



## 기계학습 기반 주택시장의 조기경보체계\*

### Early Warning System of Housing Market Using Machine Learning

박대현\*\* · 김정환\*\*\* · 류두진\*\*\*\*

Daehyeon Park · Jeonghwan Kim · Doojin Ryu

#### ■ Abstract ■

This study proposes an early warning system for risks of the housing market based on machine learning models. We adopt a signal approach to detect the housing market risk and establish the early warning system using classification methods. Considering the moment when the housing market falls into recession as a warning signal, we set the signal as the price which is more than the sum of the average and standard deviation of upcoming prices. The detected signals are consistent with empirical observations in the Korean housing market. We select the best performing function among classification models for machine learning which predicts a warning signal. As a result of an intercomparison of models including the logistic regression, the support vector machine, the random forest and the artificial neural network with the use of inputs such as housing price indices, macroeconomic variables and other housing market variables, we find that the random forest demonstrates the highest prediction performance. Our early warning system yields policy implications in terms of relevant detection of price fluctuations in the housing market.

**Keywords:** Classification, Early warning system, Housing market, Machine learning, Signal approach

\* 본 논문에 대하여 유익한 조언을 주신 익명의 세분 심사위원님께 감사드립니다.

\*\* 성균관대학교 경제학과 박사과정(주저자) | Ph.D. Candidate, Department of Economics, Sungkyunkwan University | First Author | pdh829@skku.edu |

\*\*\* SK증권 구조화금융2팀 주임 | Chief, Structured Finance, SK Securities | rlajwdghks9@naver.com |

\*\*\*\* 성균관대학교 경제학과 정교수(교신저자) | Professor, Department of Economics, Sungkyunkwan University | Corresponding Author | sharpjin@skku.edu |

## 1. 서론

주택시장의 위험을 예측하는 것은 부동산 연구에 있어서 중요한 연구주제이다. 주택가격이 내재가치(intrinsic value)와 비교하여 비정상적으로 증가한 불안정한 상황이 지속하다가, 어느 순간 과도하게 올라간 주택가격이 폭락하면 경제 전반에 큰 충격을 주게 된다. 글로벌 금융위기를 포함한 다수의 크고 작은 금융위기가 주택시장의 위험에서 초래되었으며, 주택시장의 위험은 경제 전반에 걸친 위험으로 전이된다(박원암, 2014; Brander and Koetse, 2011; Chiu, 2006, 2008). 이에 본 연구에서는 주택시장의 위험을 효과적으로 예측할 방법으로 기계학습(machine learning) 방법론에 기반을 둔 조기경보체계(early warning system)를 제시한다.

우리나라에서 주택은 역사적으로 예금보다 수익성이 높고, 주식보다 위험이 낮은 것으로 인식되며, 투자자산으로 선호되었다(Seok et al., 2020). 주택은 일반 재화처럼 사용을 통한 효용이 존재할 뿐만 아니라, 재건축 등을 통하여 감가상각을 초과하는 자본이득을 기대할 수 있어 소비재와 투자자산의 속성을 동시에 지니고 있다(강희만·김정렬, 2012). 주택가격이 하방 경직성을 갖는 것도 주택이 비교적 안전한 투자자산으로 기능할 수 있는 이유 중 하나이다(김승욱·이창석, 2009). 한편, 주택은 공급 기간이 평균적으로 수년 이상 소요되며, 주택공급은 상당히 비탄력적이다. 이로 인해 주택시장에 초과수요가 존재하면 가격이 지속해서 상승할 가능성이 있다. 이는 주택시장이 과도한 가격 상승압력에 노출될 수 있

음을 의미한다(류지수 외, 2007).

국내 부동산시장에서, 초과수요 및 투기수요, 공급을 위한 택지의 절대적 부족은 주택시장의 불안정성을 높인다. 주택시장의 불안과 위험은 주택실수요자는 물론 정책입안자에게 주요 관심사이다. 학계에서는 이러한 주택시장의 위험을 이론적 모형과 데이터에 기반하여 예측하려는 시도가 있다(이기영, 2013, 2014). 주택시장의 위험을 예측하는 기존 연구와 경보체계는 주로 고전적인 시계열 예측모형에 의존한다(손정식 외, 2003; 이기영, 2015; 이용재·박철범, 2019). 그러나, 주택 정책을 비롯한 다양한 변수의 영향이 혼재하는 주택시장은 단순한 시계열 모형으로 예측하기에 적합하지 않을 수 있다. 따라서, 진보된 방법론에 근거하여 주택시장의 위험을 예측하는 조기경보 체계에 관한 연구가 필요하다.

본 연구에서는 신호접근법(signal approach)으로 주택시장에 대한 조기경보체계를 구축한다. 주택시장에서의 위험신호를 새롭게 정의하고, 이를 기계학습 분류모형(classification model)으로 예측한다. 주택시장의 위험신호를 정의하는 방법으로 여러 통계적 방법이 사용되었으나, 기존의 방법론은 주택가격이 상승하는 시기와 하락하기 시작하는 시기를 모두 위험신호로 분류한다는 점에서 한계를 갖는다(Hagemann and Wohlmann, 2019). 본 연구에서는 주택시장의 침체기를 알려주는 위험신호를 정의하고 이를 경험적 사실과 비교한다. 경기가 침체기에 진입하면 상당 기간 가격이 지속해서 하락하는 추세를 보이기 때문에, 현재 가격이 미래가격의 평균과 표준편차의 합보다 클 경우를 위험신호로 정의한다. 이렇게 정의한 위험신호

는 우리나라 주택시장에서의 경험적 발견과 일치한다.

선행연구와는 달리, 본 연구에서는 위험신호를 기계학습 방법론을 사용하여 예측한다. 2004년 11월부터 2020년 12월까지 장기간의 주택가격지수, 거시경제변수, 주택시장변수를 활용하여 서포트 벡터 머신(support vector machine, SVM), 랜덤 포레스트(random forest, RF), 인공신경망(artificial neural network, ANN)과 같은 다양한 기계학습 방법론의 예측성결과를 상호 비교한다. 분석결과, RF가 가장 우수한 성능을 보이며, SVM과 ANN 모형이 그와 유사한 성능을 보여준다. 이러한 기계학습 방법론에 기반한 예측성 결과는 기존 연구에서 사용되던 로지스틱 회귀 분석(logistic regression)의 예측성결과를 능가함을 확인하였다.

본 연구의 주요 발견은 다음과 같다. 첫째, 미래 가격의 평균과 표준편차로 도출한 주택시장의 위험신호가 국내 주택시장의 경험적 사실과 일치함을 확인하였다. 이는 본 연구에서 정의한 위험신호가 한국 주택시장의 침체를 감지하는 신호로 적절함을 의미한다. 둘째, 예측성 결과가 우수한 기계학습 방법론으로 조기경보체계를 구축하면 현실에서의 설명력과 사용 가능성을 높일 수 있다. 이러한 결과는 기계학습 방법을 활용한 조기경보체계가 주택시장에서의 정책에 대한 참고자료로 활용될 수 있음을 의미한다.

본 연구의 구성은 다음과 같다. 제2장은 선행연구를 정리한다. 제3장은 주택시장의 위험신호를 정의하고 이를 경험적 사실과 비교하며, 제4장은 기계학습 모형을 활용하여 주택시장의 조기경보

체계를 구축한다. 마지막으로 제5장은 연구의 내용을 요약하고 결론을 제시한다.

