

논문 2022-2-19 <http://dx.doi.org/10.29056/jsav.2022.12.19>

# 빅데이터 분석을 활용한 창업의 긍정적 ESG 이미지 요인 분석

김대유\*, 배장원\*\*†

## Analyzing the positive ESG image factors of startups using big data analysis

Dae-Yu Kim\*, Jang Won Bae\*\*†

### 요 약

본 연구의 목적은 4차 산업혁명 시대의 빅데이터 분석 기술을 활용한 창업의 긍정적 ESG 이미지 요인들을 분석하여 선행 학습의 교육 및 준비에 긍정적인 솔루션을 제공한다. ESG의 측면에서 경영 활동이 사회의 지속적인 발전에 기여해야 한다는 ESG(Environmental-Social-Governance) 창업 관련 데이터를 빅데이터 클라우드 자료를 소프트웨어로 분석하고, 결과자료를 토대로 안정적인 경영 솔루션을 제공한다. 연구 결과 창업을 준비하는 MZ세대의 주요 키워드는 기업, 지원, 사업 및 사회, 투자, 경영 등의 순으로 확인되어 창업 준비에 신중한 고려를 제안한다. 창업을 준비하는 MZ세대원들에게 실패를 최소화하고 빅데이터 기술경영으로 사회적 주요 이슈인 고용 해결에 이바지하도록 한다. 향후 연구 방향은 주요 키워드를 변수 지정하여 조절 및 매개효과의 상관관계, 요인 분석, 신뢰도 분석, 회귀 분석 등으로 ESG 창업의 긍정적인 영향에 대한 증명이 필요할 것이라고 본다.

### Abstract

The purpose of this study is to analyze the positive ESG image factors of start-ups using big data analysis technology in the era of the 4th industrial revolution to provide positive solutions for education and preparation of prior learning. In terms of ESG we provide a stable management solution based on the results of big data cloud word analysis through Python coding of ESG(Environmental-Social-Governance) start-up related data that management activities should contribute to the continuous development of society. As a result of the research, the main keywords of the MZ generation preparing to start a business are identified in the order of company, support, business and society, investment, management, etc., suggesting careful consideration in preparing for starting a business. For MZ generation members preparing to start a business, we will minimize failures and contribute to solving employment, which is a major social issue, through big data technology management. In the future research direction, it will be necessary to prove the positive impact of ESG start-ups by specifying key keywords as variables, correlation of adjustment and mediating effects, factor analysis, reliability analysis, and regression analysis.

**한글키워드** : 빅데이터, 파이썬, 코딩, ESG, 창업, 경영

**keywords** : Big Data, Python, Coding, ESG, Entrepreneurship, Management

\* 남서울대학교 빅데이터산업보안학과

접수일자: 2022.12.11. 심사완료: 2022.12.13.

\*\* 한국기술교육대학교 산업경영학과

게재확정: 2022.12.20.

† 교신저자: 배장원(email: jangwon\_bae@koreatech.ac.kr)

## 1. 서론

정보통신 기술의 발전으로 산업 간 경계가 모호해지면서 융합 현상이 새로운 기술혁신의 패러다임으로 떠오르고 있다. 인공지능(AI), 빅데이터, 블록체인, 산업용 사물인터넷(IIOT) 등 ICT 기술 기반의 디지털화(Digital Transformation)로 대표되는 4차 산업혁명이 주목받고 이러한 4차 산업혁명 기술발전이 진행됨에 따라 기업들은 다양한 신규사업 기회를 접할 것으로 전망되고 해외시장 진출, 관련·비 관련 다각화를 위한 M&A, 신규사업 직접투자, 신기술 관련 투자, 혁신역량 확보 등 지속적인 성장을 위한 다양한 전략적 의사결정을 진행하고 있다[1]. 제4차 산업혁명의 시작과 알림은 2016년 1월 제46회 세계경제포럼인 다보스포럼에서 준비되고 시작했다. 유사한 혁신 모두가 시작 단계로서 물리학, 아날로그에서 디지털 변화, 생물학 분야의 경우 기술과 안전 융합을 기본으로 각각의 분야를 발전시키는 성장의 시작점에 가까워졌다고 선언하였다. 스마트팜 시스템중 설비가 융합된 원예와 과수에 활용되는 4차 산업기술은 IIOT, 클라우드 서비스, 빅데이터 활용, AI, 드론서비스, 로봇 등이 있다.

본 연구는 4차 산업시대의 빅데이터 분석 기술을 활용한 벤처기업의 긍정적 기업 이미지 요인들을 분석하여 예비 창업자가 성공적인 창업의 외부적 요건에 대하여 준비하고 선행연구를 해야 하는 교육 및 예산 편성에 미치는 영향을 분석하였다.

