

연관규칙 분석을 통한 ESG 우려사안 키워드 도출에 관한 연구

안태욱* · 이희승** · 이준서***

〈목 차〉	
I. 서론	3.4 연관규칙 분석
II. 이론적 배경	IV. 분석결과
2.1 ESG 우려사안	4.1 ESG 우려사안 관련 키워드
2.2 정보제공자로서의 미디어 역할	4.2 ESG 우려사안 키워드 분류
2.3 ESG 우려사안 데이터	4.3 키워드 신뢰도 분석
2.4 키워드 도출 관련 선행연구	4.4 키워드 네트워크 분석
III. 연구방법	V. 결론 및 시사점
3.1 자료 및 연구분석 절차	참고문헌
3.2 ESG 우려사안 사건기사 도출	<Abstract>
3.3 키워드 추출 및 정제	

I. 서론

ESG(Environmental, Social, Governance)에 대한 관심이 증대되고 있다. 환경(E), 사회책임(S), 지배구조(G)로 대변되는 ESG는 기업의 지속가능성에 대한 논의가 활발히 전개되면서 그 중요성이 강조되고 있기 때문이다. 즉 기업의 성과를 재무적 요소가 아닌 비재무적 요소로 판단하는 기준으로 ESG가 널리 활용되고 있는

것이다. 이에 따라 지난해부터 대외적으로 ESG 경영을 선포하는 기업들이 증가하고 있으며 연기금 등 기관투자자나 개인투자자들도 ESG 활동을 투자지표 중 하나로 활용하고 있다. 해외에서도 ESG에 대한 투자가 급증, 지속가능투자(sustainable investment) 잔액이 지난해 말 기준 40.5조달러로 4년 만에 두 배 증가했으며 현금유입액은 최근 3년 동안 매년 두 배씩 증가하고 있다¹⁾.

* 딥서치, tahn@deepsearch.com(주저자)

** 한국과학기술원 경영대학, hslee20@kaist.ac.kr

*** 동국대학교 경영학과, juyi@dongguk.edu(교신저자)

1) Global Sustainable Investment Alliance(GSIA), Morningstar

기업 가치를 증대시키고 지속가능성을 확보하기 위해 ESG활동은 이제 필수적인 요소로 자리 잡았다. 하지만 기업활동을 전개하다 보면 환경오염이나 산업재해, 분식회계, 배임, 횡령, 불공정거래 등 반 ESG, 즉 예상하지 못하는 우려사안(controversy)이 발생할 수 있다. 대한항공의 회항사건이나 남양유업의 대리점 갑질 및 최대주주 친인척 일탈 행위들이 대표적인 ESG 우려사안이라고 할 수 있다.

투자자들은 기본적으로 기업이 잘한 행동보다 잘못된 행동에 대해 더 민감하게 반응하기 때문에 ESG보다 ESG 우려사안이 기업가치에 미치는 영향이 더 클 수 있다. Kothari et al. (2009) 등 많은 행태재무학 연구들은 부정적 사건이 긍정적 사건에 비해 투자자 반응을 더 빠르고 지대하게 유발함을 검증한 바 있다. 또한 Krüger(2015)는 기업의 사회적 책임(CSR: Corporate Social Responsibility)의 경우 부정적인 CSR 뉴스는 주가를 강하게 하락시키는 반면 긍정적인 CSR 뉴스로 인한 주가의 상승폭은 제한적이라고 밝혔다.

그렇다면 투자자들은 어떤 정보를 ESG 우려사안으로 인식할까? ESG는 비교적 최근에 선보인 개념으로 이 보다 앞서 시장에 도입된 유사개념으로 파악이 가능할 것이다. ESG의 유사개념으로는 CSR, 지속가능성(Sustainability), 공유가치창출(CSV: Creating Social Value) 등이 존재한다. 이에 따라 투자자들은 기업들이 반 CSR, 반 지속가능성, 반 공유가치창출 관련 활동을 한 경우 ESG 우려사안으로 파악할 수 있을 것이다. 이 중 가장 대표적인 것은 반 CSR을 일컫는 기업의 사회적 무책임(CSiR: Corporate Social Irresponsibility)이다. 하지만 CSiR은 주

로 사회나 윤리 등과 관련된 기업들의 일탈행위를 의미하는 것으로 환경 및 지배구조까지 포함하는 ESG 우려사안과는 다소 상이할 수 있다. 즉 CSiR 등 유사 개념들은 기업과 이해관계자의 사회적 호혜성에 방점을 두는 반면, ESG 우려사안은 비재무적 리스크 측면 혹은 이와 연계된 투자자의사결정 및 장기적인 재무적 가치에 영향을 미칠 수 있는 요인에 초점을 맞추고 있다.

한편 ESG 우려사안에 대해서는 투자자의 인지여부가 중요하다. 기업들이 반 ESG 활동을 하더라도 이를 인지하지 못하면 투자자의사결정에 반영할 수 없기 때문이다. 이 같은 관점에서 미디어의 역할이 지대하다고 할 수 있다. 투자자 입장에서 ESG 우려사안에 대한 정보는 SNS나 풍문 등 다양한 채널을 통해 습득 가능하겠지만 정보의 신뢰성 측면에서 미디어에 대한 의존도가 높을 수밖에 없을 것이다. 더구나 부정적 활동의 경우 그 인지효과는 긍정적 활동보다 훨씬 큰 것으로 나타나고 있다. 따라서 기업의 부정적 활동에 대한 투자자의 인지여부에 미디어의 정보 전달 역할이 중대하다고 할 수 있다.

이에 따라 본 연구에서는 ESG 우려사안에 대한 키워드를 도출해 보고자 한다. 즉, ESG 등급을 부여받은 294개 상장기업을 대상으로 2013년부터 2018년까지 130개 미디어에 노출된 ESG 우려사안과 관련된 기사에 대해 머신러닝의 한 방법론인 연관규칙 분석(association rule mining)을 통해 ESG 우려사안 키워드를 도출한다. ESG 우려사안 키워드를 환경, 사회, 지배구조로 구분하여 미디어에서는 관심을 갖고 투자자가 민감하게 반응하는 키워드를 알아본다. 또한 키워드 간의 신뢰도 분석을 통해 키워드간 인과관계를 모색하며, 네트워크 분석을 통해 키

워드간 중요성, 연관성, 근접성 등을 파악한다.

본 논문은 ESG 관련 연구에 대한 기초 데이터 제공의 역할을 수행할 수 있을 것으로 기대한다. ESG 성과나 등급에 대해서는 한국기업지배구조원(KCGS) 등 일부 기관이 자료를 제공하고 있는 반면 ESG 우려사안은 그 중요성이 증대되고 있음에도 불구하고 자료 제공 기관이 존재하지 않는다. 해외의 경우 ESG 등급뿐 아니라 ESG 우려사안에 대한 DB를 제공하고 있는 점을 감안하면 본 연구의 공헌도는 적지 않을 것으로 생각한다. 또한 ESG 우려사안의 키워드 도출을 통해 투자자들이 민감하게 반응하는 기업들의 반 ESG 활동을 정의내릴 수 있을 것으로 기대한다. 기업들은 본 연구결과를 토대로 ESG 활동 시 각별히 주의해야 할 부문을 모색할 수 있을 것이다. 무엇보다도 본 연구는 연관규칙 분석을 통한 동시출현 키워드군 구축을 실시, 투자심리를 머신러닝을 기반으로 한 빅데이터 분석을 활용했다는 점에서 상당한 의의가 있다고 할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 이론적 배경을 기술하고 3장에서는 표본선정, 연관규칙 분석 등 방법론에 대해 설명한다. 4장에서는 분석결과를 기술하고 5장에서는 결론 및 시사점을 설명한다.

II. 이론적 배경

2.1 ESG 우려사안

ESG 우려사안(controversial issues)이란 기업의 환경, 사회, 지배구조와 관련하여 의심스러운 사회적 행동이나 소비자에게 피해를 끼치는 제품 스캔들로 인해 투자자가 기업에 대해 부정적 인식을 갖게 하는 사건을 의미한다²⁾. ESG 우려사안은 ESG 염려(ESG concern)라는 용어로 사용되기도 하며 술, 담배, 도박, 총기 등 사회에 해가 되는 산업 자체(controversy industry)를 지칭하기도 한다.

ESG 우려사안은 비교적 최근에 개념화되었다. ESG 우려사안 보다는 CSiR이라는 용어가 더 대중화되어있다. 왜냐하면 ESG가 CSR의 후속개념으로 볼 수 있기 때문이다. CSR은 시장자유화와 글로벌화로 많은 수익을 창출한 기업들이 윤리 중심의 경영으로 사회공헌에 앞장서야 한다는 취지에서 1980년대 후반부터 투자주요지표로 활용되었다. 즉, CSR은 사회와 윤리 중심의 투자원칙 프레임워크로서 환경은 사회책임 영역의 하나로 인식되어왔다. CSR에 관한 국제표준인 ISO 26000의 7대 핵심 주제 중 5개가 사회책임 관련이며 환경과 지배구조는 각각 1개에 불과하다³⁾. 하지만 최근 지구온난화에 따른 급속한 기후변화로 환경(E) 이슈가 제기되고, 소득 및 사회적 불평등이 확대되면서 기업의 역할수행에 지배구조(G)의 중요성이 인식되면서 ESG 개념이 확립되었다. ESG 개념이 공식화된 것은 2000년대 중반이다. 코피 아난(Kofi Annan) 전 UN 사무총장이 2004년 세계 각국의 금융회사에 지속 가능한 투자를 위한 가이드라인(UN 책임투자원칙: UN

2) Cai et al.(2012), Klein and Dawar(2004)

3) 인권, 노동관행, 공정거래, 소비자이슈, 지역사회 참여 및 발전, 기업지배구조, 환경
<https://iso26000.info/iso26000>

PRI)에 동참해줄 것을 요청하며 ESG 용어를 처음 사용한 것으로 알려져 있다.

이에 따라 ESG의 반대개념인 ESG 우려사안에 대한 선행연구는 CSR의 반대개념인 CSiR에 대한 연구에 비해 지극히 제한적이다. 국내 선행연구는 존재하지 않고 해외에서만 일부 연구가 진행되었다. 기본적으로 이해관계자이론(stakeholder theory)과 정당성이론(legitimacy theory)에 근거해 Orlitzky(2013)와 Adams(2002)는 ESG 우려사안이 발생하는 경우 기업가치는 하락한다고 주장한다. 이해관계자이론이란 이해관계자가 사회적 통제의 대리인 역할을 수행함에 따라 기업의 ESG 활동이 도덕적 자원을 구축할 수 있다는 주장이며, 정당성이론이란 기업의 행동이 규범, 가치, 신념 및 정의측면에서 바람직하고 정당해야 한다는 주장이다. 반면 Aouadi and Marsat(2018)은 전 세계 58개국 기업을 대상으로 ESG 우려사안이 기업가치에 미치는 영향을 분석, ESG 우려사안이 오히려 기업가치를 증대시킨다는 연구결과를 발표했다. 다만 CSR 활동과 결합하는 경우 ESG 우려사안과 기업가치는 무관하다고 주장했다. 이는 CSR과 CSiR이 반드시 상반되는 개념이 아니므로 CSiR이 기업가치에 미치는 영향이 다양할 수 있다는 Arora and Dharwadkar(2011)을 지지한다고 할 수 있다.

2.2 정보제공자로서의 미디어 역할

ESG 우려사안에 대해서는 특히 미디어의 관심이 증대한 역할을 하는 것으로 조사된다. ESG 우려 관련기사는 기업의 미래 전망을 불투명하게 할 뿐 아니라 기업평판에 대한 리스

크로 발전되어 결국 기업가치에 영향을 줄 것이기 때문이다. 즉 투자자 입장에서 부정적 정보를 취득할 수 있는 주요 통로가 미디어인 것이다.

