특집논문 (Special Paper)
방송공학회논문지 제25권 제2호, 2020년 3월 (JBE Vol. 25, No. 2, March 2020)
https://doi.org/10.5909/JBE.2020.25.2.176
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

서울 데이터 기반 필지별 용도전환 발생 예측

윤 성 범[a)], 문 성 철[a)], 박 순 용[a)], 김 태 현[a)‡]

Data-driven Analysis for Future Land-use Change Prediction : Case Study on Seoul

Sung Bum Yun[a)], Sungchul Mun[a)], Soon Yong Park[a)], and Taehyun Kim[a)‡]

요 약

지속적인 서울시의 발전과 쇠퇴에 따라 서울시는 정책 차원에서 도시재생을 진행하기 위해 지역별 용도전환 등의 정책을 진행하고 있지만, 이는 다양한 결과를 야기한다. 본 연구는 이런 용도전환이 발생하는 원인을 도출하고자 다양한 공공데이터를 활용하여 서울지역에서 지난 2011~2015년에 발생한 용도전환에 대한 예측 모델을 구축하고 용도전환을 야기하는 요인을 도출하고자 한다. 이를 구현하기 위해 서울시 및 국가 공공기관에서 취득한 서울시 필지에 대한 다양한 데이터를 의사결정 나무 기반 머신러닝 기법인 Random Forest에 적용하고 높은 정확도를 가지는 예측 모델을 구축하였으며, 용도전환을 야기하는 중요 요인들을 도출하였다. 해당 연구의 결과는 나아가 서울시의 당면 과제인 젠트리피케이션이 발생하는 요인연구와 예측 연구에 활용될 수 있을 것으로 판단되며, 공공의 정책 의사결정을 지원할 것으로 판단된다.

Abstract

Due to constant development and decline on Seoul areas the Seoul government is pushing various policies to regenerate declined Seoul areas. Theses various policies lead to land-use changes around numerous Seoul districts. This study aims to create prediction model which can foresee future land-use changes and while doing so, tried to derive various influential factors which leads to land-use changes. To do so, various open-data from national departments and Seoul government have been collected and implemented into random forest algorithm. The results showed promising accuracy and derived multiple influential factors which causes land-use changes around Seoul districts. The result of this study could further be implemented in policy makings for the public sectors, or could also be used as basis for studying gentrification problems happening in Seoul Area

Keyword : Land use policy, Prediction, Classification, Random Forest, Machine Learning

a) 서울기술연구원 스마트도시연구실(Seoul Institute of Techonolgy, Department of Smart City Research)
‡ Corresponding Author : 김태현(Taehyun Kim)
E-mail: thkim@sit.re.kr
Tel: +82-2-6912-0978
ORCID: https://orcid.org/0000-0001-9430-5072
※ 이 논문의 연구 결과 중 일부는 한국방송·미디어공학회 2019년 “추계학술대회”에서 발표한 바 있음.
※ 본 논문은 서울기술연구원(2020-AD-001, 서울 대도시권 데이터 사이언스 체계 구축방안)의 지원을 받아 수행된 연구임.
※ This research was funded by Seoul Institute of Technology (2020-AD-001, A study on the development of Data Science System in Seoul Metropolitan Area)
· Manuscript received December 27, 2019; Revised February 13, 2020; Accepted February 13, 2020.

Ⅰ. 서 론

1. 서론

서울의 현 당면 과제 중 개발지 부족과 현재 노후된 건축물의 재개발과 지역별 용도변경에 의한 신축 결정 여부는 많은 관심을 받고 있다. 이와 함께 5G 미디어 시대에 정보 전달의 절대량이 증가하며, 정보 수요자(시민)의 요구가 변화하고 있다. 수요가 높은 정보 중 하나로서, 다수의 시민이 미디어에 의지하고 있는 부동산 정보의 경우, 실제로 사용자가 원하는 방대한 양의 정보를 지속적으로 제공하기 위해선 실제 부동산에 대한 정보를 요구에 맞게 처리하고 분석하여 제공할 필요가 있다.

최근 5년간 서울시 내 용도변경에 의한 부동산 신축은 25,000건 이상 발생하였으며, 현재 활성화 중인 도시재생 뉴딜 정책의 경우 낙후지역에 대해 소규모 개발을 지속적으로 진행할 수 있도록 지원하고 있다. 이는 노후 건축물, 저층 건축물 밀집 지역에 신축을 유도하고 있으며 해당 정책 모델에 관한 관심이 커지고 있다. 그러나 이러한 낙후지역에 대한 지속적 소규모 개발은 지역적 개발을 일으키고 이를 통해 다수의 주거지역이 상업지역으로 변경되는 형태의 젠트리피케이션이 발생하고 있다[1].

