

외생 변수 통합과 시계열 요약 기반 수요 최적화 예측 프레임워크: 설명 가능한 트리 모델을 활용한 GS25 실증연구

강지호*¹, 김진영*², 옥준용**³, 오주희*³

A Forecasting Framework for Category-Specific Demand Optimization based on Exogenous Variable Integration and Time Series Summarization: An Explainable Tree-based Methods Case Study on GS25

Jiho Kang*¹, Jinyoung Kim*², Junyong Ok**³, and Joohee Oh*³

요약

AI 기술이 발전했음에도 유통 수요 예측은 품목별 수요 패턴의 이질성과 외생 변수의 복잡성으로 인해 여전히 불확실성이 높다. 본 연구는 Walmart M5와 GS25 데이터를 활용해 품목의 수요 변동성을 정량화하고, 군집 특성별 최적 예측 전략을 설계하였다. 시계열 데이터를 ADI(Average Demand Interval)와 CV²(Squared Coefficient of Variation) 지표로 요약해 군집화하고, 각 군집에 최적의 트리 기반 예측 모델을 적용하였다. 또한, Window 구조를 활용한 시계열 학습, 변동성 완화를 위한 Smoothing 기법, 그리고 SHAP(SHapley Additive exPlanations) 기반 변수 선택을 결합하여 예측 정밀도와 모델 해석 가능성을 동시에 확보하였다. 실험 결과, 일부 군집에서 성능이 모두 개선되었으며, 변동성이 큰 군집에서도 상대적으로 안정적 성능을 유지하였다. 이러한 결과를 통해 제안된 프레임워크는 리테일 현장에 실질적으로 적용 가능한 설명 가능한 수요 예측 솔루션임을 실증하였다.

Abstract

Although Artificial Intelligence (AI) technologies have advanced considerably, demand forecasting in the retail sector remains subject to substantial uncertainty due to the heterogeneity of product-level demand patterns and the complexity of exogenous factors. This study proposes an optimized, cluster-specific forecasting framework by quantifying demand variability using the Walmart M5 and real-world GS25 sales datasets. Time-series data are summarized using the ADI and CV² metrics for clustering, after which, tree-based forecasting models are applied to each cluster. The framework further incorporates time-series learning with a windowing structure, smoothing techniques to reduce volatility, and feature selection via SHAP to simultaneously enhance predictive accuracy and model interpretability. Experimental results demonstrate that the proposed method yields performance improvements for certain clusters while maintaining stable performance for highly volatile clusters. These findings indicate that the proposed framework is a practically deployable and explainable demand forecasting solution for retail operations.

Keywords

demand forecasting, explainable ai, time-series summarization, product segmentation, exogenous variables

* 한동대학교 경영경제학부(*³ 교신저자)
- ORCID¹: <https://orcid.org/0009-0003-1636-7405>
- ORCID²: <https://orcid.org/0009-0001-8664-0089>
- ORCID³: <https://orcid.org/0000-0003-0850-0256>
** 한동대학교 ICT창업학부
- ORCID: <https://orcid.org/0009-0009-9004-2987>

• Received: Jun. 20, 2025, Revised: Aug. 21, 2025, Accepted: Aug. 24, 2025
• Corresponding Author: Joohee Oh
Dept. of Management and Economics, Handong Global University, Korea
Tel.: +82-54-260-1420, Email: joohecoh@handong.edu

1. 서 론

1.1 연구 필요성 및 배경

최근 인공지능 기술의 비약적인 발전은 다양한 산업에서 예측 기반 의사결정 체계의 고도화를 가능케 하고 있으며, 특히 유통 산업은 이러한 변화의 중심에 놓여 있다[1]. 그러나, 급변하는 소비 트렌드, 이벤트성 수요의 증가, 기후 요인의 복잡성 증가는 전통적인 수요 예측 방식이 지닌 한계를 더욱 부각시키고 있다[2]. 유통 현장에서는 여전히 과잉재고와 품절이라는 양극단의 문제가 반복되고 있으며, 이는 불안정한 수요 예측과 정량적·정성적 해석 체계 부재에서 기인한다[1][3].

기존 통계 예측 모델은 모델 구조가 단순하고 직관적이라는 장점이 있지만, 날씨, 프로모션, 요일 등 외생 변수를 효과적으로 반영하지 못하는 한계를 지닌다[2]. 반면, LSTM(Long Short-Term Memory), RNN(Recurrent Neural Network) 등 딥러닝 기반 시계열 모델은 비선형적인 복잡한 패턴을 학습할 수 있으나, 결과 해석의 어려움과 하이퍼파라미터 최적화 및 학습 비용 부담 등의 단점으로 인해 실무 현장에서의 활용에는 제약이 크다[4][5].

