

# 전세보증사고 위험과 거시경제를 고려한 주택가격 예측\*

## Forecasting Housing Prices with Consideration of Jeonse Guarantee Incidents Risk and Macroeconomic Conditions

김 은 미 (Kim, Eun-Mi)\*\*

### < Abstract >

This study forecasts apartment prices by incorporating jeonse guarantee incidents and macroeconomic factors. Using an LSTM-based time series model with Recursive multi-step prediction, it analyzes future trends in housing prices. SHAP analysis quantifies the contribution of each variable and visually interprets what the model focuses on.

Key variables include mortgage rates, housing completion index, industrial production, construction costs, mortgage loans, jeonse guarantee incidents, and apartment price changes, using monthly data from Jan 2015 to Aug 2023.

The results show short-term price fluctuations within a mild upward trend. Mortgage rates had the highest influence, while jeonse guarantee incidents had limited impact—likely due to model limitations or complex social factors. Continued research may help stabilize the housing market.

Keyword : Jeonse Deposit Guarantee Incidents, Housing Prices, Long-Short Term Memory, Shapley Additive Explanations, Recursive Multi-Step

## I. 서론

최근 국내 전세시장의 변동성이 높아지면서 임대차 계약 만료 또는 해지 시 임대인이 임차인에게 반환해야 할 전세보증금<sup>1)</sup>이 지연되는 등 미반환 위험이 확대되고 있다. 임대차계약상 채무불이행을 뜻하는 전세보증금 미반환은 계약당사자 간의 문제를 넘어서 후속 임대차에도 영향을 끼치는 등 부동산 시장의 심각한 사회문제로 대두된다<sup>2)</sup>. 이러한 가운데 정부는 '13년

전세자금대출보증기관(이하 보증기관<sup>3)</sup>)을 통해 「전세보증금반환보증」제도를 마련하여 임차인의 전세보증금 미반환 위험을 해소하고, 주거안정을 실현하고자 하였다.

전세보증금반환보증이란 전세계약<sup>4)</sup> 종료 후 임대인이 임차인에게 반환해야 하는 전세보증금의 반환을 보증기관이 책임지는 보증상품을 말한다.

기본적으로 전세계약은 임대인과 임차인, 두 주체간에 이루어지는 계약이다. 그러나 임차인이 보증제도에 가입함으로써 보증기관이 참여하게되고 거래구조가

\* 본 연구는 2024년 하반기 부동산분석학회의 발표논문을 수정·보완한 것임을 밝힌다.

\*\* 본 학회 정회원, 주택도시보증공사 주택도시금융연구원 연구위원, kem531@khug.or.kr, 주저자

1) 임차인이 임대인에게 지급하고 전세계약 종료시 반환받을 금전으로서 전세계약서에 기입된 금액을 말한다.

2) 주택금융리서치(2023) 한국금융주택공사

3) 전세자금대출 보증기관은 주택도시보증공사, 한국주택금융공사, 서울보증보험이 있다.

4) 임대인이 임차인에게 임차목적물을 사용·수익할 수 있게 하고, 임차인이 그 대가로서 임대인에게 전세보증금을 지급하기로 하는 계약(전세보증금이 있는 월세계약을 포함)을 말한다.

변화한다. 그렇다면 위와 같은 보증제도 구조에서 전세보증금 미반환이 발생한다면 어떻게 될까?

이후 절차를 요약하자면, ‘보증기관이 보증채무 이행과 동시에 임차인의 전세보증금 반환채권 양수 및 임대인에게 채권양도 사실 통지하는 것’이 첫 순서다.

결국 임차인은 보증기관과의 전세보증금 반환채권 양수계약을 통해 전세보증금 미반환 위험에서 벗어날 수 있지만, 그 위험은 ‘소멸’이 아닌 보증기관에게로 ‘전가’<sup>5)</sup>(Risk Transfer)되는 것이다. 본 연구의 동기는 ‘소멸되지 않고 전가된 이 위험이 주택시장에 발현 되는가?’라는 의문에서부터 시작한다.

주택도시보증공사(Korea Housing & Urban Guarantee Corporation, HUG)에 따르면 전세보증금반환보증 실적(‘13년 ~)과 사고금액(‘15년 ~)은 지속적으로 증가하여 ‘23년 실적 717,676억 원, 사고금액 43,347억 원으로 나타났다<sup>6)</sup>. 이는 임차인이 보증기관에게 지속적으로 전세보증금 미반환 위험을 전가하고 있으며 전세 계약 종료 이후 임대인의 보증금 미반환으로 처리되어 보증기관이 부담하는 전세보증사고 위험의 크기가 증가함을 뜻한다.

지금까지 전세보증사고에 관한 다양한 연구가 진행되었으나 대부분은 주택가격의 변동성과 거시경제와 같은 외부 요인에 중점을 두고 역전세, 갭통전세 등 전세보증사고의 발생 원인을 다루어왔다.

그러나 본 연구는 전세보증사고 위험을 고려하여 주택가격을 예측하고, 전세보증사고 위험이 이러한 예측 결과에 어느 정도 기여하는지 다양한 측면에서 분석한다. 특히 아파트의 거래 비중이 가장 높고, 주거 공간 중 영향력이 가장 크기 때문에 아파트 가격이 주요 관심 사항인 점<sup>7)</sup>등에 착안하여 연구의 종속변수인 주택 가격은 아파트 매매가격지수로 집중하였다. 아울러 LSTM 자체는 일반적으로 단일 시점의 예측에 최적화된 구조이므로, 예측 구조 설계를 Recursive Multi-step

으로 확장했다는 점에서 분석방법의 차별성을 두었다. 추가하자면 LSTM의 시계열 학습 능력과 Recursive 구조의 단순성이 결합될 경우, 비교적 적은 연산 자원과 학습 데이터만으로도 다수 시점에 대한 예측 수행이 가능하다는 장점이 있다. 특히 예측 시점의 수가 상대적으로 짧은 경우에는 Recursive Multi-step 방식은 보다 실용적이다.

기존 주택 가격예측에 관한 연구는 시계열 분석(Autoregressive Integrated Moving Average, ARIMA)과 머신러닝(Machine Learning, ML), 딥러닝(Deep Learning, DL)을 활용한 연구방법으로 구분된다. 시계열 분석모형과 머신러닝을 이용하여 주택가격 예측을 비교한 이전 연구에서는 머신러닝과 딥러닝을 이용한 모형의 예측력이 시계열 분석모형보다 더 우수하고, 머신러닝 모형에 의해 추정된 예측값은 실제값과 상당히 유사한 움직임을 보이는 것을 확인하였다(김은미, 2022; 이지영·유재필, 2021; 전해정, 2021; 이정현, 2023).

그러나 인공지능경망을 통한 예측은 블랙박스 성질(Black-box property)<sup>8)</sup>로 인해 인과관계를 규명하는데 어려움이 있다<sup>9)</sup>. 이를 극복하기 위해 설명 가능한 인공지능(eXplainable Artificial Intelligence, XAI)이 점차적으로 주목받고 있으며 그 중 활용성이 높은 SHAP(SHapley Additive Explanations) 분석을 통해 변수의 기여도를 평가하고 정량적으로 그 영향<sup>10)</sup>력을 설명하고자 한다.

종합하자면 본 연구는 1. 주택가격에 대한 거시경제·전세보증사고 입력 데이터를 구축하여 LSTM 기반 시계열 예측 구조를 설계하고, Recursive Multi-step을 활용하여 향후 아파트 매매가격지수 변동률을 예측한다. 2. SHAP 분석을 통해 딥러닝 모형 내부에서의 변수별 예측 기여도를 정량적으로 추정하고, 모델이 어떤 입력 정보를 중심으로 예측을 수행하는지를 시각적

5) 박창복(1993)은 위험의 전가는 사고 발생에 따른 재무적 손실의 복구를 대외적으로 제3자에게 보장받는 것이며, 특정 사고의 발생에 대한 궁극적 책임은 전가되지 않고 재무적 손실의 복구 책임만이 전가되는 것으로 정의하였다. 이를 보험에 의한 전가와 보험 이외의 전가로 구분하였으며, 보험 이외의 전가로는 임대차 계약, 건설 공사 계약, 수주 계약, 무해 협약, 보증, 헛징 등으로 설명하였다.

6) 주택도시보증공사(2023)의 업무통계연보에 따르면 전세보증금반환보증 실적은 2013년부터 집계되었으며, 전세보증금반환보증 사고금액은 2015년부터 집계되었다. 2013년 실적은 765억 원, 2015년 사고금액은 1억 원이었으나 이후 지속적으로 급등패턴을 보였다. 구체적인 연도별 실적, 사고금액 현황은 2장에서 확인할 수 있다.

7) 김경민 (2018) 서울아파트시장과 거시경제변수 요인들 간 동학적 상관관계 분석

8) 블랙박스 성질(Black-box property)이란 딥러닝 모델이 어떤 입력 데이터를 받아 특정 예측을 하더라도 그 예측에 미친 구체적인 요인이나 과정에 대해 명확한 설명이 어렵다는 점을 말한다.

9) 이정현(2023) 인공지능을 이용한 주택가격 변동성 예측 모델 연구

10) 앞서 기술한 바와 같이 본 연구에서 ‘영향’은 인과관계를 뜻하는 것이 아님을 재차 밝힌다.

으로 해석하여 3. 향후 주택시장 안정화 방안 제시를 목적으로 한다.

본 논문의 구성은 총 5장으로 구성되어 있다. 제2장은 전세보증제도 및 사고, 주택가격에 관한 이론적 배경과 기존 연구를 검토할 것이다. 제3장에서는 연구에 적용한 모형 및 자료에 대해 설명하고, 제4장에서는 연구의 분석 결과를, 5장은 결론 및 시사점을 제시할 것이다.

## II. 이론적 배경

### 1. 전세보증금반환보증

전세보증금반환보증 제도는 '95년 「주택임대차신용보험」으로 국내에 처음 도입되어 대한보증보험 및 한국보증보험(SGI, 서울보증보험 전신)에서 취급하였다<sup>11)</sup>. 당시에도 임대차 계약 당사자 간 보증금 미반환에 대한 위험을 줄이고자 미반환을 임대인의 채무불이행으로 보고 이에 대한 신용보험을 제공한 것이다<sup>12)</sup>.

전세보증금반환보증은 주택임대보증 상품 중 하나이며 각 보증기관은 가입 주체에 따라 다양한 상품을 보유하고 있어 주거안정을 도모한다<sup>13)</sup>.

앞서 언급한대로 전세보증금반환보증은 임대인의 보증금 미반환 위험에 노출된 임차인을 보호하고자 전 세계약 종료 후 반환해야 할 전세보증금 지급의무를 보증기관이 책임지는 보증상품을 말한다.

이는 <그림 1>과 같이 보증계약을 체결함으로써 임차인과 임대인 외에 보증기관이 계약 주체로 참여하게

된다.

전세보증금반환보증 발급에 있어 보증기관이 가장 중요하게 고려하는 취급조건은 임차인의 전세보증금 반환채권에 대한 담보 취득이다. 보증기관은 임차인의 보증 신청이 접수된 때에 보증 심사(보증대상, 금액, 기간, 한도 등 보증신청 내용의 적정성 판단)를 통해 승인여부를 결정한 후 임차인(보증신청인)에게 통보·보증서 발급한다. 보증서 발급과 동시에 임차인은 보증료를 수납하고, 전세보증금을 돌려받을 수 있는 권리인 전세보증금반환채권을 보증기관에 양도한다.

위와 같은 과정을 통해 전세보증금의 주채권자는 임차인에서 보증기관으로 변경된다<sup>14)</sup>.

<표 1>에 따르면 전세보증금반환보증 공급 실적은 '13년 451건, 765억 원에서 시작하여 이후 급격히 증가해 '23년에는 314,456건, 71조 2,676억 원으로 집계된다. 이는 발급 기간 동안 누적 보증 실적이 973,421건, 210조 4,643억 원에 이르는 수준이다. 특히 '13년 대비 '23년의 보증 실적은 건수 기준으로 약 697배 증가, 금액 기준으로 약 93,097배 증가하여 보증 공급이 폭발적으로 확대된 것을 확인할 수 있다.

