



아기돼지 삼형 제의 집값은 얼 마일까?

부동산 집값 예측하기

국민대지 삼남매

국민대학교 빅데이터경영통계전공

이현준, 유광열, 신예주

CONTENTS

- 스토리텔링
- 문제 정의
- 모델 적용, 한계점
- 사업화 계획



스토리 텔링

<아기돼지 삼형제>

옛날 옛날에 아기돼지 삼형제가 있었습니다. 하루는 삼형제의 어머니가 모두에게 집을 사라고 돈을 준후, 10년 후 가장 집값이 많이 오른 돼지에게 자신의 전재산을 물려준다고 말했습니다.



“부동산에 맡기는게 제일 믿음직하겠지...(귀찮아)”

“나는 나의 숨겨진 인터넷 서칭 능력을 통해서 나만의 집을 구할거야!”



“나는 지난 10년간의 자료 분석을 통해서 집값이 오를 것 같은 곳을 찾아볼테야!”


과연 셋째는 어떻게
자신의 집값이 오른다는
것을 알 수 있었을까?!





문제 정의

부동산 집값 예측
하기



문제 정의

유통·부동산 > 부동산

[주간 부동산 이슈-7월 3주] 서울 전셋값 55주 연속 상승...규제 약발 어디로

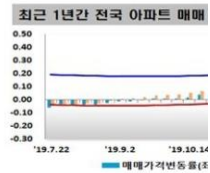


장호성 기자
hs6776@

기사입력 : 2020-07-17 17:14



[그림 1] 전국 아파트 거래 중 법인의 매입 비중 월간 추이



◇ 규제에도 오른 서울 집값

6.17 부동산대책과 7.10 부
트값은 6주 연속 상승세를
에서 이번 주 0.13%로 상승



(주1) 법인의 매입 비중은 전체 아파트 매매 건에서 법인의 매수건수가 차지하는 비중임

(주2) 법인의 매수건수는 법인이 개인/법인/기타(재외국민, 중증, 국가 등)로부터 아파트를 매수한 총 건

자료: 부동산통계정보시스템, 거래주체별 아파트 매매거래 (20.07.14 기준)

“서울시의 집값이 55주 연속 상승중”

서울 집값, 6·17대책 후 3주 연속 올랐다

KB부동산 리브온 주간 주택시장동향
서울 아파트가격 6일 기준 0.56% ↑

등록 2020-07-11 오전 6:00:00
수정 2020-07-11 오전 6:00:00

가 가

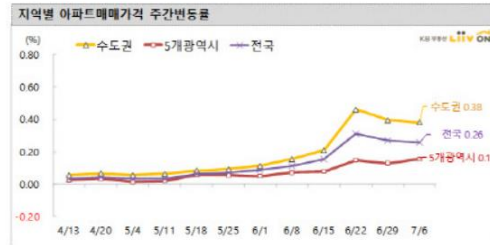


강신우 기자

N 기자구독



[이데일리 강신우 기자] 서울 아파트 매매가격이 '6·17부동산대책' 이후 3주 연속 오름세다.

