

토지 보상비 결정 요인 분석 - 건설CALS 데이터 중심으로

이상규*, 서명배, 김진욱
한국건설기술연구원

Analysis on the Determinants of Land Compensation Cost: The Use of the Construction CALS Data

Sang-Gyu Lee*, Myoung-Bae Seo, Jin-Uk Kim
Korea Institute of Civil Engineering and Building Technology

요약 본 연구는 건설 전주기 (기획, 설계, 시공, 관리) 과정에서 생성되는 건설 CALS(Continuous Acquisition & Life-Cycle Support) 시스템 내의 데이터 셋 (443개)을 활용하여 토지보상비에 영향을 주는 주요 결정 요인을 분석한다. 해당 분석을 위해 기존 토지 비용 관련 연구에서 활용된 주요 변수를 활용하였다. 이를 기반으로 8개 (토지면적, 개별 공시지가, 감정평가액, 지목, 용도지역 1, 지형 고저, 지형 형상, 도로 접면)의 주요 변수를 활용하였다. 더불어, 해당 변수는 기계학습 알고리즘 기반의 Xgboost 알고리즘을 통해 변수별 중요도 평가를 진행하였고, 해당 변수 중, 개별공시지가가 가장 중요도가 높은 변수로 확인하였다. 토지보상비 결정 요인에 대한 분석 및 검증에 대해 선형다중회귀분석을 활용하였다. 검증을 위해 구성되는 변수로 종속변수는 개별공시지가 변수를 활용하였고, 독립변수는 연속형 변수 1개 (면적), 범주형 변수는 5개 (지목, 용도지역1, 지형고저, 지형형상, 도로접면)를 활용하였다. 본 연구의 모델에 대한 검증 결과, 지목, 용도지역 1, 도로접면에 대한 독립 변수가 유의미한 것으로 확인하였다.

Abstract This study analyzed the determinants of land compensation costs using the CALS (Continuous Acquisition & Life-Cycle Support) system to generate data for the construction (planning, design, building, management) process. For analysis, variables used in the related research on land costs were used, which included eight variables (Land Area, Individual Public Land Price, Appraisal & Assessment, Land Category, Use District 1, Terrain Elevation, Terrain Shape, and Road). Also, the variables were analyzed using the machine learning-based Xgboost algorithm. Individual Public Land Price was identified as the most important variable in determining land cost. We used a linear multiple regression analysis to verify the determinants of land compensation. For this verification, the dependent variable included was the Individual Public Land Price, and the independent variables were the numeric variable (Land Area) and factor variables (Land Category, Use District 1, Terrain Elevation, Terrain Shape, Road). This study found that the significant variables were Land Category, Use District 1, and Road.

Keywords : Land Compensation Cost, Construction CALS(Continuous Acquisition & Life-Cycle Support), Xgboost Algorithm, Assessing Importance of Variables, Multiple Linear Regression Analysis

본 논문은 국토교통부/국토교통과학기술진흥원(과제번호 20SCIP-C146569-03)의 지원으로 수행되었음.

*Corresponding Author : Sanggyu lee(Korea Institute of Civil Engineering and Building Technology)
email: sanggyulee@kict.re.kr

Received July 9, 2020

Revised August 28, 2020

Accepted October 5, 2020

Published October 31, 2020

1. 서론

1.1 연구의 배경

신규 건설공사를 진행하기에 앞서 토지 보상 산정은 해당 건설사업 수행을 위한 예비 타당성 조사시 소요 예산 산정 측면에서 매우 중요한 측면이다. 이는 대규모 국가 개발사업을 시작하기 전에 반드시 예비타당성 조사를 통해 사업의 타당성 검증은 진행하기 때문이다[1].

즉, 국가에서는 정부 예산에 소요되는 대규모 공사 개발사업을 진행하기 위해 재정 지출을 보다 효율적으로 관리하기 위해 "국가재정법"에 따라 정부 총사업비관리를 규정하고 있다. 특히, 정부는 신규 개발사업에 대한 예산편성 및 기금운용계획을 수립함에 있어서 사전에 사업의 타당성을 검증 및 평가하는 절차를 두고 있는데 이러한 절차를 예비타당성조사로 정의할 수 있다[2].

최근 예비타당성과 관련한 사례로, 2019년 6월부터 GTX-A 노선 민간투자사업(파주-서울)의 차량기지 및 지상 환기구 등의 지상구간에 대한 토지보상 계획을 준비하고 있다. 해당 사업의 토지 보상 대상은 경기 파주 및 고양 등지의 토지 330필지와 서울 토지 80필지 등 총 410필지, 29만 8044㎡ 규모이다. 보상계획 공고 이후 보상 감정평가를 거친 뒤 토지 소유자에게 협의 통지를 통해 보상 절차에 따라 토지 보상이 이루어진다.

Lee (2006)에 따르면, 개별공시지가는 정부가 매년 부동산가격공시에 관한 법률에 의거하여 한국감정원에서 공표하는 표준지공시지가와 토지가격 비준표를 기초로 가격배율 조정에 의해서 산출되고 있다[3].

하지만, 이러한 개별공시지가는 실제 보상비와의 오차가 발생함으로써 이해관계자들과의 의견 수렴이 매우 어려운 상황이다. 예를 들어, 2018년 말에 발표된 3기 신도시 후보지 지역의 공시지가에 대한 큰 편차로 인해 주민들의 민원이 끊이지 않고 있다. 특히, 전국 평균이 8.03%의 공시지가 상승률에 비해 3기 신도시 후보지 2곳(남양주 왕숙과 인천 계양테크노밸리) 공시지가의 상승폭은 4.6%미만으로 공시지가 상승률의 편차로 인해 정부의 3기 신도시 건설은 정부와 지역 주민간의 큰 부담으로 다가올 수 밖에 없는 상황이다.

