

# 의류 판매량 장기 예측을 위한 트랜스포머 기반 시계열 예측 모델

이진우<sup>1</sup> · 박태남<sup>1</sup> · 조용원<sup>1</sup> · 정석윤<sup>2</sup> · 황정현<sup>2</sup> · 안태영<sup>2</sup> · 문혜리<sup>2</sup> · 김환호<sup>2</sup> · 김성범<sup>1\*</sup>

<sup>1</sup>고려대학교 산업경영공학과 / <sup>2</sup>한세실업

## Long-Term Sales Forecasting for Apparel Sales Based on Transformer

Jinwoo Lee<sup>1</sup> · Taenam Park<sup>1</sup> · Yongwon Jo<sup>1</sup> · Seokyun Jeong<sup>2</sup> · Junghyun Hwang<sup>2</sup> · Taeyoung Ahn<sup>2</sup>  
Hyeri Moon<sup>2</sup> · Hwanho Kim<sup>2</sup> · Seoung Bum Kim<sup>1</sup>

<sup>1</sup>Department of Industrial and Management Engineering, Korea University

<sup>2</sup>Hansae Co. Ltd.

Sales forecasting plays an important role in decision-making in various industries. Especially, apparel products are highly sensitive to seasons and trends, making accurate sales forecasting essential. While many studies have applied data-driven models to forecast apparel sales, many have overlooked key characteristics of apparel sales data, such as patterns and distortions. In this study, we propose a transformer-based long-term time series forecasting model using timestamps and denoised input with DILATE (PatchTSDD). The proposed method is designed to capture the unique characteristics of apparel sales. By using denoising autoencoder, the model smooths input data and effectively reflects underlying sales trends. Additionally, the inclusion of time stamp information enables the model to capture global temporal patterns. To address sudden fluctuations in sales, the DILATE loss function is applied. We present experimental results using actual sales data from a leading domestic fashion retail company. The results demonstrate that our method outperforms traditional time series forecasting methods, with each component of the model contributes to the overall performance improvement.

**Keywords:** Sales Forecasting, Time Series Forecasting, Transformer, Apparel Industry

### 1. 서론

수요 예측은 과거 데이터를 기반으로 미래 수요를 예측하는 과정이다. 이는 생산, 재고 관리, 유통 계획 등 기업 운영에 필수적인 요소로, 정확한 예측은 초과 재고나 재고 소진과 관련된 비용을 줄이고, 제품 가용성을 보장하여 고객 만족도를 높이며, 전체 공급망 효율과 기업 수익성을 향상시키는 데 중요한 역할을 한다(Seyedan and Mafakheri, 2020). 수요 예측은 주로 시계열 데이터를 활용한 예측 방법으로 이루어진다. 이는

수요 데이터가 연속적으로 수집되며, 특정 시점의 수요가 과거 수요에 영향을 받기 때문이다.

정확한 수요 예측을 위해서는 판매량 데이터의 특성을 올바르게 반영하는 예측 모델이 필수적이며, 판매량 데이터는 산업 분야와 상품에 따라 그 특성이 크게 달라질 수 있다. 의류 산업은 다양한 외부 요인과 내부 요인으로부터 영향을 받아 정확한 수요 예측이 어려운 분야 중 하나이며, 아래와 같은 특징을 갖는다. 첫째, 의류 상품은 디자인, 원자재 조달, 생산, 운송 등 여러 단계를 거쳐야 하므로 리드타임이 길다는 특징이

\* 연락처 : 김성범 교수, 02841, 서울특별시 성북구 안암로 145 고려대학교 산업경영공학부, Tel: 02-3290-3397, Fax: 02-929-5888, E-mail : sbkim1@korea.ac.kr

2024년 10월 11일 접수; 2024년 11월 13일 수정본 접수; 2024년 11월 18일 게재 확정.

존재한다. 이를 위해 장기적인 수요 예측이 필요하며, 이는 재고 관리 및 생산 계획 최적화에 중요한 영향을 끼친다. 둘째, 의류 산업은 패션 트렌드와 날씨에 민감하게 반응하여 계절성이 두드러진다. 패션 트렌드는 빠르게 변화하며, 특정 스타일이나 디자인에 대한 유행에 따라 판매량이 영향을 받는다. 따라서 매 시즌 고객 선호도가 빠르게 변화하고 트렌드에 맞는 제품이 높은 수요를 기록한다. 또한 날씨 및 온도에 적합한 옷이 달라, 추운 날씨에는 보온성이 좋은 겨울용 상품들이, 더운 날씨에는 반팔, 반바지와 같은 상품들에 대한 수요가 급증한다. 셋째, 의류 판매량은 다양한 요인의 복합적 영향을 받아 급격히 변동하기 때문에 예측이 어렵다.

시계열 예측은 다양한 분야에서 활발히 연구되어 왔으며, 전통적으로는 통계 이론에 기반한 방법론이 주로 사용되었다. 그중 대표적인 모델인 autoregressive integrated moving average(ARIMA)는 자기회귀, 이동 평균, 차분을 결합하여 비정상 시계열 데이터를 정상화(stationary)하여 예측하는 방법이다. 하지만, ARIMA는 복잡한 비선형성 시계열 데이터와 다변량 시계열 데이터에는 적합하지 않다는 단점이 있다 (Kontopoulou *et al.*, 2023). 이러한 한계점과 함께 데이터 양이 방대해지면서, 최근에는 딥러닝 모델을 활용한 시계열 예측 연구가 주를 이루고 있다.

순환 신경망(recurrent neural network, RNN)은 시계열과 같은 순차적 데이터를 처리하는 데 특화된 인공 신경망 구조이다. 그러나 RNN은 긴 시퀀스 데이터를 처리할 때 오차역전파 과정에서 기울기 소실(vanishing gradient) 문제가 발생하여 적합하지 않다. 이를 해결하기 위해 등장한 long short-term memory (LSTM; Hochreiter and Schmidhuber, 1997)는 게이트(gate)를 통해 중요한 정보만 기억하여 긴 시퀀스 시간적 의존성을 학습할 수 있게 했다. Sequence-to-sequence(Seq2Seq; Sutskever *et al.*, 2014)는 여러 LSTM 셀로 이루어진 인코더와 디코더 구성된 신경망 구조로 긴 시퀀스를 보다 효과적으로 처리할 수 있었다. 하지만 Seq2Seq 모델은 시퀀스가 길어질수록 초반에 입력된 정보가 사라질 위험이 있다.

이러한 한계를 극복하기 위해 어텐션 메커니즘(attention mechanism; Luong *et al.*, 2015)이 제안되었다. 어텐션은 입력 시퀀스 내 중요한 부분에 가중치를 부여함으로써, 긴 시퀀스에도 중요한 정보가 손실되지 않도록 한다. 최근에는 어텐션을 발전시킨 트랜스포머(transformer; Vaswani *et al.*, 2017) 계열 모델이 시계열 예측 분야에서 각광받고 있는 추세이다. Patch time series transformer(PatchTST; Nie *et al.*, 2022)는 입력 시계열 데이터를 패치(patch)들로 나누어 지역적 정보를 반영하고 연산 효율을 높인 특징을 갖고 있다. 또한, 다변량 시계열 데이터를 여러 단변량 데이터로 분리하는 채널 독립(channel independence)을 적용해 각 변수를 독립적으로 학습함으로써 장기 예측 성능을 더욱 향상시켰다.

본 연구에서는 의류 상품 판매량 데이터의 특징을 반영한 장기 시계열 예측을 수행하는 프레임워크인 PatchTSDD(PatchTST

using Time Stamp and Denoised input with Dilate)를 제안한다. 본 연구에서 사용하는 데이터셋은 크게 세 가지 특성을 지니고 있다. 첫째, 의류 상품은 판매되기까지 긴 리드타임을 갖고 있어 장기적인 수요 예측이 필수적이다. 이에 따라, 제안방법은 장기 시계열 예측에서 뛰어난 성능을 보이는 트랜스포머 계열 모델인 PatchTST를 기본 백본 모델로 활용한다. 둘째, 의류 판매량 데이터는 패션 트렌드와 날씨 변화에 민감해 추세와 계절성이 뚜렷하다. 이러한 판매량 패턴을 효과적으로 학습하기 위해 디노이징 오토인코더(denoising autoencoder)를 사용하여 입력 데이터를 스무딩(smoothing)하고, 시점 정보를 인코딩해 입력으로 사용하였다. 셋째, 판매량이 급격하게 증가하거나 감소하는 변동 구간들이 존재한다. 이러한 급격한 변동을 정확하게 예측하기 위해, 시계열 모델과 시점 변동을 고려한 DILATE(Guen *et al.*, 2019) 손실함수를 적용하였다. 본 연구에 대한 주요 기여점은 다음과 같다.