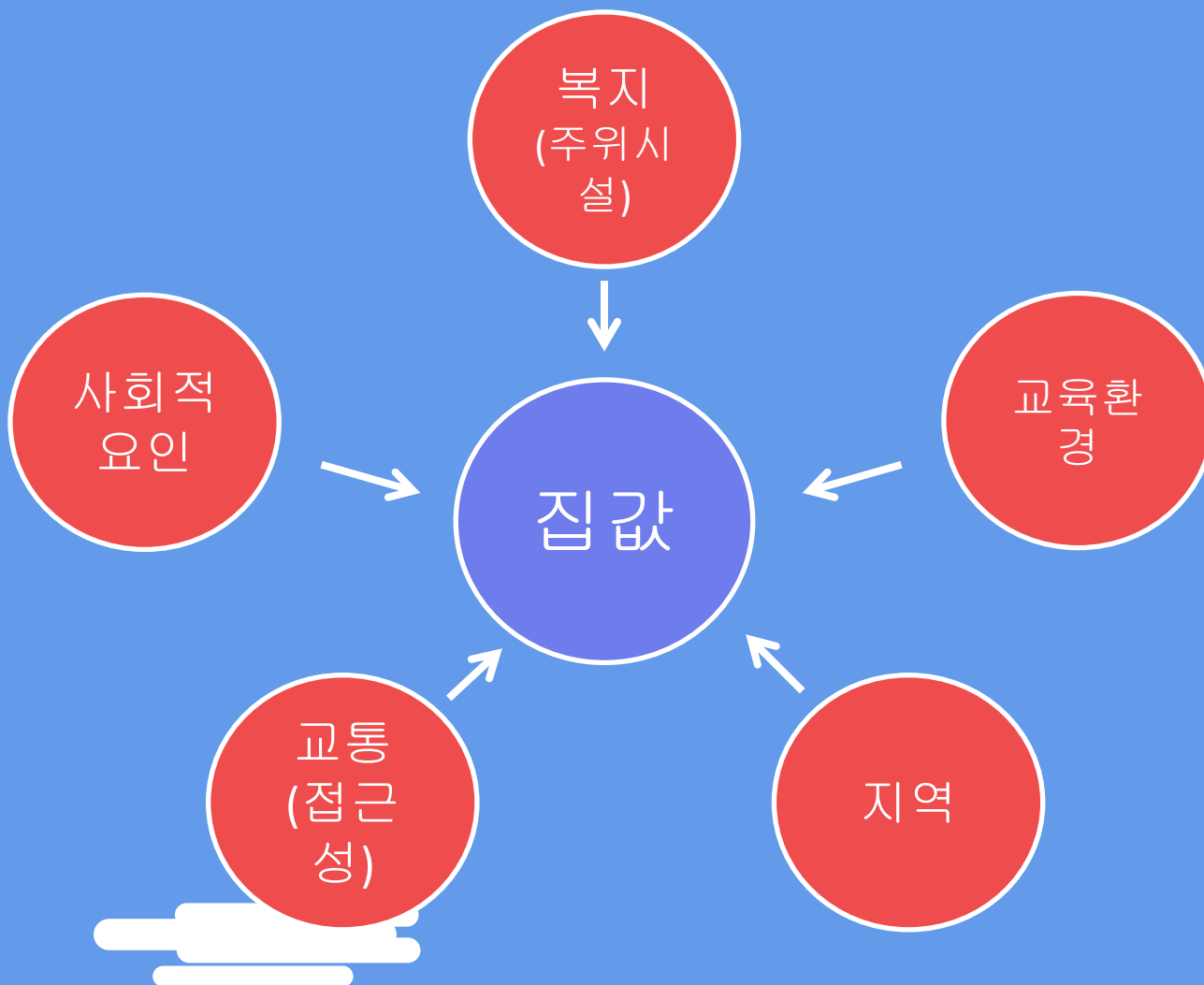


(자료=KB부동산 리브온)

모델 적용

머신러닝 - 집값의 증감율 예측
시계열 모델 - 미래의 집값을 예측

데이터 크롤링



Feature Engineering

집값에 영향을 주는
요인

행정동, 콜금리, 기준금리, 인구, 월별 매매가격
지수&변동, 상하강률,
초·중·고등학교 수, 지하
철 개통 여부, 보건소, 의
료기관, 랜드마크

