

# 표본 주택 가격 기반 부동산 가격지수 산정: 머신 러닝 방법의 활용을 중심으로

Estimating the Real Estate Price Index Based on Sample House Price:  
Focusing on the Use of Machine Learning Method

배 성 완 (Seong-Wan Bae)\* · 유 정 석 (Jung-Suk Yu)\*\*

〈 Abstract 〉

The Purpose of this study is to estimate the real estate price index based on the 'estimated price by machine learning'. The price of a sample house was estimated using the machine learning method 'Random forest' and 'Deep neural networks', and the real estate price index was calculated using the Jevons index calculation method. First, the result of the study showed that the RF index and DNN index are similar, and the variability was changed according to the learning period. Second, the RF index and DNN index showed similar long-term trends compared to the KAB index, but it was found that there was a considerable difference in short-term trends. Third, the RF index and DNN index were found to be more variable than the KAB index, KB index, and real transaction price index, and the relationship with real transaction price index could not be confirmed. If the researcher's qualitative analysis on the RF index and DNN index is added, it is expected that there is a high possibility of utilization as a new price index that can improve existing price index.

키워드 : 머신 러닝, 부동산 가격지수, 랜덤 포레스트, 심층신경망

Keyword : Machine Learning, Real Estate Price Index, Random Forest, Deep Neural Networks

\* 단국대학교 일반대학원 도시계획및부동산학과 박사수료, swbae618@gmail.com, 제1저자

\*\* 단국대학교 사회과학대학 도시계획부동산학부 부교수, jsyu@dankook.ac.kr, 교신저자

## I. 서론

국토교통부는 주택시장 판단지표로 활용하기 위해 다양한 부동산 가격지수를 발표하고 있다. 한국감정원에서 수행중인 전국주택가격동향조사에 의해 산출된 부동산 가격지수(이하 KAB지수)는 ‘조사자에 의해 산출된 표본 주택 가격에 기반한 지수’로서 지수의 정확성을 위해서는 표본 주택 가격의 정확한 산정이 필수적이다. Berry and Bednarz(1975)와 Goolsby(1997)는 과세평가사(assessor)에 의한 체계적인 오차(systematic error)를 지적한 바 있으며, 이용만·이상한(2008)은 KAB지수와 산정방식이 유사한 국민은행 가격지수(이하 ‘KB지수’)에 대해 ‘지수의 평활화’ 가능성을 지적한 바 있다. 이는 결국 ‘조사자에 의해 산출된 표본 주택 가격에 기반한 지수’는 조사자의 성향, 경험, 능력에 따라 오류 또는 편향(bias)이 발생할 수 있는 위험요인을 내재하고 있다는 것을 시사한다. 따라서 시장의 움직임에 부합하는 부동산 가격 지수 산정을 위해서는 조사자에 의한 오류 또는 편향을 제거할 수 있는 방법을 고안할 필요가 있다.

인공지능 관련 기술인 머신 러닝(machine learning 또는 기계학습)은 데이터를 분석하고 결론을 도출하는데 좋은 성과를 보여주고 있다. 배성완·유정석(2018a)은 머신 러닝을 통한 공동주택 공시가격 산정이 조사자에 의한 가격 산정과 유사한 성과를 보여준다고 주장한 바 있다. 조사자들의 성향, 능력, 경험에 따라 발생할 수 있는 오류 또는 편향을 배제할 수 있으며, 실거래가 자료나 매물 자료와 같은 부동산 가격 자료에 대한 습득이 용이하다는 점에서 부동산 가격지수 산출을 위한 표본 주택 가격 산정에 있어서 머신 러닝의 활용 가능성이 매우 높다고 할 수 있다.

본 연구의 목적은 ‘머신 러닝에 의해 산출된 표본 주택 가격’을 기반으로 부동산 가격지수를 산정하고 이것의 활용 가능성을 검토하는 것이다. 이는 기존의 평가기반 가격지수 산정에 있어서 활용된 ‘조사자에 의한 표본 주택 가격’을 ‘머신 러닝에 의한 표본 주택 가격’으로 대체하여 부동산 가격지수를 산정하는 것과 같다.

서울 강남구 아파트를 조사모집단으로 설정하여 표본 추출 후 머신 러닝 방법인 랜덤 포레스트(random forest)와 심층신경망(deep neural networks)을 이용하여 표본주택 가격을 산정하였고, 제본스 지수(Jevons index) 산정 방법을 활용하여 부동산 가격지수를 산정하였다. 표본 주택 가격 산정을 위한 학습자료는 2015년 1월부터 2017년 12월까지의

강남구 아파트 실거래가 자료를 활용하였으며, 표본 추출을 위한 조사모집단으로 부동산테크의 시세 제공 단지를 활용하였다.

본 연구는 부동산 가격 지수 산정에 있어서 기초 재료가 되는 표본 주택 가격 산정 방법에 초점을 맞췄다는 점, 머신 러닝 방법을 부동산 가격지수 산정에 활용하였다는 점, 머신 러닝이 인공지능 기술의 한 종류라는 점에서 본 연구는 '인공지능 가격에 기반한 부동산 가격 지수'라는 새로운 부동산 가격지수의 활용가능성을 검토하였다는 점에서 의의가 있다.

본 연구의 구성은 다음과 같다. 2장은 부동산 가격지수 산정방법에 대해 고찰하고 부동산 가격 지수와 관련 선행연구에 대해 검토한다. 3장에서는 본 연구에 적용할 분석모형, 분석자료 및 분석방법에 대해 고찰한다. 4장은 실증분석으로 표본 추출, 표본 가격 및 부동산 가격지수 산정을 시도하고, 5장에서는 분석결과를 바탕으로 결론과 시사점, 한계점 및 추후 과제에 대해 설명한다.

## II. 이론적 고찰 및 선행연구 검토

### 1. 부동산 가격지수 산정방법

부동산 가격지수는 산정에 활용되는 표본 주택의 특성에 따라 사용되는 모형이 달라진다. 기준시점과 비교시점의 주택표본이 동일하지 않을 수 있는 혼합 표본(mixed sample) 방식의 경우, 헤도닉가격지수모형(hedonic price index model), 반복매매가격지수모형(repeat sale price index model), SPAR지수모형(SPAR index model) 등을 사용할 수 있다. 반면, 시점간 특성변화가 발생하지 않는 동일 표본(matched sample) 방식은 물량 기준 또는 대표치의 산출 방법에 따라 다양한 지수가 만들어질 수 있는데, 바스켓 방법(basket approach)에 의한 지수는 라스파이레스 지수(Laspeyres index), 파슈 지수(Paasche index), 피셔 지수(Fisher index), 월쉬 지수(Walsh index), 톤퀴비스트 지수(Törnqvist index) 등이 있으며, 표본 추출 방법(sampling approach)에 의한 지수는 칼리 지수(Carli index), 듀토 지수(Dutot index), 조화평균 지수(harmonic average index), 제본스 지수(Jevons index) 등이 있다.

국가통계로 승인된 부동산 가격지수에는 전국지가변동률조사, 전국주택가격동향조사,

공동주택 매매 실거래가격지수, 상업용부동산 임대동향조사 등이 있다. 전국주택가격동향 조사는 제본스 지수, 공동주택 매매 실거래가격지수(이하 ‘실거래가격지수’)는 반복매매모형을 이용하여 산정된다.

