

머신러닝을 활용한 아파트 매도 호가와 매물량이 실거래가에 미치는 영향 연구

김규석*, 이소영**, 김경민***

A Study on the Impact of Asking Price and Sales Volume on the Real Transaction Price of Apartment using Machine Learning

Kyuseok Kim*, Soyoung Lee**, and Kyung-Min Kim***

요약

부동산 시장에서는 변동성이 큰 기간 동안 시장 변화를 파악하는 데 매도 호가가 도움이 된다. 이 연구는 부동산 시장의 하락과 상승이 모두 관측되는 2022년 5월부터 2023년 7월까지의 아파트 매도 호가와 매물량이 시장 가격에 미치는 영향을 조사했다. 선형 회귀와 딥러닝 기반의 DNN 모델을 활용하여 3,000세대 이상의 대규모 서울 아파트 단지의 데이터를 분석했다. 주요 변수로는 부동산 가격, 거래량, 이자율 및 부동산 특성이 포함되었다. 연구 결과, 최저 실거래 가격과 최저 매도 호가 간의 큰 차이와 매물량의 증가는 실거래 가격 변동을 크게 이끌었다. 또한, DNN 모델의 예측 정확도가 검증되었으며, 매도 호가와 매물량이 주요 요인으로 확인되었다. 이는 매도 호가와 매물량이 부동산 정책에서 중요한 시장 분석의 도구가 될 것으로 시사한다.

Abstract

In the real estate market, asking prices help gauge market shifts during volatile periods. This study examines apartment asking prices and sales volume effect on transaction prices from May 2022 to July 2023, capturing both market downturns and upswings. Using linear regression and a deep learning-based DNN model, we analyzed data from large Seoul apartment complexes with over 3,000 units. Key variables included real estate prices, transaction volume, interest rates, and property features. The findings showed that larger gaps between the lowest transaction and asking prices, and increased sales volume, led to bigger price swings. The DNN model's predictive accuracy was verified, with asking prices and inventory as its primary drivers. This suggests asking prices and sales volume will be refined as key market analysis tools in evolving real estate policies.

Keywords

real estate price, machine learning, regression analysis, bigquery

*서울대학교 환경대학원 환경계획학과 박사수료,
한국폴리텍대학 데이터융합SW과 조교수

- ORCID: <https://orcid.org/0000-0001-6613-5125>

**서울대학교 환경대학원 환경계획학과 박사과정

- ORCID: <https://orcid.org/0009-0007-9426-2691>

***서울대학교 환경대학원 환경계획학과 교수(교신저자)

- ORCID: <https://orcid.org/0000-0003-0438-7612>

· Received: Aug. 09, 2023, Revised: Oct. 06, 2023, Accepted: Oct. 09, 2023

· Corresponding Author: Kyung-Min Kim

Dept. of Environmental Planning, Graduate School of Environmental
Studies, Seoul National University, Seoul, Korea

Tel.: +82-2-880-9517, Email: kkim2@snu.ac.kr

1. 서론

대한민국 국민들의 자산 중 부동산의 비율은 다른 선진국들에 비해 매우 높은 편에 속한다[1]. 금융투자협회가 발표한 “2022 주요국 가계금융자산 비교”에 따르면, 2021년 말 한국의 가계자산 구성에서 부동산 등 비금융 자산의 비중은 64.4%였다. 이는 미국의 28.5%, 일본의 37.0%, 영국의 46.2% 보다 39.4%에서 126.0% 더 높은 수준임을 알 수 있으며, 호주의 61.2%와는 비슷한 수준이나 우리나라가 가장 큰 수치임을 알 수 있다. 경제적인 부분에 있어 대한민국 국민에게 부동산이 차지하는 비율이 높기에 부동산 정책과 관련된 것도 항상 염두에 두고 있다. 대한민국 국민에게 부동산은 결혼을 앞두고 있는 20~30대의 젊은 세대부터 중장년 그리고 노년 세대까지 모두에게 경제생활에 있어 중요한 역할을 하는 것이라고 볼 수 있다.

국내 경제적 주체의 대표 자산으로 볼 수 있는 아파트 시장은 2013년부터 상승세로 전환하여 2022년에 정점을 찍고, 4분기부터는 하락세로 전환했다[2]. 2021년에는 전국 주거용 부동산 매매가격 지수가 월별 0.7에서 1.0 사이로 증가하였다. 그러나 2021년 11월부터는 0.6으로 하락하면서 2022년 2월에는 0.0으로 시작하여 2022년 10월에는 -0.8까지 하락했다[3]. 2022년 12월, 현재 시점에서는 부동산 매매가격의 하락이 심화되고 있는 상황이다.

이러한 급격한 국면 전환과 관련하여 부동산 시장에서 매도호가의 역할에 대한 논의가 필요하다. 매도호가는 매도자가 시장에 제시하는 가격으로서 아파트 시장에서 구매자의 매물 탐색 방향을 결정하고 실거래가에 영향을 준다. 허위 매물 또는 주변 시세보다 높은 매도호가는 해당 지역 또는 인근 지역의 가격을 끌어올리는 역할을 한다. 매도호가는 매도자와 매수자가 상호 합의한 가격인 실거래가와 달리 매도자의 시장에 대한 반응을 반영하는 가격으로서 부동산 시장 분석에 유용한 지표가 될 수 있다.