2. 기존 연구

젠트리피케이션이 발생하는 사유에 관한 연구는 다방면으로 진행되고 있으며[2,3], 이러한 젠트리피케이션을 완화하고 나아가 해결하는 방안에 관한 연구도 진행되고 있다[4,5]. 이러한 도시문제를 완화하고 해결하고자 하는 기존 연구들은 대다수 회귀분석 방식에 의존하여, 해당 문제의 발생 원인을 조사하고자 함에 머물러있으며, 최근 활성화 되는 머신러닝 기법에 대한 활용도는 제한적인 수준이다.

도심지역의 젠트리피케이션 사유를 머신러닝 기법을 통해 분석하고자 하는 노력은 최근에서 진행되고 있으며[6,7,8] 실제 통계기반 머신러닝 기법을 사용하여 서울지역의 상업 젠트리피케이션 영향요인을 도출하고자 하는 방안도 연구되었다[9]. 그러나 대다수 연구의 경우 주로 상업지역에 대하여 상업과 관련된 요인들을 사용하여 요인 분석을 진행하고자 하였으며, 이는 젠트리피케이션을 유발하는 다양한 요인을 판단하기 어렵고, 이를 통해 수요자가 원하는 정보를 전달하지 못한다는 한계점을 지니고 있다.

실제 젠트리피케이션이 발생하는 과정에서는 기존 지역의 필지별 용도가 전환되는 경우가 발생하며[10], 이를 예측하는 모델을 구축하는 연구는 상대적으로 많이 진행되지 못하고 있다. 실제로 뉴욕을 대상으로 젠트리피케이션과 지역의 용도전환이 발생하는 상황에 관한 조사 연구가 진행되고 있으며[11], 실제 서울지역에 대한 예측을 진행하기 위해선 용도전환에 대한 예측과 이를 설명하는 중요 요인을 도출하는 연구가 필요하다.

3. 연구 목적

본 연구는 실제 공공기관에서 제공되는 서울시에 대한 공공데이터를 활용하여 서울시 내부 소규모 필지 단위별 용도전환이 발생한 지역에 대해 용도 변환과 유지를 일으키는 요인을 도출하고자 하며, 예측하고자 하는 종속요인과 연관성이 있는 독립변수들을 미리 선별 후 분석을 진행하는 기존 회귀분석 방식과는 다르게, 서울시 필지별 요인을 다수 활용하여 다양한 요인을 입력 변수로 구축한 뒤, 머신러닝 기반 알고리즘을 통해 스스로 연관성이 있는 요인들을 도출하고, 해당 요인들의 중요도를 판단하여 요인별 '중요도'를 판별하고자 한다. 해당 방식의 Bottom-Up 방식 분석을 수행하기 위해 트리 모델 기반 기계학습 방식인 'Random Forest' 알고리즘을 활용해 용도전환을 유발하는 요인을 도출하고, 이에 대한 예측 모델을 구축하고자 한다. 기존에 토지 이용변화에 대한 모델을 구축할 때 자주 활용되지 않은 머신러닝 기법을 활용해 토지이용변화 분류 및 예측을 할 수 있는지 판단하고자 한다. 용도전환에 대한 높은 이해도와 분석은 향후 수요자가 원하는 미디어 정보를 제공할 때 기반으로 활용될 수 있으리라 판단된다.

Ⅱ. 연구방법

1. 데이터

연구의 시간적 범위는 2011년부터 2015년까지이며, 서

표 1. 연구 활용 변수
Table 1. Data List

Variable	Categories		Factors
Target Variable	Land use change (1)		Land use change (Changed / Unchanged)
Input Variable	Parcel related variables (6)		Parcel area, Road connection, Max road width, Parcel shape, Parcel area regulation(under 90m2), Land Price
	Building related variables (3)		Deterioration, Area. Usage
	Accessibility	Mobility (5)	Bus station, Subway station, Train station, Parking lot
		Education (7)	Distance to : Kindergarten, Primary school, Secondary school, High school, University
		Green-land (5)	Distance to : Park, River, Han-river, Green-land
		Convenience (7)	Distance to Public facilities, Public area, Hospital, Cultural space, Large stores, Gym
	Development factors (Around 500m of the parcel)	Development pressure (1)	Development/construction conducted on adjacent areas to the parcel
		Socio-economic factor (6)	Population, Total employees, Single-person household, Average floating population
		Physical conditions (10)	Number of Deteriorated buildings, Regulated parcel ratio, Un- determinate parcel ratio, Road coverage, Number of bus stations, Near station ratio, Paved area ratio, Average land price, Under 4m road connection parcel ratio, Household ratio

울시 내에서 발생한 필지별 용도변경 행위를 대상으로 한다. 활용된 데이터의 경우 국토부 세움터에서 제공하는 건축 인허가대장 원자료를 가공하여 제작되었으며 이를 종속변수(Target Variable)로 활용 한다.