보증 잔액도 실적과 흐름을 같이 한다. 마찬가지로

<표 1> 연도별 전세보증금반환보증 실적 및 잔액  
(단위: 건, 억 원)

| 연도   | 보증 공급 실적 |           | 보증 잔액     |           |
|------|----------|-----------|-----------|-----------|
|      | 건수       | 금액        | 건수        | 금액        |
| '13년 | 451      | 765       | 450       | 763       |
| '14년 | 5,884    | 10,586    | 6,269     | 11,241    |
| '15년 | 3,941    | 7,221     | 9,516     | 17,301    |
| '16년 | 24,460   | 51,716    | 27,702    | 57,608    |
| '17년 | 43,918   | 94,931    | 67,163    | 144,043   |
| '18년 | 89,351   | 190,367   | 134,448   | 287,293   |
| '19년 | 156,095  | 306,444   | 243,721   | 493,930   |
| '20년 | 179,374  | 372,595   | 315,602   | 637,904   |
| '21년 | 232,150  | 515,508   | 393,162   | 850,481   |
| '22년 | 237,797  | 554,510   | 456,999   | 1,047,641 |
| '23년 | 314,456  | 712,676   | 537,775   | 1,237,635 |
| 누계   | 973,421  | 2,104,643 | 1,655,032 | 3,548,205 |

출처: 주택도시보증공사 업무통계연보(2023), 재작성

<그림 1> 전세보증금반환보증 기본구조



출처: 저자작성

11) 현재는 주택도시보증공사(HUG), 한국주택금융공사(HF), 서울보증보험(SCI)에서 전세보증금반환보증을 제공하고 있다.

12) 박희우(2021) 전세보증보험 현황과 시사점

13) 주택도시보증공사는 전세보증금반환보증(임차인), 임대보증금보증(임대인), 한국주택금융공사는 일반전세지킴보증(임차인), 특례전세지킴보증(임차인, 임대인), 서울보증보험은 전세금보장신용보증(임차인), 임대주택보증보험(임대인)을 취급한다.

14) 이후 보증기관은 보증기간 동안 보증 기한 및 해지 관리(사후관리)를 수행한다.

'14년 보증 잔액 건수와 금액은 전년 대비 가장 높은 증가율인 1,293%('13년 450건 → '14년 6,269건), 1,373%('13년 763억 원 → '14년 1조 1,241억 원)를 각각 나타냈다. 이와 같이 전세보증금반환보증 제도는 도입 초기부터 급격한 보증 수요 확대를 나타냈으나, 이는 전세계약에 따른 전세보증금 미반환 위험이 임대인에게서 보증기관에게로의 위험전이 되었음을 확인하는 지표로도 쓰일 수 있다.

전세계약 만료 시점에 임대인이 보증금을 임차인에게 정상적으로 반환하면 전세보증금반환보증 계약은 그대로 종료된다. 그러나 전세보증금 미반환일 경우, 전세보증사고로 정의되며 보증기관은 임차인에게 보증금을 우선 지급하고, 추후 전세보증금반환채권을 근거로 임대인에게 미회수 된 보증금 구상권을 청구하여 회수하는 절차를 따른다.

이를 확인하기 위해 <표 2>의 연도별 전세보증사고 추이를 살펴보면, 전세보증사고는 단순히 특정 주택 유형에서 발생하는 문제가 아닌, 전세시장의 구조적 변화와 정책적 요인 등에 의해 영향을 받으며 장기적으로 증가하였음을 명확히 파악할 수 있다.

전세보증사고는 '15년 1건 발생으로 시작하여 '23년 총 19,350건 발생하였다. 약 11년 기간 동안 급격한 증가로 인해 주택시장 전반에 위험요소로 작용하고 있다. '16년 보증사고는 전년대비 2,600% 증가하였으나 실제로는 26건 증가했기 때문에 전년대비 증가율이 큰

<표 2> 전세보증사고 통계

(단위: 건, 억 원)

| 연도   | 보증사고(건, 억 원) |        |        |        |
|------|--------------|--------|--------|--------|
|      | 건수           | 전년대비   | 금액     | 전년대비   |
| '13년 | -            | 0%     | -      | 0%     |
| '14년 | -            | 0%     | -      | 0%     |
| '15년 | 1            | 0%     | 1      | 0%     |
| '16년 | 27           | 2,600% | 34     | 3,300% |
| '17년 | 33           | 22%    | 74     | 117%   |
| '18년 | 372          | 1,027% | 792    | 970%   |
| '19년 | 1,630        | 338%   | 3,442  | 334%   |
| '20년 | 2,408        | 48%    | 4,682  | 36%    |
| '21년 | 2,799        | 16%    | 5,790  | 23%    |
| '22년 | 5,443        | 94%    | 11,726 | 103%   |
| '23년 | 19,350       | 256%   | 43,347 | 270%   |
| 누계   | 32,063       | -      | 69,888 | -      |

출처: 주택도시보증공사 업무통계연보(2023), 재작성

의미가 없다. 그러나 '19년에는 전년대비 1,258건이 늘어나면서 전세보증사고가 본격적으로 확대되기 시작했다. 이후 빠른 속도로 증가하여 '23년은 전년 대비 약 14,000건이 추가되며 역대 최대 보증사고 금액인 43,347억 원을 기록하였다. 특히 '23년의 보증사고 급증은 전세보증사고 문제가 단순히 개별 사건이 아닌 시장 전반의 위험으로 확대되었음을 보여준다.

## 2. 균형가격 이론

일반재화가 거래되는 시장과 마찬가지로 주택시장도 수요·공급 이론이 적용된다. 시장의 수요곡선과 공급곡선이 교차하여 균형 상태에 도달했을 때 형성되는 가격 수준인 시장 균형가격(Equilibrium price)은 다음과 같이 표현할 수 있다.

$$E_t = f(S_t, D_t) \quad \text{식 (1)}$$

균형가격( $E_t$ )은 다양한 수요변수( $D_t$ )와 공급변수( $S_t$ )에 의해 형성된다. 일반적으로 국민소득, 물가수준 등과 같은 거시 경제적 외생변수들은 주택의 수요에 직·간접적으로 영향을 주게 되어 가격과 거래량을 변화시키고 지가상승률, 주택 준공 실적, 공사비 등은 주택 공급에 영향을 미치는 것으로 볼 수 있다<sup>15)</sup>.

본 연구는 이와 같은 균형가격 이론을 기반으로 수요·공급 변수 외 전세보증사고 변수를 포함한다.

$$P_t = f(S_{it}, D_{jt}, V_t) \quad \text{식 (2)}$$

식 (2)의  $D_{jt}$ 와  $S_{it}$ 는 언급한대로 수요·공급 변수 벡터를 나타내고,  $V_t$ 는 전세보증사고를 나타낸다.

여기서  $P_t$ 는 본질적으로 비선형 형태를 가지는데, 이는 변수 간의 관계가 일정하지 않고 독립변수의 변화에 따라 종속변수의 변화 속도가 달라짐을 의미한다. 이를 고려한 LSTM 모델은 다음과 같다.

$$P_{t+1} = F(S_{t-k:t}, D_{t-k:t}, V_{t-k:t}) \quad \text{식 (3)}$$

식 (3)의  $t-k:t$ 는 LSTM 모델이 학습하는 과정에

서 사용하는  $t-k$ 부터  $t$ (현재 시점)까지의 시간 범위를 의미하고, 이를 입력 데이터로 활용하여 LSTM 모델은 다음 시점( $t+1$ )을 예측한다. 즉  $S_{t-k:t}$ 와  $D_{t-k:t}$ 는 과거  $k$ 시점부터 현재 시점  $t$ 까지 공급변수와 수요변수의 벡터,  $V_{t-k:t}$ 는 과거  $k$ 시점부터 현재 시점  $t$ 까지의 전세보증사고를 나타낸다.

위와 같은 독립변수를 기반으로 식 (3)은 LSTM이 학습한 비선형 함수의 향후 아파트 매매가격 ( $P_{t+1}$ )을 예측한다. 또한, 가격 결정은 수요와 공급이 독립적으로 작용할 수 있지만, 본 연구에서는 변수 간 상호작용을 통해 복합적인 영향을 미칠 수 있음을 고려한다.

연구에 적용하자면 전세보증사고와 수요·공급 변수의 상호작용이란 수요·공급 변수와 가격의 관계가 전세보증사고의 값에 따라 변화한다는 것을 뜻한다<sup>16)</sup>.

이 경우 전세보증사고가 수요·공급에 대한 효과를 조정한다고 하여 조정변수(moderator variable)라고 정의할 수 있다. 식 (4)는 본 연구의 가격을 결정하는 변수들이 상호작용 하는 과정을 수식화 한 것이다.

$$Z_t = h(S_{it} \times V_t, D_{jt} \times V_t) \quad \text{식 (4)}$$

종합하자면  $Z_t$ 는  $t$ 시점에서 아파트 매매 가격을 나타내고 공급 변수 벡터( $S_{it}$ )와 수요 변수 벡터( $D_{jt}$ ), 그리고 전세보증사고( $V_t$ )의 복합적인 상호작용을 통해 결정된다. 특히  $S$ 와  $D$ 는 각각 공급과 수요의 독립적 영향을 나타내는 동시에  $V$ 와의 상호작용을 통해 가격 형성에 비선형적이고 동시적인 영향을 미친다.

따라서 식(4)는 단순한 변수의 독립적 효과를 넘어 변수 간 상호작용의 구조적 패턴을 설명하는데 초점을 두고 있다.

본 연구에서는 이러한 이론적 틀을 바탕으로 전세보증사고와 거시경제를 적용하여 향후 아파트 매매가격을 예측하고자 한다. 이를 위해 수요 및 공급 변수 뿐만 아니라 전세보증사고와 같은 외생적 요인의 상호작용

효과를 통합하여 균형가격의 변동성을 평가할 것이다.

이러한 분석은 주택시장의 구조적 이해를 심화시키고 시장 안정화를 위한 시사점을 도출하는데 기여할 것으로 기대한다.

### III. 선행연구

#### 1. 주택가격 예측에 관한 연구

부동산에 편중된 한국의 자산구조로 인해 부동산 가격 변동은 국가·기업·가계의 경제 상황에 큰 영향을 미친다<sup>17)</sup>. 특히 주택가격의 변동성(volatility) 확대는 실물경제도 부정적인 영향을 미칠 수 있어 부동산 시장 정보 격차를 메우고 시장의 효율성 향상을 위한 주택가격 예측 연구는 다양한 방법으로 지속되어 왔다.

전해정 외(2019)는 기존 연구의 주택가격 예측 방법이 대체로 자기회귀이동평균모형 (Autoregressive Integrated Moving Average, ARIMA), 벡터자기회귀모형(Vector Autoregressive Model, VAR) 등 시계열 분석을 주로 활용하였음을 언급하였는데<sup>18)</sup>, 이는 김승현 외(2022), 이지영 외(2021), 배성완 외(2018)의 연구에서도 비슷한 견해를 확인할 수 있다<sup>19)</sup>.

저자는 비선형적인 움직임을 고려하여 딥러닝을 이용한 국내 주택가격 예측 연구를 수행하였다. 아파트 실거래가격지수에 대한 예측력을 정확도 기준으로 평가한 결과, RNN, GRU > LSTM으로 나타났다.

이보다 앞선 손정식 외(2002) 연구에서는 선행연구의 경기 예측방법을 경가지수방법과 회귀분석, 구조방정식, ARIMA, VAR 등의 계량분석방법으로 구분하였는데, 저자 또한 해당 연구의 선행연구와 마찬가지로 VAR, ARIMA모형을 활용하여 토지 및 주택 매매·전세 등 부동산 가격의 새로운 예측모형을 개발하였고, 분석 결과 실제치와 예측치 간의 괴리비와 최소자승근 오차확률을 통해 모형의 예측력을 비교하여 IMF이후,

16) 이유재(1994)에 따르면 일반적으로 독립변수인  $X_1$ 과  $X_2$ 가 상호작용한다는 것은  $X_1$ 과 종속변수  $Y$ 의 관계가  $X_2$ 의 값에 따라 변화한다는 것을 의미한다.

17) 배성완·유정석 (2018) 머신 러닝 방법과 시계열 분석 모형을 이용한 부동산 가격지수 예측

18) 전해정·양혜선 (2019) 딥 러닝을 이용한 주택가격 예측에 관한 연구

19) 김승현 외(2022)는 헤도닉 가격모형을 기반으로 한 선행회귀분석, 시계열 분석 방법이 기존연구에서 주로 활용되었다고 언급하였으며, 이지영 외(2021)는 국내 선행연구에서 주택가격지수 예측을 위해 주로 회귀분석, ARIMA, VAR 등을 이용한다고 언급하였다. 또한 배성완 외(2018)의 연구에서도 부동산시장 예측은 주로 시계열 분석 모형을 이용하여 부동산 가격지수를 예측하는 방식임을 기술하였다.