본 연구는 최근 1년 여간 부동산 시장의 가격 상승과 하락이 모두 관측 가능한 시간적 범위에서 주택 매도 호가가 아파트 매매 실거래 가격 형성에 영향을 미치는 요인을 분석한다. 기존의 다양한 연구에서도 포함되었던 기준 금리, CPI, 10년 국고채

금리 등의 거시경제 변수뿐만 아니라 인터넷에서 주기적으로 수집된 Naver 부동산 매도 호가와 매물량을 인자로 포함한 미시경제 변수를 기반으로 모형을 구성한다. 기준이 되는 부동산 실거래가는 주로 2021년 하반기에 갱신된 해당 아파트 단지, 해당 면적의 역대 최고가를 기준으로 한다.

연구 데이터의 시간적 범위는 2022년 5월부터 2023년 7월까지의 주 1회씩 수집한 Naver 부동산의 단지별 / 매물 데이터로서 하락과 상승을 모두 관측할 수 있는 기간이다. 공간적 범위는 부동산 하락기에 매매 거래가 존재할 가능성이 상대적으로 높은 서울시 소재 3,000세대 이상의 대규모 아파트 단지이다.

연구 방법론으로는 과거의 사실 관계를 분석할 수 있는 전통적인 선형 회귀분석과 미래를 예측할 수 있는 DNN(Deep Neural Network) 기반의 인공지능 모형을 활용한다. DNN에 활용되는 하이퍼 파라미터(Hyper-parameter)의 최적화를 통해 정확도가 높은 인공지능 모형을 제안하는데 역점을 둔다.

본 논문의 구성은 다음과 같다. 2장에서는 부동산 실거래가 데이터와 관련된 기존의 연구, 인공지능을 활용한 부동산 매매 가격 연구 관련 문헌을 고찰한다. 3장에서는 본 연구에서 활용하는 데이터와 연구 방법론에 대하여 설명하며, 4장에서는 연구 결과를 분석하며, 마지막으로 5장에서는 결론 및 향후 연구 과제를 설명한다.

II. 관련 연구

2.1 부동산 실거래가 및 매도호가 관련 개념 및 이론

부동산 실거래가란 부동산 거래신고 등에 대한 법률에 의하여 신고하는 가격으로서 토지 또는 건축물을 거래하는 경우 실제 성립된 가격을 의미한다. 국토교통부는 2006년 1월부터 아파트, 연립/다세대, 단독/다가구 등의 부동산 매매 거래에 대해 실거래가 신고를 의무화 하여 실거래가 공개 시스템이 공개하고 있다[4]. 또한, 2007년 6월 29일 이후부터는 아파트 분양/입주권까지도 그 대상을 넓혔고,

2011년 1월부터는 전월세 거래에 대해서도 실거래가 자료들도 포함되었다. 이러한 부동산 실거래가격 공시를 통해 부동산 거래질서의 투명성을 제고하고 부동산 시장 안정화 기여를 목표로 하고 있다[5].

실거래가 이외에 주택가격을 판단할 수 있는 자료로는 시세와 매도호가 있다. 시세는 감정평가사나 부동산중개업자를 통해 평가되는 가격으로서 금융기관 등에서 자체적으로 축적 하는 자료를 의미한다. 시세는 수집 주체, 감정평가사 또는 중개업자의 주관에 개입될 여지가 있다[6].

이에 반해 매도호가는 매도자가 시장에 제시하는 가격으로서 아파트 시장에서 구매자의 매물 탐색 방향을 결정하고 실거래가에 영향을 준다. 매도호가는 다양한 요인에 의하여 형성되는데, 일반적으로 매도자의 손실회피성향(Loss aversion) 성향, 금융 제약 등의 영향을 받는 것으로 파악되고 있다.

2.2 아파트 매도호가와 실거래가의 관계에 관한 논의

이우태 et al. (2006)은 아파트 매도호가와 실거래가의 관계를 분석하였다[7]. 연구데이터로는 1999년 6월부터 2006년 5월까지의 부동산114의 수원시 영통주공아파트로 하였으며, 연구방법론으로는 그랜저인과관계 검정방법(Granger Causality Analysis)을 활용하였다. 연구결과, 실거래 가격이 매도 호가에 영향을 주지만, 매도 호가가 실거래 가격에 영향을 주지는 않음을 확인할 수 있었다.

Han and Strange (2016)는 선행연구들은 바탕으로 매도호가와 실거래가의 관계를 검토하였으며 매도 호가가 구매자의 매물 탐색 방향을 결정하고 실거래가에 영향을 준다는 모형을 제시하고 있다[8]. 분석 결과에 따르면 매도자 수가 늘어나면 매도호가 가 하락하는 관계가 관측되며, 시장의 상승 국면보다 하락 국면에서 이러한 관계가 더욱 강하게 관측되었다. 이와 같은 논의는 매도호가와 실거래가의 형성에 직접적인 영향을 준다고 해석될 수 있다.

