한국기록관리학회지
26(3), 159-180, 2026.8
http://dx.doi.org/10.14404/JKSARM.2026.26.3.159
pISSN 1598-1487 eISSN 2671-7247

JKSARM
Journal of Korean Society of Archives and Records Management

공식·비공식 사회적 기록의 융합을 통한 기업 부정 이슈 보도량 예측 연구

Predicting the Volume of News Coverage of Corporate Negative Issues through the Convergence of Formal and Informal Social Records

한능우(Nungwoo Han)[1], 김태리(Tailee Kim)[2], 박준혁(Junhyeok Park)[3]

E-mail: nwhan2@gmail.com, taely00@naver.com, popo2693@gmail.com

[1] 제1저자 스피치로그 대표
[2] 스피치로그 선임연구원
[3] 스피치로그 책임연구원

논문접수 2026-07-13
최초심사 2026-07-25
게재확정 2026-08-18

ORCID
Nungwoo Han
https://orcid.org/0000-0002-0183-5312
Tailee Kim
https://orcid.org/0009-0004-6485-3341
Junhyeok Park
https://orcid.org/0009-0005-3314-4582

▪본 연구는 중소벤처기업부의 2024년도 창업성장기술개발사업(팁스프로그램, No. RS-2024-00508804)의 지원에 의한 연구임.

https://jksarm.koar.kr

초 록

본 연구의 목적은 공식 기록인 뉴스와 비공식 기록인 트위터(현 X)를 '사회적 기록'으로 재정의하고, 두 기록을 융합하여 이미 보도가 시작된 기업 부정 이슈의 향후 보도량을 예측함으로써 기록으로 미래를 내다볼 수 있는지 검증하는 데 있다. 2024년 5월부터 2025년 10월까지 118개 기업의 뉴스 201,811건과 트위터 게시물 69,490건을 분석하였다. 부정 이슈를 14개 카테고리로 분류하고, 대규모 언어 모델로 책임 주체·심각도 등 다섯 가지 상황맥락 정보를 추출하였다. 뉴스만 사용한 공식기록 모델과 트위터까지 결합한 융합 모델을 비교하고, 트위터의 추가 기여는 기업 블록 부트스트랩으로 검정하였다. 분석 결과 뉴스만으로도 단순 기준선보다 오차를 24.8% 줄이며 향후 보도량을 유의하게 예측하였고, 트위터 결합은 예측을 통계적으로 유의하게 더 개선하였다(sMAPE 기준 상대 약 1.4%). 그 개선은 운영·고객 불만처럼 뉴스가 잘 다루지 않는 영역에 집중되었으며, 트위터는 공식 기록의 공백을 메우는 비공식 기록으로 기능하였다. 본 연구는 기록학의 '기록의 침묵' 개념을 뉴스의 보도 공백으로 조작적으로 정의하고, 그 공백이 예측 자원이 될 수 있음을 보였다. 다만 분석은 보도가 활발한 기업에 한정되었다. 향후 대상을 넓히면 이러한 접근은 기록이 부족한 영역을 찾아 수집·보존의 우선순위 설정에 활용될 수 있다.

ABSTRACT

This study aims to redefine news—as an official record—and Twitter (now X)—as an informal record—under the shared category of "social record" and verify whether the future can be predicted through records by fusing the two sources to forecast the volume of future news coverage of corporate negative issues already under active reporting. A corpus of 201,811 news articles and 69,490 tweets about 118 companies, spanning May 2024 to October 2025, was analyzed. Negative issues were classified into 14 risk categories, and five contextual features—including responsible party and severity—were extracted using a large language model. An official records model utilizing news data alone was compared against a fusion model incorporating Twitter data, with Twitter's incremental contribution assessed using a company-level block bootstrap test. News alone reduced prediction error by 24.8% over a simple baseline, and adding Twitter yielded a further significant gain of about 1.4% in relative sMAPE, concentrated in areas rarely covered by the news, such as operational issues and customer complaints, where Twitter filled the gaps of the formal record. Defining "archival silence" as observable gaps in press coverage, the study shows that such gaps can serve as a predictive resource. Although limited to companies with active news coverage, this approach could, with broader samples, help identify poorly documented areas and inform priorities for social documentation and digital preservation.

Keywords: 사회적 기록, 계산기록학, 기록의 침묵, 이슈 예측, 대규모 언어 모델
Social records, Computational archival science, Archival silence, Issue prediction, Large language models

1. 서 론

기록은 개인과 조직이 활동하는 과정에서 남긴 증거로서, 과거에 무슨 일이 있었는지를 증명하고 설명책임을 뒷받침해 왔다. 기록학 역시 오랫동안 기록의 수집과 보존, 관리를 중심으로 발전해 왔으며, 최근에는 기관이 보존해 온 공공기록을 정보자산으로 보아 적극적으로 활용해야 한다는 논의가 이어지고 있다(김명훈, 2021). 한편 디지털 사회에서는 제도가 관리하는 기록의 범위 밖에서도 방대한 기록이 생산되고 있다. 언론이 매일 생산하는 뉴스와 이용자가 자발적으로 남기는 소셜미디어 게시물은 당대 사회가 무엇에 주목하고 무엇을 논쟁하였는지를 증언하는 사회적 기록이라 할 수 있다.

기업의 부정 이슈는 이러한 사회적 기록이 집중적으로 생산되는 대표적인 영역이다. 제품 결함이나 안전 사고, 경영진의 일탈 같은 사안은 기업의 평판과 시장 가치에 큰 영향을 미치며, 보도가 시작되면 짧은 기간에 관련 기사가 급증하곤 한다. 문제는 이러한 전개를 미리 가늠하기 어렵다는 데 있다. 뉴스가 모든 부정 이슈를 같은 비중으로 기록하는 것도 아니다. 뉴스는 취재와 편집이라는 선별 절차를 거치므로 공적 중요성이 큰 사안은 즉시 기사화되는 반면, 개별 소비자의 불만처럼 작고 사적인 사안은 보도되지 않은 채 남기 쉽다. 공식 기록이 담아내지 못하는 영역이 존재하는 것이다. 본 연구가 민간 기업의 부정 이슈를 대상으로 삼은 까닭이 바로 이 영역에 있다. 민간 기업의 활동은 공공기록물 관리 체계의 밖에 있어, 기업의 부정 이슈가 사회에 남기는 공개된 기록은 사실상 뉴스와 소셜미디어 같은 사회적 기록뿐이다. 여러 기관에 걸쳐 특정 주제나 영역을 포괄적으로 기록화해야 한다는 논의에 비추어 보면, 제도가 관리하지 않는 이 영역이야말로 사회적 기록의 가치가 가장 잘 드러나는 자리라 할 수 있다. 더구나 기업의 부정 이슈는 대형 참사부터 한 소비자의 불만까지 공적 중요성의 폭이 넓어, 무엇이 기록되고 무엇이 남지 않는지를 살피기에 알맞은 영역이다. 이러한 영역에서 확인되는 결과는 기록이 부족한 영역을 미리 찾아 수집과 보존의 우선순위를 정하는 논의에 참고가 될 수 있다.

그동안 기록학에서도 소셜미디어와 웹을 기록관리의 대상으로 검토해 왔으나, 관련 연구는 주로 기록의 수집과 보존, 관리 체계에 초점을 두어 왔다(송주형, 2014; 최두원 외, 2018). 다른 한편 대규모 디지털 데이터와 계산적 방법으로 사회 현상을 분석하는 계산사회과학(Lazer et al., 2009)을 비롯한 계산 분야에서는 뉴스와 소셜미디어를 활용해 사건과 언급량을 예측하는 연구가 축적되어 왔으나, 이들 연구에서 두 매체는 예측 모델에 투입되는 데이터일 뿐 서로 다른 절차를 거쳐 생산된 기록으로 다루어지지는 않았다. 요컨대 사회적 기록을 보존의 대상을 넘어 미래 예측의 자원으로 활용하면서, 성격이 다른 두 기록의 차이를 기록학의 관점에서 해석한 연구는 아직 시도되지 않았다.

이러한 문제의식에서 본 연구의 목적은 공식 기록인 뉴스와 비공식 기록인 트위터(현 X)[1]를 융합하여 기업 부정 이슈의 미래 보도량을 예측하고, 그 과정에서 두 기록이 수행하는 역할의 차이를 규명하는 것이다. 이는 과거를 증명해 온 기록을 미래를 내다보는 자원으로 확장하려는 시도이다. 이를 위해 본 연구는 부정 이슈 보도가 꾸준히 관측되는 118개 기업을 대상으로 뉴스와 트위터 게시물을 14개 리스크 카테고리로 나누어 분석하고, 뉴스만 사용한 모델과 트위터까지 결합한 모델의 예측력을 통계적으로 비교한다. 구체적으로는 먼저 공식 기록인 뉴스만으로 부정 이슈의 미래 보도량을 예측할 수 있는지를 검증하고, 다음으로 비공식 기록인 트위터를 결합하면 예측이 어떻게 달라지며, 트위터의 기여(뉴스가 기록하지 않는 영역의 보완과 뉴스와 함께 움직이는 동시 반응)가 영역에 따라 어떤 상이한 양상을 보이는지 살펴본다. 이 결과로부터 사회적 기록의 미래지향적 활용에 관하여 기록관리학계와 실무 현장이 고려할 수 있는 시사점을 도출하고자 한다.

1) 트위터는 2023년 7월 X로 명칭이 변경되었으나, 본 연구는 분석 자료의 통칭과 선행연구의 표기에 따라 '트위터'로 쓴다.

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

2. 연구 배경

2.1 사회적 기록으로서의 뉴스와 트위터: 공식성과 비공식성

본 연구는 뉴스와 트위터를 당대의 사회적 담론을 반영하는 사회적 기록으로 규정하고, 전자를 공식 기록으로, 후자를 비공식 기록으로 부른다.[2] 여기서 사회적 기록이란 공공기관의 업무 과정이 아니라 사회에 공개적으로 표출된 발화와 반응의 기록으로, 당대 사회가 무엇에 주목하고 무엇을 논쟁하였는지를 증언하는 자료를 뜻한다. 언론이 생산하는 뉴스와 이용자가 남기는 소셜미디어 게시물, 온라인 커뮤니티의 글이나 청원 기록이 여기에 속한다. 반면 공개되지 않는 사적 자료나 기관 내부의 업무 기록은 여기에 들지 않는다.

본 연구는 그 가운데 생산 절차의 성격이 가장 뚜렷하게 대비되는 뉴스와 트위터를 분석 대상으로 삼는다. 물론 전통적인 기록학의 관점에서 뉴스는 기록이 아니라 간행물로 분류되어 왔다. 기록은 업무 행위의 부산물로서 그 행위를 증언하는 증거성을 요건으로 하는 반면, 뉴스는 처음부터 배포를 목적으로 생산되기 때문이다. 그러나 무엇을 기록으로 남길 것인가의 기준을 기관의 업무가 아니라 당대 사회의 여론과 가치에서 찾아야 한다는 사회적 평가론(Booms, 1987)과 여러 기관에 걸쳐 특정 주제나 영역을 포괄적으로 기록화하자는 도큐멘테이션 전략(Samuels, 1986) 이래, 기록화의 시야는 조직 내부의 증거를 넘어 사회가 남기는 흔적 전반으로 확장되어 왔다. 본 연구가 뉴스와 트위터를 사회적 기록으로 규정하는 것은 이 평가론의 계보 위에 서 있다.