- 제안 방법론인 PatchTSDD는 장기 시계열 예측에 적합하며, 패턴과 변동성을 효과적으로 예측하여 기존 시계열 모델 대비 우수한 성능을 보였다.
- 기존 시계열 예측 방법에서 사용한 스무딩 기법들과는 달리 디노이징 오토인코더를 활용해 추세를 효과적으로 학습하였으며, 시퀀스 내 데이터 크기(scale)를 고려한 노이즈 주입 방식을 제안하였다.
- 일반 손실함수로는 예측하기 어려운 급격한 변화를 예측하기 위해 DILATE 손실 함수를 적용하여 예측 성능을 향상시켰다.

본 논문은 다음과 같은 구성으로 이루어져 있다. 제2장에서는 의류 판매량 예측을 수행한 관련 연구들에 대해서 소개한다. 제3장에서는 제안된 모델 구조와 각 구성 요소들을 자세히 설명한다. 제4장에서는 실험에 활용된 데이터 및 설정에 대해 소개하고, 결과를 통해 제안방법론의 우수한 성능을 보여준다. 마지막 제5장에서는 본 연구에 대한 결론 및 향후 연구 방향에 대해 제시한다.

## 2. 관련 연구

의류 상품은 긴 리드타임, 패션 트렌드, 그리고 제품 특성별로 다른 판매 패턴을 보이기 때문에 예측이 어려운 상품 중 하나이다(Swaminathan *et al.*, 2024). 기존에는 도메인 전문가가 경험을 바탕으로 미래 판매량을 예측하거나, 간단한 통계 기법을 통해 예측하는 방식이 주로 사용되었다. 그러나 이러한 방법은 예측 정확도가 떨어지고, 전문가의 역량에 따라 예측 결과가 달라질 수 있다는 한계가 있다. 최근에는 빅데이터 및 인공지능을 바탕으로 정교한 시계열 예측 기법이 발전하면서, 의류 상품 판매량 예측에도 이러한 최신 기술이 도입되고 있다.

## 2.1 정성적 기법을 활용한 의류 판매량 예측

과거에는 의류 상품 판매량 예측에 정성적 기법이 주로 활용되었다. 정성적 기법은 전문가로부터 인사이트를 통해 데이터로는 포착하기 어려운 최신 트렌드, 변화를 반영할 수 있다는 장점이 존재한다. 또한 과거 데이터가 부족한 상황에서도 유연한 대처가 가능하다(Frank *et al.*, 2003).

Green and Harrison(1973)은 과거 판매 데이터가 없는 드레스의 판매량을 예측하기 위해 사내 여성 직원들에게 여러 드레스들 중 구매할 드레스를 투표하도록 하였고, 이후 이 투표 정보와 드레스 가격 간 선형회귀분석(linear regression)을 통해 초기 판매량 추정치를 도출하였다. Mostard *et al.*(2011)은 수요 관련 정보를 보다 풍부하게 반영하기 위해 전문가뿐만 아니라 소비자의 의견도 반영하여 단일 기간 수요를 예측하였다. 이를 위해 실제 판매 이전에 소비자들로부터 사전 주문 데이터를 수집하고, 이를 구매 관리자나 전문가의 의견과 함께 사용하였다. 이러한 정성적 기법은 전문가 의견에 크게 의존하므로, 그들의 역량에 따라 예측 성능 편차가 발생할 수 있다는 단점이 있다.

## 2.2 전통 시계열 기법을 활용한 의류 판매량 예측

과거 데이터를 바탕으로 정량적으로 판매량을 예측하기 위해 ARIMA, 이동평균(moving average), 지수평활법(exponential smoothing) 등 전통적인 시계열 기법을 활용한 연구 역시 활발히 진행되어왔다(Liu *et al.*, 2013).

Anggraeni *et al.*(2015)은 무슬림 아동복 수요 예측을 위해 ARIMA와 ARIMA with exogenous variables(ARIMAX) 모델을 비교하였다. ARIMAX는 ARIMA 구조에 종교적 행사 및 기후 요인들과 같은 변수들을 사용하여 ARIMA 대비 높은 성능을 달성하였다. Kusuma *et al.*(2020)은 인도네시아 전통 의상인 바틱(batik) 판매량 예측에 최근 데이터에 높은 가중치를 부여하는 선형 지수 평활법(linear exponential smoothing)을 사용하였다. Lu *et al.*(2018)은 Seasonal ARIMA(SARIMA)모델과 홀트-윈터스 지수평활법(Holt-Winter's Exponential Smoothing)을 사용해 의류 판매량을 예측하였다. 이때 홀트-윈터스 지수평활법이 계절적 패턴을 예측하는데 효과적이며, SARIMA 모델은 단기 예측에서 높은 정확도를 보인다는 결과를 확인하였다. Choi *et al.*(2011)은 웨이블릿 변환(wavelet transform)을 통해 시계열 데이터를 분해하고, SARIMA 모델의 입력으로 활용하여, 기존 SARIMA 대비 우수한 예측 성능을 달성하였다. Lenort *et al.*(2013)은 상품 판매량을 예측하는데 있어 박스-젠킨스(Box-Jenkins) 기반 계층적 예측을 통해 예측 오차를 줄여 성능을 향상시켰다. 이 연구에서 판매량을 제품, 카테고리, 회사 전체 판매량을 계층적으로 정의하고, 이를 하향식(top down) 및 상향식(bottom up) 방식으로 예측하는 방법을 제안했다. 통계적 방법은 결과 도출이 빨라 자주 사용되었지만, 의류 판매 데이터내 불

규칙한 패턴으로 인해 정교한 예측이 어려워 우수한 예측 결과를 달성하기에는 한계가 있다(Liu *et al.*, 2013).

## 2.3 딥러닝을 활용한 의류 판매량 예측

복잡한 패턴을 갖는 의류 판매량 데이터를 보다 정확하게 예측하기 위해 인공지능이 활용되고 있다. Güven *et al.*(2020)은 기본적인 딥러닝 모델인 인공신경망(artificial neural network)을 사용하여 의류 판매량을 예측한 결과, 서포트 벡터 머신(support vector machine) 보다 우수한 성능을 보였다. 딥러닝 모델은 다변량 모델로 활용이 가능하기 때문에 판매량 데이터의 다양한 변수를 사용할 수 있다는 장점이 있다. Han *et al.*(2021)은 기상 변수와 요일 정보를 활용한 인공신경망을 통해 해당 변수들이 여성 골프 의류 판매량에 미치는 영향을 분석했다. 연구 결과, 봄과 가을에는 기상요인이, 여름과 가을에는 요일 변수가 중요한 영향을 미친다는 것을 발견하였다. Giri *et al.*(2019)은 새로운 의류 상품의 판매량을 예측하기 위해 제품 이미지에서 특징을 추출하고 이를 과거 판매 데이터와 함께 활용한 인공신경망 모델을 제안하였다. 그러나 이러한 연구들은 시계열 데이터에 특화된 모델을 사용하지 않아 비선형적이며, 어려운 패턴을 학습하는 데 한계가 있다.

판매 패턴을 보다 정교하게 예측하기 위해 복잡한 딥러닝 모델을 활용하여 높은 예측 정확도를 달성한 연구들도 존재한다. Xu *et al.*(2021)은 합성곱 신경망(convolutional neural network, CNN)과 다층 퍼셉트론(multi-layer perceptron, MLP)을 결합한 모델을 활용하여 스포츠웨어 판매량을 예측하였다. CNN은 판매량 시계열 데이터의 시간적 패턴을 학습하고, MLP를 통해 상품의 특성, 가게 정보와 같은 범주형 변수(categorical variable)로부터 정보를 추출하여 CNN과 MLP를 개별적으로 사용했을 때 대비 우수한 성능을 보였다. Luo *et al.*(2022)은 extreme deep factorization machine(XDeepFM)과 LSTM을 결합한 모델을 제안하여 향후 30일간 판매량을 예측하였다. xDeepFM 모델은 판매에 영향을 끼치는 의류 특성 간의 상관관계를 파악해 예측하고, LSTM은 잔차(residual)을 보정함으로써 전통 시계열 기법과 머신러닝, LSTM 모델보다 우수한 성능을 달성할 수 있었다. Li *et al.*(2021)은 gated recurrent unit(GRU; Cho *et al.*, 2014)과 Prophet 모델을 함께 사용한 복합 모델을 제안하였다. 나아가 중요 정보를 강조하는 어텐션 메커니즘을 추가하여 의류 판매량 예측 정확도를 향상시켰다. Skenderi *et al.*(2024)은 새로운 의류 상품 판매량 예측을 위해 이미지 데이터와 구글 트렌드 정보를 활용한 트랜스포머 기반 예측 모델을 제안하였다. 이러한 모델은 변화가 빠른 의류 산업에서 소비자의 관심을 정확하게 파악하고, 이를 반영해 예측 성능을 개선할 수 있었다. 하지만 시계열 딥러닝 모델을 활용한 기존 연구들은 데이터 내 발생하는 판매량의 급격한 증가, 감소와 같은 변동성을 고려하지 못하였다.