특히 유통 상품은 품목마다 수요 패턴이 상이하다. 자주 반복 구매되는 생수나 커피 같은 일상 소비재와, 특정 시즌이나 이벤트에만 수요가 발생하는 우산이나 핫팩 같은 계절성 상품, 변동성이 큰 한정판 상품이나 냉장 디저트류까지, 그 수요의 주기성과 변동성은 크게 다르다. 이러한 품목별 이질성을 무시하고 단일 모델을 전 품목에 일괄 적용할 경우, 과소 또는 과대 예측이 발생할 수 있으며, 예측의 신뢰성과 실효성이 저하될 수 있다[6].

또한, ‘왜 이 품목이 이렇게 팔렸는가?’, ‘내일은 얼마나 팔주어야 하는가?’와 같은 실무자의 핵심 질문에 명확히 답변하려면 단순한 수요 예측이 아니라 ‘설명 가능한 수요 예측 모델’이 필요하다. 모델이 예측한 수치 그 자체보다, 해당 예측이 어떤 변수에 의해 결정되었는지를 시각적으로 제시할 수 있어야, 비즈니스 의사결정에 실제로 활용 가능하다[3].

1.2 연구 목적 및 기여

이러한 문제의식에 기반하여, 본 연구는 국내 유통 기업 GS25의 실제 판매 데이터를 활용하여, 품목별 수요 특성을 정량적으로 분석하고 이에 최적화된 최적 예측 조합 전략을 적용함으로써, 예측 정확도와 해석 가능성을 동시에 확보하는 설명 가능한 수요 예측 프레임워크를 구축하는 것을 목적으로 한다. 기존의 일괄적 접근 방식을 탈피하고, 품목별 수요 패턴의 다양성과 외생 변수(예: 기상, 프로모션 등)의 상호작용을 함께 반영할 수 있는 체계적인 전략 설계를 통해, 실무 현장에서 직접 활용 가능한 고도화된 수요 예측 전략을 제시하고자 한다.

이를 위해, 시계열 데이터를 ADI와 CV² 지표로 요약하여 사분면 기반으로 분류하고, 각 수요 유형에 적합한 트리 기반 모델을 적용하여 맞춤형 예측 전략을 구성한다[6]. 이러한 ADI-CV² 기반 분류는 수요의 발생 빈도 및 변동성을 구조화하여 외생 변수의 예측 기여도를 차등적으로 반영할 수 있는 기반 역할을 하며, 각 군집별로 외생 변수의 영향력을 선별적으로 적용할 수 있게 해준다. 또한, Walmart M5 데이터를 활용한 사전 실험을 통해 프레임워크의 일반성과 적용 가능성을 검증하고, 이후 GS25 실제 유통 데이터에 이를 확장 적용하여 실증적 타당성을 검증하였다.

본 연구의 주요 기여는 다음과 같다:

첫째, 품목별 수요 패턴의 이질성을 ADI - CV² 기반으로 정량화하고, 사분면별로 최적 모델을 분기 적용하여 기존 단일 모델 방식의 한계를 극복하였다. 특히 각 군집의 수요 특성에 따라 Smoothing, Windowing 전략, 외생 변수 조합 등을 맞춤 적용함으로써 예측 성능을 높였다. 둘째, SHAP 분석을 통해 모델이 생성한 예측값에 대한 변수 기여도를 정량적으로 제시하고, 이를 시각화함으로써 실무자가 이해하고 활용할 수 있는 설명 가능한 AI 예측 시스템을 구현하였다. 셋째, Walmart M5 데이터에서 구축한 예측 전략을 GS25 국내 유통 데이터에 전이 학습하여 적용함으로써, 프레임워크의 범용성과 실무 확장 가능성을 입증하였다.

II. 이론적 배경 및 관련 연구

2.1 시계열 기반 수요 예측 한계와 대안

수요 예측은 시계열 데이터 기반의 대표적인 응용 분야로, ARIMA(AutoRegressive Integrated Moving Average), 지수평활법 등 전통적인 통계 모델이 오랫동안 널리 활용되어 왔다[3]. 이러한 모델들은 계절성과 추세가 명확한 시계열 데이터에 대해 안정적인 예측 성능을 보이며, 비교적 간단한 구조와 높은 해석력을 바탕으로 초기 수요 예측 시스템의 핵심이 되어왔다. 특히, 수학적 기반이 명확하여 예측 결과에 대한 직관적인 설명이 가능하다는 점에서 실무자들의 신뢰를 받아왔다.