## II. 선행연구

조기경보체계는 미래에 발생할 수 있는 위기의 신호를 사전에 포착하여 선제 대응이 가능하도록 고안한 시스템이다. 1990년대 아시아 통화위기 이후, 위기에 대한 선제 대응의 필요성이 증대되었고, 이에 따라 금융 및 부동산시장에서 조기경보체계에 관한 연구가 증가하였다. Beckmann et al.(2006)은 시계열 모형을 대체하여 선제적인 주택가격정책을 수립하는 방안으로 조기경보체계가 효과적이라고 주장한다. 조기경보체계에 관한 연구에서는 Kaminsky et al.(1998)이 제안한 신호접근법이 주로 사용된다. 신호접근법은 위험신호를 구성하고, 분류모형을 통해 이를 예측한다. Frankel and Rose(1996)가 제안한 프로빗 모형(probit model)은 이와 달리 위기 발생의 조건부 확률을 구한다. 본 연구에서는 프로빗 모형보다 신호접근법이 국내 주택시장에 대한 예측력이 우수하다는 선행연구를 참고한다(서승환, 2004).

신호접근법을 활용할 때, 위기 신호를 적절하게 정의하는 것이 중요하다. 일반적으로 내재가치보다 자산가격이 과도하게 상승한 상황을 가격 거품(bubble)이 존재하는 상황으로 정의한다(Shiller, 2000). 거품의 생성과 붕괴는 시장의 위험 상황을 판단하는 중요한 문제이므로, 거품의 존재 여부를 통계적으로 검정하는 방법에 대한 다양한 논의가 이루어지고 있다. Diba and Grossman

(1988)은 주식의 내재가치인 배당으로 주가가격의 변동을 설명할 수 있다면 두 변수 사이에 공적분(cointegration) 관계가 있을 것이라고 주장하며, 단위근 우측 검정을 통한 거품 검정법을 제안한다.

그러나 공적분에 기반을 둔 거품 검정은 주기적으로 생성 붕괴를 반복하는 폭발적인 거품을 추적하는 데는 한계가 있다(Evans, 1991). 이에 Phillips et al.(2011)은 단위근 검정을 반복하는 Supremum Augmented Dickey-Fuller(SADF) 검정 방법을 제안하고, 이 방법이 공적분 관계에 바탕을 둔 기존 방법보다 거품 붕괴와 기간의 추정에 효과적임을 보였다. Phillips et al.(2015)은 거품의 생성과 붕괴가 여러 번 나타나는 상황을 고려하기 위해 SADF 검정을 일반화한 Generalized SADF(GSADF) 검정을 제안한다. GSADF는 SADF와 달리 검정 대상이 되는 표본을 이동시키는 방법을 사용한다. GSADF는 주식시장과 기타 자산 시장에서의 거품 추정에 사용되고 있으며(Chuliá et al., 2017; Corbet et al., 2019; Geuder et al., 2019; Harvey et al., 2016), 주택 가격의 거품 분석에도 활용되고 있다(Anundsen et al., 2016; Giglio et al., 2016; Park and Ryu, 2021; Pavlidis et al., 2016).

이와 달리 시장가격이 일정 수준 이상을 넘을 때를 거품이 존재하는 시점으로 판단하는 방법도 있다. 시장압력지수(market pressure index)를 정의하고, 해당 지수의 평균과 표준편차를 활용하여 신호를 정의하는 것이 이에 해당하는데, 주로 외환시장의 거품 신호를 검정하는 데 사용되지만, 주택시장을 분석하는데 이를 활용한 사례

도 있다(Suh et al., 2011). 이와 유사한 방법으로 시장가격의 장기적 추세를 도출하고, 가격이 추세를 넘어서 시점에서 거품이 존재한다고 판단하는 방식도 있다. Jordà et al.(2015)은 HP 필터(Hodrick-Prescott filter)로 가격의 추세를 추정하고 추세 이상으로 가격이 높은 시점을 거품이 존재하는 시점으로 판단하는 방법을 제안한다. 이 방식은 거시경제모형에서 주택시장을 설명할 때 사용된다(김경훈, 2018; Knoll et al., 2017). 한편, HP 필터는 계산이 편리하고 응용력이 높아 경제변수의 순환변동을 설명하는 데 주로 사용되었으나(Yu and Ryu, 2020), HP 필터로 추출한 순환변동이 경제적 의미를 갖지 않을 수 있다고 주장하는 연구도 있다(권순신 · 최성호, 2019; Hamilton, 2018).

지금까지 논의된 방법들은 시장가격에 거품이 존재하는 상황을 효과적으로 잘 감지하는 것으로 평가받고 있지만, 거품이 형성되어 가격이 상승하는 시기와 거품이 꺼지면서 가격이 하락하고 침체에 접어드는 시기를 구분하기 힘들다는 문제가 있다. 이에 본 연구에서는 시장압력지수를 사용하는 방식과 표본을 이동하여 추정하는 방식을 참고하여 주택시장의 침체를 감지할 수 있는 위험 신호를 새롭게 정의한다.

조기경보체계에 관한 연구는 주로 금융시장에 대해 이루어지고 있으나(Aldasoro et al., 2018; Alessi and Detken, 2018), 부동산시장과 주택 시장에 대해서도 활발하게 진행되고 있다. 특히, 최근에는 예측성고가 우수하다고 평가받고 있는 기계학습 분류모형을 조기경보체계에 활용하는 연구가 나타나고 있다. Hu and Pang(2008)은

SVM이 ANN, MDA(mean decrease accuracy), 그리고 로지스틱 회귀모형과 같은 다른 분류모형보다 조기경보체제 구축에 있어 더 높은 수행능력과 예측력을 지닌다는 것을 실증적으로 보인다. Samitas et al.(2020)은 주요 선진국 금융시장에 대한 조기경보체제를 구축하는 과정에서 여러 기계학습 분류모형을 상호비교한다. Wang et al.(2020)은 LSTM(long short-term memory) 모형을 활용하여 중국 금융시장의 조기경보체제를 구축한다. Wang et al.(2020)과 Zhou et al.(2021)은 중국 부동산시장에 대한 조기경보체제를 SVM 모형을 활용하여 설계한다. 그러나 국내 주택시장을 대상으로 조기경보체제를 연구한 논문은 드물며, 특히 기계학습 방법을 활용한 연구는 아직 진행되지 못하고 있다. 우리나라의 주택시장은 전세라는 독특한 제도가 있다는 점과 강남과 같은 특정 지역이 주도하는 시장이라는 점에서 해외 사례와 차별되므로, 국내 주택시장에 관한 별도의 연구가 필요하다(Kim et al., 2018). 따라서 본 연구는 기계학습 방법론을 활용한 국내 주택시장의 조기경보체제를 제안한다.

### III. 주택시장의 위험신호

본 절에서는 주택시장의 침체를 알려주는 신호를 정의한다. 장기적인 추세에서 벗어나는 경우를 가격에 거품이 형성된 시점으로 판단할 수도 있으나, 이 경우, 가격이 급등하는 경우까지 위험신호로 판단하게 된다. 본 연구에서는 주택가격이 일정 기간 하락하는 침체기를 보여주는 신호를

구성하기 위하여, 현재의 주택가격지수가 미래의 주택가격지수보다 일정 수준 이상인 경우를 위험신호로 정의한다. 그러나 본 연구에서는 주택시장의 거품이 붕괴되고 침체기에 들어가는 시기를 감지하기 위해, 다른 방법론을 제안한다.