기록학에서 공식 기록은 본래 제도적으로 정의되는 개념이다. 「공공기록물 관리에 관한 법률」 제3조는 공공기관이 업무와 관련하여 생산하거나 접수한 자료를 기록물로 정의한다. 이 정의에 따르면 뉴스는 공식 기록에 해당하지 않는다. 본 연구의 '공식'은 이처럼 공적 권한이 부여하는 지위가 아니라, 생산 절차의 성격을 가리키는 연구 목적상의 조작적 정의다. 곧 이 명칭은 뉴스에 기록학 체계상 공식 기록의 지위를 부여하려는 것이 아니라, 생산 절차가 대비되는 두 기록을 가르기 위해 본 연구 안에서 한정적으로 쓰는 용어다. 제도화된 절차의 유무로 공식과 비공식을 가르는 것은 사회과학의 일반적 용법이다. 예컨대 제도경제학은 사회의 규칙 가운데 법률처럼 명문화된 것을 공식 제도로, 관습이나 규범처럼 명문화되지 않은 것을 비공식 제약으로 구분한다(North, 1990). 본 연구는 이 구분을 기록의 내용이 아니라 생산 절차에 적용하여, 생산 절차에 제도적 통제와 책임 소재가 존재하는 정도가 높은 기록을 공식 기록으로, 낮은 기록을 비공식 기록으로 부른다. 이는 어떤 정보가 신뢰할 수 있는 기록으로 인정받기 위해서는 엄격한 생산 절차와 통제 속에서 만들어져야 한다는 기록학의 오랜 전제를 따른 것이다(Duranti, 1995). 실제로 뉴스는 「신문 등의 진흥에 관한 법률」과 「방송법」에 따라 등록되거나 허가받은 언론사가 정해진 취재·편집 절차를 거쳐 생산하는 기록으로, 생산의 주체와 절차, 책임 소재가 모두 제도 안에 놓여 있다.

이러한 제도적 통제는 정보의 신뢰성을 뒷받침하는 기제로 작용한다. 나아가 언론사라는 법인과 편집 책임자가 보도에 대한 책임을 부담하는 만큼, 뉴스는 공표를 통해 특정 사안을 사회적 논의의 장에 등록하는 제도적 기록의 성격을 띤다. 물론 이러한 절차가 개별 보도 내용의 진위까지 보증하는 것은 아니며, 오보와 정정 보도의 존재는 뉴스 역시 불완전한 기록임을 보여 준다. 그러나 오보에 대해 정정 보도와 언론중재라는 사후 교정 절차가 제도적으로 작동한다는 사실은, 오히려 뉴스가 통제된 절차 안에서 생산되고 수정되는 기록이라는 뜻이기도 하다. 반면 트위터는 사전 승인이나 검증 절차 없이 개별 이용자가 자발적이고 무선별적으로 생산하는, 전통적 기록관리의 통제 범위를 벗어난 매체다. 그러나 역설적으로 바로 그 제도적 통제의 부재 때문에 트위터는 제도화된 뉴스가 포착하지 못하는 대중의 목소리를 담아낼 수 있고, 그런 점에서 공식 기록이 비워 둔 '보도 공백'을 메우는 비공식 기록으로서의 가치를 지닌다.

2) 공식 기록과 비공식 기록은 각각 formal records와 informal records의 역어다. 여기서 formal은 공적 권한이 부여하는 지위를 뜻하는 official과 구별되며, 생산 절차가 제도화된 통제와 격식을 갖추었다는 의미로 쓴다.

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

다만 공식성과 비공식성은 상호 배타적인 두 범주가 아니라, 정도의 차이를 지닌 연속선상의 개념으로 보는 것이 타당하다. 엄격한 절차를 거치는 뉴스가 공식 기록의 한쪽 끝에 있다면, 자유롭게 생산되는 트위터는 그 반대편 비공식 기록의 끝에 자리한다. 기존의 사회적 기록 연구는 이러한 기록을 수집하고 보존하는 데 머물러 왔다. 본 연구는 성격이 뚜렷이 다른 두 기록을 함께 활용하여 미래의 보도량을 예측하고, 사건이 앞으로 어떻게 전개될지까지 가늠한다.

2.2 기록의 침묵과 비공식 기록

모든 기록은 무엇을 남기고 무엇을 남기지 않을지를 선택하는 과정을 거치며, 그 과정에서 일부 현실은 기록되지 못한 채 배제되기 마련이다. Trouillot(2015)은 역사가 만들어지는 과정에는 사료 생성, 기록보관소 형성, 서사 구성, 역사로의 공인이라는 네 계기가 있다고 보았다. 그리고 이 네 계기마다 권력이 개입하여 특정 목소리를 지우는 침묵화가 일어난다고 지적하였다. Harris(2002) 역시 그렇게 살아남은 기록은 실재했던 과정 전체의 '조각의 조각의 조각'에 불과하며, 그마저도 권력관계 속에서 구성된 것이라고 보았다. 이처럼 '기록의 침묵'[3]에 관한 논의는 무엇이 왜 기록에서 배제되었는가를 과거형으로 성찰하는 데 집중되어 왔다.

본 연구가 말하는 '침묵의 보완'은 이 논의에서 출발한다. 여기서 침묵의 보완이란 어느 한 기록이 남기지 않은 영역을 생산 절차가 다른 기록이 대신 남기는 현상을 뜻하며, 본 연구에서는 뉴스가 보도하지 않은 사안을 트위터가 기록하는 양상을 가리킨다. 기록은 저절로 남지 않고 무엇을 남길지 선별하는 과정을 거치며, 그 선별에는 권력이 개입한다. 아카이브는 완전하지도, 중립적이지도, 객관적이지도 않으며, 절차와 업무 과정, 사회 환경에 스며든 편향이 그대로 새겨진다는 지적(Van Bussel, 2017)도 같은 맥락이다. 뉴스는 이러한 선별의 필터를 통과한 기록이다. 다만 그 필터를 논하는 층위는 구분할 필요가 있다. Trouillot과 Harris의 논의가 기록 일반에서 침묵이 발생하는 까닭을 권력의 개입으로 설명하는 이론적 층위에 있다면, 뉴스에서 그 침묵을 실제로 만들어 내는 기제는 무엇을 보도할 만한 일로 볼 것인가를 판단하는 게이트키핑이다(White, 1950). 곧 본 연구가 관측하는 보도 공백은 Trouillot이 말한 네 계기 가운데 사료가 만들어지는 첫 계기에서, 게이트키핑을 통해 발생하는 침묵에 해당한다. 반면 트위터는 이러한 제도적 필터를 거치지 않고 당사자가 직접 작성하는 기록이다. 그렇다고 트위터에 침묵이 없는 것은 아니다. 플랫폼 알고리즘, 이용자 구성의 편향, 확산 구조라는 고유한 필터가 있고, 그에 따른 트위터만의 침묵이 있다. 따라서 본 연구의 주장은 트위터가 완전한 기록이라는 것이 아니다. 두 기록이 서로 다른 필터를 거치므로, 함께 활용하면 어느 한쪽의 침묵이 부분적으로 상쇄된다는 것이다. 이 상쇄 가설이 성립하는지, 곧 뉴스만으로는 포착되지 않는 영역을 트위터가 채우는지는 4장에서 검증한다.

본 연구는 보도 공백을 바라보는 시선을 과거에서 미래로 옮긴다. 먼저 뉴스가 어떤 영역에서 침묵하고, 비공식 기록인 트위터가 그 자리를 얼마나 채우는지를 카테고리별로 측정한다. 측정에서 멈추지 않고 트위터가 채운 정보를 미래 보도량 예측에 활용한다. 곧 기록의 침묵을 미래 예측의 이론적 토대로 삼으려는 시도다.

기록에서 배제된 목소리를 둘러싼 비판적 논의는 최근 국내 기록학계에서도 이어지고 있다. 윤지수(2026)는 아카이브가 중립적 저장소가 아니라 권력이 작동하는 공간임을 지적하고 탈식민적 실천 사례를 분석하여 그동안 기록에서 소외되어 온 집단의 기억을 되살리는 방안을 모색하였다. 본 연구도 공식 기록이 특정 영역을 비워 둔다는 문제의식을 공유한다. 다만 초점이 다르다. 기존 논의가 과거에 무엇이 비워졌는가를 비판하는 데 집중해 왔다면, 본 연구는 그 비워진 영역을 비공식 기록으로 메워 미래를 예측한다.

3) '기록의 침묵'은 기록이 특정 영역을 배제해 남기지 못하는 현상을 가리키는 기록학의 정립된 개념(archival silence)으로, 사료 생산에 권력이 개입한다는 논의(Trouillot, 2015)와 살아남은 기록의 부분성에 관한 논의(Harris, 2002)에 이론적 기반을 둔다.

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

2.3 사회적 기록 기반 예측 연구 동향

사회적 기록을 활용한 예측 연구는 크게 세 흐름으로 정리된다.

첫 번째는 온라인상의 반응 규모를 모델링하는 연구다. Ng et al.(2022)는 뉴스·무력 충돌 기록 같은 외생 신호와 플랫폼 내부의 내생 데이터를 함께 학습하는 심층학습 모델을 제안하여, 전통적 시계열 기법보다 정확하게 주제별 일별 활동량을 예측하였다. 이 연구는 성격이 다른 신호를 구분하고 결합하는 것이 온라인 반응 예측에 중요함을 보여 주며, 이는 뉴스와 트위터라는 이종 기록을 결합하는 본 연구의 설계와 문제의식을 공유한다.

두 번째는 소셜미디어가 기존의 확인 채널보다 사건을 먼저 감지함을 보여 준 연구들이다. Parekh et al.(2024)는 신종 감염병 관련 게시물을 분석하여 WHO 선언보다 4~9주 앞서 유행의 징후를 포착하였고, Burns et al.(2025)는 트위터의 다국어 게시물에 토픽 모델링을 적용하여 새롭게 부상하는 지정학 이슈를 Google Trends보다 먼저 식별하였는데, 선행 폭은 사안에 따라 수 시간에서 이틀 남짓에 이르렀다. Janzen et al.(2024)는 언어모델을 결합한 신호 탐지 기법으로 에너지 분야의 위기가 발생하기 수주 전부터 트윗 빈도가 지수적으로 늘어나는 양상을 포착하였으며, Mansoor와 Ansari(2024)는 세 플랫폼에서 수집한 약 99.6만 건의 게시물로 학습한 딥러닝 모델로 정신건강 위기를 전문가 판단보다 평균 7.2일 앞서 감지하였다. 이처럼 이들 연구는 기관의 선언, 검색 통계, 전문가 판단 같은 기존 채널이 아직 침묵하는 동안 소셜미디어에는 이미 신호가 쌓이고 있음을 일관되게 보여 준다. 이는 뉴스가 다루지 않는 영역을 트위터가 기록한다는 본 연구의 '침묵 보완' 논의와 맞닿아 있다.

세 번째는 기업의 부정 이슈를 직접 다룬 연구다. Nicolas et al.(2024)는 S&P100 기업에 관한 트윗 1.14억 건을 분석하여, ESG 평판 리스크 게시물이 이례적으로 급증한 시점을 '사건'으로 정의하고 그 전후의 주가 반응을 살피는 사건연구를 수행하였다. 그 결과 게시물 급증은 통계적으로 유의한 주가 하락으로 이어졌다. 이 연구는 소셜미디어의 게시 급증이 기업 리스크에 관한 실질적 정보가치를 지님을 보였다. 트위터의 언급 급증이 부정 이슈의 신호가 될 수 있다는 본 연구의 문제의식과 직접 이어지는 결과다.

이 가운데 조기 경보에 해당하는 탐지·예측 연구들이 사용한 자료와 방법을 정리하면 <표 1>과 같다. 심층학습 시계열 모델, 언어모델 기반 분류, 토픽 모델링 등 다양한 방법이 적용되어 왔다.