김진유 (2017)는 아파트의 시세와 실거래가가 가지는 격차에 영향을 미치는 요인에 대하여 연구하였다[9]. 연구 데이터로서 시세 자료는 한국감정원

부동산테크의 수도권 아파트 가격자료로 2013년 1월부터 12월까지의 자료이며, 실거래가 자료는 국토교통부 실거래가 시스템에 공개된 2013년 1월부터 12월까지의 자료를 활용하였다. 연구 모형은 종속변수가 음수가 아닌 경우에 사용하기 적합한 토빗 모형(Tobit model)을 활용하였다. 본 연구에서는 실거래가와 시세 간의 격차를 종속변수로 하고 주택 시장 현황 등의 거시적 변수와 아파트 특성 등의 미시적 변수를 독립변수로 하였다. 연구 결과, 다음과 같이 세 가지 정도로 요약될 수 있다. 첫째, 아파트 가격의 시세와 실거래가의 격차는 시도별 월간 실거래 지수 변화율 등의 주택시장 변화에 의해 영향을 많이 받는다. 둘째, 전용면적이 넓을수록, 건축연령이 높아질수록 격차가 커짐을 알 수 있었다. 셋째, 계절별 편차는 크지 않음을 알 수 있었다.

최성호(2018)는 아파트 매도호와 실거래가의 관계를 분석하였다[6]. 2006년부터 2017년까지 부동산114의 매물들을 데이터로 하여 패널VAR모형과 동적 패널모형을 활용하여 분석하였다. 연구결과, 매도호가는 실거래가에 종속되지 않고, 과거의 실거래가만이 실거래가에 영향을 미친다는 것을 알 수 있었다.

2.3 인공지능을 활용한 부동산 가격 예측 관련 연구

전해정(2018)은 시계열 분석모형인 ARIMA (Autoregressive Integrated Moving Average), VAR(Vector Autoregression Model), BVAR(Bayesian VAR) 모형과 시계열 기반의 인공지능 모형인 LSTM(Long Short-Term Memory), RNN(Recurrent Neural Network)을 활용하여 주택가격 예측을 수행하였다[9]. 연구 데이터로는 2001년 1월부터 2018년 12월까지이며, 전국, 서울, 강남 그리고 강북으로 나누었다. 연구 결과, ARIMA, VAR, BVAR 모형의 RMSE(Root Mean Squared Error)와 MSE(Mean Squared Error)는 지역별, 시점별로 다른 양상을 나타냈으며, 0.071부터 0.732까지로 예측 정확도도 매우 다양했다. 그러나 LSTM과 RNN 모형의 RMSE와 MSE는 0.0002에서 0.0504까지였으며, RNN이 전반

적으로 더 우수한 성능을 보였다. 결론적으로 전통적인 시계열 분석 모형인 ARIMA, VAR, BVAR 보다는 인공지능 기반의 시계열 예측 모형인 RNN과 LSTM이 더 우수함을 알 수 있었고, 인공지능 기반의 예측 모형은 실제의 값과 매우 유사한 값을 나타내는 것을 알 수 있었다.

민성욱(2016)은 서울시 아파트 가격과 거시경제 변수를 기반으로 선형 회귀분석 모형, SVM(Support Vector Machine), RF(Random Forest)와 인공 신경망의 다층 퍼셉트론 모형을 활용하여 주택 가격을 예측하였다[10]. 연구 데이터는 2006년 1월부터 2016년 8월까지의 서울시 아파트 매매, 전세, 지가의 실거래가 지수를 활용하였다. 연구 결과, 선형 회귀분석 모형과 인공 신경망 모형의 MSE는 약 10.9배, 그 외 SVM, RF와 비교해도 약 2배 이상이 차이나는 등 인공지능 기반 모형의 성능이 더 우수함을 알 수 있었다.

2.4 본 연구의 차별성

본 연구는 다음과 같은 차별성이 존재한다.

첫째, 기존의 부동산 매도 호가와 실거래 가격과의 관계의 연구는 소수 존재한다. 그러나 기존 연구에서는 실거래가에 영향을 미치는 인자로 매도 호가와 매물량이 포함되지는 않았다는 한계가 존재한다.

둘째, 기존의 연구들은 인공지능을 활용한 부동산 가격 예측 관련 연구는 대부분 2010년대 중반 이후에 진행된 것들로 국내 부동산 가격이 상승세에 있던 시절에 진행된 연구들이다. 그러나 본 연구는 2021년 4분기 이후 부동산 시장 국면의 전환에 초점을 맞춰 부동산 매도 호가와 매물량이 실거래가에 미치는 영향을 파악하는 최초의 연구이다.

셋째, 본 연구는 Explainable AI를 활용하여 각 변수가 미치는 영향도를 파악해 볼 수 있는 연구이다.

III. 연구 방법

1절과 2절에서는 본 연구에서 활용하는 데이터, 연구 모형, 전처리 방법에 대해서 설명하고, 3절에서는 연구 방법론에 대해서 기술한다.

3.1 연구 데이터

2021년 9월 기준, 서울특별시 25개구 아파트 단지 수의 합은 4,479개 단지이다[11]. 2022년 1년 간 서울특별시의 아파트 매매 거래 건은 총 15,384건으로 2021년의 49,751건 대비 30.9% 수준으로 급감하였다[12]. 이는 서울특별시 4,479개의 단지가 연 평균 3.43건의 거래가 이루어졌다는 것으로 월별 한건의 매매 거래가 존재하지 않는 경우가 빈번하다는 것을 의미한다.