기업가치 변화에 대한 미디어의 역할과 관련된 논문은 다수 존재한다. Freeman and Cavusgil(1984)과 Siegel and Vitaliano(2007)는 기업가치에 영향을 미칠 만한 특정 사건이나 이슈와 관련해 미디어의 관심은 투자자나 소비자에게 정보의 비대칭을 줄여줄 뿐만 아니라 대중의 여론에 영향을 미친다고 주장한다. Du et al.(2010)도 기업의 행태에 대한 심층취재는 관련자가 기업의 활동에 대한 이성적인 기대나 판단을 형성하는데 도움을 준다고 지적한다.

또한 부정적 기사와 관련하여 Pollock & Rindova(2003)는 부정적 기사는 기업의 적법성과 명성에 손상을 주면서 기업의 미래 성과에 부정적인 영향을 미친다고 설명한다. Gurun and Butler(2012), Tetlock(2007), Tetlock et al.(2008)은 실증분석을 통해 부정적인 기사가 기업의 가치를 낮추는 결과를 보였다. 이들은 소비자가 부정적인 기사에 언급된 기업이 생산한 제품의 소비를 피하게 되며 투자자는 손상된 명성으로 인해 증대된 리스크에 대해 보상받기를 위해 더 높은 수익률을 요구하게 된다고 설명한다.

특히 ESG 우려사안에 대한 미디어의 관심은 이해관계자의 회의적 시각을 증대시키고 기업의 위선행위에 대한 인지를 촉발하여 기업에 대한 신뢰도를 추락시킨다는 의견이 제기된다. Maignan and Ferrell(2004)은 ESG 우려사안에 대한 미디어의 부정적 보도와 ESG 이슈에 대

한 이해관계자의 민감도가 증가함에 따라 의심적인 행동을 하는 기업은 정체성과 평판 변화에 대해 심각히 고민해야 한다고 주장한다. 더구나 Krüger(2015)는 부정적 ESG 뉴스에 대해 투자자가 부정적으로 반응한다는 실증분석 결과를 발표했고, Chollet and Sandwidi(2018)은 종업원과 환경, 지역사회 관련된 이슈일수록 이 같은 관계가 강하게 나타남을 보였다. 이에 따라 Jiao(2010)은 이해관계자가 ESG 정보와 기업가치를 연결하는 주요 통로가 된다고 주장한다.

2.3 ESG 우려사안 데이터

앞 절에서 ESG 우려사안이 기업 가치에 중

대한 영향을 미칠 수 있다는 사실을 보였다. 그렇다면 ESG 우려사안이 구체적으로 어떤 것일까? 어떤 사건이 ESG 우려사안에 해당되고 기업별로 어느 정도의 ESG 우려사안이 발생하며 투자자들은 이를 어떻게 해석해야 할까?

ESG 우려사안의 보다 본질적인 개념인 기업의 ESG 성과 정보는 쉽게 접근이 가능하다. 한국기업지배구조원(KCGS) 등 일부 기관에서는 기업들의 ESG 성과를 평가하여 매년 공시하고 있으며 ESG의 중요성의 증대됨에 따라 정보의 구체성이나 신뢰성이 강화되고 있다⁴⁾. 한국기업지배구조원은 2003년부터 기업의 지배구조의 평가를 실시했으며, 2011년부터 사회책임과 환경경영을 추가하여 ESG평가를 통해 국내 상

<표 1> 한국기업지배구조원 환경경영, 사회책임경영, 기업지배구조 모범규준

평가항목	대분류	소분류
환경	환경경영계획	최고경영자의 실천의지, 환경경영 전략과 방침, 환경목표 및 계획수립, 환경친화적 조직문화, 환경친화적 조직체계
	환경경영실행	환경친화적 생산, 기후변화 대응, 환경친화적 공급망 관리, 환경위험관리
	환경성과관리 및 보고	환경성과 관리, 환경회계, 환경감사, 환경정보 보고
	이해관계자 대응	이해관계자 참여활동, 환경보호 활동
사회	근로자	고용 및 근로조건, 노사관계, 직장 내 보건 및 안전, 인력 개발 및 지원, 직장 내 기본권
	협력사 및 경쟁사	공정거래, 부패방지, 사회적 책임 촉진
	소비자	소비자와 공정한 거래, 소비자 안전 및 보건, 소비자 개인 정보 보호, 소비자 및 소통
	지역사회	지역사회 참여 및 공헌, 지역경제발전, 지역사회와의 소통
지배구조	주주	주주의 권리, 주주의 공평한 대우, 주주의 책임
	이사회	이사회 기능, 이사회 구성 및 이사선임, 사외이사, 이사회 운영, 이사회내 위원회, 이사의 의무, 이사의 책임, 평가 및 보상
	감사기구	내부감사기구, 외부감사인
	이해관계자	이해관계자의 권리 보호, 이해관계자의 경영감시 참여
	시장에 의한 경영감시	공시, 기업 경영권 시장, 기관 투자자

4) 기업들의 ESG 등급은 한국기업지배구조원 이외에 서스틴베스트와 에코파트너스 등에서 제공하고 있으나 한국기업지배구조원의 ESG 등급을 가장 많이 활용한다.

장사의 지속가능경영 수준을 평가하고 있다. 한국기업지배구조원은 코스피와 코스닥 상장기업을 대상으로 환경, 사회, 지배구조, 통합 등 4개 분야에 대해 총 7개의 등급(S, A+, A, B+, B, C, D)으로 구분된 ESG등급을 발표한다⁵⁾. 평가대상부분은 <표 1>에서와 같이 환경경영 부문(E)은 환경경영, 환경성과, 이해관계자 대응, 사회책임부문(S)은 근로자, 협력사 및 경쟁사, 소비자, 지역사회에 대한 활동, 지배구조부문(G)은 주주권리보호, 이사회, 감사기구, 공시 등이다.

하지만 ESG 우려사안에 대해서는 활용할 만한 DB가 존재하지 않는다. 해외의 경우에는 ESG 평가기관이 우려사안에 대한 평가도 실시하여 우려사안 점수(controversy score)를 별도로 공시한다. ESG 대표 평가기관인 MSCI에서는 우려사안을 매우심각(very severe), 심각(severe), 보통(moderate), 경미(minor) 등으로 구분하여 평가하고 있으며 또 다른 ESG평가기관인 Refinitiv와 Morningstar에서는 우려사안 이슈 발생 시 기존 ESG 점수를 할인하는 방법을 적용하고 있다. 예컨대 Refinitiv에서는 1년 동안 발생하는 ESG관련 부정적 사건을 23개 토픽으로 분류하여 계산한 값을 제공하고 있으며 Aouadi and Marsat(2018)와 Li et al.(2019)은 이 자료를 활용하여 연구를 진행했다⁶⁾.

2.4 키워드 도출 관련 선행연구

키워드는 텍스트마이닝, 정보검색, 자연어처리 분야에서 중요한 역할을 담당한다. 특히 문서 양이 기하급수적으로 증가하는 시대에 문서에서 자동적으로 키워드를 도출하는 텍스트마이닝 분야에 있어서는 키워드가 주요 속성을 추출한다는 측면에서 반드시 필요한 과정이 되었다.

키워드를 도출하는 방법은 TF-IDF, 토픽모델링, 연관규칙 분석 등 다양하다. TF-IDF(Term Frequency-Inverse Document Frequency)는 문서 안에 존재하는 단어의 중요도를 평가하기 위해 주로 활용된다. TF-IDF모형은 벡터공간모형을 사용하여 정보검색 목적으로 문서를 표현하는 원리이다. TF는 한 문서 내에서 특정 단어가 출현하는 빈도수를 나타내며 주어진 단어가 문서 내에서 자주 출현 할수록 상대적으로 더 중요하다는 가정을 내포한다. IDF는 문서집합에 포함되어 있는 문서 수를 특정 단어가 출현한 문서의 수로 나눈 값으로 상대적으로 많은 문서에 출현한 단어의 IDF는 작은 반면 한 문서에 편중하여 출현한 단어의 IDF는 크다. 즉, IDF가 작은 단어는 보편적인 가능성이 크며, 큰 단어는 문서내에서 주요한 의미를 가진 단어일 가능성이 높다. 그러므로 문서내에서 개

- 5) 다만 A+와 A 등급비율이 지극히 낮고 기업지배구조(G) 등급을 제외하고는 B이하 등급을 통합하여 공시, 정보의 활용도 측면에서 미흡하다는 지적이다.
- 6) Refinitiv는 ESG등급과 ESG우려사안에 대한 평가데이터를 제공하는 대표적인 기관이다. Refinitiv는 2009년 ESG정보분석업체인 ASSET4를 인수하여 ASSET4 ESG Database를 Refinitiv 서비스에 포함하여 제공하고 있다. 10개의 카테고리, 450개의 기업 데이터 포인트를 기반으로 ESG점수를 산출하는데 현재 전세계 10,000개 이상의 기업에 대해 ESG점수를 제공하고 있다. 특히 ESG 우려사안에 대해서는 환경 1개, 사회 18개(지역사회, 인권, 제품책임, 근로자 등 4개 중분류), 지배구조 4개(주주와 경영자 등 2개 중분류) 등 23개 항목에 대해 최근 발생한 이슈에 대한 가중치를 적용하는 방식으로 점수를 산출한다.

별 단어의 빈도 뿐 아니라 상대적 중요도를 나타낼 때 이를 활용한다. 그러나 TD-IDF를 기반으로 한 연구들이 동시 출현하는 키워드의 연관성을 적절히 파악하지 못하는 한계점을 드러냄에 따라 Matsuo and Ishizuka (2004)는 한 문장에서 동시에 출현한 횟수를 연구하거나 Mihalcea and Tarau(2004)와 Brin and Page (1998)는 출현한 키워드로 구성된 그래프를 이용한 논문을 발표한 바 있다.

토픽 모델링은 구조화되지 않은 방대한 문헌 집단에서 추상적인 주제를 발견하기 위한 통계적 모델 중 하나로, 유사한 의미를 가진 단어들을 클러스터링하는 방식으로 주제를 추론하는 모델이다 (Steyvers and Griffiths, 2007; Blei 2012; 최병설, 김남규, 2019). 하지만 토픽 모델링은 특정 분야에 대한 동향 분석에 주로 사용되고, 키워드간 선후행관계의 연관성을 파악할 수 없으며, 키워드 분석 이전에 토픽의 개수를 임의로 설정하는 탐색적인 과정을 거쳐야 한다는 한계가 존재한다.

데이터간 유의미한 패턴을 분석하는 방법으로 유사도 분석이 활용되기도 한다. 유사도 측정은 동시출현 건수를 총 출현 빈도수로 나누는 보정작업을 통해 개체간의 연관관계를 측정하는 것으로 단순 출현 회수와 동시출현 건수를 개별적으로 활용할 때 나타나는 단점을 극복할 수 있다. 하지만 유사도 측정지수는 두 단어 간의 방향성을 나타내지 못하는 한계점이 존재한다. 즉, 두 단어가 동시 출현했을 때 빈도수에 따른 연관관계만 파악할 수 있을 뿐 한 단어가 출현했을 때 다른 단어도 같이 출현할 확률을 계산하는 것은 불가능하다.