부동산 가격이 시장기본가치에 의해 큰 영향을 받는다는 점, 그 외 자산선택이론에 의한 영향도 확인되었다. 손정식 외(2002)의 동시대 연구를 포함한 이전의 선행 연구들은 주로 1990년대 이후 국가 경제의 관점에서 부동산과 거시경제 변수 간의 인과관계를 분석하여 부동산시장 예측에 중요한 연구 성과를 이끌어냈다<sup>20)</sup>. 이러한 연구에서 활용된 계량분석 방법은 각 변수의 현재값이 다른 변수들의 과거값에 의해 설명될 수 있다는 점에서, 인과관계를 파악하는데 유용하며 이를 부동산시장 예측에 효과적으로 적용되었다고 할 수 있다.

그러나 최근 연구에 의하면 시계열 분석은 현실과는 다소 동떨어진 선형(linear)모형을 가정하기 때문에 실제 주택가격의 비선형(non linear)적인 움직임을 반영하기에는 한계가 있다<sup>21)</sup>는 비판이 제기되고 있다. 이와 같은 문제를 해결하기 위해 선형모형보다 더 유연한 비선형 기법의 필요성이 강조<sup>22)</sup>되고 있으며 변수들 간 선형성과 상호독립성의 가정, 오차항의 분포에 대한 제약과 식별의 문제 등으로부터 벗어나 머신러닝과 같은 인공지능을 이용한 높은 예측력을 지닌 모형을 제시한다<sup>23)</sup>.

여기서 ‘높은 예측력을 지닌 모형’이라 함은, 계량분석 모델과 머신러닝(또는 딥러닝) 모델을 서로 비교·분석하여 도출된 선행 연구의 결과를 함축한다.

배성완 외(2018)는 머신러닝 방법과 시계열 분석 모형을 활용해 부동산 가격지수 예측모형을 구축하고 그 예측력을 검증하였다<sup>24)</sup>. 저자는 부동산 시장 상황에 따라 모형의 예측력이 상이할 수 있음을 가정하고, 분석 기간을 안정적·불안정적 시장으로 구분하였다. 그 결과 두 시장 모두 LSTM모형이 가장 높은 예측력을 나타냈는데, 특히 시장이 급변하는 시기인 불안정적 기간에서는 머신러닝 모형이 비교적 유연하게 시장 추세를 예측하는 반면 시계열 분석 모형에서는 그렇지 못함을 밝혔다. 김승현 외(2022)는 배성완의 연구와 다르게, 분석 기간을 6(단기), 12(중기), 24개월(장기)로 설정하여 연구를 진행하였다. 저자는 아파트 실거

래가 예측을 위해 패널고정효과모형, 랜덤포레스트, 다변량 적응 회귀 스플라인을 이용하였다. 그 결과, 머신러닝 모형을 활용한 예측이 패널고정효과모형을 활용한 예측보다 높은 경우는 없었으며 이에 기반한 분석결과도 마찬가지로 RF, MARS, 패널고정효과모형 순으로 예측력이 높게 나타났다. 그러나 연구의 분석자료와 변수설정에 따라 예측결과가 달라질 수 있음을 명시하였으며 머신러닝 방법론이 항상 우수한 성능을 보인다고 일반화 할 수 없음을 지적하였다.

전해정(2020)은 금융위기 전·후를 구분하여 지역별 주택가격의 시계열분석·머신러닝 모형을 구축한 후 그 예측력을 비교한 결과 BVAR < ARIMA < VAR 순으로 시계열분석모형의 주택가격 예측력이 높았다. 반면, 머신러닝모형은 전국, 서울, 강남의 RNN 예측력이 높은 것으로 나타났으나 강북은 금융위기 이후 LSTM의 예측력이 더 우수한 것으로 나타났다. 그러나 이들의 연구 역시 배성완 외(2018) 연구결과와 일치하는 흐름을 보였으며 시계열분석모형보다 비선형자료를 효과적으로 처리할 수 있는 머신러닝을 활용했을 때 예측력이 더 향상됨을 확인하였다.

이와 같이 예측력 관점에서 비교해보면 비선형 데이터를 처리할 수 있는 머신러닝 모형이 시계열분석 모형에 비해 더 높은 예측력을 보인다는 점은 명백하다.

그러나 앞서 김승현 외(2022)의 연구에서 언급한 바와 같이 머신러닝 모형이 더 우수하다고 단정짓기에는 무리가 있다. 정원구 외(2007)는 예측력이 우수하다는 장점은 있지만, 입·출력 변수 간 관계를 명확하게 설명하지 못한다는 단점을 지적하였다<sup>25)</sup>.

이와 같이 머신러닝은 높은 예측력을 가지고 있으나 인과관계를 규명할 수 없기 때문에 소위 ‘블랙박스(Black box)’모델로 불린다. 이러한 한계를 극복하기 위해 이해인 외(2023), 박서현 외(2023), 이정현(2022)은 SHAP(Shapley Additive Explanations)분석 방법을 도입하여 문제를 해결하고자 하였다.

이해인 외(2023) 연구에 따르면 SHAP분석은 모델

20) 손정식·김관영·김용순 (2002) 부동산가격 예측모형에 관한 연구

21) 이창로·박기호 (2016) 단독주택가격 추정을 위한 기계학습 모형의 응용

22) 배성완·유정석 (2017) 딥러닝을 이용한 부동산가격지수 예측

23) 이태형·정명진 (2018) 딥러닝 모형을 활용한 서울 주택가격지수 예측에 관한 연구: 다변량 시계열 자료를 중심으로

24) 배성완 외(2018)는 ARIMA, VAR, Bayesion VAR, 서포트벡터머신(Support Vector Machine, SVM), 랜덤포레스트(Random Forest, RF), 그래디언트 부스팅 회귀 트리(Gradient Boosting Regression Tree, GBRT), 심층신경망(Deep Neural Networks, DNN), LSTM(Long Short Term Memory)를 이용하였다.

25) 정원구·이상엽(2007) 인공신경망을 이용한 공동주택 가격지수 예측에 관한 연구

을 구성하는 각 특성에 예측 중요도를 할당하여 앙상블 또는 딥러닝 모델과 같이 복잡한 모델을 보다 쉽게 해석할 수 있음을 명시하였다<sup>26)</sup>. 저자는 Random Forest모델이 선형 회귀분석모형보다 높은 예측력을 지님을 확인하였고, SHAP분석을 통해 변수 중요도와 금·부정 영향을 확인하였다.

이정현(2023)은 ARIMA, VAR, LSTM모형에 적용한 후 어떤 변수가 부동산 가격에 영향을 미치는지 분석하고자 하였다<sup>27)</sup>. 특히 LSTM모델을 기반으로 수행한 SHAP분석은 변동성이 커지는 시점에서 변수 영향도와 중요도를 확인하고 상호의존성 분석을 수행하였다. 그 결과 변동성이 큰 시기에 금리와 통화량이 중요한 영향을 미치는 것으로 확인되었다. 반면 박서현 외(2023)의 연구는 좀 더 SHAP분석에 집중했다고도 볼 수 있다. 저자는 부동산 실거래가 분포와 공간적 분포(인구기준)를 4개의 그룹으로 구분하여 독립변수의 영향 정도를 알아보고자 XGBoost와 SHAP 등을 활용한 영향도 분석을 수행하였다. 그 결과 도로의 영향 및 업종·인구 영향도가 커짐을 확인할 수 있었는데, 이는 실거래가의 그룹별 분포와 중요변수의 커널밀도함수가 유사하게 나타나 향후 실거래가의 공간적 특징 분포를 예측할 수 있는 기초자료 활용에 의의를 가졌다.

이상의 주택 가격 예측에 관한 선행 연구를 종합해 보면, 예측연구는 초기 계량분석 방법에서 점차적으로 발전하여 이후 계량분석 모형과 머신러닝 모형의 예측력을 비교하는 연구가 주를 이루게 되었다.

예측력 관점에서 머신러닝 모형이 상대적으로 더 높은 예측력을 보이지만, 변수 간 인과관계를 명확히 설명하지 못한다는 한계가 있다. 이러한 한계를 극복하기 위해 최근 연구는 SHAP분석 등을 활용해 변수 간 관계를 해석하려는 시도가 이뤄지고 있다.

선행연구를 기반으로 본 연구는 상대적으로 예측력이 높은 LSTM모형을 선정하여 예측모델을 구축하고 SHAP 분석을 통해 변수의 중요도와 기여도 및 변수 간 상호작용을 심층적으로 확인하고자 한다.

다만 예측력 비교를 주된 목표로 삼기보다 멀티스텝(Multi-step) 방식을 활용하여 향후 미래 예측값을 도출하는데 분석 방법의 차별성을 둔다.

본 장에서는 주로 '예측'에 중점을 두어 선행연구를 논의하였다. 예측이란 김승현 외(2022)와 마찬가지로 '데이터 전체 기간을 학습하여 미래 주택 가격(fitted value)을 계산하는 방식'으로 정의하였다<sup>28)</sup>. 이에 따라 본 장의 선행연구는 예측값을 계산하는 분석 방법에 초점을 맞춰 전개되었다. 한편, 다음 장에서는 수요·공급 변수를 중심으로 전세보증사고와 주택 가격에 관한 선행연구를 다루고자 한다.

## 2. 전세보증사고와 주택가격 간의 연구

최근 전세보증금 미반환 위험이 증가함에 따라 보증금 미반환과 관련된 선행연구는 '깡통전세'와 '역전세'를 두고 전세보증사고의 발생원인에 집중하였다<sup>29)</sup>.

김지혜 외(2019)는 '깡통전세: 주택가격이 전세가격과 유사하여 주택이 경매에 넘어가더라도 보증금을 돌려받지 못할 위험이 있는 상태', '역전세: 전세가격이 하락하여 발생하는 현상으로 전세 계약이 만료된 임차인이 전세금을 돌려받지 못하는 상황' 용어를 학술적으로 정의하였다. 이에 대해 민병철(2023)은 '깡통전세'는 주택가격 하락, '역전세'는 전세가격 하락과 관련이 있음을 언급, 김진유(2022)는 전세보증금이 매매가에 육박하거나 오히려 초과하는 경우를 '깡통전세'로 정의하면서 깡통전세가 많은 지역에서 전세보증사고가 집중적으로 발생함을 지적하였다.

이와 같이 전세보증사고와 주택가격 간의 연관성은 선행연구에서 지속적으로 주제가 제기된 바, 본 연구에서도 동일하게 다루고자 한다.

고제현 외(2020)의 연구는 전세가격이 하락함에 따라 역전세 위험을 인지하고, 시·군·구 단위로 대위변제율을 구분하여 전세가격 지수와 실거래 전세가격 변동과의 상관관계를 분석하였으나 유의한 관계는 없었다. 마찬가지로 민병철(2020)도 역전세의 위험성을 지적하면서 그에 따른 전세보증금반환보증 위험에 대한 정량적 측정의 필요성을 주장하였다. 특히 저자는 전세 실거래 데이터를 기반으로 조건부 실거래 가격 분포를 도출하여 축소되거나 왜곡될 가능성이 있는 가격 간 괴리를 보완했다는 점에서 차별성이 있다. 그 결과

26) 이해인·황현준(2023) 프롭테크의 소비자 평가와 머신러닝을 이용한 아파트 매매 가격 분석

27) 이정현(2023) 인공지능을 이용한 주택가격 변동성 예측 모델 연구

28) 김승현·김원혁·이윤수(2021) 머신러닝과 패널고정효과를 활용한 아파트 실거래가 예측

29) 민병철(2023)은 보증금 미반환에 영향을 미치는 요인 중 정략적으로 측정할 수 있는 것은 역전세와 깡통전세임을 언급하였다.

전세가격지수가 하락할 때 조건부 실거래 가격 분포에서 역전세 사고 발생확률이 더 높은 위험수준으로 나타남을 확인하였다. 또한 노주희(2023)의 연구도 전세 가격과 전세보증사고 간의 연관성을 연구하였으며, 그 결과 전세가격이 감소할수록, 보증발급건수가 많을수록 전세보증사고가 증가함을 주장하였다.

반면 민병철(2023)은 이전 연구와 다르게 '깡통전세'에 집중하여 주택가격 하락위험인 '깡통전세'의 발생확률을 추정하는 방안을 제시하였다. 그 결과 저자는 연구시점 이미 주택가격 하락이 시작된 대구광역시의 깡통전세 위험이 높은 것을 경고하였다.

반면 김진유(2022)는 전세사고와 전세가율 간의 관계를 확인하고자 매매가와 전세가에 중점을 두고 분석을 수행하였다. 이에 주택매매가격과 전세가격을 모두 포함하였으며 주택매매가격과 전세가격이 하락할 경우 고위험 전세가 증가하고, 향후 전세사고가 증대될 가능성이 있음을 주장하였다.

