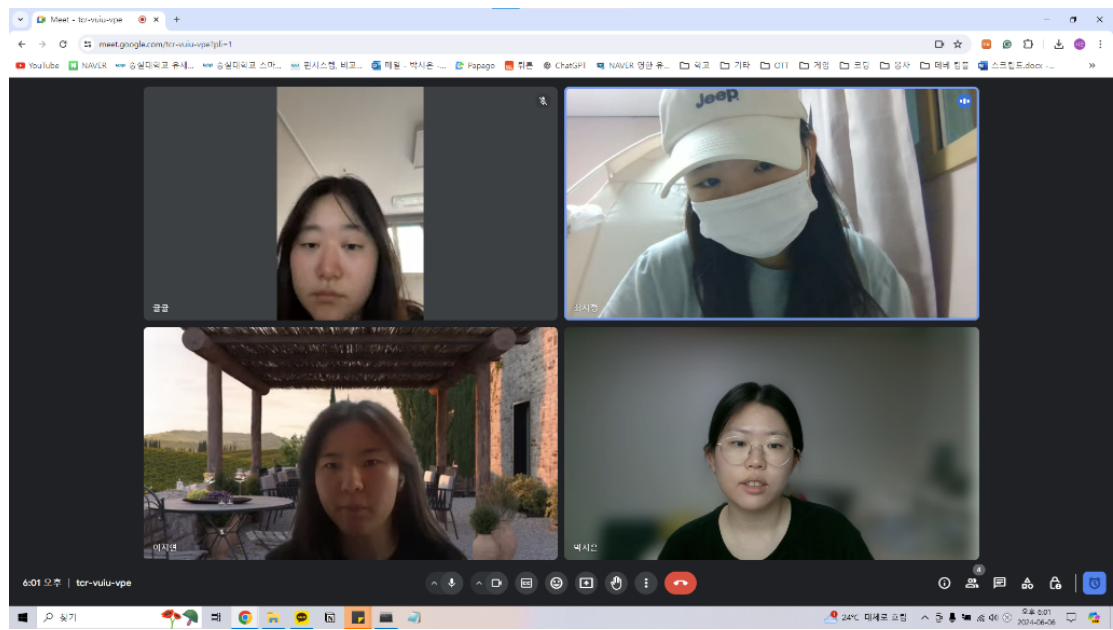


팀 회의록

팀명	베짱이들	참석자	김규린, 박시은, 이지연, 최서정
일시	6/6(목) 18:00	장소	온라인 회의
회의 내용	<p>1. 모델 학습</p> <ul style="list-style-type: none"> - 모델의 예측 성능을 높이기 위하여 하이퍼 파라미터 최적화 진행 - Gridsearch cv를 통해 하이퍼 파라미터 최적화 시도 <p>: 데이터의 객체 수가 너무 많아 시간이 오래 소요됨 => 후보 군을 줄여서 최적화 재시도</p> <p>2. 피드백</p> <ul style="list-style-type: none"> - 상관관계 문제에 대한 교수님의 피드백 반영 <p>1) 시계열 데이터의 경우 학습 데이터와 검증 데이터를 랜덤하게 분할하여 사용할 시에 성능이 높게 측정될 수 있음 : date별로 groupby를 진행하여 train/test set에 date별로 일정하게 분포 되도록 분할</p> <p>2) feature importance 분석을 통해 예측에 도움이 되지 않는 거짓 상관관계를 학습하고 있는 것은 아닌지 확인 : 확인 결과 station, time, temp 순으로 예측에 도움을 주고 있음을 확인</p> <p>3. PPT 초안</p> <ul style="list-style-type: none"> - 데이터 전처리 <p>: target data와 속성들 간의 산점도 그래프를 사용하여 상관관계가 있을 것이라고 생각한 이유 제시</p> <ul style="list-style-type: none"> - 상관관계 히트맵 <p>: 상관관계가 낮게 나오는 이유를 찾지 못해서 속성을 하나씩 제거하며 히트맵을 재구성함</p> <ul style="list-style-type: none"> - test/train set <p>: 시계열 데이터를 학습 시킬 때, 랜덤으로 set을 분리하면 학습이 높게 나오기 때문에, 날짜를 고려하여 일정한 비율로 분배되도록 구성함</p> <ul style="list-style-type: none"> - 모델 학습 <p>: randomforest 회귀 모델을 학습시킴</p> <ul style="list-style-type: none"> - feature importance <p>: 실험 설계 당시에 영향을 줄 것 같은 속성들이 예측에 영향을 많이 주고 있음을 확인</p> <ul style="list-style-type: none"> - 실시간 데이터와의 비교 <p>: 마곡나루 근처의 역 하나의 실시간 데이터를 가지고 와 적정 거치량과 계산을 통해 재배치할 따릉이 개수를 보여줌</p>		

증빙
사진



※ 증빙 사진 혹은 캡처본은 시간과 참여 인원이 나오도록 작성