



RAG-LLM(Retrieval Augmented Generation- Large Language Model) 기반 부동산 뉴스 분석을 통한 아파트 매매가격지수 예측 모델 개발

Predicting Apartment Price Indices through Real Estate News Analysis: A Retrieval Augmented Generation-Large Language Model (RAG-LLM) Approach

김진우* · 은준엽**

Jinwoo Kim · Joonyup Eun

Abstract

This study explores the effectiveness of retrieval augmented generation (RAG) with large language models (LLMs) in processing unstructured data in real estate. By connecting pretrained LLMs with external knowledge sources, the RAG methodology reduces hallucinations while generating reliable domain-specific responses. We construct an RAG-LLM model using real estate news articles to predict directional movements in Seoul apartment price indices. Analysis spanning 2015-2024 reveals that RAG-LLM predictions serve as significant leading indicators for apartment price changes, achieving 85% accuracy and 93% area under the curve - receiver operating characteristic (AUC-ROC). These RAG-LLM predictions outperform those based on the autoregressive integrated moving average (ARIMA) model by more than 11%. Employing chain-of-thought prompts with RAG further enhances performance, with improved metrics observed when evaluated using ARIMA with exogenous variables. Our findings suggest that the RAG-LLM approach represents a significant advancement in real estate analysis by effectively leveraging unstructured data and overcoming the limitations of traditional sentiment analysis through complex contextual understanding. The RAG-LLM approach also enables near real-time forecasting, significantly boosting its practical utility in real-world applications.

Keywords: Big data, Housing price forecasting, Retrieval augmented generation (RAG), Real estate news

* 고려대학교 기술경영전문대학원 박사과정(주저자) | Ph.D. Candidate, Graduate School of Management of Technology, Korea University | First Author | jinwoo6627@korea.ac.kr |

** 고려대학교 기술경영전문대학원 부교수(교신저자) | Associate Professor, Graduate School of Management of Technology, Korea University | Corresponding Author | jeun@korea.ac.kr |

1. 서론

우리나라의 부동산 가격 변동은 전 국민적인 관심사이자 주요한 경제적, 사회적 문제이다. 2023년 기준 부동산은 한국 가계 총자산의 79.7%를 차지할 정도로 큰 비중을 차지하고 있으며(신한은행, 2024), 부동산 가격의 변동은 가계의 재정적 관심 이외에도 정책 및 금융과 같은 거시적 요인에 영향을 미치는 요소이다(Mian and Sufi, 2014). 최근에는 부동산 가격의 변화를 예측하는 방안으로 뉴스 기사와 같은 미디어에서 나오는 비정형 데이터를 분석하는 접근 방식이 제안되고 있다.

부동산은 수요자들의 심리 변화가 가격 형성에 중요한 요소로 작용한다(김대원·유정석, 2013). 특히 주거용 부동산은 비전문적인 시장 참여자들이 많은 시장으로, 수요자들의 심리와 직관에 영향을 받는 것으로 알려져 있다(진창하, 2012). 미디어에서 부정적인 내용이나 어조로 보도가 이루어질 경우, 소비자들의 경제 상황에 대한 인식 역시 부정적인 영향을 받게 된다(이완수, 2007). 미디어의 영향을 분석한 부동산 연구에서는 뉴스 기사나 방송의 비정형 데이터의 감성을 측정하여 아파트 가격을 예측하고자 하는 시도가 존재하였으며(경정의·이국철, 2016; 박재수·이재수, 2019, 2021; Soo, 2018), 감성분석과 같은 방법론으로 비정형 데이터가 부동산 가격의 예측에 기여할 수 있음을 보였다.

최근 AI 기술의 발전과 함께 OpenAI사의 GPT와 같은 사전 훈련된 대규모 언어모델(large language model, LLM)이 상용으로 공개되고 있다. LLM은 감성분석과 같은 비정형 데이터 처

리에서 기존의 방법론 대비 우수한 성능을 보이고 있으며, 개체명 인식, 맥락 파악 및 추론 등 복잡한 작업을 수행할 수 있다(Brown et al., 2020). LLM에 입력하는 텍스트를 설계하는 프롬프트 엔지니어링(prompt engineering)을 통해서도 고도화된 추론과 같은 LLM의 역량을 더 끌어내 높은 품질의 답변을 생성할 수 있다(White et al., 2023). 최근 연구에서는 LLM인 GPT-4가 텍스트 이해 및 일반화나 추론 작업에서 대학생 평균보다 우수한 성능을 보이기도 하였다(Shultz et al., 2025).

다만 LLM은 특정 시점까지의 일반적인 도메인에 대해 훈련된 모델이기에 최신 데이터나 특정한 지식 도메인이 요구되는 분야에서는 성능의 저하가 관찰되며, 환각 현상(hallucination) 같이 그럴싸한 오답을 작성하는 등의 치명적인 단점이 존재한다(Gao et al., 2023). 특히, 부동산 분야와 같이 전문적인 지식이 요구되며 학습에 활용할 텍스트 데이터가 희소한 분야에서는 LLM의 효용성이 상대적으로 낮게 나타난다(Chen and Si, 2024; Gloria et al., 2024). 지식 집약적이고 답변에 명확한 근거가 필요한 분야에서 LLM이 보이는 한계를 극복하기 위하여 프롬프트에 외부 지식을 주입하는 방법론이 제안되었으며, 전문적인 지식의 제공을 통해 LLM의 환각 현상을 줄이고 생성된 답변의 신뢰도를 높일 수 있음을 보였다(Martino et al., 2023).

Lewis et al.(2020)은 지식 집약적인 분야에 검색증강생성(retrieval augmented generation, RAG)이라는 LLM과 외부의 지식 데이터베이스 검색 시스템을 결합하는 방식을 제안했다. RAG

는 LLM으로 답변을 생성하기 전 신뢰성 높은 외부 지식 데이터베이스로부터 검색(retrieval)을 통해 문제 해결에 필요한 문서와 근거를 확보하여 LLM의 역량을 증강(augmented)하여 답변을 생성(generation)함으로써 출처가 확실하고 우수한 품질의 답변을 얻을 수 있다. Gao et al. (2023)은 RAG-LLM이 검색된 정보에 근거하여 답변을 작성하고 사전훈련 때 사용한 상식의 사용을 제한함으로써, 환각 현상의 발생을 줄이고 생성된 답변의 신뢰성과 투명성을 높임을 보였다.

최근 부동산 분야에서도 LLM을 활용한 연구가 제시되고 있다. RAG-LLM을 활용하여 부동산 계약을 분석하여 유용한 정보를 추출하는 사례나(Yang et al., 2024), 부동산 가격의 예측에 LLM의 감성분석 능력을 활용하거나(김수아 외, 2023), 중국 임대주택 유사 거래 사례를 LLM이 분석하여 임대가격을 예측한 사례(Chen and Si, 2024)가 제시되는 등 LLM의 고도화된 추론 역량을 부동산 분야에도 적용하고자 하는 시도가 이어지고 있다. 다만 설명력 높고 실시간으로 접근할 수 있는 데이터인 부동산 뉴스 기사를 원천으로 하여 RAG 시스템을 구축하고, 프롬프트 엔지니어링을 활용하여 LLM을 미래 주택 가격의 예측에 활용한 사례는 전무하다.

본 연구의 목적은 주요 경제신문의 부동산 뉴스를 기반으로 RAG 시스템을 구축하고, 이를 LLM과 결합하여 미래 부동산 가격에 대한 유의미한 예측이 가능한지를 검증하는 데 있다. 뉴스 기반의 RAG-LLM을 통해 월별로 미래 부동산 가격에 대한 전망을 생성하고, 생성된 전망점수의 가격 예측의 선행성과 예측력을 분석하고자 한

다. 이를 통해 부동산 분야의 비정형 데이터를 효과적으로 처리하기 위해 LLM을 활용하고, 다음 달 전망을 신속하게 생성함으로써 예측의 시간적인 유용성을 높이는 방안을 제안한다.

본 연구에서는 2015년부터 2024년까지 수집한 부동산 뉴스 기사를 바탕으로 RAG에 활용할 수 있는 월별 데이터베이스를 구축하였으며, RAG-LLM을 통해 매월 서울 아파트의 매매가격에 대한 전망을 생성하였다. 한국부동산원의 아파트 매매가격지수를 기준으로 하여 전망점수의 유용성을 분석하였고, 생성된 전망점수는 매매가격지수 변화 대비 1~4개월 정도 유의미한 선행성을 나타내고 장기 시계열 예측에도 기여함을 보였다. 또한 RAG-LLM의 전망점수를 방향성 예측에 활용하면 데이터의 시차를 고려한 ARIMA(autoregressive integrated moving average) 시계열 모형보다 높은 정확도와 우수한 분류 성능을 확인하였다. 공표 및 산출에 한 달 이상의 시차가 존재하는 거시변수 및 매매가격지수를 활용하는 기존 방법에 비해, 제안한 방법이 시간적 제약을 극복하여 유용한 전망을 도출할 수 있음을 확인하였다.

II. 선행연구

1. 뉴스 및 미디어를 활용한 부동산 연구

뉴스와 같은 미디어는 부동산 시장 참여자들의 심리와 현실 인식에 영향을 주고(이완수, 2007), 심리적 태도의 변화가 주택의 매매에 영향을 주게 된다(김대원·유정석, 2013).

특히, 뉴스 기사에 제시된 감성이 부동산 가격의 변동과 유의미한 영향이 있다는 연구가 다수 제안되었다. 경정익·이국철(2016)은 부동산 기사의 감성이 아파트 실거래 가격과 높은 상관성을 보이며, 심리지수 및 거래량과도 유의미한 상관관계가 있음을 보였다. 박재수·이재수(2019)는 서울 권역의 아파트에서 뉴스의 감성과 가격 움직임 사이에 상관관계가 있으며, 온라인 신문 기사와 관련된 긍정적인 감성이 매매가격지수와 유의미한 인과관계가 있으며, 일부 지역에서는 감성지수가 선행함을 보였다. 서정석 외(2021)는 주택 뉴스 기사의 감성과 소비자 심리 사이에 시계열 관계가 있고, 일부 주제에서 투자심리가 주택 가격에 그레인저 인과관계가 있음을 확인하였다.

뉴스 기사에 대한 감성분석을 통해 작성된 감성지수는 매매가격 대비 선행성이 관찰되었다. 본 연구에서 핵심적으로 참고한 박재수·이재수(2021)는 뉴스 기사 기반의 심리지수가 아파트 매매가격에 대해 선행성을 지니며, 심리지수를 시계열 모형의 외생변수로 활용한 ARIMAX (autoregressive integrated moving average with exogenous variable) 모형이 ARIMA 모형 대비 RMSE(root mean square error), MAE (mean absolute error) 오차를 각각 7.9%, 6.2% 개선할 수 있음을 보였다.

미디어 데이터는 일반적으로 즉시성이 있어, 집계 후 발표 시점까지 한 달 이상의 시차가 존재하는 거시경제 데이터, 심리지수 혹은 가격지수에 비해 더 빠르게 데이터의 입수와 활용이 가능하다. 별도의 거시변수나 시차가 존재하는 데이터를 활용하지 않고 뉴스 기사 기반의 접근을 활

용하면, 데이터의 시차를 최소화하고 실시간에 가깝게 예측을 생성하여 투자 및 정책 의사결정 및 참고 자료로서의 유용성을 극대화할 수 있다.

