. S., Seong, W. J., Cho, G. J., Na, S. H., Jung, Y. M., Ko, H. S., & Park, J. S.(2022). Knowledge and Acceptability of the COVID-19 Vaccine Among Obstetricians and Gynecologists in Korea, and Factors Affecting Its Recommendation in the Pregnant Population. *Journal of the Korean Society of Maternal and Child Health*, 26(4), 245-253.
- Dalle, J. M., Den Besten, M., & Menon, C.(2017). *Using Crunchbase for economic and managerial research*, OECD Science. Technology and Industry Working Papers.
- Douzi, S., Amar, M., El Ouahidi, B., & Laanaya, H.(2017). Towards a new spam filter based on PV-DM (paragraph vector-distributed memory approach). *Procedia Computer Science*, 110, 486-491.
- Daniel, S. J.(2020). Education and the COVID-19 pandemic. *Prospects*, 49(1), 91-96.
- Felgueiras, M., Batista, F., & Carvalho, J. P.(2020). Creating classification models from textual descriptions of companies using crunchbase, In *Information Processing and Management of Uncertainty in Knowledge-Based Systems: 18th International Conference, IPMU 2020, Lisbon, Portugal, June 15-19, 2020, Proceedings, Part I 18* (pp. 695-707). Springer International Publishing.
- Hyun, B. R., & Kim, K. S.(2022). Investigating Research Trend of Korean Dance Studies Using Social Network Analysis and Topic Modeling. *Journal of Convergence and Consilience*, 5(3), 282-299.
- Ham, S. K., Kim, H. J., & Kim, Y. W.(2021). A Big-Data Analysis of Media Coverage on COVID-19: Topic Modeling and Semantic Network Analyses by Issue Cycle and Political Orientation. *Korean Journal of Journalism & Communication Studies*, 65(1), 148-189.
- He, P., Sun, Y., Zhang, Y., & Li, T.(2020a). COVID-19's impact on stock prices across different sectors-An event

