

머신러닝을 활용한 경쟁입찰토지의 낙찰 영향요인 및 예측가능성 연구* -택지개발사업지구 내 공공의 토지판매를 중심으로-

A Study on the Influence Factors and Predictability of Competitive Bidding Land
Using Machine Learning

정 재 훈 (Jung, Jae-Hun)**

전 재 범 (Jun, Jaebum)***

< Abstract >

This study analyzes the factors of winning bids and the predictability of successful bids, focusing on land sales in public housing site development project districts in Korea. First, as a result of logistic model regression analysis to analyze the factors that affect the successful bid of the land to be bid, the individual characteristic variables, the number of announcements (-), area (-), and announcement amount (+), have a significant effect on whether or not the bid is successful. and the economic characteristic variables such as housing construction permit performance (-), unsold housing status (+), real estate sentiment index (+), apartment sales price index (+), land price index (-), and loan interest rate (-) have a significant impact on the winning bid. Second, as a result of analysis using machine learning to compare and analyze the predictability of whether or not the land for the bid is successful, RandomForest (85.44%), XGBoost (91.33%), and LightGMB (91.47%) show high accuracy in that order. Finally, a comparative analysis is conducted to see if the learning period has an effect on the predictability of winning a bid. As a result of the analysis, the longer the period (more data), the higher the prediction accuracy by influencing learning.

Keyword : Machine Learning, Land Bidding, Winning Bid Factors, Winning Bid Prediction, Competitive Bidding

I. 서론

공공택지개발사업은 국가, 지자체, 관련 공사 및 공시행 주체와 주택법상 등록업자의 공동출자 법인이 사업에 필요한 용지 확보 후 단독 및 공동주택용지, 상업업무 용지 등으로 전환 후 시장에 공급하는 사업

이다. 하지만, 사업추진과정에서 공공택지 수요추정에 오류가 발생될 경우 택지수요가 과잉·과소 추정되어 택지수급이 불일치되는 문제가 발생된다. 이는 실수요가 아닌 주택수요에 의해 파생되는 수요에 기반하고 있어 주택수요 추정값의 과잉·과소 가능성이 존재하기 때문이다.

하지만, 현실에서는 이러한 주택수요 추정값의 오

* 이 논문은 2022년 한국부동산개발협회, 한국부동산분석학회 부동산개발분야 신진연구 지원사업의 지원을 받아 연구되었음.

** 본 학회 정회원, KB국민은행 차장, 서울대학교 환경대학원 도시계획학 박사, gaza8@naver.com, 주저자

*** 본 학회 정회원, 강원대학교 교양교육원 부교수, 도시및지역계획학 박사, junjb@kangwon.ac.kr, 교신저자

류 가능성이 발생하기 쉬운 구조이다. 택지개발사업의 특성상 대규모 개발에 있어서 공공목적에 따라 대상지역 내의 토지를 전면 매수하여 개발계획에 의해서 사업이 진행될 수 밖에 없기 때문에 자체적인 택지수요 추정이 불가피하기 때문이다. 따라서 택지수급의 정합성을 높이는 보완장치가 시급하다(장인석 외, 2016).

이러한 문제점의 대안으로 공공택지의 경쟁입찰 방식을 통한 토지판매를 들 수 있다. 택지수급 불일치로 발생한 수요추정의 오류로 파생된 과잉 과소 공급토지의 시장가격을 경쟁입찰의 방식으로 조정하여 수급을 조정할 수 있기 때문이다. 즉 매각토지의 경쟁입찰을 활용하여 토지가격의 조정매커니즘이 활용되는 것이다. 「주택법」 제16조에 따르면, 판매시설 용지 등 영리를 목적으로 사용될 택지의 경우 사업계획의 승인을 받아 건설하는 공동주택의 건설용지 외의 택지 등에 해당하기 때문에 경쟁입찰의 방법에 의하여 공급할 수 있다(윤성현·이성근, 2015).

하지만, 경쟁입찰 방식의 토지 판매는 공급자인 사업시행자로 하여금 유찰시 발생될 판매위험을 가질 수밖에 없다. 경쟁입찰로 판매되는 토지가 유찰되어 미매각된다면, 미매각토지자산의 증가로 인하여 사업시행자의 재무구조가 악화될 것이기 때문이다. 따라서 공공택지개발사업에 대한 사업시행자의 최소한의 수익성 보장과 택지개발사업의 영속성 확보를 위하여 경쟁입찰토지의 낙찰 영향요인을 분석하고 낙찰/유찰 예측 가능성을 분석하는 연구는 필요하다.