뉴스 기사에 대한 감성분석 방법론은 대체로 TF-IDF(term frequency-inverse document frequency)와 같은 빈도 기반의 접근이 대부분이었으나, 자연어처리 기술의 발달과 함께 문장 전체나 기사를 딥러닝 모델을 통해 처리하는 방안이 제시되었다. 박재수·이재수(2021)와 주종용·권영상(2023)에서는 언어모델로 BERT(bidirectional encoder representations from transformers)를 사용하여 문장 단위의 감성을 분석하였고, 김수아 외(2023)에서는 GPT와 같은 LLM을 통해 요약된 기사의 긍정/부정 감성을 분류하는 방법이 제시되었다.

최근 제시된 연구에서는 부동산 관련 뉴스의 전체 텍스트에 대한 감성분석보다는 세부적인 맥락의 중요성에 주목하고 있다. 홍지수(2023)는 기사의 주제에 따라 부동산의 비기초가격에 다른 방향의 영향을 미친다는 점을 제시하였다. 정책과 관련된 기사의 빈도는 가격에 부정적인 영향을 미친 반면, 광고와 관련된 기사는 가격에 긍정적인 영향을 미치는 차이를 보였다. Zhu et al. (2022)은 중국 주택 시장의 소셜 미디어를 대상으로 딥러닝 방법론을 통해 텍스트로부터 과거의 심리지수와 미래 심리지수를 분리하였다. 분리된 미래 심리지수는 주택 가격의 변화에 통계적으로 유의하며, +1 표준편차 증가할 때마다 다음 달 주택 가격이 +0.2 표준편차(0.09%)만큼 상승하였다. 이와 같은 결과는 비정형 데이터를 부동산 분석에 활용할 때, 미래 가격의 움직임에 영향을 미

칠 수 있는 부분에 집중함으로써 비정형 데이터의 유용성을 극대화할 수 있음을 시사한다.

2. LLM(Large Language Model) 및 RAG (Retrieval Augmented Generation) 연구

뉴스 기사와 같은 텍스트 기반의 비정형 데이터 처리에 있어서 LLM은 기존의 언어모델들 대비 우수한 성능을 보인다(Brown et al., 2020). 최근 발표되는 상용 LLM의 성능은 많은 벤치마크에서 인간을 능가하는 성능을 보여주고 있으며 (Shultz et al., 2025), 복잡한 맥락을 고려하여 답변을 작성하거나 단계별로 추론을 수행하는 등 고도화된 지식 작업에 활용될 가능성을 보여주고 있다.

LLM을 효과적으로 활용하기 위해서는 적절한 프롬프트 엔지니어링이 같이 진행되어야 한다. 프롬프트는 LLM이 답변하는 출력과 상호작용을 정의할 수 있는 프로그래밍 형태로서 보통 LLM에 입력하는 텍스트의 형태로 구성된다. White et al.(2023)은 LLM을 효과적으로 다룰 수 있는 프롬프트 구축을 위한 프레임워크를 제안하였는데, 작동 방식 및 제약조건과 같은 명확한 지시사항의 제공, 역할 부여, 출력의 규격화, 그리고 적절한 맥락을 제공하는 것이 답변의 품질 향상에 중요함을 밝혔다.

Wei et al.(2022)은 LLM의 답변 생성 전략에 추론 과정을 포함하도록 지정하는 방법인 Chain-of-Thought 프롬프트 엔지니어링을 통해 더 좋은 품질의 답변을 생성하는 방법을 제시하였다. 사람이 어려운 문제를 작게 나누어 풀 듯 프롬프

트를 통해 답변을 사고 단계별로 제시하도록 지시하여 LLM의 추론 역량을 활용하였다. Chain-of-Thought 프롬프트는 상식 추론 작업에서 LLM의 정답률을 69.4%에서 75.6%로 크게 끌어올렸으며, 스포츠 분야 문제의 경우 80.5%의 정답률을 95.4%까지 끌어올리는 결과가 보고되었다.

상용 LLM 및 오픈소스 LLM의 대부분은 일반적인 텍스트에 대해 훈련된 모델이기에, 최신 데이터나 전문적인 지식 도메인이 요구되는 분야에서는 성능이 크게 저하된다. LLM의 훈련용 데이터 세트에는 부동산 관련 콘텐츠가 과소 표현되는 경향이 존재하여 모델의 미세조정(fine tuning)이나 추가적인 맥락 제공 없이 LLM을 부동산 분석에 활용하면 모델의 지식 부족 등이 원인이 되어 낮은 성능을 보인다(Chen and Si, 2024; Gloria et al., 2024).

더불어 LLM의 환각 모델 특성으로 인해 LLM은 종종 그럴듯하게 들리지만, 사실이 아니거나 터무니없는 정보를 생성하는 환각 현상이라는 문제를 가지고 있다(Bender et al., 2021). 이러한 문제는 사실에 기반한 답변이 중요한 분야와 최신의 정보가 요구되는 분야에서 LLM 활용의 주된 제약 요소이다. 이 문제점을 극복하기 위하여 프롬프트에 최신 데이터나 전문적인 외부 지식을 주입하는 방안이 제시되었으며(Martino et al., 2023), 프롬프트에 적절한 외부의 맥락 데이터를 통합하면 LLM의 환각 현상을 줄이고 생성된 답변의 신뢰도를 높일 수 있음을 보였다.

외부의 전문적인 지식 데이터베이스와 연계하여 LLM을 활용하는 방법론의 대표적인 방식이 검색증강생성(RAG)이다. Lewis et al.(2020)이

제시한 RAG는 먼저 검색을 통해 문제 해결에 필요한 외부의 지식을 확보하고, 이를 기존의 프롬프트에 맥락 데이터로 통합하여 프롬프트를 증강한다. 그리고, LLM에 증강된 프롬프트를 제공함으로써 지식 집약적인 작업에 대한 효과적인 답변 생성이 가능함을 보였다. RAG는 LLM에 대한 미세조정과 같은 재훈련 방법에 비해 구축이 쉽고, 최신 데이터 사용이 용이하며 출처를 기재함으로써 답변의 신뢰도를 높일 수 있다(Gao et al., 2023). 일례로, Zhang et al.(2023)은 뉴스, 소셜 미디어 등으로 구축된 RAG-LLM을 이용하여 금융 감성분석을 수행하였다. 이는 일반적인 LLM 및 금융 분야에 미세 조정된 특화 모델에 비해서도 높은 정확도를 보였고, 시계열 시기별로 판단의 기준이 달라지는 질의에도 높은 분류 성능을 보였다.

RAG-LLM의 구축 단계는 인덱싱, 검색, 증강 및 생성의 3단계로 분류할 수 있다. 먼저, 인덱싱 단계에서는 원본 문서를 일반 텍스트로 변환하고, 특정한 크기의 텍스트 길이로 문서를 분할하고, 분할된 문서에 대한 임베딩(embedding) 벡터를 생성한다. 임베딩은 단어나 문장을 벡터로 인코딩하는 과정으로써 의미나 문맥적으로 유사한 단어의 경우 유사한 임베딩 벡터를 가진다. Reimers and Gurevych(2019)이 제안한 Sentence BERT와 같은 문장 단위의 임베딩 모델이 주로 사용되며, 문서의 텍스트와 임베딩을 향후 검색 단계에서 활용할 수 있도록 저장한다. 원본 문서의 생성일자와 같은 메타데이터(meta data)를 이용하면 더 빠른 검색과 효율적인 활용이 가능하다.

검색 단계에서는 사용자의 질의를 동일한 임베

딩 모델을 통해 변환한 다음, 높은 임베딩 유사성을 지닌 문서를 검색하게 된다. 사용자가 설정한 개수의 상위 문서가 검색 단계에서 획득되며, 이때 중복을 최소화하고 풍부한 맥락 정보의 획득을 위해 MMR(maximum marginal relevance)과 같은 검색 방법론을 활용하여 유사성과 다양성의 균형을 찾아 검색된 문서에 순위를 부여할 수 있다.

증강 및 생성 단계는 정해진 프롬프트 템플릿에 검색 단계에서 획득한 문서의 내용들을 결합하여 프롬프트를 증강하고 LLM에 보내는 단계다. 이때 프롬프트 내에 지시사항을 명시하여 자체 지식의 사용 여부와 같은 작동 방식을 통제할 수 있다.

최근 부동산 분야에서도 LLM을 활용한 연구가 제시되고 있다. 중국의 주택을 대상으로 유사한 거래 사례를 LLM으로 분석하여 임대가격을 예측한 사례(Chen and Si, 2024)에서는 규격화된 프롬프트 템플릿을 작성하고, 프롬프트 내에 명확한 지시(instruction)와 유사한 거래 사례를 맥락 데이터로 제공하여 LLM 기반의 가격 예측을 수행하였다. 실험 결과 예측 성과는 프롬프트의 사소한 표현의 변화에 강건한 모습을 보였으며, 맥락 데이터에 유사한 거래 사례가 10개 주어진 모델은 거래 사례가 없는 모델의 예측 결과 대비 MAE가 13.5% 우수하고 결정계수가 0.46에서 0.80으로 증가함을 보였다.

3. 선행연구와의 차별점

본 연구는 우수한 예측력과 도메인의 전문 지식으로 활용할 수 있는 부동산 뉴스 기사 데이터

를 대상으로, RAG-LLM 구축과 프롬프트 엔지니어링을 통해 비정형 데이터를 부동산 시장의 전망에 효과적으로 활용할 수 있는 형태로 전환했다는 점에서 의의가 있다.

먼저, LLM의 우수한 자연어처리 능력은 수집된 기사 원문에 제시된 여러 내용 중 질의 대상에 대한 명확한 내용의 추출이 가능하고, 질의와의 관계성과 중요도에 대한 파악이 가능하며, 제시된 내용이 미래 시점에 대한 전망인지 여부를 구분할 수 있다. 기존의 언어모델보다 강력한 역량을 지닌 LLM을 활용하여 비정형 데이터 분석을 고차원적으로 진행함으로써, 기존의 뉴스 및 미디어를 활용한 연구에서 제시된 여러 한계(개체별 감성 추출, 맥락 파악, 전망 정보 추출 등)를 극복하여, 기존의 뉴스 기반 예측 연구보다 미래 가격 변화에 더 높은 상관관계를 가진 전망 점수를 산출하였다.

두 번째로, RAG와 프롬프트 엔지니어링을 통해 LLM의 한계점인 환각, 전문지식 부족 그리고 사전지식에 의한 편향을 최소화하였다. RAG를 통해 LLM이 가지는 부작용을 최소화하면서 LLM이 가지는 강력한 추론 능력을 활용할 수 있는 시스템 구축 방안을 제시하고, 산출된 전망 점수의 아파트 매매가격지수 예측 성능을 검증하였다.