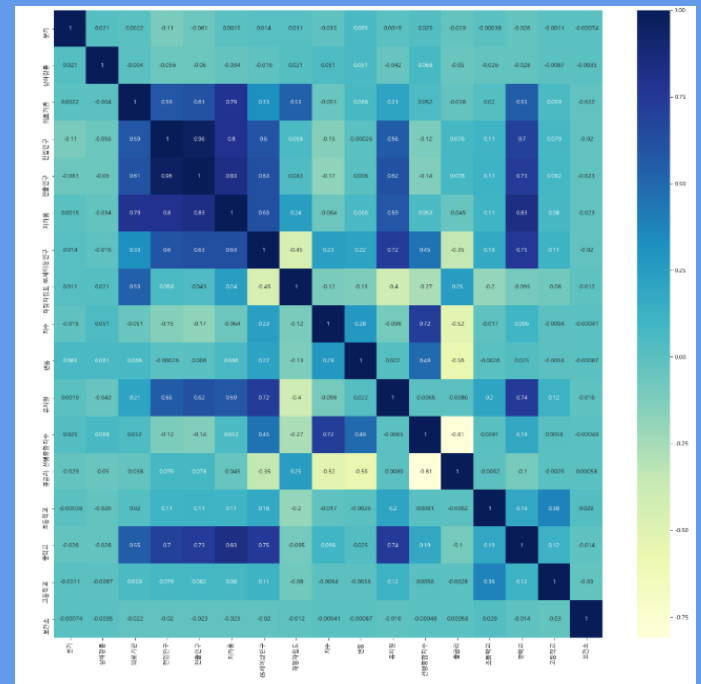
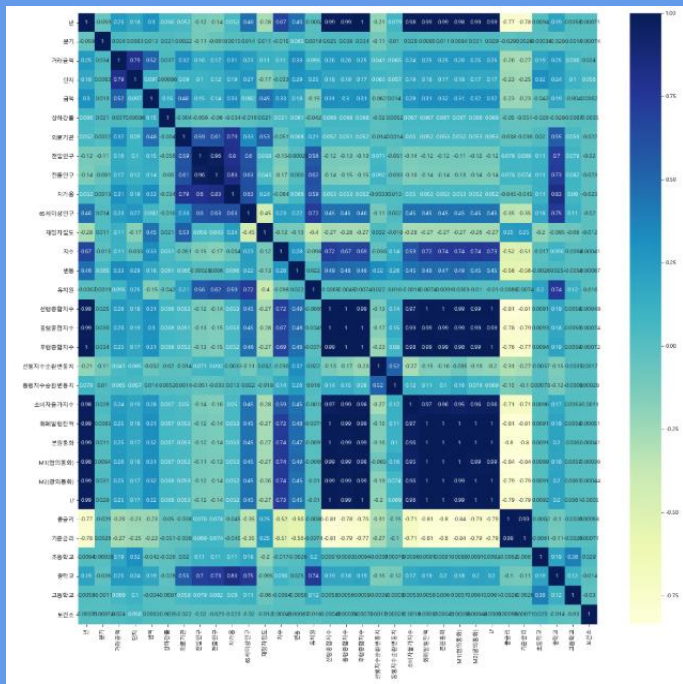


집값에 영향을 주는
요인

후행종합지수, 동행종합지수, 선행지수순환,
동행지수순환, 소비자물
가지수, 화폐발행잔액,
본원통화, M1(협의통화),
M2(광의통화), Lf

다중공산성에 의해 삭제(기준 :
0.9)