제시된 Fig. 1은 부동산개발정보 플랫폼 기업인 지존(GZone)의 시스템 자료를 기반으로 연도별 국내 토지보상금 규모 추이도(2011~2021년)를 확인하였다. 총 2011년부터 2017년까지 꾸준히 하락세를 보이다 2018년부터 본격적으로 급증하였고, 향후 2020년에는 30조 원, 2021년에는 30+α 로 꾸준히 증가할 것으로 예상하

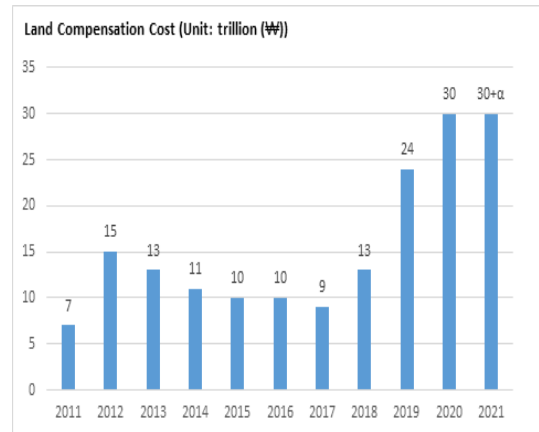


Fig. 1. Trends of Domestic's Land Compensation (2011~2021)

고 있었다.

이는 올해부터 정부가 24조원 넘는 토지보상 비용을 부담해야 하는 상황으로 향후 신도시 3기 및 각종 건설 개발사업 등에 따른 매년 30조원 넘는 토지보상 비용이 가중될 경우 정부의 재정부담은 물론 대규모 토지보상금을 노리기 위한 투자 수요로 인해 부동산 시장이 매우 불안정한 상황으로 심화될 것으로 예상된다.

따라서, 본 연구에서는 1990년대 후반부터 건설 전주기(기획, 설계, 조달, 시공, 유지보수 등)의 과정에서 발생하는 건설CALS(Continuous Acquisition & Life-Cycle Support)에서 생성된 데이터, 즉 토지면적, 개별 공시지가, 지목, 용도지역 1, 지형 고저, 지형 형상, 도로 접면 데이터를 기반으로 다중선형회귀분석을 통해 토지 보상비 결정 요인을 확인하고자 한다.

1.2 연구의 범위 및 방법

본 연구는 건설 CALS 시스템 내에서 수집되고 있는 토지 가격과 관련한 건설CALS내 주요 변수를 토대로 토지 보상비의 결정 요인을 확인한다. 이를 위해 아래와 같이 연구를 진행한다.

우선, 기존 연구 분석을 통해 토지 보상비 결정 요인을 위한 시사점을 도출한다. 이를 위해서, 토지보상비를 결정하는 주요 요인에 대한 관련 연구 및 건설CALS 시스템, Xgboost 알고리즘에 대해서 확인한다.

그리고, 기존 연구 분석에서 도출된 주요 시사점을 토대로 현재 건설CALS 시스템 내에서 필요한 데이터를 확인하고 분석을 위한 주요 변수에 관한 항목을 확인한다.

또한, 토지보상비 결정 요인 분석을 위해 Xgboost 알

고리즘을 통해 변수의 중요도를 판별하고 이를 기반으로 종속변수와 독립변수를 구성한다.

요인 검증을 위해서 본 연구에서는 다중선형회귀분석을 활용한다. 다중선형회귀분석은 독립변수가 2개 이상일 경우 활용하는 방법으로, 임의의 독립변수 외에 다른 변수들이 통제된 상황에서 종속변수를 충분히 설명할 수 있다. 토지 특성에 관한 논문 중에서 다중선형회귀분석을 통해 변수를 검증하는 연구가 활용되고 있었다[4,5]. 따라서, 본 연구에서는 건설CALIS 데이터를 기반으로 다중선형회귀분석을 통해 토지 보상비의 결정 요인을 검증 및 확인한다.

2. 관련 연구

2.1 토지 보상비 결정 요인

현재 토지 보상비를 예측하기 위해서 먼저 예비타당성 조사단계에서 구체적으로 사업이 계획되기 전 단계에서 대략적으로 추정된 토지보상비를 확인한다.

기획재정부 예비타당성 수행총괄지침 제 2조 6항에 따르면, 예비타당성 조사 대상으로 건설사업의 총 사업비 500억원 이상이면서 국가재정규모가 300억원 이상의 신규사업 대상에 대해서 타당성 검증을 반드시 진행한다 [2].

특히, 건설사업의 총 사업비는 토목, 건축 등 대부분 건설공사에 소요되는 모든 경비로 공사비, 보상비, 시설부대 경비 등으로 구성된다.

한국개발연구원 공공투자관리센터 자료에 따르면, 보상비의 구성항목으로 직접보상비와 간접보상비로 구분된다. 직접보상비는 토지보상비, 지장물 보상비, 영업보상비, 영농보상비, 어업보상비, 이주대책, 기타로 구분된다. 간접보상비는 측량비, 감정평가 수수료, 권리이전비, 세금 등으로 구분된다[6].

한국개발연구원 공공투자관리센터에서는 토지보상비 추정을 아래 Fig. 2와 같이 총 3가지 방법으로 진행한다.

첫 번째, 별도의 감정평가를 통한 보상비 추정이다. 즉, 감정평가업자에게 직접 의뢰를 통해 개발지구에 편입될 토지의 보상액을 산출하는 방법이다.