## 2. 이론적 배경

### 2.1 이미지

이미지란 사물이나 사람의 개성과 본질이 다양한 통찰과 연상을 통해 사고하고 인식되는 것으로 다른 사람의 가치관과 패러다임 등에 의해

규정이 정해진다. 어떤 대상에 대한 이미지는 일반적으로 감각적 인식과 심상의 느낌으로 구체적인 이미지의 개념을 갖기 위해서는 시각적이고 개념적인 과정이 필요하다. 이미지는 이처럼 마음속에 그려지는 영상이지만 사람의 태도에 영향을 미치며 이미지는 더 나아가 다른 행동에도 영향을 미치기 때문에 기업 대표이사의 이미지와 벤처기업 이미지는 특히 현대 기업에서 관심을 크게 가지게 되는 대상이 되고 있다[2].

사람들은 한 번 어떤 대상에 대해 이미지 개념을 형성하면 그 대상이나 사물에 대해 객관적인 지식과 정보에 의존하지 않고 관념화된 이미지에 따라 반응하게 되며, 이렇게 각자 인식한 이미지는 대상을 대체하게 되어 각 사물과 현상에 대해 개인의 반응에 크게 영향을 미치게 된다. 결국, 사람들의 행동은 각자의 내면에 형성된 세계에 대한 인상, 즉 이미지에 반영되고 각 이미지에 따라 개념이 설정되는 것이다. 그리고 이렇게 개념화된 이미지는 본질에서 인간의 다양한 소통 때문에 형성되고 수정되어 결국 변화되어 가는 것이다. 사회적 책임의 하위요인 중 경제적 책임, 윤리적 책임, 사회적 기부는 기업 이미지의 유의한 영향을 미치는 것으로 나타났다[3]. 기업은 시장 환경에서 규제를 준수하고 소비를 촉진시켜야 하는 의무를 가지고 있으므로 기업의 사회적 책임을 강조하고 있다. 기업의 사회적 책임활동(CSR)이 언급되면서 소비자들이 기업의 다양한 문제에 관하여 부정적인 평가를 하고 있다. 기업의 사회적 책임활동은 유용한 비재무적 정보를 자본시장에 전달함으로써 정보 비대칭성 문제를 완화시켜 기업의 고유위험을 감소시키는 것으로 알려져 있다[4].

일반적으로 재무적인 관점과 사회적인 관점에서 기업 성과를 판단한다. 2014년 유럽의회는 500인 이상 종사하는 기업은 관련된 다양한 정보들의 의무적으로 공개해야 하는 법안을 발의하였다. 기

업의 자선 활동은 운영하는 목적에 있어서 중요한 요인이 되었다. 세계 시장은 기업들의 사회적 책임활동에 대한 경영 투명성을 요구하고 있다. 따라서 국내 기업들은 사회적 책임활동을 체계적으로 준비해야 하는 시점에 직면해 있다[5].

그림 1은 글로벌 브랜드 컨설팅 그룹인 인터브랜드에서 발표한 글로벌 10대 브랜드로서 리더쉽, 참여, 관련성과 더불어 재무분석 및 브랜드의 역할과 강도를 분석하여 이미지 가치순위에 적용하여 발표하였다. 기업의 사회적 책임은 규범과 가치에 부합하여 의무를 다하는 것으로 볼 수 있다.



그림 1. 베스트 글로벌 브랜드 2021  
Fig. 1. Best global brands 2021

## 2.2 빅데이터

빅데이터는 일반적으로 데이터베이스 관리 시스템으로 데이터의 보관 분석 최적의 관리를 할 수 있는 범위를 넘어서는 매우 큰 데이터로서 다양한 조율의 매우 큰 데이터로부터 전략적으로 주요한 부분을 발굴하고 데이터의 초 신속 초 정확 초 안전하게 모으고, 찾아내며, 분석하도록 지원 가능토록 고안된 4차산업혁명시대의 기술로서 보통 수십에서 수천 TB(테라바이트) 이상의 매우 큰 크기를 소유하고 있으며, 여러 가지 비정형 데이터를 포함한 데이터 생성, 유통 소비가 매우 신속한 시간 단위로 변화되어 기존 방식으

로는 관리와 분석이 쉽지 않은 데이터의 집합이다. 설문조사와 사용자 인터뷰 등 정성 정량 연구와 같은 기존의 고전적 디자인 방법론은 한정된 표본을 기반으로 심층연구를 한다는 측면에서 전체 사회와 시장의 사용자 집단을 대표하여 이해하고 상호작용의 문제를 도출하는 데는 한계가 있을 뿐 아니라 조사 대상이 커짐에 따른 상대적 비용과 시간적 효율성도 낮다.

빅데이터 기반의 디자인 분야에서 데이터 자원을 효과적으로 활용하고, 빅데이터 기술과 디자인이 상생 발전할 수 있는 디자인 연구 방향을 제시하였다[6]. 빅데이터 분석에 기반한 정량적인 트렌드 연구는 의사결정의 정확도와 합리성을 높여주고, 가까운 미래를 예측하기 위한 Nowcast가 가능하며, 상황분석을 통한 새로운 비즈니스 기회 창출에 유용하게 활용되고 있다[7]. 거시적 트렌드의 동인은 사회 구성원들 변화의 총량이라는 점에서 소셜미디어 등에 나타난 개인의 언어적 집합체인 빅데이터 분석을 통하여 이러한 사회 변화 흐름을 이해하고 유의미한 통찰을 얻을 수 있다[8].