## 2. 선행연구 검토

부동산 가격 지수와 관련하여 기존 연구는 크게 부동산 가격 지수의 예측을 시도하는 연구와 부동산 가격 지수의 산정을 시도하는 연구로 구분할 수 있다. 부동산 가격 지수를 예측하는 연구는 이미 산정된 부동산 가격지수를 시계열분석 모형이나 머신 러닝 방법을 이용하여 예측하는 방식으로 이루어지고 있다. 민성욱(2017)은 딥 러닝 방법을 이용하여 예측을 시도하였으며, 배성완·유정석(2017)은 딥 러닝 방법인 DNN 및 LSTM(Long Short Term Memory networks)과 ARIMA모형의 예측력을 비교하였다. 배성완·유정석(2018b)은 머신 러닝 방법과 ARIMA, VAR 및 베이지안VAR 모형의 예측력을 비교한 바 있다. 민성욱(2017), 배성완·유정석(2017), 배성완·유정석(2018b) 모두 시계열분석 모형보다 머신 러닝 계열 방법의 예측력이 우수한 것을 보고하고 있다.

부동산 가격 지수의 산정과 관련된 연구는 헤도닉 모형(hedonic approach), 반복매매 모형(repeat sales index), 중위수모형, SPAR모형을 이용하여 지수의 활용 가능성을 검토하거나 지수간 정확성을 비교하는 방식으로 진행되고 있다. 헤도닉 모형은 Lancaster(1966)와 Rosen(1974), 반복매매모형은 Bailey et al.(1963), Case and Shiller(1987), SPAR모형은 Bourassa et al.(2006) 등에 의해 시도된바 있다.

국내에서는 반복매매모형과 관련된 연구가 활발히 이루어지고 있다. 이창무 외(2002)는 표본조사에 의한 라스파이레스 지수 산정방식인 KB지수는 가중치, 스톡의 변동성 처리, 그리고 시세 조사 자료의 정확성 등에서 한계점을 갖는다는 점을 지적하면서 이를 대체할 수 있는 지수 산정 방식으로 반복매매모형을 제안한바 있으며, 이 모형은 현재 발표되고 있는 실거래가격지수 산정의 근간이 되고 있다.

반복매매모형의 경우 표본선정의 편의(sample selection bias)와 자료활용의 비효율성, 지수의 갱신(revision)문제, 주택의 질적 변화를 반영이 어려움 등과 같은 문제점을 가지고 있다(김재익 외, 2015). 이에 따라 이를 개선하거나 대체할 수 있는 지수에 대한 연구가 진행되고 있다.

류강민 외(2009)는 비선형 회귀분석에 의한 산술평균 반복매매지수, OLS에 의한 기하평균 반복매매지수, 가치가중 산술평균에 의한 반복매매지수를 비교하여 산정방식별 차이점과 한계점을 제시하였고, 비선형 반복매매모형이 비교적 정확하게 산술평균 반복매매지수를 산정할 수 있는 것을 확인하였다.

류강민·이상영(2010)은 가치가중 반복매매지수 산정방법을 이론적으로 제시하고, 동일한 자료를 이용하더라도 산정방법에 따라 지수가 다르게 나타난다는 점을 확인하였다.

김재익 외(2015)는 반복매매모형의 문제점을 개선하기 위해 실거래자료의 이상치 제거 기준과 하위지수 병합방식으로의 지수 산정 방식 전환을 제안하였으며, 이 제안은 2015년 3월부터 발표된 실거래가격지수 산정에 적용되고 있다.

단독·다가구주택이나 연립·다세대주택과 같은 비아파트의 경우 해당 주택에 대한 정보가 많지 않기 때문에 라스파이레스 지수나 반복매매모형을 이용한 지수 산정이 어렵다. 이에 따라 비아파트를 대상으로는 중위수지수 산정과 관련된 연구가 시도되고 있다.

류강민 외(2012)는 단독·다가구주택과 연립·다세대주택의 매매 및 전세 중위수지수 산정을 시도하였다. 분석결과 매매 중위수지수는 KB지수보다 실거래가격지수와 더 유사하며, 실거래가격지수보다 평활화 되는 경향을 보이는 것으로 나타났다.

류강민 외(2017)는 미국의 Rador Logic사가 고안한 3개의 멱법칙(triple power law, TPL)을 이용하여 단독주택의 일단위 중위수지수를 산정을 시도하였다. 단독주택을 분석대상으로 하였다는 점, 자료수 부족으로 인한 지수 산정이 어려움을 TPL방법을 통해 극복할 수 있는 방안을 제시하였다는 점에서 의의가 있으나 초기 값 가정 문제, 개별 주택들의 특성을 반영할 수 없다는 점에서 한계를 가진다.

SPAR지수는 특성가격지수 산정시 나타날 수 있는 모형 설정 오류 가능성이 낮고, 반복매매지수에서 발생할 수 있는 표본추출 오류가 심하지 않으며, 좁은 지역에서도 지수 작성이 가능하다는 장점이 있다(이용만, 2007).

이해경 외(2010)는 SPAR지수모형을 이용한 경매가격지수 산정을 시도하였으며, 실거래가격지수와 변동성이 유사하고 동행하는 것으로 나타났으며, KB지수보다는 변동성이 크고 1분기 정도 선행하는 것으로 나타났다.

한편 부동산 가격지수의 특징을 탐색하거나 문제점을 고찰하는 연구도 시도되고 있는데 주요 관심사항은 지수의 평활화 여부이다.

이용만·이상한(2008)은 부분조정모형과 시간변동계수모형을 활용하여 KB지수의 평활화 여부를 검토하였으며, KB지수가 평활화로 인해 실제 거래가격 동향과 차이를 보이고 있을 가능성을 지적하고 있다.

박연우·방두완(2011)은 KB지수의 평활화 현상이 대칭적인지 아니면 비대칭적인지 검토하였으며, 분석결과 상승시장보다 하락시장에서의 평활화 정도가 큰 것으로 나타났다.

### 3. 선행연구와의 차별성

본 연구는 선행연구와 비교했을 때 다음과 같은 차별성을 갖는다. 첫째, 기존 연구는 부동산 가격 지수 산정 방법에 초점이 맞춰져 있으나 본 연구는 지수 산정의 재료가 되는 표본 가격 산정에 초점이 맞춰져 있다. 둘째, 머신 러닝 방법에 의해 산출된 표본 주택 가격은 조사자에 의한 정성적 분석과 주관적 판단이 배제된 가격으로서, 현행 부동산 가격 지수 산정 방식에 있어서 조사자의 가격 산정 역할을 머신 러닝 방법으로 대체할 수 있을지에 대한 검토를 시도하였다. 셋째, 머신 러닝이 인공지능 기술의 한 종류라는 점에서 본 연구는 ‘인공지능 가격에 기반한 부동산 가격 지수’라는 새로운 지표를 제시하였다. 넷째, 주로 부동산 가격 지수 예측에 활용되던 머신 러닝 방법을 본 연구에서는 부동산 가격 지수 산정에 활용하였다.

## III. 분석모형, 분석자료 및 분석방법

### 1. 분석모형

머신 러닝(machine learning)은 컴퓨터 또는 기계가 프로그램이나 알고리즘을 통해 경험적 데이터를 분석·학습하여 데이터 구조를 파악하거나 새로운 결과를 예측하는 것을 말한다.

