

AI

복덕방

ACORN ACADEMY 2021



프로젝트 설명

AI 복덕방은



서울 아파트의 가격을 예측한다!

부동산 가격에 대한 관심은 예로부터 뜨거웠다.
그러나 여러 요인으로 인해 부동산 가격은 변해왔고,
예측 가능성에 대한 논의가 활발히 진행됐다.
이러한 상황 속에서 부동산 투자의 핵심인 서울 아파트의 가격을
데이터 분석과 딥러닝을 통해 분석하는 프로젝트를 진행하게 됐다.



서울 아파트의 가격에 영향을 주는
요인을 분석한다.



아파트 설립일자, 전용면적,
CCTV수 등을 Feature로 잡아
모델을 학습



학습된 모델을 바탕으로
사용자가 변수를 입력해
가격을 예측

데이터 전처리 과정 및 학습



AI 복덕방

데이터 전처리
조사한 데이터를 학습할 수 있게 전처리



서울 아파트에 대한 데이터 조사 및 전처리
국토교통부 등에서 서울 아파트에 대한 설립일,
거래년월, 전용면적, 층 수 등 데이터를 수집.
범주형 데이터로 나타내기 위해 구를 특정 숫자로 매핑.



아파트 가격에 영향을 주는 변수조사 및 전처리
구별 공원 면적의 합, 어린이집 정원 당 교사 수,
CCTV의 수, K-remap지수가 아파트 가격에 영향을 준다고
가정하고 데이터 수집. 구별로 값 매핑

모델 학습

전통적 모델과 텐서플로우를 활용해 가격 예측

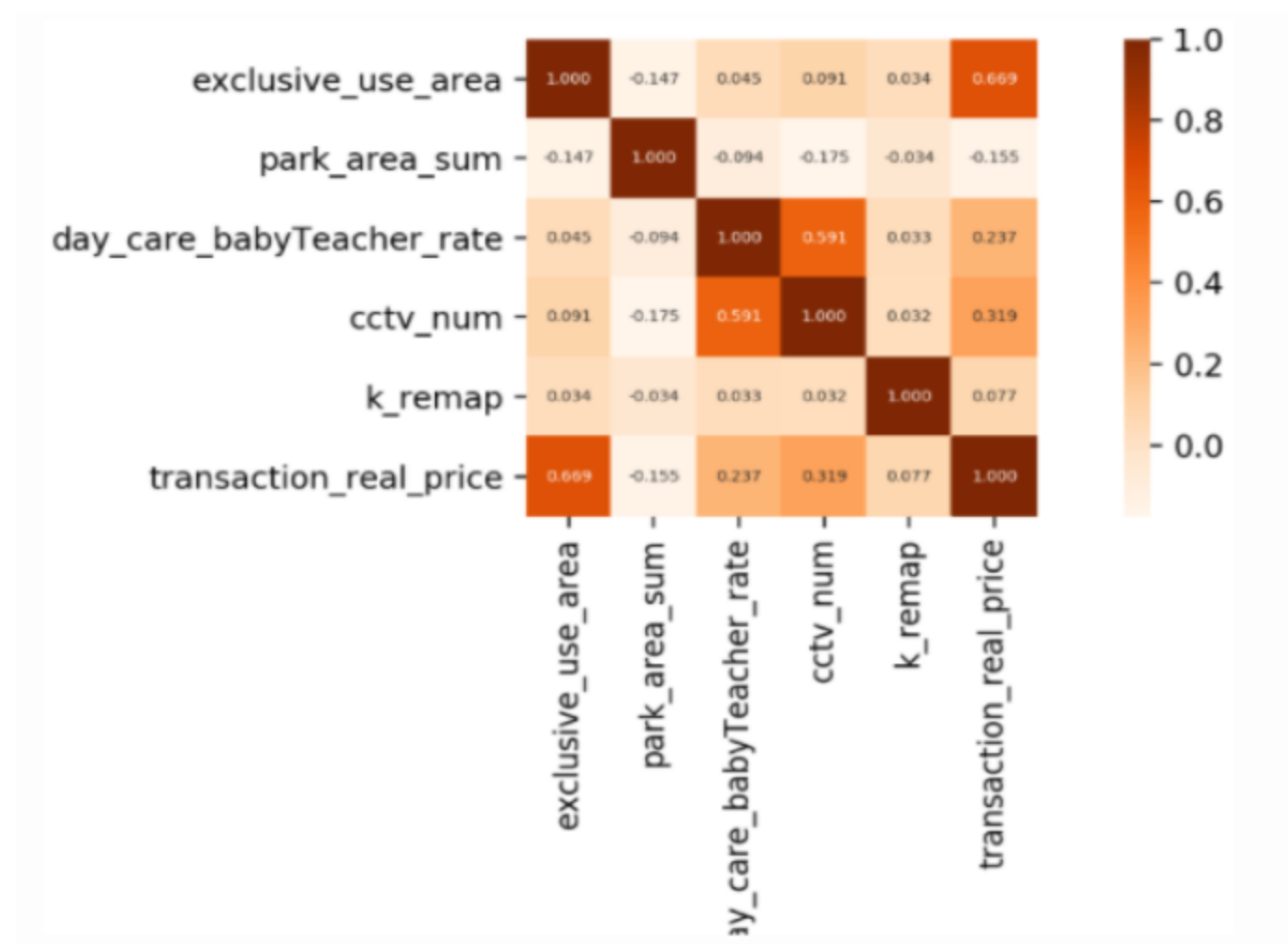


ols와 랜덤 포레스트를 사용해 분석 모델 작성
ols : 변수 10개를 사용해 다중회귀분석 실시
랜덤 포레스트 : 과적합 방지를 위해 모델 생성