그러나 최근 유통 및 소매 산업 환경은 제품 다양화, 빠른 수요 변동, 외생 변수의 영향 증대로 인해 점점 더 복잡해지고 있으며, 이로 인해 전통 모델의 한계가 두드러지고 있다. 첫째, 전통 모델은 비선형적 수요 패턴이나 급격한 불규칙적 수요 변화에 대한 적응력이 떨어진다[1]. 둘째, 외부 요인의 통합이 어려워, 외부 요인이 수요에 미치는 영향을 반영하지 못하는 구조적 한계를 지닌다[5]. 이러한 이유로 예측력이 떨어지고, 다양한 품목의 이질적인 수요 특성을 포착하지 못하는 문제가 발생한다.

이러한 한계를 극복하기 위해 최근에는 LSTM, RNN, Transformer와 같은 딥러닝 기반 예측 모델이 시계열 예측에 활발히 적용되고 있다[2][6]. 이들 모델은 장기 의존성 문제를 해결하고, 비선형적이며 고차원적인 수요 패턴을 포착하는데 강점을 가진다. 하지만 딥러닝 모델은 일반적으로 높은 연산 비용과 대량 학습 데이터 요구 조건을 수반하며, 예측 결과에 대한 직관적인 해석이 어려워 실무에서의 활용에는 제약이 있다[2]. 예를 들어, Transformer 기반 수요 예측 모델이 재고 예측 정확도를 개선한 사례가 보고되었으나, 예측 결과에 대한 설명이 부족하여 실무 적용에 제약이 있었다고 보고된 바 있다[7][8]. 특히, 모델 의사결정 근거가 불투명할 경우, 현장 실무자와의 신뢰 구축이 어렵고 의사결정 활용도가 떨어진다.

이러한 한계를 보완하기 위한 접근법으로, 최근에는 시계열 정보를 요약하여 정량적 지표로 변환

하고, 해당 지표를 기반으로 품목을 군집화하여 각 군집에 최적화된 예측 모델을 적용하는 방식이 제안된다[5]. 기존 연구에서는 간헐적 수요를 분류 기준으로 설정하여, ADI와 CV² 지표를 활용해 품목을 군집화하고, 각 군집에 다른 예측 모델을 적용하는 접근이 시도되었다. 각 그룹에 다른 모델을 적용함으로써 예측 정확도를 개선한 사례가 존재한다[9]. 이러한 방식은 수요 유형별로 모델의 적합도를 최적화할 수 있다는 점에서 주목받고 있으며, 특히 소매 유통과 같이 품목 이질성이 높은 환경에서 효과적이다. 다만, 기존 연구 대부분은 외생 변수의 통합 또는 트리 기반 모델링을 병행하지 않아, 변수 간 상호작용과 설명 가능성 측면에서 한계가 존재했다.

결과적으로, 기존 연구들은 수요 유형의 정량화나 비선형 패턴 반영 측면에서 일정 부분 기여하였으나, 외생 변수와 시계열 요약 정보를 통합하여 품목별 맞춤형 모델을 제시하고, 동시에 설명 가능성까지 확보한 통합적 접근은 부족했다. 본 연구는 이러한 관점을 결합한 전략을 통해 기존의 분절된 접근을 보완하며 통합하고자 한다.

2.2 외생 변수 통합의 효과성

소매 유통 환경에서 가격 변동, 기상 상황, 공휴일, 마케팅 이벤트 등 다양한 외생 변수는 소비자의 의사결정에 직접적인 영향을 미친다. 예를 들어, 비가 오는 날에는 우산의 판매량이 급증하고, 무더운 날씨에는 아이스크림이나 냉장 음료의 수요가 크게 증가하는 경향이 있다[1][5]. 이처럼 외생 변수는 수요 변동의 주요 원인으로 작용하므로, 이를 반영하지 못하는 예측 모델은 높은 예측 오차를 초래하여 현실성 및 예측 정확도가 현저히 저하될 수 있다.

기존 연구에서는 외생 변수를 통합한 예측 모델이 단일 시계열 기반 모델보다 예측 정확도에서 우위를 보이며, 특히 이벤트성 소비가 빈번한 유통 산업에서는 더욱 효과적임이 여러 연구에서 입증되었다[5]. 예컨대, 날씨 및 휴일 정보를 통합하여 수요 예측 성능을 크게 개선한 사례[10][11], 공휴일/프로모션 변수와 시간 정보 결합이 주말 품목 수요 예측의 정확도를 높인 연구 사례가 보고된 바 있다

[12]. 외생 변수의 대표적인 예시로는 다음과 같은 요인이 있다:

먼저 가격 및 할인 프로모션은 소비자의 가격 탄력성과 교차수요를 반영하여 매출 변동에 직접적인 영향을 준다. 날씨 요인으로는 평균 기온, 강수량, 습도 등이 있으며, 이는 품목별 수요에 즉각적이고 뚜렷한 영향을 미친다. 또한 주중·주말·공휴일 여부와 같은 캘린더 요인을 고려하면 요일별 특성과 이벤트에 따른 매출의 주기성을 효과적으로 탐지할 수 있다.

