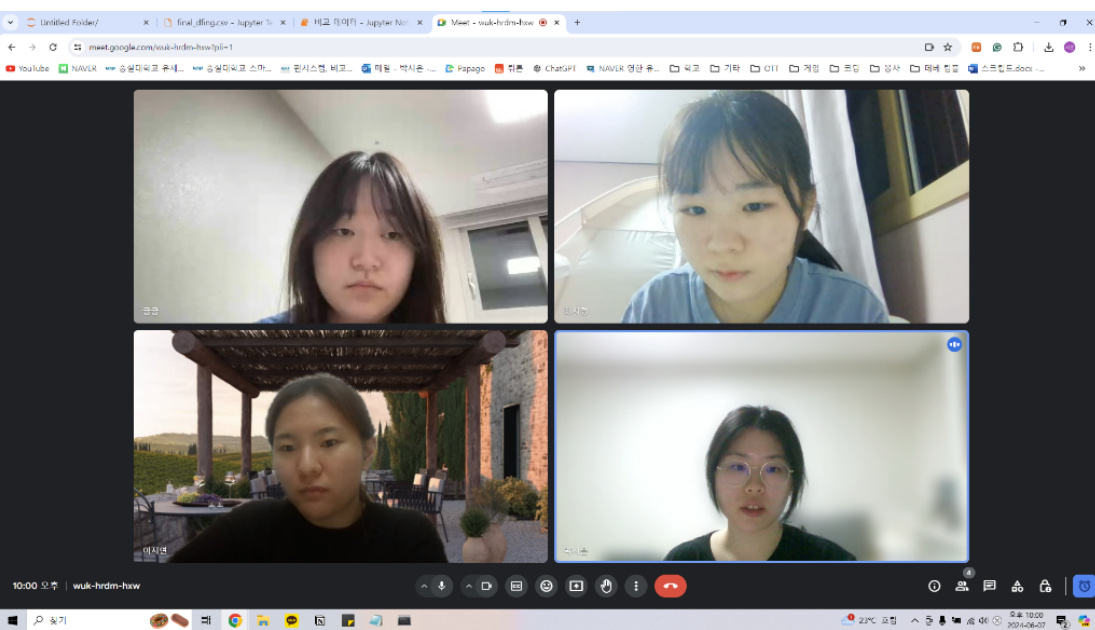


팀 회의록

팀명	베짱이들	참석자	김규린, 박시은, 이지연, 최서정
일시	6/7(금) 22:00	장소	온라인 회의
회의 내용	<p>1. PPT 수정</p> <p><배경></p> <ul style="list-style-type: none"> - 강서구 따릉이의 사용량에 대한 기사 추가 <p><속성과 target data 분석></p> <ul style="list-style-type: none"> - 속성과 대여량 간의 산점도를 나타낸 그래프 추가 <p><모델 학습></p> <ul style="list-style-type: none"> - 다항 회귀 코드 및 결과 추가 <p><인사이트></p> <ul style="list-style-type: none"> - 교통 상황, 행사 등의 요인을 고려하여 추가 분석 필요 <p>: 프로젝트 회고로 이동</p> <p>2. 대본 초안</p> <p><배경></p> <ul style="list-style-type: none"> - 강서구를 선택한 이유 <p>: 따릉이의 사용량이 가장 많은 역</p> <p><데이터 전처리></p> <ul style="list-style-type: none"> - 3, 6, 9, 12월의 데이터 사용하여 전처리한 이유 - 거치대 수를 대여소 별 거치대 기계 수만큼 매일 0시에 채워진다고 가정 <p>: 트럭이 개입하지 않은 상황을 고려하기 위한 방법</p> <p><속성과 target data 분석></p> <ul style="list-style-type: none"> - target data와 변화 패턴이 보이는 속성은 time, station, temp임 - holiday, day, season, prec은 뚜렷한 패턴이 없어 영량이 적을 것으로 예상 <p><모델 학습></p> <ul style="list-style-type: none"> - date, time 등 시계열 데이터를 학습 시키므로 train/test set을 적절히 분할 하는 것이 중요 <p>: 날짜 별로 묶어서 분할함</p> <ul style="list-style-type: none"> - 다항회귀와 랜덤포레스트 회귀 중 랜덤포레스트의 예측 성능이 좋았기 때문에 선택 <p><특성 중요도></p> <ul style="list-style-type: none"> - 모델 학습 중, 각 속성별 예측에 도움을 주는 정도를 보여줌 - 거짓 상관관계 학습 여부를 알기 위해 특성 중요도를 추출함 <p><예측 결과></p> <ul style="list-style-type: none"> - (적정 거치 대수) - (현재 거치 대수)로 실시간 트럭 적재량 계산 - 어느정도 실제 데이터와 비슷한 패턴을 보임 <p><인사이트></p> <ul style="list-style-type: none"> - Random Forest 회귀 모델을 사용하여 평균 train MSE: 40.318, 평균 test MSE: 		

	<p>44.329, 평균 train R^2: 0.877, 평균 test R^2: 0.869의 성능을 달성</p> <ul style="list-style-type: none"> - 특성 중요도를 분석한 결과, 대여소 번호(station), 시간(time), 온도(temperature) 순서대로 특성 중요도가 높다는 사실을 확인 - 모델의 예측값을 바탕으로 실시간 대여소의 거치된 따릉이 개수에서 몇 대를 가져다 주어야 적절한 따릉이 수량인지 파악함 <p><프로젝트 회고></p> <ul style="list-style-type: none"> - 대여소별 실제 거치대 수량을 알 수 있다면 더 좋은 예측이 가능할 것임 - 상관관계가 낮게 나타나는 이유를 명확히 찾지 못했으나, 전체리 과정에서 문제가 생겼을 것으로 판단 <p>: 지형/습도 등의 다른 속성을 이용해 추가 분석 진행하면 좋음</p> <ul style="list-style-type: none"> - 데이터 객체 수가 많아서 최적화 코드를 실행하지 못함 <p>: 데이터 축소나 성능이 좋은 컴퓨터를 사용해 최적화해 모델 예측도를 높이면 예측에 도움이 될 것임</p>
<p>증빙 사진</p>	

※ 증빙 사진 혹은 캡처본은 시간과 참여 인원이 나오도록 작성