〈표 1〉 사회적 기록 기반 예측·탐지 선행연구

연구	분석 대상	데이터(규모)	방법·모델	연구 결과
① 단일 소스				
Parekh et al.(2024)	감염병 유행	트위터 3억 3,100만 건	BERT-QA 모델	WHO 공식 선언보다 4~9주 선행
Janzen et al.(2024)	에너지 위기 (가격 급등·수급 불안)	트위터 46,963건	LLM 기반 하이브리드 분류 모델	위기 수주 전 징후 포착
Burns et al.(2025)	신흥 지정학 이슈	트위터 345,759건	동적 토픽 모델링(LDA)	Google Trends보다 수 시간~약 이틀 선행
② 멀티 소스 결합				
Ng et al.(2022)	베네수엘라 정치 위기	트위터·유튜브·뉴스	심층학습 시계열 (LSTM)	이종 자료 결합 시 단일 자료 모델보다 정확
Mansoor와 Ansari(2024)	정신건강 위기	트위터·레딧·페이스북 996,452건	멀티모달 딥러닝 (BERT·LSTM)	전문가 판단보다 평균 7.2일 선행

<표 1>이 보여 주듯, 선행연구는 '얼마나 먼저 포착하는가'라는 선행성을 입증하는 데 집중해 왔다. 다섯 연구 가운데 셋은 하나의 소스만을 사용하였고, 둘은 여러 소스를 결합하였다. 이 가운데 Ng et al.(2022)는 이종 자료를 결합하면 단일 자료 모델보다 예측이 정확해짐을 확인하였다. 즉 사회적 기록으로 미래를 내다볼 수 있다는 것

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

자체는 이미 충분히 입증되었다. 그러나 이들 연구는 뉴스와 소셜미디어를 모델에 투입하는 데이터로만 다루었고, 두 자료가 어떻게 만들어진 기록인지는 묻지 않았다. 기록의 관점에서 보면 뉴스는 취재와 편집이라는 필터를 거친 공식 기록이고, 트위터는 그러한 필터 없이 생산되는 비공식 기록이다. 생산 방식이 다르면 담는 내용이 다르고, 예측에서 맡는 역할도 달라질 수 있다.

본 연구는 이 지점에서 출발한다. 두 기록을 결합해 기업의 부정 이슈를 다루면서, 예측 대상을 공식 기록의 생산량인 보도량으로 삼고, 결합에 따른 예측 개선을 두 기록의 성격 차이로 해석한 연구는 찾아보기 어렵다. 이에 본 연구는 뉴스와 트위터를 결합해 기업 부정 이슈의 미래 보도량을 예측하고, 트위터가 보태는 기여를 통계적으로 검정한다. 이때 트위터에 기대하는 것은 뉴스보다 빠른 신호가 아니라, 뉴스가 담지 않는 영역을 채우는 다른 성격의 기록이다. 이 점에서 본 연구는 선행성을 겨루는 위 연구들과 문제 설정 자체가 다르다.

2.4 기록학 분야의 연구 동향과 본 연구의 위치

계산기록학은 계산적 이론·방법과 기록학적 이론·방법을 통합하는 초학제 분야로, 2016년 IEEE Big Data 학술대회의 계산기록학 워크숍을 계기로 형성되기 시작하였다(Marciano가 주도한 조직위원회 주관). Marciano et al.(2018)는 이를 대규모 기록·아카이브의 처리·분석·저장·장기보존·접근에 계산적 방법과 자원을 적용하는 분야로 정식화하면서, 단순한 기능의 전산화가 아니라 계산적 사고와 기록학적 사고의 양방향 융합으로 규정하였다. 이 창립 논문은 빅데이터 시대의 기록이 전통적인 방법만으로는 다루기 어려운 규모와 형태로 생산되고 있음을 지적하며, 기록학 교육에 계산적 사고를 이식해야 한다는 논리를 정립하였다. 계산기록학이라는 이름이 정립되기 전부터 기업의 기록은 이미 이러한 분석의 대상이 되어 왔다. 파산한 엔론이 남긴 수십만 건의 이메일 아카이브를 사회 연결망 기법으로 분석하여, 위기에 놓인 조직 내부의 소통 구조와 행태를 복원한 연구가 대표적이다(Diesner et al., 2005). 이후 국제적으로는 계산기록학을 표방한 학술지 특집호가 발간되고(Hedges et al., 2022) 정부 기록의 공개 심사에 기계학습을 적용하는 실증 연구(Baron et al., 2022)가 이어지는 등 분야가 제도화되고 있다.

반면 국내 기록학계에서 계산기록학이라는 문제틀을 명시적으로 내걸고 수행된 실증 연구는 확인하기 어렵다. 본 연구는 사회적 기록을 대상으로 대규모 언어 모델과 기계학습이라는 계산적 방법을 적용하고, 피처의 설계와 결과의 해석을 기록학의 문제틀(공식성과 침묵)로 수행하는 양방향 융합의 한 실증 사례다. 나아가 그 적용 범위를 기록의 미래예측적 활용으로 넓힌다.

기존 연구와의 거리는 두 갈래로 나뉜다. 첫째, 기록학 안에서 보면, 그동안의 사회적 기록 아카이빙은 특정 기관의 기록을 모으고 보존하는 데 초점을 두어 왔다. 송주형(2014)은 SNS를 기록관리의 대상으로 검토하였고, 최두원 외(2018)는 대통령 SNS 기록물의 수집·관리·서비스 방안을, 윤정옥(2010)은 국립중앙도서관 웹 아카이브 OASIS의 수집 현황과 개선 방안을 제시하였다. 이해영 외(2007)는 기록관들이 보존 중심으로 운영되어 기록의 활용과 서비스가 상대적으로 소홀했음을 사례연구로 확인하였고, 최근에는 기록을 정보자산으로 보아 적극 활용하자는 논의도 제기되고 있다(김명훈, 2021). 본 연구는 이 흐름의 연장선에 있되, 한 기관이 아니라 사회 전반에서 만들어진 기록을 다루고, 보존에 그치지 않고 미래를 내다보는 데 활용한다. 둘째, 계산적 예측 연구와 비교하면, 앞 절에서 살펴본 대로 이들은 뉴스와 소셜미디어를 예측에 넣는 데이터로만 다루어 왔다. 반면 본 연구는 같은 자료를 공식 기록과 비공식 기록이라는 서로 다른 성격의 기록으로 바라보고, 뉴스의 공백을 트위터가 어떤 영역에서 메우는지 카테고리별로 해석한다.

국내에서도 한승희와 이혜원(2025)이 공공기록물 관련 뉴스 기사를 텍스트마이닝으로 분석하여, 뉴스 데이터를 기록학 연구에 본격적으로 끌어들인 바 있다. 그러나 이 연구는 뉴스를 '공공기록물이라는 주제가 어떻게 보도되는가'를 보여 주는 자료로 다룬 기술적 분석이라는 점에서, 뉴스 자체를 하나의 사회적 기록으로 재정의하고

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

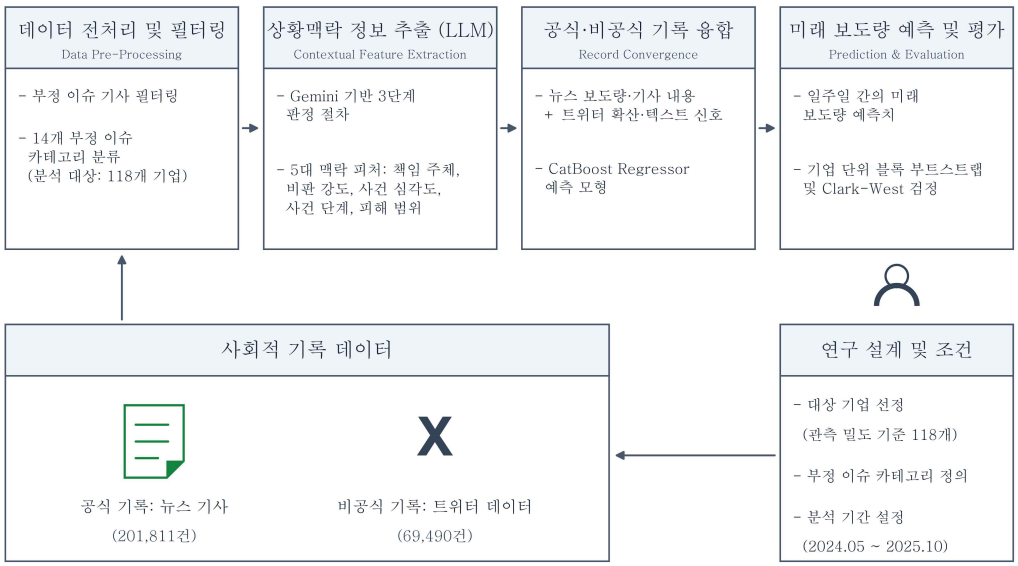
그 미래를 예측하는 본 연구와 구별된다. 이상에서 논한 본 연구의 차별성을 정리하면 <표 2>와 같다.

〈표 2〉 본 연구의 차별성

구분	사회적 기록 아카이빙	본 연구
기록을 보는 관점	수집 · 보존의 대상	공식 · 비공식 사회적 기록으로 재정의
범위	특정 기관의 기록	뉴스와 트위터
활용 방향	보존(과거지향)	보도량 예측과 보도 공백의 해석(미래지향)

3. 연구 방법

본 연구의 전체 분석 절차는 <그림 1>과 같다. 사회적 기록 데이터에서 출발하여 부정 이슈 기사를 필터링하고, LLM으로 상황맥락 정보를 추출한 뒤, 공식 · 비공식 기록을 융합하여 미래 보도량을 예측하고 평가하는 흐름이다.



〈그림 1〉 연구의 전체 분석 절차

3.1 데이터와 적법성

분석 데이터는 빅데이터 분석 전문기업 ㈜스피치로그로부터 제공받은 뉴스 · 소셜미디어 자료다. 뉴스 자료는 포털 다음(Daum) 뉴스에 게재된 제휴 언론사 기사를 사회, 경제, 정치, 국제, 문화 카테고리별로 수집한 것이다(2024년 5월 1일 ~ 2025년 10월 31일). 다음 뉴스는 국내 주요 포털의 하나로 종합일간지와 방송사, 경제지, 언론 전문지에 이르는 다양한 유형의 제휴 매체 기사가 유통되며, 본 연구의 수집분에서도 이들 유형이 모두 확인된다. 다만 포털에 제휴하지 않은 매체나 포털에 공급되지 않은 기사는 수집 범위 밖에 있으므로, 본 연구가 관측하는 보도 공백은 이 수집망 안에서 확인된 공백이며 언론 보도 전체의 부재와 곧바로 동일시할 수는 없다. 수집 단계의 기사에는 사건 보도 외에 홍보성 기사와 실적 발표, 제휴 소식, 시황 전망 등 부정 이슈 분석에 부적합한 유형이 섞여 있으므로, 기업명 매칭으로 분석 대상 기업의 기사를 추린 뒤 대규모 언어 모델을 활용한

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

다단계 필터링으로 부정 사건이 기사의 실제 주제인 경우만 선별하고 이들 유형은 명시적으로 제외하였다(판정 기준 요지는 <부록 표 4>, 단계별 절차는 3.4절 참조). 분석 기간은 두 매체가 모두 관측된 2024년 5월부터 2025년 10월까지로 통일하였으며, 이 기간에서 뉴스는 데이터 필터링과 정보 추출을 거친 기사 201,811건, 트위터는 기업 매칭과 필터링을 거친 게시물 69,490건이 분석에 사용되었다. 본 연구는 제공받은 자료의 2차 이용자로서 보도량・날짜・카테고리・트위터 비중 등 메타데이터를 분석 대상으로 하며, 기사・게시물 원문은 재배포하지 않는다. 트위터 데이터의 개인 식별정보는 비식별 처리하였고, 연구 과정에서 개인을 식별하거나 이용자와 직접 상호작용하지 않았다.

분석 대상 기업의 선정 기준은 다음과 같다. 본 연구가 제공받은 자료는 2,848개 기업을 대상으로 수집된 것으로, 상장 여부와 같은 기준으로 구성된 목록이 아니라 언론과 소셜미디어에서 언급이 관측되는 조직을 여러 산업에 걸쳐 포괄한다. 선정 지표는 보도 관측 밀도, 곧 2024년 5월부터 2025년 10월까지의 관측 기간 동안 해당 기업의 부정 이슈 보도가 하루 1건 이상 발생한 날의 비율이다. 보도가 없는 날이 과반인 기업은 예측 대상 시계열의 대부분이 0으로 채워져 학습할 신호 자체가 부족해지므로, 관측 기간의 절반 이상에서 보도가 관측된 기업, 곧 관측 밀도 50% 이상인 118개 기업(전체 2,848개 중)을 분석 대상으로 선정하였다. 이 임계값은 기준을 높여 보도가 많은 소수 기업만 남기는 것을 피하고, 가능한 한 많은 기업을 분석에 포함하기 위해 낮게 잡은 것이다.