본 연구에서는 부동산 하락기에 매매 거래가 존재할 가능성이 상대적으로 높은 3,000세대 이상의 대규모 단지의 거래 데이터를 활용한다. 표 1과 같이 서울특별시 3,000세대 이상 대단지 아파트의 실거래가, 매도 호가, 매물량을 활용하였다.

표 1. 아파트 단지 목록
Table 1. List of apartment complexes

District	Name of apartment complexes(Number of households)
Gangnam	개포주공1단지(5,040), 은마(4,424), 압구정현대(3,130), 도곡렉슬(3,002)
Gangdong	고덕그라시움(4,932), 고덕아르테온(4,066), 래미안힐스테이트고덕(3,658), 강동롯데캐슬퍼스트(3,226)
Gangbuk	SK북한산시티(3,830)
Gwanak	관악드림타운(3,544)
Nowon	월계시영고층(3,930), 중계그린(3,481), 상계보람(3,315), 월계그랑빌(3,003)
Dobong	방학신동아1단지(3,169)
Mapo	마포래미안푸르지오(3,885), 성산시영(3,710)
Seodaemun	DMC파크뷰자이(4,300), DMC래미안e편한세상(3,293)
Seocho	반포본동(3,610), 반포자이(3,410)
Seongbuk	돈암동한신한진(3,929)
Songpa	헬리오시티(9,510), 잠실파크리오(6,864), 잠실엘스(5,678), 잠실리센츠(5,563), 올림픽선수기자촌(5,540), 올림픽훼밀리타운(4,494), 잠실5단지(3,930), 잠실동트리지움(3,696), 신천장미1차2차(3,402)
Yangcheon	목동14단지(3,100), 목동센트럴아이파크위브(3,045)
Jung	신당남산타운(3,116)

3,000세대 이상의 아파트 단지가 속해 있는 지역은 강남구, 강동구, 강북구, 관악구, 노원구, 도봉구, 마포구, 서대문구, 서초구, 성북구, 송파구, 양천구, 중구 등 13개구이며, 총 34개의 단지이다. 이 중에서 강남구 압구정동 현대아파트의 경우 1차부터 13차까지 나누어져 있으나 40여년 전 비슷한 시기에 준공이 되었으며, 시점별 가격 흐름도 비슷하게 가는 점으로 인하여 한 개의 단지로 처리한다.

그림 1과 같이 부동산 정보 플랫폼인 리치고는 아파트 단지별 실거래가와 네이버 부동산에서의 최근 1년 치 매도 호가와 매물량을 제공한다[13-14]. 본 연구에서는 주별 실거래 건의 최저가, 최저 호가, 해당 평형대의 매물량을 활용한다. 또한, 평형대는 실속 평형인 25평형(±2평)과 국민 평형인 33평형(±2평)을 기준으로 한다. 시간적 범위는 최근 1년 치의 데이터를 활용하므로 시작점은 2022년 5월 23일부터 종료점은 2023년 7월 17일까지의 약 1년 2개월이다.

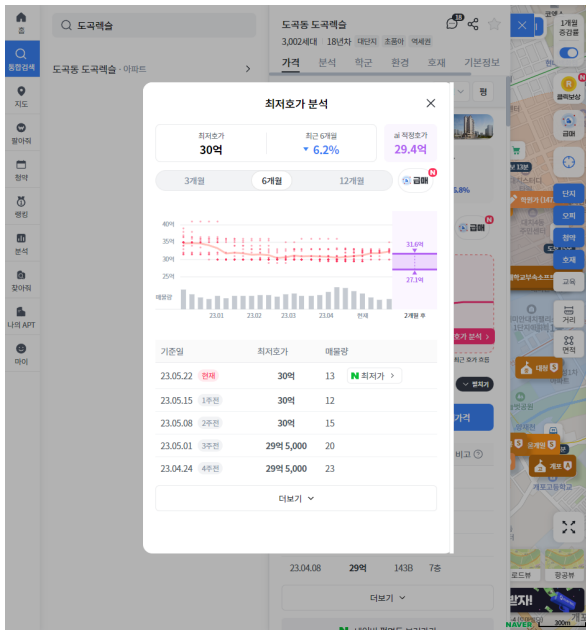


그림 1. 부동산 정보 플랫폼, 리치고
Fig. 1. Real estate information platform, richgo

3.2 연구 모형

본 연구에서는 부동산 매매 가격의 누적 하락율인 CTP를 종속변수로 하여, 매매 최저 호가의 누적 하락율인 CAP, 해당 평형대의 해당 기간 매물량을

평균 매매량으로 나눈 값인 VAV, 전용면적을 나타내는 변수인 SZ, 기준 금리 BIR, 소비자 물가지수 CPI, 10년 국채 금리인 TTY, 최근 실거래가의 증감율인 RTR, 최근 최저 실거래가와 최저 호가의 차이율인 RTA를 활용한다[14]-[16].