최근에는 딥러닝 모델에 다양한 변수를 활용하는 연구가 주

를 이루고 있다. 대부분의 연구는 한 달 이내의 짧은 기간 판매량을 예측하거나 아직 판매되지 않은 제품의 판매량을 예측하는 데 중점을 두고 있다. 이로 인해 기존 판매되던 상품에 대한 장기 판매량 예측에 활용되기엔 어려움이 있다. 제안방법은 시계열 예측에 특화된 모델을 활용하여 과거 판매량을 효과적으로 학습해 리드타임을 고려한 장기 예측이 가능하다. 또한, 전문가 의견에 의존하지 않고, 판매량 데이터만을 심층적으로 분석함으로써 의류 판매량 데이터에서 나타나는 패턴을 정확하게 포착할 수 있다. 나아가 경제 지표, 날씨 정보와 같이 판매량과 연관성이 높은 변수들은 실제 의류 산업에서 이를 주기적으로 수집하여 활용하는 것은 현실적으로 어려움이 존재한다. 제안방법은 의류 판매량이 가지는 특징을 반영함으로써 추가 변수를 확보하기 어려운 상황에서도 판매 패턴을 정밀하게 예측할 수 있다는 장점이 있다. 나아가 기존 시계열 딥러닝 연구들에서 다루지 못한 판매량 데이터 내 급격한 변동성을 효과적으로 반영하여 장기 시계열 예측 성능을 향상시키는 것을 목표로 한다.

### 3. 방법론

본 연구는 시계열 데이터의 계절성과 급격한 변동성을 반영하는 장기 시계열 예측 모델인 PatchTSDD를 제안한다. 의류 상품 패턴을 포착하기 위해, 입력 데이터를 재구성하여 스무딩하는 디노이징 오토인코더와 시점 정보를 입력 데이터로 함께 활용한다. 이 입력 데이터는 장기예측에 적합하다고 알려진 트랜스포머 모델 기반인 PatchTST에 입력되어 미래 여러 시점을 예측한다. 마지막으로 데이터내 급격한 변동성을 효과적으로 예측하기 위해 학습 시 DILATE 손실함수를 사용하였다. 제안방법은 학습 시 디노이징 오토인코더를 우선적으로 학습하고, 학습된 디노이징 오토인코더로 재구성한 입력 데이터를 활용해 PatchTST를 학습한다. <Figure 1>은 제안방법에 대한 전체적인 구조를 도식화한 그림이다.

#### 3.1 디노이징 오토인코더 입력 재구성을 통한 스무딩 효과

본 연구에서는 디노이징 오토인코더를 활용해 입력 데이터

내 잡음과 변동을 완화해 스무딩 효과를 주어 추세를 효과적으로 학습하고자 한다. 스무딩은 데이터 내 노이즈를 줄여 패턴이나 추세를 더 명확하게 표현하는 기법이다. Smyl(2020), Wibawa *et al.*(2022)은 지수평활법을 거친 입력 데이터를 딥러닝 모델에 입력으로 활용한 시계열 예측을 진행하였다. 디노이징 오토인코더는 입력 데이터에 노이즈(noise)를 추가해 저차원 특징을 생성하고, 이를 다시 입력 데이터와 같은 형태로 복원함으로써 중요한 특징을 추출하는 딥러닝 모델이다. 이때 재구성 과정에서 노이즈가 제거되는 특징과 이를 바탕으로 유용한 특징을 추출 수 있다는 장점이 있다. Xiong *et al.*(2016)은 디노이징 오토인코더로 심박수 시그널(ECG) 데이터 내 노이즈를 없앴으며, Dai *et al.*(2017)은 전력 상태 모니터링을 위해 디노이징 오토인코더를 사용해 손상된 전력 데이터 내 이상치를 제거하였다.

제안방법은 다층 퍼셉트론 구조를 기반으로 하는 디노이징 오토인코더로 사용하였다. 디노이징 오토인코더는 길이가  $T$ 인 입력 시퀀스  $(x_1, \dots, x_T)$ 에 노이즈를 추가해  $(\tilde{x}_1, \dots, \tilde{x}_T)$ 를 만들고, 이를 원본 입력 시퀀스  $(x_1, \dots, x_T)$ 를 예측하도록 하여  $(\hat{x}_1, \dots, \hat{x}_T)$ 를 복원한다. 이때 복원된 시퀀스  $(\hat{x}_1, \dots, \hat{x}_T)$ 는 원본 입력 시퀀스의 추세와 관련 없는 변동이 줄어든 입력 시퀀스이다. 디노이징 오토인코더로 재구성된  $(\hat{x}_1, \dots, \hat{x}_T)$ 은 이후 미래 시점 데이터를 예측하는 PatchTST 모델에 입력으로 사용된다.

본 연구에서 사용한 디노이징 오토인코더는 입력 시퀀스마다 데이터 크기가 다를 수 있기 때문에 입력 데이터의 분포를 고려해 아래와 같은 방식으로 노이즈를 주입하였다. 아래 식에서  $\epsilon$ 은 입력 데이터에 최종적으로 가해지는 잡음을 뜻한다. 또한  $\mu_s$ 는 입력 시퀀스 내 평균,  $\delta$ 는 노이즈 크기를 조절하는 척도인자(scale factor),  $\sigma$ 는  $N(0,1)$ 을 따르는 가우시안 노이즈(Gaussian noise)를 나타낸다.

$$(\tilde{x}_1, \dots, \tilde{x}_T) = (x_1, \dots, x_T) \times \epsilon \quad (1)$$

$$\epsilon = \mu_s \times \delta \times \sigma \quad (2)$$

최종적으로 재구성된 입력  $(\hat{x}_1, \dots, \hat{x}_T)$ 과 실제 입력  $(x_1, \dots, x_T)$  차이를 줄이기 위한  $L_{recon}$  손실함수를 최소화하는 방

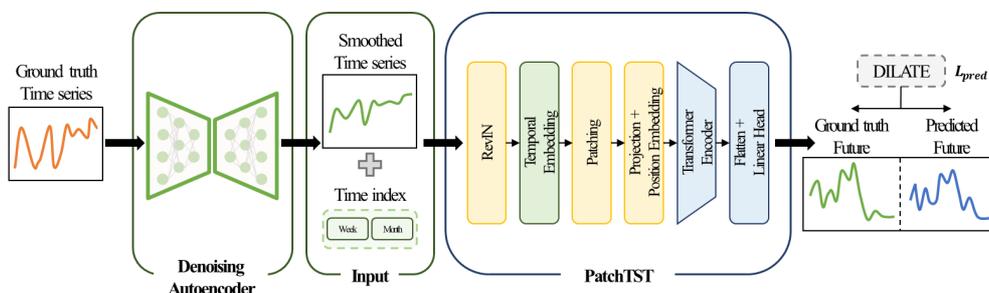


Figure 1. Overall Architecture of PatchTSDD

향으로 디노이징 오토인코더를 학습한다.  $L_{recon}$ 을 표현한 아래 식 (3)은  $g_1$ 은 디노이징 오토인코더를 나타내며,  $L$ 은 평균 제곱 오차(mean squared error)이다.

$$L_{recon} = \sum_{t=1}^T L(x_t, g_1(x_t)) \quad (3)$$

### 3.2 시점 정보를 활용한 계절성 반영

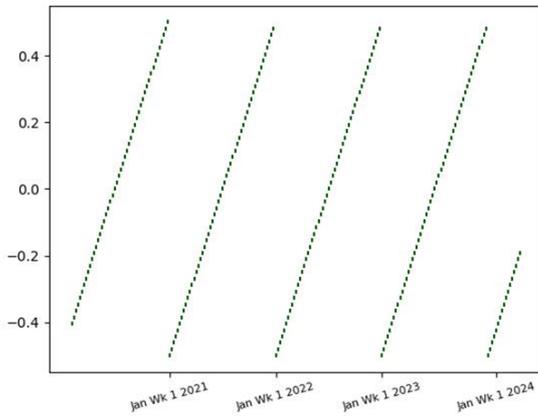
본 연구에서는 모델이 시계열적 특성과 패턴을 반영하도록 시간 정보를 입력 시퀀스와 함께 추가적으로 사용한다. 날짜, 월, 연도 등 시점 정보(time stamp)를 입력으로 사용하는 것은 각 데이터에 대한 시간적 정보를 제공하며, 시계열 패턴을 파악하는데 유용할 수 있다. Informer(Zhou *et al.*, 2021)는 장기 예측에서 전역적인 시간 정보를 파악하기 위해 시점 정보를 입력 데이터와 결합한 임베딩(embedding) 방식을 활용하였다.

