

언론 빅데이터를 활용한 건설경기체감지수(CoSI) 개발 및 활용방안

2025. 11

유위성

■ 배경	4
■ 건설산업 경기 관련 지수 조사 및 분석	7
■ 건설경기체감지수(CoSI) 개발	14
■ 시사점 및 CoSI 활용방안	29

- 건설산업은 국가와 국민 경제에서 상당한 비중을 차지하는 산업으로, 지난 3월 본 원에서 '건설산업의 재탄생(Rebirth)'을 위한 미래 전략을 제시함. 건설산업이 성공적인 전략산업으로 도약하기 위한 핵심 이슈 중 하나로 빅데이터 기술이 강조됨.
- 건설시장 환경이 빠르게 변화하면서 산업 생태계의 구조와 연관성이 더욱 복잡해지고 있음. 이러한 변화에 신속히 대응하기 위해 건설경기에 대한 체감을 포괄적으로 파악하고, 다양한 참여 주체에 대응의 적시성을 제공하는 것은 경쟁력 강화에 필수적임.
- 본 고는 매일 생산되는 언론 빅데이터를 다듬고 체계적으로 정리해, 건설경기에 대한 사람들의 체감을 숫자로 나타낸 지수를 만들고자 함. 기존 설문 기반 지수로는 충분히 설명되지 않는 부분을 보완하고, 산업의 경기 변화를 일반 사용자의 시각에서 한눈에 파악할 수 있는 도구로 제시함.
 - 기존 건설기업실사지수, 부동산시장 추이, 주택소비심리, 주택경기 등은 분야별 특화된 방식으로 활용되고 있지만, 사용자인 국민의 체감과 인식을 포괄적으로 반영하는 데 제한적임.
- 건설경기체감지수(CoSI, Construction Sentiment Index)는 언론 빅데이터에서 건설경기와 관련된 단어를 목록으로 만들고(사전화, Lexicalization), 체감 점수를 부여하여 발생 빈도와 연관 규칙으로 경기를 계량화함.
 - 2023년 10월부터 2025년 9월까지 건설산업과 관련된 언론기사에 포함된 약 2,300만개 단어에서 경기와 관련성이 있는 2,080개 단어를 설정함. 단어별 체감 점수(부정, 중립, 긍정) 부여한 후 특정 시점 언론기사에 노출된 단어의 빈도와 네트워크 분석으로 CoSI 도출함.
- 지난 12개월(2024.10.~2025.9.)간 생성된 언론 기사를 분석한 결과, '가계 → 대출', '담보 → 대출·주택', '규제→대출', '공급 → 주택', '매매 → 아파트' 등의 키워드 연관규칙 패턴이 반복적으로 나타남.
 - 또한, '인공 → 지능', '자율 → 주행' 같은 기술 규칙, '부과 → 관세' 같은 대외 변수, '사망 → 사고' 등의 안전 이슈도 빈번하게 발생함. 이는 금융, 주택, 정책, 기술, 안전 등이 동시에 건설경기 체감에 영향을 주는 복합적 구조임을 시사함.
- 주간별 CoSI 추이를 살펴본 결과, 평균 약 98.7로 100(=중립) 미만 구간이 지속되며, 2025년 2~3분기에도 여전히 중립선을 밑도는 흐름을 보임. CoSI 하락 구간에서는 가계부채, 금리, 규제 강화와 서울/아파트/매매 중심의 불안정이 결합해 지수를 하락시키고, 대형 사고와 정치 이슈가 추가적인 하방 압력으로 작용한 것으로 분석됨.
 - CoSI는 건설경기 해석에 대한 패러다임 변화, 건설 금융 및 정책 효과 포착, 사회적 이슈와 경기의 연계성 구체화, 기술 혁신과 변화 파악 등 다양한 활용성을 제시함.
- CoSI는 기존 건설경기 지수의 간극을 보완하고, 건설경기 체감을 일정한 기준으로 정리해, 대외 요인과의 관계나 통계적 흐름을 보여줌. 향후 데이터 축적이 증가할수록 건설산업의 전반적인 방향과 사용자인 국민의 수요 변화를 파악할 수 있는 보조적인 도구로 활용될 수 있음.
- 특히, CoSI는 건설기업의 분양·착공 속도 조절, 투자 시기 결정, 금융기관은 가계대출 효과, 지역·유형별 금리 영향 분석, 정부·공공은 정책 및 규제 이행에 따른 경기 체감 변동성 분석 등 건설산업 전반에 걸친 통합 대시보드 역할을 지원할 수 있음.

I 배경

■ 전(全) 산업에 걸쳐 디지털 전환이 본격화되면서 AI 기술의 원동력이 되는 빅데이터 확보와 분석의 중요성이 강조되고 있음. 지난 3월 본 원에서 제시한 ‘건설산업 재탄생(Rebirth)’의 미래 전략에 부합하는 빅데이터 기반 AI 기술 확장은 건설산업이 전략산업으로 도약하는 데 필수적인 기술로 인식되고 있음.

- 다양한 산업에서 데이터 기반 의사결정의 신뢰성이 점진적으로 향상되면서 비정형적이고 비구조적인 데이터의 활용성에 관한 관심이 급증하고 있음.
- 특히 산업의 의견을 반영하는 언론 매체가 생성하는 방대한 데이터의 구조화와 데이터화(Datafication)를 위한 데이터 공학(Data Engineering)의 가치 창출에 기대감이 증가함.
- 정보의 데이터화는 다양한 속성과 생성 속도를 고려하여 분석과 추론에 활용할 수 있는 기본 구조로 데이터의 전처리 및 정제 기술에 의해 신뢰도가 좌우됨.
- 산업에 대한 경기 변동이나 사회적 트렌드 예측에 비정형데이터 형태의 언론 기사를 활용하는 것은 특정 주체가 아닌 경제와 산업에 참여하는 포괄적인 주체의 인식과 체감을 추정하는 데 역할이 부각되고 있음.

■ 디지털 전환 시대 건설시장의 급격한 변화는 불확실성을 증가시켜, 시장 경기에 대한 체감을 정확하게 분석하는 데 어려움을 제공함. 신속한 경기 체감을 제시하여 건설산업의 참여 주체에 선제적인 대응과 적시성 제공은 경쟁력 강화에 필수적임.

- 변화 대응에 대한 적시성의 중요성이 증가하는 것은 건설산업에 참여하는 다양한 주체들의 역할과 미래 전략 수립이 시장에서 발생하는 불확실성과 관련성이 높기 때문임.
- 건설시장에 대한 트렌드와 이슈를 광범위하게 다루는 언론 기사를 활용한 지수는, 시장의 불확실성과 현안을 신속히 발굴하고, 참여 주체가 체감하거나 인식하는 경기 수준을 계량화함으로써 건설산업의 경기를 추정하는 표준이 될 수 있음.

■ 최근 한국언론진흥재단에서 제공하는 빅 카인즈 (BIG KINDS)는 국내 주요 언론에서 생산하는 기사로 이슈 도출 및 키워드 간 연관성 분석에 활용되지만, 건설산업과 같은 특정 산업에 맞춤형 경기 수준 분석과 추정에는 한계가 있음.

- 빅 카인즈를 활용한 해외건설시장 이슈 동향 분석, 미래 건설기술에 대한 연관어 분석 등은 있지만, 키워드 간 극성(Polarity)을 활용한 사례는 제한적임.
- 본고는 전(全) 산업이 아닌 건설산업에 맞춤형하고, 핵심 키워드에 대한 체감 사전을 구축한 후 경기 수준을 지수화한 사례로 차별성을 제시함.

■ 현재 전(全) 산업의 경기는 기업과 소비자 의견을 기반으로 한 대표 지수인 기업실사지수(BSI)와 소비자동향지수(CSI) 등을 통해 제시되며, 전반적인 경기의 선행 지표로 활용되고 있음.

- 이들 지수는 실질적인 통계 정보를 제공하는 역할보다 경기에 대한 선행 지수로 활용되고, 설문 등과 같은 구조적인 데이터에 의존하는 경향이 높음.
- 최근 방대한 데이터가 생산되면서 BSI, CSI와 같은 구조적인 데이터를 활용한 지수와 달리 산업에 참여하는 주체의 포괄적인 인식과 체감을 추정할 수 있는 기능이 강조됨.
- 또한, 건설산업에서 건설기업실사지수(CBSI)가 활용되고 있지만, 기업 관점에서의 건설경기를 추정하고 있어 산업의 전반적인 인식과 체감을 반영하는 데 활용도가 제한적임.

■ 건설산업의 경기 전망을 위해 건설기업실사지수, 부동산시장 추이, 주택소비심리, 주택경기 등 분야별 특화된 방식으로 분석하여 정책 방향 수립에 활용되고 있음.

- 특정 분야에서 구조화된 데이터를 활용한 경기 전망은 조사 대상의 인식과 체감을 추정하는데 효율적이지만, 언론기사 등과 같은 방대하고 포괄적인 정보를 반영하는 데 제한적임.
- 텍스트 중심의 언론기사는 대표적인 비정형데이터로 최근 정보의 데이터화 기술 발전으로 활용도가 증가하고 있음. 이는 기존의 다양한 지수들과 함께 체감경기를 반영하는 데 있어 상호보완적인 건설경기 해석을 지원해 분석의 다양성을 제공해 줄 수 있음.

■ 최근 건설산업 및 국토교통과 관련된 언론기사들이 방대하게 생산되고 있지만, 이는 텍스트 중심의 비정형데이터로 구성됨.

- 언론 빅데이터는 건설산업 경기의 전반적인 체감 정보를 제공하고 있지만, 이를 체계화한 사례는 부족한 실정임. 본고에서는 언론 빅데이터를 구조화하여 건설경기체감 수준을 계량화할 수 있는 지수를 제공하는 데 목적을 둠.

■ 본 고에서 건설산업과 관련된 방대한 언론 빅데이터를 자연어처리 등 텍스트마이닝 기술로 구조화함. 이를 경기의 체감 수준(부정, 중립, 긍정)으로 재구성하여 계량화하는 건설경기체감지수(CoSI) 개발과 활용방안을 제시함.

- 국토교통 분야와 관련된 언론기사¹⁾를 활용한 건설경기 분석은 산업에 포괄적으로 참여하는 주체의 인식과 체감 수준을 체계화함. 이를 통해 건설환경 변화에 신속히 대응할 수 있는 정책과 기업의 경영 전략 수립의 기초자료로 유용하게 활용될 수 있음.

1) 국토교통부 주요 보도자료, <https://www.molit.go.kr/clink/list.jsp>

- 건설경기체감지수(CoSI)는 건설시장의 주요 이슈와 변화를 신속히 파악하고, 필요한 정책을 시의적절하게 제시할 수 있음. 또한, 정책에 대한 효율적인 피드백을 분석하는데 맞춤형 자료로 활용될 수 있음.
 - 건설기업들은 CoSI를 활용하여 시장 변동성과 핵심 이슈를 선점하고 전략 수립과 의사결정에 활용할 수 있음. 또한, 현재 건설경기 전망에 주도적으로 활용되는 건설기업실사지수(CBSI)와 상호보완적인 분석은 신뢰도 높은 정책 방향을 제시할 수 있음.
 - 산업 차원에서 건설환경 및 수요 변화를 종합적으로 평가하고, 구조적 변화를 선제적으로 대응하기 위한 전략 수립의 효율적인 기초자료로 활용성이 기대됨.

