**데이터 가공**

1. title과 description 열 삭제 (쓸모 X)
2. near\_subways 열에서 지하철이 있는 데이터는 1, 아닌 건 0으로 변경
3. platform\_id, sale\_id, user\_id, building\_type 삭제 (no data)
4. building\_total\_floors 삭제
   1. room\_floor 확인
5. maintenance\_items 삭제
   1. maintenance\_cost 확인
   2. maintenance\_cost missing value들을 전체 data의 average value로 입력
6. room\_size (평방미터) vs room\_size\_m2 (제곱미터)
   1. 제곱미터와 평방미터의 실제값이 같을 것으로 생각, 그러나 변환을 해보니까 맞지 않음
   2. 제곱미터와 평방미터를 따로 Training해보고 더 좋은 값으로 하기로 결정
   3. 의견 1: 한국인은 ‘평방미터’ 단위를 더 선호하는 경향이 있으므로 평방미터가 더 정확할 것으로 판단
   4. 의견 2: 집을 구할 때 실제 건물을 가서 크기를 보고 구매를 결정하기 때문에 제곱미터 단위가 더 정확할 것으로 판단
7. lat, lng, lat\_m1, lng\_m1, lat\_m2, lng\_m2 삭제
   1. 위도와 경도는 addr로 대체 가능
   2. addr\_d1을 “서울특별시”로 한정하고 서울특별시가 아닌 data는 삭제
   3. addr\_d1, addr\_d2 삭제 (~동으로 구별 가능)
   4. addr\_d4, addr\_d5 삭제 (no data)
8. sort, need\_update 삭제 (no data)
9. cdate, viewcnt 삭제 (no data)
10. write\_date 년,월,일로 가공
11. option을 op0 ~ op12로 분해
12. sell\_type을 sell\_type0 ~ sell\_type2로 분해
13. is\_selling, sell\_by\_flag 이쉬끼들진짜

id, platform\_id, sale\_id, user\_id, **addr\_d1 ~ addr\_d5**, **near\_subway**, **sell\_type**, **deposit**, **rent**, **room\_type**, room\_type\_str, **room\_floor**, building\_type, building\_total\_floors, **maintenance\_cost**, maintenance\_items, **room\_size\_m2**, **room\_size**, **options**, **parking**, **write\_date**, lat, lng, is\_selling, thumbnail, need\_update, sell\_date, sell\_by\_flag, cdate, viewcnt, sort, lat\_m1, lng\_m1

linear regression : 결과값 폭망

ridge regression : 굉장히 잘 나옴

lasso regression : ridge보단 덜 하지만 잘 나옴

logistic regression : 12시간 이상 돌렸는데 결과값 X

SVM : 12시간 이상 돌렸음, 결과값 그렇게 좋지 않음

neural network : 가장 좋은 결과값, relu먼저 선택후 layer 50 선택. (relu, 50)에서 train 0.78

**학습**

1. Data의 크기가 충분히 크기 때문에 Cross-Validation은 실시하지 않음.