머신 러닝은 학습에 이용되는 데이터에 따라 지도 학습(supervised learning)과 비지도 학습(unsupervised learning)으로 구분할 수 있다. 지도 학습은 입력 값에 상응하는 결과 값을 가진 데이터를 이용한 학습을 통해 새로운 데이터에 대한 결과 값을 예측하는 학습

알고리즘이다. 반면, 비지도 학습은 입력 값에 상응하는 결과 값이 없는 데이터를 이용한 학습을 통해 데이터의 구조, 규칙 등을 밝히는데 목적이 있는 학습 알고리즘이다. 본 연구에서는 지도학습 알고리즘인 랜덤 포레스트(random forest, RF)와 심층 신경망(deep neural networks, DNN)을 활용한다.

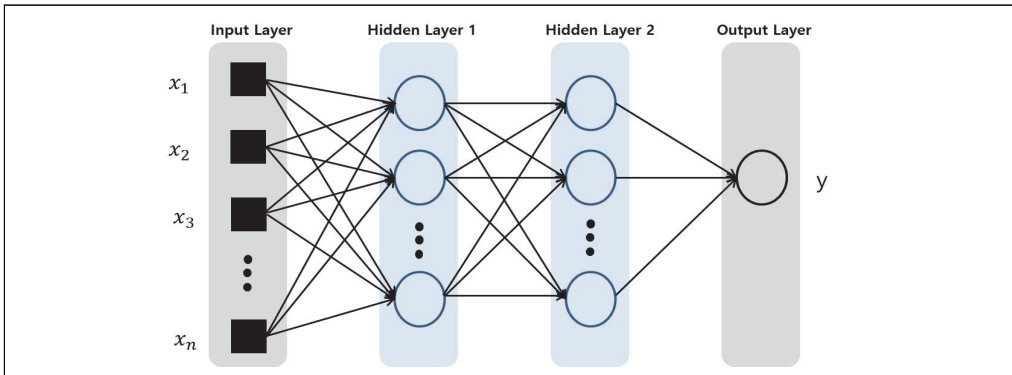
랜덤 포레스트는 부트스트랩(bootstrap) 표본을 다수 생성하고 결정트리(decision tree) 모형을 적용하여 그 결과를 종합하는 앙상블(ensemble) 방법으로 Breiman(2001)에 의해 제시되었다(유진은, 2015). 회귀트리 모형은 설명변수  $X_1, X_2, \dots, X_p$ 를  $J$ 개의 지역(region)  $R_1, R_2, \dots, R_j$ 에 서로 겹치지 않게 분할하고,  $R_j$ 지역에 속하는 관찰치에 대해  $R_j$ 지역 관찰치 평균값을 예측치로 제시하며,  $R_j$ 지역은 잔차제곱합(residual sum of squares)이 최소가 되도록 분할한다(이창로, 2015). 잔차제곱합을 최소화하는 기준으로 트리를 구성하면 과다적합의 가능성이 높아지기 때문에 트리의 규모를 최대한 키워놓고 해당 트리의 가치를 처가면서 적정규모의 트리를 결정하게 되는데 이는 식(1)을 최소화하는 것과 같다(이창로, 2015).

식(1)에서  $|T|$ 는 트리  $T$ 의 가지(terminal node) 수를,  $R_m$ 은  $m$ 번째 가지에 해당하는 분할지역,  $\alpha$ 는 동조 파라미터(tuning parameter)로서  $\alpha = 0$ 이면 아무런 패널티가 없으므로 최대 트리가 되며,  $\alpha$ 가 커질수록 트리규모는 작아지게 된다(이창로, 2015).

$$\sum_{m=1}^{|T|} \sum_{x_i \in R_m} (y_i - \widehat{y}_{R_m})^2 + \alpha |T| \quad (1)$$

심층신경망(deep neural networks, DNN)은 <그림 1>과 같이 다수의 은닉층(hidden layer)을 가지고 있는 인공신경망이다. 인공 신경망은 생물의 뇌 신경망을 모형화하여 만든 알고리즘으로서 1958년 Rosenblatt에 의해 제안된 퍼셉트론 학습 모형 또는 단층 신경망을 시작으로 다양한 형태로 개발되고 있다(구성관·홍석빈, 2017). 인공신경망은 과잉 학습, 기울기 소실(gradient vanishing)과 같은 문제점으로 상당기간 침체기를 맞이하였으나, Hinton et al.(2006)에 의해 제안된 가중치 초기화 방법을 통해 이러한 문제점은 해결되었으며, 딥 러닝(deep learning) 또는 심층신경망이라는 새로운 이름으로 활발한 연구가 이루어지고 있다.<sup>1)</sup>





자료: 구성관·홍석민(2017)

〈그림 1〉 심층신경망의 구조

## 2. 분석자료

표본 가격은 머신 러닝 방법에 따라 산정하게 된다. 머신 러닝 모형의 분석 알고리즘은 학습데이터를 이용한 훈련(train) 및 검증(validation)을 통해 모형을 최적화하고, 최적화된 모형에 시험(test)데이터를 적용하여 결과물을 예측하는 방식으로 이루어진다. 국토교통부에서 제공하는 2015년 1월부터 2017년 12월까지의 서울 강남구 아파트 실거래자료 18,020건을 머신 러닝 모형의 학습 및 검증을 위한 데이터로 활용한다. 부동산테크에서 시세를 제공하는 강남구 아파트 단지를 모집단으로 설정하여 이로부터 추출된 표본 주택은

〈표 1〉 실거래가 자료 기초통계량

구분	최소값	최대값	평균	표준편차
건물면적(단위:㎡)	22.78	264.3	86.67	37.36
입지층(단위:층)	1	68	8.89	7.12
총세대수(단위:호)	100	5,040	1,267.30	1,370.60
준공년도(단위:년)	1,976	2,016	1,995	11.29
지하철역거리(단위:m)	50	2,660	477.03	426.12
거래가격(단위:만원)	12,000	532,000	112,791	54,932

1) Hinton et al.(2006)은 제약볼츠만머신(Restricted Boltzman Machine)이라는 가중치 초기값 설정 방법을 제안하였는데, 최근에는 제약볼츠만머신보다 성능이 뛰어나면서도 사용하기 편한 가중치 설정방법이 제시되고 있다.