마지막으로, 본 연구는 RAG-LLM를 활용하여 미디어 빅데이터만으로 시간적 유용성을 높이는 예측 생성이 가능함을 보였다. 통계 데이터와 같은 거시변수나 심리지표, 매매가격지수의 경우 집계 후 공표까지 1~2개월의 시차가 발생하여 실시간으로 유용한 전망을 생성하기가 어렵다. 반면, 연구에서 제안한 접근 방식은 월말에 신속하

게 당월의 RAG를 구축할 수 있고 이를 바탕으로 다음 달 전망을 이달 말에 생성할 수 있다. 이를 통해 시간적 유용성을 높인 전망의 생성이 가능하며, 다른 지표들의 발표 전까지 의사결정을 위한 예비적인 선도 지표로 활용할 수 있다.

III. 연구 방법

1. 연구의 범위와 데이터

본 연구에서는 서울특별시의 아파트를 분석 대상으로 선정하였다. 서울은 대한민국 인구의 19%가 거주하고 있는 가장 큰 도시이고(행정안전부, 2024), 정책 목표와 및 사회적 관심이 서울 아파트시장에 집중되어 있기 때문이다(이창무, 2020). 사용한 아파트 매매가격의 기준 지표는 한국부동산원이 집계하여 작성하는 아파트 매매가격지수를 활용하였다. 월간 기준으로 공표된 매매가격지수를 대상으로 2014.12.~2025.02.의 123개월 기간에 대해 가격지수를 수집하였다. 동 지수는 집계부터 공표까지 약 2개월 이내의 시차가 존재한다(익월 15일 공표).

연구에 활용할 부동산 뉴스 기사는 금융정보 단말기(check expert)에서 제공하는 뉴스 API를 이용하여 수집하였다. 12개 경제 전문지에서 '부동산' 항목으로 분류된 뉴스 기사를 대상으로 제목, 본문 및 메타데이터(일시, 언론사)를 수집하였다. 2015.01.~2024.12.까지의 기간 동안 총 365,413건의 기사를 수집하였다. 수집한 뉴스 기사의 상세 통계는 <표 1>과 같다.

〈표 1〉 수집한 뉴스 기사의 통계

언론사	수집한 기사 수	평균 텍스트 길이
한국경제	81,674	1,297
매일경제	64,730	1,214
파이낸셜	57,580	1,115
머니투데이	35,059	1,181
News Pim	31,685	1,045
아시아경제	30,563	1,187
조선비즈	21,212	1,268
서울경제	20,481	1,123
한국금융신문	8,115	1,192
뉴스토마토	6,114	1,328
컨슈머타임스	5,086	1,018
전자신문	3,114	986

주 : 기사 수집 기간은 2015.01.01.~2024.12.31.

연구에서는 수집된 기사를 바탕으로 RAG 시스템을 구축한 다음, 서울 지역의 아파트 가격의 1개월 및 3개월 뒤 가격에 대한 미래 전망을 질의하여 답변을 RAG-LLM을 통해 생성하고 생성된 예측의 유효성을 평가하였다.

2. RAG(Retrieval Augmented Generation)

시스템 구축

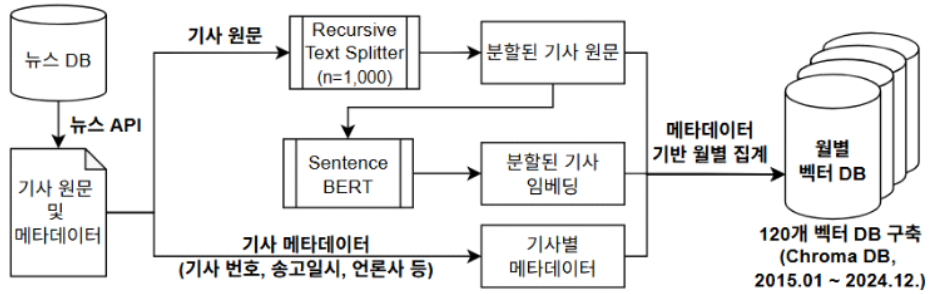
본 연구에서는 전반적인 시스템 구축을 위해 Python 환경의 LangChain 라이브러리를 활용하였다. LLM은 별도로 API key 발급 후 라이브러리에 내장된 함수들을 통해 질의 및 답변 생성 작업을 수행하였다.

RAG-LLM 구성 단계는 인덱싱, 검색, 증강 및 생성의 세 단계로 분류할 수 있다. 인덱싱 단계에

서는 먼저 분할 작업을 수행하여, 수집된 뉴스 기사와 같은 문서를 적절한 단위로 나누어야 한다. LLM은 입력 프롬프트 크기에 제한(context window size)이 있어 전체 문서를 부분으로 나누어 필요한 부분만 활용하여 프롬프트를 증강하는 절차가 효과적이기 때문이다. 다음으로 수집된 원본 문서들을 적절한 크기의 텍스트 덩어리로 분할하는 과정이 필요하다. Wang et al.(2024)은 인덱싱 단계에서 너무 긴 길이로(2,048자 이상) 문서를 분할하면 데이터의 신뢰성에 악영향을 주었으며, 겹치는 부분(overlap)을 두는 sliding window 방식으로 문서를 분할하는 게 우수한 성능을 보인다는 것을 제시하였다.

본 연구에서는 문장 단위의 분할을 수행하는 recursive text splitter를 활용하여, 뉴스 기사 본문을 1,000자 이내의 텍스트로 문단 및 구두점을 기준으로 분할하여 별도의 문서로 저장하였다. 문서 분할 과정에서는 앞뒤로 200자 이내의 겹치는 부분을 포함하여 의미의 소실을 최소화하였다. 이후, 분할된 문서를 Gan(2021)이 제시한 한국어 기반 Sentence BERT(<https://huggingface.co/jhgan/ko-sbert-nli>)를 이용하여 임베딩으로 변환하였다. 임베딩과 분할된 문서, 메타데이터를 월별로 집계하여 Chroma DB 기반의 월별 부동산 뉴스 벡터 DB를 구성하였다. 데이터 수집과 인덱싱 단계를 포함한 RAG 구축 단계를 모식으로 나타내면 〈그림 1〉과 같고, 월별 벡터 DB에 분포된 문서의 통계는 〈표 2〉와 같다.

검색 단계에서는 LLM이 해결하기 위한 문제와 가장 관련성 높은 문서를 집계하는 과정이 수행된다. 본 연구에서는 검색 단계에서 중복을 줄이고



〈그림 1〉 기사 수집 및 RAG(retrieval augmented generation) 인덱싱 단계 모식도

〈표 2〉 월별 벡터 DB 내 포함된 문서의 분포

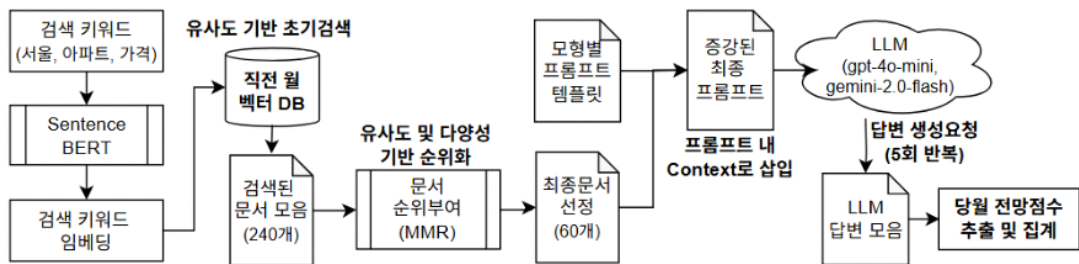
구분	값
평균	5,320
최대	3,374
최소	7,859
표준편차	950
합계	638,327

주 : Chroma DB에 구축된 총 텍스트 덩어리의 통계.

다양성을 높인 최종 검색 결과를 얻기 위하여 Carbonell and Goldstein(1998)이 제안한 MMR 방법론을 활용하였다. 직전 월의 DB에서 검색 키워

드를 임베딩으로 변환하여 쿼리를 작성하고, 작성된 쿼리와 유사도가 높은 문서를 초기 검색 풀 크기($N=240$)만큼 수집하였다. 수집된 문서를 MMR을 통해 유사도 및 다양성을 모두 고려하여 최종순위를 계산하여 최종 검색 문서 수($K=60$)만큼의 최종 문서를 선정하였다(〈그림 2〉).

MMR의 산식은 (식 1)과 같다. $Sim(d, Q)$ 은 해당 문서(d)와 쿼리(검색 키워드 임베딩, Q)의 유사도이며, $\max_{d \in D} Sim(d, d')$ 은 해당 문서(d)와 이미 선택된 문서 세트(D')에서 가장 유사한 문서와의 유사도이다. λ 는 유사성과 다양성의 상대 중요도



주 : BERT, bidirectional encoder representations from transformers; MMR, maximum marginal relevance; LLM, large language model; RAG, retrieval augmented generation.

〈그림 2〉 RAG 활용 단계 모식도(문서 검색, 프롬프트 증강, LLM 기반의 답변 생성 단계)

를 조절하는 매개변수로 연구에서는 0.5를 선정했다(〈표 3〉). MMR값으로 K 개의 문서를 순차적으로 선택하는데, 각 단계에서 이전에 선택된 문서 세트를 고려하여 MMR을 최대화하는 문서를 선택하는 과정을 K 회 반복한다.

$$MMR = \arg \max [\lambda \cdot Sim(d, Q) - (1 - \lambda) \cdot \max_{d' \in D} Sim(d, d')] \quad (\text{식 1})$$

RAG 과정에서 너무 많은 문서를 수집하여 LLM에 전달하면 LLM 활용에 드는 비용이 증가하며, 수집된 문서의 일부를 LLM이 인지하지 못하는 needle in haystack이라는 문제가 발생한다(Laban et al., 2024). 지나치게 많은 맥락 정보로 프롬프트를 증강하여 프롬프트의 크기가 길어지면 일부 문서의 내용에 대한 분석 효율이 떨어지게 된다.

Leng et al.(2024)은 최신 LLM 모델들을 대상으로 프롬프트의 토큰 길이를 증가시키면서 RAG 정답률을 비교하였다. gpt-4o-mini 및 gemini-1.5-flash LLM은 각각 3.2만 토큰과 6.4만 토큰 길이에서 성능이 최고점이 나타나는 것을 확인하였다. 본 연구에서 인덱싱한 문서가 최대 1,000 토큰

크기(1글자=1 토큰) 정도임을 고려하여, 증강된 프롬프트의 크기 한계를 6.4만 토큰 이하로 설정하고자 최종 검색 문서 수(K)를 60개로 설정하였다.

초기 검색 풀 크기(N) 설정은 LLM 기반 문서 순위 부여 연구에서 MMR이 다른 방법론보다 우수한 성과를 보인 Carraro and Bridge(2024)의 사례를 참고하였다. MMR이 다른 재순위화 방법보다 높은 성능을 보인 데이터 세트에서 검색시 N 을 K 의 4배수로 활용한 점을 고려하여 본 연구의 초기 검색 풀 크기도 60개의 4배수인 240개로 설정하였다.

증강 및 생성 단계에서는 서울 아파트 매매가격 미래 예측을 위한 프롬프트를 생성하였다. 전월 기사로 구성된 DB에서 검색 단계를 통해 수집된 문서를 프롬프트 템플릿에 맥락 정보로 증강하여 최종 프롬프트를 작성하였으며, 생성 단계에서는 프롬프트를 LLM에 전송하여 당월 미래 전망을 생성하고 이를 집계하였다. LLM 종류에 따른 효과를 최소화하기 위해 OpenAI의 gpt-4o-mini(2024.07.18.), Google의 gemini-2.0-flash(2025.02.24.)의 두 가지 모델을 답변 생성에 활용하였으며, 불안정성을 통제하기 위하여 LLM의 초매개변수 중 Temperature(출력의 무작위성을 통제하는 변수)는 0으로, Max_tokens(출력 크기 변수)은 3,000으로 설정하였다.

