트리모델을 이용한 서울 내 주택가격 예측

주 택 의 지 역 적 , 개 별 적 특 성 을 중 심 으 로

20190552 손지영, 20201365 최세환

CONTENTS

01

연구 설명

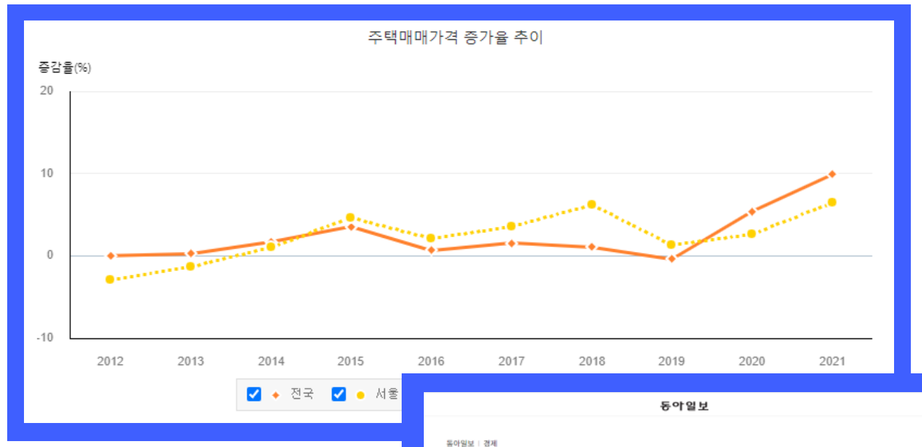
01. 주제 선정의 이유
02. 선행연구 검토
03. 연구 프로세스

01 연구 배경

주제선정 이유

주택의 지역, 특성을 이용한 가격예측

- 서울내 집값은 2019년 부터 급격히 상승
- 정부또한 부동산 시장 안정을 위해 주택 가격을 예측
- 주택 자체의 특성을 고려한 가격예측이 필요




동아일보

동아일보 | 경제

무용지물로 내몰릴 처지에 놓인 2022년 주택 공시가격

황재성기자

입력 2022-04-29 12:50 | 업데이트 2022-04-29 14:09

가

많이 본 뉴스

1 尹대통령 "김의겸-장경태 의혹, 금방 밝혀질 거짓말"

2 中 상하이서 '시진핑 퇴진, 공산당 회복' 구호 나왔다

3 '부산 모녀 사망' 유력 용의자로 끝은 빌라 이웃 여성 구속

4 농가율 최대 80mm 비…주 중반부터 '북극한파'

5 '술자리 의혹 제기' 더탐사, 한동훈 아파트까지 찾아가

6 월드컵 관중석에 나타난 욱일기[청계천 옆 사진관]

7 사업 40년 만에 '위놈은 부고'… 이제서야 그녀를 주목하는 이유는

8 대통령실, 튀 이상민 파면 '최후통첩'에 "국조 못 받아"

01 연구 배경

01. 주제선정 이유

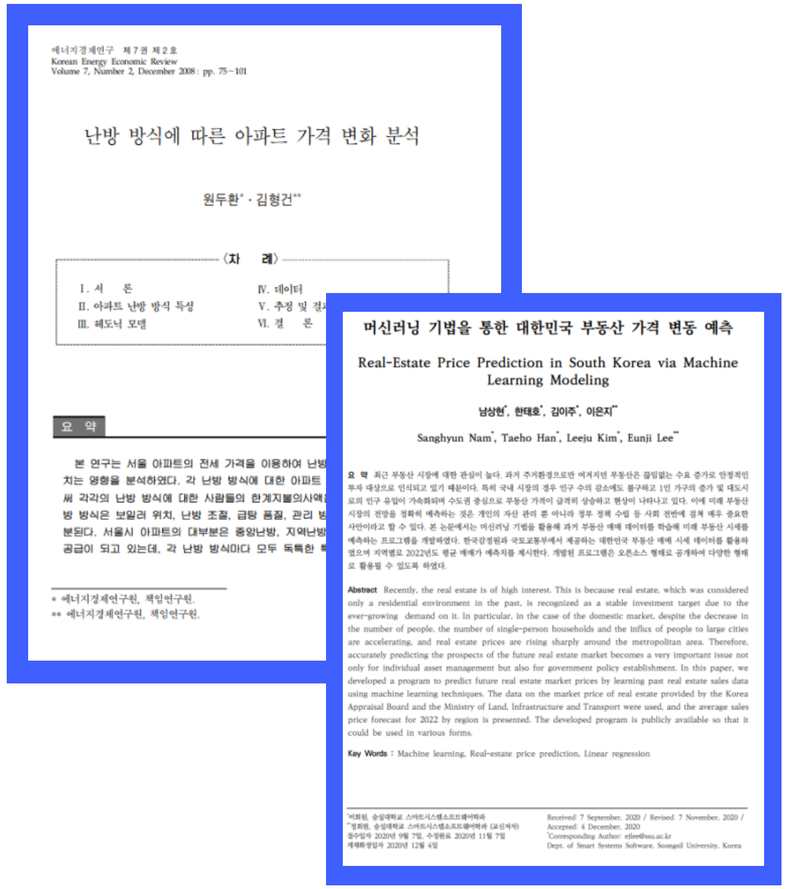
02. 선행연구 검토

03. 연구 프로세스

선행연구 검토

본 연구에 앞선 선행연구 확인

- 머신러닝을 통한 집값예측 사례들의 존재
- 가격에 영향을 미치는 요인 분석의 사례 존재
- 주택 자체의 특성을 종합한 예측이 가능하다 판단



에너지경제연구 제7권 제2호
Korean Energy Economic Review
Volume 7, Number 2, December 2008 : pp. 75~101

난방 방식에 따른 아파트 가격 변화 분석

원두환* · 김형건**

〈차 례〉

Ⅰ. 서 론
Ⅱ. 아파트 난방 방식 특성
Ⅲ. 헤도닉 모델
Ⅳ. 데이터
Ⅴ. 추정 및 결
Ⅵ. 결 론

요 약

본 연구는 서울 아파트의 전세 가격을 이용하여 난방
치는 영향을 분석하였다. 각 난방 방식에 대한 아파트
써 각각의 난방 방식에 대한 사람들의 한계지불의사액
방 방식은 보일러 위치, 난방 조절, 급탕 품질, 관리 보
분된다. 서울시 아파트의 대부분은 중앙난방, 지역난방
공급이 되고 있는데, 각 난방 방식마다 모두 독특한 특

* 에너지경제연구원, 책임연구원.
** 에너지경제연구원, 책임연구원.

머신러닝 기법을 통한 대한민국 부동산 가격 변동 예측

Real-Estate Price Prediction in South Korea via Machine Learning Modeling

남상현*, 한태호*, 김이주*, 이은지**

Sanghyun Nam*, Taeho Han*, Leeju Kim*, Eunji Lee**

요 약 최근 부동산 시장에 대한 관심이 높다. 과거 주거환경으로만 여겨지던 부동산은 끊임없는 수요 증가로 안정적인 투자 대상으로 인식되고 있기 때문이다. 특히 국내 시장의 경우 인구 수의 감소에도 불구하고 1인 가구의 증가 및 대도시로의 인구 유입이 가속화되며 수도권 중심으로 부동산 가격이 급격히 상승하고 현상이 나타나고 있다. 이에 미래 부동산 시장의 전망을 정확히 예측하는 것은 개인의 자산 관리 뿐 아니라 정부 정책 수립 등 사회 전반에 걸쳐 매우 중요한 사안이라고 할 수 있다. 본 논문에서는 머신러닝 기법을 활용해 과거 부동산 매매 데이터를 학습해 미래 부동산 시세를 예측하는 프로그램을 개발하였다. 한국감정원과 국토교통부에서 제공하는 대한민국 부동산 매매 시세 데이터를 활용하였으며 지역별로 2022년도 평균 매매가 예측치를 제시한다. 개발된 프로그램은 오픈소스 형태로 공개하여 다양한 형태로 활용될 수 있도록 하였다.