Ⅱ 건설산업 경기 관련 지수 조사 및 분석

1. 건설경기 전망 관련 지수

- 건설산업은 국가와 국민 경제에서 상당한 비중을 차지하는 산업으로 경제경기에 미치는 영향이 크고, 경기 변동에 따라 민감하게 반응하는 산업임.

 - 기존에 건설경기를 추정하는 데 대표적으로 건설기업경기실사지수(CBSI), 건설투자액, 건설수주액, 경기동행지수 등의 계량적 지표들이 활용되고 있음.
 - 이러한 지표들은 통계청, 한국은행, 한국건설산업연구원 등에서 제공되며, 계량적 데이터를 기반으로 정량적인 비교가 가능한 장점이 있음. 하지만, 건설현장이나 국민이 체감하는 경기와 차이가 발생할 수 있는 한계가 있음.
 - 특히, 건설경기는 정책, 사회적 기대감, 현장 심리 등에 크게 영향을 받기 때문에 기존의 계량적 지표만으로 경기를 충분히 설명하는 데 한계가 있음.
- 건설산업에서 큰 비중을 차지하고 있는 부동산과 관련된 한국부동산원의 매매 및 전세 가격지수, KB국민은행의 주택가격 동향지수, 부동산시장 소비자심리지수 등은 부동산 경기 추정에 효율적으로 활용되고 있음. 각각 지수는 특정 분야에 설명력이 높지만, 건설산업 전반에 걸친 경기 체감을 설명하는 데 제한적임.
- 본 장에 건설산업과 관련된 다양한 지수들의 특징을 살펴보고, 분야별 지수의 역할을 조사 및 비교하여 건설경기체감지수(CoSI)와 상호보완적 활용 방향을 제시함. <표 1>은 건설산업과 관련된 지수들의 특징과 장·단점을 요약하여 제시함.
- 기업실사지수(Business Survey Index, 이하 BSI)는 기업의 경기 전망과 체감을 조사한 지수로 한국은행이 매월 제조업과 비제조업 기업 대상 결과를 제공함.

 - 기준값 100은 경기를 긍정적으로 인식하는 기업과 부정적으로 인식하는 기업의 비중이 유사한 수준을 의미함. 또한, 100 이상은 긍정적으로, 100 이하는 경기를 부정적으로 인식하는 것으로 해석됨.
 - 기업의 체감경기를 반영하고, 경제경기의 선행지표로 활용되어 기업의 기대 심리를 신속히 파악하여 경기 전환점을 민감하게 포착할 수 있음. 하지만, 조사대상자의 기업관점과 판단에 따라 편향될 수 있고, 표본구성에도 한계가 있음.

- 소비자심리지수(Consumer Sentiment Index, 이하 CSI)는 소비자의 경제 상황에 대한 인식을 조사하는 지수로 한국은행이 매월 가구 단위를 대상으로 조사함.

 - BSI와 유사하게 100을 기준값으로 설정하고 100 이상이면 경제 상황이 낙관적, 100 이하이면 비관적인 것으로 인식함.
 - CSI는 경제활동에 있어 소비자의 소비심리와 향후 소비지출 성향을 반영하고, 경기 동행·선행지표로 사용되어 가계의 체감경기를 파악할 수 있음. 하지만, 응답자 설문 중심으로 조사되어 주관에 의존하거나 실제 소비행태와 차이가 발생할 수 있음.
- BSI와 달리 건설산업에 특화된 건설기업실사지수(Construction Business Survey Index, 이하 CBSI)는 건설기업들이 체감하는 경기와 전망을 조사한 지수임. 기준값을 100으로 설정하여, 100보다 높거나 낮음에 따라 건설 경기를 낙관 또는 비관적인 수준으로 해석함.

 - CBSI는 건설업계의 경기 전망에 특화되어 건설기업의 인식과 체감을 설문 중심으로 조사 후 직접적으로 반영하고, 공공·민간 건설의 수요 전망에 활용되고 있음.
 - 하지만, 건설기업 규모별로 건설경기에 대한 체감에 차이가 발생되어 실질적인 체감에 대한 왜곡이 발생될 수 있고, 지역별 차이를 충분히 반영하는 데 한계가 있음.
 - 또한, 전체 건설기업에 비해 조사된 기업 비중은 약 1% 미만이고, 기업관점에서 경기에 체감과 인식을 조사한 결과임. 하지만, 연관산업과 사용자인 국민 관점의 체감을 포괄적으로 반영하는 것은 제한적임.
- 매매가격지수는 주택, 아파트 등 매매가격 변동을 지수화한 것으로 한국부동산원, KB국민은행 등에서 제공함. 실거래가 또는 감정평가 데이터를 사용하여 기준점을 100으로 설정하고 매매가격의 상승 또는 하락 수준을 제시함.

 - 자산가치의 변화를 파악하고 주택시장의 경기 동향을 추정하는 데 효율적임. 하지만, 표본 및 평가 방식에 따른 오차 발생 가능성과 거래 데이터가 적은 지역에 대한 추정값의 신뢰도가 낮을 수 있음.
- 전세가격지수는 주택 전세가격의 변동을 지수화한 것으로 매매가격지수와 유사한 방식임. 시장 동향, 수급 상황, 임차인의 주거비 부담 수준 등을 파악할 수 있지만, 지역별 전세 물량 차이로 왜곡 또는 대표성 부족이 발생할 수 있음.
- 부동산시장 소비자심리지수는 부동산시장에 대한 소비자들의 심리를 조사한 지수로 한국부동산원이 매월 발표함. 100 이상이면 시장이 과열되어 매수심리가 강하고, 100 이

하이먼 매수심리가 약한 것으로 해석됨.

- 이는 매수 또는 매도 심리에 대한 의향을 파악하고, 시장 참여자들의 기대 심리를 반영함. 시장 과열 또는 침체를 조기에 감지하고, 정책 대응 판단에 활용되지만, 심리와 실제 거래가 다를 수 있고, 조사대상자의 편향 가능성이 있음.

〈표 1〉 건설산업 경기 관련 주요 지수 비교

지표명	설명	특징	장점	단점
건설기업 실사지수 (CBSI)	건설기업의 경기전망과 체감경기를 조사해 지수화	- 건설산업에 특화된 BSI - 100 이상이면 경기 양호, 100 이하면 경기침체	- 건설산업 경기 동향 파악에 적합 - 정책 대응에 유용	- 기업 중심의 주관적/정성적 평가 - 계절적, 심리적 요인의 영향 큼
기업실사 지수(BSI)	기업의 경기 인식을 조사하여 지수화	- 주관적 기대를 반영 - 100 이상이면 긍정, 100 이하면 부정	- 경기 방향의 신속한 파악 가능 - 실제 데이터가 생성되기 전에 선행 가능	- 주관적 판단으로 왜곡될 수 있음 - 계절적, 심리적 요인의 영향 큼
소비자심리 지수(CSI)	소비자의 경제 상황과 경기 전망에 대한 인식을 지수화	- 소비자의 소비의향, 경제전망 등 심리 반영 - 100 이상이면 낙관적, 100 이하면 비관적	- 소비 성향과 경기 간 관계 파악에 용이 - 소비자심리의 선행지표로 유용	- 주관적이고 감정적 요인에 의한 영향이 큼 - 단기 변동성 존재
매매/전세 가격 지수	일정 기간 주택 매매/전세 가격의 변동을 지수화	- 실거래가 또는 호가를 기준으로 계산	- 주택시장 동향을 직관적으로 파악 가능 - 장기 추세 분석에 유용 - 전세시장 수급 파악이 용이 - 월세 전환율 등과 연계 분석이 가능	- 매매/전세 거래량 적은 시기에 왜곡 - 비공식거래나 특수거래의 반영이 어려움 - 특정 지역 쏠림 현상 발생 가능
주택가격 동향 지수	매매와 전세 등을 종합하여 주택시장의 전반적 가격 동향을 지수화	- 매매+전세의 종합지표	- 주택 전체 시장 흐름 파악 가능	- 세부 시장 분석에는 한계
부동산시장 소비자심리 지수	소비자들이 체감하는 부동산시장의 심리를 지수화	- 소비자 기대와 시장 심리를 반영	- 매매·전세 수요 심리 변화 파악 가능 - 선행적 성격으로 활용	- 주관적이고 과잉반응 가능성 - 실제 거래와 괴리 발생 가능

■ BSI, CSI, CBSI, 부동산시장 소비자심리지수 등은 설문 중심의 분야별 경기 체감을 조사한 지수로 건설경기 추정에 유용하나, 일부는 관련 기업 및 특정 지역 중심으로 대표성에 한계가 존재함.

- 매매 및 전세 가격지수 등은 실거래 데이터를 기반으로 한 지표로 현실성은 반영되어 있지만, 거래량이 적거나 지역 간 편차가 크면 대표성이 낮은 특징이 있음.

2. 건설경기 지수 추이 비교와 언론 빅데이터의 활용성 검토

- **현행 활용되고 있는 건설산업과 관련된 주요 지수의 추이를 비교 및 분석하기 위해 z-score²⁾를 활용하고, 각 지수를 표준화함.** z-score는 다양한 지수의 특정 값이 평균에서 이탈된 수준을 동일한 기준에서 추이와 변동성을 비교하는 데 효과적임(〈표 2〉참조).