특히 2.3절에서 등장하는 트리 기반 모델은 이러한 외생 변수의 비선형 관계나 상호작용을 효과적으로 반영할 수 있는 구조를 갖추고 있으며, 유통 실무에서 실질적인 대안으로 주목받고 있다. 그러나 외생 변수의 통합은 변수 간 상호작용, 비선형성, 범주형·연속형 혼합 특성으로 인해 단순 통합으로는 효과적인 반영이 제한적이다. 따라서, 변수 간 복잡한 관계를 비선형적으로 포착할 수 있는 예측 구조가 필요하며, 동시에 실무 적용 가능성을 확보하기 위한 해석력도 함께 고려되어야 한다.

2.3 트리 기반 모델과 설명 가능한 AI

트리 기반 머신러닝 알고리즘은 분기 기준이 명확하고 결과 구조가 계층적이기 때문에 예측 과정의 해석이 용이하다는 장점을 가진다[5]. 대표적으로 XGBoost(Extreme Gradient Boosting), LightGBM(Light Gradient Boosting Machine), CatBoost(Categorical Boosting) 등은 높은 예측 성능과 연산 효율성, 과적합 억제 등에서 강점을 보이며, 다수의 수요 예측 연구에서 핵심 알고리즘으로 채택되고 있다[5].

최근에는 이러한 트리 모델을 설명 가능한 AI 기법과 결합하여, 단순한 결과를 제시에 그치지 않고, 예측에 기여한 변수들의 상대적 영향력을 정량적으로 시각화하는 방식이 확산되고 있다[13]. 대표적인 방법인 SHAP 분석은 각 예측값에 대해 개별 변수들이 기여한 정도를 샘플 단위로 분해하여 설명할 수 있으며, 유통 실무자가 예측 결과에 대한 논리적 근거를 받아들이는 데 큰 도움을 준다[13].

실제 유통 분야 예측 사례에서는 SHAP 분석을

통해 공휴일, 기온, 요일 등 외생 변수의 상대적 중요도를 파악하고, 이를 기반으로 재고 및 물류 전략을 사전 조정하는 방식으로 실무에 접목되고 있다 [5]. 이처럼 트리 기반 모델과 SHAP 분석을 결합한 예측 시스템은 높은 정확도, 해석 가능성, 실무 활용성이라는 세 가지 요건을 모두 충족시킬 수 있어, 설명 가능한 수요 예측 모델의 핵심 구성 요소로 자리 잡고 있다.

요약하자면, 기존 트리 기반 예측 연구들은 높은 성능을 보였음에도 불구하고 품목별 수요 유형에 따른 맞춤형 적용보다는 단일 모델로의 적용이 일반적이었다. 본 연구는 군집화 기반 수요 유형 분류와 SHAP 분석을 결합하여, 예측 정확성과 해석 가능성의 균형을 갖춘 통합적 수요 예측 전략을 제안한다는 점에서 기존 연구와 차별화된다.

III. 연구 방법론

본 장에서는 본 연구에서 제안하는 외생 변수 통합과 시계열 요약 기반 수요 최적화 예측 프레임워크의 각 단계별 데이터 구성, 시계열 요약 및 군집화, 외생 변수 통합, 예측 모델 설계, 설명 가능한 AI 적용, 성능 평가까지의 과정을 순차적으로 설명한다.

먼저, 예측 실험에 활용할 데이터셋을 선정하고 분석 목적에 맞게 구조를 통일한 뒤, ADI와 CV² 지표를 계산하여 품목별 수요 유형을 정량화한다. 이후, 각 군집 특성에 적합한 트리 기반 모델을 적용하고, 외생 변수를 통합하여 학습하고, SHAP 분석으로 변수별 기여도를 시각화하여 모델의 해석 가능성을 확보한다. 마지막으로 성능 지표를 활용하여 예측 정확도와 안정성을 검증한다. 이러한 전체 연구 절차와 데이터 분석 흐름은 그림 1에 제시하였다.

3.1 데이터 설명

본 연구는 제안한 수요 예측 프레임워크의 사전 검증과 실증 분석을 위해 두 가지 데이터셋을 활용하였다. 첫째, 사전 실험에는 미국 Walmart의 M5 Forecasting Accuracy 데이터셋을 사용하였다.