두 번째, 주변 보상 사례 가격을 통한 보상비 추정이다. 즉, 기존 사업자 주변 보상 자료를 활용하여 타당성 조사를 의뢰한 사업시행자에게 보상 전례를 확인하여 분석하는 방법이다.

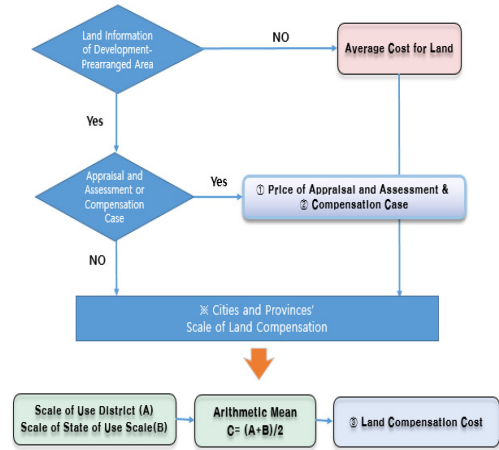


Fig. 2. Valuation Method of Land Compensation Cost (Korea Development Institute)

하지만, 앞서 제시된 감정평가를 통한 보상비 추정과 주변 보상 사례 가격의 방법은 반드시 개발 예정지의 필지정보를 정확히 확인해야하는 점이 전제가 필요하다.

마지막 방법은 시도별 보상배율 산정하여 토지보상비를 산출하는 방법이다. 만약, 개별예정지에 대한 필지정보가 없다면 표준 공시지가와 시도별 보상비율을 곱하여 토지보상비를 산출하게 된다.

아래 Table 1과 같이 기획재정부 예비타당성 수행총괄지침 제 34조 제3항 제3호에 따르면, 시도별 보상배율을 제시하고 있다[2].

하지만, 정부의 효율적인 공시지가 제도를 운영하고 있음에도 불구하고 공시지가율은 현실적으로 많은 차이가 있다[7]. 더불어, 공시지가의 시장가격 반영률은 지역적 편차를 크게 보이고 있다. 도시 지역의 표준 공시지가는 시장가치를 80% 반영하고 있지만 도시 외곽 및 농촌 지역의 경우, 현실적인 공시지가와는 거리가 먼 것으로 확인된다[8].

따라서, 토지 가격에 영향을 미치는 결정 요인을 확인할 수 있는 관련 연구 사례를 통해 토지 보상비의 주요 변수를 적용한다.

김선주, 권기욱(2013)에 따르면, 충청남도 당진시의 토지 실거래 가격을 토대로 주요 토지 변인을 확인하기 위해 종속변수에 개별공시지가와 실거래가격의 비율을 적용하였고, 독립변수에 경사도, 면적, 도로 접면 등을 적용하였다. 특히, 실증 분석을 위해 총 159건의 매매사례를 기반으로 표본자료에 활용하였다[9].

Table 1. Scale of Compensation (Cities and Provinces) (Ministry of Economy and Finance (2018))

Total	Use District					State of Use				
	Residence & Commerce	Green Area	Managem ent Area	Agricultural Area	Residence & Commerce	Commerce	farm-land	Forest-la nd	Public	
Seoul	1.66	1.59	1.84	-	-	1.23	1.52	1.29	2.77	3.66
Busan	1.90	1.87	1.93	-	-	1.86	1.61	1.90	3.00	3.90
Daegu	2.05	1.90	2.18	2.90	2.78	1.92	1.57	2.05	3.89	4.89
Incheon	2.10	1.66	1.77	3.13	2.36	1.66	1.11	2.16	2.64	3.89
Gwangju	2.13	1.54	2.71	2.57	-	1.54	1.31	2.18	2.80	3.28
Daejeon	1.59	1.59	1.83	2.00	3.00	1.59	1.57	1.60	2.59	3.81
Ulsan	2.78	2.09	3.04	2.82	3.00	1.91	1.88	2.45	5.00	4.44
Sejong	2.87	2.55	2.79	3.33	2.75	2.34	2.04	2.70	5.11	4.16
gyeonggi	1.85	1.49	1.92	2.08	2.01	1.63	1.57	1.77	2.70	2.88
Gangwon	2.44	1.89	2.65	2.71	2.68	1.90	1.64	2.38	4.46	4.62
Chungbuk	2.35	1.37	2.38	2.88	2.61	1.74	1.56	2.31	3.07	5.20
Chungnam	2.49	1.93	2.54	2.96	2.39	2.04	1.63	2.33	3.58	4.06
Jeonbuk	2.15	1.82	2.22	2.61	2.09	1.95	1.69	2.11	3.42	4.25
Jeonnam	2.50	2.03	2.75	2.62	2.47	2.17	1.72	2.40	4.50	5.00
Gyeongbuk	2.64	2.24	2.52	2.99	2.54	2.10	1.82	2.52	4.50	5.31
Gyeongnam	2.73	1.96	3.08	3.13	2.62	2.13	1.80	2.70	4.50	4.17
Jeju	2.17	1.73	2.22	2.60	2.71	1.69	1.50	2.43	3.10	4.11

이동원, 정수연(2014)에 따르면, 제주도의 올래길코스 개발에 따라 인근 토지 가격의 상승 요인을 확인하기 위해 종속변수에 개별공시지가를 적용하였으며, 독립변수로 올래길 코스와의 거리, 지목, 이용상황, 용도지역, 도로조건 등을 적용하였다[10].

