프로젝트 제안 – 건물별 주차 수요 예측

출처: 데이콘(주차 수요 예측 대회)

<https://dacon.io/competitions/official/235745/overview/description>

훈련 셋 약 3000개, 테스트 셋 약 1000개. 총 4000개 샘플로 되어 있는 데이터.

15개 columns로 구성되어 있음. 각 특성은 다음과 같음.

'단지코드', '총세대수', '임대건물구분', '지역', '공급유형', '전용면적', '전용면적별세대수', '공가수',

'자격유형', '임대보증금', '임대료', '도보 10분거리 내 지하철역 수(환승노선 수 반영)',

'도보 10분거리 내 버스정류장 수', '단지내주차면수', '등록차량수'

Target 은 등록차량수.(주차 수요)

Target이 연속형이므로 회귀모델 적용 가능.(예제코드에서는 RandomForest Regressor활용)

“임대보증금, 임대료, 도보 10분거리 내 지하철 역 수, 도보 10분거리 내 버스정류장 수”에 결측치가 존재하나,

지하철 역 수나 버스정류장 수는 결측치가 아니라 없는 것으로 볼 수 있으므로 “0” 으로 대치하면 될 것 같고.

임대보증금이나 임대료 특성의 경우 데이터를 보고 어떻게 전처리 할지 결정해도 될 것 같음. 예제코드에서는 없애는 대신 -1로 대치하여 다른 특성을 살렸음.

2개 특성 제외해도 11개 특성이 남아 있으므로(단지코드, target 제외)

결측치가 크게 문제를 일으키지는 않을 것으로 보인다.