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 대장 정보 등을 활용하였다. 이와 함께 통계청에서 구축한 추가 통계데이터를 활용하여 총 50개의 독립 변수를 구축 및 활용하였다. 표 1에 본 연구에서 활용된 변수를 분류별로 표기하였다.

2. 데이터 통계

본 연구에서 활용된 데이터는 서울시에 존재하는 총 795,268개의 필지에 대한 정보를 보유하고 있다. 이 중 본

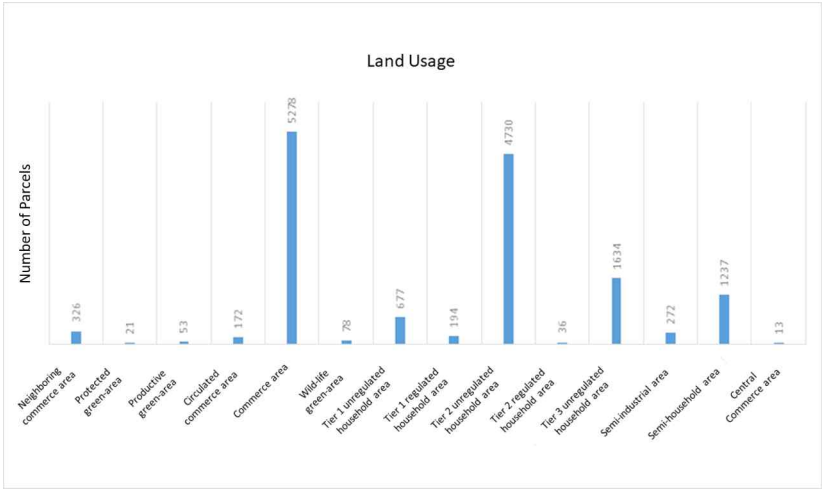

그림 1. 용도 전환 된 용도 지역 분포
Fig. 1. Distribution of Land usage

연구에서 활용될 종속변수인 '용도전환' 관련 값이 존재하지 않는 79,525개의 필지를 제외한 뒤 총 715,743개의 필지에 대한 분석을 진행하였다.

해당 데이터 중 용도 지역을 표현하는 컬럼의 경우 제1종 일반주거지역, 일반상업지역, 자연녹지지역 등 총 14종의 해당 필지의 지정 용도를 보유하고 있으며, 필지의 용도가 전환되기 전 해당 필지가 활용되었던 '종전용도'의 경우 근린생활시설, 주거시설, 사무실, 생산시설 등 총 28개의 분류로 구분되어있다.

총 715,743개의 필지 중 2016년 기준 용도전환이 진행된 필지는 14,721개로 전체의 2.1%를 차지하며, 전환이 진행된 필지의 기존 용도 분포는 그림 1과 같다.

용도전환이 된 총 14,721개의 필지 용도 중 가장 큰 비중을 차지하는 용도는 일반상업지역이며, 이후 제2종 일반주거지역, 제 3종 일반주거지역 순으로 용도전환이 진행되었다.

3. Random Forest 알고리즘

필지 내부의 건축행위 여부를 예측하기 위해 Random Forest 알고리즘을 활용하였다[12,13]. Random Forest 알고리즘은 활용되는 독립변수들을 이용해 종속변수를 가장 잘 표현할 수 있는 모델을 구축해주는 과정에서 기존 의사결정 나무(Decision Tree)알고리즘과 유사하다.

Random Forest의 경우 기존 의사결정 나무알고리즘의 한계점을 해결하고자 랜덤노드 최적화 (Randomized Node Optimization)과 배깅 (Bootstrap AGGregating, Bagging) 방식을 활용하였으며, 이를 통해 상관관계가 존재하지 않는 개별적 트리형식의 모델을 구축하고, 모든 트리에 대하여 학습 과정을 거쳐 가장 종속변수에 가장 큰 영향을 주는 독립변수들을 순차적으로 도출한다[14].

Random Forest 알고리즘은 생성되는 모델/트리를 임의(Random)로 그리고 비 상관적(Uncorrelated)으로 생성하게 되며, 이를 통해 일반화된 결과를 표출하게 된다. 이는, 모델 구축에 있어 변수 선택을 임의로 진행하는 Random Forest의 특징이며, 이를 통해 구축된 데이터 자체에 불필요 데이터(노이즈)가 포함된 데이터를 다룰 때 강점을 나타낸다. 배깅 방식을 통해 트리 모델을 구축하는 Random Forest는 지속적으로 트리 모델을 구축하며 변수를 선택/제외 시키는 과정을 진행하게 된다. 이 과정에서 불필요 변수가 포함된 모델은 필연적으로 낮은 모델 정확도를 보이게 되고, 해당 변수는 중요 변수에서 제외되고, 실제 종속변수를 가장 잘 표현하는 독립변수들이 포함된 트리 모델은 높은 정확도를 보이게 되며 중요 변수로 지정된다.

