

한국인터넷방송통신학회 논문지 제20권 제6호

ISSN : 2289-0238(Print) 2289-0246(Online)

머신러닝 기법을 통한 대한민국 부동산 가격 변동 예측

남상현, 한태호, 김이주, 이은지

To cite this article : 남상현, 한태호, 김이주, 이은지 (2020) 머신러닝 기법을 통한 대한민국 부동산 가격 변동 예측, 한국인터넷 방송통신학회 논문지, 20:6, 15-20

① earticle에서 제공하는 모든 저작물의 저작권은 원저작자에게 있으며, 학술교육원은 각 저작물의 내용을 보증하거나 책임을 지지 않습니다.

② earticle에서 제공하는 콘텐츠를 무단 복제, 전송, 배포, 기타 저작권법에 위반되는 방법으로 이용할 경우, 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

www.earticle.net

<https://doi.org/10.7236/JIIBC.2020.20.6.15>

JIIBC 2020-6-3

머신러닝 기법을 통한 대한민국 부동산 가격 변동 예측

Real-Estate Price Prediction in South Korea via Machine Learning Modeling

남상현*, 한태호*, 김이주*, 이은지**

Sanghyun Nam*, Taeho Han*, Leeju Kim*, Eunji Lee**

요약 최근 부동산 시장에 대한 관심이 높다. 과거 주거환경으로만 여겨지던 부동산은 끊임없는 수요 증가로 안정적인 투자 대상으로 인식되고 있기 때문이다. 특히 국내 시장의 경우 인구 수의 감소에도 불구하고 1인 가구의 증가 및 대도시로의 인구 유입이 가속화되며 수도권 중심으로 부동산 가격이 급격히 상승하고 현상이 나타나고 있다. 이에 미래 부동산 시장의 전망을 정확히 예측하는 것은 개인의 자산 관리 뿐 아니라 정부 정책 수립 등 사회 전반에 걸쳐 매우 중요한 사안이라고 할 수 있다. 본 논문에서는 머신러닝 기법을 활용해 과거 부동산 매매 데이터를 학습해 미래 부동산 시세를 예측하는 프로그램을 개발하였다. 한국감정원과 국토교통부에서 제공하는 대한민국 부동산 매매 시세 데이터를 활용하였으며 지역별로 2022년도 평균 매매가 예측치를 제시한다. 개발된 프로그램은 오픈소스 형태로 공개하여 다양한 형태로 활용될 수 있도록 하였다.

Abstract Recently, the real estate is of high interest. This is because real estate, which was considered only a residential environment in the past, is recognized as a stable investment target due to the ever-growing demand on it. In particular, in the case of the domestic market, despite the decrease in the number of people, the number of single-person households and the influx of people to large cities are accelerating, and real estate prices are rising sharply around the metropolitan area. Therefore, accurately predicting the prospects of the future real estate market becomes a very important issue not only for individual asset management but also for government policy establishment. In this paper, we developed a program to predict future real estate market prices by learning past real estate sales data using machine learning techniques. The data on the market price of real estate provided by the Korea Appraisal Board and the Ministry of Land, Infrastructure and Transport were used, and the average sales price forecast for 2022 by region is presented. The developed program is publicly available so that it could be used in various forms.

Key Words : Machine learning, Real-estate price prediction, Linear regression

*비회원, 송실대학교 스마트시스템소프트웨어학과
**정회원, 송실대학교 스마트시스템소프트웨어학과 (교신저자)
접수일자 2020년 9월 7일, 수정완료 2020년 11월 7일
게재확정일자 2020년 12월 4일

Received: 7 September, 2020 / Revised: 7 November, 2020 /
Accepted: 4 December, 2020
Corresponding Author: ejlee@ssu.ac.kr
Dept. of Smart Systems Software, Soongsil University, Korea

I. 서 론

최근 고도화된 IT기술은 사회 전 분야에 영향을 끼치며 4차 산업혁명 시대를 전인하고 있다. 5G 고속 네트워크 통신 기술은 IoT, 자율주행기술 등의 급성장을 도모하며 인류의 생활 방식을 급속히 변화시키고 있다^{[1][2]}. 고성능 CPU/GPU/Accelerator 등의 향상된 컴퓨팅 능력은 대용량의 데이터를 빠르게 분석하여 새로운 가치를 발견하도록 해주고 있다^[3]. 유전자 분석을 통한 표적치료제 개발 및 유전병 조기 검진 등은 대표적인 융합기술의 사례이며, 이러한 기술융합은 의료뿐 아니라 경제·정치·사회·과학 등 전 분야에 걸쳐 활발히 진행되고 있다.

