

서초구 교차로 교통량 시계열 예측

참여인원: 2명 (기여도 70%)
2023.10 ~ 2023.11

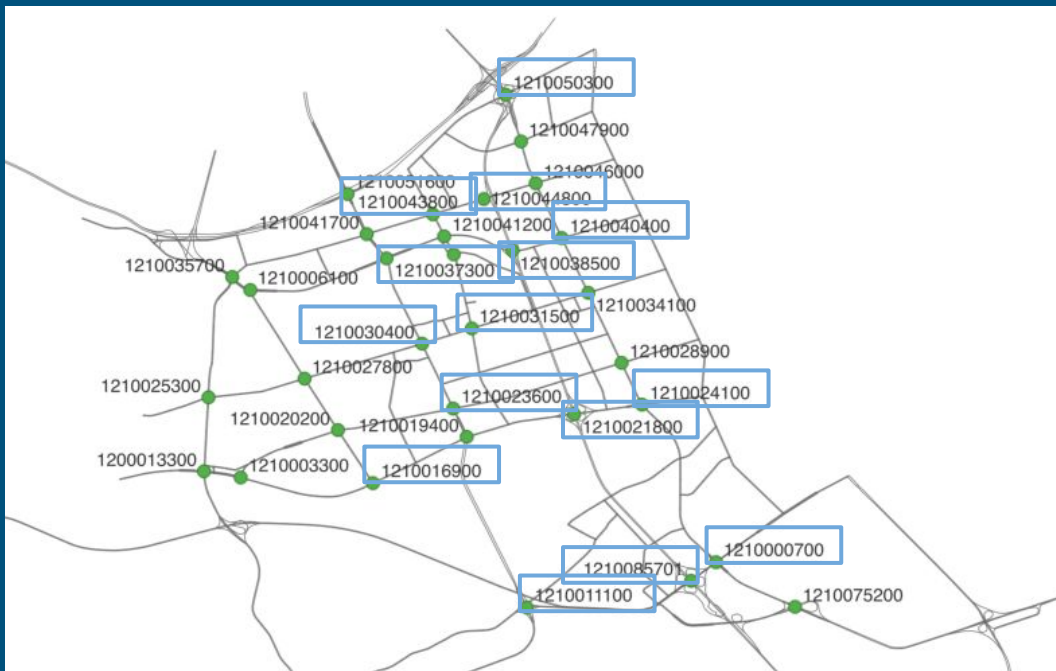
CONTENTS

1. 프로젝트 소개
 2. 데이터
 3. 변수 분석
 4. 전처리
 5. 모델링
 6. 결과
 7. 결론
-

1. 프로젝트 소개

- 프로젝트 목적
 - 서초구 교차로의 실시간 회전교통량을 예측
 - 예측값을 활용해 교통 신호를 제어하여 교통 체증을 줄이고자 함
- 기대효과
 - 운전자들이 운전하는데 드는 시간적, 비용적 부담 감소
 - 연료 감소를 통해 탄소 배출 감소 및 에너지 효율성 향상

1. 프로젝트 소개



[용어 정리]