이상 선행연구는 전술한 바와 같이 전세보증금 미반환 위험에 영향을 미치는 요인을 주택가격 또는 전세 가격으로 가정하였다. 그러나 본 연구는 역으로 전세 보증사고의 발생위험이 전세뿐만 아니라 주택시장 전체에 영향을 미칠 수 있으므로 이에 대한 예측분석을 수행하고자 하였다.

한편 주택가격은 경기상황 및 금리수준의 변동과 같이 거시환경<sup>30)</sup>에 영향으로 변동하는 것으로 알려져 있다<sup>31)</sup>. 이광택(1996)은 통화량(M1), 국민총생산(GNP), 주택건설허가량, 주거지역지가, 전국평균지가 등 거시경제변수를 적용하여 주택가격과의 관계를 실증분석하고자 하였고, 그 결과 모든 거시경제변수가 주택가격에 영향을 미치는 것으로 나타났다. 차경수 외(2019)는 거시경제의 구조적 충격이 국내 주택가격에 미치는 효과를 분석하고자 금리, 본원통화, 물가수준, 산출량 증가율 등을 적용하였는데, 그 결과 주택가격이 거시환경의 변화에 따라 금리·신용 경로를 통해

영향을 받는 것으로 나타났다.

또한 김은미 외(2019)는 거시경제변수가 전세보증 사고에 미치는 영향을 분석하였으며, 이를 위해 국고채 3년, 소비자물가지수, 주택준공실적, 주택매매시장 소비심리지수, 전산업 생산지수를 적용하여 전세보증금반환보증 사고 건수가 수취하는 영향을 확인하고자 하였다. 그 결과 장기적으로 전세보증사고는 소비자물가지수에 가장 큰 영향을 받음을 확인하였다.

본 연구는 거시경제환경 내에 전세보증사고가 발생함에 따라 주택가격이 어떻게 변화하는지 확인하고자 주택가격 예측모형을 구축하고자 한다. 선행연구에 따라 거시경제변수와 전세보증사고를 포함하여 독립변수를 선정한다. 전술한 바와 같이 아파트의 거래 비중이 주거 공간 중 가장 높고, 영향력이 가장 크기 때문에 아파트 가격은 주요 관심 사항<sup>32)</sup>일 뿐만 아니라 공시되는 아파트 가격지수는 다양한 관점의 시장 참여자들에게 의사결정 측면에서 매우 중요한 자료<sup>33)</sup>임을 근거로 예측변수는 아파트 매매가격지수 변동률로 선정한다. 또한 선행연구를 기반으로 주택가격의 수요·공급에 영향을 미치는 독립변수와 전세보증사고 금액(외생변수)을 통해 자료를 구축하고자 한다.

## IV. 자료 및 모형설정

### 1. 자료설정

주택가격은 수요와 공급이 일치하는 점에서 결정된다. 이에 따라 주택가격에 영향을 미치는 수요변수는 선행연구에서 전술한 바와 같이 통화량, 금리, 물가, 생산지수 등이 있으나 저자는 주택담보대출<sup>34)</sup>, 주택담보대출금리, 전산업 생산지수<sup>35)</sup>로 선정하여 연구를 수행하였다. 또한 공급변수는 건설공사비지수<sup>36)</sup>, 주택

30) 거시환경이란 경기상황 및 금리수준의 변동과 같이 전국의 모든 주택가격에 공통영향을 미치는 거시경제의 구조적 충격(structural shocks)을 말한다.

31) 국회예산정책처(2019) 거시경제의 구조적 충격이 주택시장에 미치는 영향

32) 김정민(2018) 서울아파트시장과 거시경제변수 요인들 간 동학적 상관관계 분석

33) 이지영·유재필(2021) 인공지능경망을 이용한 주택가격지수 예측

34) 본 연구는 비은행권의 주택담보대출을 포함하였다.

35) 이동수(2013)는 국내 거시지표인 국내총생산(GDP)은 경제 전체의 상황을 파악하기에는 적절한 지표이지만, 작성주기가 길다는 단점을 지적하였다. 마찬가지로 김정렬(2014)의 연구에서도 같은 문제를 지적한 바 산업생산지수를 소득의 대용변수로 이용하였으며 저자는 산업생산지수와 GDP의 상관관계 분석 결과 0.995으로 나타나 전산업생산지수가 GDP의 대용변수로 적합함을 주장하였다. 이에 본 연구도 GDP의 대용변수로 전산업생산지수를 선정하였다.

준공실적<sup>37)</sup>으로 선정하였으며 종속변수인 아파트 매매가격 지수(변화율)와 전세보증사고(금액)을 포함하여 총 7개의 변수를 전세보증사고가 발생한 '15년 1월부터 '24년 8월까지 총 116개의 월별데이터로 구축하였다<sup>38)</sup>.

<표 3> 변수설명

| Feature         | 변수명      | 변수정의                   | 출처    |
|-----------------|----------|------------------------|-------|
| 종속<br>(target)  | apt_buy  | 아파트 매매가격 지수<br>변화율 (%) | KB    |
|                 | incident | 전세보증사고 (억 원)           | HUG   |
|                 | loan     | 주택담보대출 (십억 원)          | KOSIS |
|                 | inter    | 주택담보대출금리 (%)           | ECOS  |
|                 | income   | 전산업생산지수                | KOSIS |
|                 | cost     | 건설공사비지수                | KOSIS |
|                 | supply   | 주택 준공 실적 (호)           | REB   |
| 독립<br>(feature) |          |                        |       |
|                 |          |                        |       |
|                 |          |                        |       |
|                 |          |                        |       |
|                 |          |                        |       |
|                 |          |                        |       |

<표 4>는 7개 변수의 116개 월별데이터 기초통계를 나타낸다. 살펴보면 전세보증사고는 최소값이 0으로 나타나 전세보증사고가 발생하지 않을 구간이 존재함을 확인할 수 있다. 또한 평균 금액은 989억 원, 표준편차는 1,555억 원으로 나타나 평균대비 표준편차가 매우 큰 것으로 나타났다. 이는 전세보증사고의 비대칭적 데이터 분포에 따른 특징 기간 보증사고 급등 가능성을 말해준다. 또한 주택 준공 실적은 평균 34,072호, 최소 15,874호, 중앙 32,689호 최대 62,354호로 나타났다. 상대적으로 평균값대비 최소값이 낮아 주택 공급 부족 구간이 존재할 수 있음을 확인하였다.

이상의 기초통계에 기인하여 <그림 2>는 시간에 따른 변수 추세를 확인할 수 있다<sup>39)</sup>.

지속적인 증가 패턴을 보여주는 주택담보대출, 전

<표 4> 기초통계

|          | mean    | std     | min     | med     | max     |
|----------|---------|---------|---------|---------|---------|
| apt_buy  | -0.003  | 0.275   | -0.932  | 0.00    | 0.931   |
| incident | 989     | 1,555   | 0       | 394     | 6,838   |
| supply   | 34,072  | 9,034   | 15,874  | 32,689  | 62,354  |
| loan     | 636,547 | 102,904 | 453,264 | 631,376 | 818,201 |
| income   | 102.55  | 8.12    | 84.8    | 101.8   | 126.5   |
| cost     | 103.8   | 15.77   | 84.89   | 98.56   | 129.41  |
| inter    | 3.3     | 0.64    | 2.39    | 3.2     | 4.82    |

산업생산지수, 건설공사비지수를 제외하고는 '22년 전후로 급상승했음을 확인할 수 있는데, 특히 금리는 '23년 최대치, 주택 준공 실적은 '22년에 최대값을 나타냈다. 또한 아파트 매매가격 지수는 '22년을 전후로 변동성이 현저히 증가하며, 짧은 주기로 급격한 등락이 반복됨을 확인할 수 있다.

본 연구는 예측해석의 신뢰성을 보완하기 위해 변수 간 계절성, 정상성, 시차를 고려하였다. 이에 LSTM 예측모형의 최적 입력데이터 설계·구축을 위해 STL 분해(Seasonality and Trend decomposition using Loess), KW검정(Kruskal - Wallis test), ADF 검정(Augmented Dickey - Fuller test)을 수행하였다.

먼저 시계열 변수의 계절성 여부는 STL분해 및 KW검정을 병행하여 평가하였고 그 결과 income의 계절성 강도비율(seasonal strength ratio)이 0.57, 월별 분산이 통계적으로 유의하여 계절성 제거 대상 변수로 판단하였다<sup>40)</sup>. 이후 정상성 확보를 위한 ADF검정을 실시하였으며 그 결과 모든 변수가 비정상적으로 나타나, apt\_buy, income, inter, supply, loan은 1차, cost, incident는 2차 차분을 적용하여 단위근 문제를 해소하고 정상 시계열을 확보하였다<sup>41)</sup>.

최적시차 설정을 위해 AIC(Akaike Information Criterion), BIC(Bayesian Information Criterion),

36) 건설공사비지수는 본 연구의 목적에 맞게 주거용 건물로 제한하였다.

37) 주택준공실적은 아파트 기준, 공공부문(국민임대, 공공임대, 공공분양)과 민간부문(민간임대, 민간분양)의 총 합으로 구성된다.

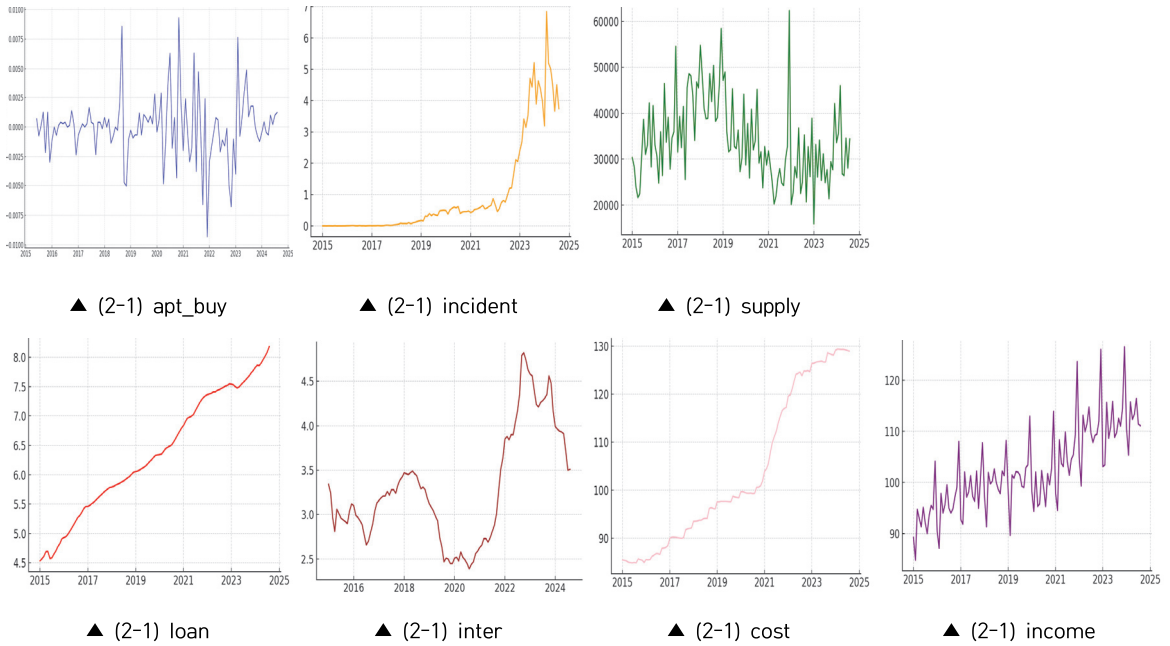
38) 최적시차 적용 이후 '변수명\_t-1'형태로 변경된다.

39) 본 연구의 시간적 범위에 따라 <그림 2>에서 전세보증사고(incident)는 2024년 8월까지 반영된 그래프이다. 연도별 총합을 산출하여 그래프를 작성했기 때문에 마치 2023년과 2024년에 최대축소폭이 존재하는 것처럼 보임에 주의하길 바란다.

40) Kruskal Wallis검정 결과, ' $H_0$ : 계절성이 존재한다.'에 대한  $p=0.002$ 로 나타나 유의한 결과를 나타냈다.