다만 이 선정 기준에는 표본 선택의 편향이 따른다는 점을 미리 밝혀 둔다. 관측 밀도를 기준으로 삼으면 보도가 드문 기업, 곧 언론이 가장 적게 다루는 기업이 구조적으로 표본에서 제외되기 때문이다. 따라서 본 연구가 관측하는 보도 공백은 보도가 활발한 기업 안에서 특정 범주가 비어 있는 공백이며, 기업 자체가 거의 보도되지 않는 더 깊은 공백은 이 설계로 관측되지 않는다. 이는 예측 모형이 성립하기 위한 최소한의 시계열 신호를 확보해야 한다는 제약에서 비롯된 것으로, 본 연구의 결과를 해석할 때 함께 고려해야 한다.

3.2 부정 이슈와 카테고리 정의

부정 이슈는 6개 대분류 아래 14개 소분류로 정의하였다(<표 3> 참조). 이 분류는 본 연구를 위해 새로 고안한 것이 아니라, 자료 제공사가 기업 리스크 모니터링 업무에서 사용해 온 운영 분류를 본 연구의 목적에 맞게 조정한 것이다. 조정은 빈도가 낮은 유형을 인접 범주로 통합하고, 성격이 다른 사안이 섞인 범주를 분리하는 방식으로 이루어졌다. 실무에서 축적된 분류를 채택한 까닭은 실제 부정 이슈 보도가 다루어지는 유형과의 정합성을 확보하는 데 있다. 다만 운영 분류는 이론적 유형론에서 연역된 것이 아니므로, 범주 구분 자체의 이론적 타당성은 본 연구가 검증한 바가 아니다. 본 연구는 그 대신 판정의 일관성을 확보하는 데 초점을 두어, 각 소분류의 판정 기준을 명문화하여 전체 기사에 일관되게 적용하고(<부록 표 4> 참조), 사람의 코딩과 대조하여 분류 신뢰도를 확인하였다(3.4절). 소분류는 기업 이슈의 다차원성을 반영한 기본 분석 단위다. 인접한 소분류 사이에는 판정이 중첩될 수 있어, 분류 신뢰도는 대분류 기준으로도 함께 보고하였다.

〈표 3〉 부정 이슈 카테고리 체계(6 대분류 / 14 소분류)

대분류	소분류
규제・법무	법적규제, 횡령배임
경영・재무	경영악화
노동・인권	노무노동, 갑질성비위, 차별인권
안전・품질	사고안전, 화재재해, 환경오염, 품질문제
정보・보안	개인정보보안, 허위조작
갈등・고객	갈등분쟁, 운영고객

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

3.3 예측 과제 정의

본 연구가 예측하려는 것은 특정 기업의 부정 이슈가 '앞으로 얼마나 보도될 것인가'이다. 구체적으로는 기업, 리스크 카테고리, 날짜의 조합마다 이후 일주일 동안의 하루 평균 보도 건수를 예측하였다. 다만 아직 보도가 시작되지 않은 사안까지 내다보기는 어려우므로, 예측은 당일에 한 건 이상 보도가 이루어진 사안으로 한정하였다. 즉 어떤 사안의 보도가 시작된 시점에서, 그 사안이 다음 일주일 동안 어느 규모로 전개될지를 가늠하는 과제다.

3.4 모델 피처 설계

예측에는 보도량뿐 아니라 기사가 사건을 어떻게 다루는가라는 내용의 신호도 필요하다. 이에 본 연구는 언론학의 프레임 이론과 위기 · 이슈 커뮤니케이션 연구(Coombs, 2007; Downs, 1972; Entman, 1993; Iyengar, 1991; Semetko & Valkenburg, 2000)를 참고하여, 기사가 사건을 다루는 방식을 책임 주체, 보도 비판 강도, 사건 심각도, 사건 단계, 피해 범위의 다섯 항목으로 조작화하고 각 기사에서 추출하였다. 모형에는 이렇게 추출한 내용 신호와 함께, 최근 보도량의 흐름을 담은 피처(이동평균과 그 변화)를 투입하였다.

분석 대상 기사가 약 20만 건에 이르러 사람이 전부 판정하기는 어렵다. 이에 연구자가 판정 기준을 정하고, 대규모 언어 모델(LLM)이 그 기준을 전체 기사에 일관되게 적용해 라벨을 부여하는 '확장된 코더' 방식을 취하였다. 이렇게 부여한 라벨로 예측 모델을 학습시키는 방식을 기계학습에서는 약한 지도학습이라 부른다.

LLM의 판정은 3단계로 수행하였다. 1차 데이터 필터링에서는 부정 이슈 기사만 골라내고, 2차 정보 추출에서는 사건 심각도와 책임 주체, 피해 범위를 측정하였다. 마지막 3차 오류 제거에서는 주변부 기사를 걸러 내면서 사건 단계와 보도 비판 강도를 측정하였다. 각 단계의 판정 난이도에 맞추어 모델을 달리 적용하였는데, 부정 이슈 여부만 가리는 단순한 1차 필터링에는 가벼운 모델을, 책임 주체 · 피해 범위 · 사건 단계처럼 맥락을 세분해 판단해야 하는 2 · 3차 추출에는 성능이 더 높은 모델을 사용하였다(사용 모델은 <부록 B>).

트위터 게시물에도 같은 방식을 적용하였다. 기업명과 별칭을 정리한 사전으로 게시물을 118개 기업에 매칭하고, 분석 기간 밖의 게시물과 정치 · 주식 시황 등 무관한 게시물을 규칙으로 걸러 낸 뒤, LLM이 기업 당사자 여부와 부정 이슈 여부를 판정하고 14개 카테고리로 분류하였다. 4장의 카테고리별 트위터 비중과 교차상관은 이 분류를 기준으로 산출한 것이다.

이러한 LLM 판정이 믿을 만한지는 사람의 판정과 대조하여 검증하였다. 층화 무작위로 50건을 추출하여 LLM 라벨을 보지 않은 상태에서 사람이 독립적으로 코딩하였고, 리스크 카테고리 분류의 일치도는 소분류 기준 Cohen's κ=0.51(95% 신뢰구간 0.35~0.67), 대분류 기준 κ=0.64(0.45~0.80)였다. 이 수준은 뉴스 내용 분류를 다룬 선행 연구와 견주어 해석해야 한다. Burscher et al.(2014)는 훈련된 코더 30명이 이중 코딩한 기사에서 프레임별 Krippendorff's α가 0.21에서 0.58에 그쳤다고 보고하며 신뢰도가 충분하지 않음을 스스로 인정하였고, Arora et al.(2025) 역시 연구자 두 명이 수행한 프레임 코딩에서 α=0.39를 보고하였다. α는 우연 일치를 보정한 지표로 두 코더의 명목 자료에서는 κ와 유사한 범위의 값을 산출하므로, 본 연구의 0.51과 0.64는 훈련된 인간 코더 사이의 일치도에 견주어 낮지 않은 수준이다. 다만 위 연구들이 다룬 프레임 판정과 본 연구의 카테고리 분류는 과제의 성격이 달라 직접 비교에는 한계가 있다. 아울러 사건 심각도와 같은 강도 판정은 사람 사이에서도 일관된 기준을 세우기 어려워 대조 대상에서 제외하였다. 이 점은 사람 코딩을 한 명의 코더가 수행하였다는 점과 함께 본 검증의 한계로 남는다.

트위터에서는 두 갈래의 피처를 추출하였다. 하나는 트윗량과 리트윗 · 좋아요처럼 반응이 얼마나 컸는지를 담는 확산 피처이고, 다른 하나는 트윗에 어떤 말이 쓰였는지를 읽는 텍스트 피처다. 텍스트 피처는 14개 리스크

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

카테고리를 가르는 위험 어휘와 루머・집단행동・감정・부인 표현을 사전으로 정리한 뒤, 그러한 표현이 최근 얼마나 늘었는지를 기업과 카테고리 단위로 측정하는 방식이다. 특히 이 계산은 뉴스 보도가 없는 날까지 포함한 전체 기간 위에서 이루어져, 공식 기록이 침묵하는 동안 트위터에 쌓인 말들도 신호로 반영된다. 이렇게 붙인 라벨 자체도 특정 모델이 특정 기준으로 생산한 기록이므로, 어떤 모델이 어떤 기준으로 판정하였는지를 <부록 B>에 연구 기록으로 남긴다.

3.5 예측 모형의 구성

본 연구는 세 수준의 모델을 비교한다(<표 4> 참조). 모델 명칭은 연구의 관점(공식・비공식 기록의 융합)을 반영하여 부여하였다.

〈표 4〉 비교 모델 구성

모델	사용한 자료
나이브 베이스라인(naive baseline)	지난 7일간의 평균 보도량
공식기록 모델	뉴스 보도량 + 기사 내용 피처
사회적 기록 융합 모델(이하 '융합 모델')	공식기록 모델 + 트위터 신호(확산량・위험 관련 단어)

나이브 베이스라인은 예측 시점 직전 7일의 평균 보도량을 그대로 예측값으로 사용하는 모형으로, 다음과 같이 정의된다.

$$\hat{y}_{i,c,t} = \frac{1}{7}\sum_{k=1}^{7} y_{i,c,t-k}$$

여기서 $y_{i,c,t-k}$는 기업 i, 카테고리 c의 t−k일 보도 건수이며, $\hat{y}_{i,c,t}$는 t일 이후 일주일 동안의 하루 평균 보도량에 대한 예측값이다.

예측 모형은 하나를 미리 정해 두지 않고, 여러 기계학습 기법을 같은 조건에서 비교한 뒤 가장 정확한 것을 선택하였다. 비교한 후보는 Ridge, ElasticNet, Random Forest, Extra Trees, HistGradientBoosting, XGBoost, LightGBM, CatBoost의 여덟 가지로, 모두 여러 변수를 입력받아 수치를 예측하는 기계학습 회귀 기법이다. 이들을 검증 구간에서 비교하여 검증 성능이 가장 우수한 CatBoost를 최종 모형으로 채택하고, 그 성능은 시험 구간에서 평가하였다(알고리즘별 검증・시험 성능은 <부록 A> 참조). 이때 비교 기준으로는 기업 간 보도량 규모 차이에 덜 민감한 sMAPE를 사용하여, 보도가 많은 소수 기업이 모형 선택을 좌우하지 않도록 하였다. CatBoost는 다수의 작은 결정트리를 순차적으로 보완해 가며 학습하는 부스팅 계열 모형으로, 보도량의 흐름과 기사 내용, 트위터 반응처럼 성격이 서로 다른 자료를 함께 다루는 데에도 적합하였다.

3.6 분석과 평가 방법

먼저 각 카테고리에서 전체 언급 가운데 트위터가 차지하는 비중을 계산하여, 뉴스가 충분히 다루지 않는 영역을 트위터가 대신 기록하고 있는지를 살펴보았다. 이 비중은 트위터 언급량을 뉴스와 트위터를 합한 전체 언급량으로 나누어 구하며, 값이 50%를 넘으면 그 영역에서는 뉴스보다 트위터에 더 많은 기록이 남았다는 뜻이다.

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

예측 모형의 평가는 시간의 순서를 지켜 이루어졌다. 2024년 5월부터 2025년 5월까지의 자료로 모형을 학습시키고 2025년 6월부터 7월까지의 자료로 조율한 뒤, 성능은 학습과 조율에 사용되지 않은 2025년 8월부터 10월까지의 시험 구간에서 측정하였다. 분할 경계에 걸친 자료는 제외하여, 미래의 정보가 학습에 스며들 여지를 차단하였다. 선정된 118개 기업 가운데 시험 구간에 예측 대상 사안이 관측된 기업은 104개로, 성능 평가와 통계 검정은 이 104개 기업을 대상으로 하였다.