Abstract Recently, the real estate is of high interest. This is because real estate, which was considered only a residential environment in the past, is recognized as a stable investment target due to the ever-growing demand on it. In particular, in the case of the domestic market, despite the decrease in the number of people, the number of single-person households and the influx of people to large cities are accelerating, and real estate prices are rising sharply around the metropolitan area. Therefore, accurately predicting the prospects of the future real estate market becomes a very important issue not only for individual asset management but also for government policy establishment. In this paper, we developed a program to predict future real estate market prices by learning past real estate sales data using machine learning techniques. The data on the market price of real estate provided by the Korea Appraisal Board and the Ministry of Land, Infrastructure and Transport were used, and the average sales price forecast for 2022 by region is presented. The developed program is publicly available so that it could be used in various forms.

Key Words : Machine learning, Real-estate price prediction, Linear regression

*이화원, 숭실대학교 스마트시스템소프트웨어학과
**정회원, 숭실대학교 스마트시스템소프트웨어학과 (교신저자)
접수일자 2020년 9월 7일, 수정완료 2020년 11월 7일
게재확정일자 2020년 12월 4일

Received: 7 September, 2020 / Revised: 7 November, 2020 / Accepted: 4 December, 2020
*Corresponding Author: eslee@ssu.ac.kr
Dept. of Smart Systems Software, Soongsil University, Korea