주택시장이 침체기에 진입할 경우, 주택가격은 일정 기간 하향하는 추세를 보인다. 이는 현재 시점의 가격이 미래 일정 기간의 가격보다 지나치게 높은 상황이라는 것을 의미한다. 이에 따라 본 연구에서는 특정 시점의 가격이 향후 1년간 가격 평균보다 일정 수준 이상 높은 경우 위험신호가 발생했다고 정의한다. 본 연구에서는 선행연구를 참고하여 다음과 같이 주택가격지수의 평균과 표준편차를 바탕으로 위험을 감지하는 신호를 정의한다.  $t$ 시점의 주택가격을  $P_t$ 라고 할 때,  $t$ 시점을 포함하여 향후 1년간 주택가격의 표본평균( $mean_{12}(P_t)$ )과 표본표준편차( $std_{12}(P_t)$ )를 각각 (식 1)과 (식 2)와 같이 계산한다.  $t$ 시점에서의 주택가격의 위험신호( $Signal(P_t)$ )는 (식 3)과 같이 정의된다.

$$mean_{12}(P_t) = \frac{1}{12} \sum_{k=t}^{t+11} P_k \quad (\text{식 1})$$

$$std_{12}(P_t) = \sqrt{\frac{1}{12-1} \sum_{k=t}^{t+11} (P_k - mean_{12}(P_t))^2} \quad (\text{식 2})$$

$$Signal(P_t) = \begin{cases} 1, & \text{if } P_t > mean_{12}(P_t) + std_{12}(P_t) \\ 0, & \text{if } P_t \leq mean_{12}(P_t) + std_{12}(P_t) \end{cases} \quad (\text{식 3})$$

본 연구에서는 주택가격의 대용변수로 2004



년 11월부터 2020년 12월까지의 월별 주택가격 지수를 사용한다. 한국 주택시장의 특성상 매매가격과 전세가격이 모두 중요한 역할을 한다는 점을 고려하여 전국 주택매매가격지수(KSI), 전국 주택전세가격지수(KCI)를 사용하였다(Kim, 2017). 또한, 강남이 한국 주택시장에 미치는 영향이 크다는 점을 고려할 필요가 있다. 주택시장에 관한 다수의 선행연구가 한국 주택시장에서 강남이 주변 지역에 미치는 효과가 크다는 점을 지적하고 있다. 강남의 주택가격이 다른 지역의 주택가격에 대해 전이 효과가 있다는 실증적인 분석이 다수 제시되고 있다(이석희 · 김수현, 2014; Al-Yahyaee et al., 2020; Hyun and Milcheva, 2018; Lee and Lee, 2019). 이를 반영하기 위하여 서울 강남의 주택매매가격지수(GSI), 서울 강남의 주택전세가격지수(GCI)를 사용하였다. 주택 수요가 계절적 요인에 영향을 받을 수 있다는 점을 고려하여 각각의 자료는 전년 동월 대비 변화율로 변환하여 사용하였다.

4개의 가격지수에서 모두 위험신호가 나타났을 때, 주택시장 전체의 위험이 나타난다고 정의한다. 주택시장의 위험신호를  $HBS_t$ 라고 할 때, 이는 4개 가격지수로 만든 위험신호를 모두 곱한 값과 같다. (식 4)와 같이 정의된다.

$$HBS_t = \prod_{i=1}^4 Signal(P_{i,t}) \quad (\text{식 4})$$

〈표 1〉은 주택가격지수와 위험신호의 기술통계량(descriptive statistics)을 정리한 것이다. 〈그림 1〉은 주택시장의 위험신호  $HBS_t$ 와 주택가격지수인  $KSI_t$ ,  $KCI_t$ ,  $GSI_t$ ,  $GCI_t$ 의 움직임을 보

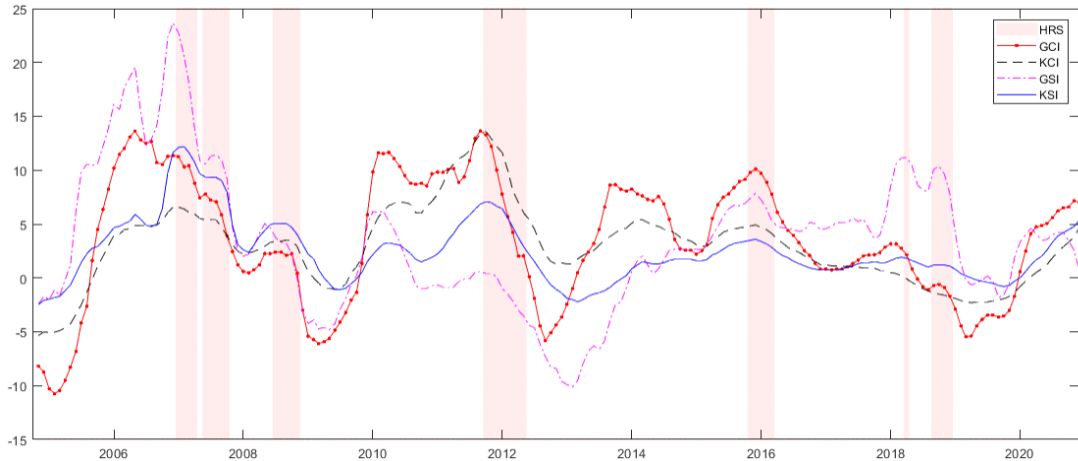
〈표 1〉 주택가격지수와 위험신호의 기술통계량

	N	Mean	Std.	Min	Max
$GCI$	194	3.55	5.78	-10.7	13.66
$KCI$	194	2.93	3.71	-5.35	13.61
$GSI$	194	3.53	6.45	-10.1	23.65
$KSI$	194	2.45	2.93	-2.46	12.20
$HBS$	183	0.17	0.38	0	1

주 : 위 표는 주택가격지수와 위험신호의 기술통계량을 제시한다. 표본 수(N)와 평균(Mean), 표준편차(Std.), 최솟값(Min), 최댓값(Max)을 나타낸다.  $GCI$ ,  $KCI$ ,  $GSI$ ,  $KSI$ ,  $HBS$ 는 각각 강남 아파트 전세가격지수, 전국 아파트 전세가격지수, 강남 아파트 매매가격지수, 전국 아파트 매매가격지수, 주택시장의 위험신호를 의미한다.

여준다. 〈그림 1〉에서 볼 수 있듯이,  $HBS_t$ 는 주택가격지수가 모두 일정 기간 하락하는 시기를 감지할 수 있다.

〈그림 1〉에서 한국 주택시장에서 위기 기간은 총 네 번 관측되는 것으로 확인할 수 있다. 첫 번째는 2008년 글로벌 금융위기 전후 기간이다. 2006년 말과 2007년 초에 정부의 부동산 대책이 발표되고, 주택담보대출 규제가 강화되면서 주택 수요가 위축되기 시작하면서 주택시장이 침체된다. 그리고 2008년 글로벌 금융위기 이후 경기 부진으로 인해 주택 수요가 감소하게 되면서 주택가격이 계속 하락하는 추세를 보인다. 두 번째는 2011년 말과 2012년 초의 기간으로, 유럽의 재정위기가 경기 악화로 이어지면서 주택시장이 위축된 기간이다. 세 번째는 2015년 말과 2016년 초로, 금융감독원의 여신심사 선진화 가이드라인 발표와 정부의 가계대출규제 시행 예정 발표 등으로 인해 주택 수요가 위축된 기간에 해당한다. 마지막으로 2018년 하반기에는 주택가격 안정을



주 : 위 그림은 2004년 11월부터 2020년 1월까지의 주택가격위험신호(HBS)와 강남아파트전세가격지수(GCI), 전국아파트전세가격지수(KCI), 강남아파트매매가격지수(GSI), 전국아파트매매가격지수(KSI)를 그림으로 나타낸 것이다.

〈그림 1〉 주택가격위험신호와 주택가격지수

위한 규제강화와 공급확대 정책이 추진되면서 전체적으로 주택가격이 하락하는 추세를 보였다. 이처럼 본 연구에서 정의한  $HBS_t$ 는 국내 주택시장의 침체기에 관한 경험적 발견과 일치한다. 따라서  $HBS_t$ 를 국내 주택시장의 위험신호로 사용할 수 있다.