반면 연관규칙 분석은 지지도, 신뢰도, 향상

도를 기반으로 키워드간 동시 발생여부 및 선후관계를 밝혀주어 단순 키워드 검색으로 인한 중복문제나 키워드 선정의 주관성을 최대한 배제할 수 있는 방법이다. Agrawal et al.(1993)에 따르면 연관규칙 분석은 어떤 두 키워드가 동시에 발생하는가를 알려주는 규칙들을 생성하는 머신러닝 알고리즘이라고 할 수 있다. ‘장바구니 분석(market basket analysis)’이라고도 명명되는 연관규칙 분석은 ‘특정 A제품을 구매할 시 B제품 역시 구매할 가능성이 높다’는 해석으로 자주 설명된다. 마케팅에서 매장 진열, 묶음상품 개발, 교차 판매등의 판매전략 수립에 활용된다. 예를 들어 ‘A→B’로 표현되는 의미는 ‘상품A를 구매한 사람들 중 많은 사람들이 상품B를 구매한다’고 말할 수 있다. 이때 연관규칙을 설명하는 3가지 지표인 지지도, 신뢰도, 향상도를 사용한다. 지지도(support)는 전체 거래의 상품A와 상품B가 동시에 거래된 비율을 의미한다. 지지도를 통하여 해당 규칙의 유용성을 확인한다. 신뢰도(confidence)는 상품A를 구매한 거래 중에서 상품A와 상품B가 동시에 나타난 비율을 의미하는데 조건부확률 형태로 연관규칙의 강도를 확인한다. 마지막으로 향상도(lift)는 상품A와 상품B가 동시에 거래된 비율에 상품A와 상품B가 각각 거래된 비율로 나눈 값이다. 이는 신뢰도를 상품B가 나타난 확률로 나눈 값과 동일하다. 상품B를 임의로 구매하는 경우와 상품A와 관계를 고려하여 구매하는 경우의 비율을 의미하는데, 이를 통해 향상도가 연관규칙이 우연히 발생한 것인지 실제 연관성이 있는 것인지 알려준다. 즉 두 상품간의 향상도가 1보다 크다면 두 상품은 서로 양(+)의 상관관계가 있는 것을 의미한다.

Agrawal et al.(1993)이 최초로 연관규칙 분석을 소개 한 후 Agrawal and Srikant(1994), Tan et al.(2005) 등 다양한 후속연구가 진행되고 있다. 국내 문헌의 경우 강경수와 박세민(2019)이 연관규칙 분석을 이용하여 KCI등제지에 게재된 경영학 논문을 분석하여 ‘기업의 사회적 책임’, ‘기업가정신’, ‘고객 만족’, ‘직무 만족’, ‘신뢰’와 같은 키워드를 도출하였다. 안민욱 등(2019)은 향상도 기반 연관규칙 네트워크를 통하여 농식품 판매전략을 제시하였다. 이밖에 황현석과 유기동(2013)은 인터넷 뉴스 사이트 방문자들의 웹로그를 활용하여 방문자가 읽은 온라인뉴스간의 연관규칙을 분석했다. 전익진과 이학연(2016)은 연관규칙 기반 동시출현단어 분석을 활용하여 기술경영 분야 연구주제 키워드 네트워크를 구축하였으며, 이태원 등(2020)은 키워드 네트워크 분석을 이용하여 국가직무능력표준 기반의 빅데이터 분석 직무역량과 국내 지방대학의 MIS 교과과정에서의 차이를 확인하였다.

Ⅲ. 연구방법

3.1 자료 및 연구분석 절차

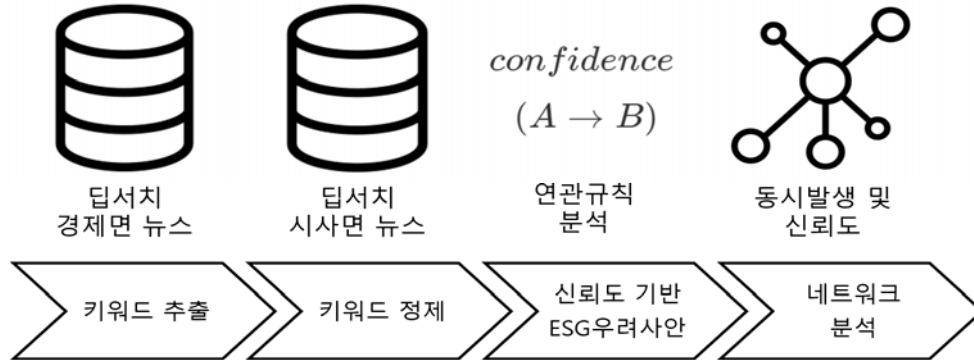
ESG 우려사안 기초 자료는 딥서치(DeepSearch)의 데이터베이스를 이용한다. 딥서치는 국내 130개 언론사에서 제공하는 신문 기사 뿐만 아니라 기업공시, 증권사리포트, IR

리포트에 이르기까지 다양한 텍스트기반 데이터를 분석할 수 있는 빅데이터 플랫폼을 제공한다. 본 연구에서는 2013년부터 2018년까지 한국기업지배구조원에서 ESG등급을 부여받은 294개 기업의 신문기사 23,837건을 대상으로 한다. 해외 선행연구에서도 ESG 관련 키워드 분석이 제한적으로 존재하였으나 환경이나 지배구조와 같은 특정 이슈 또는 배임, 횡령, 담합과 같은 특정사건에 대한 분석이었으며 특히 지나치게 적은 표본을 활용해 연구결과에 대한 신뢰성을 담보하기 어려웠다. Christensen and Murphy(2004)는 기업의 조세회피 사건에 대해 분석했으며, Demacarty(2009)는 사기사건, 가격담합, 입찰담합, 뇌물사건 등에 대한 제한된 분석을 수행했다. Capelle-Blancard and Monjon(2012)은 사회적 책임투자(SRI)에 관한 유사키워드 4개만을 이용하였고, Daugaard(2020)은 ESG투자에 대한 키워드를 조합하여 활용한 바 있다.

연구는 <그림 1>과 같은 과정을 거쳐 진행된다. 먼저 ESG 등급을 보유하고 있는 294개 기업의 신문기사 23,837건에 대해 각 기사별 다양한 키워드를 살펴본다. 기사별 키워드 중 ESG 우려사안과 관련된 잠재적 키워드를 포함하고 있는 기사를 대상으로 1,2차 태깅을 실시하여 ESG 우려사안을 포함하고 있는 신문 기사를 추출한다. 추출된 신문기사의 키워드를 대상으로 경제면과 사회면에서 ‘논란’이라는 키워드와 동시에 발생하는 경우 이를 삭제하는 정제과정을 거친다. 이후 키워드의 동시출현 여부를 분

7) 딥서치는 AI기반의 비즈니스 데이터 분석 플랫폼이다. 딥서치는 상장기업과 비상장기업(외감법 대상, 비대상 포함) 150만개 기업의 재무정보 및 이들의 기업분석 및 관련뉴스를 제공하고 있다.

www.deepsearch.com



<그림 1> 연구과정 절차도

석하는 연관규칙 분석을 통해 키워드의 결합과정을 거쳐 최종 키워드를 도출한다.

3.2 ESG 우려사안 사건기사 도출

ESG등급을 보유하고 있는 기업의 기사별 키워드를 대상으로 잠재적 ESG 우려사안 키워드를 활용하여 1차 태깅을 실시했다. 환경과 관련된 키워드인 ‘연비조작’, ‘폐수방류’, ‘환경오염’을 포함하고 있는 기사는 환경(E)으로, ‘골목상권위협’, ‘담합’, ‘불공정거래’, ‘불매운동’

등은 사회(S)로, ‘배임’, ‘주가조작’, ‘차명계좌’, ‘탈세’, ‘횡령’ 등 기업의 오너와 경영진이 저지르는 일탈행위는 지배구조(G)로 1차 분류했다. 분류가 확실하지 않은 ‘갑질’, ‘비리’, ‘밀어내기’, ‘불공정거래’, ‘논란’ 등 공통적인 키워드는 미분류(Unclassified:U)로 표기했다. 미분류 기사는 연구자가 직접 기사의 전문을 읽고 판단하는 2차 태깅을 통해 환경, 사회, 지배구조로 분류하였다. <표 2>는 1차 및 2차 태깅 후 완성한 키워드 분류표이다.

<표 2> 1차 및 2차 태깅 후 분류표

키워드	1차태깅	2차태깅	키워드	1차태깅	2차태깅
갑질	U	G	블랙기업	U	G
경영비리	G	G	비리	U	G
계약사불법지원	G	G	비자금	G	G
고액채납	G	G	살인기업	G	G
고용세습	G	G	성과급잔치	G	G
골목상권위협	S	S	소비자우롱	G	G
군대식조직	G	G	숨방망이처벌	U	G
기습인상	U	S	연비조작	E	E
내부자거래	G	G	이물질	U	E

논란	U	S	일감몰아주기	G	G
뇌물	G	G	조세회피	G	G
담합	S	S	주가지작	G	G
독과점	G	G	집단손배소	U	G
리베이트	G	G	차명계좌	G	G
미공개정보이용	G	G	차명주주	G	G
밀어내기	U	G	채용비리	G	G
배임	G	G	총수구속	G	G
부당해고	G	G	탈루	G	G
분식회계	G	G	탈세	G	G
불공정거래	S	S	폐수방류	E	E
불매운동	S	S	품질조작	U	G
불법고용	S	S	하도급문제	S	S
불법상속	G	G	환경오염	E	E
불법증여	G	G	횡령	G	G
불법과건	S	S			

U: 미분류, E: 환경, S: 사회, G: 지배구조

<그림 2> 최종 태깅 후 ESG 우려사안 사건기사 리스트

date	category	section	publisher	tag	keyword	title	content_url	ind_l	industry_name
2013-09-16	news	economy	미디어오늘	S	불법과건	노동부, 삼성AS기사에 '논란 있으나 불법과건 아냐'	Go to Website	100	분류 제외, 기타
2013-09-16	news	economy	파이낸셜뉴스	G	주가지작	셀트리온 주가지작 '진실 계명'	Go to Website	C21	의료용 물질 및 의약품 제조업
2013-10-17	news	economy	세계일보	G	비리	속속 드러나는 비리 정황에도 동양 방정관리 '대주주 입김' 유지	Go to Website	C23	비금속 광물제품 제조업
2013-10-17	news	economy	이데일리	S	담합	스마트물어 튀길래, KT, 검찰고발에 공정위 담합과장급	Go to Website	100	분류 제외, 기타
2013-10-17	news	economy	디지털타임스	S	논란	카드사 '체크카드' 포함된 일괄 평가 현황성 논란..기업계 카드사 '공공'	Go to Website	K64	금융업
2013-10-17	news	economy	연합뉴스	S	논란	<동양 계열사, 대주주 입김 속 회생절차 밝나 '논란'>(종합)	Go to Website	C30	자동차 및 트레일러 제조업
2013-10-16	news	economy	머니투데이	G	감찰	'슈퍼 감찰' 논란 대우조선 직원 금융거래 공개	Go to Website	C31	기타 운송장비 제조업
2013-10-16	news	economy	데일리안	E	환경오염	리콜 1위 논란에 르노삼성 발끈 "현대-기아차가 더 많아"	Go to Website	C30	자동차 및 트레일러 제조업
2020-11-28	news	economy	아이뉴스24	S	하도급문제	아모레퍼시픽 사정 '육설 시과...대리점 감찰은 오인'	Go to Website	C20	화학 물질 및 화학제품 제조업, 의약품 제외
2020-11-28	news	economy	내일신문	G	주가지작	셀트리온, 증권사까지 동원해 주가지작	Go to Website	C21	의료용 물질 및 의약품 제조업
2020-11-28	news	economy	아시아경제	G	감찰	아모레, 가맹점에도 '추질'	Go to Website	M71	전문 서비스업
2020-11-28	news	economy	데일리안	S	논란	막달 짜문 아모레퍼시픽, 이번엔 제품 암호작업 논란?	Go to Website	C20	화학 물질 및 화학제품 제조업, 의약품 제외
2020-11-28	news	economy	머니투데이	S	금융상권위협	LG하우시스 금융상권 침해 '논란'	Go to Website	C22	고무 및 플라스틱제품 제조업
2020-11-28	news	economy	머니투데이	G	분식회계	GS건설 투자자, 집단소송 제기 '분식논란 규명'	Go to Website	F41	종합 건설업
2020-11-28	news	economy	한겨레	G	일감몰아주기	일감 몰아주기 규제, 이중 상충 예외조항...법제대기 우려	Go to Website	100	분류 제외, 기타
2020-11-28	news	economy	국민일보	S	불매운동	[비즈카페] 불책야크 회장 공판 난동에 불매운동 조짐	Go to Website	C14	의복, 의복 액세서리 및 모피제품 제조업
2020-11-28	news	economy	머니투데이	G	배임	경영진 행정배임설 기법, 남은건 반토만 추가	Go to Website	K64	금융업
2020-11-28	news	economy	YTN	G	탈세	'국세청, 조석래 효성 회장 탈세혐의로 검찰 고발'	Go to Website	100	분류 제외, 기타
2020-11-28	news	economy	아이뉴스24	G	밀어내기	남영유업 대리점측 밀어내기 증거 '일부 조작' 의혹	Go to Website	C10	식품 제조업

1차 및 2차 태깅 후 최종 분류한 기사리스트 일부는 <그림 2>에 있다. 이 리스트는 날짜, 카테고리, 섹션, 언론사, 분류, 키워드, 제목, 본문링크, 산업라벨, 산업명 등의 정보를 포함한다.