반도체 사업장에서 사용되는 IoT Gas sensor로부터 Monitoring 기술이 적용된 시스템에서 온도 습도 데이터, Location, 적외선 및 자외선 등의 빅데이터를 의사결정 트리, 선형회귀 빅데이터 분석결과를 토대로 Alarm 발생될 때 습도와 선형 연관성을 근거로 습도를 40~60RH%로 지속해서 유지하게 AI 시스템을 적용하여 Alarm 발생률을 최소화하는 빅데이터 기술이 적용된 사례도 있다.

## 3. 연구 방법

### 3.1 빅데이터 분석과정

세계에서 수집되는 대부분 데이터는 분석에

부적합한 형태로 존재하고 있으며, 불안정성 (incomplete)으로 일부 특성이나 값 손실 혹은 기록이 되지 않기에 불필요한 데이터(noise)의 오류(error) 혹은 이상치(outlier)가 존재하기 때문에 분석에 적합한 형태의 데이터를 적합한 형태로 변환하는 과정이 필요하다.

적합한 형태의 데이터 변환을 위해서는 첫째 데이터 전처리(Data cleaning)로서 데이터에서 손실된 값을 채우거나 이상치를 제거하고 오류 평활화(smooth) 혹은 비일관성 해결이 필요하다. 예를 들어 보간법(interoperation)과 binning method 기술 등이 있다.

그림 2의 빅데이터 분석방법을 보면 기존의 통계적 관점인 데이터 특성을 분석하는 상관, 회귀 분석 등으로 분석용 데이터 준비과정에서 많이 사용되고 있으며, 데이터가 적을 때는 딥러닝보다 유리하다. 기계학습 분석은 데이터로부터 학습한 결과로 미래를 예측하는 지도학습, 비지도 학습, 강화학습으로 최근 딥러닝의 정확도 향상으로 많이 활용된다. 시뮬레이션 분석으로는 과거 혹은 미래 현상의 원인 분석 또는 가설에 대한 효과와 결과를 예측하여 과학적인 정책을 수립하고 미래변화 예측 등의 방법론으로 활용된다.

둘째 데이터 융합(Data Integration)으로서 다수의 데이터를 이용하여 일관성 있는 데이터 생성과 중복 데이터(redundant data) 제거가 필요하다. 예를 들어 스키마 융합(Schema Integration)과 상관분석 기술 등이 있다. 마지막으로 데이터 변환(Data Transformation)으로 기존 데이터의 전체 값은 새로운 값으로 대체한다. 예를 들어 정규화(normalization), IPF(Iterative Proportional Fitting) 기술 등이 있다.

### 3.2 빅데이터 분석방법

고전적 빅데이터 분석으로는 기존의 통계적 관점에서의 평균, 샘플링, 확률 등으로 데이터의 특성을 상관분석(Correlation analysis), 회귀분석(regression analysis)등으로 분석용 데이터 준비 과정에 많이 사용되고 있으며, 데이터가 적을 때는 딥러닝보다 유리한 분석 기술이다. 4차산업혁명시대의 빅데이터를 분석하기 위한 기계학습 분석으로서는 데이터로부터 지도학습(supervised) 비지도 학습(unsupervised) 그리고 강화 학습(reinforcement)한 결과로 미래를 예측한다. 시뮬레이션 분석으로는 과거 혹은 미래 현상의 원인 분석 또는 가설에 대한 효과와 결과를 거시적 미시적으로 예측한다. 예를 들어 과학기술 정책 수립과 미래변화 예측 등의 방법론에 활용된다.

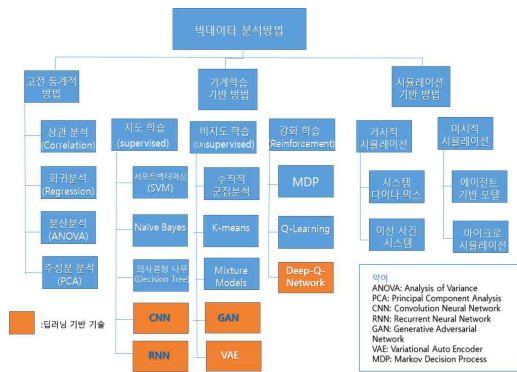


그림 2. 빅데이터 분석 방법  
Fig. 2. Big data analysis methods

### 3.3 빅데이터 분류분석 알고리즘

#### 3.3.1 의사결정트리

최적화된 답을 얻어내는 데 사용하는 근사적인 방법으로 여러 가지의 선택에 필요한 최적의 하나를 높은 가능성과 학습으로 판단해서 선택해 나가는 방식으로 진행하여 최종 해답에 도달하고 불순도(Impurity)라는 기준을 최소화하는 방향으로 성장시키며 단점으로는 과거의 데이터의 세세한 것까지 맞추려고 하는 과잉적합(Overfitting)이 발생할 수 있으나 학습의 노드의 수를 최소화