표 2. 변수
Table 2. Variable

Variable	Description
CTP	Cumulative actual transaction Price decrease rate
CAP	Cumulative Asking Price decrease rate
VAV	Volume per Annual average Volume
SZ	Size[m ²]
BIR	Base Interest Rate
CPI	Consumer Price Index
TTY	10 year Treasury Yield rate
RTR	Recent Transaction price change Rate
RTA	Ratio of the difference between the recent lowest Transaction price and lowest Asking price

CTP는 식 (1)과 같이 LTP는 해당 시점의 최저 실거래가, LTR은 기준점인 2022년 5월 23일의 실거래 최저가를 통해 구한다. 이 값은 LTP와 LTR의 차이를 LTR로 나눠 100을 곱한 것으로 기준 시점 대비 몇 퍼센트의 변동이 있었는지 파악할 수 있다.

$$CTP_i = \frac{LTP_i - LTR_i}{LTR_i} \times 100[\%] \quad (1)$$

LTP
(Lowest actual transaction price at the point)
LTR
(Lowest actual transaction price at the reference point)

CAP는 식 (2)과 같이 LAP는 해당 시점의 최저 매도 호가, LAR은 기준점인 2022년 5월 23일의 최저 매도 호가를 통해 구한다. 이 값은 LAP와 LAR의 차이를 LAR로 나눠 100을 곱한 것으로 기준 시점대비 몇 퍼센트의 변동이 있었는지 파악할 수 있다.

$$CAP_i = \frac{LAP_i - LAR_i}{LAR_i} \times 100[\%] \quad (2)$$

LAP (The lowest asking price at the point)
LAR (The lowest asking transaction price at the reference point)

해당 평형대의 시점 당시의 매물량을 평균 매물량으로 나눈 값인 VAV는 식 (3)과 같이 산출한다.

$$VAV_i = \frac{SV_i}{\frac{\sum_{i=0}^n WV_i}{n}} \quad (3)$$

SV(Sales Volume)

WV(Weekly sales Volume)

RTR은 해당 평형대의 최근 최저 실거래가의 직전 최저 실거래가 대비 증감률을 나타낸 것이며, RTA는 해당 평형대의 최근 최저 실거래가와 해당 주의 최저가와와의 차이를 비율로 나타낸 것이다.

본 연구에서 활용하는 연구 모형은 식 (4)와 같다. 아파트 실거래 최저 가격 누적 변동률인 CTP를 종속변수로 하고, 매도 최저 호가인 CAP, 주 평균 매물량인 VAV, 전용면적 SZ와 거시 경제 변수인 기준금리 BIR, 소비자 물가지수 CPI, 10년 국고채 금리 TTY, 최근 실거래가 증감률 RTR, 최근 실거래가와 호가의 차이율 RTA를 독립변수로 한다.

$$\begin{aligned} \Delta CTP = & \beta_0 + \beta_1 CAP_i + \beta_2 VAV_i + \beta_3 SZ_i \quad (4) \\ & + \beta_4 BIR_i + \beta_5 CPI_i + \beta_6 TTY_i + \beta_7 RTR_i \\ & + \beta_8 RTA_i + \varepsilon \end{aligned}$$

아파트 단지의 공간적 범위를 3,000세대 이상으로 하더라도 특정 평형대에 대해서는 실거래가 존재하지 않는 경우가 있다. 그리고 아파트 매매의 최근 실거래 기록이 2019년, 2020년 등 기간적 범위가 넓은 경우가 종종 있다. 따라서 본 연구에서는 기준점인 2022년 5월 23일 기준 최근 3개월 이내에 실거래 기록이 있는 경우에는 해당 거래 가격을 기준 가격으로 하며, 그렇지 않은 경우에는 최초 실거래 가격이 존재하는 시점의 가격을 기준 가격으로 한다.

3.3 연구 방법론

본 절에서는 전통적인 분석 방법론인 회귀분석과 딥러닝 기반의 인공지능 방법론 등 두 가지의 방법

론을 활용한 결과 해석 방법에 대하여 기술한다.

회귀분석은 독립변수가 종속변수에 영향을 미치는지 알아보려고 할 때 활용하는 전통적인 분석 방법론이다. 독립변수가 1개라면 단순 회귀분석, 2개 이상이라면 다중 회귀분석에 해당한다. 선형 회귀분석은 독립변수와 종속변수 사이의 차이를 가장 최소화한 선형 관계가 있다는 가정 하에 수행된다.

본 연구에서는 선형 회귀분석을 통해 각 변수의 계수(Coefficient)를 확인함으로써 독립변수가 종속변수에 미치는 영향도에 대한 분석을 수행한다.

DNN은 기존의 ANN(Artificial Neural Network) 대비 입력층(Input layer)와 출력층(Output layer) 사이에 여러 개의 은닉층(Hidden layer)를로 이루어진 인공 신경망이다. DNN은 Hinton et al. (2006)이 제시한 DBN(Deep Belief Network), Vincent et al. (2008)이 제시한 Denoising autoencoder 등도 포함한다 [17]-[19].

본 연구에서는 선형 회귀분석을 통해 인과관계가 파악된 모형을 기준으로 딥러닝 기반의 인공지능 모형을 실행한다. DNN(Deep neural network) 기반의 인공지능 모형의 MAPE(Mean absolute percentage error)를 계산하여 예측의 정확도를 산출한다. DNN 예측에 활용되는 모형의 Dense 레이어의 Input shape는 8개부터 시작하여 1개씩 줄여나가 총 8개의 레이어로 구성된다.