다. 본문링크를 통해 뉴스원문을 직접 확인할 수 있다.

ESG관련 우려사안 사건기사 데이터의 연도별 분류는 <표 3>에 정리되어 있다⁸⁾.

8) 사건기사 빈도 수만을 대상으로 분석할 경우 키워드 도출 연구결과에 편이가 존재할 수 있다. 동일 사건에 대한 기사를 하나로 간주한다거나, 메이저 언론사의 기사만을 대상으로 한다거나, 사건 관련

<표 3> 연도별 ESG관련 우려사안 사건기사

연도	환경(E)	사회(S)	지배구조(G)	ESG합계
2013	15	604	2,115	2,734
2014	11	427	1,674	2,112
2015	18	1,283	1,580	2,881
2016	10	1,004	3,528	4,542
2017	13	865	3,675	4,553
2018	36	3,685	3,294	7,015
합계	103 (0.4%)	7,868 (33.0%)	15,866 (66.6%)	23,837 (100.0%)

<표 4> 산업별 ESG관련 우려사안 사건기사

산업	기업수	환경(E)	사회(S)	지배구조(G)	ESG합계
제조	31	5	24	584	613(2.6%)
운송유통	65	36	4,561	4,653	9,250(38.8%)
서비스통신	26	6	565	2,459	3,030(12.7%)
철강건설전기	85	40	1,454	3,526	5,020(21.1%)
전자	33	6	767	2,749	3,522(14.8%)
소매	34	2	361	501	864(3.6%)
의료	20	8	136	1,394	1,538(6.5%)
합계	294	103	7,868	15,866	23,837(100%)

전체 구성을 보면 환경이 103개(0.4%)에 불과한 반면 사회는 7,868개(33.0%), 지배구조는 15,866개(66.6%)로 지배구조 관련이슈가 2/3를, 사회관련 이슈가 1/3을 차지하고 있다. 기존에 CSR에서 배제된 지배구조가 다수를 차지하고 있는 것은 ESG관련 연구가 CSR관련 연구와 차별화되는 이유이다. 반면 환경관련 기사가 적은 이유는 폐수방류나 환경오염 등 환경관련

이슈는 대형사건에 기인하기 때문으로 추론된다. 반면 지배구조관련 기사가 많은 이유는 비리, 뇌물, 담합, 분식회계 등 주로 재벌총수나 경영진이 주체가 되어 일어나는 사건이 기사화되기 용이하고 대중의 관심을 끌기 때문으로 추정된다.

최초 기사만을 분석대상으로 할 경우 도출되는 키워드는 상이해 질 수 있을 것으로 예상된다. 하지만 본 연구의 주요 목적 중 하나가 ‘연관규칙 분석’을 통한 키워드 도출이므로, 키워드의 동시출현 방향성도 함께 고려하는 연관규칙 분석을 위해서는 ‘출현빈도’가 가장 적절한 집계 방식이라고 판단했다. 추가적인 집계 방식을 통한 분석은 추후 연구과제로 남겨놓는다.

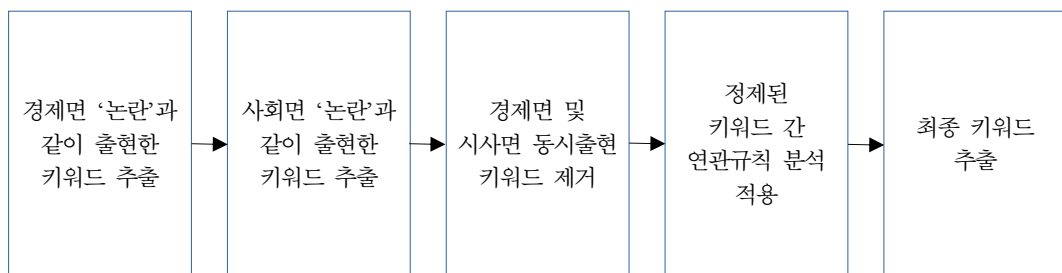
ESG 우려사안을 산업별로 구분한 결과는 <표 4>에 정리되어 있다. 산업별 ESG관련 우려사안 사건기사는 운송/유통과 철강/건설/전기가 각각 9,250개(38.8%), 5,020개(21.1%)를 차지하고 있어 전체의 60%에 해당한다. 운송유통업이 다른 산업에 비해 기업 당 사회와 지배구조 관련 우려사안 비율이 높은 이유는 대한항공의 땅콩회항과 최근의 경영권다툼 때문으로 분석된다. 또한 최근 많이 언급되는 ‘갑질’사건이나 불공정거래가 주로 계약관계에 약자에 해당하는 협력업체나 하도급업체가 많은 유통 및 건설업에서 자주 발생되기 때문으로 보인다.

3.3 키워드 추출 및 정제

기업들의 ESG 우려사안 관련 기사를 대상으로 다음 <그림 3>과 같은 과정을 거쳐 키워드를 도출한다. 먼저 답서치 데이터베이스에서 검색된 기사 중 경제면 뉴스의 각 뉴스별 단어들을 추출한다. 추출된 단어가 ‘논란’이라는 키워드와 빈번하게 동시발생할 경우, ESG 우려사안 키워드 후보로 선정한다. 이후, 답서치 데이터베이스에서 검색된 사회면 뉴스의 각 뉴스별 단어들을 추출한다. 만약 앞서 추려낸 ESG 우

려사안 키워드 후보가 사회면의 뉴스에서도 빈번히 발생한다면, 이는 ESG 우려사안에만 특정한 단어가 아니라 ‘논란’이라는 단어와 보편적으로 빈번히 나타나는 단어로 판단하여 후보에서 제외한다. 즉, 기준이 되는 단어가 매우 보편적임에 따라 이와 함께 검색되는 키워드들이 일반적으로 사용되는 단어인지, 혹은 기업의 ESG 우려사안과 관련한 단어인지 구분할 필요가 있었다. 만약 경제면과 시사면 구분없이 크게 등장하는 키워드라면 기업활동과 관련된 단어일 가능성이 낮은 반면 경제면에서 비중이 높은 키워드들은 기업활동과 관련된 단어일 가능성이 높을 것이기 때문이다. 이 같은 후보 단어들을 정제과정을 거쳐 최종적으로 ESG 우려사안 키워드를 선정한다. 기업의 ESG 우려사안을 실시간으로 보도하는 뉴스 기사를 통해 추출하였기에, ESG 우려사안에 대해 요구되는 높은 시의성을 충족하면서 연구자의 주관을 배제할 수 있는 키워드를 도출했다고 할 수 있다.

<표 5>는 키워드 추출 및 정제과정을 거쳐 도출된 키워드를 보여준다. 총 49개의 ESG 우려사안 관련 잠재적 키워드 중 정제과정을 거쳐 ‘고액체납’ ‘독과점’ ‘리베이트’ 등 21개 키워드가 삭제되고 총 28개 키워드가 도출되었다.



<그림 3> 키워드 선정과정 흐름도

<표 5> 정제 전후의 ESG우려사안 키워드

정제전	정제후	정제전	정제후
갑질	갑질	블랙기업	삭제
경영비리	삭제	비리	비리
계약사불법지원	삭제	비자금	비자금
고액체납	삭제	살인기업	삭제
고용세습	삭제	성과급잔치	성과급잔치
골목상권위협	골목상권위협	소비자우롱	삭제
군대식조직	삭제	숨방망이처벌	삭제
기습인상	삭제	연비조작	삭제
내부자거래	내부자거래	이물질	이물질
논란	논란	일감몰아주기	일감몰아주기
뇌물	뇌물	조세회피	조세회피
담합	담합	주가조작	주가조작
독과점	삭제	집단손배소	삭제
리베이트	삭제	차명계좌	차명계좌
미공개정보이용	삭제	차명주주	차명주주
밀어내기	밀어내기	채용비리	삭제
배임	배임	총수구속	총수구속
부당해고	삭제	탈루	삭제
분식회계	분식회계	탈세	탈세
불공정거래	불공정거래	폐수방류	삭제
불매운동	불매운동	품질조작	삭제
불법고용	불법고용	하도급문제	하도급문제
불법상속	불법상속	환경오염	환경오염
불법증여	삭제	횡령	횡령
불법과건	불법과건		

3.4 연관규칙 분석

본 연구의 목적 중 하나는 단순 키워드의 도출이 아닌 키워드간 관계 파악에 있다. 언어 네트워크 분야에서 키워드간 관계는 동시출현으로 설명할 수 있어 이를 위해 연관규칙 분석을 활용한다. ESG 우려사안의 경우 관련된 다양

한 키워드가 존재하고 하나의 키워드가 출현할 때 다른 키워드도 동시에 출현하는 경우가 많다. 따라서 동시 출현의 방향성을 고려한 연관관계 측정이 의미가 있을 것으로 판단한다.

연관규칙 분석은 단순 키워드 검색으로 인한 중복문제나 키워드 선정의 주관성을 배제함과 동시에 키워드 간의 관계를 파악할 수 있다는

장점이 있다. 유사한 방법론인 TF-IDF나 토픽 모델링(topic modeling)의 경우, 키워드들의 문서 내 발생 빈도를 바탕으로 유사도를 계산할 수 있지만, 개별 키워드들 간의 발생 연관성은 파악할 수 없다. 또한 연관규칙 분석은 키워드 군집화가 가능하다. 연관규칙 분석의 군집화는 사전에 군집의 개수를 설정해야 하는 토픽모델링 방식과는 다르게, 높은 연관성을 갖는 키워드들이 자연스럽게 네트워크상에서 군집을 형성하므로, 군집의 개수에 대한 연구자의 사전 설정이 요구되지 않는다는 장점도 있다.

본 연구에서 연관규칙AB란 어떤 회사에 A라는 ESG관련 우려사안이 발생하면 B라는 우려사안도 발생한다는 패턴을 의미한다. 연관규

칙 분석은 지지도(support), 신뢰도(confidence), 향상도(lift)를 통해 분석된다. 각 ESG관련 이슈 키워드가 전체 ESG 이슈에서 차지하는 비율이 지지도가 된다. A라는 ESG관련 이슈를 발생시킨 기업이 B라는 ESG 이슈 또한 발생시키는 비율은 신뢰도가 된다. 키워드A와 키워드B와 관련해서 계산된 신뢰도를 키워드B가 발생한 확률로 나눈 값은 향상도(lift)도 된다. 즉, 향상도는 키워드B가 임의로 나타나는 경우와 키워드A와의 관계를 고려하여 나타나는 경우의 비율이며, 향상도(A→B)와 향상도(B→A)는 서로 동일한 값을 가진다.