01 연구 배경

01. 주제선정 이유

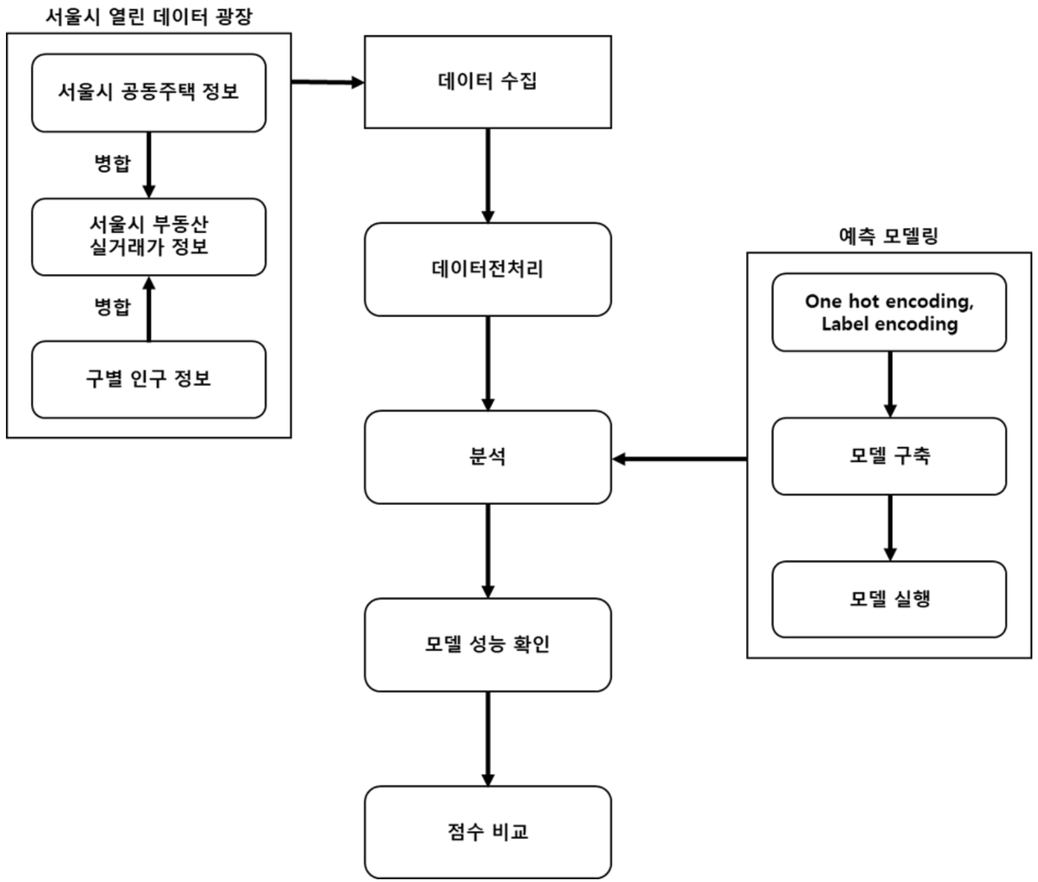
02. 선행연구 검토

03. 연구 프로세스

연구설계 및 프로세스

연구설계 및 프로세스

- 데이터 수집
- 데이터 전처리
- 분석
- 모델 성능확인
- 점수 비교



02

데이터 수집

01. 데이터 수집 방법
02. 데이터 목록
03. 데이터 별 변수

02 데이터 수집

01. 데이터 수집 방법

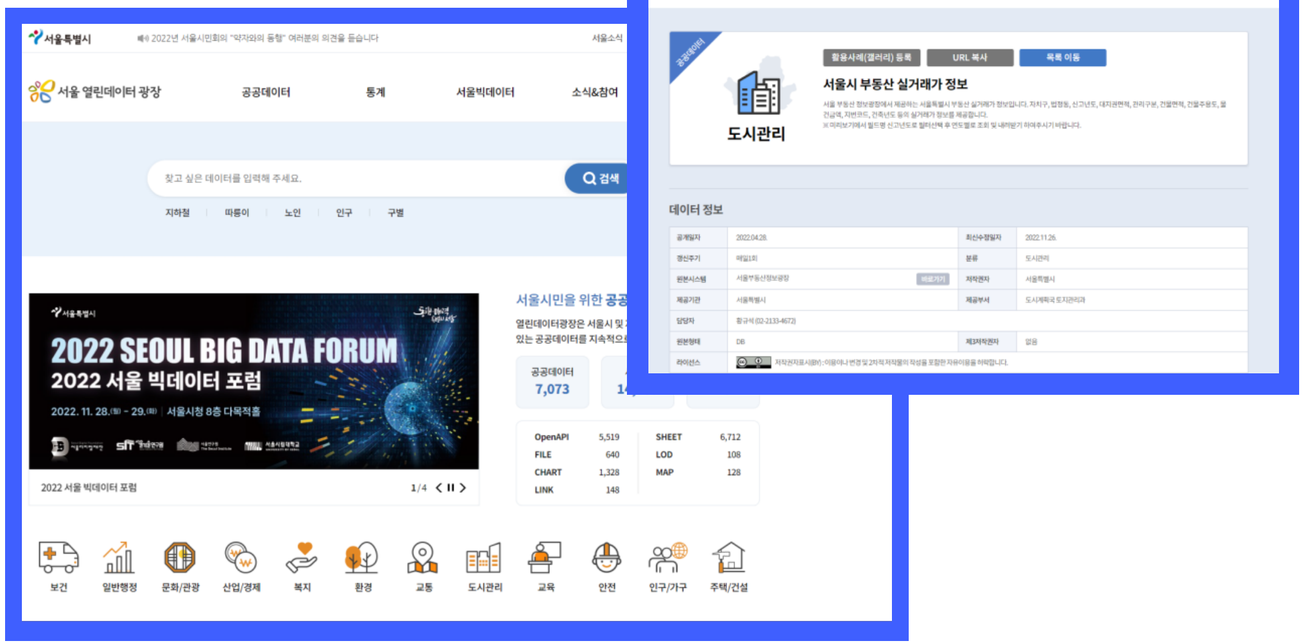
02. 데이터 목록

03. 데이터 별 변수

데이터 수집 방법

주택의 지역, 특성, 가격 데이터 수집

- 서울시에서 제공하는 데이터를 수집
- 수집된 데이터를 종합



02 데이터 수집

01. 데이터 수집 방법

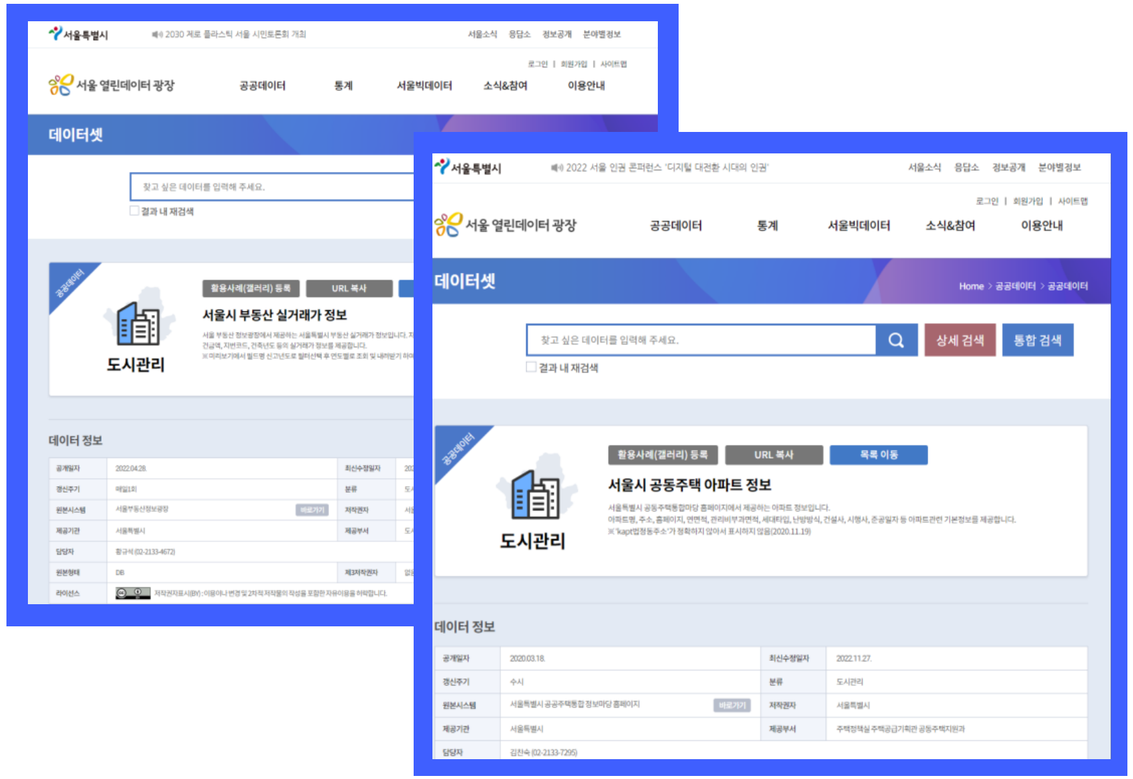
02. 데이터 목록

03. 데이터별 변수

수집 데이터 목록

수집 데이터 목록

- 서울시 부동산 실거래가
- 서울시 공동주택 아파트 정보
- 서울시 인구밀도(구별)



02 데이터 수집

01. 데이터 수집 방법

02. 데이터 목록

03. 데이터별 변수

데이터별 변수 요약

연구설계 및 프로세스

- 서울시 부동산 실거래가 - 아파트명, 거래금액, 위치정보, 면적 등
- 서울시 공동주택 아파트 정보 - 아파트명, 복도유형, 난방유형 등
- 서울시 인구밀도(구별) - 서울의 구별 인구밀도

```
0   접수연도              128580 non-null  int64
1   자치구코드             128580 non-null  int64
2   자치구명              128580 non-null  object
3   법정동코드             128580 non-null  int64
4   법정동명              128580 non-null  object
5   지번구분              118932 non-null  float64
6   지번구분명             118932 non-null  object
7   본번                118932 non-null  object
8   부번                118932 non-null  float64
9   건물명               118940 non-null  object
10  계약일               128580 non-null  int64
11  물건금액(만원)          128580 non-null  int64
12  건물면적(㎡)           128580 non-null  float64
13  토지면적(㎡)           68140 non-null   float64
14  층                 118940 non-null  float64
15  권리구분               0 non-null       object
16  취소일                0 non-null       float64
17  건축년도              128382 non-null  float64
18  건물용도              128580 non-null  object
19  신고구분               0 non-null       string
20  신고한 개업공인중개사 시군구명  0 non-null       object
```

```
0   번호                  2637 non-null   int64
1   k-아파트코드             2637 non-null   object
2   k-아파트명              2637 non-null   object
3   k-단지분류(아파트,주상복합등등)    2577 non-null   object
4   kapt도로명주소            2461 non-null   object
5   주소(시도)k-apt주소split    2637 non-null   object
6   주소(시군구)              2637 non-null   object
7   주소(읍면동)              2637 non-null   object
8   나머지주소                2113 non-null   object
9   주소(도로명)              2481 non-null   object
10  주소(도로상세주소)           2474 non-null   object
11  k-전화번호               2630 non-null   object
12  k-팩스번호               2601 non-null   object
13  단지소개기존clob           566 non-null    float64
14  단지첨부파일               184 non-null    object
15  k-세대타입(분양형태)          2614 non-null   object
16  k-관리방식               2624 non-null   object
17  k-복도유형               2621 non-null   object
18  k-난방방식               2630 non-null   object
19  k-전체동수               2610 non-null   float64
20  k-전체세대수              2633 non-null   float64
21  k-건설사(시공사)           2592 non-null   object
22  k-시행사                2579 non-null   object
23  k-사용검사일-사용승인일        2622 non-null   object
```

```
23  k-사용검사일-사용승인일          2622 non-null   object
24  k-연면적                2636 non-null   float64
25  k-주거전용면적              2607 non-null   float64
26  k-관리비부과면적             2635 non-null   float64
27  k-전용면적별세대현황(60㎡이하)     2604 non-null   float64
28  k-전용면적별세대현황(60㎡~85㎡이하)  2604 non-null   float64
29  k-85㎡~135㎡이하           2604 non-null   float64
30  k-135㎡초과              3 non-null      float64
31  k-홈페이지                773 non-null    object
32  k-등록일자                356 non-null    object
33  k-수정일자                2611 non-null   object
34  고용보험관리번호              2090 non-null   object
35  경비비관리형태               2610 non-null   object
36  세대전기계약방법              2489 non-null   object
37  청소비관리형태               2613 non-null   object
38  건축면적                2624 non-null   float64
39  주차대수                2632 non-null   float64
40  기타/의무/임대/임의=1/2/3/4    2637 non-null   object
41  단지승인일               2637 non-null   object
42  사용허가여부               2637 non-null   object
43  관리비 업로드              2637 non-null   object
44  좌표X                 2627 non-null   float64
45  좌표Y                 2627 non-null   float64
46  단지신청일               2637 non-null   object
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 4 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   구별       25 non-null     object
 1   구별인구     25 non-null     int64
 2   구별면적     25 non-null     float64
 3   구별인구밀도   25 non-null     int64
dtypes: float64(1), int64(2), object(1)
memory usage: 928.0+ bytes
```