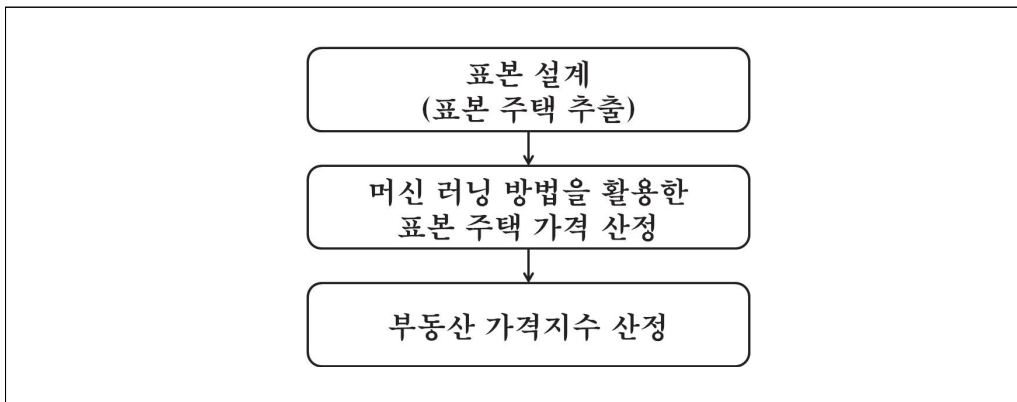
머신 러닝 모형에 적용할 시험데이터(test data)에 해당한다. 본 연구에 활용된 실거래자료와 조사모집단의 기초통계량은 각각 <표 1> 및 <표 2>와 같다. 머신 러닝을 적용함에 있어서 적용된 변수는 건물면적, 입지층, 총세대수, 준공년도, 지하철역과의 거리이다.<sup>2)</sup>

<표 2> 조사모집단의 기초통계량

구분	최소값	최대값	평균	표준편차
건물면적(단위:m <sup>2</sup> )	23.34	245.2	101.83	45.22
최고층(단위:층)	3	69	19	12
총세대수(단위:호)	100	5,040	506	593
준공년도(단위:년)	1,976	2,016	1,996	10
지하철역거리(단위:m)	50	2,660	446.83	435.63

### 3. 분석방법

본 연구에서는 동일 표본(matched sample) 방식에 기초한 지수 산정방법을 활용하여 부동산 가격지수를 산정한다. 부동산 가격 지수 산정을 위한 실증분석은 <그림 2>와 같은 절차에 진행된다.



<그림 2> 실증분석 절차

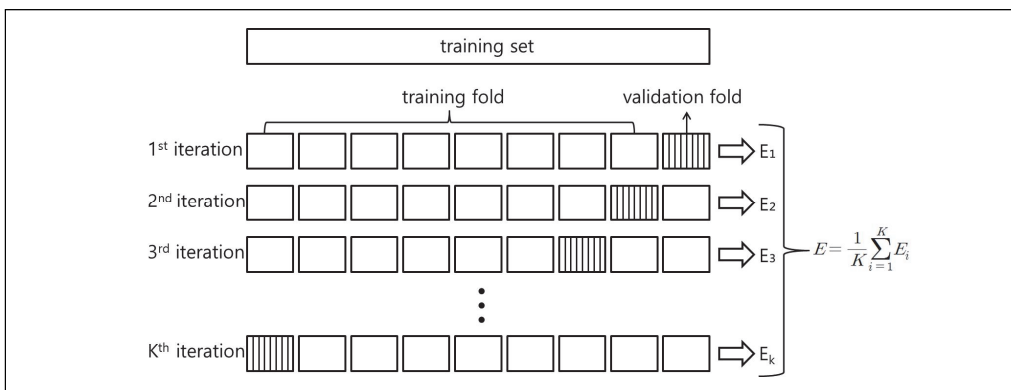
2) 아파트 실거래가 자료는 국토교통부 실거래가 공개시스템(rt.molit.go.kr), 조사모집단은 부동산테크(www.ret.co.kr)를 이용하였다. 조사모집단에 포함된 아파트의 실거래가 자료만을 활용하였다.

먼저 표본 설계를 위해 부동산테크에서 시세를 제공하는 강남구 아파트를 모집단으로 설정하였으며 이로부터 표본 주택을 추출한다.

머신 러닝 방법인 RF와 DNN을 이용하여 표본주택의 가격을 산정하였으며 학습기간은 1개월, 3개월, 6개월, 12개월로 구분하여 시점별 표본 가격을 산정하였다. 예를 들어 2017년 12월의 표본가격을 산정한다고 가정했을 때 학습기간이 12개월이면 2017년 1월부터 2017년 12월까지의 실거래자료가 활용되며, 학습기간이 3개월이면 2017년 10월부터 2017년 12월까지의 실거래자료가 활용된다. RF와 DNN의 투입변수(input variables)로는 건물면적, 입지층, 총세대수, 경과년수, 경과년수 제곱, 지하철역과의 거리, 출력변수(output variable)로는 거래가격이 활용된다.

머신 러닝 모형은 초모수(hyper-parameter)의 설정에 따라 모형의 성능이 달라지기 때문에 최적의 모형을 결정하기 위한 초모수 탐색과정이 필요하다. 본 연구에서는 최적의 모형을 위한 초모수를 찾기 위해 그리드 서치(grid search)를 이용하였다. 그리드 서치는 초모수의 범위를 설정해주면 컴퓨터가 각 초모수의 조합에 따른 오차를 계산하고 최적의 초모수를 제시해주는 방법으로 시행착오법과 유사한 방법이다.

그리드 서치를 통한 초모수 결정 과정에는 k-겹 교차검증이 함께 활용된다. k-겹 교차검증은 훈련 데이터를 k등분하고 등분된 훈련데이터 중 k-1개를 훈련 데이터로 사용하고 나머지 1개의 데이터를 이용하여 모형의 성능을 검증하는 방법이며, 등분된 숫자만큼 검증



주: E는 예측 에러(prediction error)를 의미하며, 본 연구에서는  $R^2$  값이 활용됨.

자료: <https://sebastianraschka.com/faq/docs/evaluate-a-model.html>

〈그림 3〉 k-겹 교차검증의 과정

데이터가 존재하기 때문에 k번의 검증 과정이 이루어지게 된다. 본 연구에서는 3겹 교차 검증을 이용하였으며, k-겹 교차 검증 과정은 <그림 3>과 같다.

부동산 가격 지수 산정방법으로 KAB지수와 동일하게 제본스 지수 산정 방법을 활용하였으며 이는 식(2)와 같다.  $I_j$ 는 j지역의 가격지수이며  $p_i^0$ 는 i표본의 기준시점 산정가격,

$p_i^t$ 는 i표본의 비교시점 산정가격이며  $\frac{p_i^t}{p_i^0}$ 는 표본i의 가격 변동률이며 표본의 가격 변동률

을 기하평균하여 가격 지수를 산정하게 된다. 지수 산정에 있어서 기준시점은 KAB지수와 동일하게 2017년 11월로 결정하였다.

$$I_j = \prod_{i=1}^n \left( \frac{P_i^t}{P_i^0} \right)^{\frac{1}{n}} \quad (2)$$

## IV. 실증분석

### 1. 표본 설계

부동산테크에서 시세를 제공하는 강남구 아파트 단지를 조사모집단으로 설정하였으며, 이로부터 표본 주택을 선정하였다. 먼저 동일한 규모의 가구들로 구성된 집락을 1차 추출 단위로 활용하며, 무작위로 추출된 단지를 2차 추출단위로 활용하게 된다. <표 3>과 같이 조사모집단을 규모에 따라 5개 집락으로 구분하였으며, 각 집락별로 배분된 단지 수만큼 무작위 추출을 통해 조사 단지를 선정하였다. 그리고 선정된 조사 단지별로 층을 기준으로 무작위 추출을 통해 2호씩 표본 주택을 선정하였다.<sup>3)</sup>

3) 국토해양부(2012)는 전국주택가격동향조사를 위한 아파트 표본 설계를 위해 주택규모를 기준으로 5개 층으로 구분한 후 층화 2단 집락추출법(stratified 2nd stage cluster sampling)을 사용하여 표본을 추출하였으며, 강남구의 조사모집단은 109,450호, 조사 표본은 130호이다(표본단지는 65 단지이며, 단지별 2호를 표본호로 선정, 65단지 x 2호 = 130호).