41) 최종 ADF검정 결과, 1차 차분 수준에서 income(adf통계량 = -3.88,  $p=0.00$ ), apt\_buy(adf통계량 = -3.36,  $p=0.01$ ), supply(adf통계량 = -8.70,  $p=0.00$ ), loan(adf통계량 = -5.58,  $p=0.00$ ), 2차 차분 수준에서 cost(adf통계량 = -3.26,  $p=0.02$ ), incident(adf통계량 = -4.81,  $p=0.00$ ), inter(adf통계량 = -3.88,  $p=0.00$ )로 나타났다.

&lt;그림 2&gt; 변수 추세



HQIC(Hannan Quinn Information Criterion)정보 기준 검정을 수행하였다. 그 결과 각 변수에 대해 AIC, BIC, HQIC의 값이 가장 낮은 시차를 찾았으며 모든 변수에서 최적시차는 Lag=1인 것으로 나타났다.

위와 같이 변수 간 선후 관계를 고려하여 계절성, 정상성, 시차적용 등 수행하여 LSTM의 입력데이터를 구축하였다.

## 2. LSTM · Recursive Multi - step

LSTM은 데이터의 시간적 의존성을 학습하여 과거의 features변화를 바탕으로 예측을 수행하고, 과거 중요한 정보가 시간이 지나도 소실되지 않는 장점을 갖는다. 또한 feature 간 상호작용을 학습하므로 미래의 target변수 예측 시 그 정보를 반영한다.

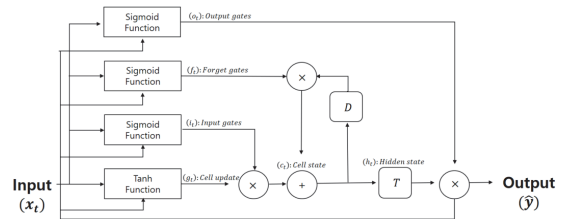
본 연구의 입력값  $x_t$ 는 계절성, 정상성, 시차를 고려하여 전처리된 6개의 독립변수로 구성되며 이를 반영한 시점 t의 LSTM 입력값은 다음과 같다.

$$x_t = [\Delta^2 apt\_buy_{t-1}, \Delta^2 incident_{t-1}, \Delta^2 income_{t-1}, \Delta^2 cost_{t-1}, \Delta^2 inter_{t-1}, \Delta^2 supply_{t-1}] \quad \text{식(5)}$$

입력값  $x_t$ 는 LSTM 셀 내부의 게이트 연산에 의해 식 (5)와 같이 상태를 업데이트 한다.

<그림 3>의 망각게이트( $f_t$ )는 입력값  $x_t$ 와 이전 은닉 상태  $h_{t-1}$ 을 입력으로 받고 시그모이드 활성화 함수를 통해 과거 입력  $c_{t-1}$ 의 일부를 현재 입력에 반영

&lt;그림 3&gt; LSTM 구조



출처 : 저자작성

$$f_t = \sigma(W_f x_t + b_{f_t} + W_{hf} h_{t-1}) \quad \text{식 (6)}$$

$$i_t = \sigma(W_i x_t + b_{i_t} + W_{hi} h_{t-1}) \quad \text{식 (7)}$$

$$g_t = \tanh(W_g x_t + b_{g_t} + W_{hg} h_{t-1}) \quad \text{식 (8)}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad \text{식 (9)}$$

$$o_t = \sigma(W_o x_t + b_{o_t} + W_{ho} h_{t-1}) \quad \text{식 (10)}$$

$$h_t = o_t \odot \tanh(c_t) \quad \text{식 (11)}$$

한다. 또한 입력게이트( $i_t$ )는 Tanh함수를 활용한  $g_t$  (Candidate for cell update)의 생성 과정을 거쳐 현재 시점의 입력을 셀 상태에 어느 정도 반영할지 결정한다.

$c_t$ 는  $f_t$ 와  $i_t$ 의 출력을 조합하여 Cell update를 수행한다. 즉  $c_{t-1}$ 에서  $f_t$ 의 출력값을 곱해 기억할 정보를 선택하고,  $g_t$ 에서  $i_t$ 의 출력값을 곱해 새롭게 추가할 정보를 결정한다. 마지막으로  $o_t$ 는 시그모이드 함수를 적용하여 셀 상태의 출력 정도를 결정하고, 정규화된 셀 상태값인  $Tanh(c_t)$ 를 출력게이트의  $o_t$ 와 결합하여 최종 은닉 상태  $h_t$ 를 출력한다.

본 연구는 LSTM모델의 Layer 1,2로 구축하였다.

Layer 1은 전 시간 단계의 은닉 상태 값을 출력하였으나, Layer 2에서는 마지막 시간 단계의 은닉 상태를 통해 최종 출력값을 산출한다.

$$\hat{y}_{t+1} = W_y h_t + b_y \quad \text{식 (12)}$$

식 (12)에 따르면 t+1에 대한 apt\_buy의 예측값을 LSTM은닉상태, 시점 t까지의 입력 정보를 요약한 벡터( $h_t$ )와 종속변수 y로 변환하는 가중치 행렬( $W_y$ ), 편향(bias)벡터 ( $b_y$ )로 정의된다<sup>42)</sup>.

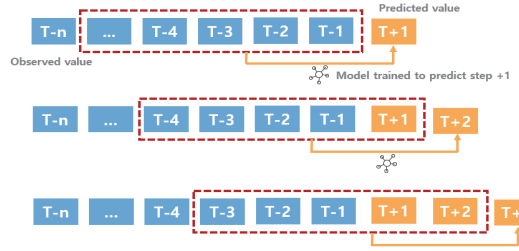
본 연구는 이상의 과정을 통해 구축된 LSTM 예측 모델에 Recursive Multi-step Prediction을 적용하여 향후 12개월의 미래 예측값을 산출하였다.

Recursive Multi-step은 과거 데이터를 입력값으로 사용한 첫 번째 예측값( $T+1$ )을 출력하고(식 13), 그 예측값을 다시 입력값으로 포함하여 두 번째 예측값( $T+2$ )를 출력한다(식 14). 이 과정을 반복하여 다수의 예측값( $T+3, T+4, \dots, T+n$ )을 순차적으로 출력한다. 본 모델은 이전 단계의 예측값이 다음 단계의 입력값으로 사용되기 때문에 재귀적 방식을 활용한 연속 time-step의 예측값을 출력한다.

즉, LSTM은 학습을 통해 시점 t까지의 은닉상태  $h_t$ 를 계산한 뒤, 이를 기반으로 다음 시점의 출력  $\hat{y}_{t+1}$ 를 예측하고 이를 재귀적으로 time-step에 포함시켜 다단계 예측을 수행한다.

마지막으로 예측값(12)은 정규화된 상태이므로, 역

<그림 4> Recursive Multi-step 구조



출처 : 저자작성

$$\hat{y}_{t+1} = f(x_t, x_{t-1}, \dots, x_{t-\rho+1}) \quad \text{식 (13)}$$

$$\hat{y}_{t+2} = f(\hat{y}_{t+1}, x_{t+1}, x_t, \dots, x_{t+1-\rho}) \quad \text{식 (14)}$$

$$\hat{y}_{t+3} = f(\hat{y}_{t+2}, \hat{y}_{t+1}, x_{t+2}, \dots, x_{t+2-\rho}) \quad \text{식 (15)}$$

$$\hat{y}_{t+k} = f(\hat{y}_{t+k-1}, \dots, \hat{y}_{t+1}, x_{t+k-1}, \dots, x_{t+k-\rho}) \quad \text{식 (16)}$$

정규화(Inverse Transform)를 통해 실제 값으로 변환하여 향후 미래시점의 최종예측값을 도출한다.

### 3. SHAP

SHAP은 머신러닝 모델에서 개별 변수가 특정 예측값에 얼마나 기여하는지를 설명하는 알고리즘이다.

이는 게임이론(Game Theory)에서 기반된 것으로 모든 변수 조합을 고려하여 SHAP value를 통해 각 변수의 기여도를 보여준다.

$$\phi_i = \sum_{S \subseteq N \setminus i} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [f(S \cup i) - f(S)] \quad \text{식 (17)}$$

SHAP value는 변수  $i$ 가 제외된 부분집합 ( $S \subseteq N \setminus i$ )에 대해 특정 변수  $i$ 가 포함되었을 경우와 포함되지 않을 경우의 예측값 차이를 계산 ( $f(S \cup i) - f(S)$ )하여 모든 가능한 변수 집합에 대해 공정하게 가중치( $\frac{|S|! (|N| - |S| - 1)!}{|N|!}$ )를 부여해 평균값을 도출한다.

가중치는 변수 순서에 대한 모든 가능한 조합을 공정하게 고려하기 위해 적용된다. 즉, 모든 가능한 변수의 조합, 변수를 추가하는 순서, 그에 따른 기여도 변화

42) 식 (12)의  $W_y$ ,  $b_y$ 는 LSTM 셀 내부의 forget gate나 input gate의  $W_f$ ,  $W_i$  등과 구분된다. 여기서  $y$ 는 예측하고자 하는 출력값(apt\_buy)에 대한 weight와 bias임을 명시하기 위한 표기이다. 즉, 이는 특정 출력 타겟에 종속되는 학습 가능 파라미터를 뜻한다.

를 고려한다. SHAP은 머신러닝 모델에 의존하여 결과를 해석하기 때문에 각 모델의 특성에 맞게 최적화된 다양한 계산방법<sup>43)</sup>이 존재한다.

본 연구는 LSTM과 같이 신경망 모델에서 주로 사용되는 Deep LIFT(Deep Learning Important Feature) 알고리즘을 기반한다. LSTM 예측모델은 각 시점의 입력 벡터에 따라 결정되는데,  $t$  시점에서의 독립변수가  $x_t$ 일 때 예측된 종속변수 값은  $f(x_t) = Y_t$ 로 표현할 수 있다.

기준값 벡터인  $x_{baseline}$ 은 평균으로 설정되어 각 독립변수의 평균값에 대해 모델이 예측하는 종속변수 값은  $f(x_{baseline}) = y_{baseline}$ 으로 표현할 수 있다.

각 독립변수의 Deep SHAP value인  $\phi_{x_i}$ 는 예측된 종속변수의 값과 각 독립변수의 평균값에 대해 모델이 예측하는 종속변수의 차이로 정의되며 이를 통해 독립변수의 기여도를 확인할 수 있다.

$$\therefore \phi_{total} = \sum_{t=1}^T [f(x_t) - f(x_{baseline})] \quad \text{식 (18)}$$

식 (18)은 본 연구에 적용된 월별데이터의 SHAP

value vector를 모두 합산하여 전 기간의 독립변수가 종속변수 예측에 얼마나 기여했는지 각 변수의 기여도를 보여준다.

## V. 분석결과

### 1. LSTM

기존 입력데이터는 2차원이지만 Input-Layer 입력을 위해서는 3차원으로 재생성해야 한다.

이에 LSTM 예측 모델에 입력하고자 기존 입력데이터를 3차원 데이터[samples, time steps, features]로 변경하였다.

samples는 LSTM 예측모델의 입력 데이터셋에서 슬라이딩 윈도우<sup>44)</sup>(Sliding window)방식으로 생성된 독립된 데이터 샘플의 수를 말한다. time steps = 12로 설정하여 최종 LSTM의 3차원 입력데이터는 [100, 12, 7]로 구성하였다.

<표 5>는 본 연구의 LSTM예측 모델의 구조를 보여준다. LSTM은 입력, 재귀가중치, 바이어스(bias)로 구

<표 5> LSTM 구조

|                  |                  |                                                  |                     |
|------------------|------------------|--------------------------------------------------|---------------------|
| Samples          | 100              | The number of samples input to the model at once |                     |
| Time steps       | 12               | The original data size time steps                |                     |
| Features         | 7                | The number of independent variables              |                     |
| Layer (type)     | Unit = 20, 20    | Epochs = 200                                     | Batch size = 64     |
|                  | Output shape     | Param                                            | Sequence            |
| Input layer      | [(None, 12, 7)]  | 0                                                | -                   |
| LSTM 1           | [(None, 12, 20)] | 2,160                                            | Return sequence = T |
| LSTM 2           | [(None, 20)]     | 3,280                                            | Return sequence = F |
| Dense            | [(None, 1)]      | 21                                               |                     |
| Total params     |                  | 16,385                                           |                     |
| Trainable params |                  | 5,461                                            |                     |
| Optimizer params |                  | 10,924                                           |                     |
|                  |                  | Optimizer                                        | Loss / Metrics      |
| Model compile    |                  | Adam                                             | MSE / MAE           |

43) 일반적인 SHAP value의 기본개념은 동일하지만, 머신러닝 모델(Linear, Tree, Gradient 등)에 따라 효율적인 계산을 위해 변형된 SHAP 추정방식을 따른다.