부동산 시장은 빅데이터 활용 가능성이 매우 높은 분야 중 하나로 논의되고 있다. 주거의 공간으로만 인식되었던 과거와는 달리, 최근에는 지속적인 수요에 따른 가치상승으로 부동산은 투자 가치가 높은 안전 자산으로 인식되고 있다. 이에 전 세계의 주요 도시 부동산을 대상으로 다양한 투자가 이루어지고 있으며 미래 전망에 대한 관심은 더욱 높아지는 추세이다. 이러한 트렌드는 국내에서도 예외가 아니다. 국내에서는 최근 수도권을 중심으로 부동산 가격이 큰 폭으로 상승하며 사회적 관심을 고조시켰다. 부동산 가격의 변동은 주거안정을 요구하는 개인부터 부동산을 자산으로 보유하고 있는 기업에 이르기까지 다양한 계층에 큰 영향을 끼친다. 본 연구에서는 이러한 사회적 파급력이 높은 부동산 시장의 가치를 머신러닝 기법을 활용해 예측하는 응용을 개발함으로써 개인의 자산설계 및 사회적 정책 수립에 정보로 활용될 수 있도록 하고자 한다.

본 논문에서는 선형회귀(Linear Regression) 방식을 활용하여 미래 부동산 가격을 전망하는 프로그램을 제안한다. 선형회귀 알고리즘은 데이터를 선형 상관관계로 모델링하여 미래의 가치를 예측하는 방식이다. 과거 매매 데이터를 기반으로 부동산 가치 변화에 대한 흐름을 선형함수 형태로 표현하고 이를 통해 미래의 부동산 가치를 예측하는 것이다. 제안된 알고리즘은 파이썬에서 제공하는 머신러닝 패키지인 사이킷런(Scikit Learn)을 통해 구현되었다^[4].

본 논문의 구성은 다음과 같다. 2장에서는 데이터의 수집 및 가공 방법과 해당 데이터를 활용해 미래 부동산 가격을 예측하는 선형 학습 모델에 대해 설명한다. 3장에서는 본 논문에서 개발된 부동산 시장 전망 예측 프로그램의 주요 기능을 설명한다. 5장에서는 제안한 머신러닝 모델로부터 획득한 부동산 전망 결과에 대해 논의하고 6

장에서는 결론을 맺는다.

II. 부동산 매매가격 예측 모델

표 1. 지역별 부동산 매매(단위: 원, 2012.1 ~ 2019.9)

Table 1. Real-estate Price

지역 \ 기간	2012.01	2012.02	2019.10
서울-강북지역	421774.5	421461.5	419423.8
서울-강남지역	588654	586395	577620.7
경기	264725.5	264532.3	263234.9
인천	193693	193374	192741
부산	195713.3	195181.5	194277.5
대구	163724	164236	165526
광주	132796	132770	133666
대전	197017	196065	195304
울산	206601	207684	209701
세종	150162	150162	150162
강원	115390	115695	116122
충청도	128436	128575.5	129378
경상도	137842.5	137850	138171
전라도	96539	96681.5	97266.5

본 논문에서는 한국감정원과 국토교통부에서 제공하는 지역별 부동산 매매가격을 사용하여 부동산 매매가의 추이를 분석하였다.^[1] 표 1은 수집된 데이터의 일부를 보여주고 있는데, 2012.1월부터 2019.10월까지 발생한 거래에 대해 월별 평균 매매가를 지역별로 나타내고 있다. 본 연구에서는 분석 결과의 가독성을 높이기 위해 해당 데이터를 기반으로 년도별 평균 매매 거래값을 추출하고 이를 머신러닝 모델의 입력값으로 활용하였다.

