

## 서울시 데이터 기반 필지별 건축행위 발생 예측

\*윤성범      \*\*김태현\*

서울기술연구원

\*yunsb33@sit.re.kr      \*\*thkim@sit.re.kr

### Data-Driven Analysis for Future Construction Prediction : Case Study on Seoul

\*Sung-Bum Yun      \*\*Tae Hyun Kim\*

Seoul Institute of Technology

#### 요약

지속적인 건축물의 노화와 개발지 부족은 현존하는 건축물의 재건축 및 활용 가능 용지에 신규 건축행위를 유도한다. 서울에서는 근 5년간 25,000여 건의 신축이 발생하였으며, 이에 대한 신규 정책 등 다양한 지원 체계가 활성화되고 있다.

본 연구에서는 2011년부터 2015년까지 발생한 필지별 건축행위 데이터와 추가적 43개의 변수를 활용하여 신규 건축행위가 발생하는 필지에 대한 예측 모델을 구축하고자 한다. 요인도출 기계학습 방식인 의사결정트리 (Decision Tree) 중 CART (Classification And Regression Tree)를 활용하여 신규 건축 예측 모델을 구축하였으며, 86.28%의 정확도와 4개의 주요 신규 건축행위 발생 요인을 도출하였다.

#### 1. 서론

서울의 현 당면 과제 중 개발지 부족과, 현재 노후된 건축물의 재개발은 많은 관심을 받고 있다. 최근 5년간 서울시 내 신축은 25,000건 이상 발생하였으며, 이중 절반 이상은 주거용 건축물이 차지하고 있다. 현재 활성화 중인 도시재생 뉴딜 정책[1]의 경우 낙후지역에 대해 소규모 개발을 지속적으로 진행할 수 있도록 지원하고 있다. 이는 노후건축물, 저층건축물 밀집지역에 신축을 유도하고 있으며 해당 정책에 대한 관심이 커지고 있다.

신규 건축이 도시변화에 미치는 영향이 매우 강함에도 불구하고, 이와 관련된 연구는 대다수가 공간적 범위가 국지적이거나 실제 신축에 대해 다루지 못하는 한계점을 지니고 있다 [2, 3].

본 연구는 실제 서울시에서 제공하는 데이터를 활용하여 서울시 내부 소규모 필지 단위별 신규 건축이 발생한 지역에 대해 요인도출 기계학습 방식 ‘의사결정나무’ 알고리즘을 활용해 신규 건축을 유발하는 요인을 도출하고, 이에 대한 예측 모델을 구축하고자 한다.

#### 2. 연구 방법

연구의 시간적 범위는 2011년부터 2015년까지 발생한 건축행위를 대상으로 한다. 활용된 데이터의 경우 국토부 세움터에서 제공하는 건축 인허가대상 원자료를 가공하여 제작되었으며 이를 종속변수 (Target Variable)로 활용한다.

독립변수 (Input Variable)의 경우 필지 단위별 데이터는 2015년 기준 국토교통부에서 제공하는 연속지적도를 활용하여 구축하였으며, 건축물 특성 관련 자료의 경우 필지 내부에 존재하는 건축물대장과 과세 대상 정보 등을 활용하였다. 이와 함께 통계청에서 구축한 추가 통

계데이터를 활용하여 총 43개의 변수를 구축 및 활용하였다. 표 1에 본 연구에서 활용된 변수를 분류별로 표기하였다.

구분	요인	
건축행위 (종속변수)	건축행위 발생 여부	
필지 (6)	대지면적, 도로접면 수, 접도 최대도로 폭, 필지형상, 과소 필지 여부, 지가	
건축물 특성 (3)	노후도, 연면적, 용도	
생활편의 접근성	교통시설 (4)	버스정류장, 지하철역, 광역철도, 주차장까지 거리
	교육 시설 (5)	유치원, 초등학교, 중학교, 고등학교, 대학교까지의 거리
	녹지/하천 (4)	공원, 하천, 한강, 녹지까지의 거리
	편의시설 (6)	공공기관, 공공시설, 대형병원, 문화시설, 대규모 점포, 체육시설까지의 거리
주변 개발여건 (필지 주변 500m )	개발압력 (1)	당해 필지 건축행위 발생 직전 분기의 주변 지역 건축행위 수
	사회경제 여건 (4)	총인구수, 총종사자 수, 1인 가구 비율, 평균 생활인구 수
	물리적 여건 (10)	노후 건축물 비율, 과소 필지 비율, 부정형 필지 비율, 도로율, 버스정류장 수, 역세권 비율, 용적률, 평균지가, 접도 4m 미만 필지 비율, 주거비율

표 1 연구 활용 변수

필지 내부의 건축행위 여부를 예측하기 위해 의사결정트리 알고리즘 (Classification And Regression Tree)을 활용하였다 [4]. CART 알고리즘은 활용되는 독립변수들을 이용해 종속변수를 가장 잘 표현할 수 있는 트리 모델을 구축해주며, 결과로 나오는 트리의 경우, 해당 종속변수를 결정하는 가장 ‘영향력’ 있는 독립변수를 순서대로 표기해 준다. 이를 통해 실제 종속변수를 예측할 뿐만 아니라, 종속변수에 영향을 주는 독립변수를 순서대로 도출할 수 있다.