〈표 2〉 건설산업 경기 전망과 관련된 주요 지수 간 z-score 비교

	건설기업 실사지수(X ₁)	기업실사 지수(X ₂)	소비자동향 지수(X ₃)	건축허가 건 수(X ₄)	건설수주액 (X ₅)	부동산시장 소비자심리 지수(X ₆)
2023년 10월	-1.83	-1.36	-0.40	0.10	-0.24	0.01
2023년 11월	0.80	-1.36	-0.56	1.69	-0.23	-1.31
2023년 12월	1.44	-1.36	-0.16	1.01	3.27	-1.70
2024년 1월	-1.16	-1.46	0.20	-0.47	-0.87	-1.09
2024년 2월	0.37	-1.56	0.25	-0.97	-0.92	-0.32
2024년 3월	0.83	-1.46	0.04	0.29	-0.36	-0.64
2024년 4월	0.89	-1.06	0.04	1.19	-0.09	-0.57
2024년 5월	-0.94	-1.06	-0.35	1.32	-0.28	-0.12
2024년 6월	-0.36	1.21	0.08	-0.03	1.57	0.52
2024년 7월	0.43	1.15	0.54	0.82	-0.26	1.34
2024년 8월	-0.49	0.89	0.06	-0.22	-0.92	1.60
2024년 9월	1.47	0.76	-0.08	-1.09	-0.03	0.96
2024년 10월	0.03	0.85	0.22	0.12	-0.28	0.05
2024년 11월	-1.19	0.79	0.04	0.49	1.07	-1.01
2024년 12월	0.25	0.34	-2.07	1.03	1.07	-1.44
2025년 1월	-0.12	0.23	-1.59	-2.59	-1.12	-1.05
2025년 2월	-1.04	0.17	-0.90	-1.77	-0.73	-0.04
2025년 3월	-0.82	0.31	-1.21	0.18	-0.30	0.61
2025년 4월	1.23	0.43	-1.14	0.64	-0.36	-0.10
2025년 5월	1.07	0.71	0.23	0.05	-0.35	0.33
2025년 6월	0.83	0.66	1.42	-0.28	1.14	1.93
2025년 7월	0.71	0.64	1.78	-0.06	0.55	0.11
2025년 8월	-0.79	0.74	1.88	-0.76	-0.55	0.42
2025년 9월	-1.62	0.80	1.66	-0.72	-0.78	1.49

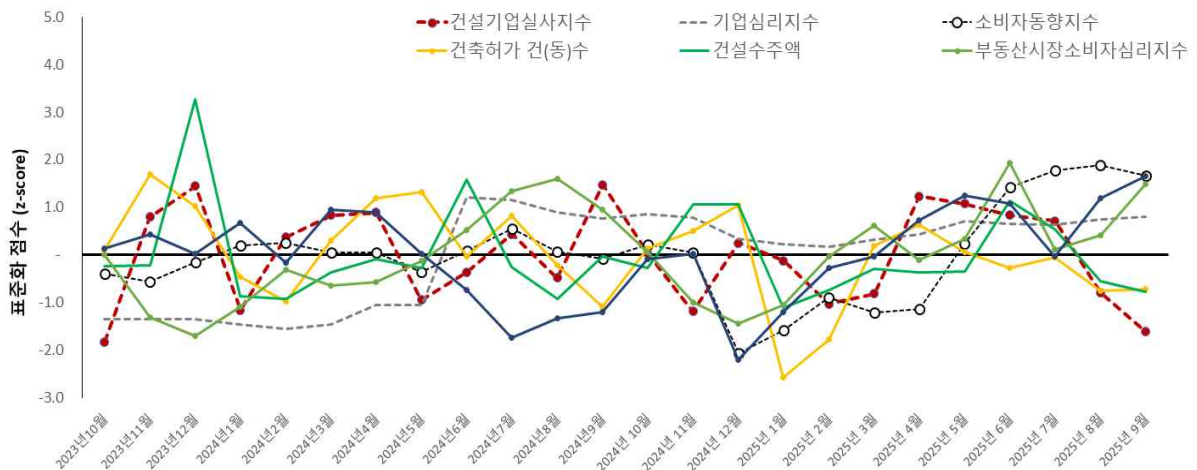
주 : 주요 지수에 대한 데이터는 한국은행, 한국부동산원, 대한건설협회, 한국건설산업연구원, 국토연구원 등에서 발표된 자료를 활용하여 연구자가 재구성함.

2) 각 지수 간 크기와 변동량에 차이가 있어, 다양한 분야에서 널리 활용되는 z-score(= (지수값 - 평균값)/표준편차) 표준화를 통해 추이를 비교함. 즉, 각 데이터 집단 내에서 어느 정도 특이한지, 평균에서 얼마나 큰 차이가 있는지 제시하여 데이터 간 추이를 살펴봄,

앞서 언급된 다양한 건설경기와 관련된 지수를 월간 기준으로 z-score 추이를 비교한 결과, 각 지수의 특징을 반영한 패턴이 다양하게 제시되고 있음. 이는 지수의 개념과 데이터 수집 대상에서 차이가 있어 건설경기의 전반적인 체감을 대표할 수 있는 지수를 설정하는 데 한계가 있음을 시사함.

- <그림 1>에 제시된 건설기업실사지수(CBSI), 기업실사지수(BSI), 소비자동향지수(CSI), 건축허가 건수, 건설수주액, 부동산시장소비자심리지수 등 각 지수의 시계열적 추이를 살펴보면, 일부 지수 간 유사한 패턴을 보이지만 변동 주기는 분야별 특징에 의해 비규칙적임.
- 한국건설산업연구원, 한국은행, 통계청, 대한건설협회 등 다양한 참여 주체에서 제공되는 건설경기 관련 지수는 정책 수립을 지원하고 산업계의 현황 분석에 활용되고 있음.

〈그림 1〉 건설경기 관련 주요 지수 간 표준화(z-score) 점수 추이 비교



〈표 3〉 지수의 표준화 점수 상관관계수 행렬

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
X ₁	1					
X ₂	-0.032	1				
X ₃	-0.041	0.214	1			
X ₄	0.249	-0.255	-0.070	1		
X ₅	0.320	0.032	0.010	0.422	1	
X ₆	-0.113	0.605	0.512	-0.240	-0.241	1

주 : X₁ : 건설기업실사지수, X₂ : 기업실사지수, X₃ : 소비자동향지수, X₄ : 건축허가 건(동)수,
X₅ : 건설수주액, X₆ : 부동산시장 소비자심리지수.

■ 건설산업의 주요 지수 간 관계성을 살펴보면, 건설기업실사지수(CBSI)와 건설수주액(0.33), 건축허가 건수와 건설수주액(0.42), 기업실사지수와 부동산시장 소비자심리지수(0.60), 소비자동향지수와 부동산 소비자심리지수(0.51) 등은 양적 규모에는 차이가 있지만, 양의 상관관계를 보여주고 있음.

- 기업실사지수와 건축허가 건수 (-0.25), 부동산시장 소비자심리지수와 건설수주액 (-0.24) 등은 음(-)의 상관관계로 기업 심리가 양호하다고 건축허가 건수가 바로 늘어나는 것은 아닌 것으로 분석됨.
- 이는 정책, 규제, 금리 등과 같은 다른 요인에 의해 추이 변화가 야기되고, 부동산 심리가 양호해진다고 해서 건설수주액이 즉각적으로 비례해서 증가하지 않음. 반대로 변동되는 경향을 보여 지수 간 변동에 대한 체계적 해석이 어려운 실정임.

■ 최근 텍스트와 같은 비정형데이터의 활용성이 급증하면서 언론기사, SNS 등의 텍스트 데이터로부터 경제경기 또는 소비자심리 동향을 분석하려는 시도가 활발함.

- 비정형데이터 분석 기술의 발전과 함께 건설경기와 관련된 언론 기사를 수집하여 주요 키워드 간 연관규칙과 네트워크 분석으로 기업 차원이 아닌 다양한 관점에서 건설경기를 추정할 수 있는 방식이 주목받고 있음.
- 이는 기존 일정 규모 이상의 기업 중심 설문 방식을 보완하여, 건설산업 내 다양한 참여 주체의 포괄적인 체감을 반영할 수 있는 역할에 기대됨.

■ 본 고에서는 매일 발간되는 언론기사에 건설경기와 관련된 단어 및 문장의 감성점수를 활용하여 특정 시점의 건설경기체감지수(CoSI)를 산정함. 또한, 기존 건설경기 관련 지표들과의 관계성을 분석함.

- CoSI는 기존의 계량적 데이터 기반 경기지수를 보완하면서, 정책 수립과 산업계의 의사결정에 활용될 수 있는 정보를 제공할 수 있음.
- 기존 건설산업에 특화된 설문과 기업 중심의 지표로 반영하기 어려운 다양한 주체의 체감 수준을 반영하여 기존 지수와 상호보완적으로 활용됨.
- 건설산업과 관련된 경기 또는 동향을 파악하는 현행 지수의 변동성이 클 때, CoSI는 키워드 간 연관규칙과 네트워크 분석으로 통계 정보와 연계하여 비구조적인 현황을 체계적으로 해석할 수 있는 특징이 있음.

Ⅲ 건설경기체감지수(CoSI) 개발

1. 건설경기 체감 키워드 간 연관규칙 분석

❖ CoSI(Construction Sentiment Index) 개발을 위해 경기에 대한 체감을 추정할 수 있는 단어 및 문장을 추출하고, 관계 분석은 연관규칙 분석(Association Rule Analysis, 이하 ARA)³⁾을 수행함.

- ARA는 대규모 데이터에서 정보를 추출하고, 연관 패턴을 발견하는 데 큰 장점이 있어, 의료, 금융, 유통 등 다양한 산업에서 활용되고 있음. 또한, 최근 건설산업에서도 텍스트 데이터 분석에 널리 이용되고 있음.
- 텍스트 중심의 대용량 자료에서 키워드 간 발생 가능한 연관규칙을 생성하는 보편적인 알고리즘에는 Apriori⁴⁾ 등이 있으며, Apriori 알고리즘은 이미 다양한 분야에서 활용되고 있어 CoSI 개발에 적용함.

❖ ARA에서 유의미한 연관규칙을 설정하기 위한 대표적인 척도는 지지도(Support), 신뢰도(Confidence), 향상도(Lift) 등으로 구성⁵⁾됨.

- 지지도 값은 전체 문장 중 특정 키워드 간 패턴이나 규칙의 집합이 포함된 비율을 나타내며, CoSI 개발에 있어 적용되는 지지도의 임계값을 0.001 ~ 0.05 내에서 설정함. 단어 및 문장의 수, 문장의 길이 등 데이터 특성과 목적을 고려함.
- 신뢰도는 키워드 A가 발견된 문장 중 키워드 B가 추가로 문장에 포함될 비율(조건부 확률)로 0.5 전후 값을 일반적으로 설정함. 신뢰도가 0.5보다 낮으면 A가 등장했을 때 단어 B가 등장할 확률이 상대적으로 낮아도 규칙으로 인정될 가능성이 커지기 때문에 유의미한 규칙이 아니라 무작위로 생성된 규칙일 수 있음.

❖ 지지도, 신뢰도, 향상도 등은 데이터의 구조 및 형태에 따라 선행적으로 탐색 구간을 설정하고, 이후 구간 내에서 실험적으로 척도의 값을 결정하는 것이 효율적임.

3) ARA는 데이터의 특정 속성(Attribute) 간에 갖는 패턴 등 연관성을 도출하는 데이터마이닝 기법으로, 특정 속성이 발생할 때 다른 속성이 발생하는 규칙을 도출함.
 4) Apriori 알고리즘은 빈번하게 발생하는 데이터 속성 집합이 모든 부분집합 역시 빈번히 출현한다는 전제하에 빈도의 동시 발생 및 종속 발생 규칙을 도출하는 특징이 있음. 키워드 간 발생 패턴이나 규칙을 비구조적 데이터에서 추출하여 속성 간 부분 집합을 선정하는 기법으로 다양한 연구에서 활용되고 있음.
 5) 신뢰도 값은 목적과 데이터 특성을 고려하여 설정되고, 향상도는 키워드 A와 B의 서로 어떠한 관계를 갖는지를 보여주는 척도임. 보통 향상도가 1보다 클 때를 유의미한 규칙으로 해석되어 1보다 큰 값을 일반적으로 설정함.