다음으로 트위터를 더한 것이 실제로 예측에 도움이 되는지를 검정하였다. 기계학습은 난수에 따라 학습할 때마다 결과가 조금씩 달라지므로, 서로 다른 난수 시드 다섯 개로 학습한 뒤 예측값을 평균하여 우연에 따른 변동을 줄였다. 트위터의 기여는 기업 단위 블록 부트스트랩으로 개선폭의 신뢰구간을 구해 평가하였다. 부트스트랩은 기업 단위로 자료를 무작위로 다시 뽑는 과정을 2,000회 반복하여, 개선폭이 우연히 나올 수 있는 범위를 직접 추정하는 방법이다. 아울러 Clark-West 검정(Clark & West, 2007)도 함께 보고하였다. 융합 모델은 공식기록 모델에 트위터 변수를 더한 것인데, 변수를 더하면 그것이 쓸모없더라도 계수를 추정하는 과정에서 잡음이 끼어 시험 구간의 오차가 오히려 커 보일 수 있다. 그래서 단순 비교는 트위터의 기여를 실제보다 낮게 평가하기 쉬운데, Clark-West 검정은 이 잡음을 보정하여 트위터가 예측을 정말로 개선했는지를 예측 오차를 제곱해 평균한 평균제곱예측오차(MSPE)를 기준으로 검정한다. 이 검정은 두 모형이 중첩 관계일 때, 곧 큰 모형이 작은 모형의 변수를 모두 포함할 때 적용되며, 본 연구의 융합 모델과 공식기록 모델이 이 조건을 충족한다. 두 모형의 예측은 동일한 시험 구간의 동일 표본에서 산출하였고, 검정은 융합 모델이 더 우수한지를 묻는 단측으로 수행하였다. 모든 유의성 검정은 기업을 단위로 묶어 수행하였는데(Cameron & Miller, 2015), 같은 기업의 자료를 개별 건으로 쪼개어 검정하면 결과가 실제보다 부풀려질 수 있기 때문이다.

끝으로 예측의 정확도는 두 가지 방식으로 측정하였다. 하나는 예측이 실제 보도량에서 평균적으로 얼마나 벗어나는지를 나타내는 오차 지표이며, 다른 하나는 그 오차가 나이브 베이스라인에 견주어 얼마나 줄었는지를 백분율로 환산한 상대 개선율이다(<표 5> 참조). 개선율이 0보다 크면 나이브 베이스라인보다 정확하게 예측하였음을 뜻하며, 값이 클수록 개선 폭이 크다. 표의 네 지표 가운데 MAE는 예측과 실제의 차이를 건수 단위로 평균한 값, RMSE는 큰 오차에 더 무거운 벌점을 주는 값이며, sMAPE와 MAPE는 오차를 백분율로 환산해 규모가 다른 대상을 같은 잣대로 비교할 수 있게 한 값이다. 한편 트위터 기여의 검정에는 기업 간 보도량 규모 차이를 표준화하는 sMAPE를 사용하여, 보도량이 많은 소수 대기업이 검정 결과를 지배하지 않도록 하였다.

4. 결 과

4.1 예측가능성: 나이브 베이스라인 대비

세 모델의 시험 구간 예측 성능은 <표 5>와 같다.

〈표 5〉 모델별 예측 성능(시험 구간, 기업 104개; 모든 지표는 낮을수록 정확)

모델	MAE	RMSE	sMAPE	MAPE	나이브 베이스라인 대비 개선율(MAE 기준)
나이브 베이스라인	4.673	14.967	60.05	87.67	–
공식기록 모델	3.515	12.109	50.39	49.27	24.8%
융합 모델	3.461	11.957	49.69	48.85	25.9%

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

개선율은 나이브 베이스라인의 MAE를 기준으로 삼아, 각 모델이 그 오차를 몇 % 줄였는지를 나타낸 값이다(=(나이브 베이스라인 MAE − 모델 MAE) ÷ 나이브 베이스라인 MAE × 100). 이하 본문에서 말하는 개선율은 모두 이 방식으로 구한 것이다. 여러 오차 지표 가운데 MAE를 택한 까닭은, 보도 건수처럼 0이나 작은 값이 많은 자료에서도 실제값으로 나누지 않아 값이 안정적으로 유지되고, 오차를 보도 건수라는 원래 단위 그대로 해석할 수 있기 때문이다. 반면 MAPE는 실제값이 0에 가까울수록 분모가 작아져 값이 비정상적으로 커지므로(나이브 베이스라인의 87.67이 그 예) 개선율의 기준으로 삼지 않았다.

이 오차가 실제로 어느 정도인지 보도 건수로 풀어 보면 다음과 같다. 융합 모델은 전체 시험 사안의 49%를 실제와 1건 이내로, 69%를 2건 이내로, 78%를 3건 이내로 예측하였다. 평균 오차가 3.5건으로 다소 크게 보이지만, 이는 보도가 폭증한 소수의 사례가 평균을 끌어올린 결과로, 전형적인 사안에서는 오차가 1건 안팎에 그쳤다.

공식기록 모델은 나이브 베이스라인을 통계적으로 유의하게 능가하였다(기업 104개 클러스터, 단측 t검정 $t=4.21$, $p=2.7\times10^{-5}$). 이는 보도량이 과거 보도의 단순 반복만으로는 설명되지 않으며, 그 안에 학습할 수 있는 예측 신호가 담겨 있음을 뜻한다. 곧 사회적 기록인 뉴스만으로도 미래의 보도량을 예측할 수 있다.

4.2 트위터의 기여

트위터를 결합한 융합 모델은 공식기록 모델보다 예측을 유의하게 개선하였다. 기업 블록 부트스트랩으로 구한 개선폭 Δ(sMAPE)=0.70의 95% 신뢰구간은 [0.125, 1.517]로, 하한이 0을 넘어 통계적으로 유의하였다. Clark-West 검정에서도 결과는 유의하였다($t=2.06$, 단측 $p=0.021$). 개선의 폭 자체는 크지 않다(sMAPE 기준 상대 약 1.4%). 이 개선이 공식기록 모델이 이미 확보한 높은 성능 위에 더해진 것임을 먼저 감안해야 한다. 아울러 이 값은 열네 개 범주를 평균한 것으로, 트위터가 채울 수 있는 자리가 뉴스의 공백에 한정되는 만큼 나머지 범주가 함께 섞이며 낮아진 값이다. 실제로 운영·고객 범주로 한정하면 개선폭 Δ(sMAPE)는 2.75로 전체 평균(0.70)의 네 배에 가깝다. 본 연구가 주목하는 지점도 개선의 크기가 아니라 그 성격이다. 개선이 어느 범주에서 발생했는가가 두 기록의 차이를 드러낸다(4.3절).

이 개선이 두 갈래의 피처 가운데 어느 쪽에서 비롯되었는지 확인하기 위해 트위터 피처를 나누어 살펴보았다. 하나는 트윗량과 리트윗처럼 반응의 크기를 담는 확산 계열이고, 다른 하나는 트윗에 쓰인 말을 읽는 텍스트 계열이다. 텍스트 계열만 결합했을 때는 개선이 유의하였으나(기업 블록 부트스트랩 95% 신뢰구간 [0.065, 1.160]), 확산 계열만 결합했을 때는 같은 기준에서 유의하지 않았다. 결국 예측을 개선한 것은 언급의 양이 아니라 그 내용이며, 텍스트 계열 신호가 추가 예측정보를 제공한 것으로 해석된다. 다만 그 추가 정보가 어떤 성격의 것인지는 이 분석만으로 확정되지 않으며, 결론에서 한계로 논한다. 이상의 검정 결과를 종합하면 <표 6>과 같다.

〈표 6〉 주요 통계 검정 결과

검증 내용	검정 방법	통계량	p값	95% 신뢰구간	판정
뉴스 기반 예측가능성 (공식기록 모델 vs 나이브 베이스라인)	클러스터 t검정	$t = 4.21$	2.7×10^{-5}	–	유의
트위터 결합의 추가 기여 (융합 모델 vs 공식기록 모델)	기업 블록 부트스트랩	Δ(sMAPE) = 0.70	–	[0.125, 1.517]	유의
트위터 결합의 추가 기여, 병행 검정	Clark-West 검정	$t = 2.06$	0.021 (Holm 보정 0.042)	–	유의

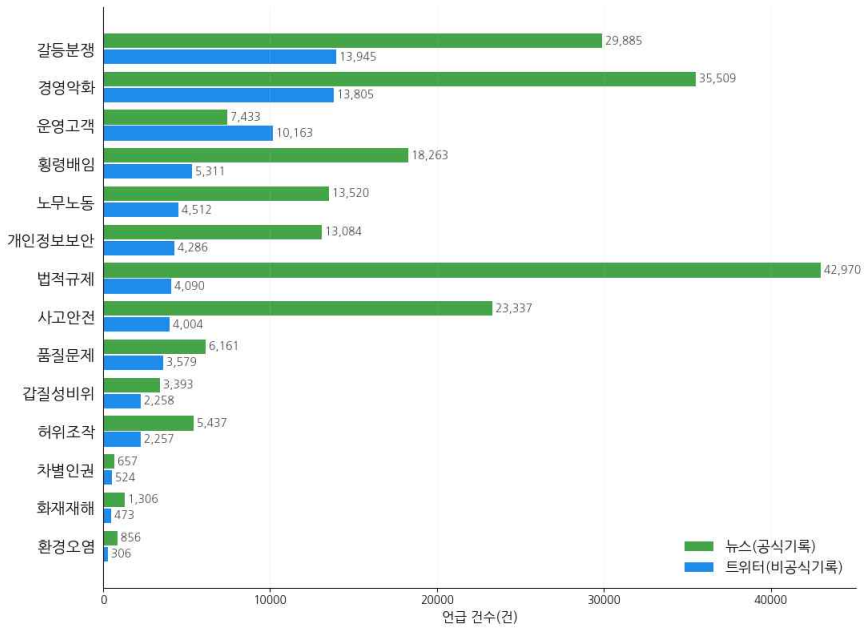
두 검정의 귀무가설(H_0)은 모두 '개선폭이 0이다'로 설정하였다. 클러스터 t검정은 기업 104개를 클러스터 단위

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

로 묶어 수행한 단측 검정이다(Cameron & Miller, 2015). 기업 블록 부트스트랩에서는 개선폭의 양측 95% 신뢰구간(하위 2.5~상위 97.5 백분위)을 구하였다. 트위터 결합이 예측을 개선한다는 방향은 결과 확인 전 연구질문에서 이미 정해진 것이므로 단측 판정을 채택하되, 그 기준을 하위 2.5 백분위가 0을 넘는지로 두어 양측 5% 검정과 동일한 엄격성을 유지하였다. 아울러 Clark-West 검정(Clark & West, 2007)을 기업 클러스터 보정과 함께 병행 수행하였다. 트위터 피처 구성별 검정 세 건(전체 · 텍스트 계열 · 확산 계열 결합)에는 Holm 다중검정 보정(Holm, 1979)을 한 묶음으로 적용하였다.

