

2023년 유통데이터 활용 경진대회 최종 제안서

SSSD 생성형 인공지능 모형을 활용한 품목별 판매 수요 예측

정순주, 조수익

I. 서론

본 프로젝트는 생성형 인공지능 모형을 활용하여 유통데이터를 기반으로 향후 판매 수요를 예측함을 목적으로 한다. 이를 위해 정형데이터를 2차원 그래프로 변형하여 모형을 구축함으로써 시계열성을 고려하는 동시에 수량을 예측할 수 있는 그래프 생성 모형을 구현하였다. 이 모형은 외부 데이터 없이 최소한의 데이터로 앞으로의 수요를 예측할 수 있으며, 높은 정확도를 보인다.

II. 본론

1. 분석 목표

실제 산업에 쉽게 활용될 수 있도록 최소한의 데이터를 활용하여 높은 예측 결과를 도출하는 것을 프로젝트 목표로 한다.

2. 데이터 분석 개요

1) Dataset 정의

- 2021년 1월 ~ 2022 6월까지의 지역 물류센터 내 판매 수량 데이터셋

2) Dataset 개괄

- 데이터 개수: 총 521,995개
- 데이터 Columns

이름	설명	예시	유형
판매일	2021년 1월 4일 ~ 2022년 7월 1일	2021-01-04	Datetime
구분	하위 카테고리로 판매유형을 구분	매출/반품	Object
우편번호	지역별 우편번호	37542	Object
판매수량	해당 판매 수량	5	Int64
옵션코드	상품별 포장단위 약어 구분	EA	Object
규격	포장단위 내 상품 용량 및 개수	50g*16*1	Object
입수	포장단위 내 상품 개수	1	Int64
상품 바코드	상품 바코드	8801077334102	Float64

상품명	상품 이름	면도기]프레쉬 708<10입>	Object
-----	-------	------------------	--------

3) Dataset 분석 결과

- 우편번호, 옵션코드, 규격, 입수 컬럼 제외
- 결측치 데이터 제외
- 상품명 기준, 최종 타겟 품목명과 관련 없는 상품 제외
- 판매주차 컬럼 추가
- 이상치 데이터 제외
 - 판매주차 기준, 판매기록이 2번 이하인 데이터 제외
 - 판매주차 기준, 판매수량의 총합이 0인 데이터 제외

	판매일	구분	판매수량	상품 바코드	상품명	판매주차
9	2021-01-04	매출	3	1701001521813	신라면 컵 6入 XX	2021-01
22	2021-01-04	매출	3	8801043045674	농심]안성탕면 컵 6入	2021-01
56	2021-01-04	매출	1	18801045522286	오뚜기]진라면매운멀티<120g>	2021-01
72	2021-01-04	매출	1	8801043015028	농심]너구리 얼큰멀티<40>	2021-01
73	2021-01-04	매출	1	8801043014847	농심]신라면 멀티<40>	2021-01
77	2021-01-04	매출	1	8801043015721	농심]신라면 컵<30>	2021-01
83	2021-01-04	매출	1	1701006157383	코카콜라<500ml*24>	2021-01
90	2021-01-04	매출	1	8801043015271	농심]짜파게티 멀티<40>	2021-01
91	2021-01-04	매출	1	68801056290308	레쓰비<175ml*30>	2021-01
104	2021-01-04	매출	5	8808244208044	삼다수2L	2021-01

[그림 1] 데이터 정제 후 데이터셋

3. 개발 모형

1) 땡땡 모형 소개

본 모형은 SSSD 모형을 기반으로 한 수요 예측 생성형 모형이다.