〈표 3〉 조사모집단 및 표본주택

구분		조사모집단		표본주택	
		세대수	비율	표본단지	표본호
소형	60㎡ 이하	27,572	28.3%	18	36
중소형	60㎡초과~85㎡이하	31,734	32.6%	22	44
중형	85㎡초과~102㎡이하	5,414	5.6%	4	8
중대형	102㎡초과~135㎡이하	18,588	19.0%	12	24
대형	135㎡초과	14,077	14.5%	9	18
합계		83,308	85.5%	56	112

## 2. 표본 가격 및 부동산 가격 지수 산정

부동산 가격 지수 산정을 위해서는 시점별 표본 주택의 가격 산정이 필요하다. 이를 위해 머신 러닝 방법인 RF와 DNN을 이용하여 표본 주택의 시점별 가격 산정을 시도하였다. 머신 러닝은 모형의 최적화를 위해 초모수에 대한 결정이 필요하다. RF의 중요 초모수는 tree의 개수이며, DNN의 중요 초모수는 은닉층(hidden layer)의 개수, 은닉층 내 뉴런(neuron)의 개수, 활성화 함수(activation function), 최적화 방법(optimizer method), 초기화(initialization) 방법, 드랍아웃(dropout) 등이 있다. 중요 초모수를 변화시키면서 그리드 서치와 k-겹교차검증을 통해  $R^2$ 값을 기준으로 최적의 초모수 조합을 결정하였다.<sup>4)</sup> 데이터의 학습기간이 1개월, 3개월, 6개월, 12개월로 상이하기 때문에 시점별 최적 초모수의 조합이 상이할 수 있다. 이에 따라 시점별로 최적의 모형을 결정한 후 시점별 표본 주택 가격을 산정하였다.<sup>5)</sup>

4) RF는 트리수를 변화시키면서 최적 모형을 결정하였고, DNN은 은닉층의 개수는 3개, 활성화 함수는 렐루 함수(ReLu function), 최적화 방법은 ADAM 알고리즘으로 결정한 상태에서 은닉층 내 뉴런의 수와 드랍아웃 비율을 변화시키면서 최적 모형을 결정하였다.

5) DNN은 하나의 초모수 조합당 100회의 시행을 반복하기 때문에 시점별로 최적모형을 선택하게 되면 분석시간이 지나치게 너무 오래 걸리는 문제점이 있었기 때문에, 전체 데이터(3년간의 데이터)를 기준으로 최적 모형을 결정하였으며, RF만 분석시점별 최적 모형을 결정하는 절차를 거쳤다.

〈표 4〉 RF 및 DNN의 평균 MAE 및 RMSE

구분	RF		DNN	
	MAE	RMSE	MAE	RMSE
1개월	1,197	1,885	1,589	2,295
3개월	604	1,030	1,457	2,121
4개월	570	899	1,381	1,988
12개월	523	806	1,327	1,886
평균	724	1,155	1,439	2,073

훈련 데이터를 기준으로 RF와 DNN에 의해 결정된 최적 모형의 평균 MAE 및 RMSE는 〈표 4〉와 같이 760~2,111만원이다. 실거래자료의 평균주택가격이 112,791만원인 점을 감안했을 때 머신 러닝 모형에 의해 산출된 가격과 실제 가격은 평균적으로 약0.6~1.8% 정도의 차이를 보이는 것으로 나타났다.

RF와 DNN에 의해 산출된 표본 주택 가격을 이용하여 산정된 부동산 가격지수는 〈표 5〉 및 〈그림 4〉와 같다. RF지수와 DNN지수는 전반적인 방향성은 대체로 유사한 모습을 보여주고 있으나 변동성은 RF지수보다 DNN지수가 더 큰 것으로 나타났다. 또한 머신 러닝 가격 지수는 학습기간이 짧아질수록 변동성이 커지는 것으로 나타났는데 이는 시점별 표본 주택 가격 산정에 있어서 학습기간이 길어질수록 학습데이터의 중복비율이 높아지기 때문인 것으로 이해된다. 예를들어 학습기간이 12개월인 경우 2015년 12월과 2016년 1월의 지수를 산정하게 되면, 학습데이터는 각각 '2015년 1월~2015년 12월', '2015년 2월~2016년 1월'을 활용하게 되어 11개월 분량의 학습데이터가 서로 중복된다.

절대적인 가격수준은 학습기간에 따라 다소 차이를 보이고 있다. 기준시점(2017년 11월)과 비교했을 때 학습기간이 6개월 또는 12개월인 경우는 KAB지수 및 실거래가격지수보다 가격 수준은 대체로 높은 것으로 나타나고 있으나 학습기간이 1개월 또는 3개월인 경우는 KAB지수 및 실거래가격지수보다 대체로 낮은 가격 수준을 보여주고 있다. 이것 역시 학습기간에 따른 과거 데이터의 영향력 차이로 인해 발생하는 것으로 판단된다.

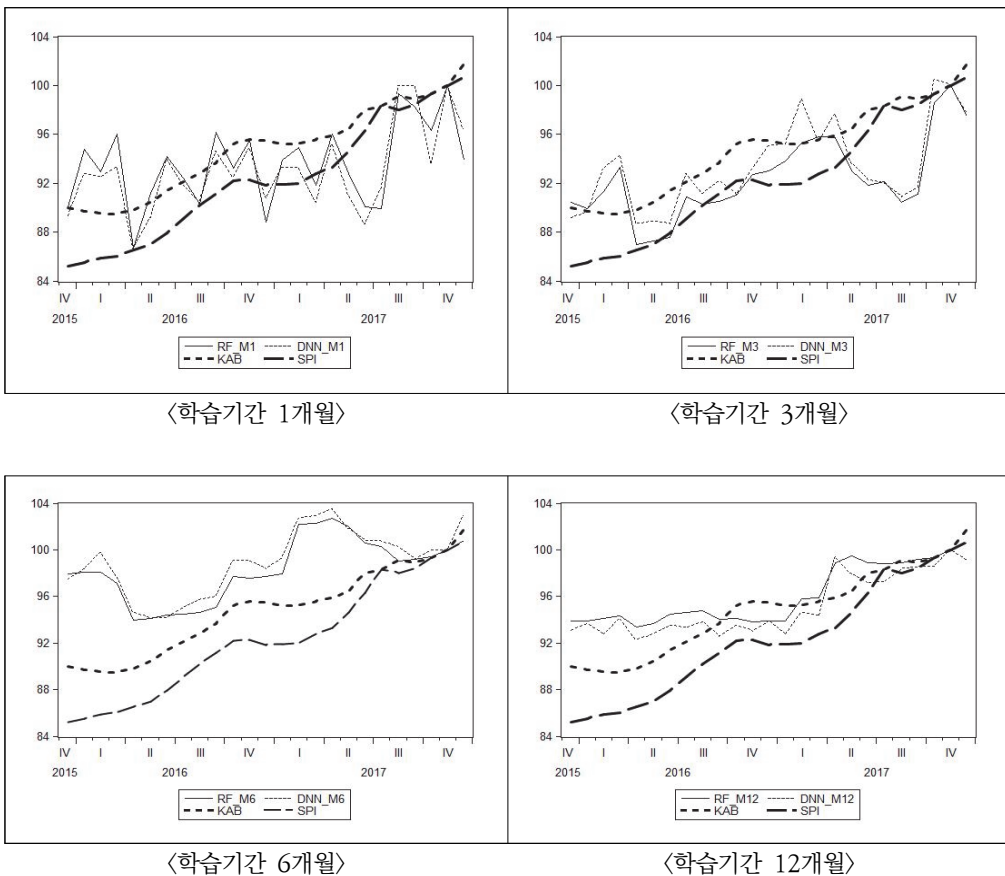
〈표 5〉 DNN지수, RF지수, KAB지수, 실거래가격지수(SPI)