- 박스권: 교차로
- 교통량 데이터 수집 영역: 회전 방향
- 예측 값: 회전 방향별 5분 누적 교통량

2. 데이터

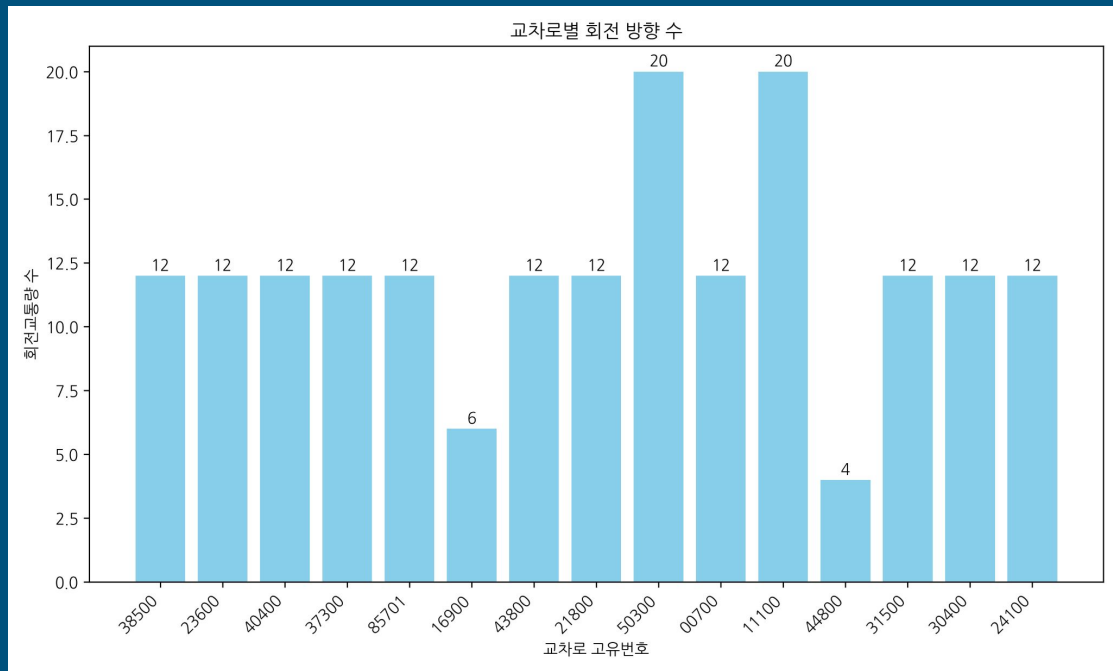
- 전체 데이터에 결측치는 존재하지 않음
- 모든 교차로의 데이터 수집 기간은 동일하나, 회전 방향 수는 차이가 있음
- 2023-07-31 23:50:00 ~ 2023-08-31 23:45:00 동안 총 8,928개 데이터가 수집됨

	Time	direction : 10 -> 20	direction : 10 -> 30	direction : 10 -> 40	direction : 20 -> 10	direction : 20 -> 30	direction : 20 -> 40	direction : 30 -> 10	direction : 30 -> 20	direction : 30 -> 40	direction : 40 -> 10	direction : 40 -> 20	direction : 40 -> 30
0	2023-07-31 23:50:00	4	75	2	0	7	10	82	46	33	1	7	3
1	2023-07-31 23:55:00	0	77	10	0	3	16	89	13	9	1	13	8
2	2023-08-01 00:00:00	3	50	3	0	5	12	67	31	23	1	15	3
3	2023-08-01 00:05:00	1	47	1	0	6	15	62	9	31	0	15	2
4	2023-08-01 00:10:00	1	58	5	0	12	7	70	18	21	2	7	2

[데이터 프레임 출력 결과]

