2019 /01/ 19

# 멘토링 회의 내용

train 데이터 기반 예측률 확인하는 과정 필요

코드 재확인(오차가 5억인건 이상한 상황임)

검증내역 시각화

기본적인 데이터셋 분석부터 시작하기

1. 랜덤하게 7:3 분할 분석 -> 검증 내역 확인할 수 있도록

2. 지역 분할해서 분석해보기.

다른 회귀 방식도 써보기.

랜덤 포레스트 예상.

행별 이전 가격의 평균(한달이든 어떤 기준별)

테스트에 트레인에 없는 아파트 있는 경우

근접거리 아파트의 가격을 활용하는 방식 사용하기(주의할 점은 아파트 가격을 요소로 하는 것이 아니라 아파트 id를 찾는 것으로 해야할듯)

테스트 추가하려면 실거래가 기반으로 해서는 못씀.

거리/준공연도 를 범주화해보기.

2010년 이전/이후 아파트 다름.(2010년 이후 아파트는 지하에서 바로 올라갈 수 있음. 등)

학교도 설립연도 기준 범주화해보기.(구간 단위로)

환승역은 가중치 필요.

이상치 outlier 정리해보기

(단, 주의할 점은 제거한 것과 제거하지 않은 것 2가지 버전으로 만들어야 함. 어느 쪽이 더 영향력 높을지는 알 수 없음)

보내주신 pdf파일 정독하기(저번주)