하여 단점을 극복할 수 있다. 의사결정 트리 생성 주요사항으로는 트리 생성 시 어떤 속성을 노드로 사용할 것인지, 분기 형태를 어떻게 할 것인지, 트리 성장을 언제 중지할 것인지 결정해야 하고, 트리에 노드를 너무 많이 붙이는 경우, 트리를 만드는 시간이 길어지고, 미래 데이터 분류를 위해, 트리의 경로를 탐색하는 시간이 길어짐에 따른 트리를 직관적으로 해석하기 어려워지고, 미래 데이터에 대한 예측 정확도가 떨어지는 과잉적합(Overfitting)이 발생하면, 오히려 미래 데이터에는 잘 안 맞을 가능성이 존재하므로 과거 데이터의 일반적인 추세만 트리로 나타내고, 특이한 경우들에 대해서는 고려하지 않는 노드수를 적게 하는 분석방법으로 진행하면 분석 효율을 높일 수 있다.

그림 3의 대표적인 빅데이터 분류 분석방법인 의사결정 분석방법으로 붓꽃의 잎에 대하여 분류된 150개 인스턴스 데이터에 대하여 분류분석을 수행하여 꽃받침, 잎의 길이와 너비 값을 기록한 자료를 수집해 붓꽃의 유형을 분류할 수 있는 분류 모델을 수립한다. 꽃잎의 너비가 값이 0.6보다

작거나 같으면 아이리스 세트사로 분류하고 0.6보다 크면서 1.7을 초과하면 아이리스 버지니카로 분류한다.

### 3.3.2 인공신경망

여러 처리장치를 사용하여 같은 시간에 처리장치를 이용하는 특징으로 하는 정보처리 시스템으로, 상호작용을 통한 접속되어 처리하는 요소들로 구성되었으며, 정보처리 구조로서 뇌의 뉴런들이 상호 연결되어 학습으로 결과를 도출하는 양성하는 생물학적 움직임을 모형화한 것으로, 인간 두뇌의 뉴런 구조를 모형화한 인공 뉴런들을 링크로 연결한 네트워크 형태를 가지고 인공신경망은 학습을 통해 특정 사건으로부터 일반화하는 능력으로 계층 수, 출력 형태, 데이터 유형, 학습 방법, 활성화 함수에 따라 다양하게 분류하고 병렬처리, 분산 정보처리, 상호적인 스스로 학습 기술 및 신속 안전 정확히 적응하는 특성을 가진다. 신경망이 커지고 복잡해질수록 더 나은 기능을 수행할 수 있음. 따라서 입력층과 출력층 사이에 새로운 은닉층(Hidden Layer)을 추가한 다층 신경망이 단층 신경망보다 더 나은 성능을 보일 수 있으며, 신경망의 연결 가중치를 조정하는 학습 방법으로 지도학습과 자율학습을 고려하고, 지도학습은 Input으로부터 데이터의 입력이 공급됨에 따라서 필요한 출력 값이 가중치에 의해서 활성화되고 조절하는 스스로 학습으로 원하는 값 없이 학습 데이터만 입력하여 스스로 연결 가중치들을 학습하는 기술로써 패턴 인식 및 분류(Pattern Recognition or Classification), 비선형 회귀 분석(Nonlinear Regression), 조건적 사후확률 추정(Estimation of the Conditional Posterior Probabilities) 분야에 활용되어 일반적으로 신경망의 가장 큰 장점은 높은 예측 능력으로, 신경망은 다른 분류 모델에서는 파악하기 어려운 입력변수와 출력변수 사이의 매우 복잡한

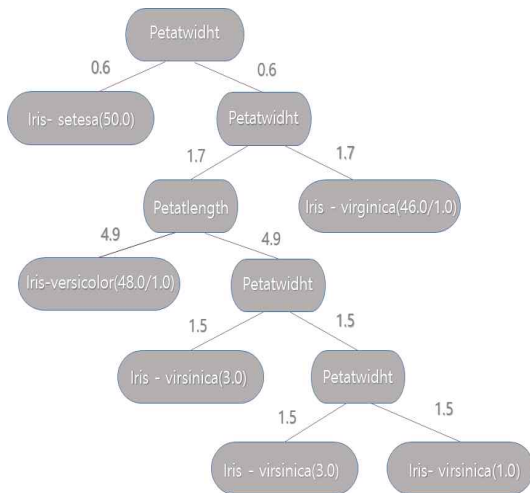


그림 3. 의사결정 분석방법  
Fig. 3. Decision analysis methods

관계를 도출한다.

그림 4는 인공신경망 알고리즘으로 입력  $x$ 의 경우 단순히 입력값을 받아들이는 노드로 구성되고, 중간층의 은닉층 노드는 입력층으로부터 입력 정보를 받아 가중치로 연결하여 출력  $y$ 에게 은닉층에 있는 노드의 출력값으로 변환한다.

