

제주도 도로 교통량

예측 AI 경진대회

팀명 : 성인ADHD



Table of Contents

01

Data

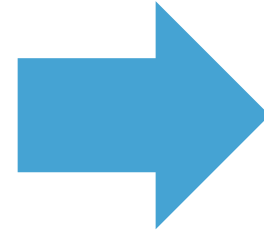
02

Modeling

01. Data

파생변수 - distance(Km 기준)

start_latitude	start_longitude	end_latitude	end_longitude	distance
33.500730	126.529107	33.504811	126.526240	0.525891
33.248505	126.569797	33.248633	126.567766	0.189391
33.485885	126.489979	33.485975	126.486409	0.331239
33.500103	126.512851	33.500132	126.512046	0.074720
33.485704	126.496451	33.483589	126.496368	0.235388



▶ 시작점, 도착점의 거리를 위도, 경도와 haversine을 이용해서 구하고 distance 컬럼 추가

01. Data

파생변수 - 금요일 / 요일별 가중치

✓ 금요일이면 True, 아니면 False를 추가

▶ isfriday 컬럼 추가

day_of_week	isfriday
목	False
일	False
금	True
금	True
토	False

✓ 요일별로 가중치를 줌

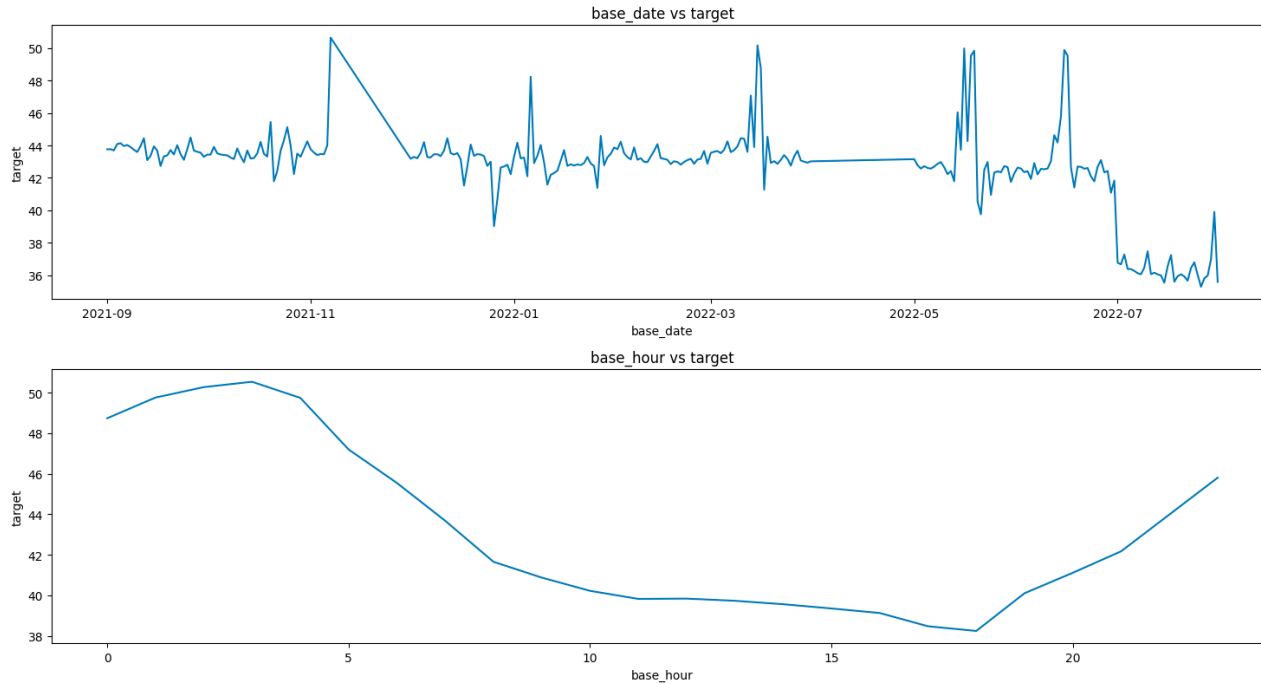
화, 수, 목 : 1 월, 금 : 2 토, 일 : 3

▶ day_weight 컬럼 추가

day_of_week	day_weight
목	1
목	1
일	3
금	2
화	1

01. Data

파생변수 - 상대 유동 지수 hour, day, road, lane, maxspeed, roadrating, roadtype



- ✓ hour, day 그래프를 살펴보면, 각각의 패턴이 존재
이를 바탕으로 각 시간, 날짜별 평균 속도를 구해
상대 유동지수 컬럼추가
그 외에 road, lane, maxspeed, roadrating,
roadtype에도 같은 방법으로 상대유동지수를
구한 후 컬럼 추가

hour_mean_speed day_mean_speed road_mean_speed lane_mean_speed maxspeed_mean roadrating_mean roadtype_mean