〈표 3〉 검색단계의 주요 변수 정리

구분	주요 변수
검색 키워드(서울)	서울, 아파트, 가격
최종 검색 문서 수(K)	60
초기 검색 풀 크기(N)	240
순위 부여 방법론	MMR

주 : MMR, maximal marginal relevance($\lambda=0.5$).

3. 프롬프트 설계와 답변 생성

연구에 활용한 프롬프트 템플릿의 세부적인 구성은 LangChain에서 제시하는 RAG 프롬프트의 권장 사항과 부동산 분야의 RAG 활용 사례인

Chen and Si(2024)와 Yang et al.(2024)에서 제시된 프롬프트 구조를 참고하여 작성하였다. 프롬프트 템플릿은 <표 4>와 같이 지시(instruction), 맥락(context), 질의(query), 출력형식지정(output formatting)의 네 가지 부분으로 구분된다.

먼저, 지시 프롬프트에서는 RAG의 명확한 작동을 위해 역할 부여, 맥락에 근거한 답변, 출력의 적시를 지시하였다. 또, 연구 모형별로 마지막 지시 사항(지시 3)을 달리하여 단순 질의 기반의

RAG 프롬프트뿐만 아니라 Chain-of-Thought 방식의 프롬프트를 활용하였다. 맥락 프롬프트에서는 RAG의 검색 단계에서 수집된 텍스트를 구분하여 통합하였다. 제시한 프롬프트 템플릿의 {context} 부분에 맥락 데이터로 검색 단계에서 수집된 60개의 문서가 들어가게 된다. 질의 프롬프트에서는 1개월 및 3개월의 두 가지 기간을 활용하여 질의 프롬프트를 달리 작성하였다.

출력형식지정 프롬프트는 LLM의 출력 분포와 편향에 직접적인 영향을 미치는 것으로 알려져 있다. Schulhoff et al.(2024)은 리커트 척도를 출력 형식으로 활용하는 것이 LLM에게 척도의 개념을 더 잘 이해할 수 있게 한다는 점을 보였고, Lu et al.(2025)은 뉴스 헤드라인을 대상으로한 감성분석 작업에서, 인간 전문가가 수행한 분석 결과와 리커트 척도를 이용해 생성한 LLM의 감성분석 사이 높은 유사도가 나타남을 보였다.

특히 답변 형식으로 '예/아니오'의 이분법적인 형식을 지정하면 다수의 모델에서 명확한 부정 편향이 발생했지만, 여러 구간을 가진 리커트 척도 형식을 활용하면 LLM은 인간과 높은 상관관계(gpt-4o-mini 기준 0.77)를 가진 답변을 생성함을 보였다. 단, 인간이 극단적인 감성으로 분류한 기사에서 LLM은 극단적인 점수를 피하는 보수적인 모습이 나타났다.

본 연구에서도 답변 형식에 따른 LLM 출력 결과의 편향을 최소화하기 위하여 리커트(Likert) 7점 척도로 출력 형식을 지정하였고, 예시 답변을 {example} 영역에 기재하여 구체화된 LLM의 답변 생성을 유도하였다.

답변 생성 단계에서는 예측 기간, 프롬프트 종

<표 4> 프롬프트 템플릿 주요 구성

구분	프롬프트 템플릿
지시 1	당신은 질문-답변을 돕는 어시스턴트입니다.
맥락	다음은 주어진 질문에 답변하기 위해 수집된 문서들입니다: {context}
질의	질문: 향후 {period}동안 서울 지역의 아파트 가격 전망이 얼마나 긍정적인가요?
지시 2	답변은 한국어로 작성해주세요. 질문에 답변할 때는 수집된 문서 정보들을 활용하세요. 만약 문서에서 질문에 대한 답을 찾을 수 없다면, "제공된 문서에서는 해당 정보를 찾을 수 없습니다."라고 답변해주세요. 답변은 반드시 주어진 문서의 내용에 기반해야 하며, 추측이나 일반적인 지식을 포함하지 마세요.
출력 형식 지정	답변의 7점 척도로 제한해주세요! 척도 기준은 다음과 같습니다. 척도 기준: 1=매우 부정적, 2=부정적, 3=약간 부정적, 4=중립, 5=약간 긍정적, 6=긍정적, 7=매우 긍정적 {example}
지시 3 (RAG)	답변의 하단에는 참고한 주요 내용의 출처에 대해 간략히 기재해주세요.
지시3 (CoT)	결론이 정확인지 확인하기 위해 단계별로 추론 과정을 작성해주세요. 처음부터 정답을 단순히 언급하는 것은 피해주세요.

주 : 1) CoT는 Wei et al.(2022)의 프롬프트 내 추론 방식인 Chain-of-Thought.
2) RAG, retrieval augmented generation.

류(Chain-of-Thought 사용 여부)를 달리하여 총 4가지의 전망 모형을 생성하였다. 모형별로 전망 생성 단계의 세부적인 차이점은 <표 5>와 같으며, 예시 답변을 포함하여 작성된 모형 2의 프롬프트 전문은 <표 6>과 같다.

답변 생성 단계에서는 LLM의 불확실성을 최소화하면서 평가를 수행하기 위하여 반복적인 답변의 생성을 요청하였다. Blackwell et al.(2024)에서는 gpt-4o 및 gemini-1.5-pro에 대해 긴 질의를 수행할 때 4회 이상 질의를 진행하면 30회 이상 질의를 한 경우와 유사한 신뢰수준이 형성됨을 보였다. 본 연구에서는 모형별로 5회의 반복적인 답변 생성을 수행한 이후 생성된 답변에서 제시된 전망점수를 집계하여 모형별 최종 전망점수를 산출하였다.

전망점수 산출 과정에서의 문서 수집 예시와 LLM의 답변 생성 메커니즘에 대한 이해를 돕기 위하여, RAG-LLM이 부정적인 전망 점수를 생성한 2022년 10월과 긍정적인 전망 점수를 생성한 2024년 7월에 수집된 주요 뉴스 기사 일부의 헤드라인을 <표 7>에, 이를 바탕으로 생성된 다음 달 가격 방향성의 답변 예시와 전망 점수를 <표

8>에 정리하였다. 예시 답변은 LLM의 답변 작성 과정이 명확히 나타나는 Chain-of-Thought 프롬프트를 사용한 모델 2번으로 제시하였다.

<표 6> 예시: 모형 2의 프롬프트 템플릿 전문

프롬프트 전문(월별로 서로 다른 맥락정보 증강)	
당신은 질문-답변을 돕는 어시스턴트입니다. 다음은 주어진 질문에 답변하기 위해 수집된 문서들입니다:	
{context}	
질문: 향후 1개월 동안 서울 아파트 가격 전망이 얼마나 긍정적인가요?	
답변은 한국어로 작성해주세요. 질문에 답변할 때는 수집된 문서 정보들을 활용하세요. 만약 문서에서 질문에 대한 답을 찾을 수 없다면, "제공된 문서에서는 해당 정보를 찾을 수 없습니다."라고 답변해주세요. 답변은 반드시 주어진 문서의 내용에 기반해야 하며, 추측이나 일반적인 지식을 포함하지 마세요.	
답변의 7점 척도로 제한해주세요! 척도 기준은 다음과 같습니다. 척도 기준: 1=매우 부정적, 2=부정적, 3=약간 부정적, 4=중립, 5=약간 긍정적, 6=긍정적, 7=매우 긍정적.	
답변의 첫 줄의 예시는 다음과 같습니다.	
[질문]: 향후 1개월 동안 서울 아파트의 가격 전망이 얼마나 긍정적인가요?	[답변]: 1점. 향후 1개월 동안 서울 아파트의 가격 전망은 매우 부정적입니다.
[질문]: 향후 1개월 동안 서울 아파트의 가격 전망이 얼마나 긍정적인가요?	[답변]: 4점. 향후 1개월 동안 서울 아파트의 가격 전망은 중립적입니다.
[질문]: 향후 1개월 동안 서울 아파트의 가격 전망이 얼마나 긍정적인가요?	[답변]: 6점. 향후 1개월 동안 서울 아파트의 가격 전망은 긍정적인입니다.
결론이 정확한지 확인하기 위해 단계별로 추론 과정을 작성해주세요. 처음부터 정답을 단순히 언급하는 것은 피해주세요.	
답변:	

주 : {context} 영역에 검색단계의 최종 문서 포함.

<표 5> 답변 생성 단계 모형 구분

구분	전망기간	프롬프트	LLM
모형 1	1개월	RAG	Gpt-4o-mini, gemini-2.0-flash 각 5회 반복 후 평균
모형 2	1개월	CoT	
모형 3	3개월	RAG	
모형 4	3개월	CoT	

주 : 1) RAG: 단순 질의응답 기반의 RAG 프롬프트.
2) CoT: Chain-of-Thought 기반의 RAG 프롬프트.
3) LLM, large language model.

〈표 7〉 긍정/부정 구간별 수집된 최종 문서 예시

구분	답변 예시
부정적 구간 (2022.10)	2022.10.15 06:00 [머니투데이] 매물 한 달에 1,000개씩 쌓이는데...서울 아파트 살 사람 계속 줄어
	2022.10.22 07:30 [아시아경제] 10명 중 7명 "집값 더 내릴 것"...원희룡 "아직 폭락 상황 아냐"
	2022.10.25 08:53 [NEWSPIM] 금리 인상 탓?...서울 아파트 중위 전세가 6억 원 밑으로
	2022.10.28 13:35 [한국경제] "오늘보다 내일이 더 싸다"...강남 대표 아파트도 6억 톱
	2022.10.31 19:21 [한국경제] 강남 3구마저...투자 안전지대 사라졌다
긍정적 구간 (2024.07)	2024.07.04 17:21 [한국경제] 서울 집값 15주째 뒤흔길...성동구 가장 많이 올랐다
	2024.07.11 15:17 [매일경제] "자고나면 또 오르려나"...서울 아파트 5년 10개월만에 최대 상승
	2024.07.18 11:14 [매일경제] 서울 집값 강세에...서울 주택사업 전망도 9개월만에 '맑음'
	2024.07.19 14:59 [파이낸셜] 당장 서울 집값 '억' 뛰는데...그린벨트에 추가 2만가구 대안?
	2024.07.29 11:32 [조선비즈] 선도아파트가 이끈 서울 집값... "집값 오른다"

주 : 수집된 문서 60개 중 헤드라인 있는 문서 임의선별.

4. 분석 방법

분석 단계는 크게 네 단계로 진행하였다. 첫 단계에서는 제시한 4가지 모형별로 월별 전망점수 집계, 조정전망점수의 생성 및 단위근 검정(unit root test)을 수행하였다. 우선 RAG-LLM을 통해 모형별로 생성된 답변의 리커트 척도 부분을 추출하여 모형별 월별 전망점수를 집계한 후 평균하였다.