입력 노드를 제외한 노드들은 선형 또는 비선형적인 매핑을 수행한다.

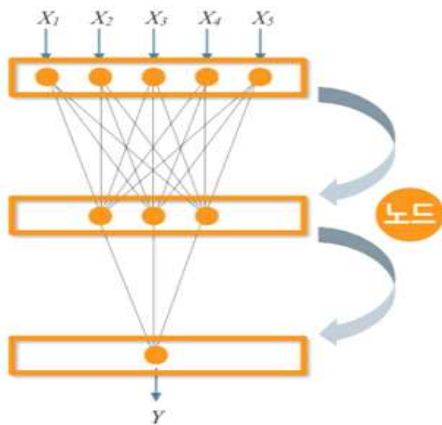


그림 4. 인공신경망 다중네트워크  
Fig. 4. Multilayer networks of artificial neural networks

### 3.3.3 서포트 벡터 머신

기계학습의 분야 중 하나로서 패턴 인식, 자료 분석을 위한 지도학습모델이며, 2개의 결과로 구성되는 데이터 집합이 주어졌을 때, 이를 바탕으로 새로운 데이터가 어느 결과에 속할지 판단할 수 있는 이진 선형 분류 모델 수립한다. 분류 또는 회귀분석에 사용 가능한 최적의 초평면 또는 초평면들의 집합으로 구성하고, 새로운 데이터 포인트가 들어왔을 때 전체 데이터 포인트와의 내적 거리를 구하지 않으며, 서포트 벡터와의 내적 거리만 구하면 되므로 계산 로드를 줄일 수 있다. 검색을 활용하여 기존의 쿼리 보다 상당히 높은 검색 정확도를 확보하는 기술이다.

그림 5는 비선형 서포트 벡터머신 알고리즘으로 클래스 간의 경계가 비선형일 때 분류성능이 저하되므로 유한 차원에서 더 높은 차원인 다항 선형커널 함수를 이용하여 분리성능을 높인다.

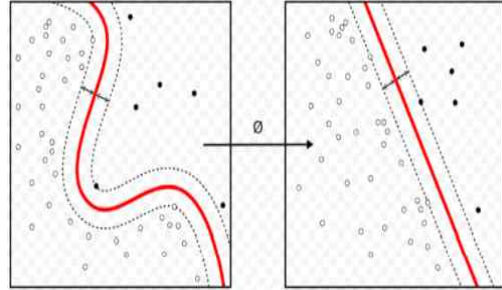


그림 5. 서포트벡터머신의 선형커널분리  
Fig. 5. Linear kernel separation of SVM

### 3.3.4 K-Means 와 EM 군집분석 알고리즘

K-Means 알고리즘은 분할, 배타, 확장, 완전 분할 등의 ahems 레코드를 K개의 군집으로 나누어 객체가 하나의 군집에 독립적으로 속하도록 군집을 형성하거나 군집 간 부분집합이나 중복 없이 상호 배타적으로 존재 분석하고 특정 객체가 특정 군집에 해당하는 것이 확정적으로 결정하며, 모든 객체가 하나의 군집에 독립적으로 속하도록 군집을 형성하는 알고리즘이다.

분석 목적이나 데이터 특성을 대략만 아는 경우에 사전에 분석자 또는 사용자에게 의해 결정 임의로 K값을 결정한다. 가장 널리 사용되는 클러스터링 알고리즘이다. K-Means 알고리즘과 사실상 유사한 접근 방식을 사용하는 EM 알고리즘은 기댓값 최대화 알고리즘으로 관측되지 않는 잠재변수에 의존하는 확률 모델에서 최대 가능도(Maximum Likelihood)나 최대 사후확률(Maximum a Posteriori)을 갖는 매개변수를 찾는 반복적인 알고리즘으로 Maximum Likelihood estimate 방법을 사용해서 관측된 데이터에 알맞은 모델의 변수를 추정한다. K-Means 알고리즘

경우에는 유클리디언(Euclidean) 거리함수 사용하여 거리를 사용하여 적합도를 평가하는 반면에 EM 알고리즘은 Log Likelihood 함수를 사용하여 모델의 적합성을 평가하여 확률을 산출하여 적합도를 평가한다.

그림 6은 K-평균 군집분석 알고리즘으로 서로 크기가 비슷한 군집들을 찾는 경향이 있어 군집간의 크기가 다른 경우에도 같게 찾아서 적용하고 K개의 원하는 클러스터 수를 설정하여 분석한다.