수집된 데이터를 사용하여 본 저자들은 머신러닝의 기본 모델인 선형 회귀분석 (Linear Regression) 방식을 사용하여 부동산 가격을 예측하였다. 선형회귀 알고리즘에서는 데이터 분포의 경향성을 예측하기 위해 선형회귀식(Formula for Linear Regression Line)과 비용함수(Cost Function)이 정의된다. 아래의 수식은 비용함수를 선형회귀식을 이용해 나타내고 있다.

$$cost = \frac{1}{m} \sum_{i=1}^m (H(x^{(i)}) - y^{(i)}) \quad (1)$$

$H(x)$ 는 선형회귀식을 나타내며 $H(x) = Wx + b$ 의 1차함수 형태로 정의된다. 비용은 m 개의 서로 다른 데이터가 있을 때, 임의의 데이터 x 에 대해 $H(x)$ 와 실제 x 에 대응되는 y 값의 차이의 평균값으로 정의한다. 제안하는

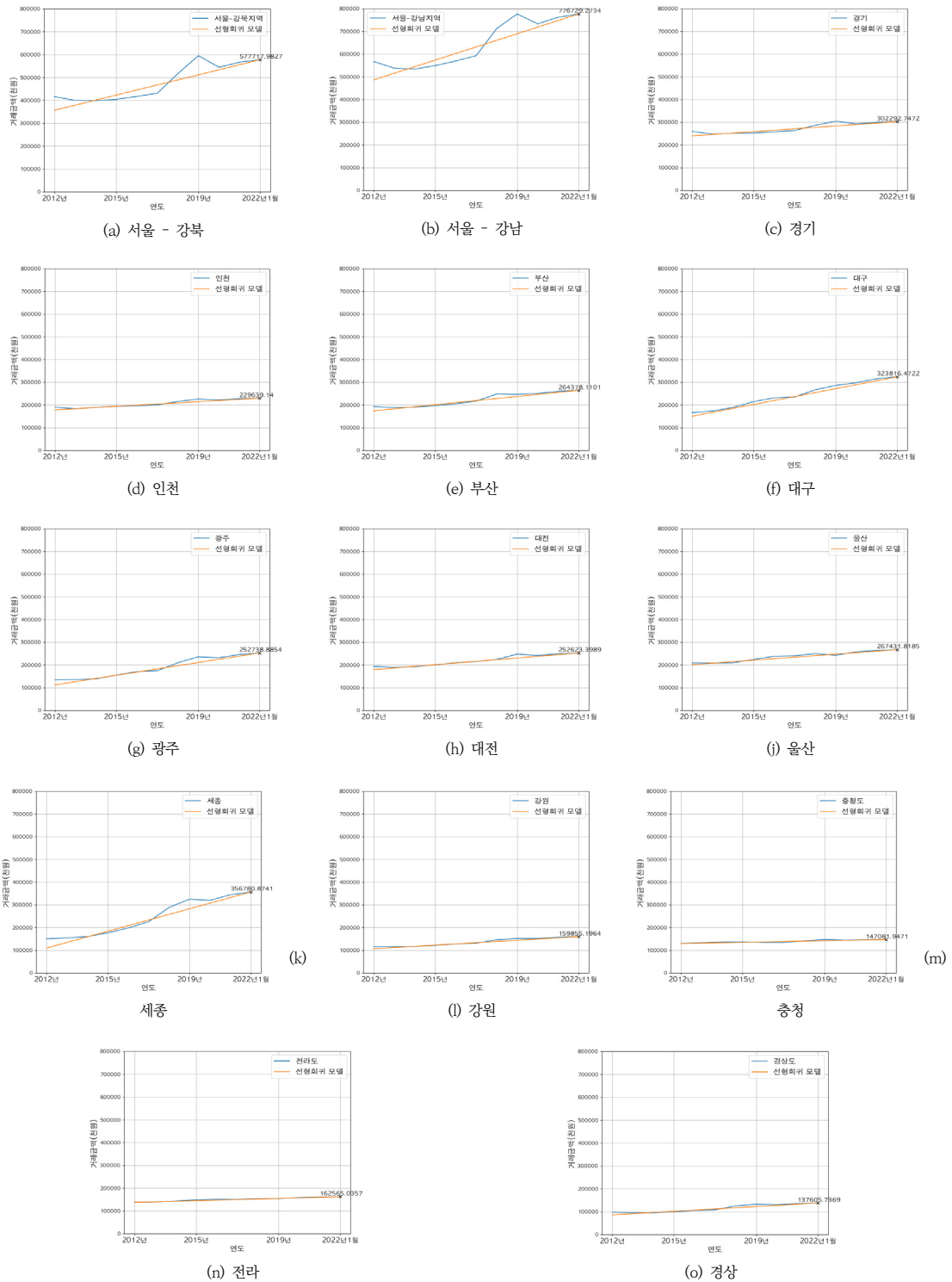


그림 1. 지역별 실제 매매가격 및 예측가격 비교

프로그램은 머신러닝을 통해 이러한 비용을 최소화하는 기울기 W 와 y 절편 b 의 값을 회귀분석을 통해 찾아내는 작업을 수행한다. 본 학습에서는 독립변수 x 에는 연도를 입력으로 받고 종속변수 y 에는 해당 연도의 평균 매매가가 입력되었다.

III. 실험결과

그림 1은 지역별 부동산 매매의 실제 가격을 바탕으로 선형 회귀분석 모델을 통한 머신러닝으로 2022년도 부동산 평균 매매가격 예측 값을 보여주고 있다. 푸른색으로 표시된 선이 실제 거래가이고 주황색으로 표시된 선이 선형 회귀분석으로 도출된 매매가격 추세선이다. 본 연구진이 개발한 예측 프로그램에서는 원하는 시점을 입력받아 해당 시점의 매매 평균가를 예측할 수 있도록 해주는데 본 논문에서는 2022년도 1월의 평균 매매가 예측치를 제시하였다.