03

데이터 전처리

01. 변수설정
02. 수치형 데이터
03. 변수 변환
04. 범주형 데이터
05. 최종 데이터 선정

03 데이터 전처리

01. 변수설정

02. 수치형 데이터

03. 변수 변환

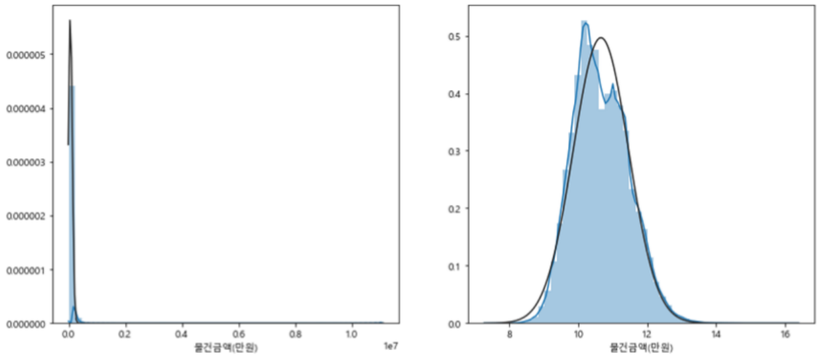
04. 범주형 데이터

05. 최종 데이터 선정

변수 설정

타겟변수 설정 및 로그변환

- 타겟변수로 "물건금액"을 설정 해당 데이터의 왜도를 로그변환을 통해 교정



데이터 가공

- 아파트의 팩스번호, 전화번호, 홈페이지와 같은 분석의 목적과 관련이 없는 변수들을 제거
- 아파트 명을 기준으로 서울시 부동산 실거래가와 서울시 공동주택 아파트 정보를 병합
- 결측치의 비율이 50%이상인 변수들을 제거
- 분양권과 같은 실물거래가 아닌 데이터와 거래취소일이 존재하는 실제 거래가 일어나지 않은 데이터 제거
- duplicate함수를 이용하여 완전히 중복되는 데이터 제거

03 데이터 전처리

01. 변수설정

02. 수치형 데이터

03. 변수 변환

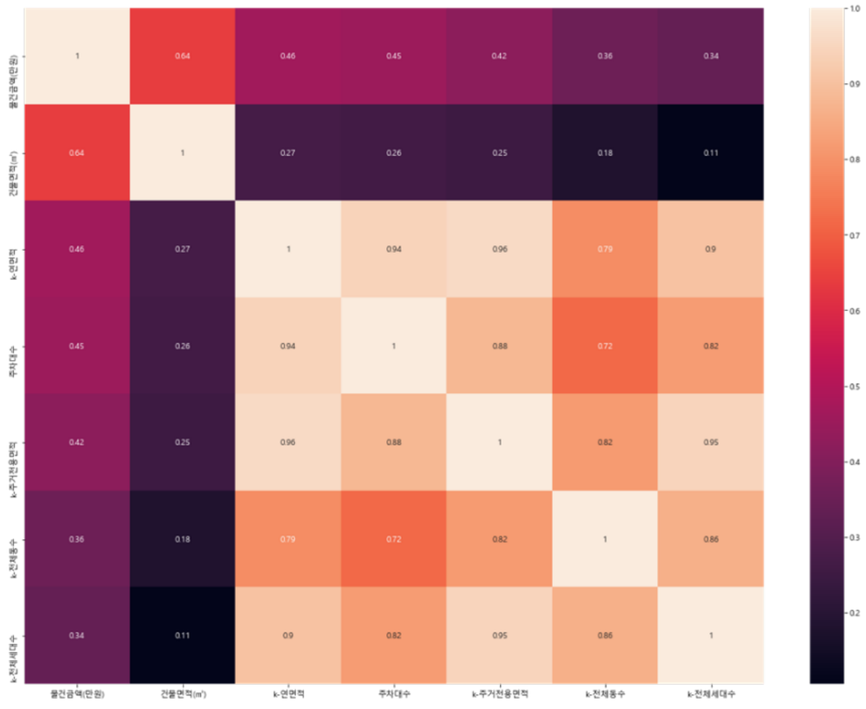
04. 범주형 데이터

05. 최종 데이터 선정

수치형 데이터 전처리

타겟변수와의 상관관계 분석

- 로그 변환된 "물건금액"과 수치형 변수들의 상관관계를 히트맵을 통해 분석



→ 상관계수가 0.40이상인 변수들 중 연면적과의 상관계수가 0.8을 넘는 변수들과 연면적의 관계 판단

03 데이터 전처리

01. 변수설정

02. 수치형 데이터

03. 변수 변환

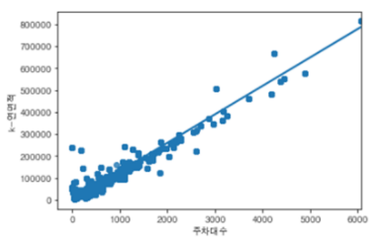
04. 범주형 데이터

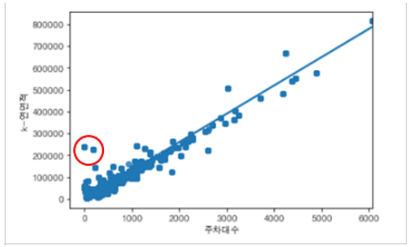
05. 최종 데이터 선정

수치형 데이터 전처리

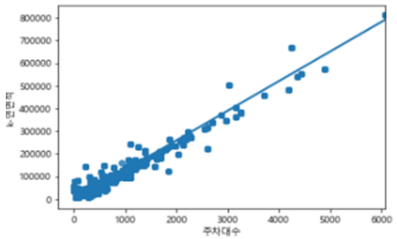
연면적과의 관계 분석 - 주차대수

- 주차대수와 연면적의 residual plot





→ 연면적이 건물의 실면적과 토지면적의 합계라는 것을 토대로 주차대수가 0에 가까운 데이터를 이상치라고 판단, 삭제



→ 이상치를 제외한 residual plot에서 주차대수는 연면적이 대표할 수 있다고 판단, 요인에서 제외한다

03 데이터 전처리

01. 변수설정

02. 수치형 데이터

03. 변수 변환

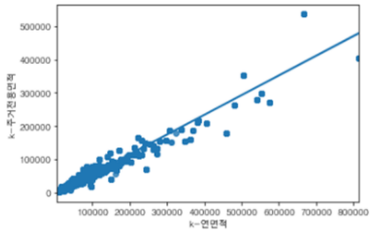
04. 범주형 데이터

05. 최종 데이터 선정

수치형 데이터 전처리

연면적과의 관계 분석 - 주거전용면적

- 주거전용면적과 연면적의 residual plot



→ 주거전용면적 또한 residual plot의 결과와 히트맵에서의 상관관계를 고려하였을 때,

연면적 변수의 특성이 주거전용면적을 대체할 수 있다고 판단하여 분석변수에서 제거

최종 수치형 변수

k-연면적	건물면적(m^2)