따라서, 본 연구는 머신러닝(Machine Learning)을 활용하여 공공택지개발사업지구 내 토지판매를 중심으로 경쟁입찰토지의 낙찰에 미치는 영향 요인을 밝히고, 낙찰 예측 가능성을 분석하고자 한다. 구체적으로 2장에서는 택지개발사업지구 내 공공의 토지판매 특성과 경쟁입찰토지의 낙찰 영향요인을 살펴보고, 3장에서는 머신러닝 학습모형과 학습에 사용할 분석자료를 확정하고 4장에서는 로지스틱회귀분석(Logistic Regression)을 통하여 경쟁입찰토지의 낙찰 영향요인 분석하고, 머신러닝 모형들을 활용하여 경쟁입찰토지의 낙찰 예측 가능성 분석하고, 학습기간에 따른 낙찰에 관한 예측 가능성 차이를 비교 분석하고자 한다.

II. 선행연구

1. 공공택지개발사업지구내 토지판매과정의 특성

택지개발사업은 국가, 지자체, 관련 공사 및 공공시행 주체와 주택법상 등록업자의 공동출자 법인이 사업에 필요한 용지 확보 후 단독 및 공동주택용지, 상업업 무용지 등으로 전환 후 시장에 공급하는 사업이다.

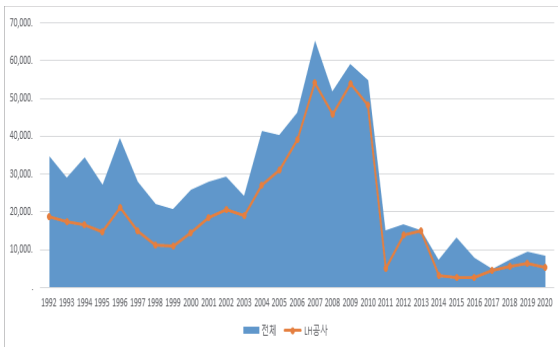
택지개발지구 공급현황을 살펴보면, 택지개발지구 공급은 공공기관(국가, 지자체, 한국토지주택공사, 각 지방공사 등)이 개발사업(택지개발사업, 도시개발사업, 국민임대단지조성사업, 보금자리주택지구, 대지조성사업, 산업단지조성사업, 유통단지조성사업, 경제자유구역조성사업, 주택단지조성사업 등)시행으로 조성한 주택건설용지와 공공시설용지를 수요자에게 공급한 실적을 의미하며, 공급면적¹⁾은 사업지구내 주택건설용지(단독주택용지, 공동주택건설용지 등)에 공공시설용지(학교, 도로, 공원, 상업용지, 종교시설용지, 주차장용지 등)를 합산한 면적으로, 사업시행자별로 택지지정 및 공급 실적을 나타내고 있으며, 지자체에는 각 지방공사 및 자체 시행분이 포함되었다. 1992년부터 2020년까지의 택지공급현황을 살펴보면, 전체 택지공급에서 LH공사가 공급하는 비중을 확인할 수 있다(그림 1 참고). 최근 2020년에는 전체 택지공급현황에서 공사가 차지하는 비중이 62.3%로 나타났다.

택지공급 실적 추이를 살펴보면, 택지공급은 주택종합계획('03-'12)에 따라 '03년부터 '12년까지 총 430km² 공급 계획되었고, 수도권에서는 '04년 이후 택지지정 증가 영향으로 택지공급이 급증하는 추세를 나타내고 있으며 이는 부족한 택지공급을 위한 정부의 주택종합계획을 차질없이 수행한 결과로 판단된다. '03~'07년 기간동안 218km²를 공급하여 주택종합계획 전체물량(430km²)의 50%을 달성, '04년 이후 4개 년간 택지공급은 '04년 이전 동일기간 공급실적 대비 80% 증가하였으며, '04~'06년에는 년평균 43km²를 공급, '07년에는 통계자료 작성이후 최대실적인 65km²를 공급하여 주택가격 안정에 기여하였고, '08년에는 전국 52km²로 금년목표 31km²의 169%로 초과 달성(혁신도시 공급(20km²))으로 초과 달성, '09년에는 전국 59km²로 금

1) 택지공급실적은 각 개별 사업법에 의한 택지공급 승인일, 공급승인제도가 없는 개발사업은 당해 택지의 주택건설사업 승인일을 기준으로 산정하여 실제 택지분양과는 다소 시간적 차이가 있음.(출처: 국토교통 통계누리)

년목표 50km²의 118%로 초과 달성, '10년에는 54.8km²를 공급하여 계획 58.8km²의 93.2%로 정상 추진 중이며, '11년에는 15.2km²를 공급하여 계획 16.7km²의 91%로 정상 추진중이고, '12년에는 16.6km²를 공급하여 계획 22.7km²의 73%로 부동산시장 상황을 고려 공급 순연되었고, '13년에는 19.1km² 공급계획이었으나 부동산경기 침체로 15.1km²(79%) 공급, '14년에는 7.5km²를 공급하여 계획 11.7km²의 64%로 부동산시장 상황을 고려 공급 순연되었고, '15년에는 13.2km²를 공급하여 계획 14.6km²의 90%를 공급, '16년에는 7.9km²를 공급하여 계획 7.7km²의 103%를 공급, '17년에는 5.0km²를 공급하여 계획 8.0km²의 63% 공급되었다(국토교통 통계누리, 2022).