인코딩 방식은 입력 데이터를 일년 단위로 처리할 때, 주(week)와 월(month) 시점 정보인  $w_t$ 와  $m_t$ 를 아래 식 (4), (5)와 같이 주기성을 갖는  $E_{w_t}$ ,  $E_{m_t}$ 로 변환하는 과정이다.

$$E_{w_t} = \left( \frac{w_t - 1}{52} \right) - 0.5 \quad (4)$$

$$E_{m_t} = \left( \frac{m_t - 1}{12} \right) - 0.5 \quad (5)$$

<Figure 2>는 본 연구에서 사용하는 데이터에 대한 주와 월 시점 정보  $w_t$ ,  $m_t$ 를 인코딩한  $E_{w_t}$ ,  $E_{m_t}$ 을 시각화한 결과이다. (a)는 주 정보, (b)는 월 정보를 나타낸 것이며, 일년을 주기로 반복되는 특징을 확인할 수 있다. 주와 달 시점 정보를 인코딩한  $E_w = (E_{w_1}, \dots, E_{w_T})$ ,  $E_m = (E_{m_1}, \dots, E_{m_T})$ 를 디노이징 오토인코더로 스무딩된  $(\hat{x}_1, \dots, \hat{x}_T)$ 와 함께 PatchTST에 대한 입력 데이터로 활용한다.



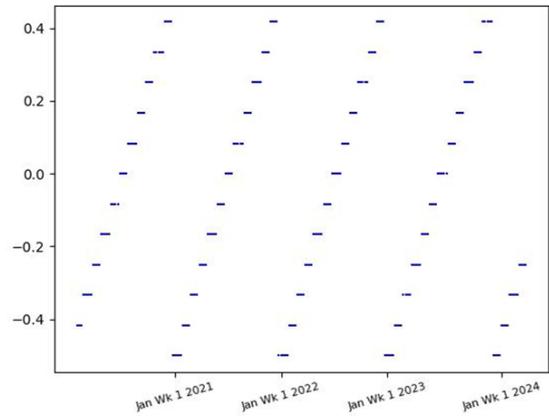
(a) Encoded week time stamp

### 3.3 PatchTST를 통한 장기 예측

본 연구에서는 리드타임으로 인한 장기 수요 예측을 위해 PatchTST(Nie *et al.*, 2022) 모델을 사용한다. PatchTST는 패칭(patching)을 통해 입력 시퀀스를 나누어 활용함으로써 지역적 정보(local semantic information)를 파악하며, 메모리 및 연산 효율적에서도 장점을 갖고 있다. 이러한 특성으로 인해 PatchTST는 다양한 분야에서 멀티스텝 예측(multi-step forecasting)에 활용되고 있다(Huang *et al.*, 2024; Lemishko and Landi, 2024; Xue *et al.*, 2024). 제안 모델에서 PatchTST는 디노이징 오토인코더로부터 재구성한 시퀀스  $(\hat{x}_1, \dots, \hat{x}_T)$ 와 시점 정보  $E_w = (E_{w_1}, \dots, E_{w_T})$ ,  $E_m = (E_{m_1}, \dots, E_{m_T})$ 를 입력 받아 미래  $k$ 개 시점 시퀀스  $(\hat{y}_1, \dots, \hat{y}_k)$ 를 예측한다.

PatchTST에 입력된 재구성 시퀀스  $(\hat{x}_1, \dots, \hat{x}_T)$ 는 시계열 예측에서 사용되는 정규화 방법인 reversible instance normalization(RevIN; Kim *et al.*, 2021)을 거친다. RevIN은 입력 시퀀스 별로 인스턴스 정규화(instance normalization)를 진행하여 비정상성(non-stationary) 정보를 제거하고, 시퀀스 간 분포 차이 문제를 해결하여 일정한 평균과 분산을 갖게 한다. 모델은 최종적으로 정규화된 입력 값을 바탕으로 예측을 수행하며, 이후 다시 비정규화(denormalization)를 통해 비정상성을 복원해 본래 시계열 데이터가 갖고 있던 특성을 반영한다.

일반적인 PatchTST와는 달리, 제안 방법인 PatchTSDD는 시점 정보와 스무딩된 입력 시퀀스를 함께 활용하기 위해 입력 데이터와 함께 인코딩하는 임베딩 레이어를 추가하였다. Zhou *et al.*(2021)에서 사용한 방식처럼 RevIN을 통해 정규화된 입력 데이터와 시점 정보를 합하는 임베딩 방식을 거친다. 입력 시퀀스와 시점 정보를 임베딩한 시퀀스  $(\hat{x}_1, \dots, \hat{x}_T)$ 는 패칭을 통해 길이가  $L$ 인  $N$ 개인 패치  $\hat{x}_i \in R^{L \times N}$ 를 생성한다. 이때 패치 개수는  $L = \left\lceil \frac{T-P}{S} \right\rceil + 2$  만큼 만들어지며 해당 수식에서  $S$ 는 보폭



(b) Encoded month time stamp

Figure 2. Examples of Encoded Time Stamp

(stride)을 뜻한다. 패칭을 활용하면 입력 시퀀스의 지역적 정보를 반영할 수 있고, 트랜스포머 모델에 입력되는 길이가  $T/S$ 만큼 감소한다. 패칭을 사용함으로써 각 패치에 대해 병렬적으로 연산이 가능하기 때문에 연산 복잡도가  $O(L^2)$ 에서  $O(\frac{L^2}{S^2})$ 로 연산 효율성이 높아진다(Nie *et al.*, 2022). 패치 시퀀스들은 선형 투영(linear projection)  $W_l \in R^{D \times N}$ 과 트랜스포머에서 사용된 포지셔널 인코딩(positional encoding)  $W_p \in R^{D \times N}$ 을 적용한다. 포지셔널 인코딩은 패치 내 데이터간 순서를 반영하며, 이 과정을 식(6)으로 표현할 수 있다.

$$\hat{x}_d = W_l \hat{x}_l + W_p \in R^{D \times N} \quad (6)$$

포지셔널 인코딩을 거친 입력 패치  $\hat{x}_d$ 는 PatchTST의 기본 트랜스포머 인코더와 셀프 어텐션을 통해 잠재 표현 벡터  $z \in R^{D \times N}$ 로 표현된다. 이후, 트랜스포머 인코더를 통해 얻은 잠재 표현 벡터  $z$ 는 디코더 부분에 해당하는 flatten 연산과 선형 계층(linear layer)을 통과해 길이가  $k$ 개인 미래 시계열  $\hat{y}_i = (\hat{y}_1, \dots, \hat{y}_k)$ 을 예측한다.

### 3.4 DILATE 손실함수를 통한 변동성 반영

본 연구에서 사용한 데이터는 급격한 변동이 포함된 구간들이 존재하기 때문에 이를 정확하게 예측하기 위해 DILATE(Guen *et al.*, 2019) 손실함수를 사용하였다. DILATE는 시계열 예측을 위한 손실함수로, 예측과 정답 시계열간 변동폭을 포함하는 전체적인 형태와 시점 차이를 고려한다. 이러한 특징으로 일반적인 평균 제곱 오차 손실함수와 달리 급격한 변동이 존재하는 시계열 데이터에 대해 더 정확한 예측을 가능하게 한다. Mazen *et al.*(2023)은 태양광 발전량 데이터의 무작위 변동(random fluctuations)을 예측하기 위해 DILATE 손실함수

로 모델을 학습하였다.

DILATE는 길이가  $k$ 인 정답 시계열  $y_i = (y_1, \dots, y_k)$ 과 예측 시계열  $\hat{y}_i = (\hat{y}_1, \dots, \hat{y}_k)$  간 형태와 시점 변동을 나타내는 dynamic time warping(DTW; Sakoe *et al.*, 1978) 기반 손실함수를 하이퍼파라미터인  $\alpha$ 로 선형결한 형태로 아래 식(7)과 같다. 여기서  $L_{shape}(y_i, \hat{y}_i)$ 는 변동을 포함한 시계열 데이터의 형태적 유사성을 나타내고,  $L_{temporal}(y_i, \hat{y}_i)$ 는 시간적 차이를 반영한 손실함수이다. 이때  $L_{shape}(y_i, \hat{y}_i)$ 는 미분 가능한 형태인 SoftDTW(SDTW; Cuturi *et al.*, 2017)를 활용해 형태 변동을 계산하며,  $L_{temporal}(y_i, \hat{y}_i)$ 는 Frías-Paredes *et al.*(2017), Vallance *et al.*(2017)에서 사용한 time distortion index(TDI)를 발전시킨 방식으로 시간적 차이를 계산한다.

$$L_{DILATE}(y_i, \hat{y}_i) = \alpha L_{shape}(y_i, \hat{y}_i) + (1 - \alpha) L_{temporal}(y_i, \hat{y}_i) \quad (7)$$

## 4. 실험

### 4.1 데이터 소개

본 연구에서는 국내 의류 제조 전문 기업인 A기업에서 수집한 실제 의류 상품(레깅스) 판매량 데이터로, 해당 데이터는 2020년 2월부터 2024년 3월까지 221주 동안 주차 별 판매량을 담고 있다. 전체 상품들 중 최소 2-3년에 대한 판매 데이터가 누적된 83개 핵심 레깅스 상품이 포함되었다. 각 레깅스 상품에 대한 4가지 상품 특성 변수가 기록되었으며 이는 상품코드, 스타일, 색상, 그리고 사이즈 정보이다. <Figure 3>은 본 연구에서 사용한 데이터를 시각화한 예시이다.