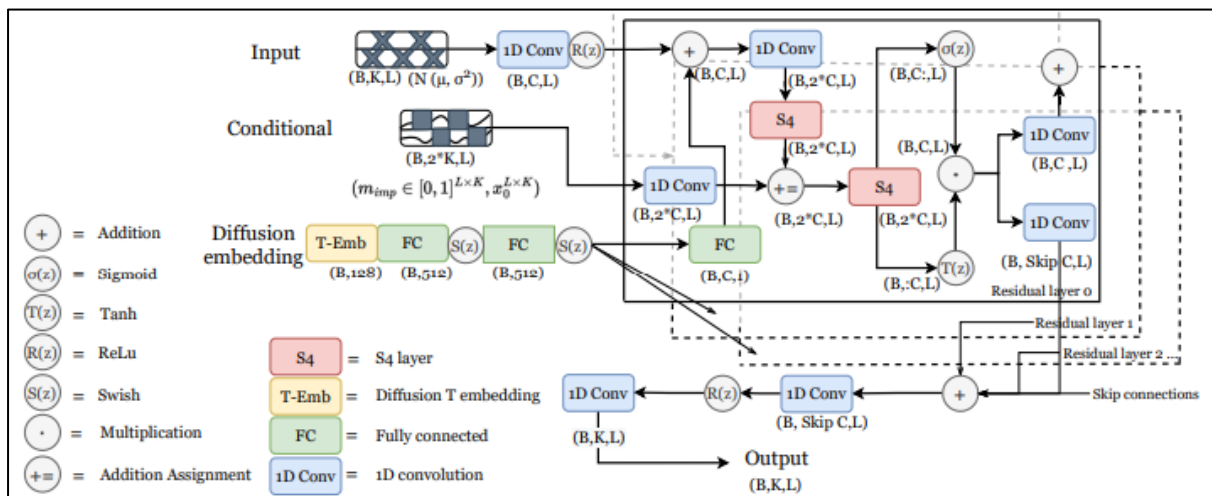
2) SSSD 모형 소개

Structured State Space Diffusion (SSSD) 모형은 생성형 모델인 Diffusion 모형을 기반으로 시계열 데이터를 예측하고 생성하는데 탁월한 모형이다. 이 모형은 Structured State Space와 Diffusion 모형의 혼합체이다.

- Structured State Space 모형: 시계열 데이터의 장기종속성을 효과적으로 모델링하기 위해 상태 공간에서의 입력과 출력 간의 관계를 표현하여 데이터 패턴을 학습시키는 모형.
- Diffusion 모형: 데이터에 잡음을 추가한 후 역방향으로 잡음을 제거하여 데이터를 생성, 예측하는 확률적 생성 모형.

