

머신러닝 기반 공동주택 분양가 예측모델 개발 기초연구

A Basic Study on Sale Price Prediction Model of Apartment Building Projects using Machine Learning Technique

손 승 현* 김 지 명** 한 범 진*** 나 영 주**** 김 태 희*****
Son, Seung-Hyun Kim, Ji-Myong Han, Bum-Jin Na, Young-Ju Kim, Tae-Hee

Abstract

The sale price of apartment buildings is a key factor in the success or failure of apartment projects, and the factors that affect the sale price of apartments vary widely, including location, environmental factors, and economic conditions. Existing methods of predicting the sale price do not reflect the nonlinear characteristics of apartment prices, which are determined by the complex impact factors of reality, because statistical analysis is conducted under the assumption of a linear model. To improve these problems, a new analysis technique is needed to predict apartment sales prices by complex nonlinear influencing factors. Using machine learning techniques that have recently attracted attention in the field of engineering, it is possible to predict the sale price reflecting the complexity of various factors. Therefore, this study aims to conduct a basic study for the development of a machine learning-based prediction model for apartment sale prices.

키 워 드 : 아파트 프로젝트, 분양가, 예측모델, 머신러닝 기법

Keywords : apartment building project, sale price, prediction model, machine learning technique

1. 서 론

지금까지 발표된 아파트 분양가 예측 연구들은 대부분 회귀 및 시계열 분석과 같은 통계적 방식을 활용한다. 그러나 이러한 기존의 분양가 예측방법은 선형모형의 가정 하에 통계적 분석이 이루어지기 때문에 현실의 복합적 영향 요인에 의해 결정되는 아파트 분양가의 비선형적인 특성을 반영하지 못한다.¹⁾ 이러한 문제점을 개선하기 위해서는 복합적 비선형 영향 요인들에 의한 아파트 분양가 예측을 위한 새로운 분석기법이 필요하다. 최근 공학 분야에서 주목받고 있는 머신 러닝(machine learning) 기법을 이용한다면, 다양한 요인들의 복합적인 작용을 반영한 분양가 예측이 가능하고, 조사자들의 능력, 성향, 경험에 따라 발생할 수 있는 오류의 최소화를 기대할 수 있다.^{2,3)} 따라서 본 연구는 머신러닝 기반 공동주택 분양가 예측모델 개발을 위한 기초연구를 하는 것을 목적으로 한다.

2. 머신러닝 기반 공동주택 분양가 예측모델 개발

데이터 수집을 위해 국내 주택시장을 대표하는 서울과 수도권으로 범위를 한정하였다. 특히, 서울에서는 강남 3구(강남구, 서초구, 송파구) 지역을 대상으로 조사하였고, 수도권 지역에서는 투기 과열이 심하지 않은 수원시(장안구, 팔달구, 권선구, 영통구)를 대상으로 수행하였다. 표 1과 같이 4가지 요인(경제, 입지, 주택, 금융환경 요인)과 실거래 자료를 조사하여 학습 데이터(train data)로 활용하였다. 각 지역별로 2017년 7월부터 2020년 7월까지의 3년간의 2,664건의 실거래 사례를 대상으로 하였다. 분양가 예측 모델은 그림 1(a)와 같이 머신러닝 알고리즘 중에서 가격예측에 가장 일반적으로 사용되는 서포트 벡터 머신(SVM), 결정트리(DT), 인공신경망(ANN) 알고리즘을 이용한다. 그 결과, 그림 1(b)와 같이 예측값과 실측값간의 상관관계를 비교하면, R^2 값이 0.9601로 기존 다중회귀분석(MRA)에 비해 매우 신뢰할만한 예측 정확도를 보였다.