또한, 배깅 과정에서 임의적으로 변수를 선택하여 모델을 구축함에 있어, 모델간의 분산(Variance)은 감소시키고 편향성(Bias)은 증가시키게 되어, 기존 의사결정 나무가 가지는 과적합(Over fitting)의 문제를 해결할 수 있다.

Random Forest는 사용자가 생성되는 트리 모델의 개수를 설정할 수 있다. Random Forest에서 생성된 다수의 트리 모델의 결과를 결합하여 실제 중요 요인을 도출하는 '앙상블 방식(Ensemble method)'을 활용하고 있다. 이를 통해 실제 종속변수를 예측하는 모델을 구축할 뿐만 아니라, 종속변수에 영향을 주는 독립변수를 순서대로 도출할 수 있다는 장점을 지니고 있다.

본 연구에서는 총 5,000개의 임의적 트리 모델을 구축하도록 설정하였다.

본 연구를 위해 수집된 총 51개의 변수는 Random Forest 알고리즘에 활용되기 위하여 Target Variable (종속변수)와 Input Variable(독립변수)로 구분되었다. Random Forest 머신러닝 방법을 통해 필지별 용도 전환 유무를 판단하고자 하는 본 연구의 목적에 따라 필지용도 전환 변수를 종속변수로 활용하였다. 나머지 50개의 변수의 경우 종속변수와의 연관관계 및 중요 변수 여부를 판단하기 위해 Random Forest의 독립 변수로 활용되었다.

활용된 독립변수의 경우 배깅 방식을 통해 임의적으로 추출되어 트리 모델 구성에 활용되었으며 이를 통해 종속변수인 필지별 용도 전환 유무를 가장 잘 표현하는 트리 모델을 구성하도록 설정 되었다. 이 후 생성된 5,000개의 임의의 트리 모델 중 종속 변수를 가장 잘 표현해주는 '중요 변수'가 중요도 순으로 추출되며, 이러한 '중요 변수'를 활용하여 종속 변수를 가장 잘 표현하는 트리 모델의 정확도를 평가하게 된다.

4. 정확도 및 통계적 일치성 평가

정확도 평가를 위하여 기존 데이터를 사용하여 구축된

Random Forest 모델을 평가하였고, 이를 Confusion Matrix (혼동행렬)로 표기하여 정확도 값을 평가하였다.

결과의 정확도를 평가하기 위해 N-fold cross validation 방식을 활용하였다[15]. 본 연구에서는 Random Forest에 적용될 데이터를 총 10개로 분류하도록 설정되었다. 분류된 10개의 데이터 클러스터 중 9개는 모델의 학습을 위해 사용되고 1개의 데이터 클러스터는 모델의 정확도 평가에 활용된다. 위 프로세스를 생성되는 5,000개의 트리에 적용하여 한 개의 트리를 구축하기 위해 선택된 데이터를 모두 학습 데이터 / 평가데이터로 활용한 뒤 총 10개의 모델에서 도출된 결과를 활용하여 혼동행렬 (Confusion Matrix)를 구축하고, 개별 트리 모델의 정확도를 평가한다.

Random Forest 모델을 구축하는 과정에서 전체 데이터를 10개로 나눈 뒤 이 중 9분할의 데이터를 Training data로, 1분할의 데이터를 Test data로 활용할 수 있도록 10-fold Cross Validation 방식을 활용하였으며, 이를 통해 생성되는 모든 트리 모델의 정확도를 평가하였다.

Cohen's Kappa 통계치[16]는 카테고리 분류 분석이 진행될 때 분류 모델간의 일치도를 판단하기 위해 활용되는 지수다. Kappa 통계치는 방정식 1. 과같이 연산된다.

$$\kappa = \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_0}{1 - p_e} \quad (1)$$

κ는 kappa 지수를 의미하여, p_0는 분류 모델 결과가 정확할 경우이며, 이는 모델의 일반 정확도와 같은 의미를 가진다. p_e의 경우 분류 모델 결과가 확률적으로 일치할 경우이다.

Kappa 통계지수가 0에 가까울수록 분석모델의 일치성이 낮다는 의미이며, 통계적으로 유의하지 않다는 의미로 볼 수 있다. 반대로 지수가 1에 가까울수록 분류 모델의 일치성이 높다는 의미로 판단할 수 있다.