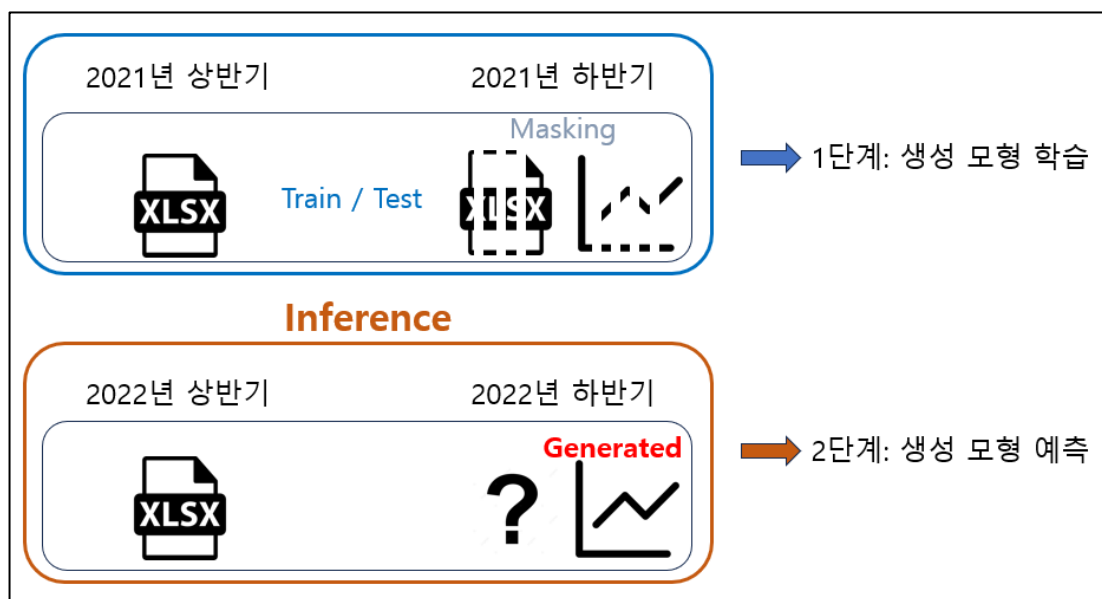
3) 입력 데이터 예시

4) SSSD 모형 구조도



[그림 2] SSSD 모형 구조도

5) 땡땡 모형 학습 구조도

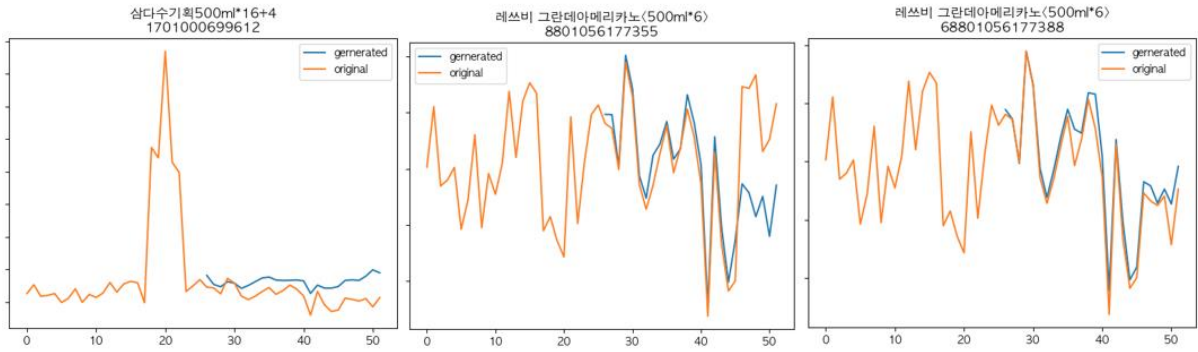


[그림 3] 우리 모형 학습 구조도

III. 결론

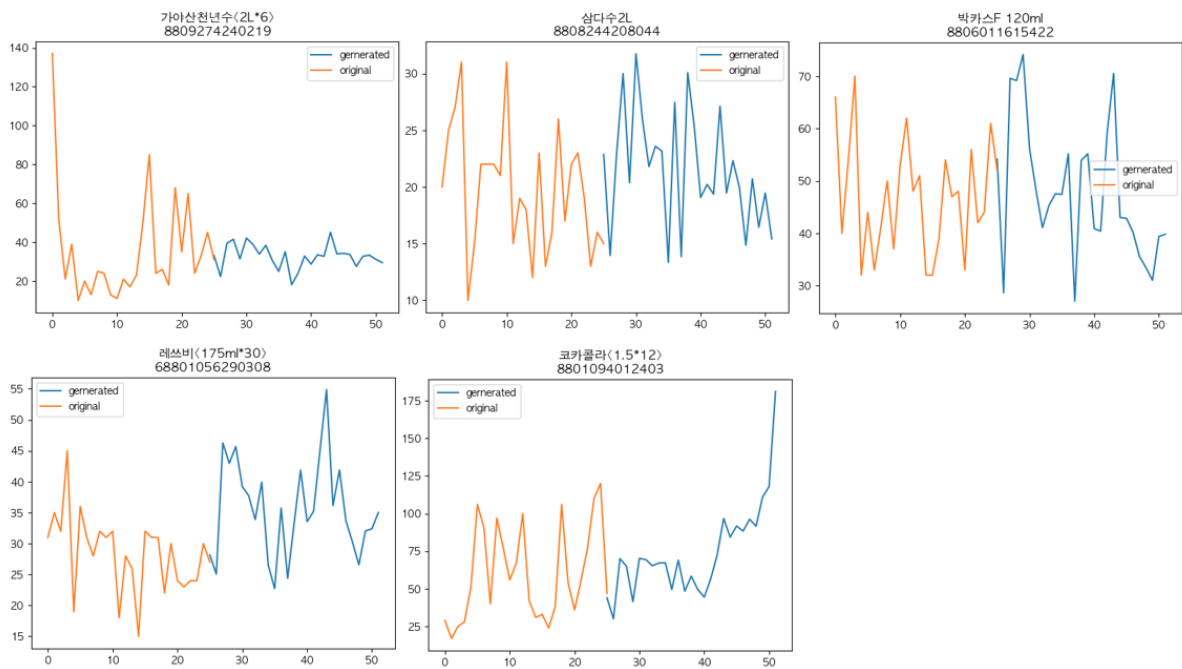
1. 품목별 결과 그래프

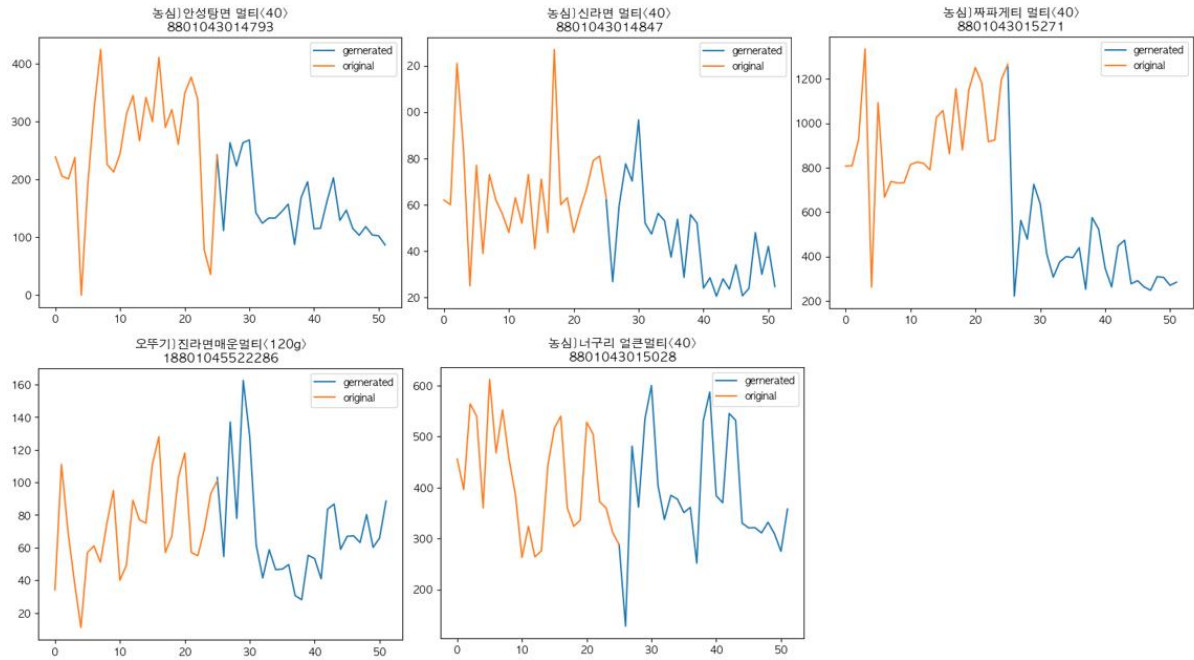
1) 1단계 모형 학습



[그림 4] 품목 중 무작위 3개의 그래프

2) 2단계 모형 예측





[그림 5] Target 품목 10개의 2022년 하반기 수요 예측 생성 그래프 (blue)

2. 성능 평가

2021년 데이터를 활용한 1단계 모형 학습에서 하반기 데이터에 Masking을 사용하여 모형의 견고함을 높이는 동시에 그래프 계형이 잘 맞아지는 학습형태를 보였다. 학습된 모형을 활용하여 2022년 상반기 데이터로 하반기 데이터를 생성하였을 때에도 그래프의 계형이 자연스럽게 생성되었다.

3. 모형 특징

- 1) 외부 데이터가 필요하지 않다.
- 2) 소량 품목의 데이터로도 원하는 품목의 수요 예측이 가능하다.
(사용된 상품 품목 수량: 80개)
- 3) 길지 않은 기간의 데이터로도 원하는 품목의 수요 예측이 가능하다.
(사용된 데이터의 기간: 18개월)

4. 모형 활용 방안

- 1) 미래 수요량에 대한 트렌드 분석뿐 아니라 판매 수량을 제시할 수 있어 재고 관리 비용 절감이 가능하다.
- 2) 시간별, 지역별 수요 예측이 가능하기 때문에 유통분야에서 새로운 접근이 가능

하다.

- 3) 그래프를 활용하여 시각화가 뚜렷하기 때문에 품목별 주기성 관측 및 코로나와 같은 특정 이벤트에 대한 수요 영향을 예측할 수 있다.

참고문헌

Juan M.L. Nils S. Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models, February 2023