실험에 사용된 예측 모델은 최소 단위인 개별 상품 단위로 판매량 예측한다. 이때 각 상품은 서로 다른 판매 패턴을 가지

product variables				weekly sales data						
Code (DPCI)	Style	Color	Size	1 <sup>st</sup> week Feb 2020	2 <sup>nd</sup> week Feb 2020	3 <sup>rd</sup> week Feb 2020	...	5 <sup>th</sup> week Jan 2024	1 <sup>st</sup> week Feb 2024	2 <sup>nd</sup> week Feb 2024
xxx- xxx - xxx	A	White	X Small	9	4	0	...	33	0	13
xxx- xxx - xxx	A	White	Small	15	23	0	...	8	0	10
xxx- xxx - xxx	B	Blue	Medium	44	13	0	...	23	0	57
xxx- xxx - xxx	B	Blue	X Large	1	6	10	...	0	0	2
⋮		⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮
xxx- xxx - xxx	C	Green	Large	37	102	354	...	212	287	349
xxx- xxx - xxx	C	Green	X Large	113	159	0	...	80	141	48

Figure 3. Example of Weekly Legging Sales Data

**Table 1.** Summary of Train/Valid/Test Dataset for Each Style Data

Style	Number of Product	Number of Train Data	Number of Validation Data	Number of Test Data	Number of Total Data
Style A	24	2,391	435	24	2,850
Style B	18	2,340	876	41	5,635
Style C	41	4,718	432	18	2,790
All	83	9,449	1,743	83	11,275

고 있어, 개별적으로 모델을 학습시키는 경우 학습 비용이 많이 들고 학습에 사용할 데이터가 부족하다는 단점이 있다. 이러한 문제를 해결하기 위해, 본 연구에서는 동일한 스타일, 즉 같은 레깅스 종류로 상품을 그룹화하여 모델을 학습시키는 방식을 사용하였다.

본 연구는 슬라이딩 윈도우(sliding window)기법을 활용하여 데이터 셋을 구축하였다. 제안방법은 윈도우 길이가 24인 판매량 데이터를 입력 받아, 향후 24주동안 판매량을 예측한다. 이는 의류 산업에서 상품이 실제 판매되기까지 걸리는 리드타임을 고려하여 24주로 설정하였다. 의류 산업은 각 계절에 따른 판매량의 변동성이 큰 분야이다. 24주는 계절 단위를 넘어서는 시간 범위이며, 공급망과 관련된 계획 조정을 수행하는데 충분한 시간적 여유를 제공한다. 따라서 장기 예측을 위해 적합한 기간이라 판단하였다. 학습, 검증, 테스트 데이터 분할 시, 마지막 24주(2023년 10월 4주 ~ 2024년 3월 4주) 데이터를 테스트 데이터로 사용하였으며, 나머지 데이터는 각각 0.85, 0.15 비율로 학습과 검증 데이터로 나누었다. 앞서 언급한 바와 같이, 실험에 사용된 모델들은 스타일 별 데이터를 학습하므로, 각 스타일별 학습, 검증, 테스트 데이터는 <Table 1>과 같이 구성되었다. 이때 <Table 1>에는 총 가지 3 스타일이 존재하는데, 이는 길이에 따른 3가지 레깅스 스타일 A(5부), B(7부), C(9부)이다.

#### 4.2 실험 설계 및 평가지표

제안방법과 다른 시계열 예측 방법론들을 비교하기 위해 동일한 학습 세부사항 및 환경에서 실험을 진행하였다. 모든 모델은 총 500 에폭(epoch)에 걸쳐 학습을 진행하였으며, 최적화를 위해 AdamW(adaptive moment estimation with weight decay)를 사용하고, 학습률(learning rate)과 배치 사이즈(batch size)는 각각 0.001과 128로 설정하였다. 모든 실험은 5회 진행하였으며, 반복 실험 성능을 평균 내어 다른 모델들과 비교하였다. 제안방법은 디노이징 오토인코더에서 노이즈에 곱해지는 척도인자  $\delta$ 로 0.1을 사용하였으며, DILATE 손실함수에서 모양과 변동 손실함수 간 가중치를 조절하는 하이퍼파라미터인  $\alpha$ 와 스무딩 파라미터(smoothing parameter)  $\gamma$ 는 각각 0.7과 0.05로 설정하였다.  $\alpha$ 를 0.7로 설정한 것은 모양 변동성, 즉 급격한 증가, 감소에 대한 가중치를 높이기 위함이다.

본 논문에서는 PatchTSDD 모델의 예측 성능을 정량적으

로 평가하기 위해 평균 절대 오차(mean absolute error; MAE)와 가중 절대 백분율 오차(weighted average percentage error; WAPE)를 지표로 사용하였다. MAE(식 (8))는 판매량 크기를 기준으로 산출하므로 오차를 직관적으로 파악할 수 있고, WAPE(식 (9))는 비율 형태로 제공되기 다른 모델들과 비교가 용이하여 두 지표를 통해 성능을 평가하였다. WAPE는 예측 구간에 대한 실제 값의 합을 사용하기 때문에, 실제 값이 0일 때도 백분율로 오차를 표현할 수 있는 장점이 있다. 식 (8)과 (9)에서  $y_i$ 는 실제 값,  $\hat{y}_i$ 는 예측 값 그리고  $n$ 은 관측치 개수를 의미한다. MAE와 WAPE는 모두 0 이상의 값을 가지며, 두 평가지표 모두 값이 작을수록 모델의 성능이 우수함을 나타낸다. 최종적으로 각 스타일별 데이터내 상품들에 대한 평균 MAE와 WAPE를 바탕으로 모델 간 성능을 비교하였다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

$$WAPE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i|} \quad (9)$$

#### 4.3 실험 결과

<Table 2> 제안방법론인 PatchTSDD가 장기 시계열 예측에서 우수한 성능을 보임을 입증하기 위해 시계열 예측에서 널리 사용되는 통계 기반 및 딥러닝 모델과의 비교 실험 결과를 제시하였다. 통계 기반 모델로는 ARIMA, 순환 신경망 계열 딥러닝 모델인 LSTM(Hochreiter *et al.*, 1997), GRU(Cho *et al.*, 2014), Seq2Seq Attention(Luong *et al.*, 2015), 합성곱 신경망을 활용한 CNN-LSTM(SHI *et al.*, 2015), Temporal Convolutional Network(TCN; Bai *et al.*, 2018), 트랜스포머 계열 시계열 모델인 Informer(Zhou *et al.*, 2021), Autoformer(Wu *et al.*, 2021), PatchTST(Huang *et al.*, 2024) 그리고 DLinear(Zeng *et al.*, 2023)을 사용한 예측 모델 성능과 비교한 결과이다. 이때 각 상품 종류 데이터 별로 성능이 가장 높은 모델을 볼드체로 강조하여 표시하였다.

<Table 2>는 본 연구에서 제안한 PatchTSDD가 다른 시계열 예측 모델들과 비교하였을 때, 가장 우수한 성능을 달성

**Table 2.** Comparison Results among Time Series Forecasting Models. Forecasting Results are Evaluated with MAE and WAPE. Bold Represents the Model with Best Performance

Model	MAE (↓)				WAPE 100(%) (↓)			
	Style A	Style B	Style C	All	Style A	Style B	Style C	All
ARIMA	1040.35	100.63	1502.34	1053.14	96.05	116.42	69.11	93.86
LSTM	827.30	173.34	1825.02	1178.33	99.01	181.60	73.10	117.90
GRU	846.45	208.65	1826.44	1192.22	93.15	216.59	71.19	127.22
CNN-LSTM	566.15	106.40	1255.16	806.80	56.65	151.81	53.62	75.79
TCN	492.91	81.00	1160.53	733.37	44.91	122.49	51.39	64.93
Seq2Seq Attention	513.97	130.19	1436.28	891.55	54.45	159.98	62.35	92.26
Informer	488.13	68.92	1211.26	754.43	50.17	102.25	53.57	68.66
Autoformer	555.00	217.10	1252.87	826.45	58.00	186.21	54.30	99.50
DLinear	679.19	322.08	1346.12	931.19	86.84	291.50	65.03	147.79
PatchTST	721.00	55.04	1208.15	817.22	64.47	77.17	56.48	66.04
PatchTSDD(Proposed)	<b>424.94</b>	<b>41.83</b>	<b>1050.72</b>	<b>650.98</b>	<b>40.94</b>	<b>59.37</b>	<b>50.9</b>	<b>49.86</b>