#### IV. 기계학습을 통한 예측

본 연구에서는  $HBS_t$ 를 예측하는 모형으로 기계학습 방법론을 검토한다. 총 4개의 분류모형으로  $HBS_t$ 를 예측하고 성과를 비교한다. 분류모형은 주어진 데이터 집합을 바탕으로 새로운 데이터가 어느 집단에 속할 것인지 판단하는 모형을 의미한다(Kim et al., 2021a). 기존 연구에서 분류

기법으로 주로 사용된 모형은 로지스틱 회귀모형이며, 기계학습 방법론 중에서는 SVM, RF, ANN 모형이 사용된다.

SVM은 서로 다른 집단으로 분류되는 데이터 간의 거리(margin)를 최대화하는 초평면(hyperplane)을 구하는 방법으로 데이터를 분류한다(Vapnik, 2000). SVM은 커널 함수(Kernel function)를 통해 비선형문제를 고차원의 공간에서의 선형문제로 대응시켜 적은 수의 표본만으로도 상대적으로 우수한 예측성고를 기대할 수 있다(안현철 외, 2006).

SVM의 커널 함수로는 다항(polynomial) 커널과 RBF(radial basis function) 커널이 대표적으로 사용된다. 본 연구에서는 성능이 좋다고 평가받는 RBF 커널을 사용한다(Hsu et al., 2016). RBF 커널은 가우시안(Gaussian) 커널이라고도 불리며, 다음 (식 5)와 같다.

$$k(x_1, x_2) = \exp(-\gamma \|x_1 - x_2\|^2) \quad (\text{식 5})$$

RF의 하이퍼 파라미터(hyper parameter)가 필요하다. 하이퍼 파라미터는 연구자가 외생적으로 조정하는 모수로, 분류에 대한 오류의 허용범위와 관련된 SVM의 비용 모수인  $C$ 와 하나의 데이터가 영향력을 행사하는 거리를 나타내는 RBF 커널 함수의 모수인  $\gamma$ 가 있다. 본 연구에서는 베이지안 최적화(Bayesian optimization)를 통해 두 파라미터를 최적화한다. 최적화된 파라미터 값은 각각  $C=937.35$ ,  $\gamma=71.28$ 이다.

RF는 Breiman(2001)이 제안한 모형으로, 의사결정나무(decision tree)에 기반을 둔 앙상블 학습(ensemble learning)모형이다. 의사결정나무는 입력변수의 조합으로 구성된 공간을 서로 겹치지 않는 영역으로 나눈 후, 새로운 관측치가 분할된 영역 중 어느 곳에 속하는지를 바탕으로 분류를 수행하는 모형이다. 예측 오차를 최소화하는 형태를 추정하는 것이 의사결정나무의 목적 함수이지만, 분할된 공간의 개수에 제약이 없으면 학습 데이터에 대해 과적합(overfitting)될 가능성이 있다(김형준 외, 2019).

앙상블 학습은 의사결정나무의 과적합 문제를 해결하는 방법의 하나이다. 앙상블 학습은 여러 모형을 결합하여 학습하는 방법으로, 크게 취합(aggregation) 방법론과 부스팅(boosting) 방법론으로 나뉜다. RF는 취합 방법론 중에서 배깅(bagging)을 활용하여 여러 개의 트리를 결합하는 학습방법을 사용한다. 배깅은 부트스트래핑(bootstrapping)을 통해 여러 의사결정나무를

생성하고, 각 의사결정나무의 예측결과와 평균을 최종 예측결과로 제시하는 방법을 의미한다. RF는 배깅 모형에서 의사결정나무를 생성할 때 사용하는 입력변수(input feature)의 개수까지 무작위로 샘플링(sampling)하는 앙상블 모형을 말한다. RF의 하이퍼파라미터는 의사결정나무의 최대 분할 수와 취합할 모형의 수, 샘플링할 입력변수의 개수가 있는데, SVM에서와 마찬가지로 베이지안 최적화 기법을 통해 최적화하였다. 최적화 결과, 최대 분할 수는 22개, 모형의 수는 10개, 샘플링할 입력변수의 개수는 17개를 사용하였다.

ANN은 생물의 뇌가 정보를 처리하는 과정에 착안하여 설계된 모형으로, 뉴런(neuron)이 정보를 순차적으로 전달하여 결과를 도출하는 과정을 단순화한 형태로 이루어져 있다(Kim et al., 2020). ANN은 실제 뉴런의 역할을 수학적으로 모형화한 인공뉴런을 여러 계층에 걸쳐 연결시킨다. ANN의 계층은 입력계층(input layer)과 은닉 계층(hidden layer), 출력 계층(output layer)으로 구성되어 있다(Kim et al., 2021b). 인공뉴런은 입력변수가 일정 기준(threshold)을 넘으면 다음 뉴런에 출력 신호를 전달한다. ANN은 각 뉴런의 가중치와 임계치를 조정하여 예측력을 높인다. ANN은 은닉 계층의 크기와 깊이에 영향을 받는다. 본 연구에서는 은닉 계층의 크기와 깊이를 조정(tuning)하여 예측력이 가장 높았던 은닉 계층이 2개이고, 각각의 은닉 계층의 크기가 10인 ANN을 사용한다.

각 모형의 예측성과는 정확도(accuracy)와 ROC(receiver operating characteristic) 곡



선, AUC(area under the curve)로 평가한다.

정확도(*Accuracy*)는 (식 6)과 같이 계산할 수 있다.  $TP$ (true positive)는  $HBS_t$ 의 예측값과 실제값이 모두 1('Risk')인 경우의 개수를,  $TN$ (true negative)은  $HBS_t$ 의 예측값과 실제값이 모두 0('No risk')인 경우의 개수를 의미한다.  $FP$ (false positive)는 예측값은 1이지만, 실제값은 0인 경우이며,  $FN$ (false negative)은 예측값은 0이지만, 실제값은 1인 경우의 개수를 나타낸다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (\text{식 6})$$

ROC 곡선은 x축이 참 양성률(true positive rate, TPR)이고 y축이 거짓 양성률(false positive rate, FPR)인 평면에 그려진다. TPR과 FPR은 각각 (식 7), (식 8)과 같이 계산할 수 있다.

$$TPR = \frac{TP}{TP + FN} \quad (\text{식 7})$$

$$FPR = \frac{FP}{FP + TN} \quad (\text{식 8})$$

주어진 거짓 양성률 비율에 대해 참 양성률이 더 높게 나타나는 것이 바람직하므로, ROC 곡선이 좌상단에 가까울수록 높은 성과를 가진 모델이라고 평가한다. AUC는 ROC 곡선 아래 영역의 넓이로, ROC 곡선이 좌상단에 가까울수록 큰 값을 갖는다.