정형식(2017)에 따르면, 생태적 보호지역 지역에 대한 토지 가격에 영향을 미치는 요인을 분석하기 위해 종속변수로 시군별 개별공시지가와 개별공시지가 상승률을 토대로 독립변수를 경사도, 인구, 도로접근성, 면적 등을 적용하였다[11].

더불어, 강영일, 김호철(2012)에 따르면, 서울역과 청량리역을 중심으로 대중교통 환승시설 설치로 인한 주변 지역의 지가변화에 영향을 미치는 요인을 분석하기 위해 독립변수로 이용상황, 용도지역, 도로접면, 지형형상, 지형고저 등을 적용하였다[12]. 안유정, 이만형 (2008)에 따르면, 단독주택의 노후화 현상을 지형적 요인과 지가의 상관관계를 통해 확인하고 있다. 해당 논문의 지형적 요인에 대한 주요 변수로는 지형고도, 경사로, 지형형상, 대지면적, 도로 폭 등으로 나타났다[13].

건설CALS시스템이 토지보상비 결정 요인에 활용된 선행연구를 아래 Table 2와 같이 정리하였다.

Table 2. Literature Review

Author (Year)	Dependent Variable	Independent Variable
Y. Ahn & M. Lee (2008)	Individual Public Land Price	Terran Evaluation, Terrain Shape, Gradient, Road, Land Area
Y. Kang & H. Kim (2012)	Land Price of the Reference Land	Use District, Road, Terrain Shape, Terrain Evaluation
S. Kim & K. G (2013)	Individual Public Land Price	Gradient, Land Area, Road, etc
D. Lee & S. Jung (2014)	Individual Public Land Price	Distance, Land Category, Use District, Road, etc
H. Jung (2017)	Individual Public Land Price	Gradient, Population, Road, Land Area

이미 많은 보상비 결정 요인에 관한 선행연구에서 개별공시지가를 종속변수로 활용하고 있었으며 독립변수는 지형, 면적, 지목, 도로접면, 용도지역, 지형고저, 지형형상 등으로 나타났다.

하지만, 본 연구는 건설CALS시스템 내에 축적되어 있는 토지보상현황 정보로서 감정평가액에 대한 변수를 추가한다. 나머지 변수로는 기존 선행연구와 동일하게 토지면적, 개별공시지가, 지목, 용도지역1, 지형고저, 지형형상, 도로접면을 활용한다.

2.2 건설CALs 시스템

건설CALs 시스템은 총 5개 단위 시스템(발주기관용 건설사업관리시스템, 계약사용 건설사업관리시스템

템, 건설인허가시스템, 용지보상시스템, 건설CALs포털시스템)으로 구성되고 있다[14].

건설CALs 시스템의 사용처는 국토교통부, 5개 지방 국토관리청, 국토교통부 소속 및 산하기관, 지자체, 건설 업체 뿐만 아니라 일반국민 역시 사용할 수 있는 대국민 포털서비스이다[15].

특히, 용지보상시스템은 건설공사 업무 프로세스에서 시공단계에서 이루어지는 보상업무를 지원하는 시스템으로 정부의 보상비에 집행에 대한 예산 관리를 효율적으로 집행할 수 있게 관리하기 위해 개발된 시스템이다[16]. 특히, 2010년부터 용지보상시스템을 통해 230개 기관에서 활용하고 있으며 타기관 시스템과의 연계를 위해 14종의 토지 대장정보를 연계하였다[17]. 아래 Fig. 3은 김성진 외 3(2019)의 현재 건설 CALs 시스템의 개념에 대한 그림을 인용하여 제시하였다[18].

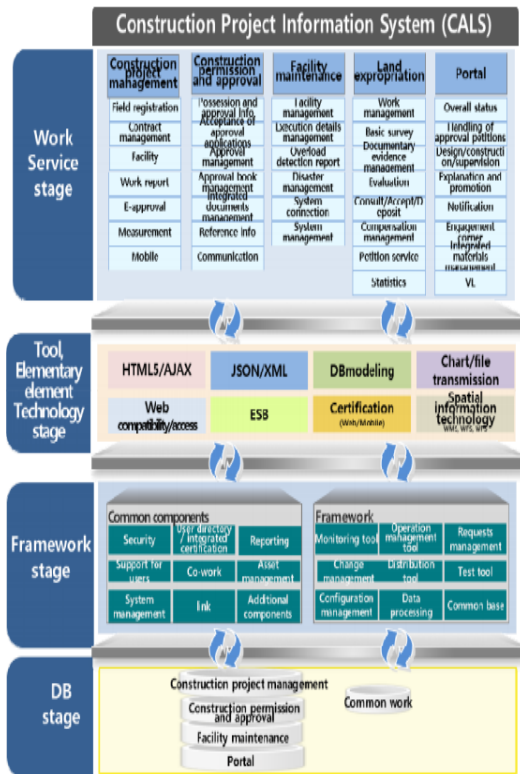


Fig. 3. Concept of Construction CALs (S. Kim et al (2019))

하지만, 건설공사의 기획·설계·시공·유지관리 전 단계에서 발생하는 많은 양의 데이터를 수집하고 있는 건설 CALs시스템은 복잡한 구조와 데이터 누락으로 인해 폭넓은 활용이 힘들다는 한계를 갖고 있다[19].

특히, 건설사업정보시스템의 활용률을 분석한 결과, 건설인허가시스템의 활용률은 높은 반면 용지보상시스템은 낮은 것으로 나타난다. 더불어, 용지보상시스템은 다른 시스템과의 연계성에도 부족한 것으로 나타난다[20].

따라서, 용지보상시스템의 활용성을 높이기 위해 기존 건설CALs 시스템 내의 방대한 양의 데이터를 활용하여 연계할 수 있는 토지보상비의 결정 요인을 확인하는 연구가 반드시 필요하다.