결과의 정확도를 평가하기 위해 N-fold cross validation 방식을 활용하였다 [5]. 본 연구에서는 데이터를 총 10개로 분류한다. 분류된 10개의 데이터 클러스터 중 9개는 모델의 학습을 위해 사용되고 1개의 데이터 클러스터는 모델의 정확도 평가에 활용된다. 위 프로세스를 총 10번 반복하여 모든 데이터를 학습데이터 / 평가데이터로 활용한 뒤 총 10개의 모델에서 도출된 결과를 활용하여 혼동행렬 (Confusion Matrix)를 구축하고, 모델의 정확도를 평가한다.

### 3. 결론

결과의 경우 총 10회의 Cross Validation 절차를 진행한 뒤, 모든 데이터를 Training 데이터로 활용하여 도출한다. 그림 1의 경우 CART 알고리즘에 의해 도출된 트리 모델로서, 해당 필지에 건축행위 발생 여부를 결정하는 요인을 중요도 순으로 표기한다.

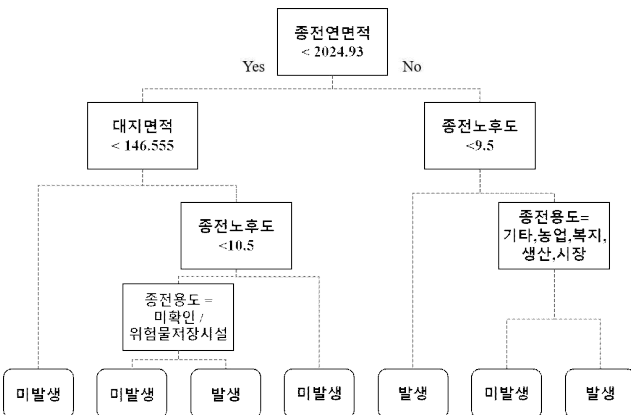


그림 1 의사결정 나무 결과

	미발생	발생	정밀도 Precision
미발생	62,059	10,288	85.78%
발생	625	6,533	91.27%
재현도 Recall	99.00%	38.84%	

표 2 혼동행렬 (Confusion Matrix)

표2의 경우 10-fold cross validation 절차를 통해 생성된 10개의 혼동행렬 (Confusion Matrix)의 평균치를 활용하여 제작한 Confusion Matrix이다. 해당 방식을 통해 도출된 모델의 정확도는 86.28%, 민감도는 0.8578 특이도는 0.9129이다. 해당 모델의 결과로서 필지의 기존 건축물의 연면적, 필지의 대지면적, 기존건축물의 노후도, 기존건축물의 용도 등이 해당 필지에 건축을 유발하는 주요 요인임을 확인할 수 있다. 이는 향후 필지별 신규 건축에 대한 선제적 도시계획에 활용할 수 있을 것이라 판단된다.

### ACKNOWLEDGEMENT

본 논문은 서울기술연구원(18-4-3, Digital 기반 도시공간 의사결정지원체계 연구)의 지원을 받아 수행된 연구임.

### 참 고 문 헌

- [1] 도시재생 뉴딜 로드맵(요약), <http://www.korea.kr/special/policyCurationView.do?newsId=148863980&pWise=mMain&pWiseMain=C4> (accessed Oct. 10, 2019)
- [2] 최주익, 2007, “서울시 일반주거지역 내 블록 단위 모델 개발연구”, *중앙대학교 대학원 석사학위 논문*
- [3] 김성보, 2017, “저층 주택지의 유형 구분 및 관리방안에 관한 연구:서울시 저층 주택지를 중심으로”, *홍익대학교 대학원 박사학위논문*.
- [4] Loh, Wei-Yin. "Classification and regression trees." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* Vol.1 No.1 pp. 14-23, 2011
- [5] Vasinek, Michal, Jan Plato, and Vaclav Snasel. "Limitations on low variance k-fold cross validation in learning set of rules inducers." *2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS)*. IEEE, 2016.