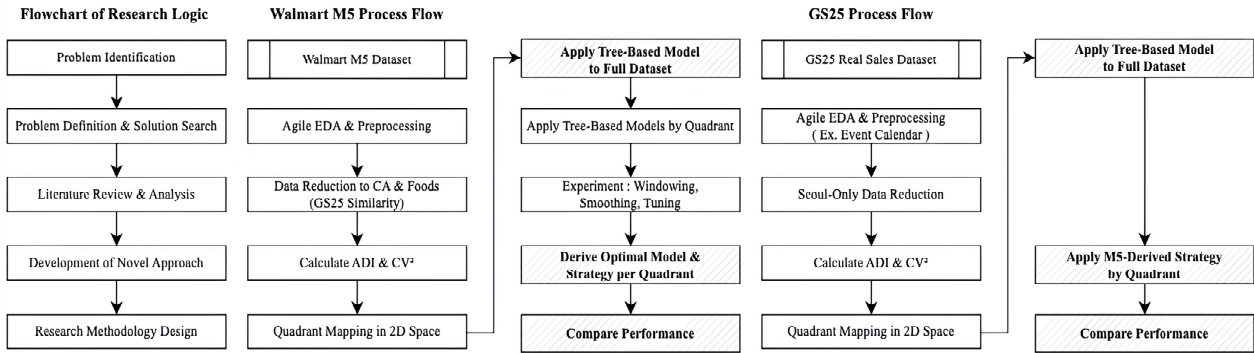


그림 1. 연구 절차 및 데이터 분석 흐름도
 Fig. 1. Research procedure and data analysis flow

이 데이터는 미국 3개 주(California, Texas, Wisconsin)의 일별 판매 데이터를 기반으로 하며, 약 30만 건의 시계열 기록을 포함한다. 특히, 이 데이터셋은 판매량(sales), 판매가격(sell_price), 이벤트 발생 여부(event_name_1, event_type_1), 저소득층 식품구매 재정 지원 프로그램 여부(SNAP, Supplemental Nutrition Assistance Program) 변수, 요일 및 월 등 캘린더 변수, 공휴일, 프로모션 정보, 가격 변동과 같은 다양한 외생 변수를 포함하고 있어, 날씨, 프로모션, 요일 등 주요 수요 영향 요인을 폭넓게 반영한 다변량 시계열 예측 실험에 적합하다[14].

둘째, 실증 분석에는 GS25 편의점의 국내 일별 판매 데이터를 활용하였다. 이 데이터는 전국 매장의 일별 판매량(adj_qty), 지역 정보(시도·시군구), 성별·연령대별 소비자 특성을 포함하며, 추가적으로 기상 데이터(기온, 강수량, 습도 등)와 결합하여 수요 예측에 활용하였다. 또한 공휴일 및 기념일 데이터를 병합해 이벤트 캘린더를 구축함으로써 국내 유통 환경에 특화된 외생 변수 구조를 마련하였다.

두 데이터셋 모두 종속 변수는 판매량(sales)으로 통일하고, 설명 변수는 공통 구조로 변환하여 예측 모델의 입력 데이터로 사용하였다. 시간 정보에는 ‘요일’, ‘월’, ‘계절(봄·여름·가을·겨울)’, ‘주말 여부(평일/주말)’를 포함하였고, 가격·판매 관련 변수로는 ‘판매가격’과 ‘할인 프로모션 여부’를 반영하였다. 기상 변수는 ‘기온’, ‘습도’, ‘풍속’, ‘강수 여부’로 구성하였으며, 이벤트·공휴일 변수에는 ‘신정’, ‘삼일절’, ‘발렌타인데이’, ‘화이트데이’, ‘광복절’,

‘한글날’, ‘크리스마스’, ‘현충일’, ‘대통령선거일’, ‘국회의원선거일’, ‘지방선거일’, ‘어린이날’, ‘석가탄신일’, ‘개천절’, ‘설날 연휴’, ‘추석 연휴’를 포함하였다. 이외에도 SNAP 여부를 포함하여 수요 변동에 영향을 미칠 수 있는 외생 요인을 최대한 폭넓게 반영하였다.

3.2 시계열 요약 및 품목 군집화

품목별 수요 패턴의 이질성을 정량적으로 분석하기 위해, 각 품목의 일별 판매 시계열을 ADI와 CV² 두 가지 지표로 요약하였다.

첫째 ADI(Average Demand Interval)는 수요가 발생하는 평균 간격을 의미하며, 수요의 빈도와 간헐성을 보여주는 지표다. 값이 작으면 해당 품목은 자주 판매되는 것이고, 값이 크면 간헐적으로 판매됨을 나타낸다. 둘째 CV²(Coefficient of Variation Squared)는 수요 변동성을 측정하는 지표로, 표준편차를 평균으로 나눈 값을 제곱하여 산출한다. 값이 클수록 수요의 불안정성이 높다는 것을 의미한다.