본 연구를 위해 사용되는 데이터는 2008년부터 2018년까지 건설 CALs시스템내의 누적된 전국의 데이터를 바탕으로 구성하였다.

2.3 Xgboost 알고리즘

기존 선형 모델 또는 의사결정나무 (Tree), 즉 트리구조 기반의 모델상 과적합 오류를 제거할 경우 매우 유용한 알고리즘으로 Xgboost 알고리즘을 본 연구의 분석방법론으로 활용하였다.

Xgboost 알고리즘은 데이터를 분류함에 있어 해당 Tree수가 많으면 많을수록 오히려 처리 속도도 빠르며, 정확도면에서 다른 기계학습에 비해 높은 것이 장점이다[21]. 특히, Ha(2017)와 Hebert(2016)은 기존 Random forest 알고리즘 보다 Xgboost 알고리즘을 통해 데이터를 분류함에 있어 높은 성능을 기반으로 정확하게 분류하고 있다[22,23].

$$Gain = \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} - \gamma \quad (1)$$

수식 (1)에서 트리구조에서 특정한 깊이에서 얻을 수 있는 L은 왼쪽 값, R은 오른쪽 값을 의미하여 이렇게 얻을 수 있는 정보의 획득량을 Gain이라고 의미한다. 즉, Xgboost 모델 결과에 대한 변수의 중요도 평가의 기준은 Gain으로 최대한 랜덤하게 트리 구조를 반복 생성하면서 마지막으로 가장 점수가 높은 값의 트리를 조합하여 가장 많은 정보를 획득한 변수를 최적의 분류 값으로 결정하는 것이다[24].

즉, 중요도가 최대가 되도록 가지가 형성됨으로서 전체 트리 구조가 형성되고 있음을 알 수 있다. 이러한 Xgboost 알고리즘에서 변수별 중요도를 평가함에 있어

기준이 되는 3가지 요인 중 gain과 coverage, frequency 가 있다[25]. 일반적으로 Xgboost 모델에서는 gain을 기준으로 변수의 중요도를 판별하고 있음을 확인하였다[26].

관련 연구를 기반으로 본 연구에서도 Xgboost 알고리즘 기반의 변수별 중요도를 적용하기 위해서 gain을 기준으로 분석 결과를 확인한다.

3. 데이터 구성

토지 보상비에 대한 결정 요인을 확인하기 위해서 구성되는 데이터는 토지 보상비 결정 요인에 관한 주요 논문을 통해 주요 변수를 적용하였다[9,10,11,12,13]. 더불어, 건설전주기 과정에서 생성된 정보시스템을 체계화한 건설CALS 데이터 내의 변수를 필터링하여 총 8개의 변수를 확보하였다. 해당 변수는 토지면적, 감정평가액, 개별공시지가, 지목, 용도지역1, 지형 고저, 지형 형상, 도로 접면이다.

특히, 8개 변수 중에서 연속형 변수는 3개 (개별공시지가, 감정평가액, 토지면적)로 확인하였다. 나머지 범주형 변수는 5개 (지목, 용도지역1, 지형고저, 지형형상, 도로접면)로 확인하였다.

구체적인 변수의 분류 및 내용은 아래 Table 3과 같이 제시한다.

Table 3. Classification and Content of Variables

Classification	Content
Continuous Variable	<ul style="list-style-type: none"> Individual Public Land Price (Unit: ₩) Appraisal and Assessment (Unit: ₩) Land Area (Unit: m^2)
Categorical Variable	<ul style="list-style-type: none"> Land Category (Farm -> 1/ Building Site-> 0) Use District 1 (Residential -> 1/ Agricultural -> 0) Terrain Elevation (Flat -> 1 / Slope -> 0) Terrain Shape (Vertical Rectangle -> 1 / Indeterminate -> 0) Road (Medium size (12m~25m) -> 1 / Small Size (8m~12m) -> 0)

건설CALS시스템 내에 누적된 전국의 데이터 중에서 랜덤하게 선택된 지역 1곳을 선정하여 분석에 필요한 데이터를 가공 및 추출을 하였다.

우선, 범주형 변수 5개는 더미변수로 설정하였다. 지

목의 경우, 전, 답, 과수원, 목장 등의 농경지를 1로 건축물이 완공되거나 설치될 수 있는 토지인 대를 0으로 설정하였다. 용도지역 1의 경우, 주거지역을 1로 농림지역을 0으로 설정하였다. 지형고저의 경우, 평지를 1로 경사를 0으로 설정하였다. 지형형상의 경우, 토지의 좁은 면이 도로에 접하거나 향하고 있는 토지 형태인 세로장방형을 1로 불규칙한 형상의 토지인 부정형을 0으로 설정하였다. 도로접면의 경우, 폭 12m~25m 미만의 도로 한 면이 접하고 있는 토지인 중로를 1로 폭 8m~12m미만의 도로 한 면에 접하고 있는 소로를 0으로 설정하였다.

이처럼, 범주형 변수(지목, 용도지역 1, 지형 고저, 지형 형상, 도로접면)는 더미변수 (1과 0)로 데이터를 구성하였다.

나머지, 연속형 변수는 더미변수를 생성할 필요가 없기에 범주형 범수를 포함한 총 443개의 데이터 셋에 구성하였다.

토지 보상비 결정 요인 분석을 위해 선형회귀분석을 사용하였다. 분석에 앞서, 토지보상비 결정 요인에 가장 영향을 주는 변수를 종속변수로 적용해야 한다. 이를 위해서, 본 연구에서는 변수의 중요도를 판별하기 위해의 사결정 트리 구조 기반의 Xgboost 알고리즘을 활용하여 종속변수를 확인하고자 한다.

