

# 미니프로젝트 6차 시계열 데이터 모델링 상품별 판매량 예측

**AI 02반 05조**

강민구, 김예진, 김채림  
서영현, 장서진, 조민지



# 판매량 예측 성능

구분	알고리즘	Hyper Parameter	Validation 성능 (R2 Score)	Test 성능 (R2 Score)	비즈니스 평가 (재고금액)
3 Beverage	CNN	LSTM(128, 64, activation='tanh', dropout=0.4, 0.4) → Dense(8, activation='relu', dropout=0.4) → Dense(1, activation='linear')   epochs = 100   Optimizer: Adam   Loss: MSE	0.3566	-0.05442	37549.425
12 Milk	LSTM	LSTM(128, 64, activation='tanh', dropout=0.3, 0.3) → Dense(8, activation='relu', dropout=0.4) → Dense(1, activation='linear')   epochs = 100   Optimizer: Adam   Loss: MSE	0.31240..	-0.13659	31090.385
42 Agricultural products	LSTM	lr=0.001, ep = 100 LSTM(128, activation='tanh', return_sequences=False), Dense(1, activation='linear') early_stopping = 10	0.7122	0.01528	37218.46

# 데이터 전처리

- get\_season 함수 정의 : 날짜를 받아 계절 정보를 반환하는 함수
  - 봄: 0[3, 4, 5], 여름: 1[6,7,8], 가을: 2[9,10,11], 겨울: 3[12,1,2]
- 시간 관련 파생 변수 추가
  - 'Weekday(요일)', 'Month(월)', 'Season(계절)', 'WeeklyQtyAvg(주간 계절성)'
- 결측치 처리
  - CustomerCount -> 같은 요일의 중앙값으로 채우기
  - WTI\_Price -> 전날 데이터로 채우되, 첫 번째 행은 다음날 데이터로 채우기
- '동일 지역 방문객 수', '동일 카테고리 판매량 합계', '7일 판매량 이동 평균', '7일 방문자 수 이동 평균 변수' 추가

# 결론

- 관련성

- 요일 : 요일별 판매량과 가장 관련이 높다.  
-> 7일 전 판매량을 반영한다.
- 날짜 : 5일 전 판매량과 상관관계가 있다.  
-> 5일 전 판매량을 반영한다.
- 유가 : 유가는 즉시 반영되지 않고, 한 달 정도 간격을 두고 판매량에 반영된다.  
• 유가가 높아지면 판매량은 감소한다.  
-> 한 달 전 유가를 참고해 판매량에 반영한다.
- 지역 : 같은 City, State의 판매량과 비슷하게 반영된다.  
-> 같은 지역의 판매량을 참고해 판매량에 반영한다.

감사합니다