본 연구에서는  $t-1$ 시점의 주택가격지수와 거시경제변수, 주택시장변수를 입력변수로  $HBS_t$ 를 예측하는 모형을 구성한다. 주택가격지수로는 앞서 사용한 KSI, KCI, GSI, GCI를 사용한다. 거

시경제변수로는 다음과 같은 변수들을 사용한다. 금리와 유동성은 주택가격에 유의미한 영향을 미친다는 연구를 참고하여, 총 유동성( $m3$ )과 고객예탁금( $dep$ ), 예대율( $dlt$ ), 3년 만기 국고채 금리( $rr3$ )를 사용한다(김중규·정동준, 2012; 송인호, 2014; 이태리·송인호, 2015; 허종만·이영수, 2018). 주가지수가 주택시장에 미치는 영향을 고려하기 위해 종합주가지수( $ps$ )를 포함하였다(이강용 외, 2015; 임병진·한성윤, 2009). 그 외에 대표적인 거시지표인 대미달러환율( $er$ ), 경기선행지수( $lir$ ), 소비자물가지수( $cpi$ ), 산업생산지수( $ip$ )를 포함하였다(전해정, 2014; 홍정효·문규현, 2009). 주택시장변수로는 관련 선행연구를 참고하여 주택시장의 수급에 영향을 미치는 변수인 상장건설업 주가지수( $pscon$ ), 대규모 건설주액( $cont$ ), 아파트 공급호수( $has$ ) 그리고 전국 주거지역 토지거래면적( $tlarea$ )을 포함하였다(민선옥·이영수, 2019; 이재율, 2004; 정건섭 외, 2011). 모든 변수는 월별 자료이며, 표본 기간은 2004년 11월부터 2020년 12월까지이다. 대부분 자료는 전년 동월 대비 증감률로 변환하여 사용하였고, 총 유동성, 국고채 금리, 예대율은 각각 전월대비, 3개월 전 대비 그리고 2개월 전 대비 증감률을 사용하였다. <표 2>는 거시경제변수와 주택시장변수의 기술통계량을 정리한 것이다.

전체구간에서 80%를 학습 데이터(training data)로 사용하고, 나머지 20%를 테스트 데이터(test data)로 사용한다. 즉, 입력변수인 주택가격지수, 거시경제변수, 주택시장변수는 2004년 11월부터 2016년 12월까지의 데이터를 학습 데이터로 사용하고, 2017년 1월부터 2019년 12월

〈표 2〉 거시변수와 주택시장변수의 기술통계량

	N	Mean	Std.	Min.	Max.
<i>m3</i>	194	0.07	0.03	-0.17	0.13
<i>dep</i>	194	0.14	0.31	-0.38	1.61
<i>dlr</i>	194	0.00	0.02	-0.05	0.05
<i>rr3</i>	194	-0.06	0.22	-0.48	0.54
<i>ps</i>	194	0.08	0.19	-0.46	0.56
<i>er</i>	194	0.01	0.12	-0.24	0.63
<i>lir</i>	194	0.04	0.02	0.00	0.09
<i>cpi</i>	194	0.02	0.01	0.00	0.06
<i>ipi</i>	194	0.03	0.04	-0.09	0.14
<i>pscon</i>	194	0.08	0.39	-0.66	1.60
<i>contl</i>	194	0.12	0.40	-0.63	1.88
<i>has</i>	194	0.11	0.49	-0.79	3.33
<i>tlarea</i>	194	0.05	0.22	-0.58	1.09

주 : 위 표는 거시경제변수와 주택시장변수의 기술통계량으로, 표본 수(N), 평균(Mean), 표준편차(Std.), 최솟값(Min), 최댓값(Max)을 정리한 것이다. 점선 위의 변수는 거시경제변수로, 총 유동성(*m3*), 고객예탁금(*dep*), 예대율(*dlr*), 3년 만기 국고채 금리(*rr3*), 종합주가지수(*ps*), 대미달러환율(*er*), 경기선행지수(*lir*), 소비자물가지수(*cpi*), 산업생산지수(*ipi*)이며, 점선 아래의 변수는 주택시장변수로, 상장건설업 주가지수(*pscon*), 대규모 건설주주액(*contl*), 아파트 공급호수(*has*), 전국 주거지역 토지거래면적(*tlarea*)이다.

까지의 데이터를 테스트 데이터로 사용한다. 예측변수인 주택가격 위험신호는 2004년 12월부터 2017년 1월까지의 데이터를 학습 데이터로, 2017년 2월부터 2020년 1월까지의 데이터를 테스트 데이터로 사용한다.

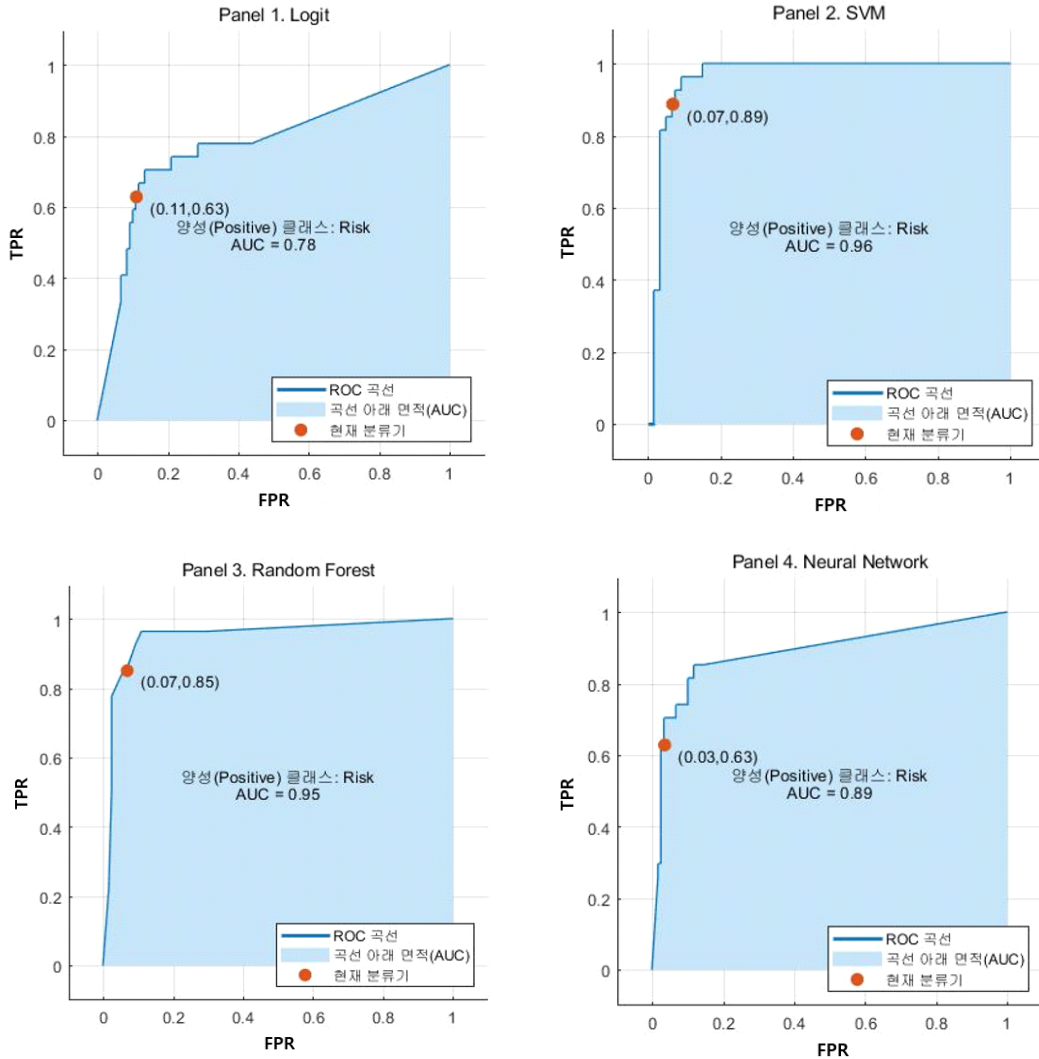
각각의 모델을 학습한 결과는 〈그림 2〉와 〈표 3〉을 통해 확인할 수 있다. 〈그림 2〉는 모형별 ROC 곡선과 AUC 값을 보여준다. ROC 곡선이 좌상단으로 붙을수록 좋은 분류모형에 해당하므로 SVM과 RF가 로지스틱 회귀모형이나 ANN에

비해 주택시장의 위험신호 예측에 더 적합한 것으로 판단된다. 또한, AUC 값도 SVM과 RF가 각각 0.96과 0.95로 큰 값을 갖지만, 신경망 모형과 로지스틱 회귀모형은 0.89와 0.78로 비교적 낮은 값을 갖는다.