그리고 이 모형에 사용되는 종속변수와 독립변수는 다중 선형 회귀분석과 같이 식 (4)의 변수들로 구성된다. 또한 학습 데이터의 크기인 batch size는 모형의 정확도에 주안점을 맞춰 1로 설정하고, 학습 횟수인 epochs는 100으로 설정한다.

이렇게 구성된 DNN 모형은 수행 시마다 학습 데이터와 검증 데이터를 7:3으로 임의 분할하여 10회 반복 수행 후 MAPE를 산출한다. 그리고 본 모형을 통한 예측의 정확도, 변수의 기여도 등을 판단하기 위하여 AI 기반의 구글 클라우드 플랫폼인 빅쿼리를 활용한다[20].

IV. 연구 결과

4.1 기술 통계

본 연구에서 활용하는 서울특별시 3,000세대 이상 단지의 총 세대 수는 139,825세대이다. 전처리를 통해 정리된 이 단지들의 최근 1년 간의 기록은 총 2,517건이다. 각 변수들의 기술 통계는 Table 3와 같다. 실거래 가격의 평균 누적 하락율인 CTP는 1.07%이며, 매도 호가의 하락율인 CAP는 0.06%이다. 실거래 가격의 최대 하락율은 -49.40%, 매도 호가의 최대 하락율은 -47.10%로 실거래 가격이 절반 이하, 매도 호가 기준으로는 2/3 수준까지 내려간 매물도 있음을 알 수 있다. 평균 매물량인 VAV의 평균은 99.71건이며, 매물의 전용면적인 SZ 변수의 평균값은 72.59이다.

표 3. 기술 통계(n=2,931)
Table 3. Descriptive statistics(n=2,931)

Variable	Average	STD.DEV	Min.	Max.
CTP	1.07	23.13	-49.40	72.41
CAP	0.06	23.24	-47.10	82.11
VAV	99.71	28.44	9.71	523.96
SZ	72.59	12.14	59	85
BIR	3.06	0.58	1.75	3.50
CPI	108.90	1.26	106.36	110.53
TTY	3.57	0.28	3.30	4.27
RTR	0.36	12.08	-31.97	97.56
RTA	1.07	23.13	-49.40	72.41

4.2 선형 회귀분석 결과

식 (4)의 모형을 활용하여 선형 회귀분석 결과는 아래 Table 4와 같다. 독립변수들 중 CPI는 유의수준이 0.05보다 크기에 종속변수에 영향을 주는 변수는 아니었다. 그 외 CAP, VAV, SZ, BIR, TTY, RTR, RTA는 모두 유의수준이 0.05보다 작아 영향을 주는 변수임을 알 수 있다. 그리고 CAP, VAV, SZ, TTY, RTR는 계수의 부호가 양수로 해당 독립변수의 값이 커질수록 종속변수에 양의 영향을 미침을 알 수 있으며, 나머지 BIR, RTA는 계수의 부호가 음수로 해당 독립변수의 값이 커질수록 종속변수에 음의 영향을 미침을 알 수 있다.

결과를 해석하면, 매도 호가(CAP) 변동률이 커질수록, 평균 매물량(VAV)이 많아질수록, 전용면적

(SZ)이 커질수록, 국고채 금리가 높아질수록, 최근 최저 실거래가와 매도호가의 차이가 심화할수록 실거래가가 높아지는 것을 확인하였다. 또한 기준금리가 하락할수록, 최근 실거래가 변동률이 하락할수록 실거래 가격이 높아지는 것을 확인하였다.

특히 최저 호가의 변동률은 실거래가의 변동률과 정(+)의 관계로 나타났으며 최저 호가가 실거래가에 미치는 영향을 확인하였다. 즉, 지난 1년간 부동산 시장의 하락과 상승 국면 모두에서 최저호가의 변동률이 증가할수록 실거래 가격의 변동폭이 증가하는 것으로 파악하였다.

또한, 해당 시점의 최저 실거래가와 최저호가 차이가 심화할수록 해당 실거래가의 변동률이 심화하는 정(+)의 관계를 확인하였다. 이 역시 시장의 상승기와 하락기 모두에서 해당 경향성이 파악되었다.

매물량과 관련해서는 평균 매물량이 증가할수록 실거래가 변동률이 커지는 것을 알 수 있었다. 이는 매물량이 많을수록 선택의 폭이 넓어지는 매수자 우위시장으로서 전환되어 실제 거래되는 금액의 하락 폭이 더 클 수 있음을 시사한다.

표 4. 선형 회귀분석 결과
Table 4. Linear regression analysis result

Variable	Coef.	t-test
CAP	0.7949***	59.706
VAV	0.0359***	4.821
SZ	0.2002***	8.161
BIR	-3.3749***	-3.743
CPI	0.6273	1.528
TTY	2.3889**	2.963
RTR	-0.2264***	-12.400
RTA	0.2451***	11.941
Adj. R2 = 0.767		
p-value : * < 0.05, ** < 0.01, *** 0.001		

4.3 인공지능 학습모형(DNN)을 통한 예측 수행 및 기여도 분석 결과

구글 빅쿼리 인공지능 모형을 구축하여 각 독립변수의 영향도를 측정해 볼 수 있다. 그림 2와 같이 기준 시점 대비 매도 호가 변동률인 CAP의 영향도

가 14.966로 상대적으로 가장 컸으며, 그 다음은 전용면적인 SZ의 영향도가 2.5293, RTA 1.6786, RTR 1.4735, BIR 1.3807, VAV 0.8029, TTY 0.6109, CPI 0.5421의 순이었다. 이를 통해 매도 호가가 가장 영향이 크다는 것을 확인하였다.