그림 1에서 보여주는 바와 같이 서울, 세종, 부산을 제외한 대부분의 지역에서는 지난 10여 년간 부동산 매매 가격이 선형적으로 증가해온 것을 확인할 수 있다. 경기, 인천 대전, 울산, 강원, 충청, 전라, 경상 지역에서는 매매가 변동이 비교적 완만한 기율기로 증가하였으며 2012년도 대비 2022년도의 부동산 가격이 15% 미만으로 나타났다. 반면 세종시에서는 지난 10년간 평균 매매가가 2배 이상 급증하며 2022년도 1월의 평균 매매가는 3억6천만원 정도에 이를 것으로 예측되고 있다.

한편, 서울은 강남과 강북지역으로 구분하여 분석하였는데 두 지역 모두 부동산 가격이 크게 상승한 것으로 나타났다. 강북지역의 경우 2012년도 평균 매매가가 4억 원이었던 반면 지속적으로 매매가가 상승하여 2022년에는 평균 매매가가 6억원 정도에 이를 것으로 나타났다. 강남지역의 경우 지난 10년간 평균 매매가가 최저일 때는 5억 중반이었으나 역시 큰폭으로 상승하여 2022년에는 평균 매매가가 7억 8천만원 정도에 이를 것으로 예측되었다.

그림 2는 머신러닝을 통한 해당 시점의 매매 예측가를 전국 지도를 통해 보여주고 있다. 해당 정보는 정보의 시각화를 통해 사용자의 편의성을 향상시키고자 프로그램에서 추가적으로 제공한다. 각 지역들은 짙은 검은색 선으로 구분이 되며 각 지역의 매매가는 흰색부터 푸른색까지의 색의 짙기로 표시하였다. 이는 그 지역의 매매 예측가가 높을수록 더 진한 푸른색으로 표시가 된다. 그림

2에서 볼 수 있듯 서울, 세종, 대구가 2022년도의 부동산 가격 상승폭이 높은 곳으로 예측된다.