이러한 결과는 〈표 3〉의 예측 정확도를 통해서도 확인할 수 있다. 학습 데이터 내에서의 검정 정확도와 학습 데이터로 학습시킨 모형의 테스트 데이터에 대한 예측 정확도를 보여준다.

과적합을 방지하기 위하여 학습 데이터에 대한 정확도는 5겹 교차검증(5-fold cross-validation)을 사용한다. 교차검증에서는 학습 데이터를 다섯 개의 데이터 집단으로 무작위로 나눠 진행한다. 검증 정확도는 RF가 93.8%로 가장 큰 값을 보여주고, SVM과 ANN이 각각 91.8%와 91.1%의 정확도를 보여준다. 로지스틱 회귀모형은 가장 낮은 88.4%의 검정 정확도를 보여준다. 테스트 데이터에 대한 예측 정확도에서는 기계학습 방법론 모두 86.5%의 정확도를 보여준 반면, 로지스틱 회귀모형은 67.6%로 가장 낮은 예측력을 보인다. 이러한 결과를 종합했을 때, 주택시장의 위험신호인  $HBS_t$ 에 대한 예측모형으로는 RF가 가장 적합할 것으로 보인다.

〈표 4〉는 모형별로 2020년의  $HBS_t$ 를 예측한 결과를 보여준다. 로지스틱 모형과 SVM의 경우, 2020년에 위험신호가 나타나지 않는다고 예측하는 반면, RF와 ANN의 예측결과는 다소 차이는 있지만 2020년 말에 위험신호가 나타났다고 예측한다. 앞선 평가에서 RF의 예측력이 우수함을 확인하였으므로, 2020년 말에 주택시장의 위험



〈그림 2〉 모형별 ROC 곡선

신호가 나타났다고 볼 수 있다. 이러한 결과는 2020년의 주택가격지수를 그린 〈그림 3〉을 볼 때 타당하다고 볼 수 있다.

〈그림 3〉에서 2020년 말에 강남 아파트매매가격지수는 하락하는 추세를 보이지만, 다른 세 가격

지수는 상승하는 추세를 보인다. 〈표 4〉에서 예측한 결과와 〈그림 3〉의 가격지수 추이를 고려하면, 주택가격은 2020년 말 고점 가까이 도달했으며, 2021년 이후에는 강남 지역을 중심으로 주택시장의 상승세가 다소 제한될 수도 있음을 의미한다.

〈표 3〉 모형별 검증 및 테스트 정확도

	Logit	SVM	RF	ANN
Validation	88.4%	91.8%	93.8%	91.1%
Test	67.6%	86.5%	86.5%	86.5%

주 : 위 표는 모형별 검증 및 테스트의 정확도(Accuracy)를 정리한 것이다. Logit, SVM, RF, ANN은 각각 로지스틱 회귀모형, 서포트 벡터 머신, 랜덤 포레스트, 인공신경망 모형을 의미한다. 정확도는 학습 데이터 내의 일부를 바탕으로 계산한 검증(Validation) 정확도와 테스트 데이터를 활용한 테스트(Test) 정확도를 각각 계산하였다.

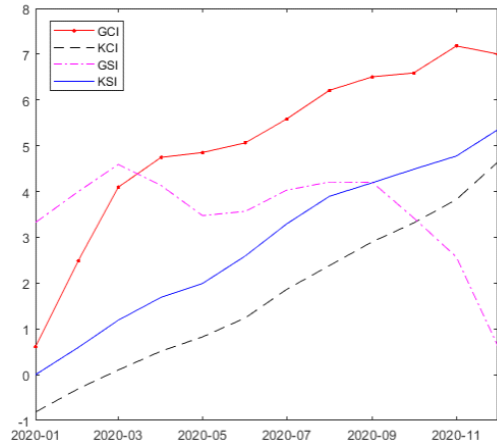
SVM, support vector machine; RF, random forest; ANN, artificial neural network.

〈표 4〉 모형별 주택가격 위험신호 예측

	Logit	SVM	RF	ANN
2020년 2월	'No risk'	'No risk'	'No risk'	'No risk'
2020년 3월	'No risk'	'No risk'	'No risk'	'No risk'
2020년 4월	'No risk'	'No risk'	'No risk'	'No risk'
2020년 5월	'No risk'	'No risk'	'No risk'	'No risk'
2020년 6월	'No risk'	'No risk'	'No risk'	'No risk'
2020년 7월	'No risk'	'No risk'	'No risk'	'No risk'
2020년 8월	'No risk'	'No risk'	'No risk'	'No risk'
2020년 9월	'No risk'	'No risk'	'Risk'	'Risk'
2020년 10월	'No risk'	'No risk'	'No risk'	'No risk'
2020년 11월	'No risk'	'No risk'	'Risk'	'No risk'
2020년 12월	'No risk'	'No risk'	'No risk'	'Risk'
2021년 1월	'No risk'	'No risk'	'No risk'	'No risk'

주 : 위 표는 모형별로 2020년 2월부터 2021년 1월까지의 주택가격 위험신호를 예측한 결과이다. Logit, SVM, RF, ANN은 각각 로지스틱 회귀모형, 서포트 벡터 머신, 랜덤 포레스트, 인공신경망 모형을 의미한다. 'No risk'는 모형에서  $HBS_t = 0$ 의 값이 나왔다는 것으로, 주택시장에 위험이 없다고 예측한 것이다. 'Risk'는 모형에서  $HBS_t = 1$ 의 값이 나왔다는 것으로, 주택시장에 위험이 있다고 예측한 것이다.

SVM, support vector machine; RF, random forest; ANN, artificial neural network.



주 : 위 그림은 2020년 주택가격지수의 추이를 나타낸다. GCI, KCI, GSI, KSI는 각각 강남 아파트 전세가격지수, 전국 아파트 전세가격지수, 강남 아파트 매매가격지수, 전국 아파트 매매가격지수를 의미한다.

〈그림 3〉 2020년 주택가격지수

## V. 결론

주택시장의 급락과 침체는 국가 경제 전반에 걸쳐 큰 영향을 미치며, 금융위기를 일으킬 위험도 있다. 역사적으로 주택시장의 위험에서 야기된 금융위기의 피해가 크다는 점을 고려하면, 주택시장의 위험을 효율적으로 감지하는 것은 위기에 더 효과적인 대응을 가능하게 할 것이다. 이에 따라 본 논문에서는 주택시장의 위험신호를 감지할 수 있는 조기경보체제를 구축하였다. 주택가격의 장기적 추세를 고려하는 기존 연구가 주택가격이 하락하여 실질적으로 침체에 접어드는 기간을 예측하는 데 한계가 있으므로, 미래가격에 대한 현재 가격의 수준을 바탕으로 위험신호를 정의한다. 이러한 위험신호는 경험적 사실과 일치하여 주택시장의 위험에 대한 예측이 가능하다.

본 연구에서는 위험신호를 예측할 모형으로 기계학습 분류기법을 사용한다. 로지스틱 회귀분석, SVM, RF, ANN의 네 모형의 성과를 비교한 결과, RF가 가장 우수한 성능을 보임을 확인하였으며, 기존에 자주 사용되었던 로지스틱 회귀분석은 가장 좋지 못한 성과를 보였다. 또한, 학습시킨 모형을 바탕으로 2020년의 주택시장 위험신호를 예측한 결과, 2021년 이후에는 주택시장의 실제 가격 상승세가 둔화할 가능성이 있다는 점을 확인하였다.