정제과정을 거쳐 선정된 ESG 우려사안 키워드를 대상으로 머신러닝 알고리즘인 연관규칙

	비자금	뇌물	횡령	비리	배임	총수 구속	불법 상속	조세 회피	갑질	주가 조작	내부자 거래	분석 회계	탈세	일감 몰아주기	불법 파견	하도급 문제	불법 고용	불공정 거래	성과급 담합	자명 주주	공목상권 위협	환경 오염	차명 계좌	불매 운동	이물질	밀어 내기	논란	
비자금	100.0	39.7	53.2	65.8	45.8	4.7	0.7	14.9	32.5	23.4	1.4	3.7	6.1	10.5	2.7	1.4	0.7	5.4	7.1	2.0	0.0	2.4	3.4	3.7	0.7	2.0	1.0	
뇌물	38.0	100.0	51.3	74.4	46.1	5.5	0.0	11.0	34.7	20.5	2.3	3.6	6.2	9.4	2.6	1.0	0.3	4.9	7.5	2.6	0.0	0.3	1.6	3.9	2.9	1.0	1.6	0.6
횡령	30.5	30.7	100.0	51.3	61.0	3.1	0.0	8.0	21.2	17.9	1.4	2.3	5.2	8.0	2.1	0.8	0.8	3.7	6.4	2.5	0.0	0.2	1.6	4.3	2.7	0.6	1.6	1.0
비리	30.0	35.4	40.8	100.0	35.1	3.2	0.5	8.7	27.4	16.2	0.9	2.5	3.9	7.7	1.5	0.9	0.5	3.2	6.2	2.2	0.0	0.0	1.4	3.2	2.6	0.6	1.2	0.9
배임	27.7	29.1	64.3	46.5	100.0	2.5	0.2	7.6	24.2	16.6	0.8	2.0	4.1	8.6	2.5	0.4	0.6	3.1	6.4	2.3	0.0	0.0	1.2	3.5	1.8	0.2	1.4	1.4
총수구속	66.7	81.0	76.2	81.0	불법상속을 비리	100.0	0.0	42.9	61.9	33.3	0.0	4.8	9.5	14.3	14.3	4.8	9.5	4.8	0.0	0.0	0.0	0.0	0.0	9.5	14.3	4.8	0.0	0.0
불법상속	66.7	0.0	0.0	100.0	비리	100.0	0.0	0.0	33.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	33.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
조세회피	49.4	38.2	46.1	62.9		0.0	100.0	50.6	29.2	2.2	2.2	7.9	12.4	1.1	1.1	0.0	6.7	7.9	4.5	0.0	0.0	1.1	4.5	1.1	0.0	0.0	0.0	0.0
갑질	19.3	21.5	21.9	35.6	23.7	2.6	0.0	9.1	100.0	9.1	0.8	2.4	5.8	9.1	3.2	0.8	0.8	5.4	6.6	1.6	0.0	0.2	1.8	3.2	3.4	0.4	1.4	0.6
주가조작	28.5	26.0	38.0	43.4	33.5	2.9	0.4	10.7	18.6	100.0	2.1	3.7	4.1	6.2	0.8	0.4	0.0	3.3	5.4	1.2	0.0	0.0	1.2	2.5	1.7	0.0	0.8	0.8
내부자거래	6.8	11.9	11.9	10.2	6.8	0.0	0.0	3.4	6.8	8.5	100.0	30.5	49.2	37.3	13.6	8.5	1.7	47.5	49.2	13.6	3.4	0.0	18.6	23.7	27.1	0.0	5.1	1.7
분석회계	6.5	6.5	7.1	9.5	5.9	0.6	0.0	1.2	7.1	5.3	10.7	100.0	27.2	26.0	7.1	5.9	1.8	21.9	39.6	11.8	1.8	1.8	10.1	13.6	11.8	0.0	2.4	1.2
탈세	7.0	7.4	10.5	9.7	7.8	0.8	0.0	2.7	11.2	3.9	11.2	17.8	100.0	37.2	11.6	6.2	3.5	28.7	43.8	15.1	1.6	2.3	12.4	25.6	23.3	0.4	1.6	1.9
일감몰아주기	7.8	7.3	10.3	12.5	10.5	0.8	0.0	2.8	11.3	3.8	5.5	11.0	24.0	100.0	11.5	6.3	2.5	18.8	34.8	13.3	0.5	1.5	8.3	13.0	13.3	0.8	1.0	1.5
불법파견	9.2	9.2	12.6	11.5	13.8	3.4	0.0	1.1	18.4	2.3	9.2	13.8	34.5	52.9	100.0	18.4	11.5	33.3	43.7	21.8	0.0	6.9	24.1	23.0	32.2	1.1	2.3	1.1
하도급문제	10.0	7.5	10.0	15.0	5.0	2.5	0.0	2.5	10.0	2.5	12.5	25.0	40.0	62.5	40.0	100.0	12.5	35.0	52.5	45.0	2.5	7.5	27.5	20.0	40.0	2.5	0.0	2.5
불법고용	14.3	7.1	28.6	21.4	21.4	14.3	0.0	0.0	28.6	0.0	7.1	21.4	64.3	71.4	71.4	35.7	100.0	42.9	78.6	42.9	0.0	21.4	35.7	64.3	64.3	7.1	0.0	0.0
불공정거래	6.3	6.0	7.5	8.3	6.0	0.4	0.0	2.4	10.7	3.2	11.1	14.7	29.4	29.8	11.5	5.6	2.4	100.0	38.9	14.3	0.4	4.0	10.3	14.3	21.4	0.4	1.6	1.6
담합	3.9	4.2	6.1	7.3	5.7	0.4	0.2	1.3	6.1	2.4	5.3	12.3	20.7	25.5	7.0	3.9	2.0	18.0	100.0	12.5	0.6	1.7	9.4	12.3	13.8	1.3	2.2	1.3
성과급잔치	4.8	6.4	10.4	11.2	8.8	0.8	0.0	3.2	6.4	2.4	6.4	16.0	31.2															
자명주주	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50.0	75.0	100.0	2	14.4	4.8	28.8	54.4	100.0	0.8	2.4	13.6	20.8	19.2	0.8	1.6	2.4	
공목상권위협	0.0	7.1	7.1	0.0	0.0	0.0	0.0	0.0	7.1	0.0	0.0	21.4	42.9	0	25.0	0.0	25.0	75.0	25.0	100.0	0.0	25.0	91.0	25.0	0.0	0.0	0.0	
환경오염	6.5	4.6	7.4	8.3	5.6	0.0	0.0	0.9	8.3	2.8	10.2	15.7	29.6	30.6	19.4	10.2	4.6	24.1	47.2	15.7	0.9	1.9	100.0	17.6	22.2	1.9	3.7	1.9
차명계좌	8.1	9.7	17.7	16.9	13.7	1.6	0.0	3.2	12.9	4.8	11.3	18.5	53.2	41.9	16.1	6.5	7.3	29.0	54.0	21.0	3.2	1.6	15.3	100.0	22.6	0.8	1.6	1.6
불매운동	7.0	5.7	8.9	10.8	5.7	1.9	0.0	0.6	10.8	2.5	10.1	12.7	38.0	33.5	17.7	10.1	5.7	34.2	47.5	15.2	0.6	5.1	15.2	17.7	100.0	1.3	1.9	0.6
이물질	2.0	3.0	3.0	4.0	1.0	1.0	0.0	0.0	2.0	0.0	0.0	0.0	1.0	3.0	1.0	1.0	1.0	1.0	7.1	1.0	0.0	1.0	2.0	1.0	2.0	100.0	29.3	25.3
밀어내기	1.6	1.3	2.1	2.1	1.9	0.0	0.0	0.0	1.9	0.5	0.8	1.1	1.1	1.1	0.5	0.0	0.0	1.1	3.2	0.5	0.0	0.3	1.1	0.5	0.8	7.7	100.0	20.2
논란	1.5	1.0	2.4	2.9	3.4	0.0	0.0	0.0	1.5	1.0	0.5	1.0	2.4	2.9	0.5	0.5	0.0	1.9	3.4	1.5	0.0	0.0	1.0	1.0	0.5	12.1	36.9	100.0

<그림 4> ESG 우려사안 키워드별 신뢰도 매트릭스

분석을 수행하여 ESG관련 부정적인 사건의 키워드들의 동시발생 관계를 분석하였다. 즉, ESG사건이 발생한 기업에서 또 다른 ESG사건이 발생하는 경우, 해당 ESG사건 키워드들을 군집화하여 유사한 ESG사건 키워드로 간주하였다. 또한 개별 ESG사건 키워드는 최초 키워드 설정 과정에서 주관성이 개입된 것으로 판단하여 분석 대상에서 제외하였다. 다만 특정 ESG사건 키워드가 제거 대상이지만 다른 키워드와의 신뢰도가 1을 나타낸 경우, 두 키워드의 발생 빈도값을 합산하였다. <그림 4>는 ESG 우려사안에 해당하는 키워드 별 신뢰도 매트릭스와 개별 키워드 신뢰도이다. 분석결과 정제 과정을 거쳐 선택된 ‘불법상속’의 경우, 연관규칙 분석에 따라 ‘불법상속’ 이슈 발생 시 ‘비리’ 이슈가 100% 발생하는 패턴을 보였다. 또한 ‘차명주주’와 ‘탈세’도 동시에 발생하는 패턴을

보였다. 따라서 ‘불법상속’ 뉴스 기사 빈도값은 ‘비리’로, ‘차명주주’ 뉴스 기사 빈도 값은 ‘탈세’로 이전하였다.

IV. 연구결과

4.1 ESG 우려사안 관련 키워드

연관규칙 분석을 활용해 설정 및 정제과정을 통해 최종적으로 도출된 ESG 우려사안 키워드는 총 26개로 <표 6>에 정리되어 있다. 분석 결과 ‘갑질’이 5,857건(24.6%)으로 압도적으로 많았다. 이는 사회구조적으로 아직도 갑의 강압적 역할이 존재한다는 의미와 동시에 투자자들이 ‘갑질’을 기업들의 반 ESG 활동 중 가장 중

<표 6> 최종 선정된 ESG우려사안 키워드 빈도수

순위	키워드	빈도수	순위	키워드	빈도수
1	갑질	5,857	14	밀어내기	337
2	비리	3,474	15	차명계좌	289
3	뇌물	2,008	16	주가조작	243
4	담합	1,766	17	불공정거래	220
5	분식회계	1,257	18	성과급잔치	211
6	비자금	1,246	19	하도급문제	206
7	일감 몰아주기	1,234	20	총수구속	180
8	배임	1,179	21	조세회피	141
9	횡령	1,064	22	논란	92
10	탈세	904	23	환경오염	55
11	불법파견	760	24	이물질	48
12	불매운동	615	25	내부자거래	21
13	불법고용	416	26	골목상권 위협	14

대하게 인식하고 있다는 것을 의미한다. 이어서 ‘비리’ 3,474건(14.6%), ‘뇌물’ 2,008건(8.4%), ‘담합’ 1,766건(7.4%)이 그 뒤를 이었다. 이들 4개 단어는 모두 지배구조와 관련된 키워드로 사회정의 실천에 기업들의 지배구조 문제가 가장 큰 걸림돌인 것으로 해석할 수 있다⁹⁾. 이들 상위 4개 키워드의 누적빈도수는 50% 이상을 넘어섰다. 즉, ESG 우려사안 키워드는 행위 주체가 대주주나 경영진의 사회적 무책임(CSiR: corporate social irresponsibility)과 관련된 사건으로 그것에 수반되는 미디어의 관심을 반영한다고 볼 수 있다.