다음으로 ARIMAX 분석을 위한 별도의 조정 전망점수를 산출하였다. 이는 ARIMAX와 같은 회귀적인 특성을 가진 분석을 수행할 때, 리커트 척도의 순위척도 특성과 극단적인 감정에 대한 출력에 보수적인 LLM의 특성(Lu et al., 2025)을 고려하여 생성된 전망점수를 재조정된 점수이다.

LLM의 출력형식지정 단계에서 리커트 척도에 제시한 키워드 특성상, 2점(부정적)과 3점(약간 부정적) 사이의 거리가 3점과 4점(중립적)의 거리보다 크게 측정되어야 한다. 이를 고려하지 않고 리커트 척도값을 단순 숫자로 변환하여 회귀분석에 활용하면 등간 척도로 가정하면서 정보의 손실이 발생할 수 있다. 본 연구에서 활용한 모형별 조정전망점수의 변환식은 (식 2)와 같다.

마지막으로 종속변수인 매매가격지수와 독립변수인 전망점수와 조정전망점수에 대해 단위근 검정을 수행하였다. ADF(augmented Dickey-Fuller) 검정을 통해 단위근의 존재 여부를 확인하였고, 단위근이 존재하는 경우 차분하여 시계열의 정상성을 확인하였다.

$$\text{조정전망점수} = (\text{평균전망점수} - 4)^3 \quad (\text{식 } 2)$$

두 번째 단계로 교차상관분석을 진행하였다. 모형별로 집계된 전망점수와 실제 매매가격 변화율 사이의 시차를 조정해 가며 시차별 교차상관관계를 확인하였고, 산출된 전망점수가 유의미한 선형성을 가지는 구간과 최적 시차를 확인하였다. 교차상관관계의 분석은 시차별로 Spearman 상관계수와 Pearson 상관계수를 계산하였다.

세 번째 단계로, ARIMAX 모델을 통해 시계열

<표 8> LLM이 생성한 답변 예시(모형 2)

구분	답변 예시
부정적 구간 답변 예시 (2022.11 전망점수)	<p>[질문]: 향후 1개월 동안 서울 아파트의 가격 전망이 얼마나 긍정적인가요? [답변]: 3점. 향후 1개월 동안 서울 아파트의 가격 전망은 약간 부정적입니다.</p> <p>추론 과정:</p> <ol style="list-style-type: none"> 1. 최근 서울 아파트 매매수급지수가 76.9로, 이는 매도자 우위의 시장을 나타내며, 매수 심리가 위축되고 있음을 보여줍니다. 이는 가격 하락 압력을 의미합니다. 2. 서울 아파트의 평균 전세가격이 하락하고 있으며, 전세 매물이 증가하고 있는 상황입니다. 이는 전세 시장의 불안정성이 매매 시장에도 영향을 미칠 수 있음을 시사합니다. 3. 금리 인상과 경기 침체 우려가 지속되고 있으며, 이는 부동산 시장에 부정적인 영향을 미치고 있습니다. 특히, 금리가 오르면서 대출 부담이 커지고 있어 매수 심리가 더욱 위축될 가능성이 높습니다. 4. 전문가들은 현재의 부동산 시장이 실수요 중심으로 변화하고 있으며, 외지인의 원정투자 비율이 감소하고 있다고 언급하고 있습니다. 이는 가격 상승 기대가 줄어들고 있음을 나타냅니다. 5. 이러한 요소들을 종합적으로 고려할 때, 향후 1개월 동안 서울 아파트 가격 전망은 약간 부정적이라고 판단할 수 있습니다.
긍정적 구간 답변 예시 (2024.08 전망점수)	<p>[질문]: 향후 1개월 동안 서울 아파트의 가격 전망이 얼마나 긍정적인가요? [답변]: 6점. 향후 1개월 동안 서울 아파트의 가격 전망은 긍정적입니다.</p> <p>**추론 과정:**</p> <ol style="list-style-type: none"> 1. **현재 가격 상승세**: 최근 서울 아파트 가격이 15주 연속 상승하고 있으며, 7월 첫째 주에는 0.20% 상승했습니다. 이는 가격 상승세가 지속되고 있음을 나타냅니다. 2. **매매가격전망 지수**: 서울의 매매가격전망 지수는 127로, 이는 상승 전망이 더 높다는 것을 의미합니다. 이 지수는 2000년 6월 이후 가장 높은 수치로, 시장 참가자들이 긍정적인 심리를 가지고 있음을 보여줍니다. 3. **수요 증가**: 서울 아파트에 대한 수요가 증가하고 있으며, 특히 전세가격 상승이 매매 수요로 전환되고 있는 상황입니다. 전세가격이 상승하면서 갭투자도 증가하고 있습니다. 4. **전문가 의견**: 전문가들은 서울 아파트 가격이 당분간 상승할 것으로 보고 있으며, 이는 실수요자들의 매수 심리가 회복되고 있다는 분석과 일치합니다. 5. **시장 심리**: 최근의 거래량 증가와 매수 심리 회복은 향후 1개월 동안 서울 아파트 가격이 긍정적인 방향으로 나아갈 것이라는 신호로 해석됩니다. <p>이러한 요소들을 종합적으로 고려할 때, 향후 1개월 동안 서울 아파트 가격 전망은 긍정적이라고 판단할 수 있습니다.</p>

주 : 1) 제시된 답변은 gpt-4o-mini 생성 답변.
 2) 예시 답변 생성에 활용된 모형 2는 Chain-of-Thought 프롬프트 활용.
 3) LLM, large language model.

모델의 외생변수(exogenous variable)로 조정
 전망점수를 사용하면 예측 정확도가 얼마나 개선

되는지 확인하였다. 우선, 준거 모델로 활용한
 ARIMA 모델은 선행연구에서 제시된 시계열 모

델에서 출발하여 사용한 매매가격지수 시계열에 대해 가장 낮은 AIC(Akaike information criterion)와 BIC(Bayesian information criterion)을 가지는 모형을 선택하였다. 이후, 동일한 차수를 가진 ARIMAX 모델을 통해 외생 변수 투입 시 개선되는 예측 지표를 확인하였다.

본 연구에서 활용한 ARIMA 및 ARIMAX 모형은 수식은 (식 3) 및 (식 4)와 같다. 식의 $\Delta Y_t = Y_t - Y_{t-1}$ 이며, Y_t 는 t 시점의 매매가격지수 관측치이다. ϕ_1 및 ϕ_2 는 자기회귀 모수이며, β 는 외생변수의 회귀계수, Z_t 는 t 시점의 외생변수로 $t-1$ 시점에 수집된 기사를 바탕으로 생성된 조정전망점수를 활용하였다.

$$\Delta Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \epsilon_t \quad (\text{식 3})$$

$$\Delta Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \beta Z_t + \epsilon_t \quad (\text{식 4})$$

시계열 모형의 예측오차를 평가하기 위한 지표로는 RMSE, MAE 지표를 활용하였고, 가장 우수한 모형에 대한 세부적인 모형 추정 결과를 확인하였다.

전체 10년의 시계열 구간을 8대 2 기준의 train/test 구간으로 나누어, 최근 2년의 test 구간에서 매월 one-step ahead out-of-sample forecast 방식으로 모형의 예측을 수행하여 예측 오차를 평가하였다. 시계열모형을 통해 차분된 시계열 ΔY_t 를 예측한 후, 이를 원 시계열로 역변환하여 최종 예측값을 구하였다. 예측오차 계산을 위한 지표는 다음의 (식 5), (식 6)과 같으며, RMSE와 MAE는 모두 값이 작을수록 모형의 예측 성과가 우수함을 의미한다. 수식의 Y_i 는 i 시점

의 실제 지수값, \hat{Y}_i 는 i 시점의 예측 지수값, 그리고 n 은 Test 구간의 데이터 수를 의미한다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (\text{식 5})$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (\text{식 6})$$

마지막 단계로, 전체 시계열 구간인 120개월에 대하여 RAG-LLM을 활용한 방향성 예측을 수행하였다. 이 단계에서는 전월 말($t-1$)에 수집된 기사를 바탕으로 생성한 전망 점수를 이용하여 미래의 아파트 매매가격 지수 변화를 예측하였다. 다시 말해, $t-1$ 까지의 기사를 바탕으로 $t-1$ 부터 목표 시점(1개월: t 시점, 3개월: $t+2$ 시점)까지의 가격 변화 방향성을 예측하였다. 리커트 7점 척도의 기준으로 구성된 전망 점수가 4점을 초과하면 상승, 4점 미만이면 하락으로 방향성 예측을 수행하였다. 정확히 4점일 경우에는 중립으로 간주하여 예측 대상에서 제외하였으며, 실제 가격지수의 변화율과 비교하여 예측 성능을 측정하였다.

방향성 예측의 벤치마크 대상으로 삼은 베이스라인 모델은 앞 단계의 ARIMA 모델을 활용하였으며, 매매가격지수의 공표에 걸리는 시차인 1.5개월을 고려하여 참조 시점에서 3-step forward와 2-step forward의 시계열을 예측하여 방향성을 예측하였다. 3개월을 예측한 모델은 미래 3개월 수익률에 대한 방향성 예측을 수행하였으며, 베이스라인 모델로 5-step forward와 2-step forward의 시계열을 예측하여 벤치마크로 활용하였다.

IV. 실증분석 결과

1. 기초 통계 및 단위근 검정

RAG-LLM을 통해 생성한 전망점수는 모형별로 120개의 데이터를 가지는 시계열로 세부 기초 통계량은 <표 9>와 같다.

정상성 검증을 위해 수행한 ADF 검정 결과는 모든 모형에서 전망점수와 조정전망점수 모두 수준 변수 단계에서 단위근을 가지지 않는 것으로 나타났다(<표 10>).

<표 9> 모형별 전망점수 기초 통계(N=120)

구분	최소	최대	평균	표준편차
모형 1	2	7	4.92	0.92
모형 2	3	6	4.94	0.91
모형 3	2	7	4.90	0.96
모형 4	2	6	4.94	0.93
평균			4.93	0.93

주 : 생성 기간은 2015.01.~2024.12.의 120개월.

<표 10> 전망점수의 단위근 검정 결과(수준변수)

구분	모형	t값	p	1%
전망점수 (수준변수)	모형 1	-4.75	6.7e-5***	-3.49
	모형 2	-4.76	6.5e-5***	
	모형 3	-4.74	7.0e-5***	
	모형 4	-4.79	5.7e-5***	
조정 전망점수 (수준변수)	모형 1	-5.58	1.4e-6***	-3.49
	모형 2	-5.35	4.2e-6***	
	모형 3	-4.83	4.6e-5***	
	모형 4	-6.56	8.5e-9***	

주 : *** 모든 p값이 0.01 이하로 형성.