03 데이터 전처리

01. 타겟변수 변환

02. 수치형 데이터

03. 변수 변환

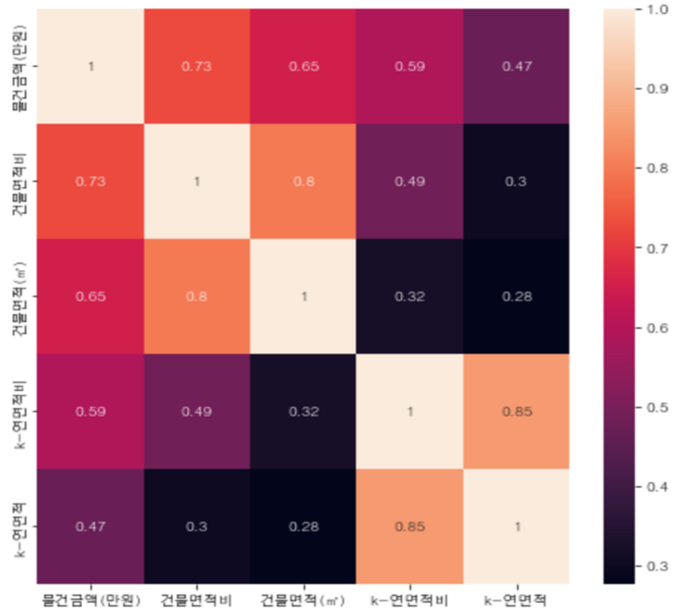
04. 범주형 데이터

05. 최종 데이터 선정

변수변환

수치형 데이터의 변환

- 기존연구의 분석에 따라 주택의 면적비를 인구밀도로 나눈값과 물건 금액의 상관관계를 분석



→ 건물면적은 0.08이 증가, k-연면적비는 0.12가 증가

기존의 변수를 '구별 인구밀도'로 나눈 값으로 대체하여 분석에 사용

03 데이터 전처리

01. 타겟변수 변환

02. 수치형 데이터

03. 변수 변환

04. 범주형 데이터

05. 최종 데이터 선정

범주형 데이터 전처리

범주형 데이터 선정

- 결측값이 50% 이상 넘는 데이터와 아파트 가격과 관련 없는 범주형 변수들을 1차 제거

자치구명	건축년도	경비비관리형태	K-단지분류
K-복도유형	건물용도	세대전기계약방법	청소비관리형태
K-세대타입	층		

03 데이터 전처리

01. 타겟변수 변환

02. 수치형 데이터

03. 변수 변환

04. 범주형 데이터

05. 최종 데이터 선정

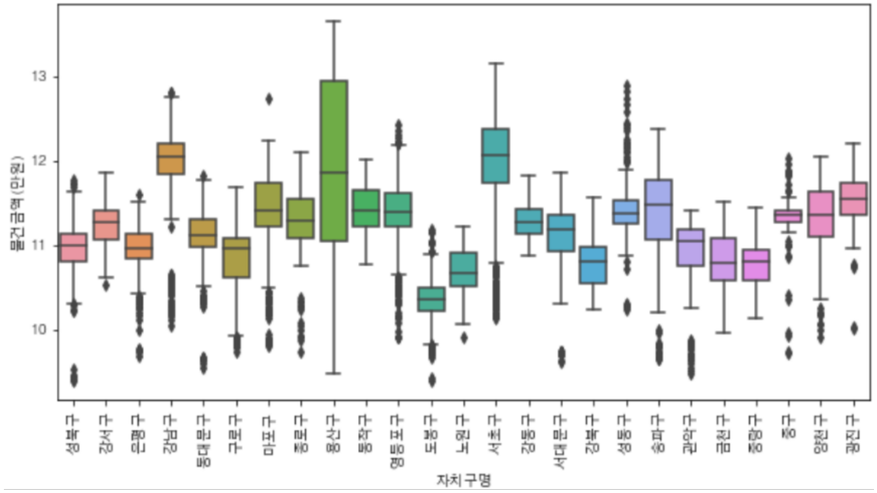
범주형 데이터 전처리

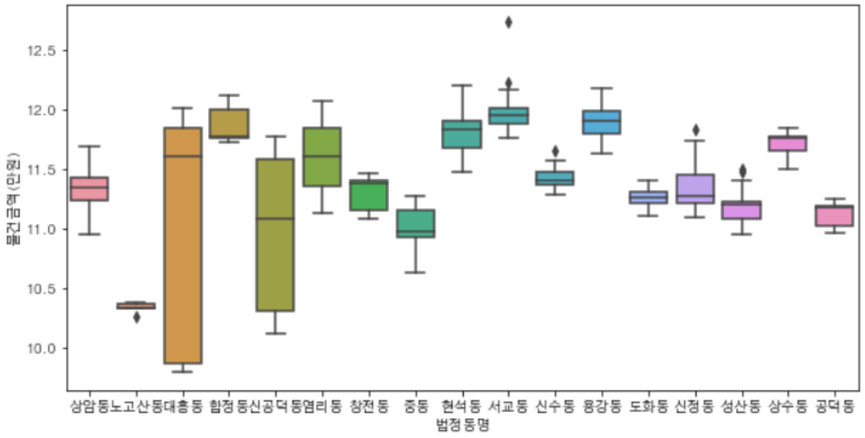
범주형 데이터 선정

- boxplot을 통해 범주형 데이터 시각화

범주형 데이터 중 지역별 변수 비교

- 범주형 데이터 내의 지역별 평균차이를 확인





03 데이터 전처리

01. 타겟변수 변환

02. 수치형 데이터

03. 변수 변환

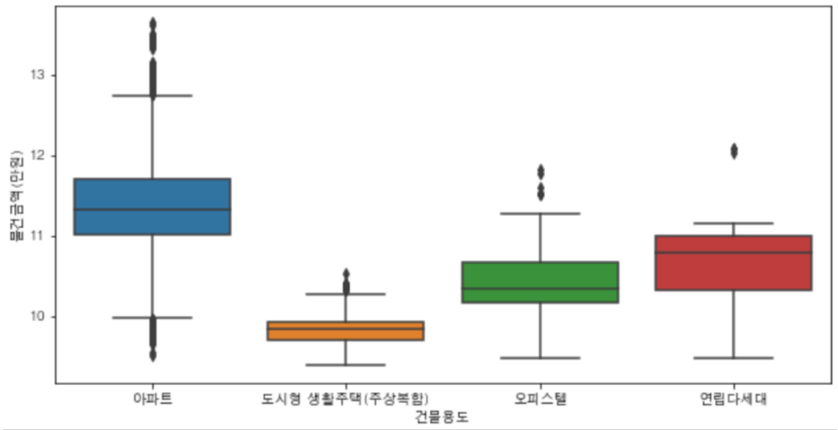
04. 범주형 데이터

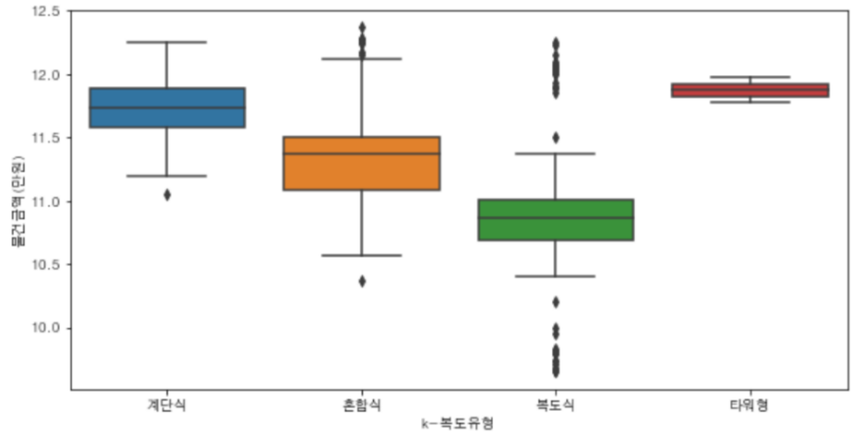
05. 최종 데이터 선정

범주형 데이터 전처리

범주형 데이터 중 주택별 특성 변수 비교

범주형 데이터 내의 주택 유형별 평균차이를 확인





→ 평균 차이가 나는 변수들을 선정, 선정된 변수 중 동일한 의미를 가지는 변수들의 삭제

03 데이터 전처리

01. 타겟변수 변환

02. 수치형 데이터

03. 변수 변환

04. 범주형 데이터

05. 최종 데이터 선정

최종 데이터 선정

최종 데이터 요약

- 타겟 변수는 '물건금액(만원)'

변수명	변수 설명	수치형/범주형	비고
K-연면적비	건축물의 각 층의 바닥면적합을 구별 인구밀도로 나눈 값	수치형	아파트명이 동일할 시 값이 동일함
건물면적비	건축물의 수직투영한 바닥면적을 구별 인구밀도로 나눈 값	수치형	아파트명이 동일하여도 세대별 특성을 가짐
자치구명	주택이 속한 서울의 구명	범주형	아파트명이 동일할 시 값이 동일함
건물용도	오피스텔, 아파트와 같은 주택의 분류	범주형	아파트명이 동일할 시 값이 동일함
건축년도	아파트가 지어진 해	범주형	년도의 특성상 수치형이 아닌 범주형으로 분류 아파트명이 동일할 시 값이 동일함

K-세대타입	분양 임대와 같은 주택의 양도방식	범주형	아파트명이 동일할 시 값이 동일함
K-난방방식	개별난방, 중앙난방과 같은 주택의 난방작동 방식	범주형	아파트명이 동일할 시 값이 동일함
법정동명	주택이 속한 서울의 동명	범주형	아파트명이 동일할 시 값이 동일함
K-복도유형	계단, 복도, 타워와 같은 주택내 세대가 배치 되어있는 방식	범주형	아파트명이 동일할 시 값이 동일함
층	건물의 층	범주형	년도의 특성상 수치형이 아닌 범주형으로 분류 아파트명이 동일하여도 세대별 특성을 가짐
세대 전기계약방법	단일, 종합과 같이 주택 내 전기공급 계약방법	범주형	아파트명이 동일하여도 세대별 특성을 가짐

04

모델링

04 모델링

01. 모델링 방식

02. 모델링

모델링 방식

사용한 알고리즘

K - NN	Decision Tree	Random Forest
Extra Tree	XGBoost	CatBoost

스케일링, 인코딩, 모델 평가방식

- Standardization 방식으로 스케일링을 진행
- 원핫인코딩과 레이블 인코딩 두가지 방법을 모두 사용하여 범주형 데이터 처리
- MAE값, RMSE값 R2 score값을 이용하여 모델 간의 성능을 비교

04 모델링

01. 모델링 방식

02. 모델링

원핫 인코딩 후 모델링

원핫 인코딩 후 모델링

- 자치구 + 법정동명 -> 총 222개의 변수

```
KNeighborsRegressor
        Training time: 0.021s
        Prediction time: 0.860s
        Total time: 0.881s
        MAE: 7049.487290643645
        RMSE: 11798.247704383497
        R2 score: 0.9691059085204098
```

```
DecisionTreeRegressor(random_state=42)
        Training time: 0.479s