텐서플로우를 사용해 분석 모델 작성
값의 단위가 다르기 때문에 표준화 작업 진행
4개의 Dense층을 사용해 모델 생성

변수 간 상관관계 분석



exclusive_use_area가 0.669로
상관관계가 가장 높다.

cctv_num이 0.319로 두번째로 높다.

day_care_babyTeacher_rate이
0.237로 세번째로 높다.

0.669

exclusive_use_area
상관관계가 가장 높다

0.077

K-remap
상관관계가 가장 낮다

OLS 모델 분석

[742285 rows x 10 columns]

summary :

OLS Regression Results

Dep. Variable:	transaction_real_price	R-squared:	0.544
Model:	OLS	Adj. R-squared:	0.544
Method:	Least Squares	F-statistic:	9.830e+04
Date:	Thu, 15 Apr 2021	Prob (F-statistic):	0.00
Time:	15:02:32	Log-Likelihood:	-7.6200e+05
No. Observations:	742285	AIC:	1.524e+06
Df Residuals:	742275	BIC:	1.524e+06
Df Model:	9		
Covariance Type:	nonrobust		

거래액을 종속변수로 설정하고,
전용면적, 층 수, 공원면적 등을 독립변수로 설정

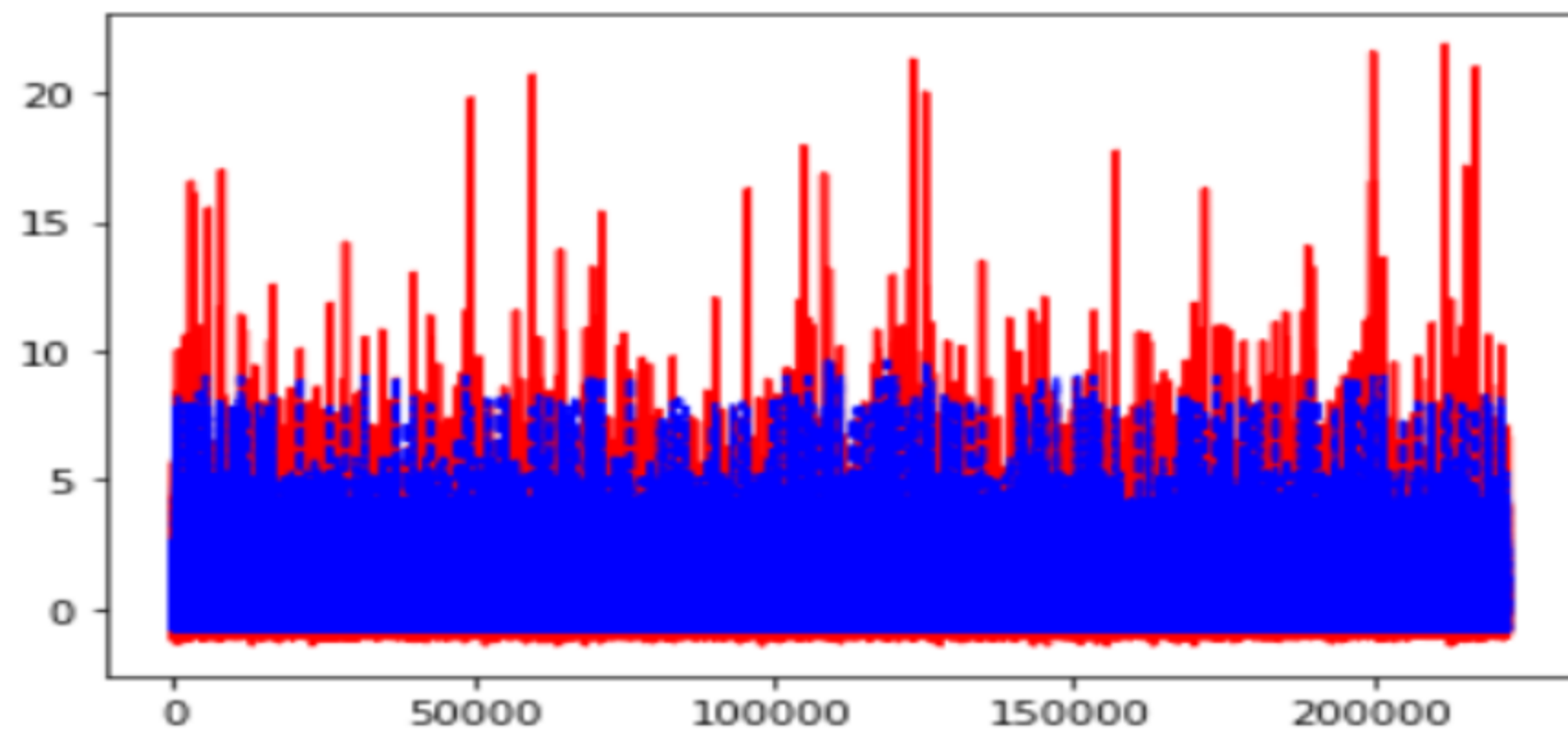
다중 분석으로 조정된 R-squared값으로 보면
설명력은 54%정도이다