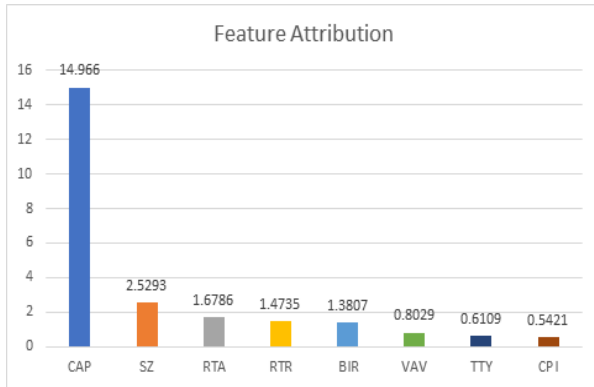


그림 2. 변수별 영향도
Fig. 2. Influence by variable

표 3의 데이터와 식 (4)의 모형을 기반으로 인공지능 예측 모형을 수행하여 최적의 값을 얻은 결과는 표 5와 같다. 학습된 데이터를 기반으로 예측된 결과 값의 MSE, RMSE, MAE는 각 약 9.40, 3.07, 1.98이었으며, 평균 잔차의 표준편차는 약 9.40이었다.

특히 예측값과 실제값의 절대 오차율을 의미하는 MAPE 값이 8.3885%로 도출되어 매우 양호한 예측력을 가지는 것으로 나타났다.

표 5. 인공지능 모형 예측력 분석 결과
Table 5. Result from DNN model

Variable	Value
MAE (Mean Absolute Error)	7.7056
MSE (Mean Squared Error)	124.9583
RMSLE (Root Mean Squared Error)	11.1684
MSLE (Mean Squared Log Error)	1.1826
MAPE	8.3885
Mean residual deviance	9.4
Model : Gradient boosting machine	

V. 결 론

5.1 연구 결론

본 연구는 부동산 가격 하락기에 있던 부동산 시장 상황에 맞게 매도 호가와 매물량이 아파트 매매 실거래 가격과의 관계를 분석하였다. 연구 모형으로는 기존의 다양한 연구들에서도 포함되었던 기준 금리인 BIR, 소비자 물가 지수인 CPI, 10년 국고채 금리인 TTY 등의 거시경제 변수뿐만 아니라 최저 매도 호가와 매물량을 인자로 포함한 모형을 구성하였다.

연구 데이터는 서울특별시 3,000세대 이상의 34개 아파트 단지를 활용하여, 선형 회귀분석과 딥러닝 기반의 인공지능 모형을 구성하였다. 독립변수 중 BIR은 종속변수에 영향을 미치지 않았음을 알 수 있었으며, 그 외 CAP, CPI, TTY, SZ, VAV 등의 모든 변수는 종속변수에 영향을 미침을 알 수 있었다.

특히 최저 호가의 변동률은 실거래가의 변동률과 정(+)의 관계임을 파악하여 최저 호가가 실거래가에 미치는 영향을 확인하였다. 즉, 지난 1년간 부동산 시장의 하락과 상승 국면 모두에서 최저호가의 변동률이 증가할수록 실거래 가격의 변동폭이 증가하는 것으로 파악하였다.

또한, 해당 시점의 최저 실거래가와 최저호가 차이가 심화할수록 해당 실거래가의 변동률이 심화하는 정(+)의 관계를 확인하였다. 이 역시 시장의 상승기와 하락기 모두에서 해당 경향성이 파악되었다.

매물량과 관련해서는 평균 매물량이 증가할수록 실거래가 변동률이 커지는 것을 알 수 있었다. 이는 매물량이 많을수록 선택의 폭이 넓어지는 매수자 우위시장으로서 전환되어 실제 거래되는 금액의 하락 폭이 더 클 수 있음을 시사한다.

한편, 변수별 인공지능 모형에 영향도를 측정 결과, 매매호가 하락율인 CAP의 영향도가 가장 크며 이하 SZ, RTA, RTR, BIR, VAV, TTY, CPI 순으로 영향을 미침을 알 수 있었다.

본 연구에서는 부동산 시장의 국면 전환에 있어서 실거래가 변동률에 영향을 주는 요소로서 매도 호가의 변동률, 매물량의 영향력을 확인하였다.

본 연구는 매도 호가의 누적 하락율과 매물량을 변수로 포함한 최초의 연구로서 의의를 갖는다.

5.2 연구의 보완점 및 추후 연구

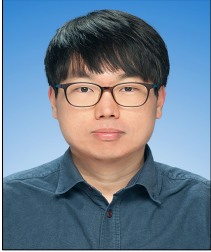
본 연구에서는 아파트 매매 가격 하락에 영향을 미치는 요인들에 대한 연구를 수행하였다. 이는 부동산 시장의 정체기와 상승기는 정밀하게 반영하지 못했다는 점이 한계로 남는다. 추후 연구에서는 부동산 시장의 정체기와 상승기를 구분하여 시계열적인 인과관계를 더욱 정밀하게 반영하고, 부동산 정책의 종류 등의 다양한 독립변수를 추가 반영하여 모형의 정확도를 높이는 연구를 진행할 것이다.