        Prediction time: 0.013s
        Total time: 0.492s
        MAE: 7047.171530689977
        RMSE: 12476.538685255751
        R2 score: 0.9717408821209225
```

```
RandomForestRegressor
        Training time: 13.737s
        Prediction time: 0.104s
        Total time: 13.841s
        MAE: 5972.066017254148
        RMSE: 10374.272613625 16
        R2 score: 0.9821516562859197
```

```
ExtraTreesRegressor
        Training time: 17.538s
        Prediction time: 0.112s
        Total time: 17.650s
        MAE: 6083.7366610593635
        RMSE: 10175.285896178028
        R2 score: 0.9819654642185772
```

```
XGBRegressor
        Training time: 4.355s
        Prediction time: 0.016s
        Total time: 4.370s
        MAE: 6179.3876865931925
        RMSE: 9962.829380351592
        R2 score: 0.9815123058869758
```

```
CatBoostRegressor
        Training time: 8.823s
        Prediction time: 0.015s
        Total time: 8.838s
        MAE: 6124.355144268549
        RMSE: 9869.169546117368
        R2 score: 0.9823561924950127
```

04 모델링

01. 모델링 방식

02. 모델링

레이블 인코딩 후 모델링

레이블인코딩 후 모델링

- 자치구 + 법정동명 -> 총 11개의 변수

```
KNeighborsRegressor
        Training time: 0.061s
        Prediction time: 0.195s
        Total time: 0.256s
        MAE: 8516.844416716749
        RMSE: 15336.538490798104
        R2 score: 0.9557081830845495
```

```
DecisionTreeRegressor
        Training time: 0.036s
        Prediction time: 0.016s
        Total time: 0.052s
        MAE: 7481.490773252005
        RMSE: 13458.858832290798
        R2 score: 0.9680318654195414
```

```
RandomForestRegressor
        Training time: 3.365s
        Prediction time: 0.149s
        Total time: 3.514s
        MAE: 6476.412435788055
        RMSE: 10870.44990633121
        R2 score: 0.9779374251236124
```

```
ExtraTreesRegressor
        Training time: 3.350s
        Prediction time: 0.096s
        Total time: 3.446s
        MAE: 6546.813190181937
        RMSE: 11013.428513001445
        R2 score: 0.9798707460520708
```

```
XGBRegressor
        Training time: 0.680s
        Prediction time: 0.012s
        Total time: 0.692s
        MAE: 6211.649528544072
        RMSE: 10154.196569148598
        R2 score: 0.9824818887713793
```

```
CatBoostRegressor
        Training time: 5.461s
        Prediction time: 0.006s
        Total time: 5.467s
        MAE: 6134.224240733634
        RMSE: 9981.688005249121
        R2 score: 0.982748737578792
```