이 두 지표를 각각 스칼라 변수로 계산한 뒤 2차원 공간에 매핑하여 시각화하였고, Syntetos & Boylan[4]의 기준을 참고하여 기준값 ADI=1.32, CV²=0.49를 중심으로 사분면을 구분하여 다음 네 가지 군집으로 분류하였다[15].

먼저 Q1(Low CV², Low ADI)은 변동성과 간격이 모두 낮아 꾸준히 판매되는 안정적 품목군으로, 생수나 커피처럼 기본적인 소비재가 대표적이다. Q2(Low CV², High ADI)는 전반적으로 안정적이지만

판매 간격이 길어 계절이나 특정 상황에서만 수요가 발생하는 품목군으로, 우산이나 핫팩 등이 여기에 속한다. Q3(High CV², High ADI)는 수요 변동성과 간헐성이 모두 높아 예측이 까다로운 품목군으로, 시즌 한정 상품처럼 불규칙한 패턴을 보인다. 마지막으로 Q4(High CV², Low ADI)는 판매 빈도는 높지만 수요의 변동성이 커서 복잡한 패턴을 보이는 품목군으로, 냉장 디저트나 도시락이 대표적이다.

이러한 군집화는 품목의 수요 특성을 사전에 분류하여, 예측 전략 및 모델 구조를 군집별로 최적화할 수 있는 기초 데이터를 제공한다.

3.3 외생 변수 통합

시계열 정보만으로 설명하기 어려운 수요 변동성을 보완하기 위해 가격·프로모션, 캘린더 변수(요일, 공휴일, 주요 이벤트), 기상 정보(기온, 강수량, 습도, 풍속 등) 등 다양한 외생 변수를 통합하였다[2]. 이러한 변수들은 품목별 일별 판매 데이터와 날짜 및 지역 기준으로 병합하였다. 필요에 따라 범주형 변수는 더미 변수화하고 연속형 변수는 스케일링하였다. 이를 통해 군집별 모델 성능을 비교하고, 변수별 기여도를 분석할 수 있는 기반을 마련하였다.

3.4 예측 모델 설계 및 학습 방법

본 연구는 품목별 수요 특성에 따라 군집(Q1~Q4)에 최적화된 수요 예측 모델과 데이터 처리 전략을 적용하였다. 이는 기존 단일 모델 방식을 적용할 때 발생할 수 있는 과적합 및 성능 저하 문제를 보완하고, 계산 효율성과 예측 정밀도를 동시에 확보하기 위함이다. Window 기법은 모든 군집에서 공통적으로 적용되었으며, Window 크기는 30일로 설정하였다. 이는 소매 산업의 월간 발주 주기 및 M5 대회의 평가 주기(28일)를 참고하여 월 단위 수요 패턴과 재고 관리 주기를 동시에 반영하기 위함이다. 동일한 기준을 국내 GS25 데이터에도 적용하여 두 데이터셋 간의 비교 가능성과 분석 일관성을 확보하였다.

각 군집별 도출된 전략은 다음과 같다:

Q1은 CatBoost에 7일 지수이동평균(EMA_7)으로 스무딩을 더하고, 공휴일·날씨 변수를 통합해 안정적인 패턴을 정교하게 학습하며, 고정 윈도우로 예측 구간을 관리한다. Q2는 XGBoost를 기반으로 주중 여부, 프로모션, 월(계절) 정보를 핵심 특징으로 삼고 롤링 평균을 병행해 간헐적이지만 규칙성이 있는 수요를 잡는다. Q3는 하이브리드 접근으로 먼저 “판매 발생 여부”를 분류한 뒤, 발생 시에만 회귀 모델로 수요량을 예측하는 이단 구조를 채택해 높은 간헐성과 변동성을 처리한다. Q4는 LightGBM에 사후 스무딩(Post-Smoothing)과 원-핫 인코딩을 조합해 빈번하지만 변동성이 큰 복합 패턴을 안정화하며, 카테고리 특성을 세밀하게 반영한다.

각 군집별 기계학습 모델은 수요 발생 빈도와 변동성 수준에 따라 차별적으로 선정되었다. Q1은 외생 변수의 영향이 크고 범주형 변수가 다수 포함되어 있어, 이를 효과적으로 처리하고 복잡한 비선형 패턴을 학습할 수 있는 CatBoost를 채택하였다. Q2는 계절성과 주기성이 뚜렷해 RollingMean과 같은 추세 기반 특징이 중요하며, 비교적 안정적인 수요 패턴에서 높은 예측력을 보이는 XGBoost를 적용하였다. Q3는 0 값이 많은 희소 데이터 특성상 단일 회귀 모델로는 평균 오차 왜곡이 발생하므로, 판매 여부를 먼저 분류(XGBClassifier)한 뒤 판매량을 예측(LGBMRegressor)하는 하이브리드 구조를 통해 예측 정확도를 개선하였다. 마지막으로 Q4는 판매량 변동성이 크고 불규칙성이 높아 단순하고 해석 가능한 LightGBM 단일 회귀를 사용하되, EMA 기반 Post-Smoothing으로 노이즈를 완화하고 피크·저점의 과도 예측을 줄였다.