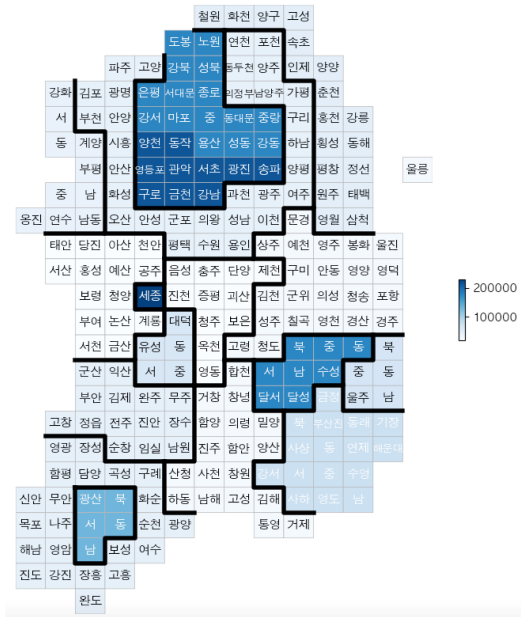


그림 2. 예측매매가를 바탕으로 한 전국 지도

IV. 논 의

본 연구에서는 선형회귀 분석 모델을 사용하여 대한민국의 부동산 시장을 전망하는 연구를 수행하였다. 공개되어 있는 공공데이터를 기반으로 학습하여 모델을 도출하였기 때문에 데이터에 대한 신뢰성은 높다고 할 수 있다. 그러나 본 연구는 초기 단계의 연구로 아래와 같은 한계점이 있다. 먼저 일차함수 형태로 데이터를 모델링하는 선형회귀 모델에서는 가격이 양방향으로 변하거나 특정 시점에 급격히 변하는 데이터에 대한 모델링은 정확도가 낮아질 수 있다. 본 연구에서는 서울지역과 세종, 광주 등지에서 실제 데이터와 매매가 예측 그래프 간의 격차가 나타나는 것을 관찰할 수 있다. 향후 다양한 모델을 부동산 전망 예측에 적용해 봄으로써 변동성이 증가되고 있는 부동산 시장을 보다 정확하게 예측할 수 있는 머신러닝 기법을 찾고자 한다.

두 번째로는 현재 결과는 다양한 주거 형태를 포함한 총체적인 매매가 평균에 대한 예측이라면 향후에는 아파트, 단독주택, 다세대주택 등 형태에 따른 가격 변동성을 분석해보고자 한다. 또한 매매가 뿐 아니라 전월세에 대

한 시세도 분석하여 구체적이고 실효성 높은 부동산 전망 데이터로 활용될 수 있도록 하고자 한다.

또한 부동산 시장은 경제, 정치, 사회 등 매우 다양한 요소와 밀접한 관련이 있다. 이에 향후에는 단순히 과거의 부동산 거래 내역만으로 시장을 전망하기 사회의 다양한 요소들과 부동산 가격 간의 관계를 모델링하고 이를 바탕으로 향후 부동산 가격의 방향성을 심도 있게 예측할 수 있는 딥러닝 기술에 대해서도 연구하고자 한다.

IV. 관련연구

고도의 반도체 집적도 기술은 컴퓨터 시스템의 계산능력을 폭발적으로 향상시켜 머신러닝과 같이 많은 양의 계산을 기반으로 데이터를 학습하여 미래를 예측하는 기법들을 대중화시켰다. 이에 다양한 분야에서 머신러닝을 활용하기 위한 연구가 활발히 진행되고 있다. 최근 연구에서는 인공지능 기법을 활용해 선박에 대한 신수요를 예측함으로써 현재 대내외 정세 변화로 위기에 처해있는 조선산업 분야의 신성장동력을 창출하고자 하였다^[5]. 또한 머신러닝 기법을 양식업에 활용해 수요가 지속적으로 증가하고 있는 수산 시장에 효율적으로 대응하고자 하는 연구도 수행된 바 있다^[6]. 양식업의 경우 양식장의 위치가 실제 생산성에 큰 영향을 끼치므로 자연재해 등을 고려하여 최대의 생산성을 낼 수 있는 양식장 위치를 k-means 알고리즘을 활용해 탐색하는 기법을 제안하였다. 한편, 최근 연구에서는 급성장하고 있는 간편 결제 서비스 이용 의사결정에 영향을 끼치는 요인을 머신러닝 기법을 사용해 분석한 바 있다^[7].

데이터 분석을 통해 부동산 가치 추정 및 평가를 위한 연구도 진행된 바 있다. 머신러닝 기법을 사용해 강남구를 중심으로 상업용 건물의 용도와 매매가 간의 상관관계를 분석한 연구가 진행되었으며^[8], 최근 연구에서는 온라인 데이터를 분석해 부동산 관련 기사의 감성지수와 아파트 매매가격 간의 상관관계를 분석하였다^[9]. 본 연구에서는 특정 용도나 건물에 국한되지 않고 전국 부동산을 대상으로 평균 부동산 매매 시세의 트렌드를 분석하고 가격을 예측했다는 측면에서 기존 연구와 차별적이다.