54.4%

거래액에 대한 모델의 설명력

랜덤포레스트 모델 분석

r2_train: 0.6640853438910082
r2_test: 0.6615561269542667
평균제곱근오차MSE: 0.5855342110045365



과적합 방지를 위해 선택한 모델

66.4%

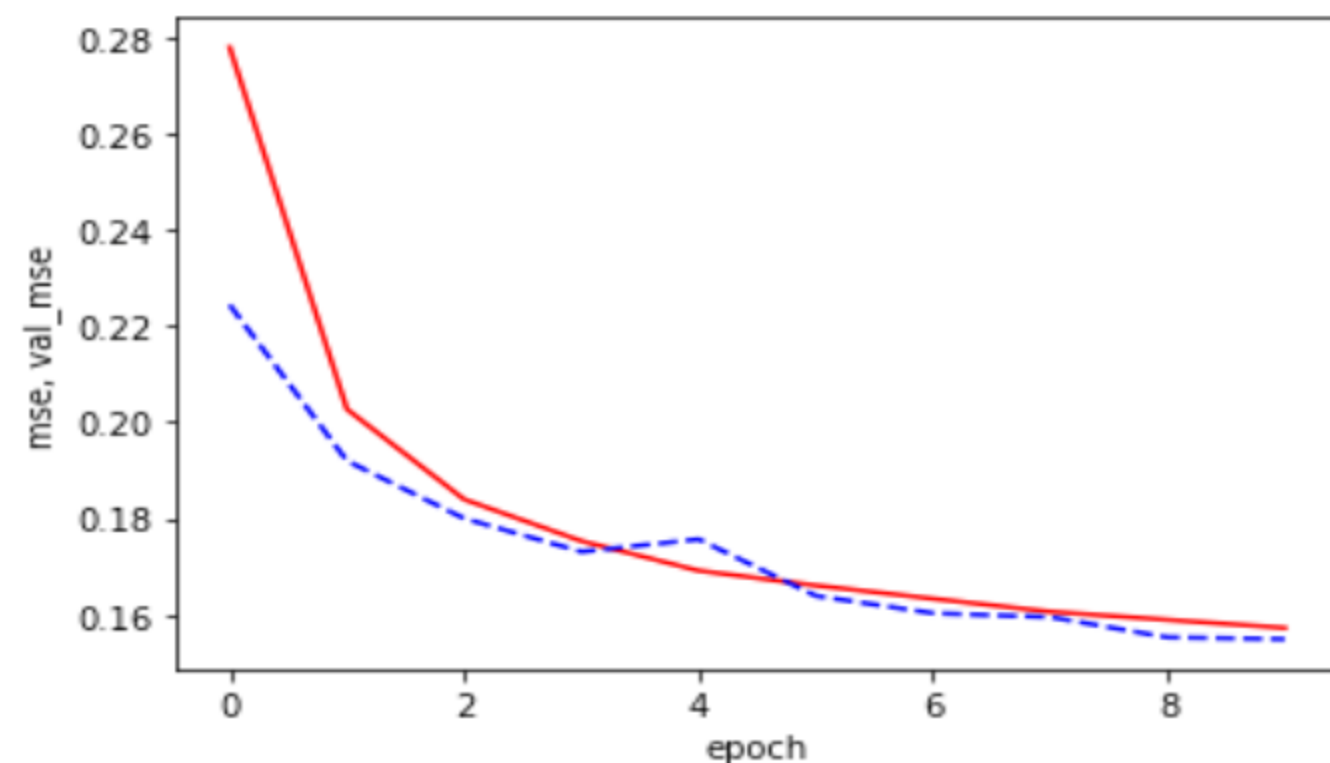
거래액에 대한 모델의 설명력

텐서플로우 모델 분석

loss: 113300232.0000 - mse: 113300232.0000 - val_loss: 708155072.0000 - val_mse: 708155072.0000

→ 처음 모델 학습 시 loss, mse값

```
mse : 0.19932036101818085
6959/6959 [=====] - 7s 1n
loss : 0.19932036101818085
```



설명력 : 0.8425529014345188

모델 수정 후 mse, val_mse값 추이 그래프

84.3%

거래액에 대한 모델의 설명력

보완할 지점들

모델 학습과 웹사이트 제작 시에 보완지점

모델 예측력 향상

변수의 상관관계 향상

마이페이지 및 로그인
기능 구현

사용 언어 및 기술

Python, Django, Ajax,
Jquery, MariaDB, Tensorflow