〈표 4〉 연관규칙분석(ARA)을 위한 핵심 척도

척도	설명	수식
지지도 (Support)	전체 문장 중 키워드 A, B가 동시에 포함된 문장의 비율	$Pr(A \cap B)$
신뢰도 (Confidence)	키워드 A가 출현한 문장에서 키워드 B가 추가로 해당 문장에 포함될 비율(조건부 확률)	$\frac{Pr(A \cap B)}{Pr(A)}$
향상도 (Lift)	키워드 A와 B가 같이 문장에 기술되는 빈도가 독립적으로 발생하는 것보다 얼마나 더 높거나 낮은지를 나타내는 척도 - 향상도 = 1 : 독립적인 관계 - 향상도 > 1 : 양의 관계 (무작위 발생 확률보다 높음.) - 향상도 < 1 : 음의 관계 (무작위 발생 확률보다 낮음.)	$\frac{Pr(A \cap B)}{Pr(A) \times Pr(B)}$

2. 연관규칙 기반 네트워크 분석

- 소셜네트워크분석(Social Network Analysis, 이하 SNA)은 건설경기와 관련된 키워드의 관계와 연관규칙을 노드와 링크를 통해 네트워크 형태로 구축함. 키워드 관계의 구조와 특성을 파악하는 방법으로 건설산업에서도 활용성이 증가하고 있음.
- 본 고에서 SNA는 연관규칙분석에서 생성된 유의미한 규칙 결과를 활용하여 네트워크를 구축하고, 키워드 간 관계와 패턴을 시각적으로 분석하는 데 활용함.

 - 네트워크의 특성을 확인하는 지표로 평균연결 정도(Degree)⁶⁾, 밀도(Density)⁷⁾, 포괄성(Inclusiveness)⁸⁾ 등이 있음. 네트워크의 구조적 특징을 파악할 수 있는 대표적인 지표는 중심성(Centrality)⁹⁾으로 정의함.
 - CoSI는 SNA에서 널리 활용되는 네트워크 내 노드와 연결된 링크의 수로 산정되는 중심성을 활용함.
- CoSI 산정을 위한 자료로 주요 언론에서 매일 제공하는 건설과 관련된 언론기사에서 추출된 키워드이고, 키워드 간 연관규칙 결과와 이를 활용한 네트워크 분석이 CoSI 산정에 선행됨.

6) 평균연결 정도는 네트워크의 링크 수를 노드 수로 나눈 값으로, 한 노드와 직접 연결된 이웃 노드의 수가 평균적으로 어느 정도 인지를 제시함.
 7) 밀도는 가능한 총 링크 수와 실제로 연결된 링크 수의 비율로, 노드 간의 상호작용이 얼마나 밀접하게 발생하는지를 측정하는 지표로 일반적으로 네트워크의 크기와 반비례함.
 8) 포괄성은 한 네트워크의 총 노드 수에서 연결이 없는 노드의 개수를 제외한 비율로 도출됨.
 9) 중심성(Centrality)은 네트워크를 구성하는 노드들이 형성하는 관계를 나타내는 측정 지표로, 네트워크 내에서 중심성이 높은 노드(키워드)를 파악하고, 해당 노드가 전체 네트워크에서 다른 노드와 연결되는 수준을 이해하는 데 이용됨.

- ARA에서 하나의 문장은 하나의 속성 집합(키워드 집합)에 해당하므로, SNA에서 한 문장 안에 키워드 조합이 네트워크에서는 두 노드(키워드) 간 링크로 연결되는 형태로 제시됨.
- ARA에서 키워드 간 규칙이 생성된 경우, 각각의 키워드는 네트워크에서 노드로 표현되며, 규칙은 링크로서 설명됨. 이때, 노드(키워드)의 중심성이 높을수록 네트워크 내 노드는 크게 나타나고, 두 노드가 갖는 규칙의 지지도가 높을수록 링크는 굵게 나타남.

3. ARA 및 SNA 결과 기반 건설경기 동향 분석

■ CoSI 산정을 위해 텍스트 비정형데이터에서 문장 단위의 묶은 자료로 구성하고, 단어 추출 후 건설경기와 관련된 단어(키워드) 간 연관규칙 및 네트워크 분석을 수행함.

- 건설경기와 관련된 단어 추출은 파이썬의 TIKa 라이브러리를 이용하여 추출한 후 KoNLpy 패키지의 OKT 모듈을 활용해 단어(명사)를 추가로 추출함.

■ CoSI 산정에 필수적인 건설경기와 관련된 단어의 사전화(Lexicalization)는 단어의 발생 빈도, 연관규칙, 네트워크 관계성 등을 고려한 경기 체감 단어의 체계적인 집합임.

- 낮은 발생 빈도 단어의 경우 건설경기와 연관성이 낮거나 의미 파악에 한계가 있어 제외하고, 연관 규칙 기반 네트워크 중심성에서 유의미한 상위 2,080개 단어(키워드)를 선정함.
- 이는 전체 키워드 중 약 70%의 빈도 수준에 해당하고, <표 5>는 전체 중 상위 빈도의 50개 키워드를 내림차순으로 정렬한 예시를 보여줌.

<표 5> 건설산업 관련 키워드 중 빈도 기준 상위 50개('23년 10월 ~ '25년 9월)

순위	구분	순위	구분	순위	구분	순위	구분	순위	구분
1	사업	11	시장	21	단지	31	정책	41	청약
2	건설	12	가구	22	산업	32	지원	42	교통
3	주택	13	종합	23	금융	33	전국	43	경기
4	서울	14	기준	24	공사	34	전세	44	관리
5	부동산	15	계획	25	증가	35	안전	45	업계
6	대출	16	분양	26	공급	36	도시	46	규모
7	정부	17	전기차	27	거래	37	기술	47	투자
8	아파트	18	가격	28	추진	38	수도권	48	확대
9	경제	19	금리	29	개발	39	은행	49	문제
10	지역	20	기업	30	사회	40	상승	50	수요

■ 지난 1년간 건설경기와 관련된 키워드만 포함해서 문장 단위로 구축한 데이터의 연관 규칙을 분석한 결과, 신뢰도(Confidence) 기준 상위 30개 연관규칙을 도출하여 <표 6>과 같이 제시함.

- 선행과 후행 사건(키워드) 간 높은 지지도 값을 갖는 규칙 중 ‘가계’, ‘대출’, ‘서울’, ‘아파트’, ‘주택’, ‘공급’, ‘거래’, ‘허가’, ‘부동산’, ‘시장’, ‘은행’, ‘대출’, ‘증가’, ‘상승’ 등과 같은 키워드가 포함된 연관규칙이 빈번하게 생성됨.
- ‘대출’, ‘아파트’ 키워드의 경우 후행 사건(키워드)으로 생성되는 규칙이 빈번하고, ‘거래’, ‘가격’, ‘규제(허가)’ 키워드는 주로 선행 사건(키워드)으로 규칙이 생성됨.
- 유의미한 키워드 간 연관규칙을 생성한 기준으로 지지도, 신뢰도, 향상도를 각각 0.01, 0.1, 1 이상으로 설정함.

<표 6> 키워드 간 ‘신뢰도’ 기준 상위 30개 연관규칙

규칙 번호	선행 사건	후행 사건	지지도	신뢰도	향상도	규칙 번호	선행 사건	후행 사건	지지도	신뢰도	향상도
1	재테크	금융	0.0015	0.98	110.0	16	인하	금리	0.0009	0.33	67.7
2	시중	은행	0.0008	0.81	147.3	17	규제	대출	0.0034	0.32	16.3
3	가계	대출	0.0037	0.79	39.9	18	허가	토지	0.0010	0.32	94.8
4	인공	지능	0.0009	0.78	607.1	19	매매	아파트	0.0009	0.30	22.4
5	자율	주행	0.0021	0.67	202.5	20	매매	가격	0.0008	0.30	33.8
6	담보	대출	0.0015	0.64	32.1	21	완화	규제	0.0008	0.29	27.0
7	담보	주택	0.0014	0.62	29.0	22	은행	대출	0.0015	0.26	13.2
8	신용	대출	0.0008	0.54	27.1	23	공급	주택	0.0027	0.25	11.8
9	한도	대출	0.0011	0.50	25.4	24	전세	대출	0.0012	0.25	12.7
10	당국	금융	0.0018	0.50	56.4	25	토지	주택	0.0008	0.24	11.3
11	부과	관세	0.0008	0.41	69.1	26	금리	대출	0.0011	0.24	11.8
12	정비	사업	0.0019	0.40	18.7	27	아파트	서울	0.0030	0.22	11.1
13	허가	거래	0.0012	0.38	44.2	28	가계	증가	0.0009	0.19	23.9
14	허가	구역	0.0011	0.36	69.1	29	발생	사고	0.0011	0.17	17.6
15	토지	거래	0.0012	0.35	40.2	30	증가	대출	0.0013	0.16	8.2

■ 지지도와 신뢰도가 높은 키워드 간 연관규칙을 살펴보면, ‘가계’와 ‘대출’ 관련해 이슈가 가장 빈번하게 동반되는 것을 알 수 있음.

- 이는 가계 부채와 관련하여 건설산업의 위험 조기 경보 시스템의 필요성과 소득 대비 대출 비율에

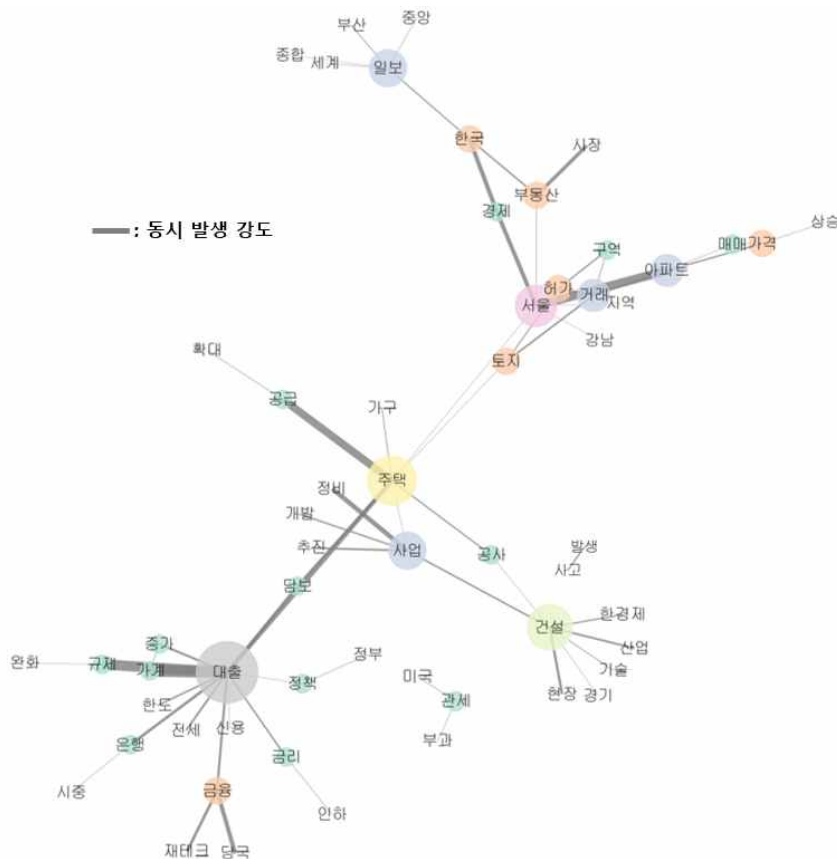
대한 모니터링 강화가 강조됨.