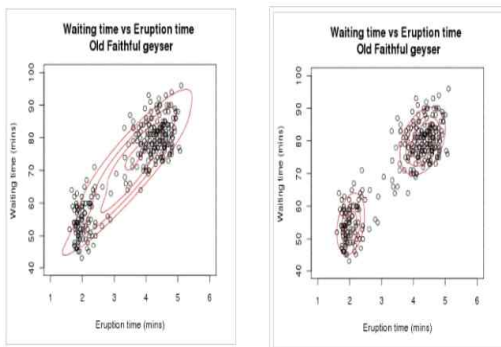


그림 6. 분산적합의 분포 변경  
Fig. 6. Change distribution of variance to fit

### 3.3.5 베이지안 네트워크 알고리즘

베이지안 네트워크는 속성을 의미하는 노드(Node)와 노드 간 독립적인 관계를 가지는 호(Arc)로 구성된 네트워크로서 파라미터 추정 시 베이스 확률론이나 베이지안 방법들을 이용하지 않고도 훈련이 가능하고 분류에 필요한 파라미터를 추정하기 위한 트레이닝 데이터의 양이 매우 적으며, 많은 복잡한 실제 상황에서 분석이 용이하다.

그림 7은 베이지안 네트워크 조건부 확률 계산식으로 관심 있는 사건의 조건부 확률을 계산하는 베이스 정리에 기반하여 확률값을 산출한다. 산출된 평균과 표준편차를 가지는 정규분포의 확률 밀도를 수립할 수 있고 이를 통해서 C가 주어

졌을 때 속성 A에 대한 조건부 확률 P값을 산출한다.

### 조건부 확률 (Conditional Probability)

$$P(C \setminus A) = \frac{P(A, C)}{P(A)} \quad P(A \setminus C) = \frac{P(A, C)}{P(C)}$$

### 베이즈 정리 (Bayes Theorem)

$$P(C \setminus A) = \frac{P(A \setminus C)P(C)}{P(A)}$$

그림 7. 베이즈안 이벤트정리 A,C 조건부 확률  
Fig. 7. Bayes' theorem of events A,C conditional probability

### 3.3.6 로지스틱 회귀 분류분석 알고리즘

로지스틱 회귀모델은 종속 변수가 범주형 데이터인 경우, 해당 데이터의 결과가 특정 분류로 나누어지는 확률을 구하고 독립 변수값에 무관하게 종속 변수 값이 항상 범위 [0,1] 사이에 있으며, 독립 변수 x가 주어졌을 때 종속 변수가 1의 범주에 속할 확률을 의미한다.

그림 8은 로지스틱 선형회귀 함수로 이항 데이터에 적용하였을 때 종속 변수 y의 결과 범위가 0 또는 1로 제한되고 이진적이므로 정규분포 대신 이항 분포를 따른다.

$$\text{logistic function} = \frac{e^{\beta \cdot X_i}}{1 + e^{\beta \cdot X_i}}$$

그림 8. 로지스틱 선형회귀 함수  
Fig. 8. Logistic regression functions

## 4. 연구 방법

### 4.1 데이터 수집

#### 4.1.1 빅데이터 비정형 데이터

한국 언론진흥재단에서 제공하는 빅카인즈는

비정확화된 54개의 언론사의 뉴스 수집하여 뉴스 검색 분류하고 정형화된 3,493개의 빅데이터를 제공 분석한다.

데이터의 수집 기간은 2003년 10월 21일~2022년 10월 15일로서 ESG 창업과 관련된 키워드를 분석하였다.

표 1은 ESG 창업 관련 키워드 빅데이터의 정형화 된 데이터로 파일명은 NESGBigdata.csv이다.

표 1. NESG.CSV 파일  
Table 1. NESG.CSV files

뉴스 식별자	일자	언론사	기고자	제목	통합 분류 인물	키워드
1100201.202	20221015	국민일보	이택현	스타트업 모사경제>취업_창업	스타트업,건설업계,ESG경영,원인,	
2100201.202	20221014	머니투데이	김택현	기자나니아랩스 인IT_과학>과강남우,한나니아랩스,엔서,테크,스타트업,용		
1100401.202	20221014	동아일보	권영관	[스케일업] 문IT_과학>콘김,장,김보:브랜딩,마케팅,목적,스케일업코리		
2100201.202	20221014	머니투데이	류준영	기자오늘 'K-테크 경제>산업 서동희,김,테크,스타트업,왕중왕전,결선,최고		
1400351.202	20221013	중도일보	이유나	기가소진공, 소외 지역>충남 박성호,이동소진공,지원,소외,이웃,배리어프리		
1100801.202	20221013	조선일보	안상현	기자[WEEKLY BIZ]경제>산업 해리,파트너우울증,1400조,제2,팬데믹,정신건		
1101101.202	20221013	한국일보	최연진	SK브로드밴드경제>산업 이주환	SK브로드밴드,지속가능경영,스타	

#### 4.1.2 빅데이터 파이선 분석

파이선은 컴퓨터 그래픽 사용자들이 가장 관심을 갖는 스크립트 언어이다. 사용자가 컴파일 하지 않고 바로 실행할 수 있으며 한 줄 단위로 실행되어서 사용자가 쉽게 결과를 확인할 수 있다는 것뿐만 아니라 파이선에서는 다양한 라이브러리를 사용 가능하다는 점 때문이다. 또한, 파이선은 사용하기 쉽고 강력한 제어 기능을 제공하므로 3Ds Max, Houdini, Blender, Rhino 등 다른 3D 툴들에서도 파이선을 지원하고 있다.[9]

창업에 긍정적 ESG 이미지 분석을 파이선 빅데이터 워드 클라우드로 코딩하여 분석하였다.