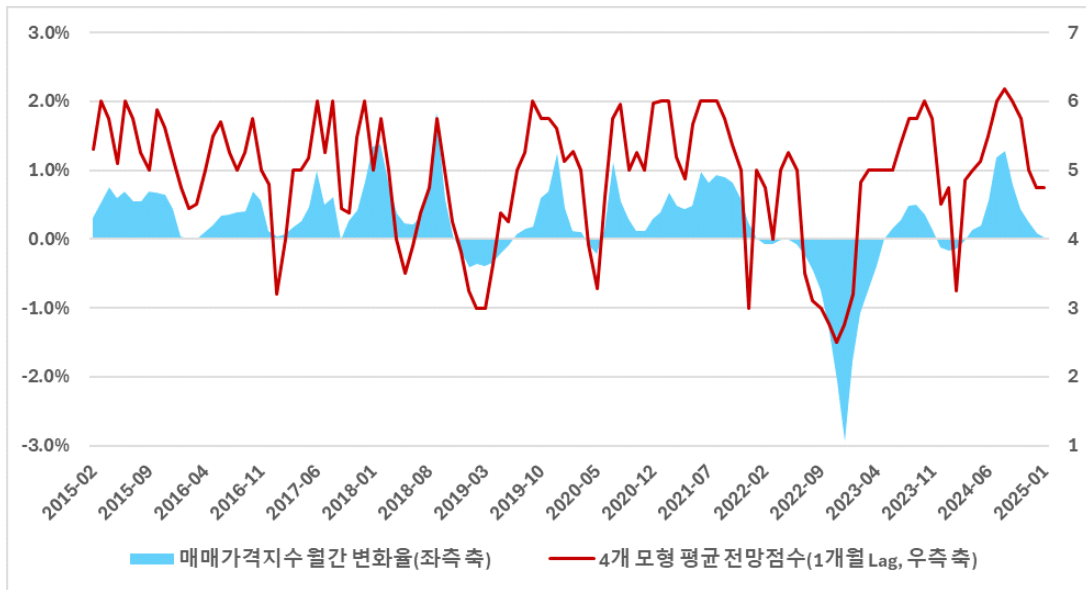
분석의 대상인 아파트 매매가격지수의 기초 통계와 단위근 검정 결과는 <표 11>과 같다. 매매가격지수는 수준변수 단계에서 단위근이 존재하며, 1차 차분 후에는 단위근 귀무가설이 유의수준 1%에서 기각되어 정상성을 가지는 것으로 확인되었다.

2015년부터 2024년 말까지의 분석 기간에 매매가격지수의 월간 변화율과 모형이 생성한 평균적인 전망점수의 그래프는 <그림 3>과 같이 나타난다. 그래프의 전망점수는 모형 1~4에서 산출된 점수를 평균하여 4개 모형 평균 전망점수 시계열을 형성하였다. 전망점수는 리커트 7점 척도 기준으로 부정적(2점)에서 긍정적(6점) 사이에 분포하며, 2022년 4분기에 가장 부정적인 점수가 형성되었고 2024년 3분기에 가장 긍정적인 점수가 형성되었다.

<표 11> 매매가격지수 통계 및 단위근 검정 결과

구분	최소	최대	평균	표준편차
서울 (수준변수)	74.56	104.38	91.18	8.21
서울 (1차 차분)	-2.95%	1.85%	0.23%	0.62%
구분		단위근 검정 결과		
수준변수	t값	-1.82		
	p	0.37		
	1%	-3.49		
	5%	-2.89		
1차 차분	t값	-4.16		
	p	0.00***		
	1%	-3.49		
	5%	-2.89		

주 : *** p<0.01.



〈그림 3〉 매매가격지수 월간 변화율(좌측 축) 및 모형 평균 전망점수(우측 축, 1개월 후행) 그래프

2. 교차상관분석

교차상관분석 수행 시 t 시점의 매매가격 변화율(전월 대비 가격 증감)과 t 시점까지의 기사로 전망한 전망점수를 시차 0으로 비교하였다. 예를 들어, 모형 1, 2의 경우 2023년 12월의 가격과 2024년 1월의 가격을 비교하여 2024년 1월 기준의 1개월 수익률을 작성하였다. 그리고, 이 수익률을 2024년 1월에 수집한 기사로 생성한 RAG-LLM의 전망점수와 비교하는 것을 시차 0으로 설정하여 선행 및 후행시차 6개월에 대한 교차상관관계 분석을 수행하였다.

프롬프트의 질의를 3개월 이후의 가격 전망으로 설정한 모형 3, 4의 경우 기사 집계 시점 이후 3개월 동안의 수익률과 분석을 진행하였다. 이때, 2023년 12월의 가격과 2024년 3월의 가격을 비

교하여 2024년 1월 기준의 3개월 수익률을 작성하였다. 그리고, 이 수익률을 2024년 1월에 수집된 기사로 구축된 벡터 DB를 활용하여 RAG-LLM을 통한 전망점수를 비교하는 것을 시차 0으로 설정하였다.

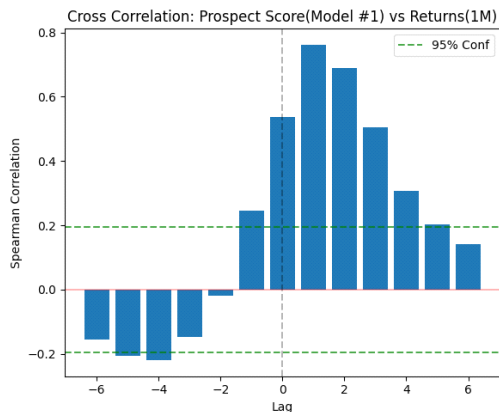
모형 1에 대해 실시한 교차상관관계 분석의 결과는 〈표 12〉 및 〈그림 4〉와 같다. 모형 1의 교차상관분석 결과는 시차 -1부터 +5까지 매매가격지수의 변화율과 전망점수 사이에 유의한 상관관계가 있는 것으로 나타난다.

특히, RAG-LLM 기준으로 긍정적인 전망이 집계된 다음 달에 가격 변화율에 대한 상관관계가 가장 높게 나타나며, 시차가 양수인 구간에서 높은 양의 상관관계를 가진다. 다시 말해 전망점수가 선행성을 지니는 것을 볼 수 있으며, 이는 전망점수를 활용한 유의한 방향성 예측이 가능함을 의미한다.

〈표 12〉 모형 1의 교차상관분석 결과(1개월)

시차	Spearman corr.		Pearson corr.	
	Lead	Lag	Lead	Lag
0	0.54		0.51	
1	0.75	0.24	0.69	0.24
2	0.66	-0.04	0.66	0.02
3	0.47	-0.18	0.54	-0.13
4	0.29	-0.25	0.42	-0.20
5	0.20	-0.23	0.31	-0.20
6	0.15	-0.17	0.23	-0.16

주 : 1) 95% 유의수준은 0.178, 99% 유의수준은 0.235.
 2) 매매가격지수의 변화율은 2015년 1월부터 2024년 12월까지의 120개 구간 변화를 활용.



〈그림 4〉 모형 1의 전망점수와 1개월 매매가격 변화율의 교차상관관계 분석 그래프

교차상관분석 결과는 참고한 선행연구인 박재수·이재수(2021)에서 제시한 신문 기사 기반의 감성지수와 매매가격지수 사이에 교차상관관계가 대칭에 가깝게 나타난다는 점에서는 비슷하나, 전망을 기반으로 RAG-LLM을 구축한 특징으로 인해 음의 시차를 지닌 구간에 비해 양의 시차

를 지닌 구간에서 더 큰 상관계수를 보인다.

RAG-LLM을 기반으로 추출한 전망점수가 부동산 기사에서 과거의 가격 변동보다는 미래의 가격 변동에 더 많은 영향을 끼치는 정보를 추출했음을 알 수 있다. Zhu et al.(2022)도 소셜 미디어에서 추출한 텍스트에서 미래에 대한 전망과 관련된 감성의 변화가 매매가격 변화에 큰 예측력을 가진다는 점을 보였다. 선행 1개월 시차에 대해서 가장 높은 교차상관관계가 나타난다는 점은 동일하나, 선행연구의 교차상관계수 최대값인 0.50에 비해 본 연구의 교차상관계수는 0.75로 더 높게 나타나며 선행 2개월 시차의 교차상관계수도 0.66으로 높게 나타난다. 이는 단기적인 예측에 있어 본 연구에서 제시한 전망점수의 유용성이 더 크다는 점을 시사한다.

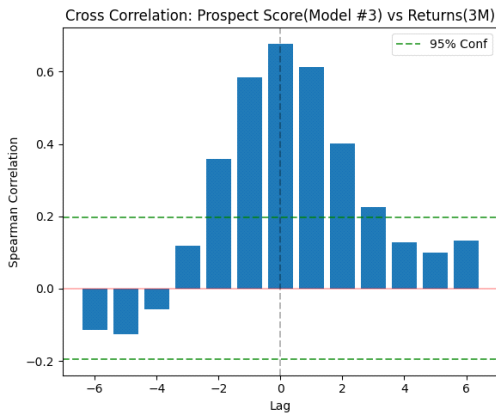
3개월 수익률을 대상으로 예측한 모형 3의 분석결과는 〈표 13〉 및 〈그림 5〉와 같다.

기사 수집 시점으로부터 3개월 뒤의 가격 변화

〈표 13〉 모형 3의 교차상관분석 결과(3개월)

시차	Spearman corr.		Pearson corr.	
	Lead	Lag	Lead	Lag
0	0.68		0.70	
1	0.61	0.59	0.69	0.58
2	0.40	0.36	0.56	0.36
3	0.22	0.12	0.42	0.14
4	0.13	-0.06	0.30	-0.01
5	0.10	-0.13	0.21	-0.09
6	0.13	-0.11	0.17	-0.11

주 : 1) 95% 유의수준은 0.178, 99% 유의수준은 0.235.
 2) 2024년 12월의 가격변화율은 2024년 11월 가격 대비 2025년 2월의 가격을 비교하여 계산.



〈그림 5〉 모형 3의 전망점수와 3개월 매매가격 변화율의 교차상관관계 분석 그래프

전망을 질의하여 산출한 전망점수와 3개월 수익률에 대한 교차상관관계 분석 결과 역시 기준시점을 포함하여 대칭으로 나타난다. 다만, 3개월 수익률은 더 먼 미래시점에 대한 예측인 만큼 선행성 자체는 줄어드는 것으로 나타난다. 시차 -2부터 +2까지 유효한 수준의 상관관계가 나타나며, 전망점수 생성 시점 이후인 시차 +1 및 +2구간에서 3개월 수익률에 유의미한 선행성을 나타냈다.

제시된 모형의 교차상관관계를 포함하여, 전체 모형에 대한 교차상관분석 결과는 〈부록〉에 수록하였다.