구분	KAB	SPI	RF 가격 지수				DNN 가격 지수			
			1개월	3개월	6개월	12개월	1개월	3개월	6개월	12개월
'15.01	82.3	145.0	87.1				87.4			
'15.02	82.9	146.4	92.3				89.7			
'15.03	84.0	147.9	91.7	86.8			91.8	87.2		
'15.04	84.8	149.2	93.9	91.7			94.1	93.2		
'15.05	85.5	150.1	84.9	86.8			84.6	88.6		
'15.06	86.4	151.0	84.6	84.3	92.0		84.1	85.3	92.4	
'15.07	87.3	152.5	91.0	86.4	92.8		89.8	88.3	93.7	
'15.08	88.1	154.1	86.6	86.7	93.3		86.4	87.8	93.1	
'15.09	88.7	154.9	89.6	90.7	94.7		89.8	91.9	95.5	
'15.10	89.1	156.1	92.8	90.7	94.9		92.0	91.9	95.3	
'15.11	89.7	156.1	90.6	91.0	95.2		89.7	91.5	96.0	
'15.12	90.0	155.2	90.1	90.4	97.9	93.9	89.3	89.2	97.5	93.1
'16.01	89.7	155.8	94.8	89.9	98.1	93.9	92.8	89.6	98.4	93.7
'16.02	89.5	156.5	92.9	91.3	98.1	94.1	92.5	93.2	99.8	92.7
'16.03	89.5	156.7	96.0	93.3	97.1	94.3	93.4	94.3	97.7	94.1
'16.04	89.8	157.6	86.7	87.0	94.0	93.4	86.7	88.7	94.6	92.3
'16.05	90.4	158.6	91.2	87.3	94.1	93.7	89.2	88.9	94.1	92.8
'16.06	91.4	160.2	94.2	87.6	94.4	94.5	94.0	88.7	94.2	93.5
'16.07	92.1	162.3	92.4	90.9	94.5	94.6	91.9	92.8	95.1	93.3
'16.08	92.8	164.3	90.3	90.3	94.6	94.8	90.5	91.2	95.7	93.8
'16.09	93.7	166.2	96.2	90.5	95.1	94.0	94.6	92.2	96.0	92.6
'16.10	95.2	167.9	93.2	91.0	97.7	94.1	92.5	91.1	99.1	93.5
'16.11	95.6	168.1	95.5	92.7	97.6	93.8	94.9	93.2	99.0	93.0
'16.12	95.4	167.3	88.8	93.0	97.7	93.9	90.8	95.1	98.4	93.9
'17.01	95.2	167.4	93.9	93.8	97.9	93.9	93.3	95.2	99.3	92.7
'17.02	95.2	167.7	94.9	95.3	102.2	95.8	93.2	98.9	102.7	94.6
'17.03	95.6	169.0	91.8	95.8	102.3	95.9	90.4	95.4	102.9	94.3
'17.04	95.9	170.0	96.0	95.7	102.7	98.9	95.2	97.7	103.5	99.4
'17.05	96.4	172.4	92.8	93.0	102.0	99.5	91.0	93.6	101.8	97.9
'17.06	98.0	175.4	90.1	91.8	100.6	98.9	88.7	92.3	100.8	97.1
'17.07	98.3	179.1	89.9	92.1	100.3	98.8	91.6	92.0	100.7	97.3
'17.08	99.1	178.5	99.3	90.4	99.0	98.9	100.0	90.9	100.2	98.4
'17.09	98.9	179.3	98.2	91.1	99.2	99.2	100.0	91.7	99.2	98.5
'17.10	99.3	181.0	96.3	98.6	99.4	99.3	93.6	100.5	100.0	98.6
'17.11	100.0	182.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
'17.12	101.7	183.4	93.9	97.5	100.8	100.6	96.4	97.8	103.0	99.0

주: 기준시점은 2017년 11월임.

### 3. DNN지수 및 RF지수의 적정성 검토

〈그림 4〉에서 2015.12~2017.12까지의 장기 추세를 보면 대체로 우상향하는 모습을 보여주고 있으며 특히 학습기간이 12개월인 경우 KAB지수 및 실거래가격지수와 비교적 유사한 양상을 보여주고 있다. 하지만 월별 변동성에서는 KAB지수 및 실거래가격지수와 상당한 차이를 보여주고 있다.<sup>6)</sup>



〈학습기간 1개월〉

〈학습기간 3개월〉

〈학습기간 6개월〉

〈학습기간 12개월〉

주: 분석기간이 중복되는 2015년 12월부터 2017년 12월까지의 가격 지수이며, 실거래가격지수(SPI)는 2017년 11월을 기준시점으로 변환 후 그래프 작성하였음.

〈그림 4〉DNN지수, RF지수, KAB지수 및 실거래가격지수(SPI) 비교

6) 비교대상인 KAB지수는 강남구의 아파트매매가격지수이며, 실거래가격지수는 서울 전체 아파트를 기준으로 작성된 지수이다.



이용만·이상한(2008)은 ‘지수의 평활화’는 정책자 및 투자자들에게 위험을 과소평가함으로써 문제를 야기할 수 있다는 점을 지적한바 있으며, 이런 측면에서 부동산 가격지수의 평활화 여부는 지수의 적합성 및 활용 가능성을 판정하는데 있어서 중요한 기준이 될 수 있다. 지수의 평활화 여부는 실거래가격지수 변동률과의 관계를 분석하여 판단하는데 표준편차 분석, 부분조정모형, 시간변동계수모형 등을 통해 확인할 수 있다. 박연우·방두완(2011)은 실거래 아파트가격지수의 변화율인  $(\frac{P_t - P_{t-1}}{P_{t-1}})$ 이 평가기반 아파트가격지수의 변화율인  $(\frac{Q_t - Q_{t-1}}{Q_{t-1}})$ 에 미치는 영향을 추정한 식(3)과 같은 부분조정 회귀모형을 이용하여 지수의 평활화 여부를 분석하였다. 식(3)의  $\beta$ 값은 통해 평가기반 지수가 실거래지수 변화율을 얼마만큼 반영하는지를 확인할 수 있다.  $\beta$ 값이 클수록 상대적으로 평활화 정도가 낮은 것으로 이해할 수 있다.

$$\left(\frac{Q_t - Q_{t-1}}{Q_{t-1}}\right) = \alpha + \beta\left(\frac{P_t - P_{t-1}}{P_{t-1}}\right) + \epsilon_t \quad (3)$$

〈표 6〉은 실거래가격지수, KAB지수, KB지수, RF지수 및 DNN지수의 지수 변동률에 대한 평균 및 표준편차이다. RF지수와 DNN지수는 KAB지수 및 KB지수보다 변동성이 더

〈표 6〉 지수 변동률에 대한 평균 및 표준편차

구분	평균	표준편차
실거래가격지수	0.010118	0.010473
KAB 지수	0.006047	0.005317
KB 지수	0.005092	0.004009
RF1	0.002148	0.048124
RF3	0.003523	0.028223
RF6	0.003045	0.013352
RF12	0.002872	0.008333
DNN1	0.002800	0.043318
DNN2	0.003476	0.033314
DNN6	0.003620	0.014940
DNN12	0.002560	0.015020

큰 것으로 나타났다. 실거래가격지수와 비교하면, 학습기간 12개월을 적용한 RF지수를 제외한 모든 경우에서 RF지수 및 DNN지수의 변동성이 더 큰 것으로 나타났다.