본 연구는 주택시장의 침체를 예측할 수 있는 위험신호를 새롭게 정의했다는 점에서 기존 연구와 차별점을 가진다. 또한, 기존에 충분히 연구되지 않았던 한국 주택시장에 대한 조기경보체계를 기계학습 방법론을 사용하여 설계했다는 점에서 학술논문으로써의 기여와 정책적 함의를 갖는다. 그러나 현 정부에서 급변해 온 부동산 정책이나 급격한 외부 충격으로 인한 주택시장의 구조적 변화를 전부 반영하지는 못했다는 한계를 갖는다. 특히, 2020년 이후 세계적 전염병 대유행으로 인해 주택시장에 구조적 변화가 일어났을 가능성이 있다. 추후 연구에서 외부적 충격과 해외시장의 충격과 같은 요인을 모형에 반영한다면 예측력이 더 높은 조기경보체계를 만들 수 있을 것이다.

## ORCID

박대현 <https://orcid.org/0000-0001-7438-7760>

김정환 <https://orcid.org/0000-0002-4400-9088>

류두진 <https://orcid.org/0000-0002-0059-4887>

## 참고문헌

1. 강희만 · 김정렬, 2012, 「투기적 특성이 아파트가격에 미치는 영향 분석」, 『산업경제연구』, 25(4): 2841-2858.
2. 권순신 · 최성호, 2019, 「주택가격 순환주기의 지역간 관계 분석」, 『부동산분석』, 5(2): 1-15.
3. 김정훈, 2018, 「주요 선진국 부동산 버블과 금융위기간의 관계: 글로벌 패널자료 분석을 중심으로」, 『국제금융연구』, 8(2): 5-32.
4. 김승욱 · 이창석, 2009, 「부동산의 투자와 투기에 관한 이론적 고찰」, 『부동산학보』, 36: 213-226.
5. 김중규 · 정동준, 2012, 「유동성과 금리가 부동산 가격 변동에 미치는 영향 분석」, 『주택연구』, 20(1): 105-125.
6. 김형준 · 류두진 · 조훈, 2019, 「기업부도예측과 기계학습」, 『금융공학연구』, 18(3): 131-152.
7. 류지수 · 임규채 · 기석도(2007), 「주택가격 및 토지 가격의 거품존재 가능성 검토」, 『산업경제연구』, 20(6): 2245-2264.
8. 민선옥 · 이영수, 2019, 「부호제약을 이용한 한국 주택시장의 변동요인 분석」, 『부동산분석』, 5(1): 1-13.
9. 박원암, 2014, 「글로벌 금융위기와 우리나라 주택시장: 특성과 거품검정」, 『국제경제연구』, 20(4): 71-95.
10. 서승환, 2004, 「주택시장의 조기경보체계에 관한 연구」, 『지역연구』, 20(3): 17-36.
11. 손정식 · 김관영 · 김용순, 2003, 「부동산가격 예측 모형에 관한 연구」, 『주택연구』, 11(1): 49-76.
12. 송인호, 2014, 「주택담보인정비율(LTV) 규제가 거시경제에 미치는 영향」, 『경제전망』, 31(1): 4-17.
13. 안현철 · 김정재 · 한인구, 2006, 「다분류 support



- vector machine을 이용한 한국 기업의 지능형 기업채권평가모형, 『경영학연구』, 35(5): 1479-1496.
14. 이강용 · 이종아 · 정준호, 2015, 「주택시장과 주식 시장의 동적 네트워크 구조 비교: 시가총액 상위 자산을 중심으로」, 『부동산학보』, 61: 195-207.
15. 이기영, 2013, 「신호접근법에 의한 중국 주택시장 조기경보시스템 연구」, 『한중사회과학연구』, 26: 173-195.
16. 이기영, 2014, 「EWS 기법을 사용한 중국주택시장 위기에 대한 영향요인 분석: 정책영향을 위주로」, 『중국지역연구』, 1(1): 77-109.
17. 이기영, 2015, 「위기발생원인 차이에 따른 중국 주택시장 조기경보체계 반응변화 연구」, 『중국학연구』, 73: 363-394.
18. 이석희 · 김수현, 2014, 「한국 주택체제의 성격과 변화: 동아시아 발전주의 국가의 특성을 중심으로」, 『공간과 사회』, 24(2): 5-37.
19. 이용재 · 박철범, 2019, 「한국 주택시장의 거품 검정과 주택시장 정책」, 『한국경제학보』, 26(2): 277-299.
20. 이재울, 2004, 「투기와 거품 그리고 부동산가격의 변동: 한국과 일본의 사례를 중심으로」, 『경영경제』, 37(1): 315-334.
21. 이태리 · 송인호, 2015, 「통화정책의 주택시장 파급경로 연구」, 『주택연구』, 23(3): 31-62.
22. 임병진 · 한성윤, 2009, 「주식시장 지수와 부동산 시장 지수의 시계열 특성비교와 관계에 관한 실증적 연구」, 『산업경제연구』, 22(4): 2065-2083.
23. 전해정, 2014, 「글로벌 금융위기 전후로 거시경제 변수와 부동산시장 간의 관계에 대한 연구: 동적 패널분석을 이용하여」, 『부동산학보』, 58: 33-44.
24. 정건섭 · 김성우 · 이상엽, 2011, 「그랜저인과분석을 통한 매매와 전세시장의 주택가격 결정구조 분석: 부산지역 주택시장의 수급요인을 중심으로」, 『정책 분석평가학회보』, 21(2): 179-198.
25. 허종만 · 이영수, 2018, 「이자율변동의 주택시장 파급효과 분석」, 『부동산분석』, 4(1): 55-69.
26. 홍정효 · 문규현(2009), 「국내 부동산시장과 주요 거시경제지표들간의 선-후행성 연구」, 『금융공학연구』, 8(2): 97-125.
27. Aldasoro, I., C. Borio, and M. Drehmann, 2018, "Early warning indicators of banking crises: Expanding the family," *BIS Quarterly Review*, 29-45.
28. Alessi, L. and C. Detken, 2018, "Identifying excessive credit growth and leverage," *Journal of Financial Stability*, 35: 215-225.
29. Al-Yahyaee, K. H., W. Mensi, H. U. Ko, M. Caporin, and S. H. Kang, 2020, "Is the Korean housing market following Gangnam style?" *Empirical Economics*, 1-32.
30. Anundsen, A. K., K. Gerdrup, F. Hansen, and K. Kragh-Sorensen, 2016, "Bubbles and crises: The role of house prices and credit," *Journal of Applied Econometrics*, 31(7): 1291-1311.
31. Beckmann, D., L. Menkhoff, and K. Sawischlewski, 2006, "Robust lessons about practical early warning systems," *Journal of Policy Modeling*, 28(2): 163-193.
32. Breiman, L., 2001, "Random forests," *Machine Learning*, 45(1): 5-32.
33. Brander, L. M. and M. J. Koetse, 2011, "The value of urban open space: Meta-analyses of contingent valuation and hedonic pricing results," *Journal of Environmental Management*, 92(10): 2763-2773.
34. Chiu, R. L. H., 2006, "Globalization and localization: Economic performance and the housing markets of the Asian tigers since the financial crisis," *Housing Finance International*, 20(3): 12-17.
35. Chiu, R. L. H., 2008, "Government intervention in housing: Convergence and divergence of the