3. ARIMAX(Autoregressive Integrated Moving Average with Exogenous Variable) 모델을 통한 분석결과

ARIMAX 분석을 위해 먼저 매매가격지수 시계열 구간에 대한 준거모형을 확인하였다. 김동환(2023)에서는 Box-Jenkins 방법에 따라 한국

부동산원이 집계한 전국 아파트 매매가격 지수를 대상으로 2011년부터 2023년까지의 시계열을 검증하였고, 상수항이 없는 ARIMA(2,1,1) 및 ARIMA(2,1,0) 모형이 모두 타당한 모형임을 보였다. 이에 기초하여 본 연구에서도 2015년부터 2024년까지의 시계열에 대한 ARIMA 최적 모형을 탐색하였다. 탐색 결과 및 최적 모델의 세부 추정 결과는 〈표 14〉와 같이 나타난다. 추정 결과 ARIMA(2,1,0) 모형이 가장 낮은 AIC 지표를 가

〈표 14〉 ARIMA 모형 후보 및 최적 모델 추정 결과

ARIMA(p, q, d)	AIC	BIC
(1, 1, 0)	71.86	77.42
(1, 1, 1)	64.51	72.85
(1, 1, 2)	64.44	75.56
(1, 1, 3)	66.44	80.34
(2, 1, 0)	62.63	70.97
(2, 1, 1)	64.54	75.65
(2, 1, 2)	66.40	80.30
(2, 1, 3)	68.39	85.07
(3, 1, 0)	64.51	75.63
(3, 1, 1)	66.40	80.29
(3, 1, 2)	68.40	85.08
(3, 1, 3)	68.52	87.97

구분	Coef.	Std	Z	Prob	
ARIMA (2,1,0)	AR.L1	1.134	0.078	14.312	0
	AR.L2	-0.299	0.123	-2.433	0.015
	SIGMA	0.093	0.007	12.464	0
	BIC	70.97			
	AIC	62.63			

주 : 1) p(AR 항), q(차분), d(MA 항). 1차 차분만 집계.
 2) ARIMA, autoregressive integrated moving average; AIC, Akaike information criterion; BIC, Bayesian information criterion.

지는 시계열 모형으로 나타나 준거 모형으로 활용하였다.

다음으로 RAG-LLM 기반의 조정전망점수를 외생변수로 투입하여 ARIMAX 추정을 수행하고 예측 성능을 나타내는 지표인 RMSE, MAE를 측정하였다. 예측 지표 산출 시 모델의 초기화 문제를 고려하여 처음 1개의 데이터를 제외하고 지표를 생성하였다.

예측은 전체 구간을 시계열 순서로 8대 2로 길이를 가지도록 train/test 분리 후, test 구간에 대한 one-step ahead out-of-sample 방식으로 추정치를 생성하였다. 추정 결과 상수항을 가진 ARIMAX 모형이 더 큰 개선치를 보였으며, 제시한 모형별로 예측오차의 개선을 확인하면 <표 15>와 같다.

제시한 4개 모형 모두 우수한 test 구간에서 개선된 시계열 예측 성능을 보였다. 3개월 후의 가격 전망을 질의한 3번 모형은 준거 모형 대비 10.52% 개선된 RMSE와 6.03% 개선된 MAE를 기록하였으며, 1개월 후의 가격 전망을 질의한 1번 모형은 준거 모형 대비 7.30% 개선된 RMSE와 12.96% 개선된 MAE를 기록하였다. 이는 기존 신문 감성지수 기반의 예측 연구인 박재수·이재수(2021)에서 제시한 RMSE 7.9%, MAE 6.2% 대비 소폭 개선된 결과이다.

우수한 예측 성능을 보인 모형 3의 ARIMAX 모형의 추정 결과는 <표 16>과 같다. 동 모형에서 투입된 외생변수인 조정전망점수의 설명력은 1% 이내로 유의한 수준임을(X1) 확인하였으며 준거 모형의 AIC 값(62.63) 대비 감소한 AIC 값(57.58)이 나타났음을 확인하였다.

<표 15> 모형별 주요 예측오차(ARIMAX)

외생변수 생성모형	Metric	
	RMSE	MAE
ARIMA 2,1,0 (준거모형)	0.1814	0.1217
모형 1	0.1681	0.1059
모형 2	0.1744	0.1184
모형 3	0.1623	0.1143
모형 4	0.1718	0.1183
오차 개선을 (모형 1)	-7.30%	-12.96%
오차 개선을 (모형 3)	-10.52%	-6.03%

주 : 1) 전체 구간을 8/2로 train/test로 분할 후, test 구간 one step ahead out-of-sample 예측 결과.
2) ARIMAX, autoregressive integrated moving average with exogenous variable; RMSE, root mean square error; MAE, mean absolute error.

<표 16> ARIMAX 모델 추정 결과(모형 3)

구분	Coef.	Std	Z	Prob	
모형 3 ARIMAX (2,1,0)	Intercept	0.040	0.029	1.384	0.166
	X1	0.011	0.004	3.205	0.001***
	AR.L1	1.131	0.077	14.664	0.000***
	AR.L2	-0.334	0.120	-2.784	0.005***
	SIGMA	0.086	0.007	12.715	0.000***
	BIC	71.48			
	AIC	57.58			

주 : 1) *** p(0.01).
2) ARIMAX, autoregressive integrated moving average with exogenous variable; BIC, Bayesian information criterion; AIC, Akaike information criterion.

4. 전망점수를 활용한 방향성 예측

RAG-LLM을 통해 산출한 전망점수는 기존 가

격 데이터 및 거시 데이터를 이용하는 예측 방법론에 비해 우수한 적시성을 보인다. 먼저 집계 및 처리에 시일이 걸리는 심리지표나 거시경제 지표, 통화 관련 지표에 비해 뉴스 데이터는 빠른 입수가 가능하며, RAG의 구축 및 LLM 활용도 빠르게 진행할 수 있다. 본 연구에서 활용한 매매가격지수의 경우 집계부터 공표까지 1.5개월의 시차가 존재하기에, 다음 달의 가격 변동을 위해 사용할 수 있는 데이터는 2개월 전의 매매가격지수를 최신의 데이터로 사용해야 한다. 반면, 뉴스 데이터의 집계 및 검색 DB의 구축과 LLM 호출은 빠르게 수행될 수 있는 만큼 월말 기준으로 생성된 전망점수는 다음 달 예측에 실시간에 가깝게 활용될 수 있다.

이와 같은 제약조건을 반영하여 방향성 예측을 수행한 결과는 <표 17>과 같이 나타난다. 전망 점수가 4점을 초과하면 1개월 후(모형 3, 4는 3개월

후)의 가격이 상승할 것으로 예측하였으며, 3점 이하일 경우에는 가격이 하락할 것으로 예측하여 전체 구간의 정확도 및 예측 성능 지표를 집계하였다. 모든 모형이 비교를 위한 베이스라인 ARIMA 모형보다 우수한 정확도를 보이며, Chain-of-Thought 프롬프팅 기법을 활용하여 전망점수를 산출한 모델들이 단순 RAG 프롬프트 모형 대비 소폭 우수한 성과를 보이는 것을 확인하였다.

1개월 예측의 경우 정확도가 최대 85.6%로 베이스라인 모델인 74.2%에 비해 11.4%가량 더 정확한 예측을 수행하였고, 3개월 예측의 경우에도 정확도가 최대 83.8%로 베이스라인 모델보다 10.7% 개선된 정확도를 보였다.

특히, 예측의 정확도뿐만 아니라 분류 성능을 나타내는 지표인 AUC-ROC(area under the curve - receiver operating characteristic)가 대부분 0.9 이상으로 형성되어 예측 모형의 분

<표 17> 모형별 전망 점수의 예측 정확도 비교(중립 제외)

모형 구분	Prompt type ^a	LLM ^b	예측 개수	Accuracy	Precision	Recall	F1_score	AUC-ROC	
1개월 예측 (N=120)	모형 1	RAG	GPT, Gemini	111	0.852	0.963	0.859	0.908	0.932
	모형 2	CoT		113	0.856	0.965	0.863	0.911	0.930
	베이스라인(ARIMA, 1개월 방향성)			120	0.742	0.835	0.807	0.821	0.685
3개월 예측 (N=119)	모형 3	RAG	GPT, Gemini	116	0.817	0.909	0.860	0.884	0.852
	모형 4	CoT		112	0.838	0.941	0.860	0.899	0.912
	베이스라인(ARIMA, 3개월 방향성)			119	0.731	0.845	0.789	0.816	0.670

- 주 : 1) 모형의 합산 점수가 정확히 중립(4점)인 경우 제외, 명확한 극성(긍정/부정)을 보인 경우만 예측.
 2) 1개월 베이스라인 모형은 매매가격지수 시차를(1.5개월) 고려하여 2-step forward 예측값 대비 3-step forward 예측값 증감 폭 이용. 3개월 베이스라인 모형은 2-step forward 예측값 대비 5-step forward 예측값 증감 폭 이용.
 3) 2015년 1월 시점의 베이스라인 모형의 추정을 위해 2015년 이전 데이터 추가로 공급.
 4) ^a) RAG는 단순 RAG로 프롬프트 구성, CoT는 Chain-of-Thought로 프롬프트 구성.
 5) ^b) LLM은 gpt-4o-mini(2024.07.18.), gemini-2.0-flash(2025.02.24.) 활용.
 6) LLM, large language model; AUC-ROC, area under the curve - receiver operating characteristic; RAG, retrieval augmented generation; ARIMA, autoregressive integrated moving average.

류 성능이 매우 우수하게 나타났다.

V. 결론 및 제언

본 연구에서는 비정형 빅데이터인 부동산 뉴스 기사를 활용해 RAG-LLM 시스템을 구축하고 2015년부터 2024년 구간에 대해 모형이 생성한 리커트 척도 기반의 전망점수의 예측 성능을 검증하였다. LLM의 여러 편향 및 한계를 고려하여 부동산 뉴스를 활용할 수 있는 RAG-LLM 구조를 제안하였으며 세부적인 시스템 구성 방법론을 제시하였다.

RAG-LLM으로 산출한 전망점수는 선행구간에 높은 교차상관관계를 가지며, 단기적인 방향성 예측에서도 시계열 모형보다 10%가량 높은 성능을 보였다. 리커트 척도의 특성을 고려하여 변환된 전망점수를 외생변수로 활용한 ARIMAX 기반의 예측도 준거 모형 대비 우수한 예측 성능을 보였다. 유사 선행연구 대비 개선된 예측 성능을 보였으며, 이는 LLM 기반의 접근이 기존 텍스트 분석 방법론의 한계점인 개체명 인식이나 미래 감정 분리와 같은 다양한 한계를 개선하여 보다 나은 미래 지표를 생성할 수 있음을 시사한다.

RAG-LLM 모형이 산출한 전망점수는 가격 데이터에 의존하지 않고 비정형 빅데이터인 뉴스 데이터를 기반으로 생성된다. 제시된 방법을 활용하면 뉴스 데이터의 특성상 실시간에 가깝게 다음 달에 대한 유효한 예측을 수행할 수 있다는 장점이 있다. 이는 다른 지표들의 발표 전까지 선도적인 지표로써 활용할 수 있을 뿐만 아니라, 실제 투

자나 정책 결정에 있어 유용한 의사결정 지원 도구로써 활용될 수 있는 잠재성을 지닌다. 특히, 1개월 및 3개월 방향성 예측 모형의 정확도가 높게 나타나고, 분류 성능 지표인 AUC-ROC 역시 매우 우수한 수준인 0.9 이상을 기록하며 부동산 빅데이터를 활용한 예측의 유용성을 보였다.

본 연구에서 한계점은 다음의 두 가지로 정리할 수 있다. 첫 번째로, 수집한 부동산 기사의 언론사를 경제신문으로만 한정하였다. 일간지와 경제신문의 논조가 다를 뿐만 아니라, 일간지별로도 서로 다른 성향을 보인다. 또한, 최근 중요한 빅데이터의 원천으로 부상한 소셜 미디어 데이터에 대한 고려를 수행하지 않았다. 데이터 원천을 분화하고 특화하여 생성한 RAG-LLM을 구축하고 분석을 진행할 경우, 본 연구에서 제시된 결론보다 더 다양한 결론이 제시될 것으로 기대한다.