2. 데이터

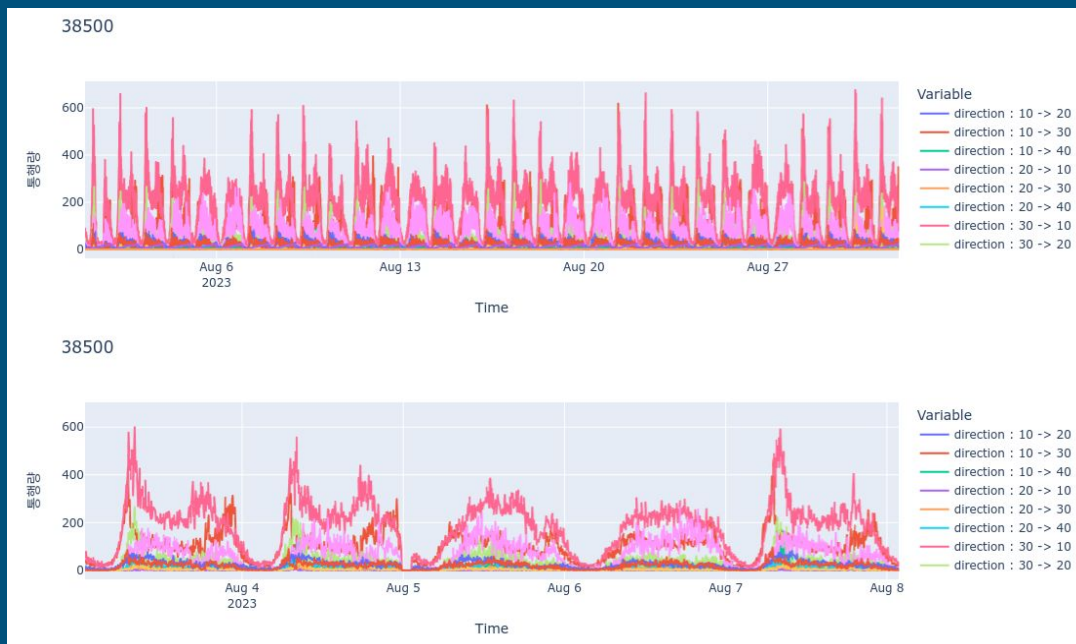
- 교차로별 회전 방향 수 시각화



[교차로별 회전 방향 수 비교]

2. 데이터

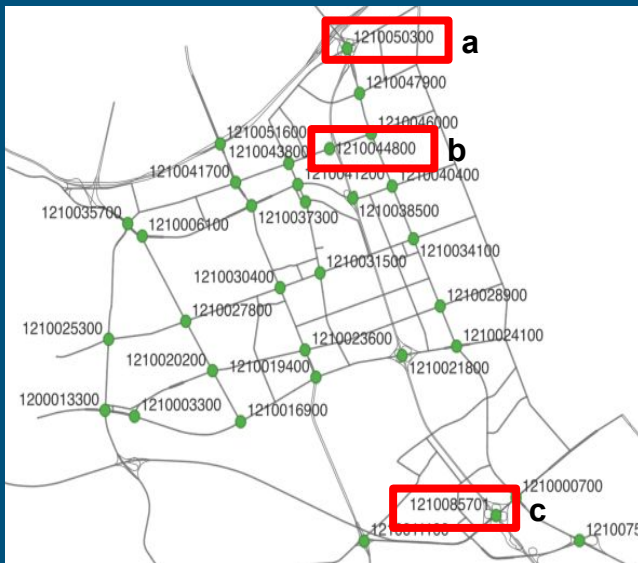
- 38500 교차로의 전체 회전 방향 교통량 시각화



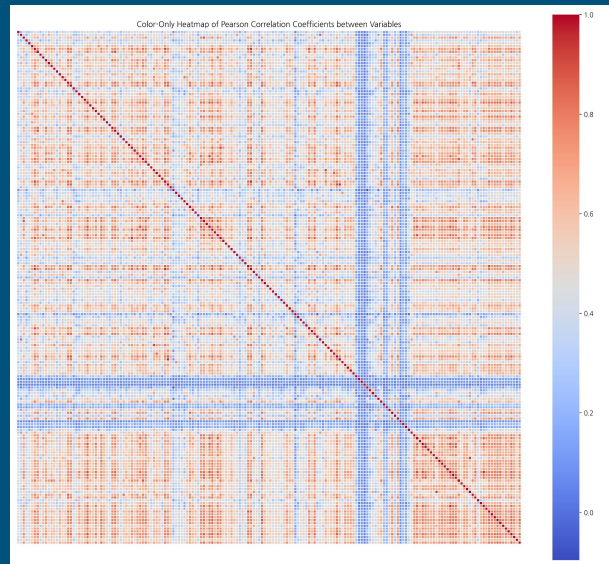
[38500 교차로 교통량 시각화]
위 : 전체 기간 / 아래 : 일부 구간 확대

3. 변수 분석

- [분석 1] 거리가 가까운 교차로 간의 상관관계와 거리가 먼 교차로 간의 상관관계는 차이가 있을까?
 - 차이가 없음. 모든 교차로의 교통량은 대체로 양의 상관관계가 있으며, 위치와는 관계가 없음
 - 교통량이 적은 회전 방향은 상관관계가 낮음