Ⅲ. 결 과

Random Forest로 구축된 총 5,000개의 임의적 트리 모델

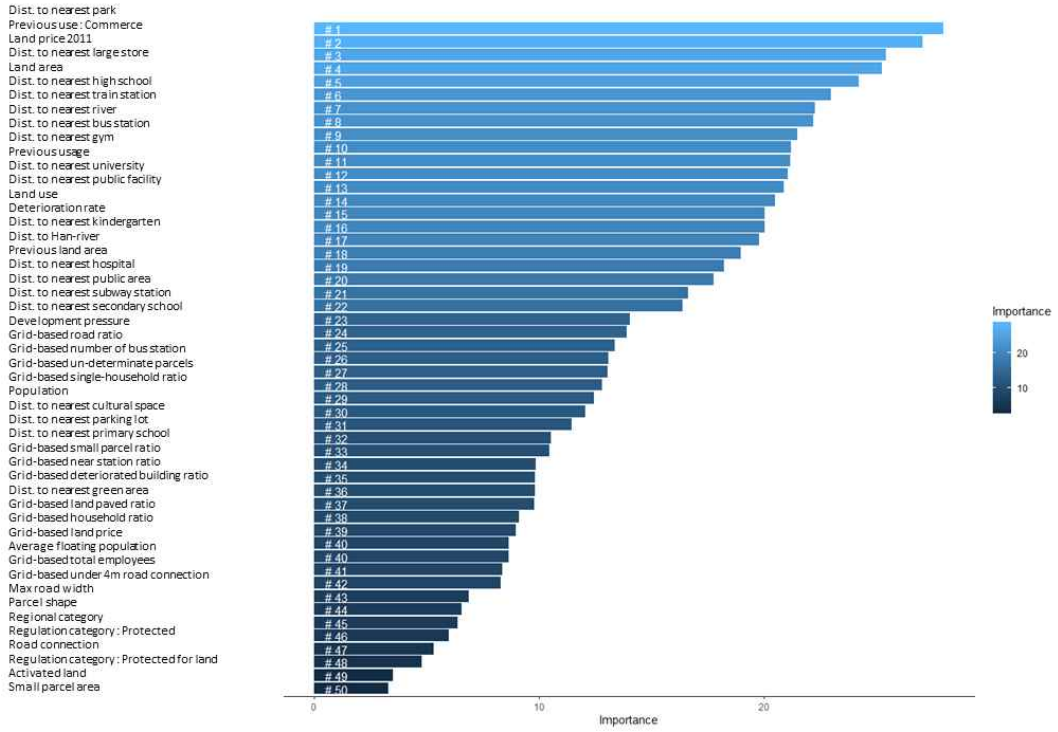

그림 2. 랜덤포레스트 결괏값 (Mean Decrease Accuracy)
Fig. 2. Random Forest Result (Mean Decrease Accuracy)

을 활용하여 용도전환에 가장 큰 영향을 미치는 요인들을 도출하였다. 그림 2는 Mean Decrease Accuracy 방식과 Mean Decrease Gini 방식을 활용하여 주요 요인들을 도출한 결과를 표현한다.

Mean Decrease Accuracy(MDA) 방식은 구축된 트리 모델에서 특정 변수를 제거했을 시 정확도가 줄어드는 척도를 활용하여 중요 요인을 도출하는 방식으로서, 종속변수에 큰 영향을 미치는 요인일수록 제거된 후 트리 모델이 생성되었을 시 정확도가 크게 떨어진다는 점을 활용한 방식이다. Mean Decrease Gini(MDG) 의 경우 트리 모델을 구축하는 나뭇가지가 생성될 때 선택되는 변수들이 불순도를 감소시키는 양을 활용하여 트리 모델상에서 중요한 요인을 도출하는 방식이다[17]. 본 연구에서 사용된 'Positive' Class는 필지의 용도전환 여부가 '미전환'인 변수이며, 이는 도출된 주요 결과 요인이 필지의 용도전환이 발생하지 않도록 영향을 준다고 판단될 수 있다.

MDG 방식의 경우 모델이 구축되며 '전환' / '미전환' 분류 값에 대한 불순도가 가장 낮은 요인들을 추출하지만, 본 연구에서 활용된 데이터셋의 '전환' / '미전환' 분류의 개수 차이가 크기에, 불순도 분류는 상대적으로 분석모델의 유의성을 저하한다고 판단할 수 있다[18]. 이러한 사유로 본 연구에서는 MDA 방식을 통해 추출된 중요 요인들을 채택하고자 한다.

그림 3은 모든 50개의 독립변수들이 종속변수에 영향을 미치는 순서대로 작성한 그래프로서, 용도전환이 발생하지 않도록 하는 중요 요인들이 '최인접 공원까지의 거리', '종전용도 상업지역', '지가', '최인접 대규모 점포까지의 거리', '대지면적' 등임을 확인할 수 있다.

반대로 용도전환이 발생('Negative' Class)하도록 하는 요인의 경우 해당 그래프 하단에서 확인할 수 있으며 이는 '과소 필지 여부', '활성화 지역 여부', '경관지구 여부', '도로접면 수', '미관지구 여부' 등이 있으며, 이와 유사한 사유로서, '개발압력', '노후 건축물 비율' 등이 추가로 용도전환을 유발하는 사유로 확인할 수 있다.