두 번째로, 문서 검색 단계에서의 추가적인 연구 확장이 필요하다. RAG를 주제로 삼은 연구에서는 문서 검색 방법론의 다양화와 질의의 확장을 통해 응답의 품질을 개선할 수 있는 사례가 다수 보고되었다. 본 연구에서도 검색 단계의 주요한 변수를 바꿔가며 검색 방식에 따른 연구의 강건성을 검증할 필요가 있으며, LLM 자체를 질의와 검색 단계에 활용하는 생성형 Reranker와 같은 방안을 검토하면 비정형 데이터로부터의 지식 추출이 더욱 고도화될 수 있을 것으로 전망한다. 본 연구와 같이 부동산 분야의 비정형 데이터를 RAG-LLM 방법론을 이용하여 다양한 데이터와 주제에 적용한 연구가 지속되기를 기대한다.

ORCID 

김진우 <https://orcid.org/0009-0009-7352-4669>

은준엽 <https://orcid.org/0000-0002-4970-8224>

참고문헌

1. 경정익, 이국철. (2016). Textmining에 의한 부동산 빅데이터 감성분석 모형 개발. *주택연구*, 24(4), 115-136.
2. 김대원, 유정석. (2013). 주택가격에 대한 심리적 태도가 주택 매매 거래량에 미치는 영향 분석. *주택연구*, 21(2), 73-92.
3. 김동환. (2023). ARIMA 모형을 이용한 전국 아파트 시장 가격 예측. *대한부동산학회지*, 41(4), 5-27.
4. 김수아, 권미주, 조수빈, 김은수, 김현희. (2023). 생성 AI 기반 뉴스 기사 심리지수를 활용한 부동산 가격 예측 모델. *한국정보처리학회 학술대회논문집*, 30(2), 1198-1199.
5. 박재수, 이재수. (2019). 아파트 매매가격과 부동산 온라인 뉴스의 교차상관관계와 인과관계 분석: 온라인 뉴스 기사의 비정형 빅데이터를 활용한 감성분석 기법의 적용. *국토계획*, 54(1), 131-147.
6. 박재수, 이재수. (2021). 부동산 감성지수의 주택가격 예측 유용성: 뉴스기사와 방송뉴스 빅데이터 활용 사례. *국토계획*, 56(4), 99-111.
7. 서정석, 오지훈, 김정섭. (2021). 부동산 뉴스와 주택 경기의 동적 관계에 대한 고찰: 언론보도노조지수 개발을 중심으로. *한국지역개발학회지*, 33(1), 89-112.
8. 신한은행. (2024). *보통사람 금융생활 보고서 2024*. 신한은행.
9. 이원수. (2007). 한국 경제뉴스의 속성(attributes) 프레임효과 연구. *언론과 사회*, 15(1), 86-122.
10. 이창무. (2020). 문재인 정부 부동산정책의 비판적 평가. *한국행정연구*, 29(4), 37-75.
11. 주종용, 권영상. (2023). 주택가격 상승기의 부동산 규제정책이 투자심리와 주택가격에 미치는 영향: 서울시를 중심으로. *부동산분석*, 9(2), 1-31.
12. 진창하. (2012). 신문기사 내용과 주택가격: 인식, 사유, 그리고 투자심리. *부동산학연구*, 18(2), 125-142.
13. 행정안전부. (2024). *2023년 주민등록 인구통계*. 행정안전부.
14. 홍지수. (2023). 부동산 비기초가격과 부동산 신문 담론의 시계열 관계 분석: 서울 아파트 매매가를 중심으로. *국토계획국토계획*, 58(3), 117-136.
15. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). Association for Computing Machinery.
16. Blackwell, R. E., Barry, J., & Cohn, A. G. (2024). *Towards reproducible LLM evaluation: Quantifying uncertainty in LLM benchmark scores*. <https://arxiv.org/abs/2410.03492>
17. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). *Language models are few-shot learners*. <https://arxiv.org/abs/2005.14165>
18. Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering

- documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 335–336). Association for Computing Machinery.
19. Carraro, D., & Bridge, D. (2024). *Enhancing recommendation diversity by re-ranking with large language models*. <https://arxiv.org/abs/2401.11506>
 20. Chen, T., & Si, S. (2024). *Predicting rental price of lane houses in Shanghai with machine learning methods and large language models*. <https://arxiv.org/abs/2405.17505>
 21. Gan, J. (2021). *ko-sbert-nli: Korean Sentence-BERT trained with NLI data*. <https://huggingface.co/jhgan/ko-sbert-nli>
 22. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., & Wang, H. (2023). *Retrieval-augmented generation for large language models: A survey*. <https://arxiv.org/abs/2312.10997>
 23. Gloria, B., Melsbach, J., Bienert, S., & Schoder, D. (2024). Real-GPT: Efficiently tailoring LLMs for informed decision-making in the real estate industry. *Journal of Real Estate Portfolio Management*, 31(1), 1–17.
 24. Laban, P., Fabbri, A. R., Xiong, C., & Wu, C. S. (2024). *Summary of a haystack: A challenge to long-context LLMs and RAG systems*. <https://arxiv.org/abs/2407.01370>
 25. Leng, Q., Portes, J., Havens, S., Zaharia, M., & Carbin, M. (2024). *Long context rag performance of large language models*. <https://arxiv.org/abs/2411.03538>
 26. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-augmented generation for knowledge-intensive NLP tasks*. <https://arxiv.org/abs/2005.11401>
 27. Lu, Y. L., Zhang, C., & Wang, W. (2025). *Systematic bias in large language models: Discrepant response patterns in binary vs. continuous judgment tasks*. <https://arxiv.org/abs/2504.19445>
 28. Martino, A., Iannelli, M., & Truong, C. (2023). Knowledge injection to counter large language model (LLM) hallucination. In *The Semantic Web: ESWC 2023 Satellite Events* (pp. 182–185). Springer Nature.
 29. Mian, A., & Sufi, A. (2014). What explains the 2007–2009 drop in employment? *Econometrica*, 82(6), 2197–2223.
 30. Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence embeddings using siamese BERT-networks*. <https://arxiv.org/abs/1908.10084>
 31. Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., Li, Y., Gupta, A., Han, H., Schulhoff, S., Dulepet, P. S., Vidyadhara, S., Ki, D., Agrawal, S., Pham, C., Feilean Li, G. K., Tao, H., Srivastava, A., Da Costa, H., Gupta, S., Rogers, M. L., Goncarenco, I., Sarli, G., Galynker, I., Peskoff, D., Carpuat, M., White, J., Anadkat, S., Hoyle, A., & Resnik, P. (2024). *The prompt report: A systematic survey of prompting techniques*. <https://arxiv.org/abs/2406.06608>
 32. Shultz, T. R., Wise, J. M., & Nobandegani, A. S. (2025). Text understanding in GPT-4 versus humans. *Royal Society Open Science*, 12(2), 241313.
 33. Soo, C. K. (2018). Quantifying sentiment with news media across local housing markets. *The Review of Financial Studies*, 31(10), 3689–3719.
 34. Wang, X., Wang, Z., Gao, X., Zhang, F., Wu, Y., Xu, Z., Shi, T., Wang, Z., Li, S., Qing, Q., Yin, R., Lv, C., Zheng, X., & Huang, X.

- (2024). *Searching for best practices in retrieval-augmented generation*. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 17716–17736). Association for Computational Linguistics.
35. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). *Chain-of-thought prompting elicits reasoning in large language models*. <https://arxiv.org/abs/2201.11903>
36. White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with chatGPT*. <https://arxiv.org/abs/2302.11382>
37. Yang, S., Zhao, Y., & Gao, H. (2024). *Using large language models in real estate transactions: A few-shot learning approach*. https://osf.io/preprints/osf/kb9cu_v1
38. Zhang, B., Yang, H., Zhou, T., Ali Babar, M., & Liu, X. Y. (2023). Enhancing financial sentiment analysis via retrieval augmented large language models. In *ICAF '23: Proceedings of the Fourth ACM International Conference on AI in Finance* (pp. 349–356). Association for Computing Machinery.
39. Zhu, E., Wu, J., Liu, H., & Li, K. (2022). A sentiment index of the housing market in China: Text mining of narratives on social media. *The Journal of Real Estate Finance and Economics*, 66(1), 77–118.

논문 접수일: 2025년 4월 29일
 심사(수정)일: 2025년 6월 3일
 게재 확정일: 2025년 7월 18일

국문초록

본 연구는 검색증강생성(retrieval augmented generation, RAG) 기반의 대규모 언어모델(large language model, LLM)이 부동산 분야의 비정형 빅데이터를 효과적으로 처리하여 유용한 전망을 생성함을 보였다. LLM을 전문적인 외부 지식과 연결하는 RAG는 환각 현상을 줄이고 신뢰할 수 있으면서 분야에 특화된 답변을 생성할 수 있다. 연구에서는 비정형 빅데이터인 부동산 뉴스 기사를 활용하여 RAG-LLM을 구축하고 서울 아파트매매가격 변동에 대한 전망 점수를 생성하였다. 2015년부터 2024년까지 10년에 걸쳐 생성된 전망 점수는 가격 변화에 선형성을 보였고, 가격 방향성 예측에서 시계열 모형 대비 11% 향상된 85%의 정확도와 93%의 AUC-ROC 성능을 보였다. RAG 과정에서 추론(chain-of-thought) 프롬프팅을 이용하면 더 우수한 품질의 답변을 얻을 수 있었고, 시계열 모형의 외생변수로 활용할 때도 예측 성능의 개선을 보였다. 연구 결과는 RAG-LLM 방법론이 비정형 데이터로부터 복잡한 맥락을 이해하고 지식을 추출할 때 기존 감성분석 방법론의 한계를 극복할 수 있고, 데이터를 유용하게 활용하기 위한 돌파구를 제공할 수 있음을 시사한다. 또한, 이 접근법은 시차 지연 없이 다음 기의 시장 예측을 형성함으로써 투자나 정책 의사결정에 유용한 도구로 활용될 잠재성을 지닌다.

주제어 : 빅데이터, 주택 가격 예측, 검색증강생성(retrieval augmented generation, RAG), 부동산 뉴스

부록

〈표 A-1〉 모형별 교차상관관계

시차	모형 1	모형 2	모형 3	모형 4
-6	-0.167	-0.155	-0.115	-0.130
-5	-0.230	-0.206	-0.126	-0.147
-4	-0.246	-0.220	-0.058	-0.080
-3	-0.176	-0.147	0.118	0.094
-2	-0.036	-0.020	0.358	0.343
-1	0.242	0.246	0.585	0.581
0	0.537	0.538	0.678	0.685
+1	0.749	0.762	0.612	0.635
+2	0.658	0.690	0.401	0.441
+3	0.466	0.506	0.225	0.267
+4	0.293	0.309	0.127	0.160
+5	0.202	0.204	0.099	0.126
+6	0.154	0.142	0.132	0.145

주 : 1) Spearman 상관계수, 시차 0은 월초부터 월 말까지 집계한 뉴스데이터와 해당 월의 수익률을 비교.

2) 95% 유의수준은 0.178, 99% 유의수준은 0.235.