- 또한, 수도권 중심의 아파트 가격 및 공급 정책에 대한 맞춤형 전략을 구축하여 주택 공급 정책에 대한 수요 예측뿐 아니라 공급 계획과 시기에 대한 이슈가 부각되고 있음.
- 건설산업과 관련된 이슈는 수도권 중심으로 강조되고 있어 지역 균형발전을 위한 정책과 경제구조에 대한 논의가 강조됨. 즉 도시 정책, 부동산 공급과 거래, 가계 재정관리, 지역경제 균형 전략 수립 등이 건설산업 경기에 영향을 크게 미치는 것으로 분석됨.

■ 연관규칙을 활용한 키워드 간 네트워크 결과는 <그림 2>와 같이 '대출', '서울', '주택', '아파트', '가격', '거래' 등의 중심성이 가장 높고, 다음으로 '부동산', '매매', '가계' 등이 높게 나타남.

- '24년 10월부터 '25년 9월까지 데이터를 대상으로 도출된 결과로 건설경기에 여전히 큰 비중을 차지하고 있는 부동산 관련 핵심 키워드(노드) 간 중심성이 높음.

<그림 2> '24년 10월 ~ '25년 9월까지 주요 키워드 간 네트워크 결과



<표 7> 중심성 기준 상위 10개 키워드(노드)

순위	키워드	중심성 (Centrality)
1	대출	0.1579
2	주택	0.1053
3	서울	0.0789
4	아파트	0.0526
5	금융	0.0395
6	부동산	0.0395
7	허가	0.0395
8	토지	0.0395
9	가격	0.0395
10	가계	0.0263
11	규제	0.0263
12	공급	0.0263
13	담보	0.0263
14	금리	0.0263
15	관세	0.0263

- 중심성과 지지도(선 굵기) 기반 키워드의 군집성을 살펴보면, ‘대출’과 연관된 키워드는 ‘가계’, ‘금융’, ‘은행’, ‘규제’, ‘금리’, ‘증가’ 등으로 가계 부채 및 금융 규제 이슈와 밀접하여 규제, 저소득층 금융 안전망 확대 등의 정책이 필요함을 의미함.

- ‘서울’과 연관된 키워드는 ‘아파트’, ‘공급’, ‘주택’, ‘재건축’, ‘청약’, ‘집값’ 등으로 서울 지역 주택 공급 정책 및 가격안정 대책이 핵심이고, 지역 맞춤형 공급 정책과 규제 병행의 필요성이 강조됨.

- <그림 2>와 같이 단어 간 연결성과 군집성은 금융자산관리, 주택·부동산 공급 정책의 필요성, 인허가 시스템 고도화 등의 현안과 수요가 강조되고 있음을 시사함.

- 이슈 간 관계를 살펴보면, 가계 → 대출 → 금융 관계로 자산관리의 군집성은 강하게 나타나며, 가계 부채 정책, 금융 지원, 규제 등 통합적 이슈가 강조된 것으로 분석됨.
- 서울 → 아파트/공급/주택 관련 군집성은 특정 지역 키워드와 분리되어 있으며, 서울 중심의 공급 정책의 독립적 접근이 필요함을 시사함.
- 또한, 경제, 거래, 허가 등의 군집성은 행정/정책 처리 절차와 밀접하게 연결되어 있어, 행정 절차 자동화, 허가 및 거래 연계 시스템 고도화의 필요성이 부각됨.
- 노드 간 연결성이 낮은 외부 키워드는 특정 이슈에만 국한되어, 맞춤형 정책 모니터링 대상으로 분석됨.

- 향후 데이터가 지속적으로 축적될 경우 연관규칙 및 네트워크 분석 결과를 활용하여 핵심 키워드 간 비중(%)을 독립변수, 건설경기 수준을 종속변수로 설정하여 학습 모형으로 활용성을 확장할 수 있음.

- 본 고에서 건설경기체감지수(CoSI)는 인식 및 감성을 계량적으로 추정하는 감성 분석(Sentiment Analysis)¹⁰⁾을 활용함. 건설경기와 관련된 단어의 감성 또는 체감 수치화는 앞서 설명된 경기 체감 사전이 구축되어야 함.

- 건설경기에 대한 체감은 텍스트에 반영된 의미를 감성으로 정의하는 데 체감을 긍정적 혹은 부정적으로 확인하는 극성(Polarity) 분석 또는 긍정이나 부정적인 인식의 강도를 측정하는 감성 강도(Intensity) 분석 등으로 도출됨.

- CoSI에 적용되는 건설경기 체감 사전 구축은 수동 접근법¹¹⁾과 말뭉치기반 접근법¹²⁾

10) 감성분석은 기술적으로 특정 도메인과 관련된 단어 사전 또는 렉시콘(Lexicon)을 기반으로 수동적 분석과 지도학습(Supervised Learning) 기반 분석으로 분류됨. CoSI는 수동적 감성 분석 기법이 활용됨.

11) 수동 접근법(manual approach)은 연구자가 직접 단어를 모아 사전을 구축하는 방법으로, 특정 주제나 도메인에 대한 전문가의 경험과 지식이 중요함. 특히, 주제에 대한 감성을 포함하는 단어를 선별하고 그 감성의 극성과 강도를 단어에 부여하는 방법으로 상당한 노력이 요구되지만, 신뢰도와 정확도가 높은 사전을 구축할 수 있음.

을 활용하고, 체감을 계량화하는 데 있어 PMI(Point-wise Mutual Information)¹³⁾를 적용함.

- PMI 기반 건설경기 체감 계량화는 단어의 상대적인 관계를 나타내므로, 절대적인 체감 점수나 임계값을 설정하기에 어려움. 하지만, 데이터에서 직접 추출한 통계 정보를 활용하기 때문에 특정 영역에 특화된 사전을 구축하는 데 효율적임.¹⁴⁾

4. 건설경기체감지수(CoSI) 산정 및 분석

■ 2023년 10월부터 2025년 9월까지 건설산업과 관련된 주요 언론기사에 수록된 텍스트 수는 약 3,200만개로 특수문자를 제외한 약 2,300만개 단어를 추출함.

- 명사, 형용사, 동사 등 중심으로 토큰화(Tokenization)¹⁵⁾한 결과 약 320만개이지만, 명사 중심의 키워드 유형만 추출한 경우 1만 279개이고, 건설산업의 경기 수준과 관련된 키워드는 2,080개로 선정함.
- 토큰화된 키워드의 빈도 분석으로 CoSI에 대한 불용어를 제외함. 건설경기과 관련된 추출된 단어에 체감 점수(부정, 중립, 긍정)를 부여한 자료를 생성하고, 점수 목록을 하나의 체감 사전(Lexicon)으로 구축함.

■ 본고는 단어 단위로 구축된 사전을 활용하지만, 향후 축적되는 데이터 규모가 다양해질 경우 문장 단위의 체감 사전으로 확대하여 CoSI의 신뢰성을 제고할 수 있음.

- 각각의 단어가 경기에 미치는 예상 영향(부정, 중립, 긍정)에 따라 점수를 부여함. 긍정적으로 영향을 미칠 것으로 판단되는 경우 2점, 모호한 경우 1점, 부정적인 영향을 미칠 것으로 판단되면 0점을 부여한 사전(Lexicon)에 점수별 비중은 <그림 3>과 같음.
- <표 8>은 비중 기준 상위 30개의 긍정, 중립, 부정 단어를 제시한 결과로, 긍정 단어의 경우 '계획', '공급', '추진', '지원', '확대' 등 경기 부양과 관련된 단어이고, 부정 단어의 경우는 '문제', '발생', '규제', '우려', '사고' 등 대응이 필요한 단어로 구성됨.

12) 말뭉치 기반 접근법(corpus-based approach)은 기계학습이나 통계적 방법을 활용하여, 말뭉치 내에서 단어와 그 극성 및 강도를 자동으로 추출하는 방법으로 자료의 크기가 충분할 경우 비지도 학습을 통해 라벨링 등의 노력을 줄일 수 있으나, 감성단어를 정확하게 분류하기 위한 학습 데이터와 알고리즘 선정이 중요함.

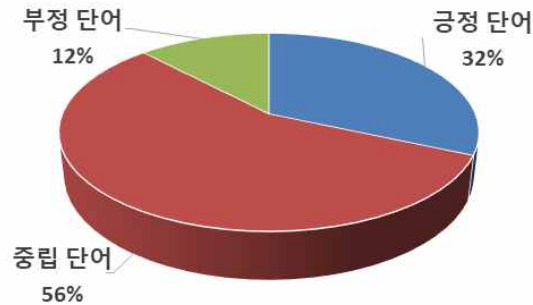
13) 두 단어의 확률적 연관성을 나타내며, 두 단어 사이의 유사성을 분석하는 데 활용되며, 두 단어의 감성 극성이 유사한 경우 동일 문서 내에서 등장할 확률이 커진다는 가정을 포함함.

14) PMI는 두 단어가 동일한 문서 내 동시에 등장하는 확률과 각 단어가 문서에 등장하는 확률을 이용하여 산정됨. w_1 과 w_2 는 연관성을 측정하고자 하는 두 단어의 확률변수로, 두 단어가 문서 내 등장할 사건이 서로 독립적이라면 0, 양의 연관성을 가지면 양수로, 음의 연관성을 가지면 음수로 나타내지만, PMI 값은 기준단어의 선택에 따라 분석 결과가 크게 변동한다는 단점이 있어 이를 보완하기 위해 SO-PMI 방법을 이용하여 각 단어에 감성을 부여함.

$$PMI(w_1, w_2) = \log \frac{p(w_1, w_2)}{p(w_1)p(w_2)}, \quad SO-PMI(w) = \sum_{pw \in PW} PMI(w, pw) - \sum_{nw \in NW} PMI(w, nw)$$

15) 토큰화는 방대한 텍스트 데이터에서 분석 가능한 작은 단위(토큰)로 분류하는 과정으로 점수가 부여된 단어가 문장에 등장하면 1, 등장하지 않으면 0으로 정의한 이진값(Boolean 데이터셋) 구축에 적용함.

〈그림 3〉 건설경기 체감 사전의 키워드 점수별 비중



〈표 8〉 건설경기 체감별 키워드(비중 기준 상위 30개) 예시

긍정						중립						부정					
순위	단어	순위	단어	순위	단어	순위	단어	순위	단어	순위	단어	순위	단어	순위	단어	순위	단어
1	계획	11	대응	21	활성화	1	정부	11	사회	21	계약	1	문제	11	제한	21	하자
2	공급	12	지속	22	개정	2	경제	12	정책	22	이용	2	발생	12	부실	22	사태
3	추진	13	효과	23	스마트	3	시장	13	전국	23	방안	3	규제	13	위험	23	손실
4	지원	14	미래	24	선도	4	종합	14	기술	24	설명	4	우려	14	재해	24	사망
5	확대	15	조정	25	기대	5	기준	15	청약	25	운영	5	사고	15	참제	25	악화
6	개선	16	안정	26	해결	6	기업	16	경기	26	주민	6	부담	16	중단	26	비상
7	검토	17	적극	27	이익	7	산업	17	전망	27	대책	7	하락	17	자연	27	불안
8	활용	18	준비	28	해소	8	공사	18	평가	28	영향	8	피해	18	부족	28	심화
9	확보	19	인하	29	보호	9	거래	19	국민	29	정비	9	사기	19	갈등	29	금지
10	완화	20	수익	30	노력	10	개발	20	공공	30	생산	10	위기	20	차별	30	위반

■ 언론기사에 포함된 텍스트 데이터를 전처리 후 구조화된 건설경기 체감 사전은 특정 시점에 긍정, 중립, 부정 단어의 빈도와 관계를 고려한 체감 수준을 지수화하는 데 활용됨. 기존 설문 기반의 경기 추정과 달리 건설산업에 참여하는 다양한 주체의 인식을 반영할 수 있는 특징이 있음.