<그림 1> 공공택지개발사업지구내 연간 택지공급현황



2. 경쟁입찰토지의 낙찰영향요인

대규모 택지개발사업을 진행할 때, 판매시설용지 등 영리를 목적 택지의 경우에는 「주택법」 제16조에 따라서 경쟁입찰의 방법에 의하여 공급할 수 있다. 경쟁입찰토지의 낙찰영향요인은 토지의 가격에 미치는 영향요인에서 이론적인 근거를 찾아볼 수 있다. 가격에 영향을 미치는 요인으로 전반적으로 토지의 가격에 영향을 미치는 경우와 지역적으로 토지에 영향을 미치는 요인 마지막으로 개별 토지 가격에 영향을 주는 요인들로 구분할 수 있다(윤성현·이성근, 2015; 최기현·이상엽, 2017).

우선 전반적으로 국토 전반에 걸쳐서 영향을 주는 요인을 살펴보면, 토지의 가격에 영향을 주는 사회적 환경으로 인구와 가구수, 공공시설 서비스 현황, 교육 수준 등을 말한다. 이들 요인 외에 토지의 가격에 영향

을 주는 경제적 변수들도 포함된다. 거시경제 변수중에 저축수준, 무역수지, 물가수준, 임금수준, 재정상황, 조세부담정도 등을 포함할 수 있다. 마지막으로 토지의 가격에 영향을 주는 요인중에서 법제도와 관련된 행정요소들을 들 수 있다(김보미·장희순, 2009; 김상희, 2019; 한동근, 2011).

두 번째로 지역적으로 영향을 주는 요인을 살펴보면, 토지의 지리적 위치에 따른 영향요인들을 들 수 있다. 구체적으로 개별 토지의 물리적 상태, 자연환경, 자연자원 포함 유무와 토지 개별적 특성인 부동산의 용도 및 이용 상황 등을 들 수 있다(이희수 외, 2010; 정상국·오염석, 2010).

마지막으로 개별 토지의 특성에 따른 영향요인을 살펴보면, 토지의 위치, 지반, 통풍, 고저, 면적, 지반, 일조 등의 변수들을 생각할 수 있다. 구체적으로 살펴보면, 용도지역을 고려한 토지의 접한 도로폭, 토지의 배후세대로의 역할이 가능한 아파트 등 주거지와와의 최단거리, 근린상가 등을 포함한 중심상업지구 등과의 접근성 등이 존재한다. 「표준지의 선정 및 관리지침」에서도 제2조를 통하여 표준지 공시지가의 가격에 영향을 주는 요인들을 밝히고 있다. 지역의 토지가격에 영향을 주는 사회적, 행정적, 경제적 요인들을 의미하며, 개별 토지의 가격에 영향을 직접적으로 주는 면적, 형상, 이용상태, 위치 등의 개별요인들이 토지의 가격에 영향을 미치는 변수들이라고 말한다(김시백, 2002; 김용순 외 2011; 윤성현·이성근, 2015).

III. 연구방법

1. RandomForest, XgBoost, LightGMB 학습모형

RandomForest는 2001년 Leo Breiman이 소개한 방법으로 여러개의 의사결정나무를 결합하고, 하나의 모형으로 만드는 방법이다. 즉, 붓스트랩(bootstrap) 표본을 여러개 생성하고, 의사결정 나무모형을 사용하여 분석하고 결과들을 종합하여 사용하는 방법이다. 이를 앙상블(Ensemble) 모형이라고 하는데, RandomForest의 특징은 붓스트랩을 통한 표본들을 만드는 과정에서 임의성(random)을 이용하고, 의사결정 나무모형들을 연결할 때마다 설명변수들을 선택하는 과정 또한 임의성을 사용한다는 점이다. 이러한 방법론