[교차로 간 거리 비교 예시]



[전체 회전 방향의 교통량 상관관계 시각화]

3. 변수 분석

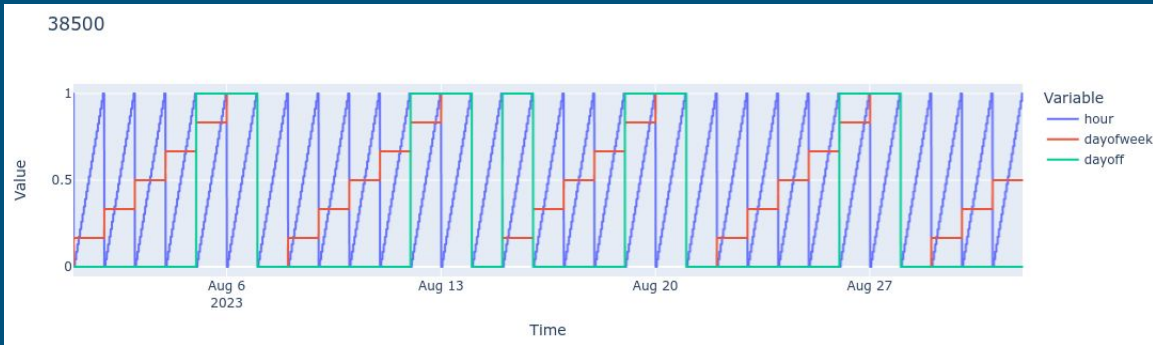
- [분석 2] 그레인저 인과관계 검정(Granger Causality Test)을 통해 예측에 유의미한 변수를 선별하는 것이 효과적일까?
 - 효과적이지 않음
 - 모든 교통량 변수들이 서로 유사한 정보를 공유하고 있으므로, 변수의 다양성이 매우 떨어짐
 - 그레인저 인과관계 검정은 다중공선성을 고려하지 않음
 - 모델 구조의 간소화 측면에서는 효과가 있음
 - 비효율적인 입력 변수 구조를 개선하여 모델의 성능은 유지하면서 모델 구조를 간소화 함

4. 전처리

- 교통량 예측에 유의미한 영향을 미칠 것으로 예상되는 파생변수 생성
 - ‘요일’, ‘시’, ‘휴일’ 총 3가지의 파생변수 생성

```
df_38500['Time'] = pd.to_datetime(df_38500['Time'])

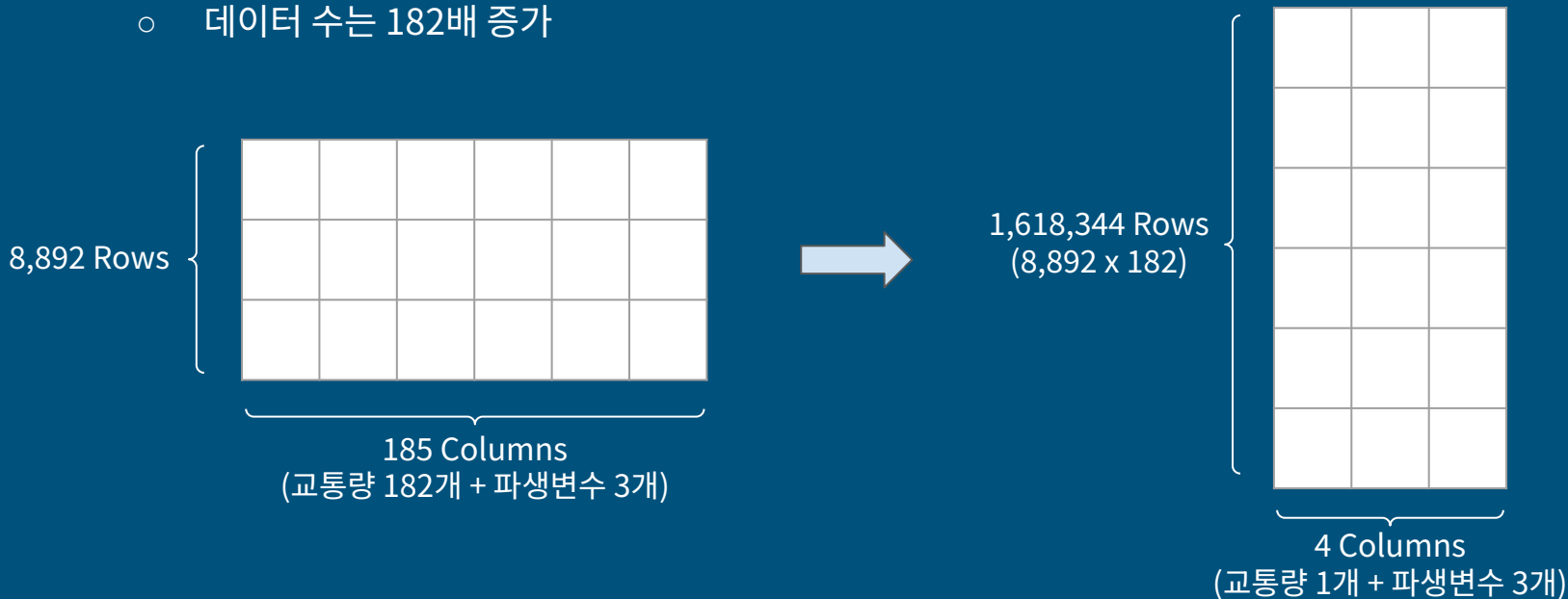
df_38500['dayofweek'] = df_38500['Time'].dt.dayofweek # 0=월요일 6=일요일
df_38500['hour'] = df_38500['Time'].dt.hour # 시간 정보(Hour)
df_38500['dayoff'] = [0 if day < 5 else 1 for day in df_38500['dayofweek']] # 0=평일 1=휴일
idx = df_38500['Time'].dt.day == 15 # 8월 15일은 휴일
df_38500.loc[idx, 'dayoff'] = 1
```