- 기존의 건설경기를 추정하는 다양한 기법들은 설문 및 수치형 데이터 중심으로 활용되고 있음. 하지만, 텍스트 형태의 비정형데이터 비중과 활용도가 급증하면서 텍스트 추출 및 정제 기술 기반의 지수화 방법이 널리 활용되고 있음.
- 아래 식은 언론기사에 건설경기와 관련된 단어가 문장에 포함되는 빈도 및 연관규칙에 따라 수치화하는 방식을 설명한 개념으로 체감 사전 기반으로 활용됨.

$$match(t, s) = \begin{cases} 1 & : \text{단어 't'가 문장 's'에 포함된 경우} \\ 0 & : \text{그 외의 경우} \end{cases}$$

$$sentence - score(s) = \frac{\sum_t (match(t, s) \times word - score(t))}{\sum_t match(t, s)}$$

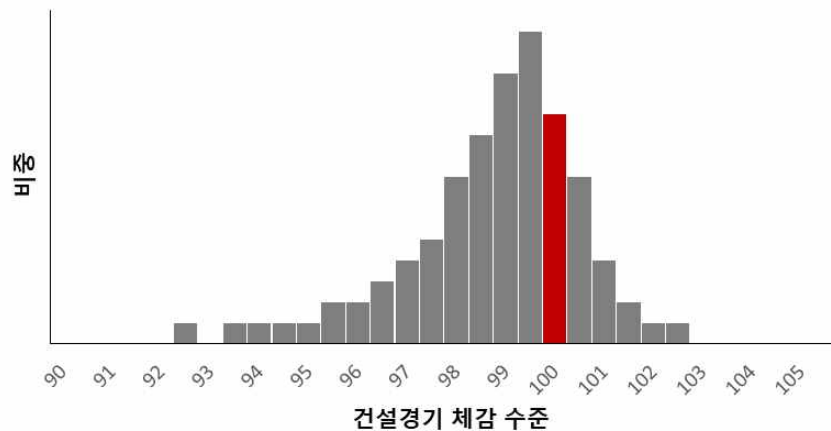
$$document - score(d) = \frac{\sum_{s \in d} sentence - score(s)}{S(d)}$$

$S(d)$: 특정 시점의 언론기사에 등장하는 문장의 수

❖ <그림 4>를 살펴보면, 2023년 10월부터 주간별 생성된 언론 기사를 활용한 CoSI 값의 분포를 제시한 결과임.

- 주간별 건설경기체감지수(CoSI)는 중립적인 값 100 미만에서 반복적인 등락을 보여, 건설경기가 여전히 전반적으로 낮은 것으로 분석됨.
- 하지만, 부동산, 건설기술 정책 중심으로 건설경기가 100 이상을 유지하는 시점도 조사되어 특정 시점에 키워드 간 연관규칙과 네트워크 분석으로 입체적인 해석이 가능함.

<그림 4> 2023년 10월부터 2025년 9월까지 주간별 건설경기체감지수(CoSI) 분포

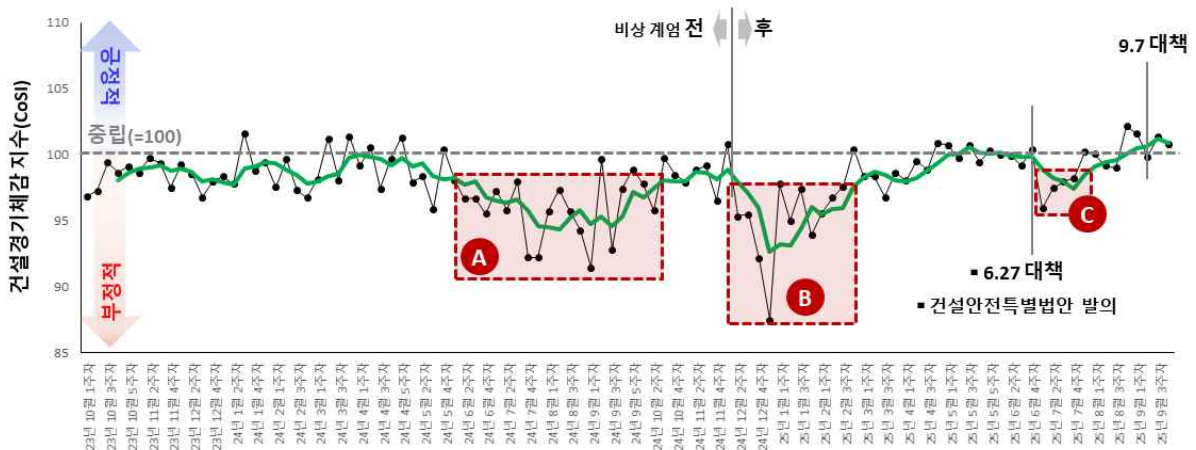


❖ <그림 5>는 구축된 주간별 CoSI를 시계열적 추이로 제시한 결과로 시기별 건설경기 수준과 주요 정책에 따라 등락이 발생하는 것으로 분석됨.

- 지난 2년간 평균 주간별 건설경기체감지수는 100(또는 '중립=1') 미만 수준(약 98.7)으로 다소 부정적인 것으로 해석됨.

- 특히, 2024년 6월부터 9월('A' 시점), 2024년 1월부터 2025년 2월('B' 시점), 2025년 7월('C' 시점)은 CoSI 값이 두드러지게 낮은 시점으로 해당 시점별 키워드 연관규칙과 네트워크 분석으로 입체적이고 구체적인 분석이 가능함.
- <그림 5>와 같이 CoSI는 변동성이 큰 특정 시점에 대해 독립적으로 이슈 및 관계 분석을 가능케 하고, 이슈로 인한 파급효과를 파악하는 데 용이한 특징이 있음.

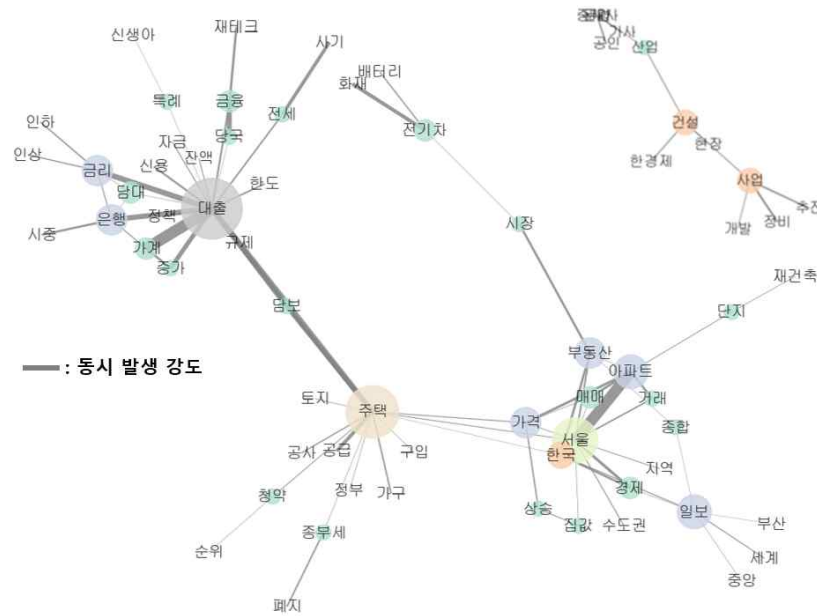
<그림 5> 2023년 10월부터 2025년 9월까지 주간별 CoSI 추이(이동 평균 포함)



❖ <그림 6>은 <그림 5>에서의 'A' 시기 건설경기체감지수가 하락한 구간에 대한 네트워크 분석 결과로, 대표적으로 부동산 금융 리스크가 증가한 것으로 분석됨. 또한, 가계 대출, 대출 금리 등 가계 경제에 대한 부담이 증가하여 정책 변화가 건설산업에 미치는 영향이 큰 시점으로 해석됨.

- 건설산업 경기 관점에서 부동산 대출 및 금융에 대한 우려가 확산하여, 건설경기 전반적으로 영향을 미쳐 산업 내 참여 주체의 금융 불안정이 확대된 상황으로 분석됨.
- 정책적으로 가계 부채 관리, 금융 안정 마련의 필요성이 강조되고, 수도권 주택에 편중되는 경향에 대응 방안의 필요성과 스마트 건설기술 등 신기술의 연계성과 확장 가능성이 강조됨.

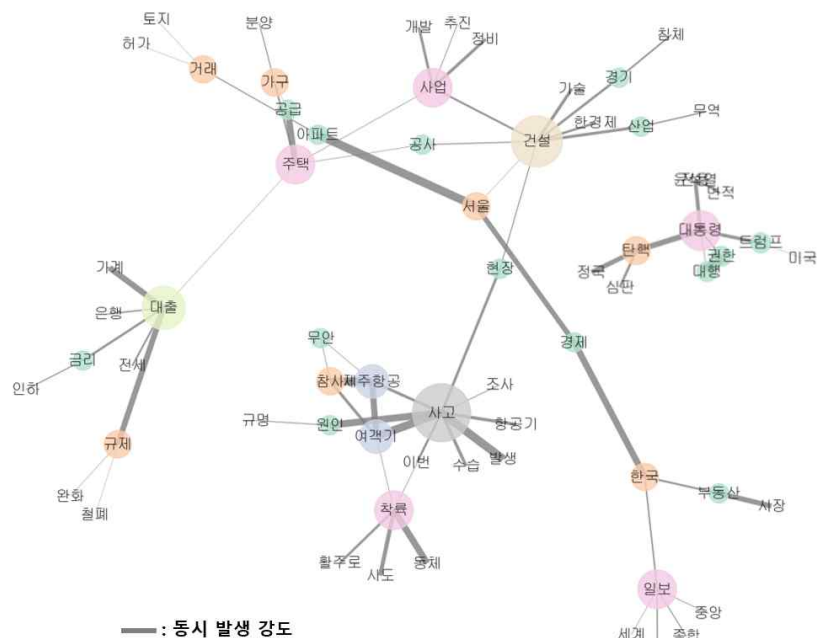
〈그림 6〉 'A' 시기 키워드 간 네트워크 결과



〈표 9〉 중심성 기준 상위 10개 키워드(노드)

순위	키워드	중심성 (Centrality)
1	대출	0.1809
2	주택	0.1383
3	서울	0.1064
4	아파트	0.0638
5	금리	0.0638
6	가격	0.0532
7	부동산	0.0532
8	건설	0.0532
9	가계	0.0532
10	금융	0.0426