했음을 보여준다. 특히 스타일이 A와 B 상품에 대해 제안방법 PatchTSDD가 월등한 성능을 보인다. 스타일 A와 B는 여름철에 판매량이 높은 상품으로, 이는 PatchTSDD가 이러한 계절성 및 패턴을 정확히 반영하였다는 것을 의미한다. 또한 제안방법은 PatchTST 대비 스타일 A, B, C 상품에 대해 각각 41.06%, 24%, 13.03%의 MAE 성능 향상을 보여, PatchTST에 추가된 구성요소들이 성능 향상에 효과적임으로 입증하였다. <Figure 4>는 각 스타일별 특정 상품에 대한 예측 결과를 시각화한 그래프이다. <Figure 4>에서 스타일 A와 C 상품에 대한 그래프를 보면, PatchTST는 실제 값 대비 적은 수량으로 예측하는 반면 PatchTSDD는 테스트 구간에서 판매량이 최고치인 부분을 근사한 값으로 예측하였다. 이는 PatchTSDD가 DILATE 손실함수를 사용함으로써 PatchTST 보다 변동을 더 정확하게 예측했음을 의미한다. 전반적으로 트랜스포머 계열 모델들과 합성곱 신경망 계열 모델들이 순환 신경망 계열 모델 대비 좋은 성능을 보여주었다. Informer와 TCN은 스타일 A와 C 상품에 대해 높은 예측 성능을 보였지만, 스타일 B 상품에 대해 WAPE가 100% 이상으로 매우 낮은 예측 정확도를 보였다. 이는 <Figure 4> 3, 4행에서 나타나듯, 스타일 B 상품의 학습 데이터와 테스트 데이터 간에 분포 차이가 존재하기 때문이다. 따라서 이러한 분포 차이 문제를 완화하기 위해 정규화 기법인 RevIN을 적용한 PatchTSDD와 PatchTST는 스타일 B에 대해 우수한 성능을 보여 주었다. 또한, <Figure 4>에서 스타일 A 상품의 예측 결과를 통해 PatchTSDD는 판매량이 증가하는 부분도 정확하게 예측하며, 실제 판매량과 유사한 예측 결과를 도출했다. 이는 PatchTSDD가 디노이징 오토인코더와 시점 정보를 효과적으로 활용하여 다른 모델들에 비해 판매량 데이터 내 패턴을 더 잘 학습했음을 의미한다.

#### 4.4 구성 요소별 성능 기여도

<Table 3>은 제안방법론인 PatchTSDD에 사용한 디노이징 오토인코더, 시점 정보, DILATE 손실함수가 최종 성능에 미치는 개별적인 효과를 검증하기 위해 각 요소를 제거하며 진행한 절제 실험(ablation study) 결과를 나타낸다. 디노이징 오토인코더를 사용하지 않은 모델(w/o DAE)은 입력 데이터를 재구성하지 않고 원본 입력 데이터를 그대로 사용하였다. <Figure 5>는 실제 입력 데이터와 디노이징 오토인코더로 재구성한 입력 데이터를 시각화한 결과로, 재구성된 데이터는 실제 입력 데이터 대비 변동성이 줄어들어 스무딩된 모습을 보였다. 실험 결과 디노이징 오토인코더를 사용하지 않을 때 전반적으로 예측 성능이 하락하는 모습을 보였다. 특히 스타일 A 상품의 경우, 디노이징 오토인코더를 제거했을 때 MAE와 WAPE가 각각 40.91%, 34.07% 증가하였다. 이는 디노이징 오토인코더를 통한 스무딩 효과가 예측 성능 향상에 효과적이라는 것을 보여준다.

시점 정보를 사용하지 않은 모델(w/o time stamp)의 경우, PatchTSDD 대비 전체 상품 데이터(ALL)에 대해 WAPE 차이가 62.92%로 가장 크게 나타났다. 또한 스타일 B 상품에 대해서는 성능차이가 약 3배로 나타났다. 이러한 결과는 시점 정보를 활용함으로써 모델이 시계열 데이터 내 전역적인 시간 관계를 더 잘 반영해 우수한 성능을 달성하였음을 보여준다.

DILATE 손실함수가 아닌 일반적인 MSE 손실함수로 학습한 모델(w/o DILATE)은 제안방법과 비교했을 때 스타일 A와 B에 MAE가 각각 12.8%, 38.18% 높게 나타났다. 또한 전체 상품(ALL)에 대해서도 DILATE를 제거한 경우 WAPE가 8.55% 증가하여, DILATE 역시 PatchTSDD의 예측 정확도를 높이는 데 기여했음을 확인할 수 있었다.

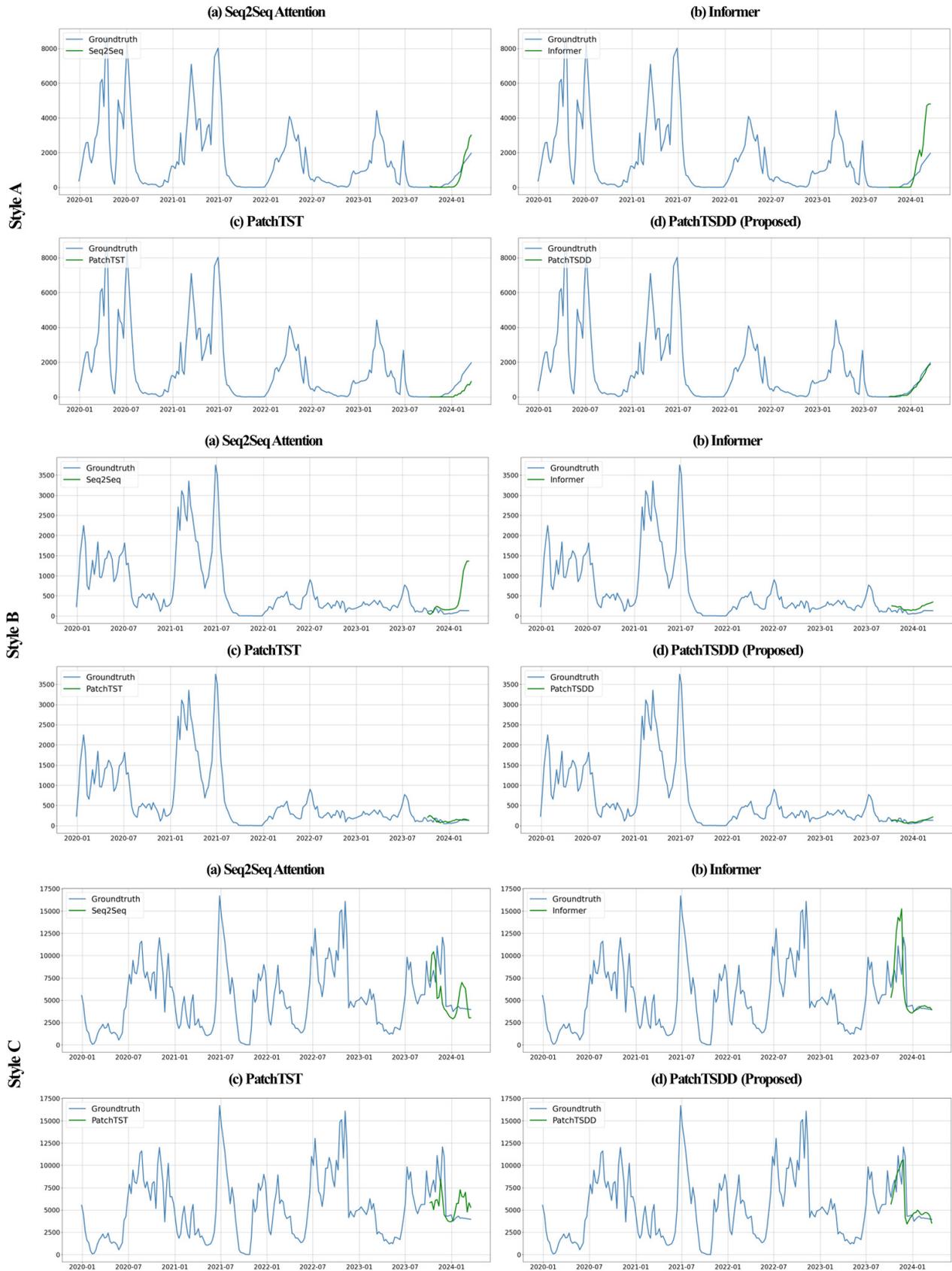
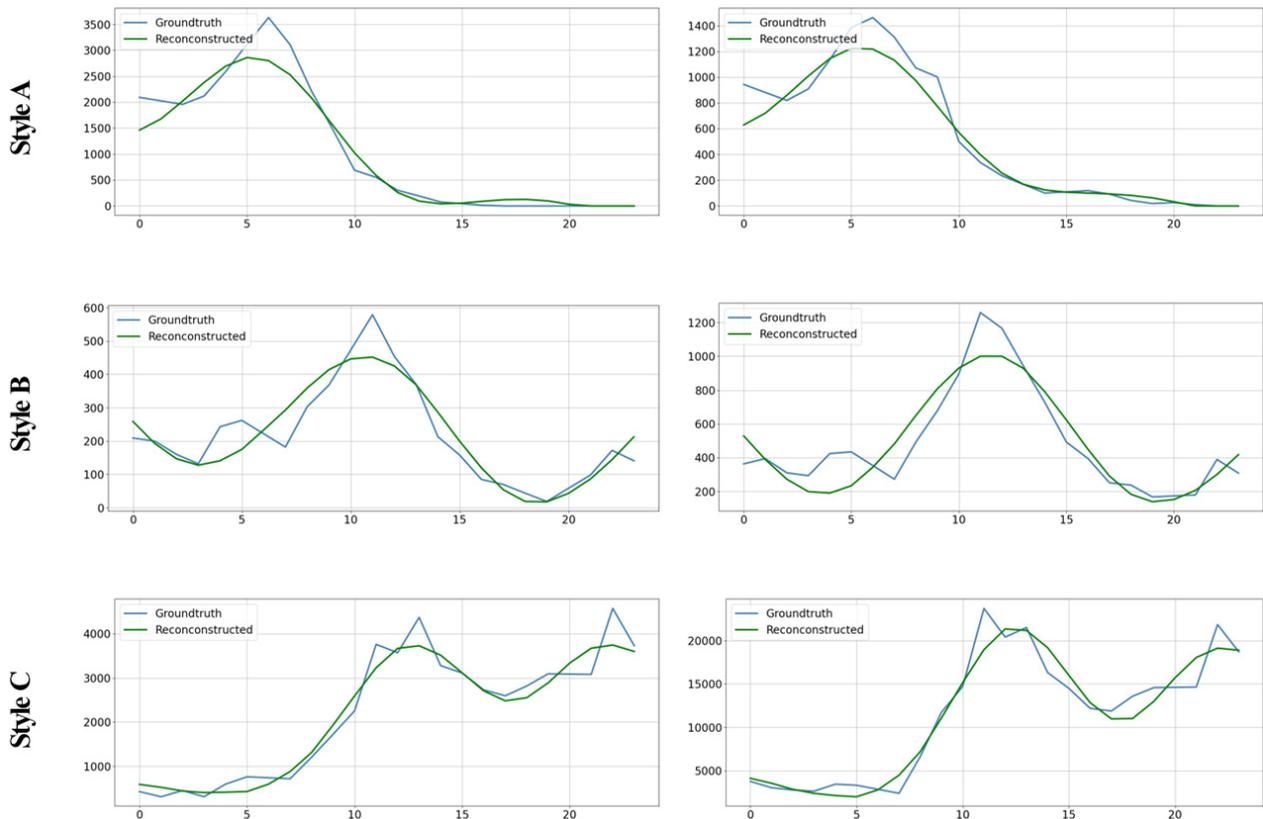