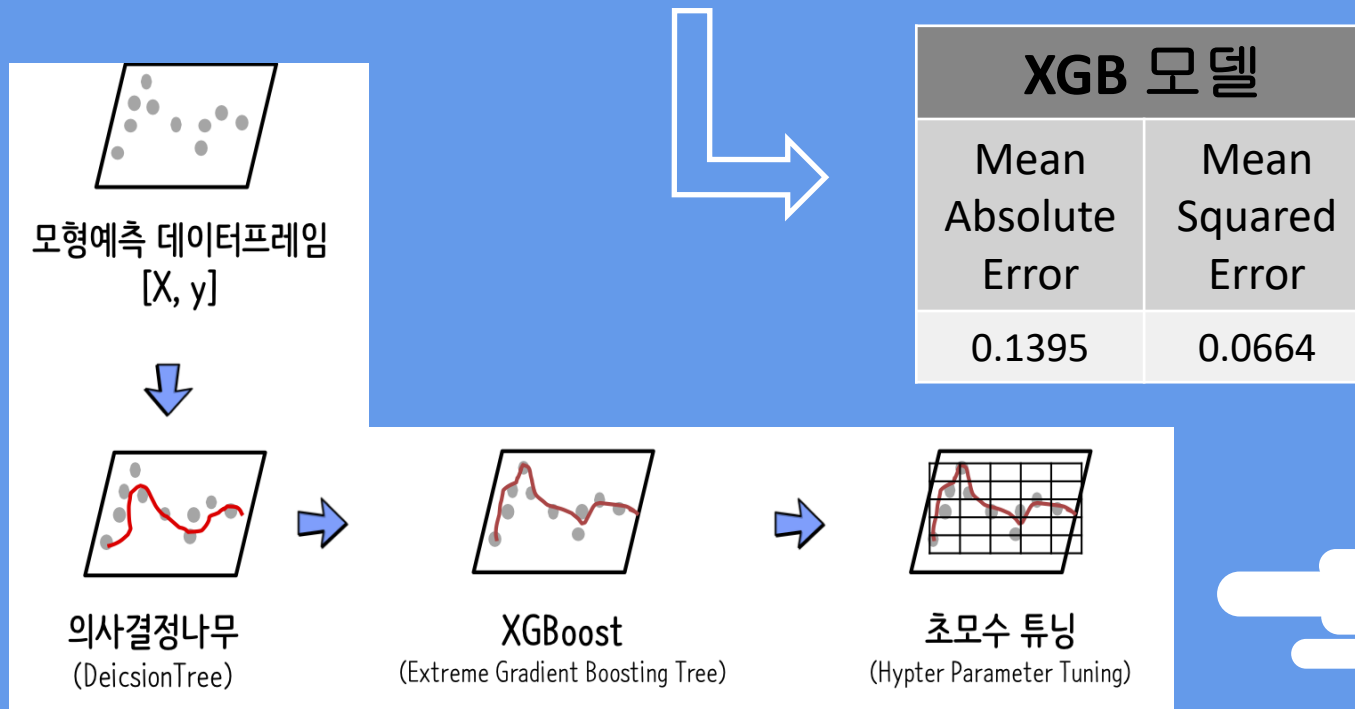
Feature Engineering



모델 적용

•머신러닝

외부요인에 대한 변동 → XGB 모델을 사용 → 증감률 예측 → 가중치 도출



모델 적용

- 머신러닝에서 XGB모델 선택 이유

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

Training loss Complexity of the Trees



모델의 성
능

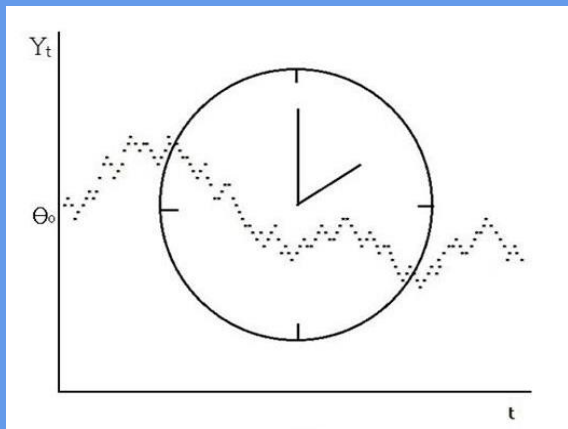


모델의 복잡
성

모델 적용

•시계열 분석

시간의 흐름에 영향을 받는 시계열 데이터 → ARIMA 모델 → 미래 예측값

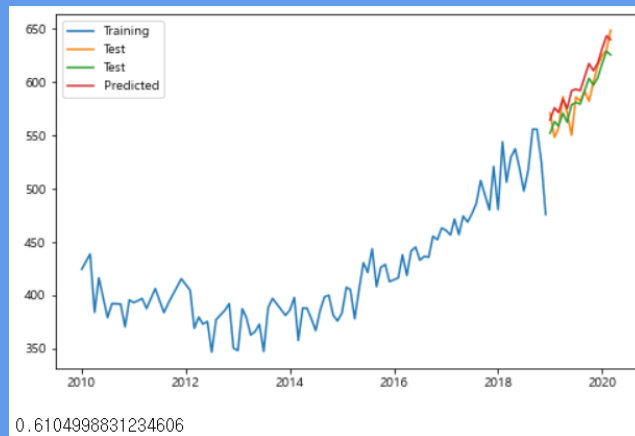
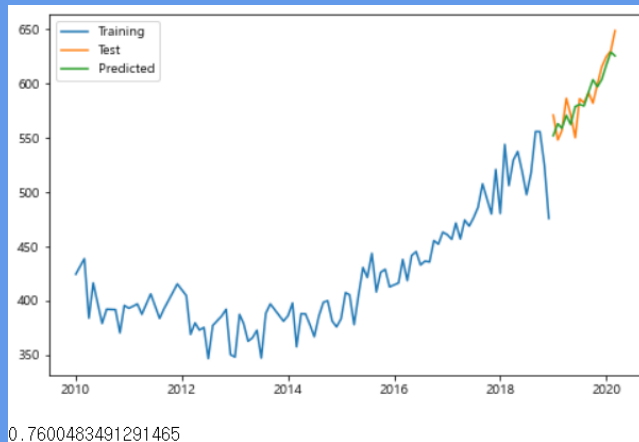


- 자기 회귀 모델과 이동 평균 모델을 모두 고려하는 모형.

- 시계열의 비정상성을 설명하기 위해 관측치의 차분을 이용.

모델 적용

•시계열 분석 결과



- 시계열 분석으로 도출해낸 예측값

Date	predicted_price	weight	predicted_price_w
2019-01-01	551.947590	0.022868	564.569528
2019-02-01	562.980363	0.022868	575.854598
2019-03-01	558.710040	0.022868	571.486621
2019-04-01	570.712031	0.022868	583.763073
2019-05-01	562.331928	0.022868	575.191335
2019-06-01	578.703722	0.022868	591.937518