또한 이들 키워드는 ‘성과급잔치’나 ‘이물질’과 같이 비교적 낮은 수준의 기업의 일탈행위부터 ‘총수구속’과 같이 검찰 등 국가기관의 조사가 이루어지는 수준까지 광범위한 반 ESG 관련 활동을 포함한다고 볼 수 있다. 특히 부정적인 ESG사건 범위의 유동성과 시의성이 반영되어 “갑질”이나 “밀어내기”와 같은 키워드가 ESG 우려사안에 편입된 것으로 보인다. 이를 통해 투자자들은 ESG사건과 관련된 사회적 파장과 심각성, 재발 가능성 등을 폭넓게 고려하고 있는 것으로 추론된다.

이 같은 키워드는 해외 ESG정보 제공업체인 Refinitiv에서 사용하는 키워드와 비교했을 때 비즈니스윤리(business ethics)나 불공정거래(anti-competition), 인권(human right), 회계(accounting), 내부자거래(insider dealings) 등의 우려사안에 대해서는 공통부분이 있으나, 소비자(consumer), 지적재산권(intellectual property), 사생활(privacy), 공공보건(public

health)등의 우려사안에 대해서는 아직 국내에서는 크게 이슈가 되지 않는 것으로 나타났다.

4.2 ESG 우려사안 키워드 분류

ESG 우려사안으로 도출된 26개 키워드에 대해 환경, 사회, 지배구조에 대한 구분과 <표 1>에서 제시된 ESG 모범기준에 따른 대분류는 <표 7>에 정리되어 있다. 최종 선정된 키워드 중 지배구조 관련 키워드는 16개, 사회관련 키워드는 8개였으나 환경관련은 2개에 불과했다.

한 가지 흥미로운 사실은 ESG 우려사안 상위 10개 키워드들은 모두 ESG 모범기준 상 주주와 관련된 것이라는 점이다. 소분류로 보면 주주의 책임과 관련된 사항이다. 즉 일명 오너로 일컬어지는 최대주주 또는 지배주주들이 사회적 규범을 벗어난 행동을 하는 경우 투자자들은 기업의 지속가능성을 해치는 주요 요인으로 판단하고 있는 것이다. 지배구조 대분류 상 가장 많은 키워드를 포함한 것은 ‘분식회계’, ‘배임’, ‘횡령’ 등 감사기구와 관련된 것이다. 최근 ‘주식회사 등 외부감사에 관한 법률’이 강화된 이유와 맥을 같이 한다고 할 수 있다. 기업들이 회계투명성을 제고하고 내부통제체도를 강화할 필요성이 제기되는 대목이다.

사회부문과 관련해서는 ‘담합’, ‘불공정거래’ 등 공정거래와 관련된 키워드가 가장 많았다. 반면 최근 중요성이 증대되고 있는 소비자 권익증대와 관련된 키워드는 ‘불매운동’ 하나에 그쳤다. 환경부문은 ‘환경오염’이나 ‘이물질’ 등 대분류 상 환경경영실행과 관련된 것이었다.

9) 갑질이나 비리, 뇌물 등은 지배주주가 아닌 사회문제로도 분류할 수 있으나 관련기사에 등장하는 대부분의 행위주체가 대주주나 경영진이어서 지배주주로 분류한다.

<표 7> 최종 선정된 ESG우려사안 키워드 분류

	키워드	분류	대분류		키워드	분류	대분류
1	갑질	지배구조	주주	14	밀어내기	지배구조	이해관계자
2	비리	지배구조	주주	15	차명계좌	지배구조	감사기구
3	뇌물	지배구조	주주	16	주가조작	지배구조	시장에 의한 경영감시
4	담합	사회	협력사 및 경쟁사	17	불공정거래	사회	협력사 및 경쟁사
5	분식회계	지배구조	감사기구	18	성과급잔치	지배구조	이사회
6	비자금	지배구조	주주	19	하도급문제	사회	협력사 및 경쟁사
7	일감 몰아주기	지배구조	주주	20	총수구속	지배구조	이사회
8	배임	지배구조	감사기구	21	조세회피	지배구조	감사기구
9	횡령	지배구조	감사기구	22	논란	사회	지역사회
10	탈세	지배구조	감사기구	23	환경오염	환경	환경경영 실행
11	불법과건	사회	근로자	24	이물질	환경	환경경영 실행
12	불매운동	사회	소비자	25	내부자거래	지배구조	이해관계자
13	불법고용	사회	근로자	26	골목상권 위협	사회	협력사 및 경쟁사

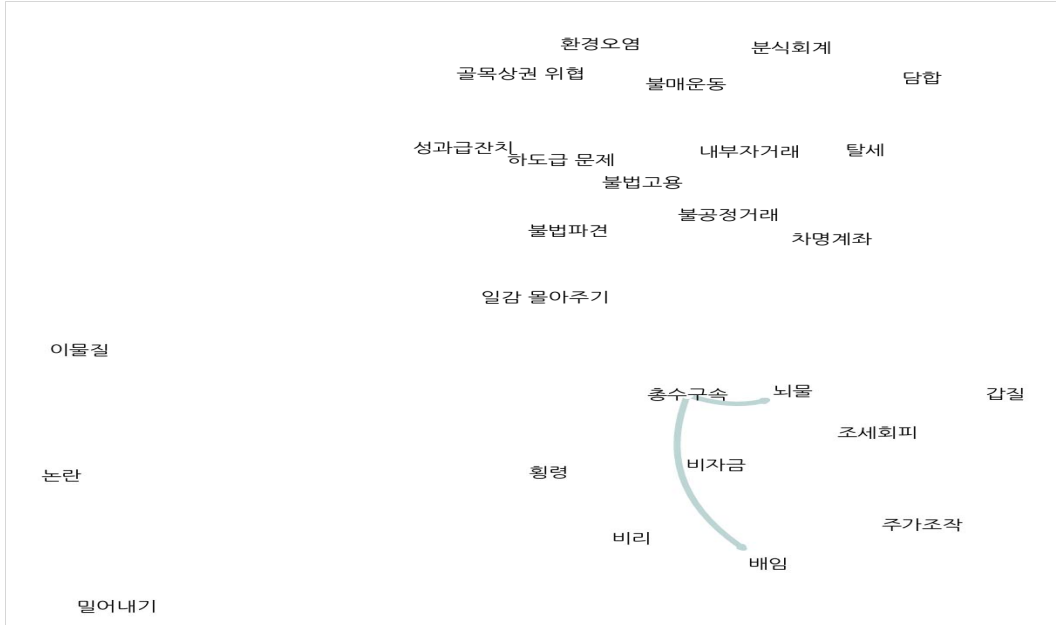
4.3 키워드 신뢰도 분석

ESG 우려사안 키워드 26개의 신뢰도 분석은 <그림 5>부터 <그림 8>까지 도식화 되어있다. <그림 5>는 신뢰도 80% 이상, <그림 6>은 신뢰도 50% 이상, <그림 7>은 신뢰도 30% 이상, <그림 8>은 신뢰도 10% 이상의 관계를 보이는 키워드를 연결하여 나타낸다.

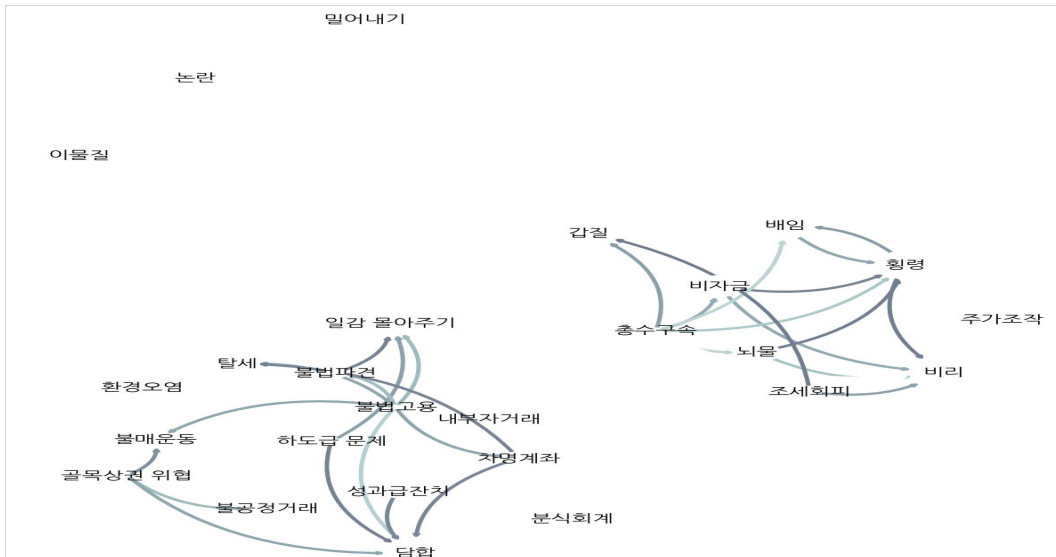
<그림 5>에 따르면, 80% 이상의 신뢰도를 보이는 키워드는 총수구속과 배임, 뇌물만이 존재한다. 이들의 관계는 총수구속이 발생하는 경우 80% 이상이 배임과 뇌물 키워드를 동반하는 것으로 배임과 뇌물이 총수구속의 주 원인으로

작용하고 있음을 시사한다. <그림 6>은 신뢰도 50% 이상의 키워드 관계를 보여주는 것으로 크게 사회책임 부문과 지배구조 부문 등 두 군집으로 구분된다. 사회책임 부문은 근로자 및 공정거래와 관련된 이슈들이고, 지배구조 부문은 비자금, 횡령, 비리 등 최대주주나 경영자들의 일탈행위와 관련된 이슈들이다. 한 가지 흥미로운 사실은 사회책임 이슈와 지배구조 이슈가 각각 집단화되어 있으나 일감몰아주거나 차명계좌의 경우에는 지배구조가 아닌 사회책임 이슈와의 신뢰도가 더 높게 나타나 이들 이슈는 지배구조와 사회책임 이슈를 모두 포함하는 것으로 추론할 수 있다.