〈그림 7〉 'B' 시기 키워드 간 네트워크 결과



〈표 10〉 중심성 기준 상위 10개 키워드(노드)

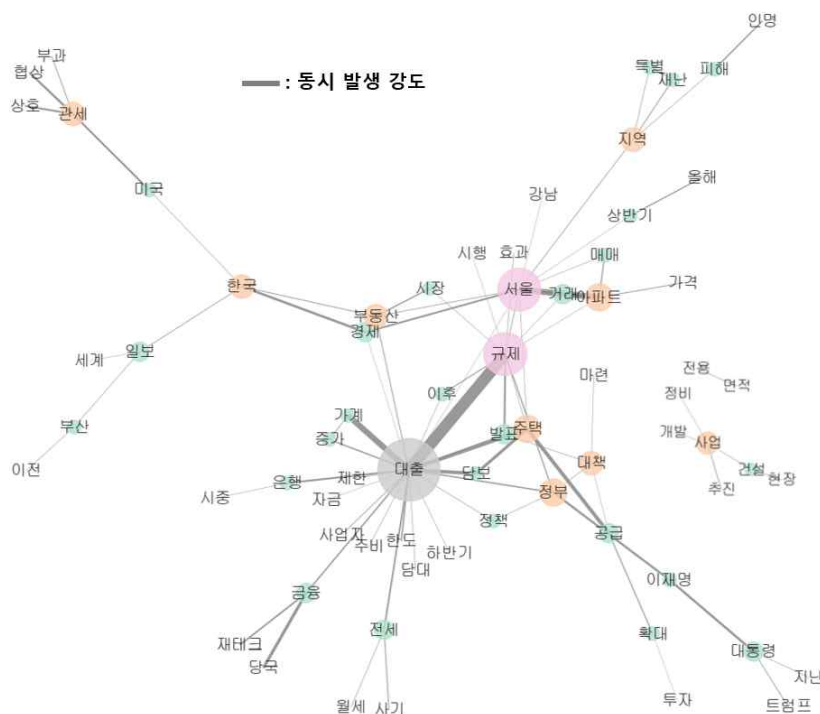
순위	키워드	중심성 (Centrality)
1	사고	0.1031
2	건설	0.0825
3	대출	0.0619
4	착륙	0.0515
5	주택	0.0515
6	대통령	0.0515
7	사업	0.0515
8	여객기	0.0412
9	제주항공	0.0412
10	참사	0.0309

■ ‘B’ 시기 CoSI 하락은 국내외 정치적 이슈, 가계 부채 증가, 안전사고 리스크 등에 의한 영향을 크게 받은 것으로 분석됨.

- 가계 부채, 금리 규제 등으로 주택시장의 불안정성이 증가하여 개발 허가, 대출 규제 등의 제도적 변화에 의한 하락 효과가 발생한 것으로 분석됨.
- 또한, 국내외의 정치적 이슈로 건설산업 정책의 일관성에 대한 우려가 증가하고, 항공 사고 등의 대형 안전사고로 인한 사회적 불안감 증가도 건설경기 하락의 직간접적 원인으로 분석됨.

■ 특히, 항공 사고 등 안전 및 재난 이슈가 두드러졌고, 이 시기에 부동산 개발 및 정비 사업과 공급 정책에 대한 대응을 위해 시장의 불안정성 완화와 가계대출의 효율적인 규제의 필요성이 요구된 것으로 분석됨(〈그림 7〉 참조).

〈그림 8〉 ‘C’ 시기 키워드 간 네트워크 결과



〈표 11〉 중심성 기준 상위 10개 키워드(노드)

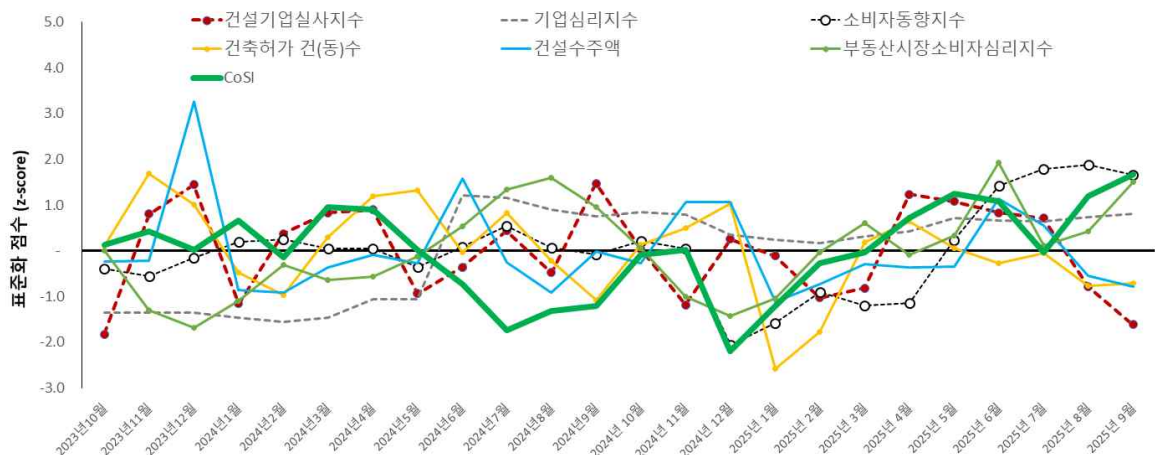
순위	키워드	중심성 (Centrality)
1	대출	0.2444
2	규제	0.1222
3	서울	0.1222
4	아파트	0.0556
5	주택	0.0556
6	정부	0.0556
7	한국	0.0444
8	관세	0.0444
9	부동산	0.0444
10	지역	0.0444

■ ‘C’ 시기에 건설경기체감지수의 일시적인 하락은 국내의 정책적/정치적 금융 규제, 서울 중심의 주택 매매가격 변동성 등에 영향을 받은 것으로 분석됨.

- 특히, ‘대출’과 ‘규제’ 간 키워드의 강한 연결성은 제도 변화에 대한 이슈가 부각되고, 글로벌 경제 환경(무역 관세 등) 변화, 건설산업 신기술(시 인프라 등) 추이 급변 등이 CoSI 하락의 직간접적 원인으로 해석됨.

- 가계부채 관리, 전세대출 제도개선, 지방 도시 주거 정책 및 균형발전 전략, 주택시장에서 사회안전망과 금융/주거 정책의 연결성 유지 등으로 건설산업의 건전성 향상을 위한 대책 마련이 강조됨(〈그림 8〉 참조).

〈그림 9〉 CoSI와 주요 경기전망 지수 간 표준화(z-score) 점수 추이 비교



- 〈그림 9〉는 CoSI와 주요 경기 전망 지수들과의 추이를 비교한 결과로, 이는 향후 건설기업과 소비자의 심리를 아우르는 통합적 척도로 활용할 가능성을 제시함.

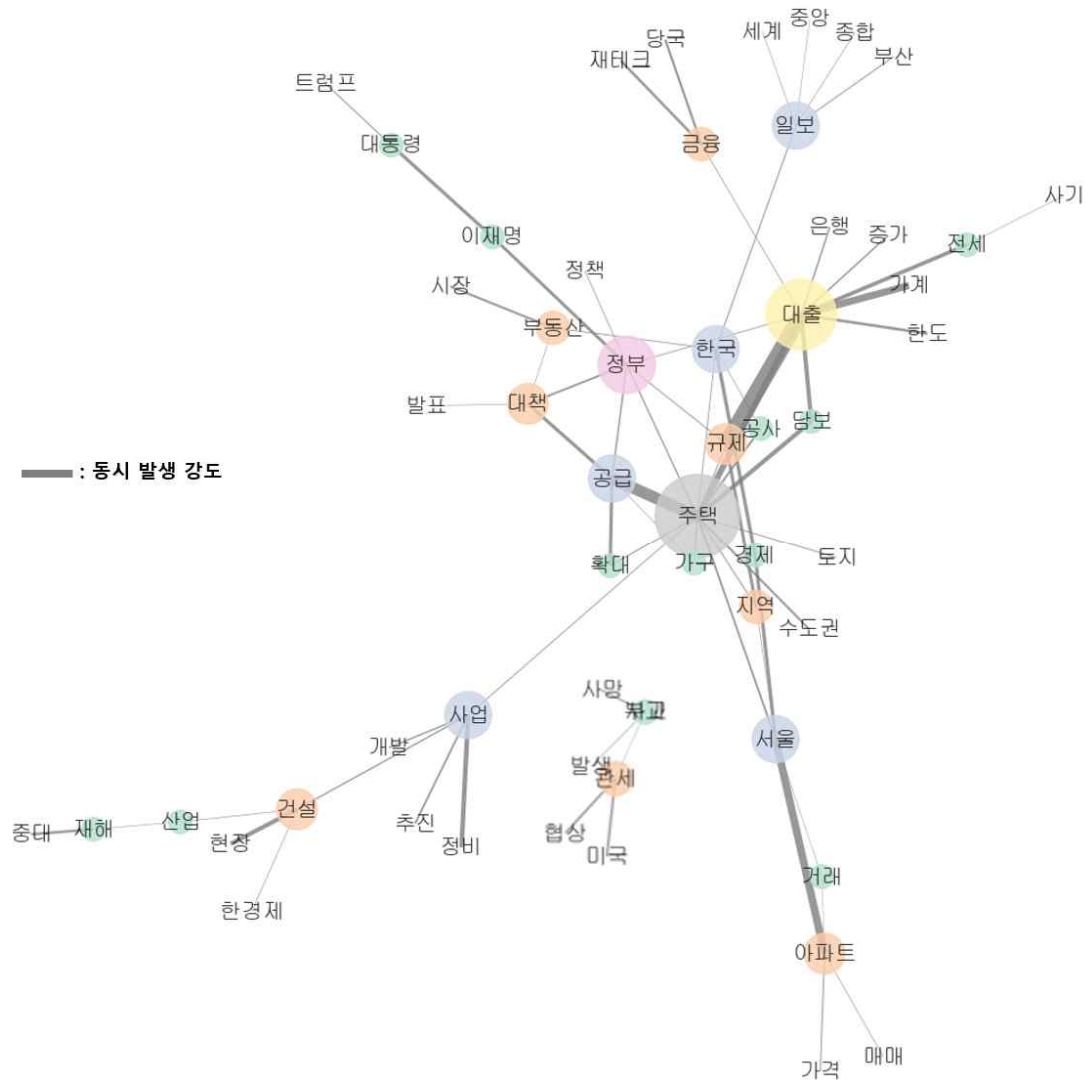
- CoSI와 CBSI 추이는 조사 대상에 차이가 있어 기업의 기대치와 건설산업의 다양한 참여 주체의 체감 수준을 상호보완적으로 활용할 수 있음.
- CoSI는 건축허가, 건설수주액 추이의 선행적 신호를 제공하고, 부동산시장 소비자심리지수와 민감하게 반응하고 있음.

