

<박사학위논문 소개>

주택 시장 분석을 위한 머신러닝 알고리즘에 관한 연구: 서울시 사례*

A Study on Machine Learning Algorithms for a House Market Analysis: A Case of Seoul

저 자: 주 영 아 (Ju, Young A)**

지도교수: 김 갑 성 (Kim, Kabsung)***

본 연구는 서울시 주택 시장의 복잡한 구조와 주요 영향 요인을 규명하기 위해 머신러닝 알고리즘을 활용한 분석을 수행하였다. 기존의 통계적 접근 방식에서 벗어나 물리적 특성, 입지 요인, 거시경제 변수 등을 포함한 다차원적 데이터를 수집하고, 이를 머신러닝 알고리즘에 적용함으로써 시장 동향을 보다 정밀하게 이해하고자 하였다. 데이터는 2017년부터 2024년까지 약 7년간 수집된 서울시 아파트 거래 자료 및 관련 변수를 포함하며, 모델 학습에 앞서 데이터 품질을 향상시키기 위해 체계적인 전처리와 검증 과정을 수행하였다.

연구에서는 XGBoost와 랜덤 포레스트를 포함한 다양한 앙상블 기법을 활용하여 분석을 진행하였으며, SHAP(Tree SHAP) 분석을 통해 각 변수의 중요도를 정량화하고 시장에 미치는 영향을 시각적으로 해석하였다. 분석 결과, 서울 아파트 가격에 영향을 미치는 주요 요인으로는 물리적 특성(예: 전용면적, 주차대수)과 입지 요인(예: 인근 지하철역, 학군)뿐만 아니라, 금리 및 고용률과 같은 거시경제 변수가 포함되었다. 특히 물리적 특성과 입지 요인의 상호작용이 가격 결정에 중요한 역할을 하는 것으로 나타났다.

머신러닝 기반 분석은 기존 통계 모델 대비 예측력과 설명력이 우수하여 서울시 주택 시장의 동태적 변화와 복잡한 패턴을 효과적으로 포착하였다. 이 과정에서 SHAP 분석은 주요 요인에 대한 직관적 해석을 가능하게 하여 주택 시장의 구조를 보다 깊이 이해하는 데 기여하였다. 예를 들어 아파트 전용면적과 주차대수는 물리적 특성 중 가장 중요한 변수로 나타났으며, 인근 학군과 대중교통 접근성은 입지 요인 중에서 가장 큰 영향을 미쳤다. 거시경제 변수로는 금리 수준과 고용률

이 시장 전반에 걸쳐 지속적인 영향을 미쳤다.

본 연구는 머신러닝 기술을 활용하여 서울시 주택 시장을 체계적으로 분석한 국내 사례로서, 빅데이터 기반 분석 방법론을 제시하고 서울 주택 시장의 특수성과 복잡성을 구체적으로 규명하였다. 서울의 주택 시장은 물리적 제한과 급격한 인구 구조 변화로 인해 다양한 요인이 복합적으로 작용하고 있으며, 이를 정량적으로 이해하기 위해 머신러닝 접근법의 활용이 효과적임을 입증하였다. 나아가 글로벌 도시와의 비교 분석, 데이터 품질 개선, 모델 투명성 강화 등의 추가 연구 필요성을 제기하며 향후 연구를 위한 기초를 제공한다. 이러한 분석은 서울 주택 시장의 구조적 특성과 주요 요인을 명확히 밝힘으로써 주택 시장의 동태적 변화를 이해하는 데 기여할 것이다.

주제어 : 머신러닝, 서울 주택 시장, 앙상블 기법, 주택 가격 결정 요인, 모델 해석 가능성, 예측 성능 향상

Keyword : Machine Learning, Seoul Housing Market, XGBoost, SHAP Analysis, Ensemble Methods, Housing Price Determinants, Model Interpretability, Predictive Performance Improvement, Urban Data Analysis

* 본 논문은 필자의 연세대학교 박사학위논문인 주영아(2024), 'A study on Machine Learning Algorithms for a House Market Analysis: a Case of Seoul'를 요약한 것임

** 연세대학교 도시공학박사, j.purplerose@yonsei.ac.kr

*** 연세대학교 도시공학과 교수, kabsung@yonsei.ac.kr