상의 차이로 인하여 다른 앙상블 모형들과 달리 Random Forest모형의 경우 최대한 랜덤한 방법을 통하여 의사결정나무 간에 상관관계를 낮출수 있으며 이를 통해서 예측오차를 줄어들게 할 수 있는 것이다. 또한 의사결정나무 수가 많으면 많을수록 예측오차가 줄어들고 과적합의 결과를 피해갈 수 있다는 장점이 있는 모델이다. 일반적으로 모형을 검증하는 방법으로 데이터를 분리하여 검증을 한다. 분리방법으로 학습과 테스트 데이터로 나누는 방법과 cross-validation 등을 사용하여 분류하는 방법이 있다. 하지만, Random Forest 모형의 경우 데이터를 학습과 테스트 데이터로 분리할 필요가 없다. 붓스트랩 표본을 생성하는 시점에서 붓스트랩 표본으로 선정되지 않은 데이터들(OOB, Out-Of-Bag)을 테스트 데이터로 사용하여 모형을 검증할 수 있다. RandomForest 모형의 경우 기존 모형에서 보여주는 모형의 설명력이 높지만, 예측력이 낮은 단점을 보완하였다. 임의성을 최대로 높여서 붓스트랩을 통한 모형 생성으로 이를 해결하였으며, 설명변수가 많을수록 예측력이 높아지는 결과를 보여주며 안정적인 모형을 보여주고 있다. 많은 연구자들이 RandomForest를 활용하여 연구를 수행하였고, 우수한 결과를 보여주고 있다(W. Chen, 2014).

XGBoost는 선형 모델과 의사결정기반 나무모델에서 발생하는 과적합 문제를 해결하고자 모형의 발전이 시작되었으며, 많은 규모의 데이터셋에 관한 안정성을 확보하고 훈련속도를 높이기 위한 방법으로 Tianqi Chen과 Carlos Guestrin이 처음 소개한 머신러닝방법이다. XGBoost는 eXtreme Gradient Boosting의 약자이다. 크게 보면, Goosting algorithm 기반의 모델이라고 생각할 수 있으며, 회귀분석과 분류모형을 수행하고, 순위를 구분할 수 있으며, 사용자 정의의 objective을 지원할 수 있는 다양한 목적에 사용 가능한 유연한 모델이라고 할 수 있다. Boosting모형의 경우 분석을 진행해가며 오차를 수정하는 Sequential 프로세스로, 이전의 나무들로부터 얻은 정보들을 다음 나무들을 만드는데 활용하며 오차들을 교정하며 분석을 진행한다. Gradient boosting모형에서는 나무의 의사결정과정에서 부정적인 손실(negative loss)이 발생하면 그 과정에서 작동을 멈추는 프로세스를 가지고 있는데 반해, XGBoost 모형의 경우는 분석을 수행시 파라미터로 지정한 반복수행수준(max_depth)까지 진행을 하고, 오차율에서 개선이 일정 수준 못 미친

다고 판단될 경우 다시 역방향으로 프로세스를 진행하여 적합성을 확보하기 위한 분석을 진행한다. XGBoost 모형의 특징은 결측치를 내부저장하여 처리하고, 나무를 생성할 때에도 직렬이 아닌 병렬적으로 생성하여 새로운 데이터에 대한 이전의 학습결과 모델을 반영한다는 점이다. 이러한 특성으로 성능을 더욱 향상시키는 방법으로 훈련시키는 연속적인 훈련 알고리즘을 지니고 있다는 특징을 가지고 있다. 또한 학습과정에서 CPU 코어들을 효과적으로 사용할 수 있는 parallel computing을 실시하며, 메인 메모리 공간에 적합하지 않은 데이터들을 처리하기 위하여 cache optimization과 같은 알고리즘 방법을 사용하고 있다. XGBoost모형의 장점은 처리 속도가 빠르다는 점이며, 모델의 이전 결과들을 활용하기 때문에 모델을 계속적으로 개선시킬 수 있다. 성능이 뛰어난 Google, MS Azure, Alibaba 등에서도 많이 활용되고 있는 모델이다(T. Chen, 2016).

2. 분석자료

경쟁입찰토지의 분석자료는 LH청약센터에 공고된 토지 공급계획을 바탕으로 2010년 1월부터 2020년 12월까지 공사가 일반 실수요자들에게 경쟁입찰 방식으로 공급한 전국 공공택지개발사업지구 내 28,454필지를 대상으로 한다.

본 연구에서는 입찰대상 토지의 낙찰 유무를 종속변수로 설정하고, 개별특성(공고금액, 판매소요시간, 면적), 경제적특성(대출금리, 토지가가지수, 아파트매매가격지수, 부동산심리지수, 인허가실적, 토지거래량, 주택미분양률)등 외부요인들의 지역별 데이터를 활용하여 독립변수로 선정한다.