[위: 파생변수 생성 코드 / 아래: 정규화된 파생변수 시각화]

4. 전처리

- 데이터 양이 부족하고(수집기간: 한 달) 모든 교통량 변수 간 상관관계가 높아, 각 회전 방향의 교통량 변수들을 **하나의 '교통량' 변수로 통합**
 - 입력 변수의 수는 182개에서 1개로 감소
 - 데이터 수는 182배 증가



4. 전처리

- 데이터 전처리 주요 정보
 - Window Size: 36(3시간)
 - Target Size: 12(1시간)
 - Step: 1
 - Number of Features: 4(교통량, 요일, 시, 휴일)
 - Data Splitting
 - 훈련 세트: 2023-07-31 23:50:00 ~ 2023-08-22 16:45:00
 - 검증 세트: 2023-08-22 16:50:00 ~ 2023-08-27 08:15:00
 - 테스트 세트: 2023-08-27 08:20:00 ~ 2023-08-31 23:45:00

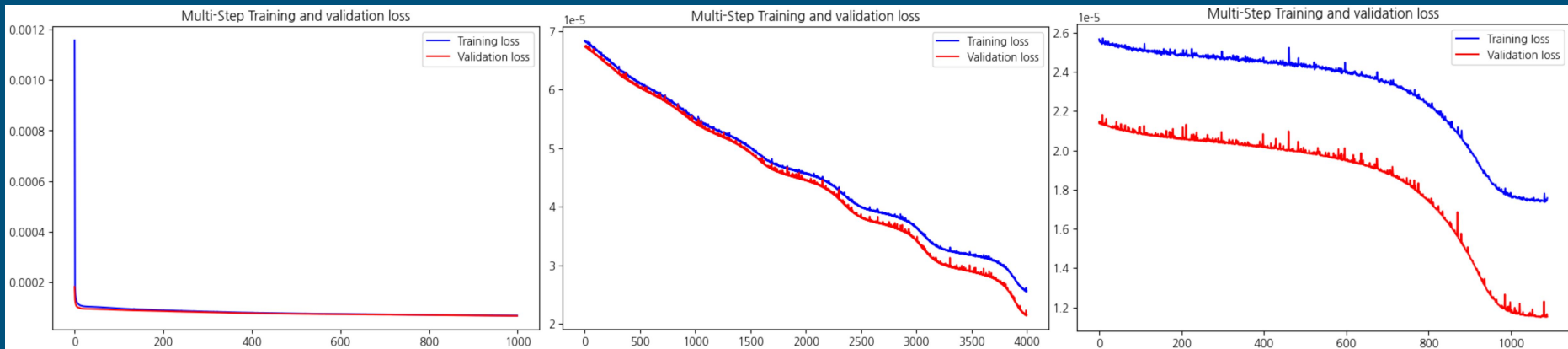
5. 모델링

- 변수를 통합하고 Feature 수가 줄어, 모델 구조는 매우 간소화됨
- 구성한 시계열 모델의 상세 설정은 다음과 같음

모델 주요 구성 요소	상세 내용
Algorithm	LSTM
Input Shape	(36, 4)
Output Shape	(12)
Number of Layers	6
Dropout	0.05
Activation Function(LSTM)	tanh
Activation Function(Dense)	ReLU
Optimizer(Learning Rate)	Adam(0.001)
EarlyStopping	yes(patience=15)
Minimum Learning Rate	0.00001
Loss	mse

5. 모델링

- 기타 세부 정보
 - 훈련용 데이터 세트 수: 1,131,312 개 (Batch Size: 17,676)
 - 검증용 데이터 세트 수: 243,516 개
 - 테스트용 데이터 세트 수: 243,516 개
- 아래는 Epoch에 따른 Training Loss와 Validation Loss를 3개의 구간으로 나누어 시각화 한 그림



[Epoch에 따른 Loss 시각화]