〈표 7〉은 식(3)을 이용한 부분조정 회귀모형의 추정 결과이다. KAB지수, KB지수는  $\beta$ 가 모두 유의한 것으로 나타났다. KAB지수는 KB지수보다  $\beta$ 값이 더 크기 때문에 실거래가격의 변동추세를 더 많이 반영하고 있는 것으로 나타났다. 반면 RF지수 및 DNN지수는  $\beta$ 값이 모두 유의하지 않은 것으로 나타났으며, 조정된  $R^2$ 값 역시 매우 낮거나 음수 값이 산출되어 〈식 3〉을 통해서는 실거래가격지수의 변동률과 RF지수 및 DNN지수 변동률과의 관계를 분석하는 것이 적합하지 않은 것으로 나타났다.<sup>7)</sup>

부분조정 회귀모형에서는 RF지수 및 DNN지수는 실거래가격지수와 관련성을 확인하기 어렵다. 이러한 결과는 첫째, RF지수 및 DNN지수 산정에 있어서 재료가 되는 표본 주택 가격이 정성적인 분석없이 산출되었다는 점, 둘째, 머신 러닝에 의한 표본 가격 산정시 적용된 거래사례들이 시점에 따른 가중치가 반영되지 않은 점 때문인 것으로 판단된다.

〈표 7〉 부분조정 회귀모형 분석 결과

구분	$\alpha$	$p - value$	$\beta$	$p - value$	$adj.R^2$
KAB	0.0007	0.005 ***	0.3123	0.000 ***	0.2918
KB	0.0006	0.000 ***	0.2232	0.000 ***	0.2755
RF1	0.0025	0.317	-0.9889	0.260	0.0090
RF3	0.0015	0.331	-0.3444	0.509	-0.0175
RF6	-0.0010	0.147	-0.2010	0.408	-0.0102
RF12	0.0007	0.165	-0.0551	0.734	-0.0398
DNN1	0.0019	0.391	-0.6452	0.417	-0.0095
DNN3	0.0020	0.257	-0.6149	0.314	0.0014
DNN6	0.0009	0.246	-0.0855	0.754	-0.0320
DNN12	0.0012	0.186	-0.3084	0.286	0.0084

7) 실거래가격지수, KAB지수는 한국감정원 부동산 통계정보(www.r-one.co.kr), KB지수는 KB부동산 Live ON(nland.kbstar.com)에서 확인 가능하다.

## V. 결론

본 연구는 지수 산정 방식에 초점이 맞춰져 있던 기존 연구와 달리 지수 산정의 재료가 되는 표본 가격 산정에 초점을 맞췄다는 점, 부동산 가격지수 산정 과정에 있어서 머신 러닝 방법을 활용하였다는 점, 새로운 부동산 가격지수로서 머신 러닝 가격 기반 부동산 가격지수의 활용 가능성을 검토하였다는 점에서 의의가 있다.

본 연구의 결과를 요약하면 다음과 같다. 첫째, RF지수와 DNN지수는 서로 유사한 모습을 보이고 있으며, 학습기간에 따라 변동성이 달라지는 것으로 나타났다. 둘째, RF지수 및 DNN지수는 KAB지수 비교했을 때 장기적인 추세는 유사하나 단기적 추세에 있어서는 상당한 차이를 보이는 것으로 나타났다. 셋째, RF지수 및 DNN지수는 KAB지수, KB지수, 실거래가격지수보다 변동성이 큰 것으로 나타났으며, 실거래가격지수와의 관련성은 확인할 수 없었다.

본 연구의 시사점은 다음과 같다. RF지수 및 DNN지수는 장기간 추세에 있어서는 KAB지수 및 실거래가격지수와 유사성이 있으며 상당한 변동성을 보여주고 있으나, 실거래가격지수 변동률과의 관련성을 확인할 수 없기 때문에 시장상황을 적절히 반영한다고 인정하기 어렵다. 이러한 결과는 RF지수 및 DNN지수는 단순히 정량적인 분석만을 기반으로 작성되었기 때문인 것으로 이해된다. 하지만 RF지수 및 DNN지수는 기존 가격 지수와 비교했을 때 변동성이 크고, 표본 주택 가격 산출에 있어서 조사자의 주관이 개입될 여지가 없기 때문에 해당 시점의 가격을 비교적 객관적으로 산출 할 수 있는 방법이라고 할 수 있다. 따라서 현행 부동산 가격지수 산출에 있어서 참고자료로 활용함으로써 시장 조사자 또는 분석가들의 판단에 도움을 줄 수 있을 것으로 기대된다. 또한, 머신 러닝에 의해 산출된 표본 가격에 대한 조사자의 정성적 분석이 더해진다면 기존 가격지수를 개선할 수 있는 새로운 가격 지수로서의 활용 가능성 역시 높을 것으로 기대된다.

지수의 타당성은 산출된 부동산 가격지수가 실제 부동산 가격 추세를 얼마나 적절히 반영하고 있는지 여부이다. 본 연구는 적용된 학습기간에 따라 지수의 추이가 상이한데 적절한 학습기간 결정에 대한 기준을 제시하지 못하고 있다는 점에서도 한계점이 있다. 학습데이터의 기간이 짧아질수록 최근의 가격 추세를 반영할 가능성은 높으나 상대적으로 데이터의 양이 부족하기 때문에 가격의 변동성이 증가할 가능성이 높다. 반면 학습데이터의 기간

이 길어지는 경우 데이터의 양이 증가하기 때문에 가격의 변동성은 감소할 가능성은 높으나 상대적으로 최근의 가격 추세를 반영하기 어렵거나 지수가 평활화 될 가능성이 높아진다. 따라서 안정적인 가격의 변동성을 확보하면서 시장의 가격 추세를 반영하기 위해서는 적절한 학습기간의 결정 방법과 최근 자료의 영향력을 높일 수 있는 방법에 대한 추가적인 연구가 필요하다. 본 연구는 모형을 어떻게 최적화 할 것인가 하는 머신 러닝의 근본적인 문제점과 함께 최소한의 특성 자료만을 이용하여 표본 주택 가격을 산정하였다는 점에서 한계점이 있다. 또한 본 연구는 실거래자료를 활용하여 부동산 가격 지수를 산정하였는데 실거래자료의 노출시간을 고려하면 적시성 있는 가격 지수 산정이 어렵다는 문제점이 있다. 다만, 최근에는 부동산 정보 사이트를 통한 매물 자료의 수집이 가능하다는 점에서 실거래자료 뿐만 아니라 매물자료를 함께 활용한다면 적시성 있는 가격 지수 산정이 가능할 것으로 기대된다.