- Asian Dragons,” *Urban Policy and Research*, 26(3): 249–269.
36. Chuliá, H., M. Guillén, and J. M. Uribe, 2017, “Measuring uncertainty in the stock market,” *International Review of Economics & Finance*, 48: 18–33.
37. Corbet, S., B. Lucey, A. Urquhart, and L. Yarovaya, 2019, “Cryptocurrencies as a financial asset: A systematic analysis,” *International Review of Financial Analysis*, 62: 182–199.
38. Diba, B. T. and H. I. Grossman, 1988, “Explosive rational bubbles in stock prices?” *American Economic Review*, 78(3): 520–530.
39. Evans, G. W., 1991, “Pitfalls in testing for explosive bubbles in asset prices,” *American Economic Review*, 81(4): 922–930.
40. Frankel, J. A. and A. K. Rose, 1996, “Currency crashes in emerging markets: An empirical treatment,” *Journal of International Economics*, 41(3–4): 351–366.
41. Geuder, J., H. Kinatader, and N. F. Wagner, 2019, “Cryptocurrencies as financial bubbles: The case of bitcoin,” *Finance Research Letters*, 31: 179–184.
42. Giglio, S., M. Maggiori, and J. Stroebel, 2016, “No-bubble condition: Model-free tests in housing markets,” *Econometrica*, 84(3): 1047–1091.
43. Hagemann, D. and M. Wohlmann, 2019, “An early warning system to identify house price bubbles,” *Journal of European Real Estate Research*, 12(3): 291–310.
44. Hamilton, J. D., 2018, “Why you should never use the Hodrick–Prescott filter,” *Review of Economics and Statistics*, 100(5): 831–843.
45. Harvey, D. I., S. J. Leybourne, R. Sollis, and A. M. R. Taylor, 2016, “Tests for explosive financial bubbles in the presence of non-stationary volatility,” *Journal of Empirical Finance*, 38(B): 548–574.
46. Hsu, C. W., C. C. Chang, and C. J. Lin, 2016, “A practical guide to support vector classification,” National Taiwan University.
47. Hu, Y. and J. Pang, 2008, “Financial crisis early-warning based on support vector machine,” in *2008 IEEE International Joint Conference on Neural Networks*, Hong Kong, 2435–2440.
48. Hyun, D. and S. Milcheva, 2018, “Spatial dependence in apartment transaction prices during boom and bust,” *Regional Science and Urban Economics*, 68: 36–45.
49. Jordà, Ò., M. Schularick, and A. M. Taylor, 2015, “Leveraged bubbles,” *Journal of Monetary Economics*, 76: S1–S20.
50. Kaminsky, G., S. Lizondo, and C. M. Reinhart, 1998, “Leading indicators of currency crises,” *IMF Staff Papers*, 45(1): 1–48.
51. Kim, H., H. Cho, and D. Ryu, 2018, “Characteristics of mortgage terminations: An analysis of a loan-level dataset,” *Journal of Real Estate Finance and Economics*, 57(4): 647–676.
52. Kim, H., H. Cho, and D. Ryu, 2020, “Corporate default predictions using machine learning: Literature review,” *Sustainability*, 12(16): 6325.
53. Kim, H., H. Cho, and D. Ryu, 2021a, “Forecasting consumer credit recovery failure: Classification approaches,” *Journal of Credit Risk*, Forthcoming.
54. Kim, H., H. Cho, and D. Ryu, 2021b, “Corporate bankruptcy prediction using machine learning methodologies with a focus on sequential data,” *Computational Economics*, Forthcoming.

55. Kim, H. M., 2017, "Ethnic connections, foreign housing investment and locality: A case study of Seoul," *International Journal of Housing Policy*, 17(1): 120–144.
56. Knoll, K., M. Schularick, and T. Steger, 2017, "No price like home: Global house prices, 1870–2012," *American Economic Review*, 107(2): 331–353.
57. Lee, H. S. and W. S. Lee, 2019, "Cross-regional connectedness in the Korean housing market," *Journal of Housing Economics*, 46: 101654.
58. Park, D. and D. Ryu, 2021, "A machine learning-based early warning system for the housing and stock markets," *IEEE Access*, Forthcoming.
59. Pavlidis, E., A. Yusupova, I. Paya, D. Peel, E. Martínez-García, A. Mack, and V. Grossman, 2016, "Episodes of exuberance in housing markets: In search of the smoking gun," *Journal of Real Estate Finance and Economics*, 53(4): 419–449.
60. Phillips, P. C. B., S. Shi, and J. Yu, 2015, "Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500," *International Economic Review*, 56(4): 1043–1078.
61. Phillips, P. C. B., Y. Wu, and J. Yu, 2011, "Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?" *International Economic Review*, 52(1): 201–226.
62. Samitas, A., E. Kampouris, and D. Kenourgios, 2020, "Machine learning as an early warning system to predict financial crisis," *International Review of Financial Analysis*, 71: 101507.
63. Seok, S. I., H. Cho, and D. Ryu, 2020, "The information content of funds from operations and net income in real estate investment trusts," *North American Journal of Economics and Finance*, 51: 101063.
64. Shiller, R. J., 2000, "Measuring bubble expectations and investor confidence," *Journal of Psychology and Financial Markets*, 1(1): 49–60.
65. Suh, S. H., K. Kim, and J. S. Jeon, 2011, "Housing market early warning system: The case of Korea," *European Journal of Scientific Research*, 56(4): 539–547.
66. Vapnik, V. N., 2000, *The Nature of Statistical Learning Theory*, 2<sup>nd</sup> edition, Springer.
67. Wang, X. J., G. T. Zeng, K. X. Zhang, H. B. Chu, and Z. S. Chen, 2020, "Urban real estate market early warning based on support vector machine: A case study of Beijing," *International Journal of Computational Intelligence Systems*, 13(1): 153–166.
68. Wang, P., L. Zong, and Y. Ma, 2020, "An integrated early warning system for stock market turbulence," *Expert Systems with Applications*, 153: 113463.
69. Yu, J. and D. Ryu, 2020, "Effects of commodity exchange-traded note introductions: Adjustment for seasonality," *Borsa Istanbul Review*, 20(3): 244–256.
70. Zhou, W., M. Chen, Z. Yang, and X. Song, 2021, "Real estate risk measurement and early warning based on PSO-SVM," *Socio-Economic Planning Sciences*, 101001, Forthcoming.

---

논문접수일: 2021년 2월 16일

심사(수정)일: 2021년 4월 22일

게재확정일: 2021년 4월 23일

## 국문초록

본 연구는 주택시장 위험을 진단하기 위하여, 다양한 외생변수를 반영하는 기계학습에 기반한 조기경보체계를 제안한다. 신호접근법을 바탕으로 주택시장의 위험신호를 정의하고, 기계학습 분류모형을 사용하여 새로운 형태의 주택시장 조기경보체계를 구축한다. 주택시장이 침체에 접어드는 시점을 위험신호로 정의하기 위하여 미래가격의 평균과 표준편차의 합보다 가격이 높은 시점을 위험신호가 있는 시점으로 정의한다. 이러한 위험신호는 국내 주택시장의 경험적 사실과 일치한다. 그리고 위험신호를 예측하는 기계학습의 분류모형을 상호 비교하여 가장 우수한 성능을 보이는 모형을 선정한다. 주택가격지수와 거시경제변수, 주택시장변수를 입력변수로 사용하여 로지스틱 회귀모형과 서포트 벡터 머신, 랜덤 포레스트, 인공신경망 모형을 상호 비교한 결과, 랜덤 포레스트가 가장 높은 예측성적을 보인다는 것을 확인할 수 있다. 본 논문에서 제시한 조기경보체계는 미래 주택시장의 가격변동을 유의미하게 감지할 수 있다는 점에서 정책적 함의를 갖는다.

주제어 : 기계학습, 분류, 신호접근법, 조기경보체계, 주택시장