44) 슬라이딩 윈도우 방식이란 시계열 데이터에서 일정한 time step(윈도우 크기)를 지정하고 데이터를 한 칸씩 이동(sliding)하며 연속된 구간을 샘플로 추출하는 방식을 말한다. 본 연구는 해당 방식에 근거(29=116-88+1)하여 samples를 산출하였다.

성되고 각 게이트(입력, 망각, 출력, 후보)에 대해 적용된다. 본 연구의 LSTM예측 모델은 2개 층과 Dense로 설계하였다. LSTM 1은 모든 time steps의 은닉 상태를 출력하기 때문에 Output shape = (None, 12, 20), Param = 2,160으로 나타났다.

또한 LSTM 2는 마지막 time steps의 은닉 상태만 출력하기 때문에 Output shape = (None, 20), Param = 3,280이고, Dense의 Output shape = (None, 1), Param = 21로 나타났다.

즉, 주어진 조건에서 LSTM 1은 Unit = 20, Param = 2,160을 가짐에 따라 입력가중치 = 140, 재귀가중치 = 400, 바이어스 = 20로 구성되고, LSTM 2는 Unit = 20, Param = 3,280에 따라 입력가중치 = 400, 재귀가중치 = 400, 바이어스 = 20로 구성된다. 이에 Trainable params값은 5,461이며, Total params = 5,461+10,924 = 16,385으로 나타났다.

한편 모델 컴파일을 위해 Optimizer = 'Adam', 예측력 평가방법은 평균제곱오차(Mean Squared Error, MSE), 평균절대오차(Mean Absolute Error, MAE),  $R^2$ (R-squared)를 설정하였다<sup>45)</sup>. 변환된 3차원 입력데이터는 Train, Test(8:2)로 분리하여 예측모델에 학습하였다.

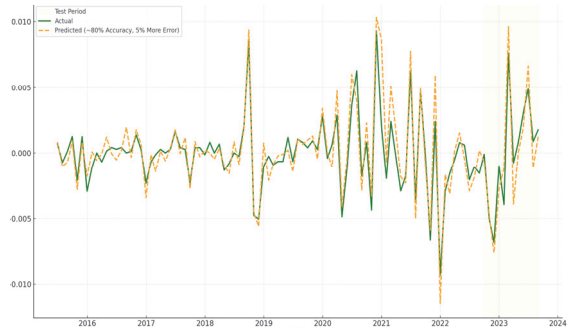
본 모델의 예측력은 MSE = 1.87, MAE = 1.91,  $R^2$  = 0.77인 것으로 나타났다. Train period에서 feature를 기반으로 아파트 매매 가격 지수의 변동 패턴을 학습하고, Test period(노란색 음영)에서 모델이 학습하지 않은 데이터에 대한 예측 결과를 보여준다.

<그림 5>의 예측값은 전체적으로 실제값의 방향성과 주요 추세를 일정 수준 따라가고 있으며, 특히 고점 및 저점 형성 시기와 시계열 변화 패턴이 비교적 유사하게 재현되었음을 확인할 수 있다. 또한 월별 변화율 수준의 비교적 작은 수치 변화에 대해서도 예측값은 민감하게 반응하나, 고빈도 진동 구간에서 예측값의 진폭이 실제값의 미세한 등락 흐름에 비해 일정 수준 이상 추적하는 현상을 일부 구간에서 관찰된다.

이는 LSTM이 단기적인 시계열 변화까지 부분적으로 민감하게 반영함을 시사한다. LSTM 예측 학습 결과를 기반으로 현재 상황이 지속될 경우를 가정하여

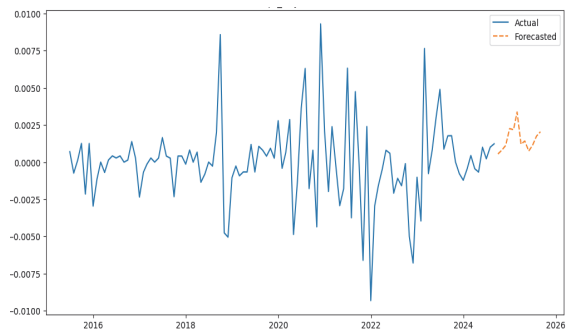
Recursive Multi-step forecasting을 수행하였다.

<그림 5> LSTM 예측 학습 결과



<그림 6>의 주황색 점선은 '24년 9월부터 '25년 8월까지 아파트 매매가격 지수 변동을 예측값을 나타낸다. '24년부터 등락을 반복하며 상승추세를 보여주던 변화율은 이후 '25년 2월까지 지속되다가 '25년 3월과 5월 두 차례에 걸쳐 하락, 이후 다시 반등패턴을 나타냈다<sup>46)</sup>.

<그림 6> Recursive Multi-step 예측 결과



## 2. SHAP

LSTM 모델의 특정 예측 시 어떤 입력(features)이 중요한 역할을 했는지, 모델의 해석(Explainability)을 위해 SHAP 분석을 시도하였다.

먼저 전체 데이터를 기준으로 전역적인(feature

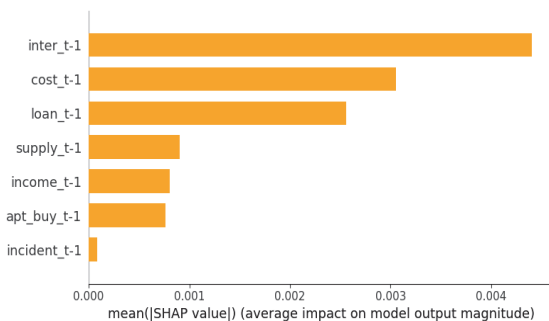
45) MSE: 실제값과 예측값의 차이를 제곱하여 평균한 값, MAE: 실제값과 예측값의 차이의 절대값을 평균한 값,  $R^2$ : 예측값이 실제값의 분산을 얼마나 설명하는지를 나타내는 지표이며 1에 가까울수록 설명력이 높음을 의미한다.

46) Recursive Multi-step 결과 예측값은 (2024) 9월 0.05%, 10월 0.08%, 11월 0.11%, 12월 0.23%, (2025) 1월 0.23%, 2월 0.34%, 3월 0.12%, 4월 0.14%, 5월 0.07%, 6월 0.11%, 7월 0.17%, 8월 0.2%인 것으로 나타났다.

wise) 중요도를 확인하고자 Global plot인 Bar plot과 Beeswarm plot을 시각화하였다.

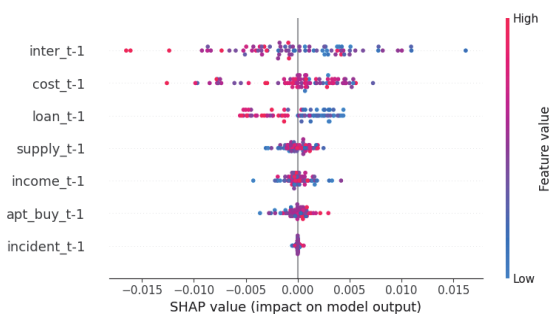
<그림 7>의 Global Bar plot은 모든 샘플(samples)에 대해 각 feature SHAP value 평균 절대값을 보여준다. 그 결과 가장 중요한 변수는 inter\_t-1로 나타났으며, cost\_t-1, loan\_t-1, supply\_t-1, income\_t-1, apt\_buy\_t-1, incident\_t-1 순으로 변수의 SHAP value 평균 중요도가 높은 것으로 나타났다.

<그림 7> Global Bar plot



Bar plot은 개별 샘플에서 해당 feature가 양 · 음의 영향인지 확인하기엔 한계가 있으므로 Beeswarm plot을 통해 각 feature에 관한 SHAP value의 분포를 확인하였고 그 결과는 <그림 8>과 같다.

<그림 8> Global Beeswarm plot



Global Beeswarm plot의 x축은 SHAP value, y축은 feature를 나타낸다. 해당 feature의 원시 값은 color bar 색상으로 표현된 점(point)을 통해 SHAP value와 함께 시각화된다.

inter\_t-1은 전체 변수 중 SHAP value 분산 폭이 가장 넓게 나타났다. inter\_t-1값이 높은 경우 변동률이 하락하는 방향으로, inter\_t-1값이 낮은 경우 변동

률이 상승하는 방향으로 예측값에 기여한 바, 해당 변수는 본 모형에서 가장 높은 예측 기여도를 가지는 핵심 변수임을 확인할 수 있다.

한편 cost\_t-1값과 SHAP value 간의 뚜렷한 단조적 관계는 나타나지 않았으나, 다수의 샘플에서 예측값에 대한 비중 있는 기여가 확인된다. 즉, cost\_t-1의 절대 수준보다는 해당 시점의 다른 변수들과의 상호작용 혹은 비선형 구조 하에서 예측값에 영향을 준 것으로 해석할 수 있다.

또한 loan\_t-1은 그 값이 클수록 예측된 변동률이 낮아지는 일관된 방향으로 기여했음을 확인할 수 있다. 마지막으로 incident\_t-1은 <그림 7>의 결과와 마찬가지로 SHAP value가 전 구간에서 0에 수렴하였다. 비록 SHAP plot에서 incident\_t-1의 평균적인 예측 기여도는 미미하게 나타났으나, 이는 전체 데이터 분포를 기반으로 한 평균적 해석에 불과하며, 특정 조건이나 변수 조합 하에서의 국지적 영향력 등을 완전히 배제할 수는 없다. 이에 본 연구는 incident\_t-1을 특정 변수로 활용한 SHAP Dependence Interaction plot과 개별 설명변수에 대한 Dependence plot을 비교 · 분석함으로써 다른 변수와의 상호작용 하에서 발생할 수 있는 비선형적 예측 기여 가능성을 추가적으로 검토하고자 한다.

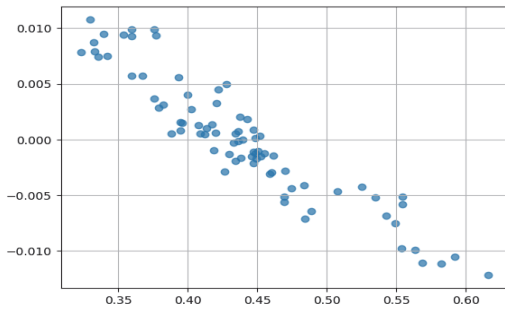
<그림 9-좌>는 단일 feature값에 따른 예측 기여도 변화 양상을 나타내는 Dependence plot, <그림 9-우>는 incident\_t-1(color bar) 조건하에서 feature 값과 상호작용 형태의 예측 기여도를 확인할 수 있는 Dependence Interaction plot이다.

먼저 Dependence plot에 따르면 inter\_t-1은 가장 높은 예측 기여도를 보여줄 뿐만 아니라, SHAP value분포도 넓고 명확한 음의 선형관계를 보여준다. 반면 supply\_t-1은 공급이 증대될수록 SHAP value가 증가하는 뚜렷한 양의 기여 추세를 보여준다. 특히 중요도가 가장 높은 변수는 아니지만, SHAP value의 분포 폭이 넓고 일관된 방향성을 갖는 점에서 예측 성능 향상에 기여하는 설명변수임을 확인할 수 있다.

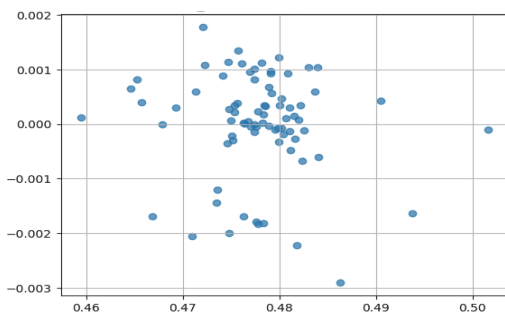
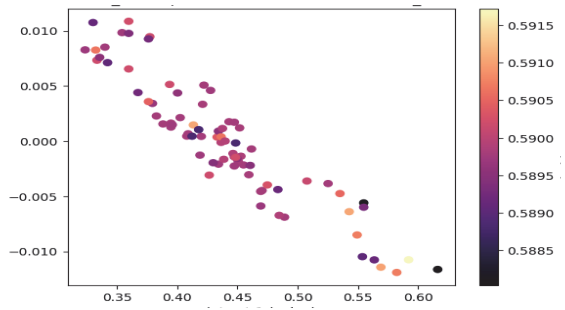
income\_t-1은 그 값이 커질수록 SHAP value가 음의 상관성 요인으로 작용함을 확인할 수 있었고, apt\_but\_t-1은 완전한 선형은 아니지만 대체로 양의 관계를 보였다.