표 2의 빅데이터 분석결과 1은 기업, 지원,

ESG, 창업 사업, 사회 순으로 키워드가 언론에 노출되었던 것으로 확인되었다.

표 2. 빅데이터 분석결과 1  
Table 2. Big data analysis results 1

Identifier	word	count
60	기업	15172
374	지원	11234
148	ESG	9724
51	창업	7933
349	사업	7857
44	사회	7391
95	투자	6521
129	경영	6222
227	스타트업	5637
565	대표	4969
146	환경	4938
710	성장	4776
1020	지역	4445
1481	혁신	3964
353	분야	3710
436	산업	3615
32	회장	3557
1066	경제	3350
571	시장	3304

#### 4.2 자료의 특성

웹크롤링(Web Crawling)을 통해 모바일 사용 행태와 액세서리에 관련한 사용자의 댓글 수집(Text Mining)하고, 이를 R Programming을 기반으로 의미형태소관점에서 분류 및 정제한 후, 연관어 분석과정을 거쳐 워드 클라우드(Word Cloud)로 시각화하는 사례도 확인되었다. 블로그 및 소셜 네트워킹 서비스(SNS)와 같은 소셜미디어는 고객 의견과 관련된 풍부하고 좋은 소스의 정보를 실시간 데이터로 제공한다는 측면에서 매우 효율적이며, 최근 사회적 상호작용과 정보 교환을 위한 도구로써 빠르게 제품에 대해 고객이 생생한 목소리를 전달하는 채널로써 데이터양의 폭발적 증가 추세의 측면에서 볼 때 동시대적 관점의 동향 파악에 매우 유용하다[10].

본 연구에서 제시한 빅데이터 분석하기 위해

전국지역의 언론사에서 제공하는 키워드를 수집하여 뉴스 검색 분류하고 정형화하여 빅데이터를 파이선 빅데이터 클라우드 워드로 시각화 분석을 수행하였다. 분석에서 사용되었던 ESG와 창업의 키워드를 제외하고 결과물을 도출하였다.

그림 9의 워드별 카운트 분석은 파이선 프로그램을 사용하여, 입력된 키워드인 ESG와 창업을 제외하고 키워드별 노출 카운트를 시각화하였다.

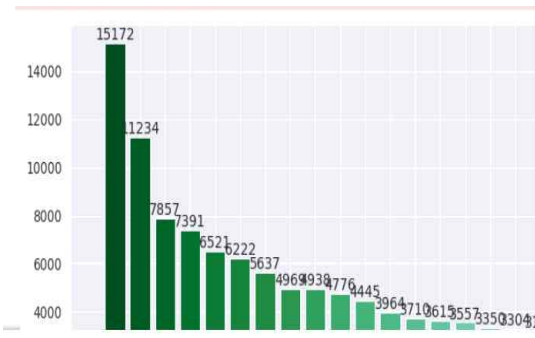


그림 9. 워드별 카운트 분석  
Fig. 9. Word type count analysis

그림 10의 빅데이터 클라우드 워드의 ESG 이미지 키워드 분석결과 창업에 긍정적 키워드는 기업, 지원, 사업, 사회, 투자, 경영 등의 글자가 크게 나타나 긍정적 이미지로 확인되었다. 반면 협약, 소비자, 인프라, 규제, 공유 등은 작은 글씨로 확인되어 긍정적 이미지에 영향이 큰 글씨에 비해 미비했다.

빅데이터 워드 클라우드 분석결과 가장 많은 15,172번의 언론 기사에 노출된 단어는 기업으로서 일반적으로 수익을 중심으로 기업에서의 지속해서 판매되는 Item 즉 캐시카우를 확보 활성화해야 하고, 두 번째로 많이 검색된 11,234번의 지원으로서 금융권 비금융권의 현금흐름을 안정적으로 검토하고, 정부에서 지원하는 과제를 수주하여 금융지원 및 컨설팅 지원프로젝트를 진행하고 특허 및 디자인등록 등의 기술성과와 더불어

고객에게 최선의 서비스를 제공하여 충성도 높은 고객을 확보해야 한다고 판단된다. 세 번째로 많이 검색된 7,857번의 사업으로서 과거의 시장과 현재의 시장을 분석하고 미래시장을 예측하여 사업의 진입 시기를 결정하고 성장시키는데 중점적인 역할이 필요하다. 이외에도 7,391번의 검색된 사회, 6,521번의 검색된 투자 등도 관심이 필요하다고 판단된다.