42.173431

42.768345

26.400712

44.915713

42.316287

46.388468

41.160930

50.278658

43.179300

26.352858

43.570562

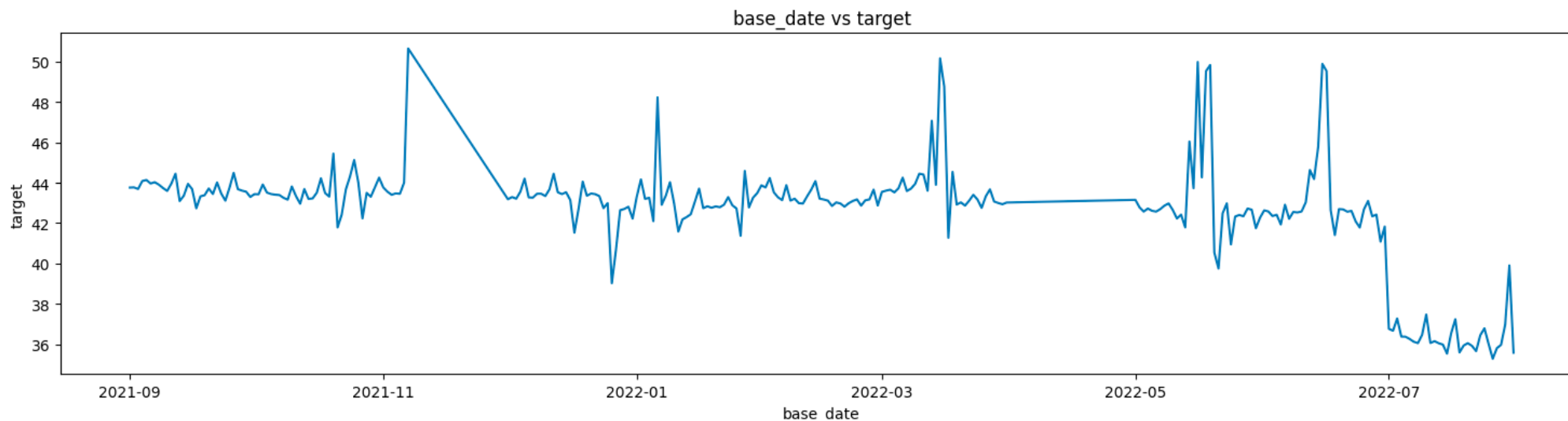
35.492231

34.701408

41.160930

01. Data

#7월 데이터



✓ base_date : 2022년 7월 기준 교통량 증가

✓ 7월의 lane, maxspeed, roadrating, roadtype에 대해서 상대유동지수를 구한 후 컬럼 추가

lane_mean_july_speed maxspeed_mean_july roadrating_mean_july roadtype_mean_july

35.776806