<그림 9>의 Dependence Interaction plot결과, incident\_t-1을 특정변수(color bar)로 포함한 feature

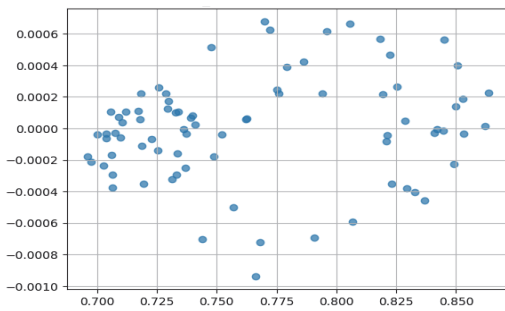
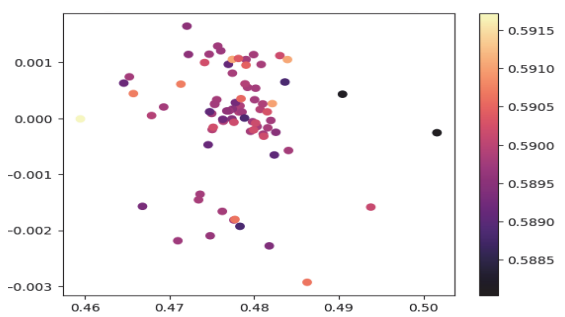
<그림 9> Dependence plot(좌), Dependence Interaction plot(우)



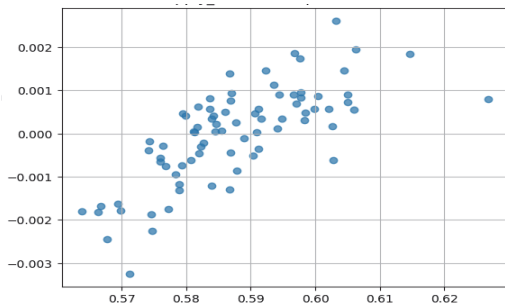
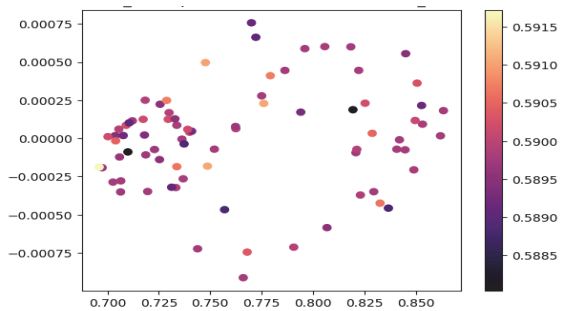
▲ (9-1) inter\_t-1



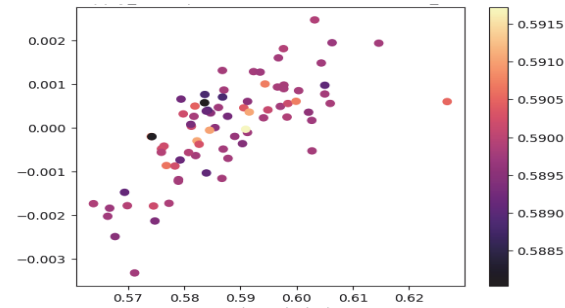
▲ (9-2) cost\_t-1



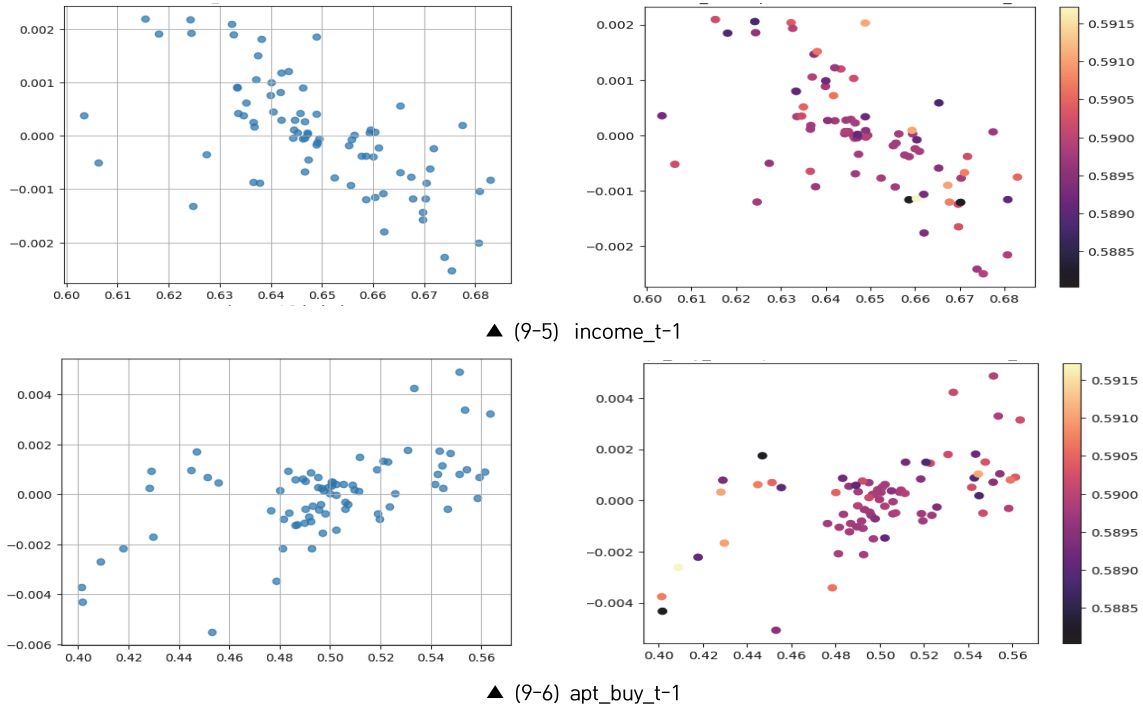
▲ (9-3) loan\_t-1



▲ (9-4) supply\_t-1



&lt;그림 9&gt; Dependence plot(좌), Dependence Interaction plot(우)\_계속



에서 기존 Dependence plot과 유사한 분포 양상이 반복되었으며 그 외 패턴이나 구조적 변화는 관찰되지 않았다. 이는 단기에 아파트 매매가격지수 변화율 예측에 대한 전세보증사고와 각 feature의 상호작용이 본 예측모형 내에서 정보제공력 또는 예측 기여 수준이 제한적임을 보여준다.

## VI. 결론

본 연구는 '전세보증금반환보증을 통해 소멸되지 않고 전가(Risk Transfer)된 전세보증사고 위험이 향후 주택시장에 발현되는가?'라는 연구적 관심에서 출발한다.

전세보증사고는 다양한 원인에 기반하여 임대차 계약 만료 또는 해지 시 임대인이 임차인에게 반환해야 할 전세보증금의 미반환을 의미하는데, 임차인이 사고 위험을 피하고자 전세보증금반환보증 상품에 가입함으로써 임대인, 임차인, 보증기관으로 채무구조가 변화하여 보증기관이 보증의무를 부담하게 된다.

전세보증금반환보증은 '13년 도입 초기부터 소비자들에게 신뢰를 얻으며 시장에서 빠르게 자리를 잡고, 이는 급격한 보증 수요 증대로 나타났다. 그러나 전세보증사고 발생도 빠른 속도로 증가였고, 특히 '23년은 최대 보증사고로 집계되어 시장 전반의 위험으로 확대되었다. 이에 대한 전세보증사고의 다양한 발생 원인이 존재하겠으나, 선행연구의 대부분은 거시경제 등 외부요인에 중점을 두고 주택가격의 변동성(역전세, 강통전세 등)과의 인과관계를 분석하였다.

본 연구는 1. 아파트 매매가격지수 변동률에 대한 Recursive Multi-step 예측을 위해 거시경제·전세보증사고 데이터를 구축하여 LSTM 기반 시계열 예측 구조를 설계한다. 2. SHAP 분석을 통해 딥러닝 모형 내부에서의 변수별 예측 기여도를 정량적으로 추정하고, 모델이 어떤 입력 정보를 중심으로 예측을 수행하는지를 시각적으로 해석하여 3. 이를 기반으로 주택시장 안정화 방안을 제시하고자 하였다.

저자는 선행연구를 기반으로 예측력이 높은 LSTM 모형을 선정하여 예측모형을 구축하고, Recursive Multi-step을 통해 향후 아파트 매매가격 변화율을 예측하였다. 또한 SHAP 분석을 통해 예측 모형 내에서 변수의 중요도·기여도 및 변수 간 상호작용을 분석하

였다. 단순히 예측 모델의 예측력을 비교한 것이 아닌, 향후 예측값을 도출하기 위해 Recursive Multi-step을 활용했다는 점에서 분석 방법의 차별성과 주택가격을 기준으로 전세보증사고를 예측·분석한 이전연구와 달리 전세보증사고를 고려하여 예측했다는 점에서 접근 방식의 차별성을 두었다.

LSTM 모델의 예측 결과, 예측력은  $MSE=1.87$ ,  $MAE=1.91$ ,  $R^2=0.77$ 인 것으로 나타났다. 예측값은 전체적으로 실제값의 방향성과 주요 추세를 일정 수준 따라가고 있으며, 특히 고점 및 저점 형성 시기와 시계열 변화 패턴이 비교적 유사하게 재현되었음을 확인할 수 있었다. 월별 변화율에 대해 예측값은 민감하게 반응하나, 고빈도 진동 구간에서 예측값의 진폭이 실제값의 미세한 등락 흐름에 비해 일정 수준 이상 추적하는 현상이 일부 구간에서 관찰되었다. 이는 LSTM이 단기적인 시계열 변화까지 부분적으로 민감하게 반영함을 시사한다.

LSTM 예측 학습 결과를 기반으로 Recursive Multi-step forecasting을 수행한 결과, '24년부터 등락을 반복하며 상승추세를 보여주던 변화율은 이후 '25년 2월까지 지속되다가 '25년 3월과 5월 두 차례에 걸쳐 하락, 이후 다시 반등패턴을 나타냈다. 향후 예측 추세가 상승과 하락이 반복되는 완만한 진동 형태의 상승흐름을 보여주고 그 주기가 비교적 짧다는 점에서 단기적 시장 변동 가능성이 존재함을 시사한다.

본 예측모델의 Global Bar Plot 결과 주택담보대출금리, 건설공사비 지수, 주택담보대출, 주택 준공 실적, 전산업 생산지수, 아파트 매매가격 지수 변화율, 전세보증사고 순으로 변수 중요도가 높음을 확인하였다. 이는 SHAP value의 평균 절대값을 기준으로 한 변수 중요도이고 개별 샘플에서 각 feature에 관한 SHAP value의 분포를 확인하고자 Global Beeswarm Plot을 수행하였다.

먼저 높은 예측 기여도를 갖는 주택담보대출금리는 SHAP value 분산 폭이 가장 넓게 나타났고 feature value가 커질수록 SHAP value가 음수로 나타나 주택매매가격지수 변화율과 음의 상관성 요인으로 작용함을 확인할 수 있었다. 반면 건설공사비 지수는 SHAP value 간 뚜렷한 단조 관계는 나타나지 않았으나, 다수의 샘플에서 예측값에 대한 비중 있는 기여가 확인되었다. 즉, 건설공사비는 단독 변수로서보다는, 다른 변수들과의 결합 하에서 예측값에 대한 설명력을 갖는

형태로 작용했을 가능성이 있음을 보여주었다. 또한 주택담보대출은 그 값이 클수록 SHAP value가 낮아져 일관된 방향으로 예측력에 기여했음을 확인할 수 있다. 특히 본 연구의 주요 변수인 전세보증사고는 Global Bar Plot과 Global Beeswarm Plot에서 SHAP 값이 0에 수렴하는 양상을 보였다. 이를 단편적으로 해석하면 전세보증사고는 아파트 매매가격 변화율에 대한 예측 기여도가 매우 낮은 변수로 간주될 수 있다. 그러나 이러한 결과는 SHAP value의 평균적 해석이며, 특정 조건이나 변수 조합 하에서 나타날 수 있는 국지적 영향력은 배제되었기 때문에 저자는 전세보증사고와 거시경제 간 상호작용은 아파트 매매가격지수 변화율에 어떤 형태로 예측 기여도를 행사하는지 Dependence Plot과 Dependence Interaction Plot을 추가 분석·비교하였다.