05

모델링 결과 비교

01. 인코딩 유형별 성능 비교
02. 수행 시간 비교
03. 모델 성능비교
04. 최종 모델 선정

05 모델링 결과 비교

01. 인코딩별 성능 비교

02. 수행 시간 비교

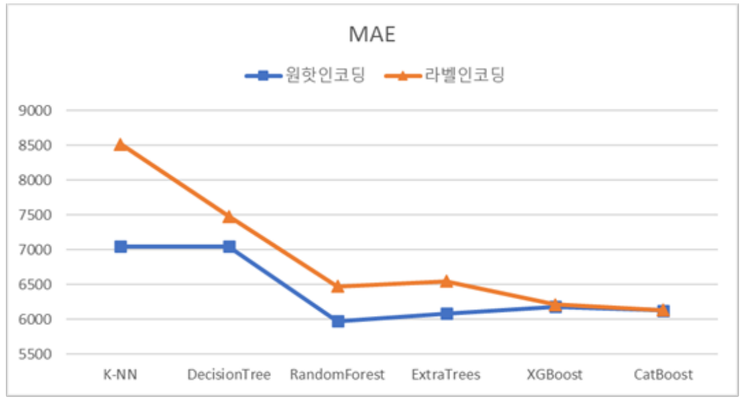
03. 모델 성능비교

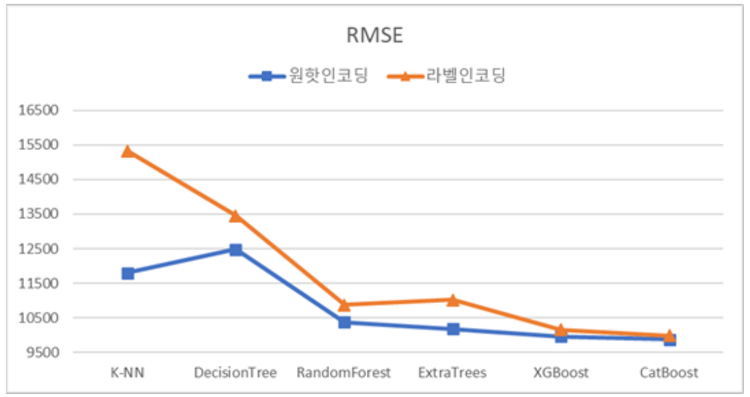
04. 최종 모델 선정

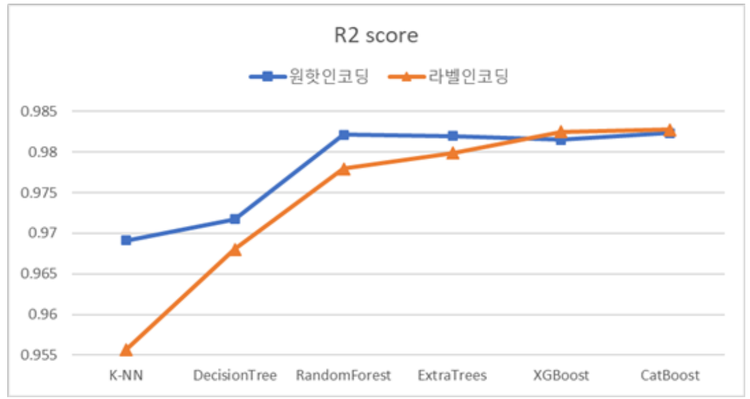
인코딩 유형별 성능 비교

평균적인 모델 성능

- 원핫인코딩이 레이블인코딩에 비해 뛰어남







05 모델링 결과 비교

01. 인코딩별 성능 비교

02. 수행 시간 비교

03. 모델 성능비교

04. 최종 모델 선정

인코딩 유형별 수행 시간 비교

인코딩 유형별 수행 시간 비교

- 원핫인코딩이 레이블인코딩에 비해 오랜 시간이 소요됨

05 모델링 결과 비교

01. 인코딩별 성능 비교

02. 수행 시간 비교

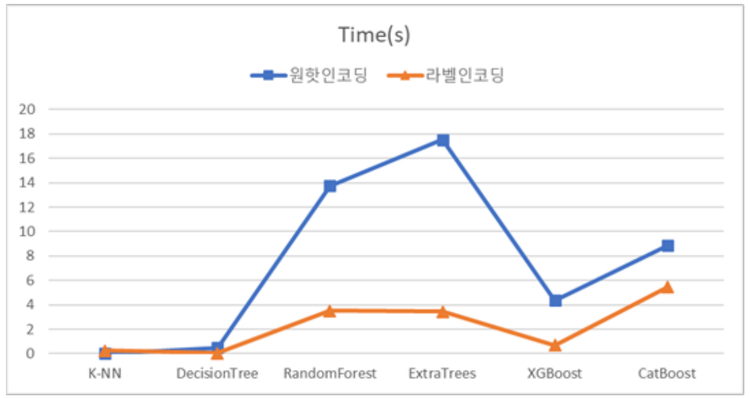
03. 모델 성능비교

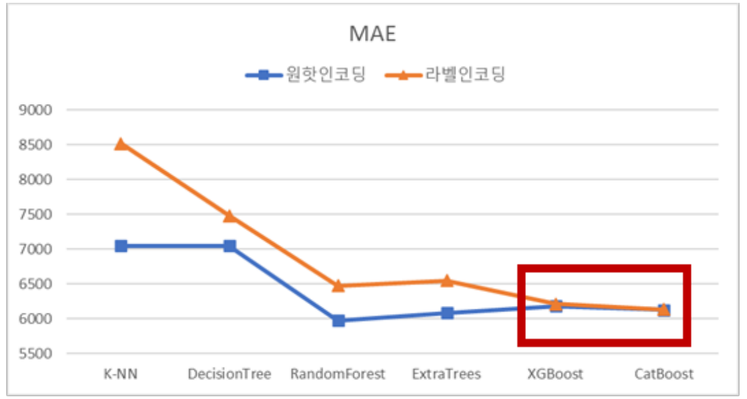
04. 최종 모델 선정

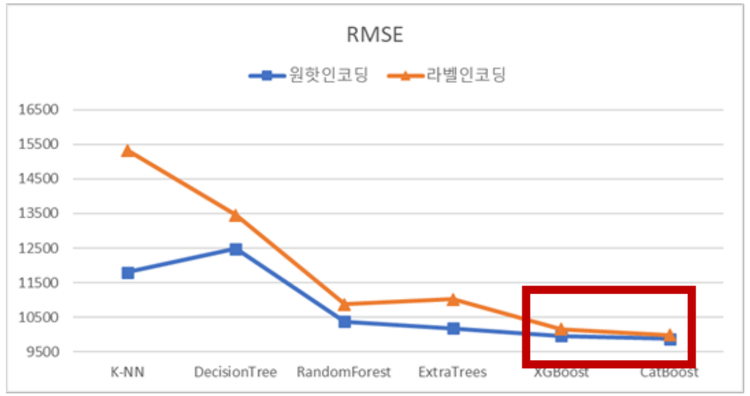
모델별 성능 비교

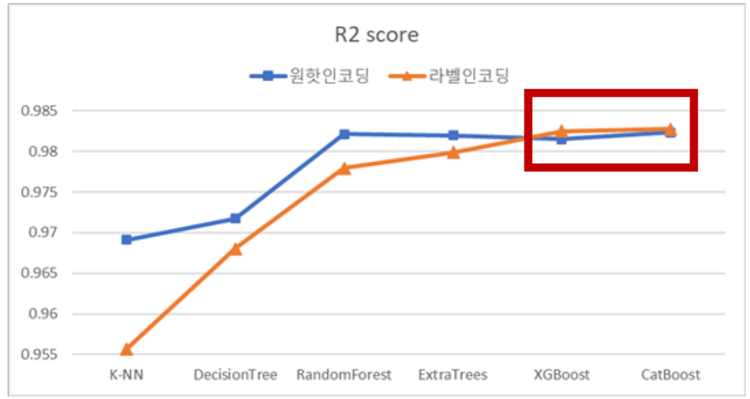
모델별 성능 비교

- 레이블 인코딩의 경우, XGBoost, CatBoost가 성능이 가장뛰어나며, 원핫 인코딩과의 차이가 가장 적음









05 모델링 결과 비교

01. 인코딩별 성능 비교

02. 수행 시간 비교

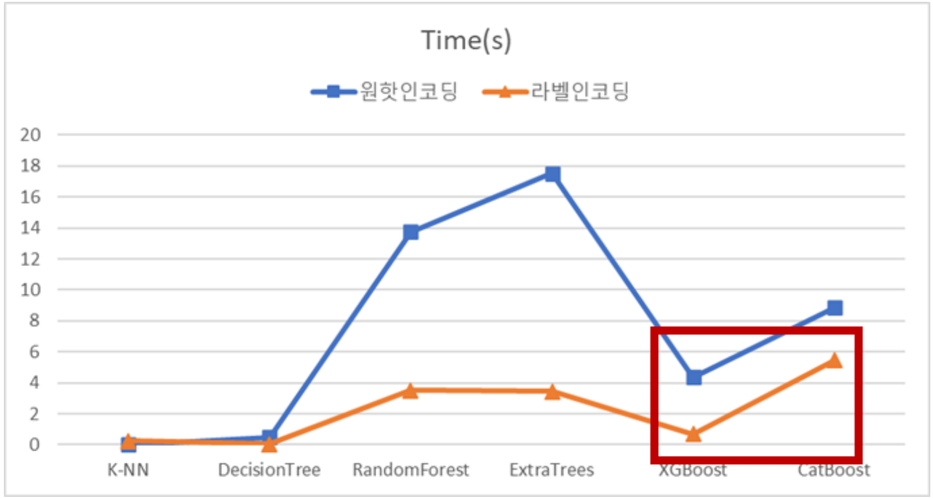
03. 모델 성능비교

04. 최종 모델 선정

최종 모델 선정

최종 모델 선정

- 레이블 인코딩을 사용한 XGBoost가 CatBoost 대비 우수한 처리시간을 가진다.