4.3 트위터의 특징: 침묵의 보완

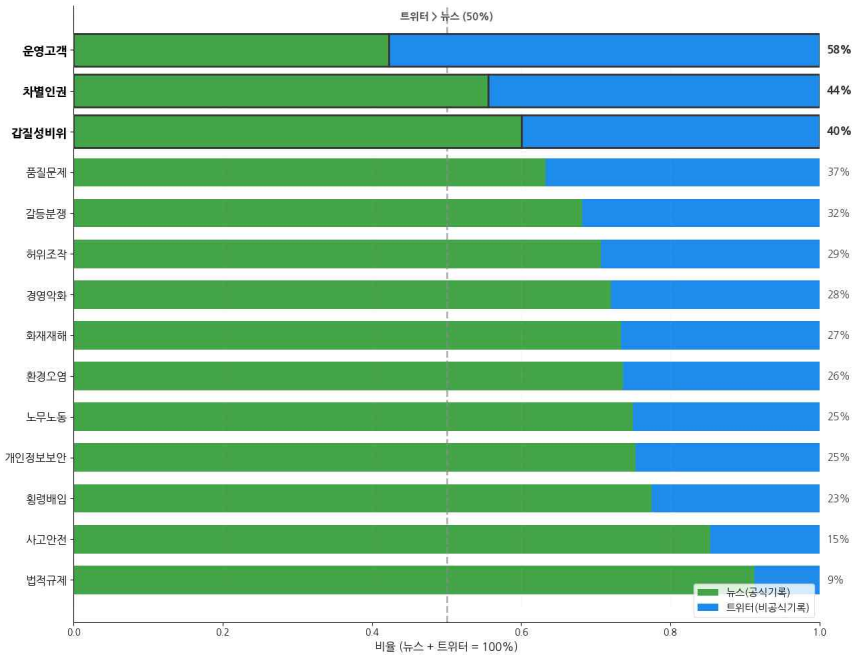
트위터는 뉴스가 좀처럼 다루지 않는 영역에서 그 공백을 메우는 비공식 기록으로 기능하였다. 이 점은 절대 언급량과 비중을 함께 볼 때 드러난다. 절대 건수로는 갈등 · 분쟁이 트위터에서 가장 많이 언급되었으나(<그림 2> 참조), 이는 뉴스 역시 그만큼 비중 있게 다룬 '큰 사안'이기 때문으로 보인다. 트위터의 고유한 기여는 비중에서 확인된다.



〈그림 2〉 카테고리별 뉴스 · 트위터 절대 언급 건수

카테고리별 트위터 비중을 보면(<그림 3> 참조), 운영 · 고객 관련 불만에서 전체 언급의 약 58%가 트위터에서 나와 14개 범주 가운데 유일하게 트위터가 뉴스를 앞질렀다. 갑질성비위(약 40%)와 차별 · 인권(약 44%) 사안에서도 트위터 비중이 상대적으로 높았으나, 두 범주는 여전히 뉴스 언급이 더 많았다. 다만 차별 · 인권은 표본이 작아(뉴스 · 트위터 합산 약 1,200건) 비중 해석에 신중할 필요가 있다. 반면 사고 · 안전(약 15%)이나 법적 규제(약 9%)처럼 사회적 파장이 큰 사안에서는 뉴스의 비중이 뚜렷하게 높았다. 이러한 양상은 대분류 수준에서도 유지되어, 운영 · 고객이 속한 갈등 · 고객 대분류의 트위터 비중(약 39%)이 여섯 대분류 가운데 가장 높았고, 규제 · 법무(약 13%)가 가장 낮았다. 대분류의 비중은 소분류 비중의 단순 평균이 아니라, 해당 대분류에 속한 소분류의 언급 건수를 모두 합산하여 산출한 값이다.

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159



〈그림 3〉 카테고리별 뉴스 대 트위터 비율(트위터 비중 내림차순, 50% 점선은 두 매체 균형선)

이러한 범주 간 비중의 차이는 보도할 만한 일로 여기는가 하는 언론의 선별 기준과 맞닿아 있는 것으로 보인다. 대형 사고나 법적 분쟁은 공적 중요성이 커 곧바로 기사화되지만, 한 소비자의 서비스 불만이나 직장 내 부당한 처우처럼 작고 사적인 일은 그 자체로 기사가 되기 어렵다. 그러나 이러한 사안은 소셜미디어에서 당사자의 목소리로 직접 기록된다. 실제로 운영·고객 범주의 트위터 게시물에서는 배달 앱이나 온라인 쇼핑, 디지털 서비스의 장애와 불편을 직접 토로하는 사례가 적지 않았다.

이러한 보도 공백의 실물을 확인하기 위해, 본 연구의 뉴스·트위터 데이터에서 동일 기업·동일 범주 기준으로 뉴스 보도가 없거나 희소한 기간에 트위터 게시물이 집중된 국면을 추출하였다. 다음 두 사례는 통계적 대표성을 갖는 표본이 아니라, 앞서 확인한 비중의 차이가 실제로 어떤 모습인지를 보여 주는 예시다. 첫째, 국내 대형 웹툰 플랫폼[4]에서 특정 작품의 게재를 둘러싸고 이용자 불매 운동이 일어난 2024년 10월, 운영·고객 범주의 게시물은 한 달 동안 812건 쌓였고 공감·재확산 반응은 172만 회를 넘었으며, 게시물 열에 아홉이 이 사안을 직접 언급하였다. 같은 기간 같은 범주의 보도는 24건에 그쳤다. 트위터에는 오프라인 행사 예약을 취소하고 서비스에서 탈퇴했다는 인증, 탈퇴와 환불의 구체적인 방법 공유, 이용자 수 감소 추이의 추적, 문제 작품과 플랫폼의 과거 운영 사례를 대조하는 이용자들의 자체 검증까지, 운동의 실행 과정 자체가 당사자의 목소리로 기록되었다. 반면 이 사안에 대한 첫 보도는 게시물이 쌓이기 시작한 지 나흘 뒤에야 나왔고, 반발이 정점을 지난 지 보름 뒤에야 환불과 탈퇴 인증이 잇따른다는 요약과 이용자 수 감소 수치를 전했으며, 탈퇴 방법의 공유나 이용자들의 자체 검증 같은 운동의 내부 과정은 보도에서 확인되지 않았다. 같은 사안을 두고 비공식 기록은 행위의 과정을 남겼고, 공식 기록은 그 결과를 뒤늦게 요약해 등록한 것이다. 둘째, 한 지상파 방송사의 기상캐스터가 직장 내 괴롭힘 끝에 사망한 사실이 알려진 2025년 1월 말부터 약 3주간, 노무·노동 범주의 게시물은 86건 쌓이며 평상시의 스무 배가 넘는 수준으로 이어졌고, 그 대부분이 이 사건에 관한 것이었다. 게시물은 진상 규명과

4) 본 절의 사례에서 기업명을 밝히지 않은 것은, 본 연구의 관심이 특정 기업의 잘잘못이 아니라 같은 사안을 두 기록이 어떻게 남기는가에 있기 때문이다. 이미 종결된 개별 기업의 부정 이슈를 다시 지목할 필요가 없다고 보았으며, 이는 기사와 게시물의 원문을 재배포하지 않는다는 본 연구의 자료 이용 원칙과도 일관된다.

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

책임자 조사를 요구하는 목소리와 함께, 타사의 직장 내 괴롭힘은 보도해 온 언론사가 자사의 사안에는 침묵한다는 지적을 담고 있었다. 실제로 공분이 표출되기 시작한 첫 일주일 동안 그 방송사 자신의 보도는 한 건도 확인되지 않았고, 같은 3주 동안 같은 범주의 보도도 4건에 그쳤다. 기록을 생산하는 언론사가 스스로 당사자가 된 사안에서 보도 공백이 나타난 경우로, 무엇을 보도할 만한 일로 여기는가라는 선별이 작동하는 지점을 가장 뚜렷하게 보여 주는 일례다. 적어도 이 두 사례 모두에서 시민들의 목소리는 공식 기록이 침묵하거나 뒤처진 동안 비공식 기록에 먼저 쌓였다. 공식 기록이 같은 사안을 다룰 때에도 그것은 반응의 결과를 요약하는 기록이었고, 행위의 과정을 남긴 것은 비공식 기록이었다.

다만 이 해석에는 신중할 필요가 있다. 자동 분류 과정에서 운영·고객 범주에 성격이 다른 사안, 이를테면 특정 콘텐츠를 둘러싼 논란 등이 일부 섞여 들어가, 이를 순수한 소비자 불만이라고 단정하기는 어렵다. 아울러 여기서 비교한 비중은 수집 단위가 서로 다른 뉴스 기사와 트위터 게시물의 건수를 견준 것으로, 매체 간 언급량의 상대적 차이를 나타낼 뿐 뉴스가 특정 사안을 배제하는 과정을 직접 측정한 것은 아니다. 따라서 본 연구가 말하는 '침묵의 보완'은 이 언급량 차이를 공식 기록의 공백으로 해석한 것이며, 기록학적 의미의 침묵, 곧 뉴스 생산 과정의 선별과 배제 그 자체를 관측한 결과는 아니다. 그럼에도 전체적인 흐름은 뉴스가 침묵하는 사적 불만의 영역을 트위터가 메우는 양상을 일관되게 보여 준다.

실제로 트위터 결합의 개선폭을 카테고리별로 나누어 보면, 개선이 가장 큰 범주는 운영·고객(Δ(sMAPE)=2.75)이었다. 트위터 비중이 가장 높았던 범주에서 예측 개선 폭 역시 가장 컸다. 반면 뉴스 보도가 이미 풍부한 범주들에서는 개선이 거의 나타나지 않아, 트위터의 기여가 뉴스가 비워 둔 자리에서 나온다는 해석과 일관된다. 범주 단위의 표본이 작아 이 분해는 방향의 확인에 해당한다.

트위터가 예측에 더하는 고유한 가치는 이 침묵의 보완에 있다. 뉴스만으로는 잡히지 않는 영역의 신호를 트위터가 채워 줌으로써, 두 기록을 함께 활용한 모델은 더 넓은 범위의 이슈를 내다볼 수 있게 된다.

4.4 사건의 동시성

침묵의 보완과는 대조적으로, 사고·안전 카테고리에서는 트위터가 뉴스의 공백을 메우기보다 뉴스와 함께 움직이는 동시 반응 양상이 나타났다. 뉴스 보도량과 트위터 언급량의 교차상관은 0.79로 열네 개 카테고리 가운데 가장 높았고, 두 곡선이 시차 없이 거의 동시에 치솟았다. 카테고리별로 뉴스와 트위터의 최대 교차상관과 그 시차를 구하면 <표 7>과 같다. 차별·인권 한 범주를 제외한 열세 개 범주에서 최대 상관의 시차가 0으로 나타났다. 일 단위로 집계한 자료에서 트위터 언급과 뉴스 보도가 같은 날 함께 움직인다는 뜻이다.

다만 이 결과를 읽을 때에는 두 가지 제약을 함께 고려해야 한다. 첫째, 본 분석은 일 단위로 집계한 자료에 기반하므로 하루보다 짧은 선행은 원리상 관측되지 않는다. 앞서 살펴본 Burns et al.(2025)가 시간 단위 분석에서 약 3시간의 선행을 보고한 것처럼, 두 기록의 선후는 하루 안에서 갈릴 수 있다. 따라서 시차 0은 두 기록이 정확히 같은 시점에 반응하였다는 뜻이 아니라, 하루 이상의 체계적인 선행이 관측되지 않았다는 뜻이다. 둘째, 교차상관은 원계열을 표준화한 뒤 산출한 것이어서, 두 계열에 남아 있는 자기상관과 추세가 상관의 크기를 부풀리고 시차 0에서 최댓값이 나타나도록 작용하였을 수 있다. 이를 통제하려면 사전백색화나 잔차 기반 교차상관이 필요하며, 이는 후속 과제로 남는다. 이러한 제약을 고려하여 본 연구는 <표 7>을 두 기록이 같은 국면에서 함께 움직인다는 양상의 확인으로 한정해 해석하고, 인과의 방향이나 상관의 정확한 크기를 주장하는 근거로 삼지 않는다.

이러한 한정 위에서 자료를 보면, 대형 참사나 산업 현장의 사망 사고와 같이 공적이고 충격이 큰 사건에서는 뉴스 보도와 동시에 트위터에서도 추모와 분노, 뉴스 리포스트가 함께 쏟아지는 양상이 관찰되었다. 실제로 항공기 참사, 물류센터 사망 사고, 제철소 감전 사고 등이 이러한 동반 급증의 대표적 사례로 확인되었다.