가장 높은 예측 기여도를 가진 주택담보대출금리는 SHAP value분포도 넓고 명확한 음의 선형관계를 나타내 금리가 낮을수록 모델은 아파트 매매가격 상승을 강하게 예측하는 경향을 보였다. 이는 저금리 환경이 주택 수요를 자극하여 가격 상승으로 이어질 수 있다는 기존 시장 논리와도 일치한다. 반면 주택 준공 실적은 값이 증대될수록 SHAP value가 증가하는 뚜렷한 양의 상관성을 나타냈다. 이는 공급 확대가 가격 상승으로 이어질 수 있다는 해석 보다는, 수요 예측을 반영하여 선행적으로 공급이 증가했을 가능성을 보여준다. 즉 공급지표는 아파트 가격 상승률 예측 시 선행지표적 역할을 수행함으로 볼 수 있다. 아파트 매매가격지수 변동률은 완전한 선형은 아니지만 대체로 양의 관계를 보였다. 이는 과거의 수요 강도가 미래의 가격형성에 예측 기여함으로 볼 수 있다.

추가적으로 전세보증사고를 특정변수(color bar)로 선정하여 Dependence Interaction Plot결과, feature에서 기존 Dependence Plot과 유사한 분포 양상이 반복되었으며 그 외 패턴이나 구조적 변화는 관찰되지 않았다. 이는 단기에 아파트 매매가격지수 변화율 예측에 대한 전세보증사고와 각 feature의 상호작용이 본 예측모형 내에서 정보제공력 또는 예측기여 수준이 제한적임을 보여준다.

본 연구의 결과를 바탕으로 도출한 시사점은 다음과 같다. 첫째, 예측 구간에서의 진동 폭은 과거 급변 구간에 비해 상대적으로 작지만, 일정한 간격으로 등락이 반복되는 단기 주기의 파동 형태가 관찰된다. 이는 예

측 모델이 과거의 변동 패턴을 일부 반영하고 있음을 보여주는 것으로, 향후에도 유사한 단기 변동 가능성 존재를 시사한다. 따라서 예측 결과를 활용한 정책 수립 시에는, 평균적 추세뿐 아니라 단기적 변동성에도 대응할 수 있는 정책적 유연성이 고려될 필요가 있다. 둘째, 연구의 중요 변수는 주택담보대출금리로 확인되었으며 이에 금리 상승 속도를 신중하게 조율하여 주택가격의 급격한 변동과 시장 불안의 최소화가 진행되어야 할 것이다. 셋째, 전월의 아파트 매매가격지수 변화율이 다음 달의 가격 변동을 예측에 지속적으로 기여하는 것으로 나타났다. 즉, 외부 충격이 없는 한 수요 흐름이 급격히 전환되기보다는 일정 기간 관성을 가지고 유지됨을 말한다. 이에 정책 개입의 시기와 강도를 판단함에 있어 매매가격 변화를 중요한 선행지표로 고려할 필요가 있다. 특히 변동은 시장 분위기 전환의 신호일 수 있으므로 해당 지표의 급격한 변화는 정책 대응의 시점을 조율하는데 실질적인 판단 기준이 될 수 있다. 넷째, SHAP 분석 결과에서 주택 준공 실적의 일관된 예측 기여 방향성(+)은 공급이 단순히 가격을 상승시킨다는 해석보다는 시장 내 수요 압력을 반영하여 선행적으로 공급이 확대되었을 가능성을 시사한다. 따라서 공급지표는 단지 공급량 그 자체를 의미하는 것이 아니므로 정책 수립에 있어 공급지표의 추이를 통해 시장 전환 국면을 탐지하는 보완적 지표로 활용될 수 있다.

한편, 전세보증사고는 본 연구의 SHAP 분석 전반에서 예측 기여도가 낮고, 분포 또한 극히 제한적인 양상을 보여주었다. 이는 모델 내부에서 해당 변수가 주택 매매가격 변동을 예측에 있어 역할이 미미하다는 결과를 의미하지만, 곧바로 '전세보증사고가 주택시장에 영향을 미치지 않는다' 결론으로 일반화하는 것은 오판단일 수 있다. 단편적으로 본다면 전세보증사고는 아파트 주택유형보다는 오피스텔·빌라 등에서 더 높은 발생 추이를 보여준다. 또한 전세보증사고와 같은 사회적 위험 요인은 시장의 비정형적 반응이나 지역적 요인, 혹은 장기적 신뢰도 저하와 같은 심리적 영향을 수반할 수 있는데, 이는 단기적인 예측모델에는 반영되지 않을 가능성도 존재한다. 따라서 전세보증사고의 낮은 기여도는 변수 자체의 무의미함을 뜻하기보다는 현실의 복잡한 사회적 요인을 충분히 반영하지 못한 모형의 구조적 한계 또는 지역·주택유형 등의 제약을 시사한다고 보아야 할 것이다. 결국 전세보증사고와

같은 사회적 위험요인을 보다 입체적으로 이해하기 위해서는 사회·정책적 요인을 보다 정교하게 통합·확장하는 장단기 프레임이 요구된다.

저자는 위와 같은 한계점을 보완하고, 예측모델을 통한 주택시장의 예측 시스템을 구축하여, 전세보증사고의 구조 및 변화를 선제적으로 대응하는 연구를 지속한다면 향후 안정적인 주택시장을 유지하는데 크게 기여할 것으로 기대한다.

논문접수일 : 2025년 2월 13일

논문심사일 : 2025년 3월 18일

게재확정일 : 2025년 6월 4일

## 참고문헌

- 고제현 · 민병철 · 백인걸, “전세가격 변동에 따른 전세보증 영향 분석”, 「한국주택금융공사 연구보고서」, 2020
- 국회예산정책처, “거시경제의 구조적 충격이 주택시장에 미치는 영향”, 2019
- 김경민, “서울아파트시장과 거시경제변수 요인들 간 동학적 상관관계 분석”, 「부동산학보」, 제2018권 제73호, 2018, pp. 115-129
- 김승현 · 김원혁 · 이윤수, “머신러닝과 패널고정효과를 활용한 아파트 실거래가 예측”, 「주택연구」 제30권 제1호, 2022, pp. 43-69
- 김은미 · 김시연, “거시경제변수가 보증사고에 미치는 영향 분석”, 「주택도시금융연구」 제4권 제1호, 2019, pp. 51-81
- 김은미, “거시경제변수가 지역 별 아파트 전세가격에 미치는 영향 및 예측모델 구축에 관한 연구”, 「지적과 국토정보」 제52권 제2호, 2022, pp. 211-231
- 김정렬, “유동성 및 소득 변수가 주택가격에 미치는 영향에 관한 연구”, 「경제연구」 제32권 제1호, 2014, pp. 127-146
- 김지혜 · 이길제 · 하서진, “주택 역전세 현황과 임차인 보호를 위한 정책 개선 방안: 전세보증보험제도 개선방안을 중심으로”, 「국토연구원」, 2019
- 김진유, “고위험 전세와 전세보증금 미반환 위험의 상관관계 분석: 서울시 전세보증사고를 중심으로”, 「부동산학연구」 제28집 제4호, 2022, pp. 55-69
- 노주희, “전세보증금반환보증 사고 결정요인에 관한 연구”, 서울대학교 박사학위논문, 2023
- 민병철, “깡통전세의 발생 확률 추정”, 「주택도시금융연구」 제8권 제1호, 2023, pp. 5-18
- 민병철, “전세가격 변동률 분포를 통한 역전세 위험 추정”, 「한국주택금융공사 연구보고서」, 2020
- 박서현 · 김도형, “머신러닝을 활용한 부동산 실거래가 요인 분석”, 「한국지적학회지」 제39권 제3호, 2023, pp. 199-210
- 박창복, “위험의 처리 방법”, 「방재와 보험」 제59권, 1993, pp. 30-35
- 박희우, “전세보증보험 현황과 시사점”, KIRI 리포트, 2021, 8, 17
- 배성완 · 유정석, “딥러닝을 이용한 부동산가격지수 예측”, 「부동산연구」 제27권 제3호, 2017, pp. 151-161
- 배성완 · 유정석, “머신 러닝 방법과 시계열 분석 모형을 이용한 부동산 가격지수 예측”, 「주택연구」 제36권 제1호, 2018, pp. 107-133
- 손정식 · 김관영 · 김용순, “부동산가격 예측모형에 관한 연구”, 「주택연구」 제11권 제1호, 2002, pp. 49-75
- 이광택, “주택가격과 변동요인간의 인과성에 관한 상관분석”, 제2권, 1996, pp. 79-105
- 이동수, “전산업생산지수를 이용한 분기 성장률 예측”, 「통계연구」 제18권 제2호, 2013, pp. 66-88
- 이유재, “상호작용 효과를 포함한 다중회귀분석에서 주효과와 검증에 관한 연구”, 「경영학 연구」 제23권, 1994, pp. 183-210
- 이정현, “인공지능을 이용한 주택가격 변동성 예측 모델 연구”, 건국대학교 박사학위논문, 2023
- 이지영 · 유재필, “인공신경망을 이용한 주택가격지수 예측”, 「한국산학기술학회논문지」 제22권 제4호, 2021, pp. 228-234
- 이창로 · 박기호, “단독주택가격 추정을 위한 기계학습 모형의 응용”, 「대한지리학회지」 제51권 제2호, 2016, pp. 219-233
- 이태형 · 정명진, “딥러닝 모형을 활용한 서울 주택가격지수 예측에 관한 연구: 다변량 시계열 자료를 중심으로”, 「주택도시연구」 제8권 제2호, 2018, pp. 39-56
- 이해인 · 황현준, “프롭테크의 소비자 평가와 머신러닝을 이용한 아파트 매매 가격 분석”, 「부동산분석」 제9권 제1호, 2023, pp. 23-42
- 전광섭, 「주택금융론」, 부연사, 2007
- 전해정, “기계적 학습을 이용한 주택가격 예측”, 「부동산경영」 제24집, 2021, pp. 223-243
- 전해정, “시계열분석모형과 머신러닝을 이용한 주택가격 예측력 연구”, 「주거환경」 제18권 제1호, 2020, pp. 49-65
- 전해정 · 양혜선, “딥 러닝을 이용한 주택가격 예측에 관한 연구”, 「주거환경」 제17권 제2호, 2019, pp. 37-49
- 정원규 · 이상엽, “인공신경망을 이용한 공동주택 가격지수 예측에 관한 연구: 서울지역을 중심으로”, 「주택연구」 제15권, 2007, pp. 39-64
- 차경수 · 배정환, “거시경제의 구조적 충격이 주택시장에 미치는 영향”, 2019, 국회예산정책처
- 주택도시보증공사, 「업무통계연보」, 2023
- 한국주택금융공사, 「주택금융리서치」, 2023

## <국문요약>

# 전세보증사고 위험과 거시경제를 고려한 주택가격 예측

김 은 미 (Kim, Eun - Mi)

본 연구는 전세보증사고와 거시경제를 고려하여 주택가격을 예측하고, 전세보증사고 위험과 거시경제 요인이 예측 결과에 어느정도 기여하는지 다양한 측면에서 분석한다. 이에 1. LSTM 기반 시계열 예측모형을 학습하고, Recursive Multi-step을 활용하여 향후 아파트 매매가격지수 변동률을 예측한다. 2. SHAP 분석을 통해 딥러닝 모형 내부에서의 변수별 예측 기여도를 정량적으로 추정하고, 모델이 어떤 입력 정보를 중심으로 예측을 수행하는지를 시각적으로 해석하여 3. 향후 주택시장 안정화 방안 제시를 목적으로 한다. 주택가격의 수요·공급변수인 주택담보대출금리, 주택 준공실적, 전산업 생산지수, 건설공사비 지수, 주택담보대출, 전세보증사고, 아파트 매매가격 지수 변동율을 선정하였으며 전세보증사고 발생 시기인 '15년 1월부터 '23년 8월까지 총 116개의 월별자료를 구축하였다.

예측 결과, 향후 상승과 하락이 반복되는 완만한 진동 형태의 상승흐름과, 그 주기가 비교적 짧은다는 점에 기인하여 단기적 시장 변동 가능성이 존재함을 확인하였다. 또한 예측 기여도가 높은 변수는 주택담보대출 금리로 나타났으나 상대적으로 전세보증사고는 제한적인 것으로 나타났다.

이는 변수 자체의 무의미함이 아닌 현실의 복잡한 사회적 요인을 계량적으로 충분히 반영하지 못한 한계 또는 모형의 구조적 제약을 시사한다고 보아야 할 것이다. 이와 같은 한계점을 보완하여 지속적인 연구를 수행한다면 향후 주택시장의 안정성에 큰 기여를 할 것으로 기대한다.

주 제 어 : 전세보증사고, 주택가격, Long-Short Term Memory, Shapley Additive exPlanations, Recursive Multi-step