분석정확도의 경우, Random Forest 알고리즘의 특성상 트리 모델의 개수가 높아질수록 오차율이 낮아지는 경향을

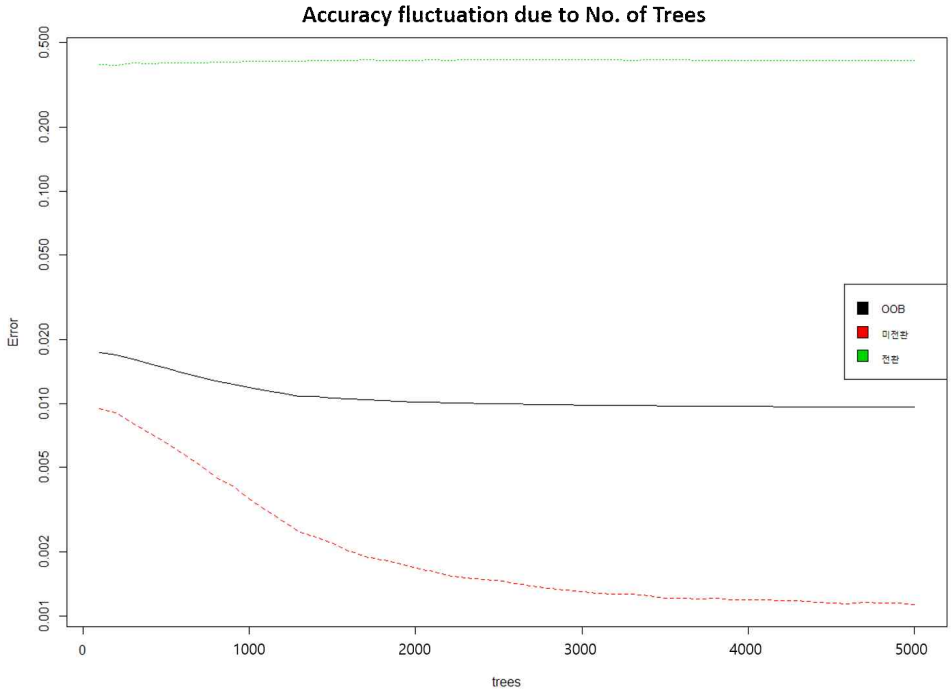

그림 3. Random Forest 트리 모델 개수별 정확도 변화
Fig. 3. Accuracy Fluctuation due to number of Trees

보인다. 그림 3에서는 트리 모델이 1개에서 5000개까지 생성되며 추출되는 중요 요인으로 인해 전체 모델의 오차율이 감소함을 볼 수 있다.

표 2. 혼동행렬
Table 2. Confusion Matrix

	Unchaged	Changed	Precision
Unchaged	77,810 (TP)	662 (FP)	99.16%
Changed	91 (FN)	962 (TN)	91.36%
Recall	99.88%	59.24%	

표 2의 경우 10-fold cross validation을 활용한 Random Forest 모델의 Confusion Matrix이다. 이를 활용하여 정확도와 균형정확도 (Balanced Accuracy)를 연산하고, 모델의 통계적 일치도를 판단 하기 위해 Cohen's Kappa 계수를 도출한다.

표 3. 결과 테이블
Table 3. Result Table

Accuracy	Balanced Accuracy	Kappa Statistics	Elapsed Time
99.05%	79.56%	0.7141	83,353sec (23.11 hrs)

결과로 도출된 일반 정확도의 경우 99.05%로서 매우 높은 정확도를 보인다. 그러나 표 3에서 확인할 수 있듯, 데이터 자체의 용도변경 발생 횟수와 미발생 횟수의 차이가 크다는 한계점이 있으며, 그림 3에서 확인 할 수 있듯 미전환에 대한 오차율은 트리 모델의 개수가 증가하며 지속적으로 감소하지만, 전환에 대한 오차율은 트리 모델의 개수가 증가해도 큰 차이를 보이지 않고 있다. 상대적으로 적은 '전환' 표본에 대한 편차를 보정 하기 위해 정확도를 균형정확도 (Balanced Accuracy)로 연산할 시 79.56%의 모델 정확도를 보인다[19,20].

Kappa 통계치의 경우 0.7141로 도출되었으며 이는 '상당한(Substantial) 일치도'[21] 이며 해당 분석이 통계적으로도 유의함을 나타낸다.

Ⅳ. 결 론

본 연구는 공공기관에서 제공하는 서울시에 대한 다양한 공공데이터를 활용하여 서울시 용도전환을 야기하거나 방해하는 중요 요인을 도출하였다. 해당 연구를 진행하기 위해 Random Forest 알고리즘을 활용하였으며, 이를 통해 높은 정확도를 가진 예측 모델을 구축하였다.