V. 결 론

본 논문에서는 최근 관심이 고조되고 있는 대한민국

부동산 시장 전망에 대해 머신러닝 기법을 사용해 예측해보는 프로그램을 개발하였다. 지난 10년간 발생한 매매 데이터를 활용해 선형회귀 분석 방식을 통해 부동산 매매 가격 변동을 모델링 하고 이를 활용해 미래 부동산 가격을 예측하였다. 또한, 모델링 및 예측 결과를 분석함으로써 지역별 가격 변동 폭 향후 전망에 대해서도 분석하였다. 향후에는 해당 연구를 바탕으로 부동산 가격 예측 프로그램을 고도화시켜 정확하고 구체적인 시장 전망 데이터를 생성할 수 있도록 하고자 한다.

References

- [1] T. Pardi, "Fourth industrial revolution concepts in the automotive sector: performativity, work and employment", *Journal of Industrial and Business Economics*, Vol. 46, pp.379-389, 2019.
DOI: <https://doi.org/10.7298/x4057cvd>
- [2] J. Moavenzadeh, "The 4th Industrial Revolution: Reshaping the Future of Production", Presented at DHL Global Engineering & Manufacturing Summit, Amsterdam, The Netherlands, 6-8 October 2015.
- [3] Y. You, A. Buluc, J. Demmel, "Scaling deep learning on GPU and knights landing clusters", In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, 2017.
DOI: <https://doi.org/10.1145/3126908.3126912>
- [4] <https://scikit-learn.org/stable/>
- [5] S. Lee, I. Jung, "Development of a Platform Using Big Data-Based Artificial Intelligence to Predict New Demand of Shipbuilding", *The Journal of The Institute of Internet, Broadcasting and Communication*, Vol. 19, No. 1, pp.171-178, 2019.
DOI: <https://doi.org/10.7236/IIBC.2019.19.1.171>
- [6] M. Kang, Y.-G. Jung, D.-H. Jang, "A Study on the Search of Optimal Aquaculture farm condition based on Machine Learning", *The Journal of The Institute of Internet, Broadcasting and Communication*, Vol. 17, No. 2, pp.135-140, 2017.
DOI: <https://doi.org/10.7236/IIBC.2017.17.2.135>
- [7] H. Park, "Factors of Simple Payment Service Use: Application of Machine Learning Prediction Model", *Journal of Digital Contents Society*, Vol. 21, No. 5, pp.921-929, 2020.
DOI: <https://doi.org/10.9728/dcs.2020.21.5.921>
- [8] S.-U. Kim, S.-Y. Ji, H.-J. Jun, "Estimating and Valuing Commercial Building Sales Prices - Focused on the Gangnam-gu -", In *Proceedings of the Autumn Annual Conference of Architectural Institute of Korea*, 2017, Vol. 39, No. 2, pp. 100-101. 2019.

- [9] J. Park, J. Lee, "An Investigation into the Causal Relationship and the Cross Correlation between Apartment House Sales Prices and Real Estate Online News", Journal of Korea Planning Association, Vol. 54, No. 1, pp. 131-147, 2019.
DOI: <https://doi.org/10.17208/jkpa.2019.02.54.1.131>

저 자 소 개

남 상 현(비회원)



- 2020년 숭실대학교 스마트 시스템 소프트웨어학과 학사과정
- 주관심분야 : 키밸류 스토어, 분산 시스템, 스토리지 시스템

한 태 호(비회원)



- 2020년 숭실대학교 스마트 시스템 소프트웨어학과 학사과정
- 주관심분야 : 데이터분석, 자율주행 시스템

김 이 주(비회원)



- 2020년 숭실대학교 스마트 시스템 소프트웨어학과 학사과정
- 주관심분야 : 키밸류 스토어, 분산 시스템, 스토리지 시스템

이 은 지(정회원)



- 2005년 이화여자대학교 컴퓨터학과 학사
- 2012년 서울대학교 컴퓨터공학부 박사
- 2014~2018년 충북대학교 소프트웨어학과 전임교수
- 2019년 숭실대학교 스마트시스템 소프트웨어학과 조교수

- 주관심분야 : 데이터 시스템, 분산 시스템, 클라우드 컴퓨팅

※ 이 연구는 미래창조과학부의 재원으로 한국연구재단(No.NRF-2019R1A2C1090337)의 지원을 받아 수행된 연구임.