기초 통계분석 결과, 경쟁입찰토지는 28,446건이었으며, 이중 낙찰된 토지는 9,562건이었다. 경쟁입찰토지의 공고되는 평균 토지판매 면적은 2,439㎡로 나타나고 있다. 이를 월별 데이터로 분석해보면, 토지판매 공고액은 월별 평균 1조 3,767억원이었으며, 월별 최솟값은 480억원 최댓값은 4조 5,097억원으로 나타났다. 입찰토지대상으로 실제 공급된 금액은 월별 공급액은 8,295억원으로 나타났다. 토지판매 공고일로부터 낙찰일까지의 토지판매소요시간은 평균 730일이 걸렸으며, 최소값은 100일 최댓값은 2,362일이 소요됐다. 나머지 경제적특성 변수들의 기초통계분석 결과는 <표 1>과 같다.

<표 1> 기초통계분석 결과표

구분	평균	표준편차	최소값	최대값
공고금액(억원)	38.7	136	0.00	8,030
공고횟수(횟수)	4	3	1	18
면적(m ²)	2,439	15,163	0.1	691,083
거래량(천m ²)	165,904	24,780	114,964	229,557
주택건설인허가실적(건)	355,327	178,864	14,841	885,394
부동산심리지수(지수)	112.3	9.2	87.7	133.1
미분양주택현황(건)	65,477	20,665	19,005	119,125
지가지수(지수)	84.5	7	77.5	101.3
매매가격지수(지수)	85.2	4.2	78.4	95.2
대출금리(%)	4.37	0.95	2.63	5.94

IV. 분석결과

1. 경쟁입찰토지의 낙찰 영향요인 분석

경쟁입찰토지의 낙찰여부에 영향을 미치는 요인 분석을 위하여 로지스틱모형 회귀분석을 실시하였다. 회귀분석결과를 살펴보면, 낙찰여부에 영향을 미치는 요인으로 공고횟수와 면적, 공고금액, 거래량, 주택건설인허가실적, 부동산심리지수, 미분양주택현황, 토지 지가지수, 아파트매매가격지수, 대출금리가 영향을 주는 변수로 확인되었다. 개별특성 중에서는 공고횟수, 면적, 공고금액은 낙찰여부에 미치는 영향에 1%내에서 유의미한 결과를 보여주었으며, 이는 경제적 특성 등 외부요인을 통제하였을 때 토지 낙찰여부가 토지 개별특성에 모두 1%내에서 유의미한 영향을 받는 것을 의미한다. 외부요인인 경제적 특성 변수들을 살펴보면, 주택건설인허가실적, 부동산심리지수, 미분양주택현황, 지가지수, 매매가격지수, 대출금리는 낙찰에 미치는 영향도 모두 1%내에서 유의미한 결과로 나타났다. 토지 개별특성을 통제하여도 낙찰 여부에 미치는 경제적 특성 변수들이 유의미하다는 것을 의미한다(표2 참조).

개별특성 변수를 자세하게 살펴보면, 공고횟수는 낙찰여부에 부(-)의 영향을 미쳤는데 유찰이 거듭될수록 공고횟수가 늘어나며 낙찰이 안되고 있는 것으로 해석된다. 면적은 적을수록 공고금액은 클수록 낙찰가

능성이 높아졌는데, 토지의 감정평가금액으로 평가되어 공고금액으로 공시되는 최저낙찰금액이 높다는 것은 해당토지의 가치가 상대적으로 높은 용도(상업용지 등)에 포함되어 있다는 것을 의미하고, 위의 실증결과는 개발이 용이한 면적의 높은 가치일수록 낙찰가능성이 높다고 해석할 수 있다(홍진석 · 김종진, 2018).

경제적 특성 등 외부요인을 자세히 살펴보면, 주택건설인허가실적이 줄어들수록 미분양주택현황이 늘어날수록 토지 낙찰에 양(+)의 영향을 주었다. 이는 주택건설인허가실적의 증가는 건설투자 증가 및 건설경기 상승을 의미하여, 토지의 판매가 늘어나지만, 건설인허가실적과 토지시장은 금리, 정부정책 등의 영향으로 인해 양(+)의 상관관계를 나타내기까지 시차가 존재한다는 기존의 연구결과와 일치하는 실증결과이다(김용순 · 이현림, 2011).

부동산심리지수와 아파트매매가격지수는 토지 낙찰에 양(+)의 영향을 주었으며, 토지 지가지수와 대출금리는 토지 낙찰에 부(-)의 영향을 주었으며, 이는 토지 수요량에 영향을 주는 변수들이라고 해석된다(박천규 · 김태환, 2015; 김은영, 2017).