초기 실험 단계에서는 Grid Search 기반 하이퍼파라미터 튜닝을 적용하였으며, 이를 통해 기본 성능을 확인하였다. 이후 최종 모델 학습에서는 Bayesian Optimization을 활용하여 탐색 효율성을 높이고 계산 비용을 절감하였다. 탐색 대상 파라미터는 learning_rate, num_leaves, max_depth, min_child_samples 등 LightGBM 성능에 핵심적인 요소로 한정하였다.

데이터는 시간 순서를 유지한 상태에서 학습(70%)·검증(20%)·테스트(10%) 세트로 분할하였으며, 최적화 과정에서는 시계열 특성을 고려한 5-fold 교차검증(rolling-window 방식)을 통해 일반화 성능을 평가하였다. 또한 각 군집별로 효과적인 변수 조합이 상이하게 나타나, 수요 유형별 맞춤형 feature 설계가 예측 성능 향상에 중요한 역할을 함을 확인하였다.

3.5 SHAP 분석: 설명 가능한 AI 구현

트리 기반 모델의 해석 가능성을 확보하기 위해 SHAP 기법을 적용하여 변수 기여도를 정량적으로 분석하였다[13]. SHAP는 게임 이론에 기반한 기여도 산정 방법으로, 각 예측값을 생성하는 데 개별 변수가 기여한 정도를 샘플 단위로 분해하여 제시한다. 이를 통해 모델의 블랙박스 특성을 완화하고, 실무자가 예측 결과를 수용·활용할 수 있는 근거를 마련한다.

군집별 분석 결과, Q1(CatBoost) 유형에서는 'rolling_mean_30', 'ema_7', 'avg_temp' 등이 주요 변수로 나타났으며, 이는 단기 변동성과 외부 요인에 민감한 수요 특성을 반영한다. 반면 Q2(XGBoost) 유형에서는 'weekday', 'category', 'rolling_mean_30', 'lag_7' 등의 반복적 패턴과 카테고리 구분이 중요한 변수로 부각되었다. Q3(Hybrid) 유형은 판매 발생 여부 분류에서 'promotion', 'holiday_flag', 'rain_flag' 등이 높은 기여도를 보였으며, 이는 불규칙적·간헐적 수요가 주로 특정 이벤트나 날씨 변화에 의해 발생함을 의미한다. Q4(LightGBM) 유형에서는 'rolling_mean_30', 'max_temp', 'promotion' 등이 상위에서 위치해, 단기 판매 추세와 기온·관측 활동이 변동성이 큰 품목군의 핵심 요인임을 보여주었다.

또한 SHAP Summary Plot과 Dependence Plot 분석 결과를 바탕으로 변수별 영향 방향을 시각적으로 확인하였다. 예를 들어, avg_temp는 일정 범위까지는 판매량 증가에 기여했으나, 기온이 과도하게 상승하면 오히려 수요가 감소하는 패턴을 보였다.

SHAP 분석 결과를 바탕으로 주요 기여 변수만으로 구성된 경량 모델을 설계한 결과, 예측 성능 저

하 없이 모델의 해석력을 유지할 수 있었고, 모델 복잡도를 약 35% 줄일 수 있었다. 이는 향후 현장 적용을 위한 모델 경량화와 설명력 확보 간의 균형 가능성을 시사한다.

3.6 성능 평가 지표

예측 모델의 정확도와 군집별 전략의 효과를 정량적으로 비교하기 위해 Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Scaled Error (RMSSE) 지표를 활용하였다[5].

RMSE와 MAE는 절대 오차를 기반으로 하며, RMSE는 큰 오차에 가중을 부여해 예측 안정성 평가에 유리하다. MAPE는 상대 오차를 백분율로 표현해 모델의 일반적 예측 정확도를 직관적으로 비교할 수 있으나, 실제값이 0에 가까울 경우 값이 불안정해질 수 있다. RMSSE는 시계열 데이터 간 스케일 차이를 보정하여, 다른 품목이나 기간 간의 예측 성능을 공정하게 비교할 수 있다. 특히 M5 대회에서도 채택된 지표로, 시계열 예측에서 참값의 크기나 변동성이 큰 경우에도 비교 가능성을 유지할 수 있다.

본 연구는 이들 지표를 종합적으로 활용해 절대·상대 정확도와 시계열 스케일 보정 성능을 함께 검증하였으며, 이를 통해 군집별 모델 전략의 유효성을 다각도로 검증하였다.