용도전환을 야기하는 중요 요인 중 '활성화 지역 여부'의 경우 현재 도시재생활성화지역 지정 여부를 의미하며 이는 기존 용도에서 타 용도로 전환되는 사유 중 큰 요인으로 작용한다. 기존 도시재생 활성화 지역의 경우 낙원상가 일대, 세운상가 일대, 신촌, 서울역 등이 있었으며, 현재 도시재생 활성화 지역의 경우 영등포, 용산, 서울역 등의 지역이 도시재생 활성화 지역으로 지정되어있다. '도로접면 수'의 경우 도로를 기반으로 개별 필지에 접도 해 있는 면수를 필지별로 구축한 자료로서, 해당 지역에 대한 접근성을 나타내는 요인으로 판단할 수 있으며, 이는 향후 해당 지역에 유입되는 인구가 많아질 수 있음을 표현한다고 볼 수 있다. '경관지구' 및 '미관지구'의 경우 경관이나 미관의 관리를 위해 건축물 높이를 규제하는 등 다양한 제한사항이 존재하는 지역을 의미하지만, 이는 반대로 기존에 존재하는 다양한 종류의 건축물을 상업적 용도로 변경하게 되는 사유로 적용될 수 있으며, 실제 경관/미관지구로 지정되어있는 서울 서촌지역의 도시재생 정책이 진행되며, 젠트리피케이션 현상이 발생하고 있는 현상과 일치한다고 볼 수 있다[23,24]. 반대로 용도 미전환 지역의 중요 요소로서 도출된 '최인접 공원까지의 거리'의 경우 실제 주거지역의 특성으로 볼 수 있으며, 해당 지역은 주거지역으로 유지되는 현상으로 판단할 수 있다. '종전용도 - 상업지역' 요인의 경우 현재 용도가 이미 상업지역이기에 전환이 필요 없다는 이유로 볼 수 있으며, '지가', '최인접 대규모 점포까지의 거리', '대지면적' 등은 상업적 요인으로서 용도전환을 방해하는 요소로 판단된다.

본 연구에서는 실제 젠트리피케이션 관련 요인을 변수로 활용하지 않았다는 한계점을 지니며, 또한 Random Forest 뿐만 아닌 다양한 머신러닝 알고리즘과 현재 화두 되는 딥러닝 알고리즘을 활용하지 않았다는 한계점을 지니고 있다.

젠트리피케이션과 관련된 요소의 경우 실제 상업용 매장

의 영업신고내역을 활용하여 실제 상업매장의 개점과 폐점 내역을 변수로 사용할 수 있을 것으로 판단되며 향수 연구에서 이에 관한 내용을 다루고자 한다.

추가적으로 본 연구에서는 용도전환의 예측과 함께 용도전환을 야기하는 중요 요인을 도출하기 위하여 모델 정확도가 높으며 사건을 야기하는 중요 요인을 도출할 수 있는 Random Forest 알고리즘을 활용하였으나, 향후 딥러닝과 같은 더 높은 차원의 분석 알고리즘을 통해 모델의 정확성을 향상하고자 한다.

참 고 문 헌 (References)

[1] D. Kim, K. Kim, G. Kim, “The Impact of Commercialization-induced Gentrification on Poor UIrban Neighborhoods: A case study of Dongja-dong Jjok-bank district”, Seoul Studies, Vol. 18, No 2, pp 159-175, 2016.

[2] HONG, Shijian, and Xianchun ZHANG. "Rent gap, gentrification and urban redevelopment: The reproduction of urban space driven by capital and right." Urban Development Studies Vol 3, No 15, 2016

[3] HELBRECHT, Ilse. Gentrification and displacement. Gentrification and Resistance. Springer VS, Wiesbaden, p. 1-7. 2018.

[4] Ferm, Jessica. "Preventing the displacement of small businesses through commercial gentrification: are affordable workspace policies the solution?." Planning Practice & Research Vol 31. No 4, pp.402-419. 2016

[5] Ghaffari, L., Klein, J. L., & Angulo Baudin, W. Toward a socially acceptable gentrification: A review of strategies and practices against displacement. Geography Compass, Vol 12, No 2, pp: e12355. 2018.

[6] Reades, Jonathan, Jordan De Souza, and Phil Hubbard. "Understanding urban gentrification through machine learning." Urban Studies 56.5 (2019): 922-942.

[7] Knorr, David. Using Machine Learning to Identify and Predict Gentrification in Nashville, Tennessee. Diss. Vanderbilt University, 2019.