References

- [1] Korea Financial Investment Association, "2022 Comparison of Household Financial Assets in Major Countries", <https://kofia.or.kr/> [accessed: Jul. 1, 2023]
- [2] REB, <https://www.reb.or.kr.> [accessed: Jul. 1, 2023]
- [3] Statistics Korea, "2022 Trends in Housing Sales Price", <https://www.index.go.kr/> [accessed: Jul. 1, 2022]
- [4] Ministry of Land, "Infrastructure and Transport", <https://rt.molit.go.kr/> [accessed: Jul. 1, 2022]
- [5] Ministry of Land, "Infrastructure and Transport Real Estate Transaction Management System", <https://rt.molit.go.kr/> [accessed: May 1, 2022]
- [6] S. H. Choi, "An Analysis of Relationship between Asking Price and Transaction Price - Focusing on Apartment Market in the Seoul Capital Area", *Housing Studies Review*, Vol. 26, No. 1, pp. 89-106, May 2018.
- [7] W. T. Lee and S. J. Kim, "Empirical Analysis on Causal Relationship between Real Sale Prices and Asking Prices of Apartment", *Space and Environment*, No. 26, pp. 183-219, Jul. 2006.
- [8] L. Han and W. C. Strange, "What is the role of the asking price for a house?", *Journal of Urban Economics*, Vol. 93, pp. 115-130, May 2016. <https://doi.org/10.1016/j.jue.2016.03.008>.
- [9] J. Y. Kim, "Determinants of Price Gap between Asking Price and Real Transaction Price of Apartment in Seoul Metropolitan Area", *Journal of the Korea Real Estate Analysts Association*, Vol. 23, No. 1, pp. 73-85. Mar. 2017. <http://dx.doi.org/10.19172/KREAA.23.1.6>.
- [10] H. J. Chun, "Prediction of Housing Price Using Time Series Analysis and Machine Learning Methods", *Journal of The Residential Environment Institute of Korea*, Vol. 18, No. 1, pp. 49-65, May 2020.
- [11] S. W. Min, "A Study on the Forecasting of Housing Price Using Deep Learning - Focusing on Apartment Price Index in Seoul", Graduate School of Kangnam University, 2017.
- [12] Seoul Open Data Plaza, <https://data.seoul.go.kr/dataList/172/S/2> [accessed: May 1, 2023]
- [13] Richgo, <https://m.richgo.ai/pc> [accessed: May 23, 2023]
- [14] Naver Real Estate, <https://land.naver.com> [accessed: May 23, 2023]
- [15] Korea Bank, <https://www.bok.or.kr> [accessed: May 23, 2023].
- [16] Korea Statistics, <https://www.index.go.kr> [accessed: May 23, 2023].
- [17] G. E. Hinton, S. Osindero, and Y. W. Teh, "A Fast Learning Algorithm for Deep Belief Nets", *Neural Computation*, Vol. 18, No. 7, pp. 1527-1554, Jul. 2006. <https://doi.org/10.1162/neco.2006.18.7.1527>.
- [18] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders", *Proceedings of the 25th International Conference on Machine Learning(ICML)*, pp. 1096-1103, Jul. 2008. <https://doi.org/10.1145/1390156.1390294>.
- [19] S. W. Bae and J. S. Yu, "Predicting the Real Estate Price Index using Deep Learning", *Korea Real Estate Review*, Vol. 27, No. 3, pp. 71-86, Sep. 2017.
- [20] BigQuery, <https://cloud.google.com/bigquery> [accessed: May 23, 2023].

저자소개

김 규 석 (Kyuseok Kim)



2011년 2월 : 한국항공대학교
정보통신공학(공학사)
2019년 8월 : 아주대학교
정보통신공학(공학석사)
2019년 9월 ~ 현재 : 서울대학교
도시계획학 박사수료
2019년 6월 : LG전자(주)

선임연구원

2020년 2월 : (주)LG유플러스 책임

2020년 ~ 현재 : 한국폴리텍대학 분당융합기술교육원
데이터융합SW과 조교수

관심분야 : 데이터 수집, 데이터 분석, 머신러닝, 딥러닝,
도시계획, 근거리 무선통신, 상황인지

이 소 영 (Soyoung Lee)



2017년 8월 : 한국예술종합학교
(건축학사)
2021년 8월 : 서울대학교 도시계획
전공(석사)
2021년 ~ 현재 : 서울대학교
도시계획 전공 박사과정
관심분야 : 데이터분석,
부동산시장분석, 머신러닝, 딥러닝, 도시계획

김 경 민 (Kyung-Min Kim)



1995년 2월 : 서울대학교 지리학
학사
2002년 5월 : UC Berkeley School
of Information 정보 시스템 전공
석사
2008년 6월 : Havard University
도시계획 · 부동산 전공 박사

2002년 5월 : Oracle, US 소프트웨어 엔지니어

2001년 1월 : Hyosung Data Systems 소프트웨어
엔지니어

2009년 ~ 현재 : 서울대학교 환경대학원 환경계획학과
교수

관심분야 : 도시계획, 부동산 시장 분석, 인공지능,
머신러닝