2019-07-01	580.676755	0.021732	593.296022
2019-08-01	579.361314	0.021732	591.951994
2019-09-01	592.136278	0.021732	605.004584
2019-10-01	603.485787	0.022868	617.286300
2019-11-01	597.053583	0.022868	610.707005
2019-12-01	603.759152	0.022868	617.565916
2020-01-01	616.997321	0.022868	631.106816
2020-02-01	628.908193	0.022868	643.290066
2020-03-01	625.515969	0.022868	639.820269

- 최종 결과값

- 예측한 값과 가중치



한계점



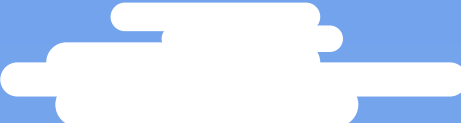

- 데이터가 부족하여 학습량이 적어, 정확도가 떨어짐. 거래량이 부족한 아파트는 예측이 어려움
- 데이터가 자세히 나와있지 않아서 원하는 형태로 가공하는 데 오랜 시간이 소요됨





사업화 계획

이 프로그램의 시장성 및 구체적인 사업
화 계획



사업화 계획

•개인사업적 측면

다방, 직방과 같은 부동산 플랫폼과 연결하여 부동산 정보와 매물 뿐만 아니라 집값 예측정보를 제공함으로써 소비자가 더 좋은 선택을 할 수 있도록 도와줌



사업화 계획

- 공익사업적 측면

정부와 협약을 맺어,
시민들이 좀 더 자신
의 집값에 대한 정보
를 얻을 수 있고, 더
나아가 정부가 앞으
로의 부동산 정책을
결정하는 데 있어 참
고 가능하다





THANK
YOU

국민대지 삼남
매