그림 10. ESG 이미지 키워드 분석결과  
Fig. 10. ESG image keyword analysis results

## 5. 결론

본 연구를 기반으로 전국에서 창업을 준비하는 MZ세대는 기업, 지원, 사업 및 사회, 투자, 경영 등의 순으로 확인되어 창업 준비에 요구되는 핵심 키워드를 신중히 고려하는 방안을 제안하였다.

인과관계에 중점을 두는 전통적 분석방법과는 달리 상관관계의 개념을 가진 빅데이터 분석은 통계적으로 유의미한 관계 규명은 가능하나 이러한 결과의 원인적 배경과 통찰을 제공하는 데는 한계가 있으며, 전통적인 연구방법과 병행하지

않을 경우, 충분한 현상 이해와 이를 통한 기회 발굴에는 부족할 수 있다.

연구의 한계점으로 본 연구에서 사용된 데이터들은 전국을 대상으로 데이터를 분석하였으나 향후 연구 방향은 전국적으로의 더욱 다양한 빅데이터를 확보 및 정형화하여 창업의 양적 질적인 발전에 대한 논의가 필요하다고 생각하고, 주요 키워드를 변수 지정하여 조절 및 매개효과의 상관관계, 요인 분석, 신뢰도 분석, 회귀 분석 등으로 ESG 창업의 긍정적인 영향에 대한 증명이 필요할 것이라고 본다.

This paper was supported by Education and Research promotion program of KOREATECH in 2022

## 참 고 문 헌

- [1] Curran, C.-S., & Leker, J. (2011). Patent indicators for monitoring convergence examples from NFF and ICT. *Technological forecasting and social change*, 78(2), 256 - 273  
<https://doi.org/10.1016/j.techfore.2010.06.021>
- [2] S. H. Goo. (2022). A study on the Influence of CEO Image on Brand Value -Focused on the Legal Process of Samsung CEO and Change of Corporate Value-Volume. *The Journal of the Korea Contents Association*, 22(2), 237-254  
<https://doi.org/10.5392/JKCA.2022.22.02.237>
- [3] C. I. Eun. (2016). A Study on the Relations between Social Responsibility Corporate Image and Consumer Attitude of Design Social Enterprises Focusing on the Mediating Effect of Corporate Image. *Journal of the Korean Society of Design Culture*, 22(2), 399-414.  
<https://doi.org/10.18208/ksdc.2016.22.2.399>
- [4] S. H. Kim. (2021). Corporate Social Responsibility and Liquidity. *Journal of the Korean Data Analysis Society*, 23(2), 673-685.  
<https://doi.org/10.37727/jkdas.2021.23.2.673>
- [5] J. H. Joo. (2019). The Effect of Corporate Social Responsibility on Organizational Trust, Job Satisfaction and Turnover Intention. *Korean journal of construction engineering and management* 20(5), 83-94.  
<https://doi.org/10.6106/KJCEM.2019.20.5.083>
- [6] H. J. Lee. (2021). Data-Driven Design Methodology based on Data Science Paradigm Focused on Design Research Case Study of Fine Dust Information App Service. *The Journal of the Korea Contents Association*, 21(10), 103-114.  
<https://doi.org/10.5392/JKCA.2021.21.10.103>
- [7] S. H. Park. (2018). Design Thinking Methodology for Social Innovation using Big Data and Qualitative Research. *Asia-Pacific journal of business and venturing*, 13(4), 169-181.  
<https://doi.org/10.16972/apjbve.13.4.201808.169>
- [8] S. M. Jung. (2012). A Study on the Classification system and fundamental notions of the Peripheral Devices and Accessories in Product Design for Smartphone. *Journal of Korea design knowledge*, 24(1), pp. 20-29  
<https://doi.org/10.17246/jkdk.2012.24.003>
- [9] Adam Mechtley. Ryan Trowbridge. (2012). *Maya Python for games and film a complete reference for the Maya Python and the Maya Python API*. Morgan Kaufmann. Amsterdam Boston, 24(1), 20-88. <https://doi.org/10.9708/jksci.2015.20.4.049>
- [10] Brooks, S. (2015). Does personal social media usage affect efficiency and well-being?. *Computers in Human Behavior*, 46, 26-37.  
<https://doi.org/10.1016/j.chb.2014.12.053>

저 자 소 개



김대유(Dae-Yu Kim)

2012.2 한국산업기술 대학교 IT융합 졸업  
2015.2 서강대학교 기술경영 석사  
2022.8 한국기술교육대학교 산업경영학과  
기술경영전공 박사수료  
현재 : 남서울대학교 빅데이터산업보안학과  
조교수  
<주관심분야> 빅데이터, 기술경영, 통계



배장원(Jang Won Bae)

2007.9 고려대학교 전기전자전파공학 졸업  
2009.9 한국과학기술원 전자공학과 석사  
2015.2 한국과학기술원 시스템공학과 박사  
현재 : 한국기술교육대학교 산업경영학과  
조교수  
<주관심분야> 모델링, 시뮬레이션, 빅데이  
터, 인공지능