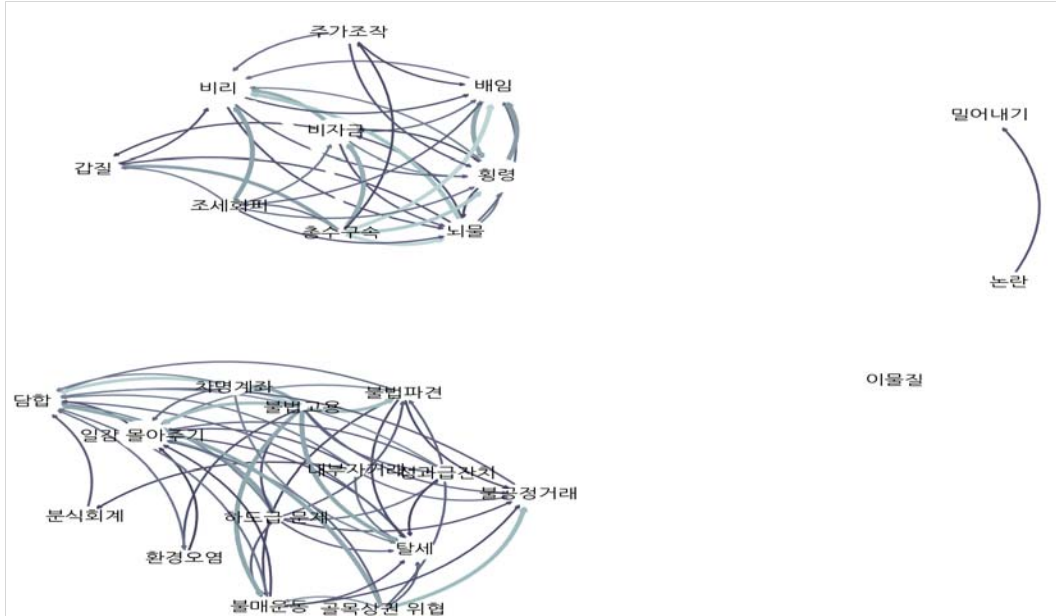
<그림 5> 신뢰도 80% 이상의 키워드 네트워크



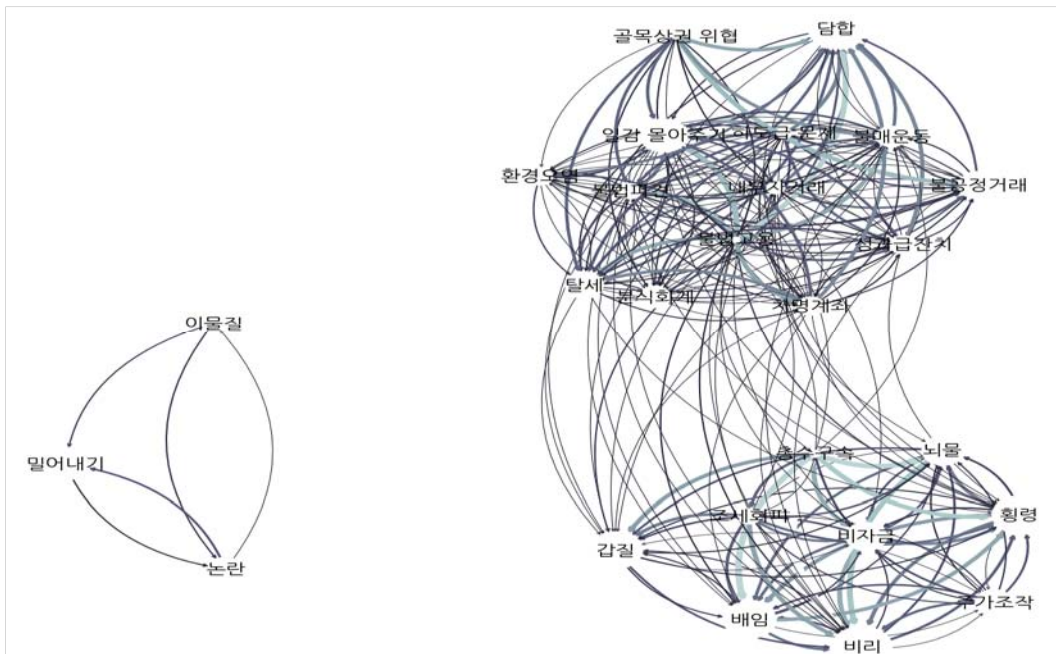
<그림 6> 신뢰도 50% 이상의 키워드 네트워크



<그림 7> 신뢰도 30% 이상의 키워드 네트워크



<그림 8> 신뢰도 10% 이상의 키워드 네트워크



신뢰도 30% 이상의 네트워크를 도식화한 <그림 7>도 <그림 6>과 유사하게 지배구조와 사회책임 두 부문이 각각 집단화되어 나타나고 있으나 <그림 6>에 비해 관련도가 있는 키워드들이 증가한 모습을 볼 수 있다. 지배구조 집단에서는 ‘비자금’이 중심에 위치하여 다른 키워드와 많은 신뢰도를 보이고 있는 반면 사회책임 집단에서는 ‘불공정거래’와 ‘담합’이 양 극단에 위치하며 다른 키워드와의 많은 신뢰도 관계를 보이고 있다.

<그림 8>은 신뢰도가 10% 이상인 키워드 관계를 도식화 한 것으로 <그림 6>, <그림 7>과 같이 기본적으로 지배구조와 사회책임 두 부문으로 집단화 되어있다. 다만 신뢰도가 낮아진 만큼 네트워크 연결선은 증가된다. 또한 사회책임을 나타내는 키워드와 지배구조를 의미하는 키워드가 상호 연결되어 있어 신뢰도 10% 수준에서는 지배구조 이슈와 사회책임 이슈가 상호 관련성이 높다고 볼 수 있다.

한편 ‘논란’, ‘이물질’, 그리고 ‘밀어내기’ 키워드는 이들 두 집단과 동떨어져 별도의 신뢰도 네트워크를 구성하고 있다. 사회책임관련 키워드인 ‘논란’, 지배구조관련 키워드인 ‘밀어내기’, 그리고 환경관련 키워드인 ‘이물질’이 상호 연결되어 있는 구조는 특이하다고 할 수 있다. 또한 환경오염은 환경관련 키워드임에도 불구하고 신뢰도 면에서는 사회부문 키워드와 연결관계를 보이는 것도 흥미롭다고 할 수 있다.

4.4 키워드 네트워크 분석

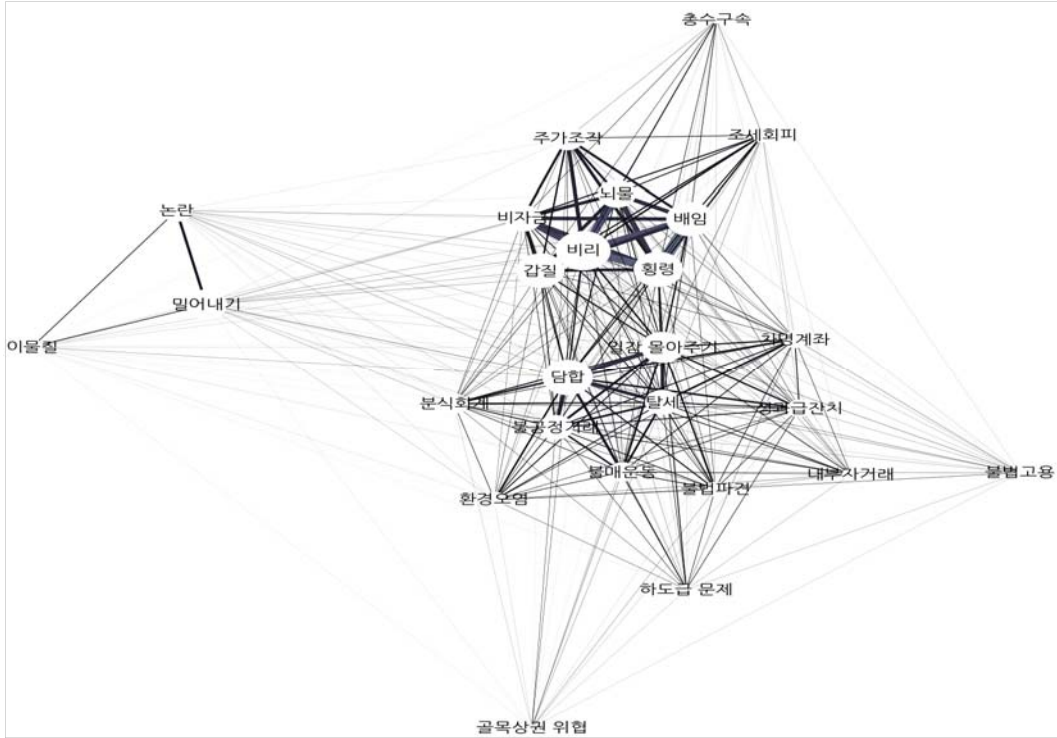
ESG 우려사안과 관련되어 도출된 26개 키워

드들의 빈도에 따른 네트워크 분석 결과는 <그림 9>에 도식화되어 있다. 분석 결과 예상과 같이 빈도수가 많았던 ‘비리’, ‘횡령’이나 ‘담합’, ‘일감 몰아주기’가 네트워크의 중심부에 위치하고 있다. 네트워크 분석 상 중심부에 위치하는 키워드일수록 다른 키워드들과의 연관성이 높고 주변부의 키워드보다 이슈의 파생성이나 보편성이 높다는 점을 의미한다.

개별 키워드의 단독 발생빈도를 나타내는 원의 크기는 ‘비리’가 발생빈도가 가장 높은 ‘갑질’보다 큰 것으로 나타났다. 이는 ‘갑질’은 다른 키워드와 동반하여 발생하는 반면 ‘비리’는 독립적으로 발생하는 경우가 많음을 의미한다.

또한 키워드 간의 동시 발생빈도를 나타내는 선의 굵기와 키워드간 연관성을 나타내는 거리를 감안할 때, ‘비리’, ‘횡령’, ‘배임’, ‘뇌물’ 등 지배구조관련 키워드는 키워드간 연관성이 상당히 높고 동시발생하는 경우도 많은 반면 ‘몰아주기’, ‘담합’, ‘불공정거래’ 등 사회책임관련 키워드는 지배구조관련 키워드에 비해 연관성과 동시발생빈도가 낮은 것으로 나타났다. 특히 가장 많은 빈도수를 기록한 ‘갑질’의 경우는 ‘비리’와 가장 연관성이 높은 것으로 나타났다. 이러한 연관성 분석은 투자자로 하여금 기업이 어떠한 ESG 우려사안을 발생시켰을 때 후속적으로 발생할 수 있는 ESG 우려사안을 유추할 수 있게 한다. ESG 우려사안별 기업의 가치나 주가 등에 미치는 영향의 크기가 다를 수 있으므로 과거 ESG 우려사안 발생 여부는 리스크 관리관점에서 미래에 있을 기업에 대한 부정적 사건에 대한 신호로 활용될 수 있다.

<그림 9> ESG 우려사안 키워드 빈도의 네트워크



V. 결론 및 시사점

본 연구는 최근 관심이 집중되고 있는 ESG와 관련, 투자자들이 인지하는 기업들의 반 ESG 활동이 무엇인가를 검증했다. ESG 등급이 존재하는 294개 상장기업의 2013년부터 2018년까지 130개 미디어에 노출된 반 ESG 활동 관련 기사 23,837건을 대상으로 머신러닝의 한 방법인 연관규칙 분석을 통해 ESG 우려사안에 대한 키워드를 도출했다.

분석결과 총 26개의 ESG 우려사안 키워드가 도출되었고 ‘감질’이 가장 많은 빈도수를 기록했다. 이어 ‘비리’, ‘뇌물’, ‘담합’ 순으로 빈도수가 높았다. 26개 키워드 중 지배구조관련은

16개, 사회책임관련은 8개, 환경관련은 2개 이었다. 상위랭크를 기록한 키워드는 대부분 지배구조관련으로 투자자들은 후진적 지배구조가 기업의 지속가능성을 저해하는 주요 위험요인으로 파악하고 있는 것으로 밝혀졌다. 또한 지배구조관련 ESG 우려사안 키워드는 대부분 주주의 책임과 관련된 것으로 최대주주나 지배주주의 일탈행위가 기업가치를 훼손시킬 수 있는 것으로 밝혀졌다. 사회책임관련 ESG 우려사안 키워드는 불공정거래와 관련된 것이 많아 아직도 사회의 공정성이 기대수준에 미치지 못하고 있음을 보여주었다.

신뢰도 분석결과 지배구조관련과 사회책임관련 키워드는 각각 군집을 이루어 각 집단 내

에서 키워드간 상호 발생가능성이 높음을 보였다. 특히 ‘충수구속’의 경우 80%의 신뢰도로 ‘뇌물’과 ‘배임’이 그 원인 역할을 하는 것으로 드러났다. 네트워크 분석결과에서는 ‘비리’가 중앙부에 위치하면서 단독 발생가능성이 가장 높고 ‘배임’, ‘횡령’, ‘뇌물’과 높은 관련성이 있음을 보여주었다.

본 연구는 거시적인 수준에서 기업이 발생시키는 예상하지 못한 ESG관련 우려사안에 대한 인지뿐 아니라 미시적인 키워드 수준에서 동시출현 단어의 연관성 분석을 바탕으로 키워드간 방향성과 우려사안 발생 시 초래할 수 있는 결과를 미리 예측해 볼 수 있는 근거를 제공함에 의의가 있다. 또한 추상적인 개념의 ESG 우려사안을 키워드 네트워크를 통해 구조를 쉽게 파악할 수 있게 하여, 향후 실무적으로 투자자나 기업들이 이를 리스크관리 수단으로 활용할 수 있을 것으로 예상된다. 특히, 국민연금이 적극적인 주주활동을 할 수 있는 근거가 되는 ESG 우려사안에 대한 중대성 평가 지표로 활용 가능할 것으로 기대한다.