<표 2> 로지스틱모형 회귀분석 결과표

낙찰여부	Coef.	Std.Err.	z	P> z
공고횟수	-0.839055**	0.014069	-59.64	0.00
면적	-0.000064**	0.000007	-9.72	0.00
공고금액	0.000000**	0.000000	7.68	0.00
거래량	0.000000	0.000001	0.63	0.53
주택건설인허가실적	-0.000001**	0.000000	-11.16	0.00
부동산심리지수	0.005675**	0.002337	2.43	0.01
미분양주택현황	0.000015**	0.000002	8.25	0.00
토지 지가지수	-0.211056**	0.007803	-27.05	0.00
아파트매매가격지수	0.376199**	0.016982	22.15	0.00
대출금리	-0.867514**	0.042892	-20.23	0.00
constant	-10.23296**	1.392082	-7.35	0.00

2. 경쟁입찰토지의 낙찰 예측 가능성 분석

경쟁입찰토지 낙찰/유찰 여부에 관한 예측 가능성 비교분석을 위하여 전체 28,446필지의 데이터를 9:1의 훈련 데이터와 테스트 데이터로 분류하여 데이터를 구축하였다. 이후 로지스틱 회귀모형과 RandomForest, XgBoost, LightGMB 머신러닝 모형으로 연구모형을

설정하고, 예측 가능성을 비교분석을 실시하였다.

분석결과, 로지스틱 회귀모형의 경우 경쟁입찰토지의 낙찰/유찰 여부에 대한 예측 가능성을 평가한 결과 84.159%의 정확성을 보였으며, RandomForest 머신러닝모형의 경우는 튜닝을 통하여 기존 66.613%에서 85.44%로 정확성을 향상시켰다.²⁾ XGBoost 머신러닝모형은 91.33%의 예측 정확성을 보여주었으며, Light GMB 머신러닝모형은 91.47%의 가장 높은 정확성을 보여주고 있다.³⁾

요약하면, 전체기간 데이터의 90%로 학습을 실시하여 나머지 10%를 예측한 결과 낙찰여부에 관한 예측 가능성을 91.47%까지 향상시킬 수 있었다(표3 참조).

<표 3> 경쟁입찰토지의 낙찰여부에 관한 예측 정확성 분석표

연구 모형	예측 정확성
Model1	로지스틱 회귀모형 84.15%
Model2	RandomForest 머신러닝모형 85.44%
Model3	XgBoost 머신러닝모형 91.33%
Model4	LightGMB 머신러닝모형 91.47%

3. 학습기간에 따른 낙찰에 관한 예측 가능성 차이 분석

앞서 전체기간인 2010.1~2020.12 까지의 경쟁입찰 토지 낙찰/유찰 여부에 관한 예측 가능성 비교분석을 실시하였다. 이번 장에서는 기간을 달리하면 낙찰여부에 관한 예측 가능성을 높일 수 있을지 비교 분석해보고자 한다.

이를 위하여 전체 28,446필지의 데이터를 최근 5년간(2016.1~2020.12), 최근 3년간(2018.1~2020.12), 최근 1년간(2020.1~202012)으로 데이터를 분리 생성하고, 전체기간의 모델링 방법과 마찬가지로, 9:1의 훈련 데이터와 테스트 데이터로 분류하여 데이터를 구축하였다. 이후 로지스틱 회귀모형과 RandomForest, XgBoost, LightGMB 머신러닝 모형으로 연구모형을 설정하고, 예측 가능성을 비교분석을 실시하였다.

분석결과를 다음과 같다. 최근 5년간(Type5) 데이

터를 분석하여 낙찰여부를 예측한 결과, 로지스틱 회귀모형의 경우 83.02%의 정확성을 보였으며, RandomForest 머신러닝 모형의 경우는 80.75%, XGBoost 머신러닝모형은 88.87%, LightGMB 머신러닝모형은 88.97%의 예측 정확성을 보여주었다. 최근 3년간(Type3)의 경우 로지스틱 회귀모형은 81.04%, RandomForest 머신러닝모형은 80.31%, XGBoost 머신러닝모형은 87.63%, LightGMB 머신러닝모형은 87.70%의 예측 정확성을 보여주었다. 마지막으로 최근 1년간(Type1)의 예측 가능성 분석결과, 로지스틱 회귀모형은 80.19%, RandomForest 머신러닝모형은 80.19%, XGBoost 머신러닝모형은 87.62%, Light GMB 머신러닝모형은 87.31%의 예측 정확성을 보여주었다.

요약하면, 학습의 기간이 길수록 낙찰여부에 관한 예측정확성이 높아졌으며, LightGMB 머신러닝모형이 높은 정확성을 보여주는 모형으로 분석되었다(표4 참조). 이는 데이터가 많을수록 학습에 도움을 주어 예측정확성을 높일수 있다는 기존의 연구와 동일한 실증결과이다(W. Chen, 2014).