[8] ALEJANDRO, Yesenia; PALAFOX, Leon. Gentrification Prediction Using Machine Learning. In: Mexican International Conference on Artificial Intelligence. Springer, Cham, pp. 187-199. 2019

[9] Kim, Gyoung-Sun, and Dong-Sup Kim. "The Study on the Influential Factors on Commercial Gentrification in Seoul." The Journal of the Korea Contents Association Vol 19. No 2 pp. 340-348. 2019

[10] Seifolddini, Faranak, and Michael Harris. "Incentive-based Land Use Policies and Strategies for Land Acquisition in Gentrification Process." International Journal of Physical and Social Sciences Vol 6. No 2 pp. 64-83. 2016

[11] McCabe, Brian J. "Protecting Neighborhoods or Priming Them for Gentrification? Historic Preservation, Housing, and Neighborhood Change." Housing Policy Debate Vol 29. No 1 pp. 181-183. 2019

[12] Loh, Wei Yin. "Classification and regression trees." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery Vol.1 No.1 pp. 14-23, 2011

[13] Y. Eun, “Random Forest“ Journal of Educational Evaluation, Vol 28, pp. 427-448, 2015

[14] Leo Breiman, “Random Forests”, Machine Learning, Vol.45, No.1, pp.5‐32, 2001 doi:10.1023/A:1010933404324

[15] 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.

[16] Ben-David, A. (2008). Comparison of classification accuracy using Cohen’s Weighted Kappa. Expert Systems with Applications, 34(2), 825-832.

[17] Han, Hong, Xiaoling Guo, and Hua Yu. "Variable selection using mean decrease accuracy and mean decrease gini based on random forest." 2016 7th ieee international conference on software engineering and service science (icsess). IEEE, 2016.

[18] CALLE, M. Luz; URREA, Víctor. Letter to the editor: stability of random forest importance measures. Briefings in bioinformatics, 2010, 12.1: 86-89.

[19] Brodersen, Kay Henning, et al. "The balanced accuracy and its posterior distribution." 2010 20th International Conference on Pattern Recognition. IEEE, 2010.

[20] García, Vicente, Ramón Alberto Mollineda, and José Salvador Sánchez. "Index of balanced accuracy: A performance measure for skewed class distributions." Iberian conference on pattern recognition and image analysis. Springer, Berlin, Heidelberg, 2009.

[21] LANDIS, J. Richard; KOCH, Gary G. The measurement of observer agreement for categorical data. biometrics, 159-174, 1977.

[22] H. Lim, Future Development Directions through the Evaluation of the Policies in Seochon, Soeul, The Seoul Institute, 2012

[23] H. Doh, B. Byun., A Study of the Factor Analysis about the Gentrification of Seo-Chon in Seoul. The Geographical Journal of Korea, Vol 51 No. 3, pp.311-322. 2017

저 자 소 개

윤 성 범

- 2016년 2월 : 연세대학교 토목환경공학과 공학사
- 2019년 2월 : 연세대학교 토목환경공학과 공간정보컴퓨팅 공학석사
- 2019년 8월 ~ 현재 : 서울기술연구원 스마트도시연구실 위촉연구원
- ORCID : https://orcid.org/0000-0001-6813-3692
- 주관심분야 : GIS, 공간빅데이터, 머신러닝, AI

문 성 철

- 2015년 2월 : 한국과학기술연구원(UST) HCI & Robotics 공학박사
- 2015년 3월 ~ 2016년 11월 : 한국과학기술연구원 국가기반기술연구본부 박사후연구원
- 2016년 12월 ~ 2017년 6월 : 골프존뉴딘그룹 전략사업실 책임연구원
- 2017년 7월 ~ 2019년 6월 : CJ Hello Future Engine Lab. 부장
- 2019년 6월 ~ 현재 : 서울기술연구원 스마트도시연구실 연구위원
- ORCID : https://orcid.org/0000-0003-4596-9889
- 주관심분야 :Digital Healthcare, Human Factors, HCI, Deep Learning

박 순 용

- 2000년 2월 : 단국대학교 토목공학과 공학사
- 2004년 8월 : 단국대학교 토목공학과 교통공학전공 공학석사
- 2008년 9월 ~ 2013년 5월 : 단국대학교 토목환경공학과 연구원
- 2013년 2월 : 단국대학교 토목환경공학과 교통공학전공 공학박사
- 2013년 6월 ~ 2015년 7월 : 한국건설기술연구원 도로연구소 박사후연구원
- 2015년 8월 ~ 2019년 6월 : 도로교통공단 교통과학연구원 선임연구원
- 2019년 7월 ~ 현재 : 서울기술연구원 스마트도시연구실 수석연구원
- ORCID : https://orcid.org/0000-0002-0434-1836
- 주관심분야 : 교통공학, 교통운영, 교통신호제어, 자율주행

김 태 현

- 2004년 6월 ~ 2010년 6월 : 서울시청 도시계획·정책연구위원 도시계획국
- 2010년 6월 ~ 2018년 6월 : 서울연구원 연구위원 도시공간연구실
- 2018년 6월 ~ 현재 : 서울기술연구원 선임연구위원 스마트도시연구실장
- ORCID : https://orcid.org/0000-0001-9430-5072
- 주관심분야 : 스마트 도시계획 및 정책개발, 빅데이터 기반 도시진단, NLP기반 법률 추론, AI 기반 전문가 시스템