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

〈표 7〉 카테고리별 뉴스·트위터 교차상관과 최대 상관 시차

카테고리	최대 교차상관	시차(일)
사고안전	0.79	0
개인정보보안	0.75	0
경영악화	0.64	0
법적규제	0.59	0
갈등분쟁	0.50	0
노무노동	0.48	0
횡령배임	0.43	0
갑질성비위	0.41	0
운영고객	0.40	0
화재재해	0.37	0
품질문제	0.35	0
차별인권	0.20	+5
허위조작	0.18	0
환경오염	0.11	0

그러나 이러한 동시성은 침묵의 보완과 성격이 다르다. 앞서 밝힌 집계 단위의 제약 때문에 뉴스가 트위터를 이끌었는지, 트위터가 뉴스를 이끌었는지, 아니면 두 기록이 같은 외부 사건에 각각 반응한 것인지는 가려낼 수 없다. 그럼에도 열세 개 범주에서 하루 이상의 선행이 관측되지 않았다는 결과는, 두 기록이 같은 사건에 하루 안에서 함께 반응하였다는 해석과 가장 부합한다. 이 국면에서 두 기록은 서로를 대신하기보다 중복된다. 같은 사건이 뉴스에는 보도로, 트위터에는 그 보도를 퍼 나르는 리포스트와 추모·분노의 반응으로 나란히 남아, 기록의 형태는 달라도 가리키는 사건은 하나다. 따라서 예측의 관점에서 이 국면의 트위터는 뉴스가 갖지 못한 새로운 정보를 더해 주기 어렵다. 더욱이 사고·안전 범주의 높은 상관에는 특정 대형 참사 한 건의 영향이 크게 반영되어 있어, 수치를 그대로 일반화하기는 어렵다.

물론 <표 7>의 상관은 분석 기간 전체를 평균한 값이므로, 4.3절의 사례처럼 특정 국면에서 비공식 기록이 먼저 쌓이는 전개와 모순되지 않는다. 전체 기간의 평균에서 범주 간 차이를 만드는 것은 상관의 크기다. 시차는 열세 개 범주에서 모두 0으로 같았기 때문이다. 사고·안전(0.79)처럼 상관이 높은 범주에서는 두 기록이 같은 사건에 함께 반응하는 동시성이 지배적이다. 반면 운영·고객(0.40)처럼 상관이 낮은 범주에서는 두 기록이 같은 리듬으로 움직이지 않으며, 이는 앞 절에서 확인한 침묵의 보완과 부합한다.

종합하면 비공식 기록이 공식 기록에 대해 수행하는 역할은 사건의 규모와 공공성에 따라 두 갈래로 나뉘는 것으로 보인다. 작고 사적인 사안에서 비공식 기록은 공식 기록이 남기지 않는 것을 대신 남기는 상보적 기록으로 작동한다. 반면 크고 공적인 사안에서는 공식 기록과 같은 사건을 같은 날 겹쳐 남기는 중복적 기록에 머문다. 그리고 예측에 실질적으로 기여하는 것은 전자, 곧 상보적 기록으로 작동하는 경우다.

5. 결 론

본 연구는 118개 기업의 부정 이슈를 대상으로, 공식 기록인 뉴스와 비공식 기록인 트위터를 결합하여 향후 보도량을 예측하고 그 과정에서 두 기록이 수행하는 역할의 차이를 규명하였다. 여기서 공식·비공식 기록이라는 명칭은 생산 절차의 제도화 정도에 따른 연구 목적상의 조작적 정의로, 기록학의 제도적 공식 기록 개념과는 구별된다(2.1절). 기존 연구가 사회적 기록을 수집과 보존이라는 과거지향적 단계에서 다루거나 뉴스와 SNS를

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

예측을 위한 입력 데이터로만 취급해 왔다면, 본 연구는 뉴스와 트위터를 공식・비공식 기록으로 재정의하고 미래 예측에 활용하였다는 점에서 차별성을 갖는다.

분석 결과는 다음과 같다. 먼저 뉴스만으로도 미래 보도량을 유의하게 예측할 수 있었다(나이브 베이스라인 대비 오차 개선율 24.8%). 여기에 트위터를 결합한 융합 모델은 예측 오차를 다시 낮추었으며(sMAPE 기준 상대 약 1.4%), 이 개선은 통계적으로 유의하였다. 이 값의 비교 대상은 나이브 베이스라인이 아니라 이미 오차를 24.8% 줄인 공식기록 모델이다. 개선의 크기보다 눈여겨볼 것은 그 성격이다. 트위터가 더하는 기여는 영역에 따라 질적으로 달랐다. 운영・고객 불만처럼 작고 사적인 사안에서 트위터는 뉴스가 다루지 않은 사안을 대신 남기는 상보적 기록으로 기능하였다. 반면 대형 사고처럼 크고 공적인 사안에서는 같은 사건을 같은 날 함께 기록하는 중복적 기록에 가까웠다. 두 역할 가운데 예측을 실질적으로 개선한 것은 전자, 곧 상보적으로 작동한 경우였다.

이러한 결과는 트위터가 뉴스를 보강하는 부수적 자료에 그치지 않고, 영역에 따라서는 공식 기록이 놓치는 사회적 현실을 담아내는 질적으로 다른 기록일 수 있음을 시사한다. 나아가 본 연구는 기록학이 과거를 되짚는 개념으로 다루어 온 '기록의 침묵'을 관측할 수 있는 형태, 곧 뉴스의 보도 공백으로 조작적으로 정의하고, 그 공백이 미래 보도량을 내다보는 자원이 될 수 있음을 보였다. 이는 기록의 쓰임을 보존에서 예측으로 넓히려는 시도다.

이러한 관점은 기록화 실무에도 함의를 갖는다. 어떤 범주에서 트위터 언급이 전체의 절반을 넘고(운영・고객 약 58%) 같은 기간의 보도가 드물다면, 뉴스만 모아 둔 기록에는 그 범주의 사회적 반응이 남지 않는다. 트위터 언급 비중과 보도 빈도라는 지표는 기업의 기록관리 부서나 사회적 기록을 수집하는 기관이 무엇을 수집할지 판단하는 참고 신호가 될 수 있다. 개별 기관의 업무에서가 아니라 사회적 관심의 분포에서 무엇을 남길지 묻는다는 점에서 이는 거시평가의 문제의식(Cook, 2005)과 맞닿으며, 기록관리 체계를 갖춘 경우가 드문 기업 영역에서 아카이브의 필요성을 제기하는 근거도 될 수 있다. 다만 예측의 자원이 된다는 점과 기록화 대상 선별의 기준이 될 수 있다는 점은 서로 다른 차원의 문제다. 후자는 본 연구가 검증한 결과가 아니라 제언이며, 기록관리 현장의 판단 기준과 함께 별도로 검토되어야 한다. 아울러 트위터 역시 이용자 구성의 인구통계학적 편향과 플랫폼 알고리즘, 노이즈가 강하게 개입되는 매체이므로(2.2절), 소셜미디어 언급이 많고 보도가 드물다는 이유만으로 수집 우선순위를 부여한다면 소셜미디어의 편향이 아카이브에 그대로 옮겨 새겨질 위험이 있다. 따라서 이 지표는 기록관리 전문가의 평가와 선별을 보조하는 신호로 쓰여야 하며, 이를 대신할 수는 없다.

본 연구의 한계와 범위를 분명히 해 둔다. 예측이 성립하려면 보도가 일정 수준 이상 쌓여 있어야 하므로 분석은 보도가 비교적 활발한 118개 기업으로 한정되었다. 따라서 여기서 포착한 것은 그 기업들 안에서 뉴스가 특정 범주를 비워 둔 공백이다. 표본에 들지 못한 기업의 공백이나 뉴스와 트위터 어느 쪽에도 남지 않은 영역은 이러한 방법으로는 관측할 수 없다. 예측 대상 또한 그날 이미 보도가 시작된 사안으로 한정하였으므로, 새로운 사안의 발생 자체는 예측하지 못한다. 본 연구의 예측은 이미 시작된 사안의 전개를 가늠하는 데 그치며, 사안의 발생 자체를 미리 알리는 조기 경보는 되지 못한다. 아울러 트위터 게시물의 어휘에서 얻은 신호(4.2절의 텍스트 계열)가 예측을 개선한 까닭도 본 분석만으로는 가려내기 어렵다. 트위터가 뉴스에 담기지 않은 새로운 사실이나 맥락을 실어서일 수도 있고(내용의 독자성), 뉴스와 같은 내용이더라도 소셜미디어 특유의 위험 어휘와 감정 표현이 예측에 유용한 신호로 잡혀서일 수도 있다(표현 방식과 신호가 잡히는 시점의 차이). 두 가능성을 가려내는 일은 향후 과제로 남긴다.

끝으로 본 연구의 결과는 트위터라는 단일 플랫폼과 기업 부정 이슈라는 단일 영역에서 얻은 것이므로 다른 소셜미디어나 영역으로 일반화에는 신중해야 한다. 향후 연구가 대상과 기간을 넓혀 표본에 들지 못한 기업의 보도 공백까지 포괄하고, 분류와 신뢰도 검증을 고도화하며, 비공식 기록이 공식 기록의 공백을 메우는 양상을 더 많은 사례로 확인한다면 본 연구의 발견은 한층 정교해질 것이다.

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

참고문헌

김명훈 (2021). 기록의 정보자산화: 각국 국립기록청의 정보자산 관리 동향 분석. 한국기록관리학회지, 21(4), 45-63. https://doi.org/10.14404/JKSARM.2021.21.4.045

송주형 (2014). 기록관리 대상으로서 SNS 연구. 기록학연구, 39, 101-138. https://doi.org/10.20923/kjas.2014.39.101

윤정옥 (2010). 웹 아카이브 OASIS에 관한 고찰. 한국문헌정보학회지, 44(2), 5-27. https://doi.org/10.4275/KSLIS.2010.44.2.005

윤지수 (2026). 아카이브의 탈식민화: 기억 주권, 제도적 복원, 그리고 디지털 윤리. 한국기록관리학회지, 26(2), 1-18. https://doi.org/10.14404/JKSARM.2026.26.2.001

이해영, 김영은, 김은영, 김현지, 남경희, 이미라, 이은화, 전혜영, 최정윤 (2007). 기록정보서비스의 평가 및 개선 방향. 한국기록관리학회지, 7(2), 25-42. https://doi.org/10.14404/JKSARM.2007.7.2.025

최두원, 이수진, 윤은하, 오효정 (2018). 대통령 SNS 기록물 관리방안에 관한 연구. 한국기록관리학회지, 18(2), 29-59. https://doi.org/10.14404/JKSARM.2018.18.2.029

한승희, 이혜원 (2025). 텍스트마이닝을 활용한 공공기록물 관련 뉴스 기사 분석. 한국기록관리학회지, 25(1), 103-127. https://doi.org/10.14404/JKSARM.2025.25.1.103

Arora, A., Yadav, S., Antoniak, M., Belongie, S., & Augenstein, I. (2025). Multi-modal framing analysis of news. Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, 31531-31553. https://doi.org/10.18653/v1/2025.emnlp-main.1606

Baron, J. R., Sayed, M. F., & Oard, D. W. (2022). Providing more efficient access to government records: A use case involving application of machine learning to improve FOIA review for the deliberative process privilege. Journal on Computing and Cultural Heritage, 15(1), 1-19. https://doi.org/10.1145/3481045

Booms, H. (1987). Society and the formation of a documentary heritage: Issues in the appraisal of archival sources. Archivaria, 24, 69-107.