<표 4> 학습기간에 따른 예측 정확성 차이 분석표

연구 모형	예측 정확성		
	Type5 (5년)	Type3 (3년)	Type1 (1년)
Model1	로지스틱 회귀모형 83.02%	81.04%	80.19%
Model2	RandomForest 머신러닝모형 80.75%	80.31%	80.19%
Model3	XgBoost 머신러닝모형 88.87%	87.63%	87.62%
Model4	LightGMB 머신러닝모형 88.97%	87.70%	87.31%

V. 결론

본 연구는 머신러닝을 활용하여 공공택지개발사업 지구 내 토지판매를 중심으로 경쟁입찰토지의 낙찰 영향요인과 예측 가능성을 비교분석하였다.

우선, 경쟁입찰토지의 낙찰여부에 영향을 미치는

2) RandomForest의 하이퍼패러미터는 max_depth=50, random_state=20을 적용하여 기존 정확성을 향상시켰음.

3) XgBoost, LightGMB의 하이퍼패러미터는 n_estimators=200, max_depth=50, random_state=20을 적용하였음.

요인분석을 위하여 로지스틱모형 회귀분석을 실시하였다. 경쟁입찰 방식으로 공급한 전국 공공택지개발사업지구 내 입찰대상 토지의 낙찰 유무를 종속변수로 설정하고, 개별특성(공고금액, 판매소요시간, 면적), 경제적특성(대출금리, 토지지가수, 아파트매매가격지수, 부동산심리지수, 주택건설인허가실적, 토지거래량, 미분양주택현황)등 외부요인들을 독립변수로 선정하여 분석하였다. 구체적으로, 공고횟수는 낙찰여부에 부(-)의 영향을 미쳤으며, 면적은 적을수록 공고금액은 클수록 낙찰여부에 양(+)의 영향을 미쳤다. 이는 유찰이 거듭될수록 낙찰이 안되고, 면적이 적을수록 공고금액은 클수록 낙찰가능성이 높아졌는데, 개발이 용이한 면적의 높은 가치의 신규공고 토지일수록 낙찰가능성이 높다고 해석할 수 있다. 경제적 특성 변수들의 경우, 주택건설인허가실적은 토지 낙찰여부에 부(-)의 영향을 주었고, 미분양주택현황은 양(+)의 영향을 주었는데, 이들 변수들과 토지시장은 금리, 정부정책 등의 영향으로 인해 양(+)의 상관관계를 나타내기까지 시차가 존재한다고 해석할 수 있다. 부동산심리지수와 아파트매매가격지수는 토지 낙찰여부에 양(+)의 영향을 주었으며, 토지지가수와 대출금리는 토지 낙찰에 부(-)의 영향을 주었다. 이들 변수들은 토지 수요량에 영향을 주는 미치는 변수들로 해석된다.

둘째, 경쟁입찰토지 낙찰 여부에 관한 예측 가능성을 비교분석하기 위하여 전체 28,446필지의 데이터를 훈련 데이터와 테스트 데이터로 분류하여 다양한 모형을 통해서 학습시켰다. 구체적으로, 로지스틱 회귀모형과 RandomForest, XgBoost, LightGMB 머신러닝 모형으로 연구모형을 설정하고, 예측 가능성을 비교분석을 실시하였는데, 분석결과 로지스틱 회귀모형의 경우 토지의 낙찰 여부에 대한 예측 가능성을 평가한 결과 84.159%의 정확성을 보였으며, RandomForest 머신러닝모형의 경우는 85.44%로 정확성을 향상되었고, XGBoost 머신러닝모형은 91.33%의 예측 정확성을 보여주었으며, LightGMB 머신러닝모형은 91.47%의 가장 높은 정확성을 보여주었다.

마지막으로, 기간이 낙찰 여부에 관한 예측 가능성에 영향을 주는지 비교 분석하였는데, 이를 위하여 전체 데이터를 최근 5년간(2016.1~2020.12), 최근 3년간(2018.1~2020.12), 최근 1년간(2020.1~2020.12)으로 데이터를 분리 생성하여 학습시켜 예측 가능성을 비교분석하였다. 분석결과, 데이터가 많을수록 학습에

도움을 주어 예측정확성을 높일수 있었다.

본 연구의 한계점은 자료의 한계로 인하여 토지 개별특성을 입지환경변수 등으로 보다 세부적으로 분석하지 못했다는 점이다. 토지는 도로, 위치 등 개별 입지특성에 따라서 큰 영향을 받기 때문이다. 개별 토지에 대한 입지특성을 추가분석하여 진행된다면 낙찰에 미치는 영향을 보다 정확하게 예측할 수 있을 것이라 판단된다.