본 연구 결과는 향후 급증할 것으로 예상되는 ESG 연구에 중요한 기초자료로 활용될 수 있을 것으로 기대한다. 즉, ESG 활동으로 인한 기업가치 상승폭보다 반 ESG 활동으로 인한 기업가치 하락폭이 훨씬 큰 현실을 감안할 때 ESG 우려사안이 무엇인가에 대한 정의가 시급하다고 할 수 있다. 또한 ESG등급은 자료를 제공하는 기관이 존재하는 반면 ESG 우려사안에 대한 정보제공기관은 존재하지 않아 본 연구결과를 활용해 기업들의 ESG 우려사안 등급을 설정할 수 있을 것이다. 이와 함께 기업관점에서 앞 다투어 ESG 경영을 선포하고 있는 상

황에서 ESG와 관련하여 특별히 주의해야 할 행동규범을 인지할 수 있을 것이다.

참고문헌

- 강경수, 박세민, “웹 크롤링과 기계학습 기법을 이용한 경영학 분야 KCI 저널의 주제어 분석,” 대한경영학회지, 제32권, 제4호, 2019, pp. 597-615.
- 안민욱, 원은송, 김성용, 유동희, “향상도 기반 연관규칙 네트워크를 활용한 농식품 판매 전략 개발: 대형마트와 전통시장을 중심으로,” 인터넷전자상거래연구, 제19권, 제3호, 2019, pp. 105-127.
- 이태원, 성행남, 김은정, “키워드 네트워크 분석을 이용한 MIS 교과정보와 NCS 기반 빅데이터 분석 직무역량에 대한 연구,” 정보시스템연구, 제29권, 제4호, 2020, pp. 101-121.
- 전익진, 이학연, “연관규칙 기반 동시출현단어 분석을 활용한 기술경영 연구 주제 네트워크 분석,” 기술혁신연구, 제24권, 제4호, 2016, pp. 101-126.
- 최병설, 김남규, “감정 딥러닝 필터를 활용한 토픽 모델링 방법론,” 정보시스템연구, 제28권, 제4호, 2019, pp. 271-291.
- 황현석, 유기동, “온라인 뉴스 웹사이트의 로그를 이용한 연관규칙 발견에 관한 연구,” 한국산업정보학회논문지, 제18권, 제2호, 2013, pp. 47-57.
- 홍서연, “데이터마이닝을 활용한 성공적 노후 예측 키워드 분석,” 한국콘텐츠학회논문

- 문지, 제20권, 제3호, 2020, pp. 223-234.
- Adams, C. A., "Internal Organisational Factors Influencing Corporate Social and Ethical Reporting: Beyond Current Theorising," *Accounting, Auditing & Accountability Journal*, Vol. 15, No. 2, 2002, pp. 223-250.
- Agrawal, R, T. Imieliński, and A. Swami., "Mining Association Rules between Sets of Items in Large Databases," in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 1993, pp. 207-216.
- Agrawal, R., and Srikant, R., "Fast Algorithms for Mining Association Rules," in *Proceedings of the International Conference on Very Large Databases*, 1994, pp. 125-131.
- Aouadi, A., and S. Marsat., "Do ESG Controversies Matter for Firm Value? Evidence from International Data," *Journal of Business Ethics*, Vol. 151, No. 4, 2018, pp. 1027-1047.
- Arora, P., and R. Dharwadkar., "Corporate Governance and Corporate Social Responsibility (CSR): The Moderating Roles of Attainment Discrepancy and Organization Slack," *Corporate Governance: An International Review*, Vol. 19, No. 2, 2011, pp. 136-152.
- Blei, D., "Probabilistic Topic Models," *Communications of the ACM*, Vol. 55, No. 4, 2012, pp. 77-84.
- Brin, S., and L. Page., "The Anatomy of a Large-Scale Hypertextual Web Search Engine," in *Proceedings of the 7th International World Wide Web Conference*, 1998, pp. 107-117.
- Cai, Y., H. Jo, and C. Pan., "Doing Well While Doing Bad? CSR in Controversial Industry Sectors," *Journal of Business Ethics*, Vol. 108, No. 4, 2012, pp. 467-480.
- Capelle-Blancard, G., and S. Monjon., "Trends in the Literature on Socially Responsible Investment: Looking for the Keys under the Lamppost," *Business Ethics European Review*, Vol. 21, 2012, pp. 239-250.
- Chollet, P., and B. W. Sandwidi., "CSR Engagement and Financial Risk: A Virtuous Circle? International Evidence," *Global Finance Journal*, Vol. 38, 2018, pp. 65-81.
- Christensen, J., and R. Murphy., "The Social Irresponsibility of Corporate Tax Avoidance: Taking CSR to the Bottom Line," *Development*, Vol. 47, No. 3, 2004, pp. 37-44.
- Demacarty, P., "Financial Returns of Corporate Social Responsibility, and the Moral Freedom and Responsibility of Business Leaders," *Business and Society Review*, Vol. 114, No. 3, 2009,

- pp. 393-433.
- Daugaard, D., “Emerging New Themes in Environmental, Social and Governance Investing: A Systematic Literature Review,” *Accounting and Finance*, Vol. 60, No. 2, 2020, pp. 1501-1530.
- Du, S., C. B. Bhattacharya, and S. Sen., “Maximizing Business Returns to Corporate Social Responsibility: The Role of CSR Communication,” *International Journal of Management Reviews*, Vol. 12, No. 1, 2010, pp. 8-19.
- Freeman, S, and S. T. Cavusgil., “Strategic Management. A Stakeholder Approach,” *Journal of International Marketing*, Vol. 15, No. 4, 1984, pp. 1-40.
- Gurun, U. G., and A. W. Butler., “Don’t Believe the Hype: Local Media Slant, Local Advertising, and Firm Value,” *Journal of Finance*, Vol. 67, No. 2, 2012, pp. 561-598.
- Jiao, Y., “Stakeholder Welfare and Firm Value,” *Journal of Banking and Finance*, Vol. 34, No. 10, 2010, pp. 2549-2561.
- Klein, J. and N. Dawar., “Corporate Social Responsibility and Consumers’ Attributions and Brand Evaluations in a Product-Harm Crisis,” *International Journal of Research in Marketing*, Vol. 21, No. 3, 2004, pp. 203-217.
- Kothari, S. P., S. Shu, and P. D. Wysocki., “Do Managers Withhold Bad News,” *Journal of Accounting Research*, Vol. 47, No. 1, 2009, pp. 241-276.
- Kruger, P., “Corporate Goodness and Shareholder Wealth,” *Journal of Financial Economics*, Vol. 115, No. 2, 2015, pp. 304-329.
- Li, J., Z. A. Haider, X. Jin, and W. Yuan., “Corporate Controversy, Social Responsibility and Market Performance: International Evidence,” *Journal of International Financial Markets, Institutions and Money*, Vol. 60, 2019, pp. 1-18.
- Maignan, I. and O. C. Ferrell., “Corporate Social Responsibility and Marketing: An Integrative Framework,” *Journal of the Academy of Marketing Science*, Vol. 32, No. 1, 2004, pp. 3-19.
- Matsuo, Y., and M. Ishizuka., “Keyword Extraction from a Single Document Using Word Co-Occurrence Statistical Information,” *International Journal on Artificial Intelligence Tools*, Vol. 13, No. 1, 2004, pp. 157-169.
- Mihalcea, R., and P. Tarau., “TextRank: Bringing Order into Text.” In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 2004, pp. 404-411.
- Orlitzky, M., “Corporate Social Responsibility, Noise, and Stock Market Volatility,” *Academy of Management Perspectives*,

- Vol. 27, No. 3, 2013, pp. 238-254.
- Pollock, T. G., and V. P. Rindova., “Media Legitimation Effects in the Market for Initial Public Offerings,” *Academy of Management Journal*, Vol. 46, No. 5, 2003, pp. 631-642.
- Siegel, D. S., and D. F. Vitaliano., “An Empirical Analysis of the Strategic Use of Corporate Social Responsibility,” *Journal of Economics and Management Strategy*, Vol. 16, No. 3, 2007, pp. 773-792.
- Steyvers, M., and T. Griffiths., “Probabilistic Topic Models. in *Handbook of Latent Semantic Analysis*,” Vol 427, No. 7, 2007, pp. 424-440.
- Tan, P. N., Steinbach, M., and Kumar, V., “Introduction to Data Mining,” *Intelligent Systems Reference Library*, Vol. 12, 2005, pp. 1-43.
- Tetlock, P. C., “Giving Content to Investor Sentiment: The Role of Media in the Stock Market,” *Journal of Finance*, Vol. 62, No. 3, 2007, pp. 1139-1168.
- Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy., “More Than Words: Quantifying Language to Measure Firms’ Fundamentals,” *Journal of Finance*, Vol. 63, No. 3, 2008, pp. 1437-1467.

안 태 욱 (Ahn, Tae Wook)



Drexel University에서 경영학석사(MBA)학위를 취득하고 현재 딥서치에서 CSO (Chief Strategy Officer)로 재직하고 있다. 주요 관심분야는 Big Data Analytics, ESG, Text Mining 등이다.

이 희 승 (Lee, Hee Sung)



경희대학교에서 경영학 학사를 취득하고 현재 카이스트 경영대학에서 석사과정에 재학 중이다. 주요 관심분야는 금융 데이터, 딥러닝, ESG 투자, 자산배분 등이다.

이 준 서 (Yi, June Suh)



Syracuse대학교에서 경영학박사(재무)학위를 취득하였다. 현재 동국대학교 경영학과 교수로 재직하고 있으며, 주요 관심분야는 펀드, APM, 지배구조 등이다.

<Abstract>

A Study on the Keyword Extraction for ESG Controversies Through Association Rule Mining

Ahn, Tae Wook · Lee, Hee Seung · Yi, June Suh

Purpose

The purpose of this study is to define the anti-ESG activities of companies recognized by media by reflecting ESG recently attracted attention. This study extracts keywords for ESG controversies through association rule mining.

Design/methodology/approach

A research framework is designed to extract keywords for ESG controversies as follows: 1) From DeepSearch DB, we collect 23,837 articles on anti-ESG activities exposed to 130 media from 2013 to 2018 of 294 listed companies with ESG ratings 2) We set keywords related to environment, social, and governance, and delete or merge them with other keywords based on the support, confidence, and lift derived from association rule mining. 3) We illustrate the importance of keywords and the relevance between keywords through density, degree centrality, and closeness centrality on network analysis.

Findings

We identify a total of 26 keywords for ESG controversies. ‘Gapjil’ records the highest frequency, followed by ‘corruption’, ‘bribery’, and ‘collusion’. Out of the 26 keywords, 16 are related to governance, 8 to social, and 2 to environment. The keywords ranked high are mostly related to the responsibility of shareholders within corporate governance. ESG controversies associated with social issues are often related to unfair trade. As a result of confidence analysis, the keywords related to social and governance are clustered and the probability of mutual occurrence between keywords is high within each group. In particular, in the case of “owner’s arrest”, it is caused by “bribery” and “misappropriation” with an 80% confidence level. The result of network analysis shows that ‘corruption’ is located in the center, which is the most likely to occur alone, and is

highly related to ‘breach of duty’, ‘embezzlement’, and ‘bribery’.

Keyword: ESG, Controversies, Association Rule Mining, Network Analysis

* 이 논문은 2021년 3월 4일 접수, 2021년 3월 9일 1차 심사, 2021년 3월 30일 게재 확정되었습니다.