- study based on the Chinese stock market. *Emerging Markets Finance and Trade*, 56(10), 2198-2212.
- He, P., Niu, H., Sun, Z., & Li, T.(2020b). Accounting index of COVID-19 impact on Chinese industries: A case study using big data portrait analysis. *Emerging Markets Finance and Trade*, 56(10), 2332-2349.
- Jung, S. J.(2022). A study on changes in the food service industry about keyword before and after COVID-19 using big data. *International Journal of Internet, Broadcasting and Communication*, 14(3), 85-90.
- Jeon, S. R., & Kim, H. Y.(2021). The Effects of Stress Experienced in the Prolonged COVID-19 Situation on Psychological Distress: The Moderating Effect of Social Support. *The Korean Journal of Stress Research*, 29(4), 207-219.
- Jeong, M. K., & Choi, K. W.(2021). Analysis of Major Topics for Platform Services for Delivery Orders before and after Covid19 with the Use of Text Mining Techniques. *Regional Industry Review*, 44(3), 283-305.
- Kim, D. W., & Koo, M. W.(2016). Categorization of Korean News Articles Based on Convolutional Neural Network Using Doc2Vec and Word2Vec. *Proceedings of the Korean Society of Computer Information Conference*, 67-71.
- Kim, S. M.(2020). Analysis of press articles in Korean media on online education related to COVID-19. *Journal of Digital Contents Society*, 21(6), 1091-1100.
- Kim, S. K., & Jang S. Y.(2016). A Study on the Research Trends in Domestic Industrial and Management Engineering using Topic Modeling. *Journal of the Korea Management Engineers Society*, 21(3), 71-95.
- Kim, K. K.(2020). Major Challenges and Solutions for Korean Higher Education in the Era of Change. *Korean journal of sociology of education*, 30(4), 1-32.
- Kim, E. M., Shi, B., & Min, Y. K.(2022). A Study on Organizational Resilience of Social Economy in COVID-19 Situation: Focusing on Cooperatives and Social Cooperatives in Education Service Industry for the Developmentally Disabled. *Social Enterprise Studies*, 15(1), 165-209.
- Kim, J. O.(2022). The effect of job satisfaction of service industry workers on life satisfaction in the context of the COVID-19 pandemic: Focusing on the mediating effect of family support. *International Journal of Tourism and Hospitality Research*, 36(7), 97-108.
- Kim, K. W., Han, S. H., Lee, J. S., Kim, N. Y., & Oak, J. W.(2020). Analysis of domestic education trends after COVID-19: focusing on non-face-to-face education. *Journal of Next-generation Convergence Information Services Technology*, 9(4), 377-390.
- Kim, S. H.(2021). The Analysis of Global Research Trends in Education in the Era of COVID-19: The Asia Pacific Region. *Asia Review*, 11(2), 187-215.
- Kwon, Y. H.(2022). Long COVID and Aviation Safety. *The Korean Journal of Aerospace and environmental Medicine*, 32(3), 80-84.
- Kim, K. J.(2021). *Effect of family restaurant delivery service quality factors in COVID-19 on brand satisfaction, brand trust, brand loyalty*. Master's thesis, The Graduate School of Tourism Sejong University.
- Kim, J. S., & Lee, H. R.(2022). A Comparative Analysis on Public Perception about Hotel before and after the Covid-19 Outbreak using Big Date on Social Media: Text Mining and Semantic Network Analysis. *Korean Journal of Hospitality & Tourism*, 31(4), 1-15.
- Kim, Y. H., Jeong, J. H., Kang, D. B., Park, K. M., & Kim, S. M.(2015). Trend Analysis of Research Topics in Journal of Lifelong Learning Society: Using Network Text Analysis. *Journal of Lifelong Learning Society*, 11(1), 291-315.
- Kim, J., Jeon, W., & Geum, Y.(2023). Industry Convergence for Startup Businesses: Dynamic Trend Analysis Using Merger and Acquisition Information. *IEEE Transactions on Engineering Management*, 70(4), 1468-1489.
- Lee, S. S.(2016). A Study on the Library Marketing Research Trends through Keyword Network Analysis: Comparative Analysis of Korea and Other Countries. *Journal of the Korean Society for Library and Information Science*, 50(3), 383-402.
- Lee, S. Y., Yun, S. W., Park, J. S., Lee, J. Y., Kim, N. H., Moon, S. Y., & Kim, N. Y.(2022). Influencing Factors the Knowledge and Fear of COVID-19 on Job-Seeking Anxiety of Nursing Students. *The Korean Society of Applied Science and Technology*, 39(3), 407-416.
- Lee, W. H.(2020). Strategy for Overcoming Environmental Change Risks Facing the Distribution and Logistics Industry Caused by COVID-19 Pandemic. *Journal of Distribution and Management Research*, 23(3), 81-95.
- Lee, C. Y.(2007). *A Study on the Effects of Job Stress and Superior Leadership on Turnover*. Master's thesis, Graduate School of Business Administration Hanyang University.
- Lee, Y. S., & Shin D. K.(2020). An Investigation of the Implementation of Online Classes in the Untact Era Caused by the COVID-19 Pandemic. *The Journal of Curriculum Evaluation*, 23(4), 39-57.
- Lee, Y., & Geum, Y.(2023). Identifying patterns of mergers and acquisitions in startup: An empirical analysis using Crunchbase data. *IEEE Access*, 11, 42463-42472
- Liang, Y. E., & Yuan, S. T. D.(2016). Predicting investor funding behavior using crunchbase social network features. *Internet Research*, 26(1), 74-100.
- Moon, S. H., Chung, S. H., & Chi, S. H.(2018). Topic Modeling of News Article about International Construction Market Using Latent Dirichlet Allocation. *Journal of The Korean Society of Civil Engineers*, 38(4), 595-599.
- Moon, M. J., & Koh, I. K.(2022). Study on Air Logistics Service Provider's Performance Under Post Covid19 Situation: Focusing on Trust and Reciprocity. *Asia-Pacific Journal of Business Venturing and Entrepreneurship*, 17(3), 135-145.
- Oh, S. E., Park, J. H., & Choi, C. H.(2020). Research Trends in Senior Citizen studies in Physical Education and Exercise Science: Knowledge Network Analysis of Keywords in the Aging and Aged Society in Korea.