4. 토지 보상비 결정 요인 분석

4.1 Xgboost 알고리즘 적용

앞서 제시한, 건설CALS내 8개 변수별 데이터를 토대로 기계학습 기반의 Xgboost 알고리즘과 회귀분석 방법을 적용하여 토지 보상비 추정에 영향을 주는 주요 변수의 중요도를 확인하고자 한다.

해당 데이터를 바탕으로 분석 프로그램은 R을 통해 Xgboost 알고리즘을 적용하였다. 적절한 분석 결과를 위해 Train set과 Test set의 비율은 7:3으로 구분하여 학습하였다. 해당 알고리즘 내 부스팅 파라미터(eta, max_depth, min_child_weight 등)의 튜닝을 통해 가장 적합한 설정값을 기반으로 교차 검증 (Cross Validation)을 진행하였다.

앞서 언급한 것처럼, Xgboost알고리즘의 변수별 중요도는 트리구조를 통해 획득한 gain 값으로 결정한다. Xgboost 알고리즘을 기반으로 건설 CALS내 8개 변수를 분석 결과, 아래 그림 Fig. 4과 같이 본 연구에서 토지

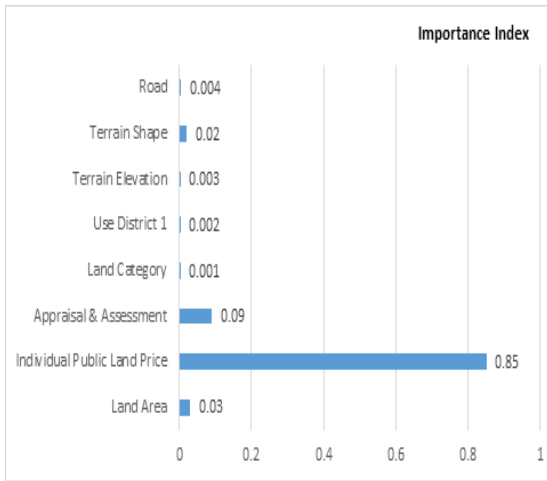


Fig. 4. Assessing the Importance of Variables Based on Xgboost Algorithm

보상비의 주요 변수 중 개별공시지가가 가장 높은 값인 총점 1점 중 0.85로 나타났고, 다음은 감정평가액, 면적, 지형 형상 순으로 중요도 값이 나타났다.

결국, 건설 CALS 시스템 내의 8개 주요 변수 중 개별 공시지가의 변수가 매우 높은 영향을 끼치는 것으로 확인하였다.

4.2 다중선형회귀분석 적용

앞서 제시된 Xgboost 알고리즘을 기반으로 건설 CALS 내 변수의 중요도를 확인한 결과, 개별공시지가가 가장 높은 변수임을 확인하였다. 이로써 개별공시지가를 종속변수로 구성하면서, 그 외에 6개(토지면적, 토지지목, 용도지역1, 지형고저, 지형형상, 도로접면) 변수는 독

립변수로 구성하였다.

토지 비용을 산정하기 위해서 활용되는 데이터 변수는 기존 사례 연구를 기반으로 종속변수와 독립변수를 확인하였다[9,10,11,12,13]. 그뿐만 아니라, 본 연구에서는 기계학습 기반의 Xgboost 알고리즘을 통해 변수별 중요도를 산정하는 방법으로도 종속변수와 독립변수를 구성하는 것이 가능하다는 것을 검증하였다.

다만, 앞서 진행된 Xgboost 알고리즘을 기반으로 변수의 중요도를 평가함에 있어 감정평가액에 대한 변수가 들어갔지만, 실제 토지 보상비의 주요 요인으로는 감정평가액에 대한 변수를 적용하지는 않는다.

이러한 이유로, 감정평가액은 국토부에서 정한 토지의 27가지 지목을 포함하여, 용도지역, 형상, 도로접면의 수, 토지의 형태 등의 여러 가지 변수들을 기준으로 산정하고 있기 때문에 결정요인 분석을 위해서 독립변수로서 적용하는 것은 적절하지 않다고 판단한다. 따라서, 본 연구에서 Xgboost 알고리즘을 기반으로 8개의 변수를 활용하여 변수별 중요도에 대해서만 확인한다.

다중선형회귀분석을 진행한 결과, 제시된 모델의 설명력이 0.79 (R^2 : R Square)로 모집단의 79%를 설명하고 있는 것을 확인하였다. 더불어, 해당 모델 분석을 토대로 독립변수 6개 중 지목, 용도지역 1, 도로접면에 대한 변수가 유의수준(0.05)을 기준으로 유의미한 것으로 확인하였다.

더불어 다중선형회귀분석을 진행할 경우, 독립변수 간의 다중공선성의 문제가 발생할 수 있기 때문에 각 변수별 독립변수에 대한 VIF (Variance Inflation Factor: 분산팽창요인)값을 통해 확인하였다.

Table 4. Analysis on the Determinants of Land Compensation Cost

Independent Variable	Unstandardized Coefficients		T-value	Significance Probability *(5%)	VIF (Variance Inflation Factor) Value
	β	Standard Error			
Land Area	-1.5184	1.7336	-0.875	0.38155	1.250907
Land Category	-25093.0	7617.88	-3.293	0.00106*	1.130251
Use District 1	95332.68	6241.51	15.273	0.00001*	2.393940
Terrain Elevation	108.8329	9431.22	0.0115	0.99079	1.195076
Terrain Shape	-9011.76	5285.29	-1.705	0.08889	2.156304
Road	38278.81	5702.91	6.712	0.00001*	1.021589

위 Table 4와 같이, 각 독립 변수별 VIF 값은 1.00~2.50미만으로 변수간의 다중공선성의 가능성은 없는 것으로 확인하였다.