- CoSI는 다양한 주체들의 의견을 반영하는 언론 기사를 활용하기 때문에 건설기업과 소비자의 체감을 포괄적으로 설명하는 데 보완적인 지수임. 건설기업 및 금융기관은 CoSI 변동 추이의 긍정적 또는 부정적 신호를 활용해 리스크관리와 투자 시점에 대한 기초자료로 활용할 수 있음.

- CoSI는 정책 및 규제 변화, 금리 변동, 경기침체 등에 민감하게 반응하여 국가 및 산업 차원의 제도 변화가 건설경기에 미치는 영향을 설명하는 데 효율적이고, 정책 효과(규제 및 금리 변화 등)를 점검하는 데 활용할 수 있음.
- 또한, 건설기업과 소비자의 심리적 변동을 반영하고 있어 건설시장 안정화 정책 수립의 기초자료로 활용되어 경기 변화를 설명하는 선행 지수로서의 가능성을 확인할 수 있음.

- 건설기업은 CoSI 하락기에 개발 및 투자 자제에 대한 조기경보를 제공하고, 금융기관은 대출 정책 조정의 위험관리에 활용할 수 있으며, 정부는 정책 발표 시점의 적시성 확보를 기대할 수 있음.

〈그림 10〉 2025년 3분기(7~9월) 주요 키워드 네트워크 결과



〈표 12〉 2025년 3분기(7~9월) 중심성 기준 상위 10개 키워드(노드)

순위	키워드	중심성 (Centrality)
1	주택	0.1556
2	대출	0.1111
3	정부	0.0778
4	공급	0.0556
5	서울	0.0556
6	사업	0.0556
7	규제	0.0444
8	아파트	0.0444
9	대책	0.0444
10	금융	0.0333

❖ 〈그림 10〉, 〈표 12〉는 ‘25년 3분기(7~9월) 건설산업의 주요 키워드 간 네트워크 분석 결과임. 이를 살펴보면, ‘주택’ 문제는 단순 공급이 아니라 ‘대출’과 같은 건설 관련 ‘금융’, ‘정책’에 전반적인 강한 연관성을 제시하고, 건설경기에 큰 영향을 미치는 것으로 분석됨.

- 특히, ‘정부 → 금리 → 규제 → 대출 → 주택시장’으로 연결된 구조로 정부 정책 변화가 대출 규제, 금리 변동과 연관성이 크고, 이는 주택 매매 심리와 건설경기에 영향을 미치는 구조임.
- 수도권 집중 현상이 심화되면서 정부의 규제 및 정책 변화는 시장 전반에 큰 파급효과를 미치는 것으로 해석됨.
- 건설/사업/재개발/정비/재해, 사고/사망/관세 등의 연결 구조는 규제·인허가, 원자재 가격 변동 등의 이슈로 건설기업의 비용 증가와 연관성을 제시함.
- 건설경기와 밀접하게 연관된 주택시장의 대출, 금리 등의 변동성 증가가 가속되어 주택, 금융, 건설, 정책 등 다차원적인 구조 속에서 상호 연결되어 있고, 서울 중심의 불균형성과 금융 리스크에 대한 현안이 부각된 것으로 분석됨.

❖ 〈표 13〉에서 제시된 키워드 간 연관규칙 분석 결과를 살펴보면, 미래 건설 스마트 기술 부각, 부동산 금융의 리스크 헷지를 위한 구조화, 중대 재해 및 전세 사기 등의 산업의 부정적 이미지 증가 등이 강조됨. 또한, 국제 무역 환경 변화에 의한 국내 주택 공급 대책 등의 구체적인 이슈가 건설경기에 영향을 미친 것으로 분석됨.

- 또한, 인공지능, 자율주행 등 신기술 발전이 본격화되면서 건설산업의 정책과 활용 환경에 대한 이슈가 강조됨. 중대 재해, 안전사고 등 건설산업의 지속적인 개선 사항이 사회적 현안으로 부각됨.
- 건설산업에서 큰 비중을 차지하고 있는 주택시장의 금융 규제, 대출 정책, 전세 사기 등의 현안이 사회적 문제로 우려되고, 국제 무역의 환경 변화에 반응하여 건설경기 위축을 일으킨 것으로 분석됨.

〈표 13〉 2025년 3분기(7~9월) 키워드 간 상위 20개 연관규칙 결과

규칙 번호	선행 사건	후행 사건	지지도	신뢰도	향상도	규칙 번호	선행 사건	후행 사건	지지도	신뢰도	향상도
1	인공	지능	0.0014	0.85	491.4	16	한강	버스	0.0008	0.40	133.6
2	가계	대출	0.0030	0.84	37.8	17	규제	대출	0.0059	0.38	17.1
3	중대	재해	0.0015	0.77	229.3	18	부과	관세	0.0008	0.36	41.3
4	담보	주택	0.0021	0.64	25.6	19	재개발	재건축	0.0009	0.32	83.7
5	담보	대출	0.0020	0.62	28.2	20	전세	대출	0.0019	0.31	14.1
6	자율	주행	0.0017	0.61	219.4	21	사망	사고	0.0012	0.31	42.3
7	사기	전세	0.0008	0.60	97.4	22	매매	아파트	0.0009	0.30	22.7
8	한도	대출	0.0016	0.51	23.2	23	협상	관세	0.0013	0.30	33.9
9	당국	금융	0.0014	0.46	47.8	24	공급	주택	0.0045	0.28	11.2
10	정비	사업	0.0022	0.42	19.3	25	데이터	센터	0.0008	0.27	69.6

IV 시사점 및 CoSI 활용방안

- 건설산업은 국가와 국민 경제에 밀접한 산업이지만, 기존 생태계는 급변하는 환경에 대한 대응력의 한계에 직면하고 있음. 또한 전(全) 산업의 디지털 전환이 가속되면서 건설산업 생태계의 재탄생(Rebirth)은 선택이 아닌 필수로 인식되고 있음.
- 건설산업 재탄생 전략의 핵심 기술로 강조되는 빅데이터는 급변하는 환경에 대한 선제적 알람(Alarm) 기능 제공을 기대할 수 있음. 이는 기업과 산업 차원의 대응 방향을 제시하는 도구로 중요성이 증가하고 있음.

 - 디지털 전환과 AI 기술 확산은 빅데이터 기반 의사결정의 중요성을 비약적으로 확대시키고 있음. 특히, 건설산업은 경기 변동과 정책 변화에 민감하게 반응하는 특성이 있으며, 이러한 불확실성을 신속히 파악하고 대응하는 체계가 요구됨.
- 전(全) 산업에 걸쳐 기업실사지수(BSI), 소비자심리지수(CSI) 등을 활용한 경기 추정 지수들은 특정 시점에서 경기 체감에 대한 비교가 가능함. 하지만, 국민과 산업에 참여하는 주체들의 포괄적인 인식과 체감을 충분히 반영하는 데 제한적임.

 - 특히 건설산업에 특화된 경기는 정책, 사회적 기대감, 현장 심리 등에 크게 영향을 받기 때문에 기존의 계량적 지표와 함께 CoSI는 건설경기 해석을 지원할 수 있음.
 - 건설경기를 설명하는 CBSI와 건설수주액, 건축허가 및 착공면적, 건설기성액, 건설업 취업자 수 등에 언론 빅데이터에 포함된 다양한 의견을 반영할 경우 산업 전반에 걸친 경기 체감을 추정하는 데 효과적임.
- 본 고에서 제시된 건설경기체감지수(CoSI)는 국민이 건설산업에서 체감하는 경기 수준을 언론 기사를 활용하여 제공함. 기존 설문 중심의 건설기업실사지수(CBSI)와 상호보완적으로 활용될 경우 정책 방향과 기업의 입체적 전략 수립에 효율적임.

 - CoSI는 건설산업의 다양한 참여 주체들이 제공하는 언론기사 중심의 빅데이터를 활용하여 특정 시점의 체감 수준과 장기적인 추이 예측에 나침반 역할을 제공할 수 있음.
 - CoSI 개발 및 활용은 건설경기에 대한 해석의 패러다임 변화, 건설 금융 및 정책 리스크의 선제적 포착, 사회적 이슈와 건설경기의 연계성 강화, 기술 혁신과 변화의 적시성 강화 등 기존 지수와의 차별성을 제시함.

- 또한, 건설경기체감지수(CoSI)는 언론기사의 방대한 비정형데이터를 구조화 및 계량화하여 건설산업의 체감경기를 측정하는 보조적인 도구임. 이는 기존 CBSI, BSI, CSI 등 설문 중심의 지수와 상호보완적인 도구로 활용성이 기대됨.

 - CoSI는 언론기사라는 대규모 비정형데이터를 활용하여 특정 주체에 국한되지 않고 사회 전반의 ‘건설경기’에 대한 인식을 반영함.
 - 또한, 경기 흐름의 구조적 패턴 파악을 지원하고 기존의 계량적 지수(CBSI, 건설수주액, 건축허가 건수 등)와 상호보완적으로 활용될 수 있음.
- 건설산업에 맞춤형하여 주요 정책 및 제도 개선에 따른 경기 변화를 표준화된 기준으로 계량화할 수 있고 국가와 국민 경제에 미치는 영향 분석을 지원함.
- CoSI는 건설산업 경기 추정을 보강하는 도구로, 설문 기반 한계를 넘어 사회 전반의 체감 또는 인식을 반영하는 계량적 지수라는 점에서 의미가 큼.

 - 본 고에서 제시된 CoSI는 주별, 월별, 분기별 등 맞춤형된 주기로 결과를 도출할 수 있고, 건설경기의 변동성이 기간을 대상으로 독립적으로 분석이 가능한 특징이 있음. 이는 기존의 건설경기 관련 지수와 차별화됨.
 - 한국언론진흥재단의 BIG KINDS 서비스는 이슈 도출 및 관계 분석에 집중되어 있어, 경기 체감의 지수화에 대한 이론적 또는 체계적 모형 제공에 미흡한 한계가 있음.
- 특히 금융·주택·정책 리스크와의 연계성, 사회적 이슈에 시의적절한 대응 등 건설기업, 금융기관, 정부 차원의 실질적 활용 가능성에 긍정적인 시사점을 제공함.

 - CoSI는 단순한 경기지수를 넘어, 산업의 정책 수립 시점을 지원하고, 나아가 디지털 전환 시대에 데이터 기반 의사결정을 효율적으로 지원하는 도구로 발전될 수 있음.
 - 건설기업 차원에서 CoSI 변동을 기반으로 투자·개발 시점, 경기 하락기에 위험관리 및 의사결정 지원, 기술 키워드(스마트 건설, AI 등) 추적으로 신성장 영역 발굴 등 지원
 - 금융기관 차원에서 CoSI를 가계부채·대출 건전성 지표와 연계, 금리 정책, 규제 강화 등과 연동된 수요 변화 감지, 건설 금융 시장의 위험관리 지표 등과 연계
 - 정부 및 정책 당국 차원에서는 정책 발표 시기와 내용의 적시성 제고, CBSI 및 BSI 등과 병행한 정책 효과 검토 자료로 활용, 안전사고 등 구조적 문제 식별 지원 등 맞춤형 정책 수립에 활용할 수 있음.

유위성(연구위원·wsyoo@cerik.re.kr)