Figure 4. Visualization of Apparel Sales Forecasted by Comparative Time Series Forecasting Models. The Blue Line Indicates Ground-Truth Test Data, and Green Line Indicates Predicted Values Using Forecasting Model.

**Table 3.** Ablation Study for PatchTSDD to Show Effectiveness of Component by Removing each Component from the Model. The Best Result is Highlighted in Boldface

Model	MAE (↓)				WAPE 100(%) (↓)			
	Style A	Style B	Style C	All	Style A	Style B	Style C	All
PatchTST	721.00	55.04	1208.15	817.22	64.47	77.17	56.48	66.04
w/o DAE	598.78	45.85	<b>959.28</b>	656.95	54.89	<b>58.09</b>	49.94	53.14
w/o Time stamp	524.37	118.01	1164.94	752.67	52.17	184.54	52.89	81.23
w/o DILATE	485.04	57.80	992.34	<b>642.98</b>	44.93	79.38	<b>49.22</b>	54.52
PatchTSDD (Proposed)	<b>424.94</b>	<b>41.83</b>	1050.72	650.98	<b>40.94</b>	59.37	50.9	<b>49.86</b>



**Figure 5.** Visualization of Reconstructed Input Sequence with Denoising Autoencoder. The Blue Line Indicates Ground-Truth Input Data, and Green Line Indicates Reconstructed Input Data Using Denoising Autoencoder

#### 4.5 모델링 방식에 따른 성능 평가

본 연구에서 사용된 데이터셋에서 각 상품은 스타일, 색상 그리고 사이즈 3가지 특징으로 구분된다. 유사한 판매 패턴을 보이는 상품들을 함께 학습해 예측 성능을 높이고, 연산 효율을 개선하기 위해 위 세 가지 특징들을 기준으로 상품을 나누어 모델을 학습하였다. <Table 4>는 상품 특징인 스타일, 색상, 사이즈별로 상품을 구분해 모델을 학습했을 때 예측 성능을 보여준다. 실험 결과, 스타일별로 모델 학습 시 MAE, WAPE가 650.98, 49.86로 가장 우수한 성능을 보였다. 특히 데

이터 크기에 영향을 받지 않는 WAPE 성능이 스타일별 모델 학습에서 모든 상품에 대해 안정적인 예측이 가능하다는 것을 보여준다. 또한 스타일, 색상, 사이즈별로 모델을 학습할 경우 각각 3개, 6개, 15개 모델이 필요하므로, 학습 비용 측면에서도 효율적임을 실험을 통해 확인하였다. 이 결과는 스타일별로 모델을 학습할 때 제안방법이 시계열 패턴을 효과적으로 학습했음을 입증한다. 따라서 본 연구에서는 유사한 시계열 패턴을 보이는 상품들을 묶어 효과적으로 모델을 학습하기 위해 상품을 나누는 가장 기본적인 특징인 스타일별로 모델을 구축하였다.

**Table 4.** Experimental Results on Training the Model Using a Grouped Approach with Different Apparel Features. The Best Result is Highlighted in Boldface.

Model	Number of models	MAE (↓)	WAPE 100%(↓)
		All	All
All product	1	697.37	69.73
Color-wise	6	704.59	71.33
Size-wise	15	665.70	64.64
PatchTSDD (Style-wise)	3	<b>650.98</b>	<b>49.86</b>

## 5. 결론

많은 산업에서는 제품 생산부터 실제 판매까지 긴 리드타임이 존재하므로, 장기 판매량 예측이 필요하다. 특히 의류 산업은 트렌드와 계절성으로 인해 판매량 변동이 심하다는 특징이 있다. 본 연구는 이러한 장기 판매량 예측을 위해 딥러닝 기반 시계열 예측 모델인 PatchTSDD를 제안하였다. PatchTSDD는 시계열 패턴을 정확하게 파악하기 위해 디노이징 오토인코더와 시점 정보를 활용하였다. 디노이징 오토인코더는 입력 데이터를 스무딩하고, 시점 정보 주입은 시계열 데이터 내 시간 관계를 파악하는데 기여하였다. 또한 변동성 예측을 위해 DILATE 손실함수를 사용하여 모델을 학습하였다. 실험결과, 제안 방법이 시계열 예측에 사용되는 기존 딥러닝 모델들 대비 우수한 성능을 보이는 것을 확인할 수 있었다. 아울러 절제 연구를 통해 제안방법에 사용된 구성 요소들이 성능 향상에 크게 기여함을 입증하였다.

제안방법은 의류 산업에서 현업 전문가가 원자재 주문과 상품의 재고 관리를 하는데 유용한가이드 역할을 수행할 수 있을 것으로 생각한다. 또한, 예측된 미래 판매량을 바탕으로 특정 시즌 인기가 있을 디자인을 예측하고 이에 맞는 맞춤형 디자인 및 마케팅 전략을 수립하는데 도움을 줄 수 있다. 현재 제안방법은 의류 판매량에 영향을 끼치는 다양한 변수를 고려하지 못했다는 한계점이 존재한다. 향후에는 날씨, 환율 등과 같은 판매량과 관련된 변수를 활용한다면 제안방법의 성능을 향상시킬 수 있을 것이다. 또한 시계열 군집화 기법을 적용해 유사한 상품들을 묶어 학습한다면 시계열적 패턴을 보다 효과적으로 학습할 수 있을 것으로 기대한다.

## 참고문헌

Cho, K. (2014), Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078.