5. 결론

신규 건설공사에 앞서 예비타당성 조사단계에서 보다 정확한 토지 보상의 비용 추정을 위해서 토지 보상비의 주요 결정 요인을 확인해야 한다.

이를 위해, 본 연구는 건설CALS 시스템내의 데이터를 활용하여 토지 보상비의 주요 결정 요인을 확인하였다.

특히, 기존에 용지보상시스템을 포함하고 있는 건설CALS시스템과의 낮은 연계성으로 인해 데이터 활용 및 누락 등의 많은 문제점이 부각되고 있었기에 본 연구를 통해 건설CALS시스템에서 활용되고 있는 데이터를 기반으로 토지 보상에 대한 요인 분석 연구를 진행하였다.

더불어, 기존 토지 가격 결정에 주요 영향을 미치는 요인에 관한 사례연구를 통해 토지 보상비의 주요 요인을 탐색 및 적용하는 것과 동시에 기계학습 기반의 Xgboost 알고리즘을 통해 건설CALS 데이터에 대한 중요도 평가를 통해 독립변수와 종속변수를 판단하였다.

해당 변수는 토지 보상비 결정 요인 분석을 위해 다중 선형회귀분석을 사용하였다.

분석 결과, 건설CALS 데이터를 기반으로 토지 보상비 결정 요인으로 개별공시지가에 영향을 미치는 지목, 용도 지역 1, 도로접면에 대한 변수가 유의미한 것으로 확인하였다. 이는 앞서 토지 보상비 결정 요인으로 제시된 기존 문헌과 동일하게 도출되었다.

지목이 부(-)의 영향을 미치고 있기에 농경지역보다 건축물이 포함된 토지 보상비에 영향을 미치는 것으로 예상할 수 있다.

하지만, 용도지역 1은 정(+)의 영향을 미치고 있기에 도시지역에 해당하는 주거, 상업, 공업, 녹지 등의 용지지역이 농림지역보다 토지 보상비 상승에 긍정적인 영향을 미치는 것으로 나타난다. 더불어, 도로접면 역시 정(+)의 영향을 미치고 있기에 폭이 넓은 중로에 포함된 토지가 폭이 좁은 소로에 비해 토지 보상비에 영향을 미치는 것으로 확인하였다.

하지만 본 연구에서 종속변수를 개별공시지가만을 사용하는 것은 무리라고 판단한다. 즉, 건설CALS 시스템 데이터 뿐만 아니라 외부 데이터와의 연계를 통해 보다 정확한 분석이 필요하다. 예를 들어, 개별공시지가와 시

계열 시도별 지가상승률을 반영할 수 있는 개별공시지가 현실화율을 통해 신속하게 반영될 수 있는 지가 기준이 필요하다.

더불어, 건설CALS 내의 다양한 변수를 토대로 토지 보상비 산정에 영향을 미치는 변수를 확인해야 한다. 본 논문에서는 건설CALS내 랜덤하게 선택된 A 지역의 토지 특성에 대해서 주요 요인 분석을 진행하였다. 하지만, A 지역으로 제한하여 분석을 제시하는 것보다는 많은 양이 구축되어 있는 건설CALS 데이터를 활용하여 전국 17개 시도에 대한 토지 특성 분석이 필요하다.

본 연구를 기반으로 향후 건설CALS 시스템 데이터와의 지속가능한 시스템 연계 활용 방안과 함께 보다 정확한 토지 보상비 예측 모델을 개발 할 수 있는 기틀을 마련하고자 한다.

References

- [1] S. Ryu, J. Yoo, "A Study on Improvement of Preliminary Feasibility Studies for Road Project," Gyeonggi Research Institute, 2017.
- [2] "Guidelines of Preliminary Feasibility Study Operating Costs," Ministry of Economy and Finance, 2018.
- [3] S. Lee, S. Park, S. Hong, "A Similar Price Zone Determination of Public Land Price Using a Hybrid Clustering Technique," Journal of the Korean Geographical Society, Vol. 4, No. 2, pp. 121-135, 2006.
- [4] S. Lew, J. Kang, "A Study on the Influence Factors of Land Value by Urban Spatial Constitution," Journal of the Korean Society of Civil Engineers, Vol. 32 No. 1, pp. 61-69, 2012.
- [5] J. Jeong, C. Kim., "A Study on Application Transfer of Development Rights(TDR) for Preservation of Historic Cultural Resources - Focusing on the Areas of Joong-gu, Daegu," Journal of the Urban Design Institute of Korea Urban Design, Vol. 17 No. 5, pp. 5-21, 2016.
- [6] "Valuation Method of Land Compensation Cost Included Question and Answer," Korea Development Institute PIMAC(Public and Private Infrastructure Investment Management Center), 2018.
- [7] G. Lee, K. Kim, "A Study on the Spatial Mismatch between the Assessed Land Value and Housing Market Price: Exploring the Scale Effect of the MAUP," Journal of the Korean Geographical Society, Vol. 48, No. 6, pp. 879-896, 2013.
- [8] Y. Kim, H. Kim, "A Policy Direction for the Readjustment of Officially Announced Standard Land Price," Journal of the Korean Urban Management