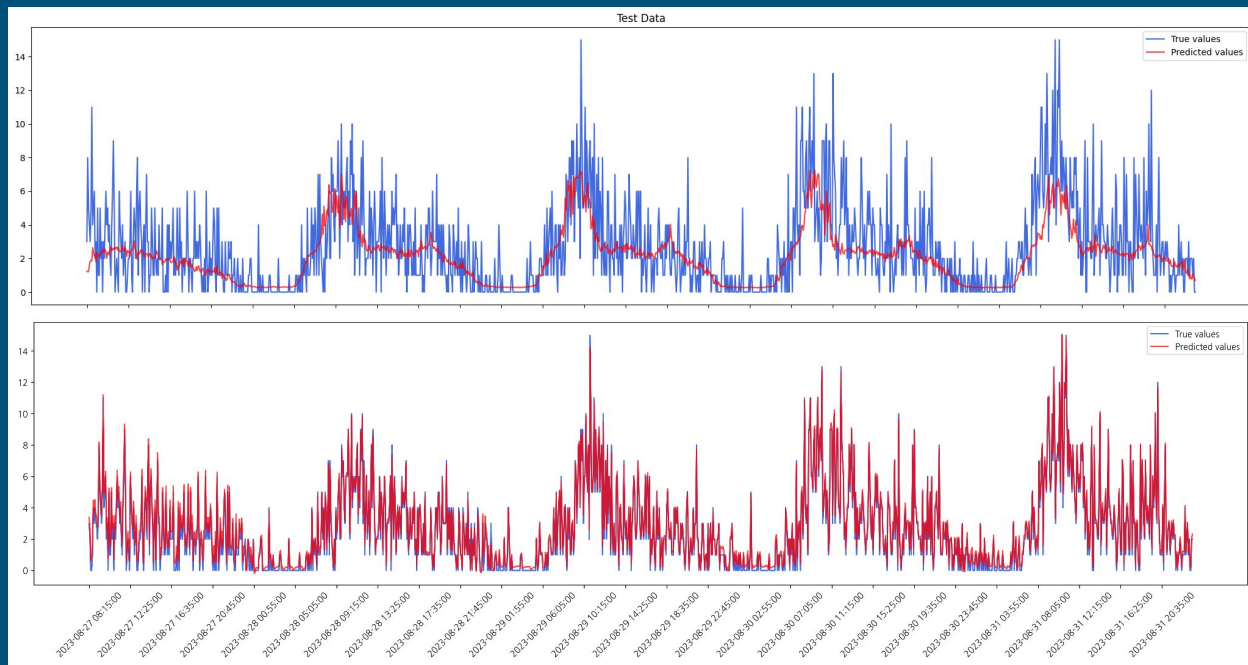
6. 결과

- 5분 후 누적 교통량 예측 성능 비교
 - 모델 1: 전체 교통량 변수를 그대로 사용한 기본 모델
 - 모델 2: 전체 교통량 변수를 통합하여 사용한 변수 통합 모델

분류	비교	기본 모델	변수 통합 모델
모델별 특징	예측 가능 범위	하나의 교차로의 하나의 회전방향	전체 교차로의 전체 회전방향
	알고리즘	GRU	LSTM
	파라미터 수	1,114,641	298,124