- The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 22(4), 111-121.
- Park, J. T.(2022). Analysis of Research Trends in Physical Education for the Disabled after COVID-19 Using Text Mining. *Korea Sport Society*, 20(3), 241-249.
- Park, H. R., & Lee, S. J.(2022). Analysis of Anxiety, Stress and Coping Type of Health-Care Worker in the COVID-19 Pandemic. *Journal of the Korea Academia-Industrial cooperation Society*, 23(2), 57-66.
- Park, M. Y., Jeong, S. H., Kim, H. S., & Lee, E. J.(2022). Images of Nurses Appeared in Media Reports Before and After Outbreak of COVID-19: Text Network Analysis and Topic Modeling. *Journal of Korean Academy of Nursing*, 52(3), 291-307.
- Park, J. S.(2017). A Study on the Domestic Research Trends Related with Education for Persons with Disabilities Through a Keyword Network Analysis. *Journal of Special Education & Rehabilitation Science*, 56(2), 107-126.
- Persada, S. F., Prasetyo, Y. T., Suryananda, X. V., Apriyansyah, B., Ong, A. K., Nadlifatin, R., Setiyati, E. A., Putra, R. A. K., Purnomo, A., Triangga, B., Siregar, N. J., Carolina, D., Maulana, F. I., & Ardiansyahmiraja, B.(2022). How the Education Industries React to Synchronous and Asynchronous Learning in COVID-19: Multigroup Analysis Insights for Future Online Education. *Sustainability*, 14(22), 15288.
- Park, C. W., & Kang, G. L.(2020). A Study on the Impacts of Entrepreneurship and Transformational Leadership on Employment and Startup Awareness : Mediating Effects of Corona 19 Social Issue. *Asia-Pacific Journal of Business Venturing and Entrepreneurship*, 15(5), 51-63.
- Savin, I., Chukavina, K., & Pushkarev, A.(2023). Topic-based classification and identification of global trends for startup companies. *Small Business Economics*, 60(2), 659-689.
- Song, H. G.(2022). A study on the trends of meal-kits industry using big data: focusing on comparison of time before and after the outbreak of COVID-19. *Journal of Foodservice Management*, 25(5), 433-457.
- Seo, E. G.(2010). Trends Analysis on Research Articles in the Journal of Korean Society for Information Management. *Journal of the Korean Society for Information Management*, 27(4), 7-32.
- Spoorthy, M. S., Pratapa, S. K., & Mahant, S.(2020). Mental health problems faced by healthcare workers due to the COVID-19 pandemic: A review. *Asian journal of psychiatry*, 51, 102119.
- Scott, J.(2012). What is social network analysis?. *Bloomsbury Academic*.
- Shin, Y. J.(2020). The Role and Economic Effects of the Aviation Transportation Service Industry on the Korean Economy: To cope with the COVID-19 Pandemic. *Journal of the Aviation Management Society of Korea*, 18(3), 115-136.
- Song, M. K., Lee S. J., & Cho, J. H.(2022). Analysis for Economic Impacts of the Tourism Industry in Daegu and Gyeongbuk after the COVID-19 Outbreak: Using Regional Input-Output Analysis. *The Korean Society of Management Consulting*, 22(6), 239-277.
- Song, J. J., & Kang, S. S.(2019). Automatic Classification of Patent Documents Using Doc2Vec. In *Proceedings of the Korea Information Processing Society Conference*, 26(1), 239-241.
- Yoon, K. J.(2021). An exploratory study on the development plan of the medical tourism industry in the context of the COVID-19 pandemic. *The Journal of the Convergence on Culture Technology (JCCT)*, 7(4), 577-582.
- Yoon, O. H.(2020). Exploring the Direction of Liberal Arts Education After the Coronavirus Pandemic. *Korean Journal of General Education*, 14(4), 25-34.
- Yoon, S. H.(2021). COVID-19 Impact on Three Northeast Asian Countries: Focusing on GDP and Service Industries. *Koreanische Zeitschrift fuer Wirtschaftswissenschaften*, 39(2), 97-113.
- Yoon, M. G., & Chang, Y. H.(2020). Air Travel Market Recovery Process for the post COVID-19 Era. *Journal of the Aviation Management Society of Korea*, 18(4), 89-106.
- Yoon, J. A., Kim, S. J., & Seo, H. A.(2016). Investigating Trends of Gifted Education in Domestic and Foreign Countries through Social Network Analysis from 2010 to 2015. *Journal of Gifted/Talented Education*, 26(2), 347-363.
- Yang, L., Li, X., & Hernández-Lara, A. B.(2022). Scientific collaboration and thematic analysis of the tourism industry in the context of COVID-19: a bibliometric approach. *International Journal of Contemporary Hospitality Management*, (ahead-of-print).
- Zhao, J. W., & Park H. J.(2022). The Influence of the COVID-19 Anxiety and Dietary Lifestyles on the Drone Food Delivery Service Attitude. *Journal of Digital Convergence*, 20(2), 175-184.

Analysis of Global Entrepreneurship Trends Due to COVID-19: Focusing on Crunchbase*

Shinho Kim**
Youngjung Geum***

Abstract

Due to the unprecedented worldwide pandemic of the new Covid-19 infection, business trends of companies have changed significantly. Therefore, it is strongly required to monitor the rapid changes of innovation trends to design and plan future businesses. Since the pandemic, many studies have attempted to analyze business changes, but they are limited to specific industries and are insufficient in terms of data objectivity. In response, this study aims to analyze business trends after Covid-19 using Crunchbase, a global startup data. The data is collected and preprocessed every two years from 2018 to 2021 to compare the business trends. To capture the major trends, a network analysis is conducted for the industry groups and industry information based on the co-occurrence. To analyze the minor trends, LDA-based topic modelling and word2vec-based clustering is used. As a result, e-commerce, education, delivery, game and entertainment industries are promising based on their technological advances, showing extension and diversification of industry boundaries as well as digitalization and servitization of business contents. This study is expected to help venture capitalists and entrepreneurs to understand the rapid changes under the impact of Covid-19 and to make right decisions for the future.

KeyWords: Covid-19, Business Trend, Trend Analysis, Entrepreneurship Trends, Industry Trends, Crunchbase

* This study was conducted with the support of the Korea Research Foundation's mid-sized research (NRF-2023R1A2C1004752) and was conducted with the support of the Korea Industrial Technology Promotion Association in 2023 (P0017123, 2023 Industrial Innovation Talent Growth Support Project).

** First Author, Master's Degree, Data Science, Seoul National University of Science and Technology(SeoulTech), shk240@seoultech.ac.kr

*** Corresponding, Department of Industrial Engineering/Department of Data Science, Seoul National University of Science and Technology(SeoulTech), yjgeum@seoultech.ac.kr