→ 추후 수치형 변수의 추가 가능성을 고려하여 레이블 인코딩을 사용한 XGBoost를 최종 모델로 선정

06

한계점

01. 환경적 요인 고려

02. GridSearchCV

06 한계점

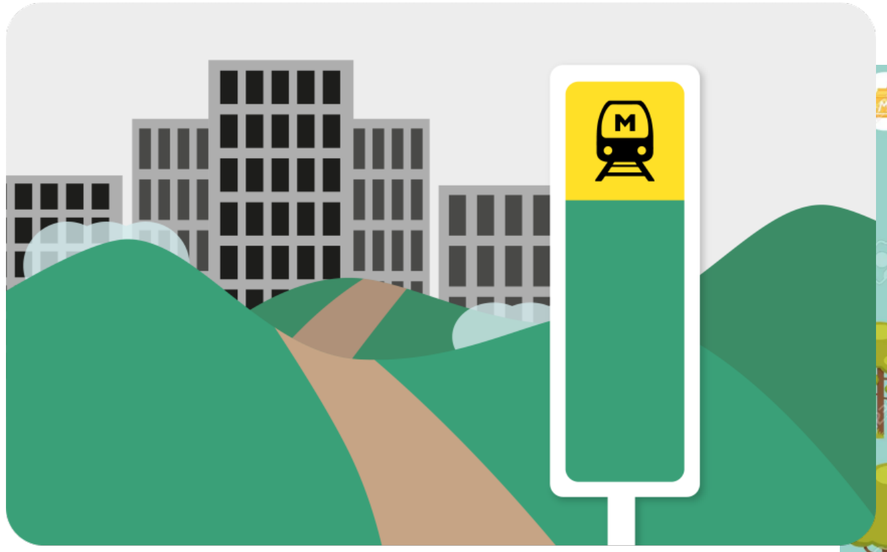
01. 환경적 요인 고려

02. GridSearchCV

환경적 요인 고려

거리변수의 활용

- X, Y좌표를 활용한 지하철역과의 거리, 학교와의 거리 변수에서의 유의미한 상관관계가 나타나지 않음
- 기존연구와 일반적 인식을 고려하였을 때, X,Y좌표가 주택단지의 중심점인 것,
  지역의 특징에 따라 주택의 가격에 영향을 주는 환경적인 요인이 다르다는 점에 의한것으로 추정



06 한계점

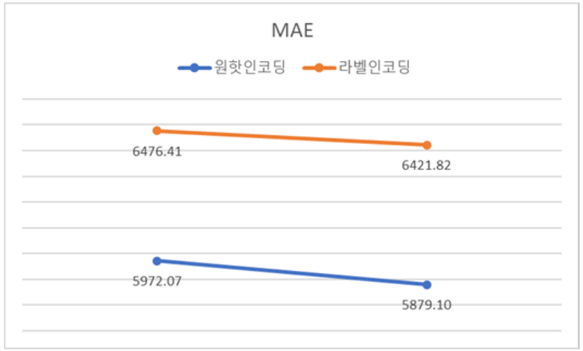
01. 환경적 요인 고려

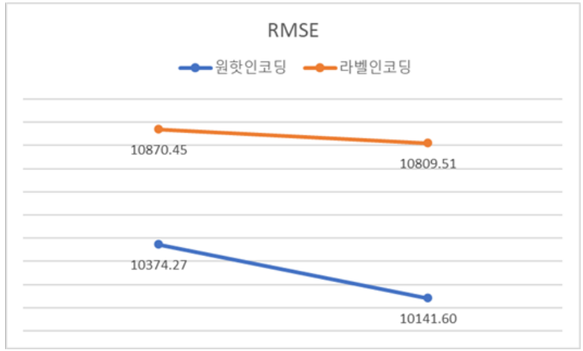
02. GridSearchCV

하드웨어적 한계 - GridSearchCV

GridSearchCV를 통한 하이퍼파라미터 도출 실패

- 랜덤포레스트에서의 GridSearchCV 수행결과 파라미터를 통해 MAE, RMSE값에서 유의미한 감소를 이룸
- 하드웨어적 한계로 인해 XGBoost와 CatBoost 알고리즘의 하이퍼파라미터 도출을 이루어내지 못함





→ 추후 GridSearchCV를 통한 하이퍼파라미터 도출을 통해 모델을 향상 시킬 수 있을 것으로 예상됨

07

결론

07 결론

결론

주택의 지역적, 개별적 특성을 이용한 예측모델 도출

- 서울시의 지역적 특징과 주택의 특징의 상관관계를 확인하고 이를 통한 예측 모델을 도출
- 레이블인코딩, XGBoost 알고리즘을 활용하여 98.25%의 예측률 달성

개선방향

- 추후 변수의 추가를 통한 주변 인프라와의 연관성을 추가
- GridSearchCV를 활용한 모델 보정을 통해 모델의 오류 값과 예측 값 향상이 가능할 것으로 예상