- Association, Vol. 19, No. 3, pp. 3-22, 2007.
- [9] S. Kim, K. Kwon, "A study on determinant factors for the realization rate of publicly notified individual land price," Journal of the Korean Cadastre Information Association, Vol. 15, No. 2, pp. 149-162, 2013.
- [10] D. Lee, S. Jung, "Research into the Effect of Jeju Olle Trails on Nearby Land Prices using Feasible Generalized Least Squares," Journal of Korea Real Estate Review, Vol. 24, No. 1, pp. 63-76, 2014.
- [11] H. Jung, "Opportunity Cost of Protected Area Designation for Conservation of Terrestrial Ecosystem," Master's Thesis of Seoul National University, 2017.
- [12] Y. Kang, H. Kim, "The Impact of Transit Facilities on Land Prices: Experiences at Seoul Station and Cheongnyangni Station," Journal of the Korean Urban Management Association, Vol. 25, No. 1, pp. 139-159, 2012.
- [13] Y. Ahn, M. Lee, "The Impact of Topographical Factors on Land Price of Dilapidated Residential Complex : The Case of Cheong-ju," Journal of the Architectural Institute of Korea Planning & Design, Vol. 24, No. 6, pp. 265-272, 2008.
- [14] I. Chung, "Improvement Direction of Facility Management System(FMS) through Related Systems Analysis," Korea Academy Industrial Cooperation Society, Vol. 16, No. 10, pp. 6751-6758, 2015.
DOI: <https://doi.org/10.5762/KAIS.2015.16.10.6751>
- [15] H. Ok, J. Kim, "A Research on Development Measures of Information Services for Construction Technology," Journal of the Korea Academia-Industrial cooperation Society, Vol. 16, No. 8, pp. 5707-5715, 2015.
DOI: <https://doi.org/10.5762/KAIS.2015.16.8.5707>
- [16] M. Seo, N. Kim, "Research about sharing of compensation data between owner and contractor for construction CALS improvement," The Korean Institute of Information Scientists and Engineers, Vol. 37, No. 1, pp. 144-147, 2010.
- [17] J. Kim, N. Kim, "The Status and Development Direction of Construction CALS," Proceedings of Korean Institute of Information Technology Conference, pp. 1-5, 2012.
- [18] S. Kim, B. Kim, T. Kim, N. Kim, "A Study on Project Information Integrated Management Measures Using Life Cycle Information in Road Construction Projects," Korea Academy Industrial Cooperation Society, Vol. 20, No. 11, pp. 208-216, 2019.
DOI: <https://doi.org/10.5762/KAIS.2019.20.11.208>
- [19] S. Yang, N. Kim, "A plan of improvement and DB analysis for the elevation of CALS System performance," The Korean Institute of Information Scientists and Engineers, Vol. 39, No. 1, pp. 31-33, 2012.
- [20] S. Kim, H. Ok, "A Study of Establishing the Development Strategy of Construction Project Management System Using SWOT Analysis," Korea Academy Industrial Cooperation Society, Vol. 13, No. 10, pp. 4837-4844, 2012.
DOI: <https://doi.org/10.5762/KAIS.2016.17.11.86>
- [21] H. Lee, J. Heo, D. Hwang, "Performance Comparison of Machine Learning Algorithms for Malware Detection," The Korean Society Of Computer And Information, Vol. 26, pp. 143-146, 2018.
- [22] J. Ha, H. Shin, Z. Lee, "Korean Text Classification Using Randomforest and XGBoost Focusing on Seoul Metropolitan Civil Complaint Data," The Korean journal of bigdata, Vol. 2, No. 2, pp. 95-104, 2017.
- [23] J. Hebert, "Predicting Rare Failure Event using Classification Trees on Large Scale Manufacturing Data with Complex Interactions," IEEE International Conference on Big Data, pp. 2024-2028, 2016.
- [24] T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016.
DOI: <https://doi.org/10.1145/2939672.2939785>
- [25] Y. Jia, S. Jin, P. Savi, Y. Gao, J. Tang, Y. Chen, W. Li, "GNSS-R Soil Moisture Retrieval Based on a XGboost Machine Learning Aided Method: Performance and Validation," Remote Sensing, Vol. 11, No. 14, pp. 1655, 2019.
DOI: <https://doi.org/10.3390/rs11141655>
- [26] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, Y. Si, "A data-driven design for fault detection of wind turbines using random forests and xgboost," IEEE Access, Vol. 6, pp. 21020-21031, 2018.
DOI: <https://doi.org/10.1109/ACCESS.2018.2818678>

이 상 규(Sang-Gyu Lee)

[정회원]



- 2015년 2월 : 서울대학교 경영대학원 경영정보학 (경영학석사)
- 2016년 12월 ~ 현재 : 한국건설기술연구원 전임연구원
- 2019년 3월 ~ 현재 : 성균관대학교 경영대학원 경영정보학 (경영학박사과정 중)

<관심분야>

Big Data, 데이터마이닝, 기술수용모델(TAM)

서 명 배(Myoung-Bae Seo)

[정회원]



- 2001년 2월 : 조선대학교 전자계산학과 졸업 (이학석사)
- 2019년 8월 : 세종대학교 컴퓨터공학과 박사수료
- 2003년 2월 ~ 현재 : 한국건설기술연구원 수석연구원

<관심분야>

BIM, 3D프린팅, 가상건설, 건설정보, 빅데이터, 인공지능

김 진 욱(Jin-Uk Kim)

[정회원]



- 1991년 2월 : 충남대학교 계산통계학과 (이학석사)
- 1999년 2월 : 충남대학교 컴퓨터과학과 박사 수료
- 1991년 10월 ~ 현재 : 한국건설기술연구원 연구위원

<관심분야>

Big Data, 데이터베이스, 정보검색