예측 성능	MSE	156.55 (± 1206.34)	1.02 (± 30.61)
	RMSE	12.51 (± 12.50)	1.01 (± 1.01)
	MAE	5.81 (± 11.08)	0.48 (± 0.89)

6. 결과

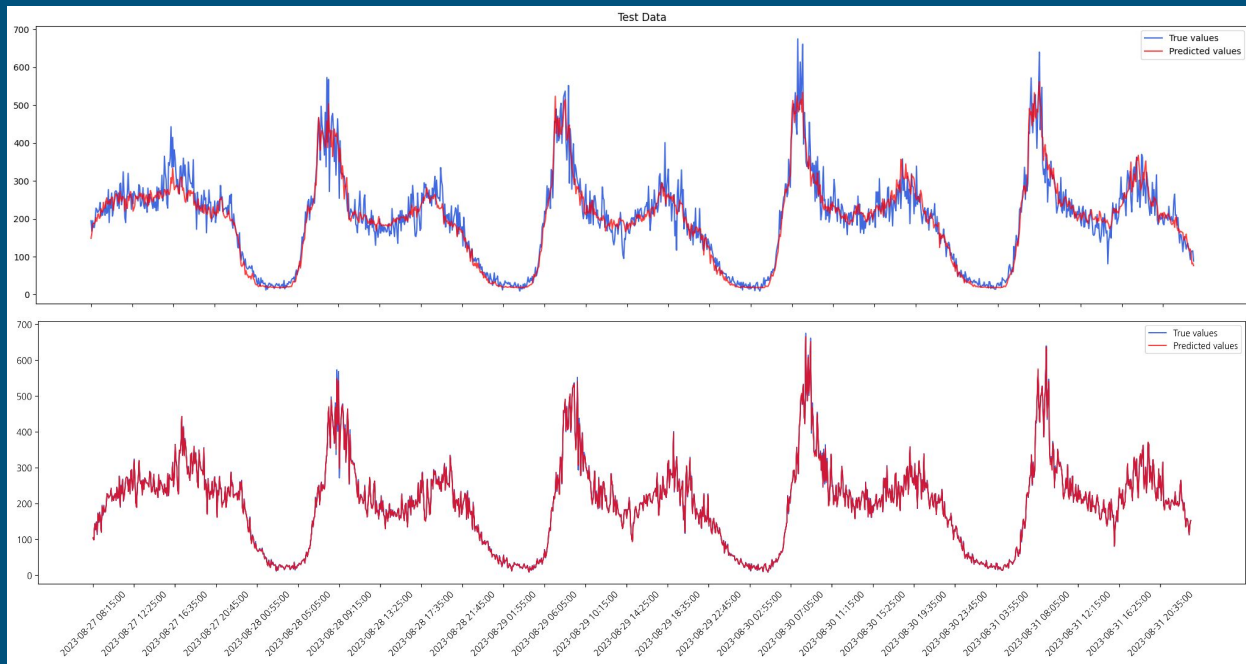
- 50300 교차로, direction : 70 -> 20 회전 방향 예측결과 시각화



[위 : 기본 모델 / 아래 : 변수 통합 모델]