37.977766

39.353222

34.320044

38.590217

31.048812

29.995480

34.320044

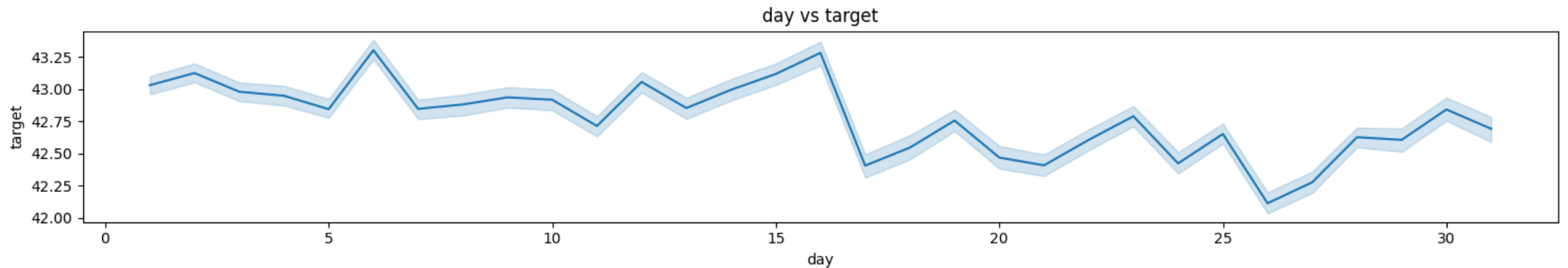
01. Data

학습

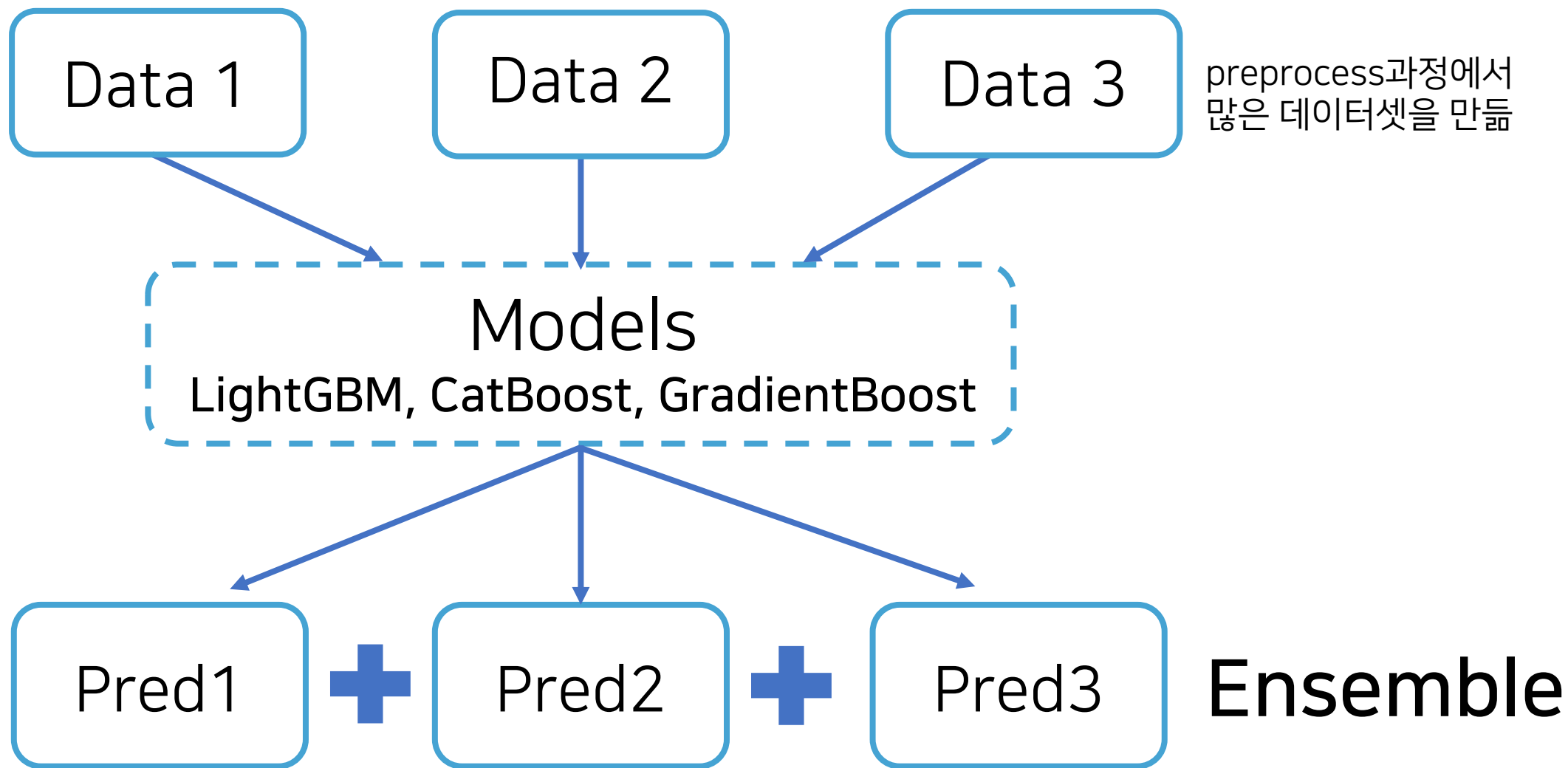
✓ 각 월별로 나눈 후 학습

▶ 7월의 데이터셋으로 학습했을 때가 제일 성능이 좋음

✓ 7월 데이터셋에서 16일 기준으로 나눠봄(16일 기점으로 추이가 다르다고 판단)



02. Modeling



02. Modeling

optuna를 이용해서 최적의 하이퍼파라미터 찾기

LightGBM, CatBoost, GradientBoost 이용

ensemble

결과 값의 mean, median값을 이용해서 도출

감사합니다.