* 목포대학교 건축공학과 강사, 공학박사

** 목포대학교 건축공학과 조교수, 공학박사

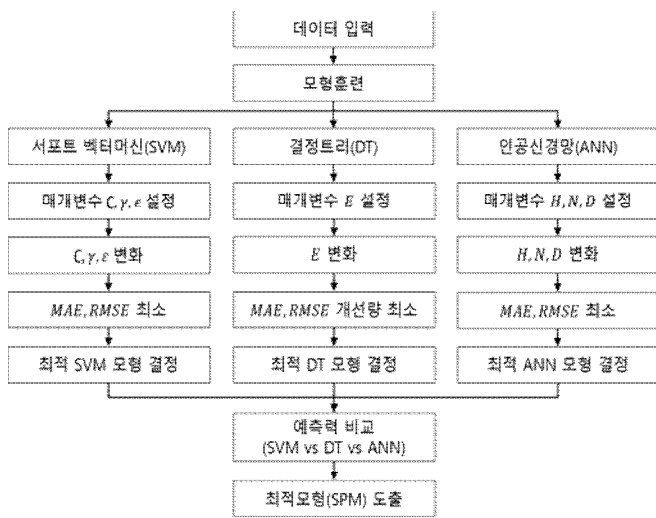
*** 국토교통연구원프라운영원 연구개발실 책임연구원, 박사수료

**** U1대학교 건축공학과 조교수, 공학박사

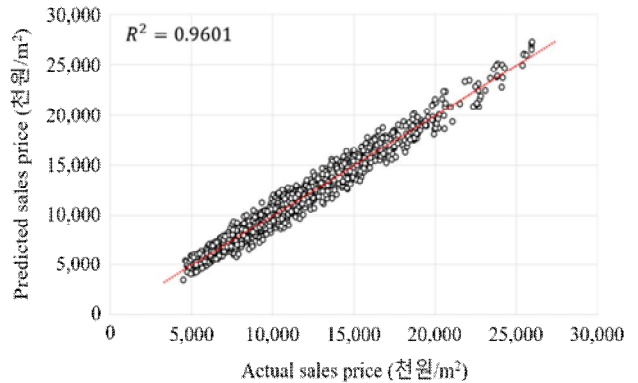
***** 목포대학교 건축공학과 교수, 교신저자(thkim@mokpo.ac.kr)

표 1. 분양가 예측 변수 설정 및 데이터 수집

구분	영향요인	변수	정의	자료수집 출처
독립변수 (입력변수)	경제요인	아파트가격상승률(%)	월별 아파트 가격 상승률	국토교통부 통계자료
		입지요인	의료시설(개소)	의료시설 수
	편의시설(개소)		백화점과 전통시장 수	통계연보
	교육시설(개소)		보육시설 수	통계연보
	주택요인	교통환경(개소)	초, 중, 고등학교 수	교육청 통계
		총세대수(세대)	인근지하철역 수	KB 부동산
		공급면적(m ²)	총 세대수	KB 부동산
	금융환경	대출금리(%)	공급면적	KB 부동산
		아파트 실거래가격(천원)	월별 대출금리	한국은행 통계자료
	종속변수(출력변수)		m ² 당 판매가격	KB 부동산



(a) 모델 개발 프로세스



(b) 실측값 vs 예측값 using SVM

그림 1. 모델 개발 프로세스 및 SVM을 이용한 분양가 예측 결과

3. 결 론

본 연구의 결과는 학술적으로 공동주택 분양가 예측을 위한 보다 설명력 높은 새로운 방법론을 제시하며, 실무적으로 사업을 성공적으로 계획하기 위한 근거로 활용될 것이다.

Acknowledgement

This research was supported by a grant (NRF-2019R1A2C1009398) from the National Research Foundation of Korea by Ministry of Science, ICT and Future Planning.

This research was supported by a grant (NRF-2018R1C1B6004123) from the National Research Foundation of Korea by Ministry of Science, ICT and Future Planning.

참 고 문 헌

1. Jung-ho Kim, "A study on the decision variables of sales price in housing construction project, Master dissertation", pp.3-35, Choongnam National University, 2010
2. Fan, C., Cui, Z., & Zhong, X. (2018, February). House prices prediction with machine learning algorithms. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing, pp.6~10
3. Fan, G. Z., Ong, S. E., & Koh, H. C. (2006). Determinants of house price: A decision tree approach. Urban Studies, Vol.43, No.12, pp.2301~2315.