6. 결과

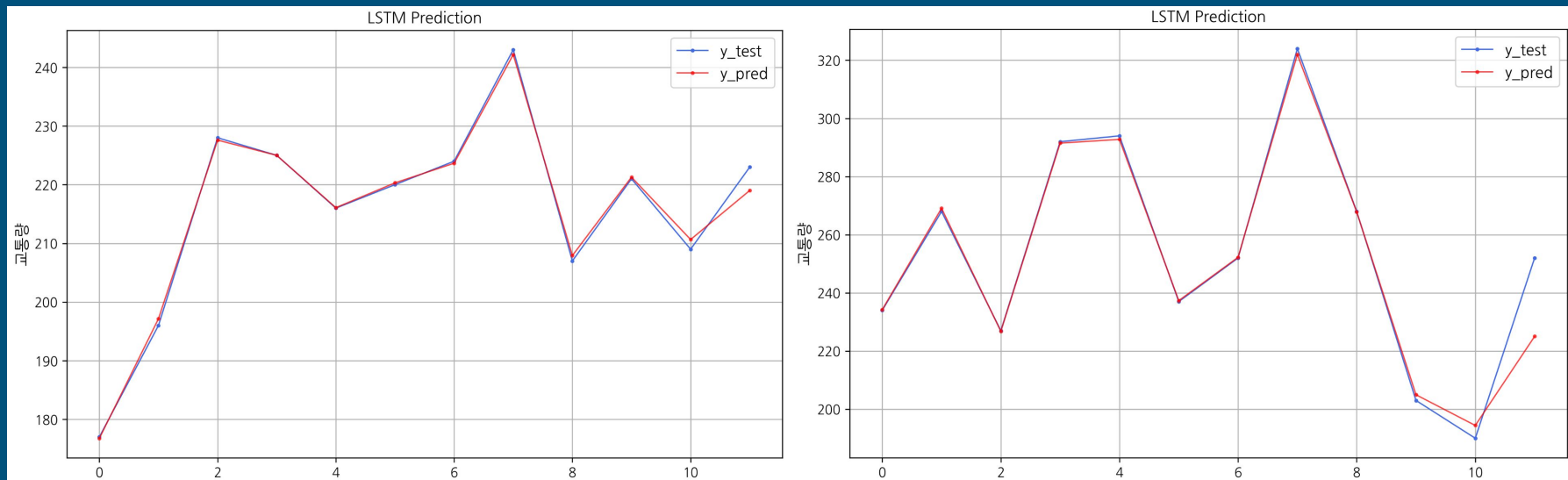
- 38500 교차로, direction : 30 -> 10 회전 방향 예측결과 시각화



[위 : 기본 모델 / 아래 : 변수 통합 모델]

6. 결과

- 최대 예측 구간(1시간)의 예측 결과 시각화
 - 실제값과 예측값을 세부적으로 비교하여 과소예측 및 추종여부 확인
 - 과소예측 및 추종 문제는 발견되지 않음



[1시간 동안의 예측추이 비교 예시]

7. 결론

- 본 프로젝트에서는 서초구 교차로의 실시간 회전교통량을 예측하는 인공지능을 개발함
- 유사한 변수의 수를 줄이고 이를 통해 데이터 양을 증폭시키는 방법이 모델 성능 향상에 크게 기여함
- 예측 결과를 종합하여 살펴보았을 때, 분석에 사용한 서초구 교차로의 5분 누적 교통량은 **예측 가능한 패턴이 존재**하는 것으로 보임
- 그러나 데이터의 수집기간이 짧기 때문에, 계절 변화 등 외부 요인에 따른 통행량 변동을 예측하기 어려움.
또한, 모델 성능의 신뢰성도 아직 충분하지 않음
- 수집 기간 1년 이상의 데이터를 확보한다면 다음과 같은 이점이 있음
 - 파생변수 추가생성
 - 계절 또는 월
 - 샌드위치 데이(휴일 사이의 평일)
 - Target Size를 높일 수 있음
 - 모델 성능의 신뢰도를 향상시킬 수 있음

THANK YOU!