논문접수일 : 2023년 5월 5일

논문심사일 : 2023년 7월 4일

게재확정일 : 2023년 7월 27일

참고문헌

1. 김보미·장희순, “단독주택 용지의 가격형성 요인 분석”, 「감정평가학 논집」 제8권 제2호, 2009, pp. 31-42
2. 김상희, “공영개발사업 지구 내 상업용도 토지의 낙찰가를 결정요인에 관한 연구”, 한양대학교 석사학위논문, 2019
3. 김시백, 생존분석을 통한 상업용지 필지 특성에 따른 분양성 연구, 서울대학교 석사학위논문, 2002
4. 김은영, “행태적 부동산소비심리지수 대비 PLS 방법론을 사용한 부동산소비심리지수의 수익률 예측 가능성과 분석”, 「도시연구」 제9권 제2호, 2017, pp. 103-117
5. 김용순·이현림, “주택건설시장의 진단 및 활성화 대책”, 「건설경제」 제67권, 2011, pp. 61-73
6. 고현림·송선주·신종철, “아파트 분양계약률에 영향을 미치는 특성요인들에 관한 분석”, 「주거환경」 제16권 제4호, 2018, pp. 159-177
7. 박천규·김태환, “주택시장 소비심리지수를 활용한 시장 진단 및 활용 방안”, 「부동산분석학회 학술발표논문집」 2015, pp. 53-66
8. 이희수·안유정·이만형, “택지개발지구 내 상업용지 토지이용계획 수립방안”, 「부동산학보」 제42권, 2010, pp. 299-313
9. 윤성현·이성근, “공공택지개발사업지구내 경쟁입찰토지의 낙찰가율에 미치는 영향요인 분석”, 「부동산연구」 제25권 제3호, 2015, pp. 41-52
10. 장인석·윤정란·이상훈, “택지 적정 공급을 위한 모니터링 지표 개발 연구”, 「부동산학보」 제59권, 2014, pp. 246-257
11. 정상국·오염석, “택지개발사업지구의 상업용지 분양률 결정요인에 관한 연구”, 「부동산학보」 제40권, 2010, pp. 324-337
12. 한동근, “토지거래량과 토지가격의 관계에 관한 실증분석”, 「경제연구」 제29권 제2호, 2011, pp. 23-47
13. 홍진석·김종진, “공공택지개발사업지구 내 상업용지의 공급가격과 낙찰가격 비교연구”, 「주거환경」 제16권 제3호, 2018, pp. 17-30
14. 최기현·이상엽, “택지개발지구 공동주택용지의 매각결정 요인에 관한 연구”, 「부동산학연구」 제23권 제2호, 2017, pp. 63-73
15. 국토교통 통계를누리, <https://stat.molit.go.kr/portal/main/portalMain.do>. 2023.5.1.
16. T. Chen, C. Guestrin, “XGBoost: A Scalable Tree Boosting System”, KDD’16, 2016, 785-794
17. W. Chen, S. Li, et al., “Forested Landslide Detection using LiDAR data and the Random”, 2014

<국문요약>

머신러닝을 활용한 경쟁입찰토지의
낙찰 영향요인 및 예측가능성 연구
-택지개발사업지구 내 공공의 토지판매를 중심으로-

정 재 훈 (Jung, Jae-Hun)

전 재 범 (Jun, Jaebum)

이 연구는 한국의 공공택지개발사업지구 내 토지판매를 중심으로 경쟁입찰토지의 낙찰요인과 낙찰여부의 예측 가능성을 분석하였다. 첫째, 경쟁입찰토지의 낙찰여부에 영향을 미치는 요인분석을 위하여 로지스틱모형 회귀 분석한 결과, 개별특성변수인 공고횟수(-), 면적(-), 공고금액(+)으로 낙찰여부에 유의미한 영향을 미쳤으며, 경제적 특성변수인 주택건설인허가실적(-), 미분양주택현황(+), 부동산심리지수(+), 아파트매매가격지수(+), 토지지가 지수(-), 대출금리(-)으로 낙찰에 유의미한 영향을 주었다. 둘째, 경쟁입찰토지 낙찰 여부에 관한 예측 가능성을 비교분석하기 위하여 머신러닝을 활용하여 분석한 결과, RandomForest (85.44%), XGBoost (91.33%), LightGBM (91.47%)순으로 높은 정확성을 보여주었다. 마지막으로, 학습기간이 낙찰 여부에 관한 예측 가능성에 영향을 주는 지 비교 분석하였는데, 분석결과 기간이 길수록(데이터가 많을수록) 학습에 영향을 주어 예측정확성을 높일수 있었다.

주 제 어 : 머신러닝, 토지낙찰, 낙찰영향요인, 낙찰여부, 경쟁입찰