- Cuturi, M. and Blondel, M. (2017), Soft-dtw: A differentiable loss function for time-series, *In International Conference on Machine Learning*, 894-903.
- Dai, J., Song, H., Sheng, G., and Jiang, X. (2017), Cleaning method for status monitoring data of power equipment based on stacked denoising autoencoders, *IEEE Access*, **5**, 22863-22870.
- Frias-Paredes, L., Mallor, F., Gastón-Romeo, M., and León, T. (2017), Assessing energy forecasting inaccuracy by simultaneously considering temporal and absolute errors, *Energy Conversion and Management*, **142**, 533-546.
- Le Guen, V. and Thome, N. (2019), Shape and time distortion loss for training deep time series forecasting models, *Advances in Neural Information Processing Systems*, 32.
- Hochreiter, S. (1997), *Long Short-term Memory*, Neural Computation MIT-Press.
- Huang, X., Tang, J., and Shen, Y. (2024), Long time series of ocean wave prediction based on PatchTST model, *Ocean Engineering*, **301**, 117572.
- Kim, T., Kim, J., Tae, Y., Park, C., Choi, J. H., and Choo, J. (2021), Reversible instance normalization for accurate time-series forecasting against distribution shift, *In International Conference on Learning Representations*.
- Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., and Matsopoulos, G. K. (2023), A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks, *Future Internet*, **15**(8), 255.
- Luong, M. T. (2015), Effective approaches to attention-based neural machine translation, arXiv preprint arXiv:1508.04025.
- Mazen, F. M. A., Shaker, Y., and Abul Seoud, R. A. (2023), Forecasting of Solar Power Using GRU-Temporal Fusion Transformer Model and DILATE Loss Function, *Energies*, **16**(24), 8105.
- Nie, Y., Nguyen, N. H., Sinthong, P., and Kalagnanam, J. (2022), A time series is worth 64 words: Long-term forecasting with transformers, arXiv preprint arXiv:2211.14730.
- Sakoe, H. and Chiba, S. (1978), Dynamic programming algorithm optimization for spoken word recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **26**(1), 43-49.
- Seyedan, M. and Mafakheri, F. (2020), Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities, *Journal of Big Data*, **7**(1), 53.
- Smyl, S. (2020), A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting, *International Journal of Forecasting*, **36**(1), 75-85.
- Sutskever, I. (2014), Sequence to Sequence Learning with Neural Networks, arXiv preprint arXiv:1409.3215.
- Vallance, L., Charbonnier, B., Paul, N., Dubost, S., and Blanc, P. (2017), Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric, *Solar Energy*, **150**, 408-422.
- Vaswani, A. (2017), *Attention is all you need*, Advances in Neural Information Processing Systems.
- Wibawa, A. P., Utama, A. B. P., Elmunsyah, H., Pujianto, U., Dwiyanto, F. A., and Hernandez, L. (2022), Time-series analysis with smoothed Convolutional Neural Network, *Journal of Big Data*, **9**(1), 44.
- Wu, H., Xu, J., Wang, J., and Long, M. (2021), Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, *Advances in Neural Information Processing*

- Systems*, **34**, 22419-22430.
- Xiong, P., Wang, H., Liu, M., Zhou, S., Hou, Z., and Liu, X. (2016), ECG signal enhancement based on improved denoising auto-encoder, *Engineering Applications of Artificial Intelligence*, **52**, 194-202.
- Ailing, Z., Muxi, C., Lei, Z., and Qiang, X. (2023), Are transformers effective for time series forecasting, In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. (2021), Informer: Beyond efficient transformer for long sequence time-series forecasting, In *Proceedings of the AAAI Conference On Artificial Intelligence*, **35**(12), 11106-11115.
- Swaminathan, K. and Venkatasubramony, R. (2024), Demand forecasting for fashion products: A systematic review, *International Journal of Forecasting*, **40**(1), 247-267.
- Frank, C., Garg, A., Sztandera, L., and Raheja, A. (2003), Forecasting women's apparel sales using mathematical modeling, *International Journal of Clothing Science and Technology*, **15**(2), 107-125.
- Mostard, J., Teunter, R., and De Koster, R. (2011), Forecasting demand for single period products: A case study in the apparel industry, *European Journal of Operational Research*, **211**(1), 139-147.
- Liu, N., Ren, S., Choi, T. M., Hui, C. L., and Ng, S. F. (2013), Sales forecasting for fashion retailing service industry: A review, *Mathematical Problems in Engineering*, **2013**(1), 738675.
- Pongdatu, G. A. N. and Putra, Y. H. (2018, August), Seasonal time series forecasting using sarima and holt winter's exponential smoothing, In *IOP Conference Series: Materials Science and Engineering*, **407**(1), 012153.
- Choi, T. M., Yu, Y., and Au, K. F. (2011), A hybrid SARIMA wavelet transform method for sales forecasting, *Decision Support Systems*, **51**(1), 130-140.
- Lenort, R. and Besta, P. (2013), Hierarchical sales forecasting system for apparel companies and supply chains, *Fibres & Textiles in Eastern Europe*, **21**(6).
- Güven, I. and Şimşir, F. (2020), Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods, *Computers & Industrial Engineering*, **147**, 106678.
- Han, K. H. (2021), Prediction of women's golf wear sales using machine learning: Focused on weather factors and days of the week, *J Korean Soc Cost*, **71**, 17-33.
- Green, M. and Harrison, P. J. (1973), Fashion forecasting for a mail order company using a Bayesian approach, *Journal of the Operational Research Society*, **24**(2), 193-205.
- Giri, C., Thomasse, S., Balkow, J., and Zeng, X. (2019, August), Forecasting new apparel sales using deep learning and nonlinear neural network regression, In *2019 International Conference on Engineering, Science, and Industrial Applications (ICESI)*, 1-6.
- Luo, T., Chang, D., and Xu, Z. (2022), Research on Apparel Retail Sales Forecasting Based on xDeepFM-LSTM Combined Forecasting Model, *Information*, **13**(10), 497.
- Skenderi, G., Joppi, C., Denitto, M., and Cristani, M. (2024), Well googled is half done: Multimodal forecasting of new fashion product sales with image-based google trends, *Journal of Forecasting*, **43**(6), 1982-1997.
- Green, M. and Harrison, P. J. (1973), Fashion forecasting for a mail order company using a Bayesian approach, *Journal of the Operational Research Society*, **24**(2), 193-205.
- Bai, S., Kolter, J. Z., and Koltun, V. (2018), An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv preprint arXiv:1803.01271.
- Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., and Woo, W. C. (2015), Convolutional LSTM network: A machine learning approach for precipitation nowcasting, *Advances in Neural Information Processing Systems*, 28.
- Anggraeni, W., Vinarti, R. A., and Kurniawati, Y. D. (2015), Performance comparisons between arima and arimax method in moslem kids clothes demand forecasting: Case study, *Procedia Computer Science*, **72**, 630-637.
- Kusuma, N., Roestam, M., and Pasca, L. (2020), The analysis of forecasting demand method of linear exponential smoothing, *International Journal of Educational Administration, Management, and Leadership*, 7-18.
- Xu, J., Zhou, Y., Zhang, L., Wang, J., and Lefloch, D. (2021), Sportswear retailing forecast model based on the combination of multi-layer perceptron and convolutional neural network, *Textile Research Journal*, **91**(23-24), 2980-2994.
- Li, Y., Yang, Y., Zhu, K., and Zhang, J. (2021), Clothing sale forecasting by a composite GRU-Prophet model with an attention mechanism, *IEEE Transactions on Industrial Informatics*, **17**(12), 8335-8344.

## 저자소개

**이진우** : 고려대학교 산업경영공학부에서 2023년 학사 학위를 취득하고, 고려대학교 산업경영공학과에서 석사과정에 재학 중이다. 연구 분야는 Time Series Forecasting, Diffusion Models이다.

**박태남** : 인하대학교 통계학과에서 2022년 학사 학위를 취득하고, 고려대학교 산업경영공학과에서 석사과정에 재학 중이다. 연구 분야는 Generative Models, Diffusion Models이다.

**조용원** : 고려대학교 정보통계학과와 산업경영공학부에서 2019년 학사학위를 취득하고, 고려대학교 산업경영공학과에서 박사과정에 재학 중이다. 연구 분야는 Image Segmentation Algorithms and its Applications, Deep Semi and Self-Supervised Regression이다.

**정석윤** : 한세실업 부서장으로 재직 중이다.

**황정현** : 한세실업 부문장으로 재직 중이다.

**안태영** : 한세실업 부서장으로 재직 중이다.

**문혜리** : 한세실업 팀장으로 재직 중이다.

**김환호** : 한세실업 팀장으로 재직 중이다.

**김성범** : 고려대학교 산업경영공학부 교수로 2009년부터 재직

하고 있으며, 인공지능공학연구소 소장, 기업산학협력센터 센터장, 한국데이터마이닝학회 회장을 역임했다. 미국 University of Texas at Arlington 산업공학과에서 교수를 역임하였으며, 한양대학교 산업공학과에서 학사학위를 미국 Georgia Institute

of Technology에서 산업시스템공학 석사 및 박사학위를 취득하였다. 인공지능, 머신러닝, 최적화 방법론을 개발하고 이를 다양한 공학, 자연과학, 사회과학 분야에 응용하는 연구를 수행하고 있다.