## 참고문헌

1. 구성관·홍석민, “심층신경망을 활용한 활주로 가시거리 예측 모델 개발,” 『한국항해학회 논문지』, 제21권 제5호, 한국항해학회, 2017, pp. 435-442.
2. 김재익·김달호·서안나·조애정, “반복매매모형에 기반한 아파트실거래가격지수의 산정 방식비교 : 표본산정방식과 하위지수 병합방식을 중심으로,” 『주택연구』, 제23권 제4호, 한국주택학회, 2015, pp. 47-71.
3. 국토해양부, 『전국주택가격동향조사 표본 및 통계 개편 방안 연구』, 2012.
4. 류강민·박유미·이창무, “비선형 회귀분석을 이용한 산술평균 반복매매지수 산정방법에 관한 연구,” 『주택연구』, 제17권 제4호, 한국주택학회, 2009, pp. 259-278.
5. 류강민·이상영, “S&P/Case-Shiller 반복매매모형을 이용한 주택가격지수 산정에 관한 연구,” 『주택연구』, 제18권 제2호, 한국주택학회, 2010, pp. 183-204.
6. 류강민·최성호·이상영, “서울시 단독다가구 연립다세대의 중위수 지수 산정에 관한 연구,” 『부동산학연구』, 제18권 제2호, 한국부동산분석학회, 2012, pp. 57-72.
7. 류강민·한제선·정상준·이창무, “TPL을 이용한 일단위 실거래 가격지수 산정방법에 관한 연구,” 『주택연구』, 제25권 제2호, 한국주택학회, 2017, pp. 5-23.
8. 민성욱, “딥 러닝을 이용한 주택가격 예측모형 연구,” 강남대학교 박사학위 논문, 2017.

9. 박연우·방두완, “평가기반 아파트가격지수에서의 비대칭 평활화 현상에 관한 연구,” 『주택연구』, 제19권 제2호, 한국주택학회, 2011, pp. 23-46.
10. 배성완·유정석, “딥 러닝을 이용한 부동산가격지수 예측,” 『부동산연구』, 제27집 제3호, 한국부동산연구원, 2017, pp.71-86.
11. 배성완·유정석, “기계 학습을 이용한 공동주택 가격 추정: 서울 강남구를 사례로,” 『부동산학연구』, 제24권 제1호, 한국부동산분석학회, 2018a, pp. 69-85.
12. 배성완·유정석, “머신 러닝 방법과 시계열 분석 모형을 이용한 부동산 가격지수 예측,” 『주택연구』, 제26권 제1호, 한국주택학회, 2018b, pp. 107-133.
13. 유진은, “랜덤 포레스트 : 의사결정나무의 대안으로서의 데이터 마이닝 기법,” 『교육평가연구』, 제28권 제2호, 한국교육평가학회, 2015, pp. 427-448.
14. 이창로, “비모수 공간모형과 앙상블 학습에 기초한 단독주택가격 추정,” 서울대학교 박사학위 논문, 2015.
15. 이창무·김병욱·이현, “반복매매모형을 활용한 아파트 매매가격지수,” 『부동산학연구』, 제8권 제2호, 한국부동산분석학회, 2002, pp. 1-19.
16. 이용만, “주택가격지수의 목적과 방법을 둘러싼 쟁점 - 실거래가격에 기초한 지수를 중심으로,” 『부동산학연구』, 제13권 제3호, 한국부동산분석학회, 2007, pp. 147-167.
17. 이용만·이상한, “국민은행 주택가격지수의 평활화 현상에 관한 연구,” 『주택연구』, 제16권 제4호, 한국주택학회, 2008, pp. 27-47.
18. 이해경·방송희·이용만, “경매시장의 주택가격지수 추정에 관한 연구 - 강남3구의 아파트를 중심으로,” 『부동산학연구』, 제16권 제2호, 한국부동산분석학회, 2010, pp. 101-117.
19. Bailey, M. J., R. F. Muth and H. O. Nourse, “A Regression Method for Real Estate Price Index Construction,” *Journal of American Statistical Association*, Vol. 58, 1963, pp. 933-942.
20. Berry, B. and R. Bednarz, “A Hedonic Model of Prices and Assessments for Single-Family Homes: Does the Assessor Follow the Market or the Market Follow the Assessor?,” *Land Economics*, Vol. 51 No. 1, 1975, pp. 21-50.
21. Bourassa, S. C., M. Hoesli and J. Sun, “A Simple Alternative House Price Index Method,” *Journal of Housing Economics*, Vol. 15, 2006, pp. 80-97.

22. Brieman, L., "Random forests," *Machine learning*, Vol. 45 No. 1, 2001, pp. 5-32.
23. Case, K. E, and R. J. Shiller, "Prices of Single Family Home Since 1970: New Indexes for Four Cities," *New England Economics Review*, 1987, pp. 45-56.
24. Goolsby, W., "Assessment Error in the Valuation of Owner-Occupied Housing," *Journal of Real Estate Society*, 1997, pp. 33-45.
25. Hinton, G. E., S. Osindero and Y. W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation*, Vol. 18 No. 7, 2006, pp. 1527-1554.
26. Lancaster, K. J., "A New Approach to Consumer Theory", *Journal of Political Economics*, Vol. 74 No. 2, 1966, pp. 132-157.
27. Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", *Journal of Political Economy*, Vol. 82, 1974, pp. 34-55.
28. nland.kbstar.com, KB부동산 Live ON
29. rt.molit.go.kr, 국토교통부 실거래가 공개시스템
30. www.ret.co.kr, 부동산테크
31. www.r-one.co.kr, 한국감정원 부동산통계정보
32. sebastianraschka.com/faq/docs/evaluate-a-model.html, Dr. Sebastian Raschka

- 
- 접수일 2018. 06. 05.
  - 심사일 2018. 06. 18.
  - 심사완료일 2018. 10. 12.

## 국문요약

### 표본 주택 가격 기반 부동산 가격지수 산정: 머신 러닝 방법의 활용을 중심으로

본 연구의 목적은 '머신 러닝에 의한 표본 주택 가격'을 기반으로 부동산 가격 지수를 산정하는 것이다. 머신 러닝 방법인 랜덤 포레스트(random forest)와 심층신경망(deep neural networks)을 활용하여 표본 주택의 가격을 산정하였으며, 제본스 지수(Jevons index) 산정 방법을 이용하여 부동산 가격 지수를 산정하였다. 연구결과 첫째, RF지수와 DNN지수는 서로 유사한 모습을 보이고 있으며, 학습기간에 따라 변동성이 달라지는 것으로 나타났다. 둘째, RF지수와 DNN지수는 KAB지수와 비교했을 때 장기적 추세는 유사하나 단기적 추세는 상당한 차이를 보이는 것으로 나타났다. 셋째, KAB지수, KB지수, 실거래가격지수보다 변동성이 큰 것으로 나타났으며, 실거래가 지수와의 관련성은 확인할 수 없었다. RF지수와 DNN지수에는 대한 조사자의 정성적 분석이 더해진다면 기존 가격 지수를 개선할 수 있는 새로운 가격 지수로서의 활용 가능성이 높을 것으로 기대된다.