Burns, J. C., Kelsey, T., & Donovan, C. (2025). Extracting emerging events from social media: X/Twitter and the multilingual analysis of emerging geopolitical topics in near real time. Journal of Social Media Research, 2(1), 50-70. https://doi.org/10.29329/jsomer.14

Burscher, B., Odijk, D., Vliegenthart, R., de Rijke, M., & de Vreese, C. H. (2014). Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. Communication Methods and Measures, 8(3), 190-206. https://doi.org/10.1080/19312458.2014.937527

Cameron, A. C. & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. Journal of Human Resources, 50(2), 317-372. https://doi.org/10.3368/jhr.50.2.317

Clark, T. E. & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics, 138(1), 291-311. https://doi.org/10.1016/j.jeconom.2006.05.023

Cook, T. (2005). Macroappraisal in theory and practice: Origins, characteristics, and implementation in Canada, 1950-2000. Archival Science, 5, 101-161. https://doi.org/10.1007/s10502-005-9010-2

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. Corporate Reputation Review, 10(3), 163-176. https://doi.org/10.1057/palgrave.crr.1550049

Diesner, J., Frantz, T. L., & Carley, K. M. (2005). Communication networks from the Enron email corpus "It's always about the people. Enron is no different". Computational and Mathematical Organization Theory, 11, 201-228. https://doi.org/10.1007/s10588-005-5377-0

Downs, A. (1972). Up and down with ecology—the "issue-attention cycle". The Public Interest, 28, 38-50.

Duranti, L. (1995). Reliability and authenticity: The concepts and their implications. Archivaria, 39, 5-10.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. Journal of Communication, 43(4), 51-58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Harris, V. (2002). The archival sliver: Power, memory, and archives in South Africa. Archival Science, 2, 63-86. https://doi.org/10.1007/BF02435631

Hedges, M., Marciano, R., & Goudarouli, E. (2022). Introduction to the special issue on computational archival science. Journal on Computing and Cultural Heritage, 15(1), 1-2. https://doi.org/10.1145/3495004

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2), 65-70.

Iyengar, S. (1991). Is Anyone Responsible? How Television Frames Political Issues. Chicago: University of Chicago Press.

Janzen, S., Saxena, P., Baer, S., & Maass, W. (2024). "Listening in": Social signal detection for crisis prediction. Proceedings of the 57th Hawaii International Conference on System Sciences, 2096-2105. https://doi.org/10.24251/HICSS.2024.260

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational social science. Science, 323(5915), 721-723. https://doi.org/10.1126/science.1167742

Mansoor, M. A. & Ansari, K. H. (2024). Early detection of mental health crises through artificial-intelligence-powered social media analysis: A prospective observational study. Journal of Personalized Medicine, 14(9), 958. https://doi.org/10.3390/jpm14090958

Marciano, R., Lemieux, V., Hedges, M., Esteva, M., Underwood, W., Kurtz, M., & Conrad, M. (2018). Archival records and training in the age of big data. In Percell, J., Sarin, L. C., Jaeger, P. T., & Bertot, J. C. eds. Re-Envisioning the MLS: Perspectives on the Future of Library and Information Science Education. Advances in Librarianship, 44B. Bingley: Emerald Publishing Limited, 179-199. https://doi.org/10.1108/S0065-28302018000044B010

Ng, K. W., Horawalavithana, S., & Iamnitchi, A. (2022). Social media activity forecasting with exogenous and endogenous signals. Social Network Analysis and Mining, 12, 102. https://doi.org/10.1007/s13278-022-00927-3

Nicolas, M. L. D., Desroziers, A., Caccioli, F., & Aste, T. (2024). ESG reputation risk matters: An event study based on social media data. Finance Research Letters, 59, 104712. https://doi.org/10.1016/j.frl.2023.104712

North, D. C. (1990). Institutions, Institutional Change and Economic Performance. Cambridge: Cambridge University Press.

Parekh, T., Mac, A., Yu, J., Dong, Y., Shahriar, S., Liu, B., Yang, E., Huang, K.-H., Wang, W., Peng, N., & Chang, K.-W. (2024). Event detection from social media for epidemic prediction. Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), 5758-5783. https://doi.org/10.18653/v1/2024.naacl-long.322

Samuels, H. W. (1986). Who controls the past. The American Archivist, 49(2), 109-124. https://doi.org/10.17723/aarc.49.2.t76m2130txw40746

Semetko, H. A. & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. Journal of Communication, 50(2), 93-109. https://doi.org/10.1111/j.1460-2466.2000.tb02843.x

Trouillot, M.-R. (2015). Silencing the Past: Power and the Production of History, 20th Anniversary Edition. Boston: Beacon Press.

Van Bussel, G. J. (2017). The theoretical framework of the 'Archive-as-Is': An organization oriented view on archives. Part I. Setting the stage: Enterprise information management and archival theories. In Smit, F., Glaudemans, A., & Jonker, R. eds. Archives in Liquid Times. s-Gravenhage: Stichting Archiefpublicaties, 16-41.

White, D. M. (1950). The "gate keeper": A case study in the selection of news. Journalism Quarterly, 27(4), 383-390. https://doi.org/10.1177/107769905002700403

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

Choi, Doo-Won, Lee, Su-Jin, Youn, Eun Ha, & Oh, Hyo Jung (2018). A study on a presidential SNS records management

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

method. Journal of Korean Society of Archives and Records Management, 18(2), 29-59. https://doi.org/10.14404/JKSARM.2018.18.2.029

Han, Seunghee & Lee, Hyewon (2025). An analysis of news articles related to public records using text mining. Journal of Korean Society of Archives and Records Management, 25(1), 103-127. https://doi.org/10.14404/JKSARM.2025.25.1.103

Kim, Myoung-hun (2021). Information assetization of records: Analysis of information asset management trends of the national archives in each country. Journal of Korean Society of Archives and Records Management, 21(4), 45-63. https://doi.org/10.14404/JKSARM.2021.21.4.045

Rieh, Hae-Young, Kim, Young Eun, Kim, Eun-young, Kim, Hyun-Jee, Nam, Kyoung Hee, Lee, Mira, Lee, Eunhwa, Jeon, Hye-Young, & Choi, Jung-Yun (2007). Evaluation and improvement direction of information services in records centers and archives: A case study. Journal of Korean Society of Archives and Records Management, 7(2), 25-42. https://doi.org/10.14404/JKSARM.2007.7.2.025

Song, Zoo-Hyung (2014). A study on SNS records management. The Korean Journal of Archival Studies, 39, 101-138. https://doi.org/10.20923/kjas.2014.39.101

Yoon, Cheong-Ok (2010). Research on the OASIS, a web archive in Korea. Journal of the Korean Society for Library and Information Science, 44(2), 5-27. https://doi.org/10.4275/KSLIS.2010.44.2.005

Yoon, Jisu (2026). Decolonizing the archive: Memory sovereignty, institutional repair, and digital ethics. Journal of Korean Society of Archives and Records Management, 26(2), 1-18. https://doi.org/10.14404/JKSARM.2026.26.2.001

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

[부록 A] 모델 선택 상세: 8개 알고리즘 비교

본 연구는 여덟 가지 회귀 알고리즘을 동일 조건에서 비교하여 CatBoost를 최종 모형으로 채택하였다. 모형 선택의 기준인 검증 구간 sMAPE는 <부록 표 1>과 같으며, CatBoost가 공식기록 모델과 융합 모델 두 구성 모두에서 가장 낮아 최종 모형으로 채택되었다. <부록 표 2>는 이 여덟 알고리즘의 시험 구간 성능을 두 구성 각각에 대해 보인 것이다. 시험 구간에서도 CatBoost가 두 구성 모두에서 MAE · sMAPE · MAPE가 가장 낮았으며(굵게 표시), RMSE는 Random Forest가 가장 낮았으나 차이는 크지 않았다. 아울러 여덟 알고리즘 모두에서 트위터 결합이 MAE · RMSE · sMAPE를 개선하여, 트위터의 기여가 특정 알고리즘에 의존하지 않음을 확인하였다.

〈부록 표 1〉 알고리즘별 검증 구간 sMAPE(모형 선택 기준)

알고리즘	공식기록 모델	융합 모델
Ridge	51.36	49.92
ElasticNet	51.64	50.29
Random Forest	47.09	47.18
Extra Trees	46.44	46.45
HistGradientBoosting	46.11	45.89
XGBoost	45.91	45.76
LightGBM	45.98	45.80
CatBoost (채택)	**45.72**	**45.67**

값은 검증 구간 sMAPE이며, 두 구성 모두 CatBoost가 가장 낮아(굵게 표시) 최종 모형으로 채택되었다.

[부록 표 2] 모델 구성별 8개 알고리즘의 예측 성능

알고리즘	MAE	RMSE	sMAPE	MAPE
[공식기록 모델: 뉴스 보도량 + 내용 피처]				
Ridge	3.644	12.281	53.45	57.73
ElasticNet	3.652	12.305	53.71	57.77
Random Forest	3.617	11.873	52.32	60.73
Extra Trees	3.569	11.941	51.52	57.51
HistGradientBoosting	3.588	11.975	51.66	53.14
XGBoost	3.554	12.032	51.24	51.41
LightGBM	3.570	11.982	51.08	51.11
CatBoost (채택)	**3.515**	12.109	**50.39**	**49.27**
[융합 모델: 공식기록 모델 + 트위터 신호]				
Ridge	3.535	11.927	51.55	59.39
ElasticNet	3.537	11.970	51.80	57.85
Random Forest	3.485	11.471	51.45	59.16
Extra Trees	3.490	11.746	50.71	56.01
HistGradientBoosting	3.509	11.732	50.68	52.64
XGBoost	3.519	11.909	50.61	51.08
LightGBM	3.495	11.803	50.10	49.98
CatBoost (채택)	**3.461**	11.957	**49.69**	**48.85**

모든 값은 동일 시험 구간(2025년 8~10월) · 동일 전처리 기준이다.

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159

[부록 B] 연구 기록: LLM 판정의 생산 맥락

본 연구의 LLM 판정은 뉴스에 대해 세 단계로, 트위터에 대해 한 단계로 수행되었으며, 각 판정의 모델과 항목은 <부록 표 3>, 판정 기준의 요지는 <부록 표 4>와 같다. 모든 판정은 temperature를 0으로 고정하고 배치 API로 수행하여 판정의 재현성을 확보하였으며(1차: gemini-2.5-flash-lite, 2·3차와 트위터: gemini-3.1-flash-lite), 프롬프트는 <부록 표 4>의 판정 기준을 판단 항목·판단 순서·출력 형식과 함께 명시하는 방식으로 작성하였다.

〈부록 표 3〉 처리 단계별 사용 모델과 판정 항목

단계	판정 항목	사용 모델
1차 데이터 필터링	기업 당사자·부정 이슈 여부, 카테고리	Gemini 2.5 Flash-Lite
2차 정보 추출	사건 심각도·책임 주체·피해 범위, 카테고리 재판정	Gemini 3.1 Flash-Lite
3차 오류 제거	비즈니스 리스크 여부, 사건 단계·보도 비판 강도	Gemini 3.1 Flash-Lite
트위터 판정	기업 당사자·부정 이슈 여부, 카테고리	Gemini 3.1 Flash-Lite

〈부록 표 4〉 판정 기준 요약

항목	LLM에 제시한 판정 기준
부정 이슈 선별	사고·분쟁·위법·악화 등 부정 사건이 본문의 실제 주제인가. 홍보·실적·제휴·시황은 제외
기업 당사자 여부	기업이 사건의 당사자·책임 주체·직접 피해자인가. 우연 언급·플랫폼 언급·계열사 사건은 제외
카테고리(14종)	사건의 본질로 판정. 행위 본질이 처리 절차에 우선, 인명 피해가 사고 원인에 우선
책임 주체	개인(한 사람의 일탈) / 민간(법인이 조직으로서 주체) / 국가(공공기관·공기업)
사건 심각도(1~5)	1=경미 불만 ~ 3=행정처분·리콜·수사 착수 ~ 5=다수 사망·대형 참사·파산. 보도량이 아니라 사건 자체로 판정
피해 범위	전 국민(불특정 다수의 직접 피해) / 일부 집단
사건 단계	발생(첫 보도) / 전개(수사·공방) / 결말(판결·처분 확정) / 수습(기업의 사후 대응)
보도 비판 강도(1~3)	사건이 아니라 기사의 논조: 1=중립 단신, 2=비판, 3=고발·규탄

http://dx.doi.org/10.14404/JKSARM.2026.26.3.159