IV. 실험 결과

본 장에서는 앞서 제안한 수요 예측 프레임워크를 실제 GS25 유통 판매 데이터에 적용하여 실험을 수행한 결과를 제시한다. Walmart M5[14] 데이터 기반의 사전 실험을 통해 도출된 군집별 맞춤형 예측 전략과 최적 모델 조합을 실제 국내 유통 데이터에 확장 적용하여 검증하고, 그 성능을 정량적으로 비교·분석함으로써 프레임워크의 실효성을 평가하였다.

4.1 실험 대상 설정

GS25는 전국적으로 분포한 편의점 체인을 기반으로 하며, 입지 특성상 접근성과 편의성을 중시하는 일상 소비 중심 품목이 주를 이룬다. 이에 따라, 분석 대상 품목을 수요 발생 빈도와 수요 변동성 기준으로 군집화한 결과, Q1과 Q2 유형이 큰 비중을 차지하였다. 반면, Q3와 Q4 유형은 전체 비중이 낮고, 판매 데이터의 불균형이 두드러졌으나, 희소·불규칙 수요군의 특성을 반영한 최적화 모델도 구축하여 실험을 병행하였다. 따라서, 모든 군집에 대해 특성 기반 맞춤형 모델을 구성하고 동일한 평가 체계에서 비교·분석하였다.

4.2 모델 구성 및 실험 설계

각 군집별로 앞서 도출해낸 최적 모델을 기반으로 Q1에는 CatBoost, Q2에는 XGBoost, Q3에는 Hybrid(XGBClassifier → LGBMRegressor), Q4에는 LightGBM + EMA Smoothing을 적용하였으며, 추가적으로 입력 변수 구성 수준에 따라 세 가지 모델 유형으로 파생시켜 성능 수준을 비교하였다.

Model 0 (Baseline)은 시계열 변수만을 활용한 기본 구조로, 외생 변수의 효과를 배제하고 성능 비교의 기준선 역할을 한다. Model 1 (Optimized)은 외생 변수와 함께 EMA, 롤링 평균 등 스무딩 특성을 통합한 확장형 모델로, 예측 정확도를 최대화하는 데 초점을 둔다. 마지막으로 Model 2 (Simplified)는 SHAP 분석을 통해 주요 기여 변수만을 선별하여 구성한 경량화 모델로, 해석 용이성과 계산 효율성 간의 균형을 추구한다.

4.3 예측 성능 비교 결과

도출된 세 가지 모델 유형(Model 0: Baseline, Model 1: Optimized, Model 2: Simplified)을 Q1, Q2, Q3, Q4 각 군집에 적용한 결과, 군집별 특성에 따라 예측 성능의 차이가 관찰되었다. 아래 표 1은 각 군집별 모델 수요 예측 성능 결과를 요약한 것이다.

Q1 군집은 수요의 단기 변동성이 크고, 외생 변수의 영향이 높은 품목군으로 분류된다[15]. 이 군집에서는 CatBoost 기반 Optimized 모델 (Model 1)이 가장

우수한 성능을 보였으며, RMSE 62.56, MAPE 27.58로 나타났다. Baseline 모델 대비 RMSE는 약 5.3% 개선되었으며, 이는 Smoothing Feature(예: ema_7, rolling_mean_30)와 외생 변수 통합이 예측 정밀도 향상에 효과적임을 시사한다[2]. Simplified 모델(Model 2)의 경우 주요 변수만을 활용한 결과, RMSE는 63.01로 Optimized 모델 대비 소폭 저하되었으나, 복잡도를 줄이면서도 비교적 예측력을 유지함을 확인할 수 있었다[3]. 이는 모델 경량화와 해석력 확보를 병행할 수 있는 가능성을 보여준다[13].

Q2 군집은 안정적인 반복 수요를 갖는 품목군으로, 예측이 상대적으로 용이한 구조를 갖는다[15]. 본 군집에서는 XGBoost 기반 Optimized 모델 (Model 1)이 RMSE 52.67, MAPE 22.69로 가장 나은 결과를 보였으나, Baseline 대비 RMSE는 0.19% 소폭 향상, MAPE는 오히려 0.31% 소폭 악화되었다. 이는 Q2 품목의 예측 가능성이 본질적으로 높아, 외생 변수 및 파생 변수 추가의 효과가 제한적임을 보여준다[6]. Simplified 모델(Model 2)은 RMSE 52.81, MAPE 22.73으로 Baseline과 유사한 수준의 성능을 유지하였으며, 이는 불필요한 변수 추가 없이도 기본적인 예측 성능을 확보할 수 있다는 실증적 근거로 해석된다[3].

추가적으로, Optimized 모델의 성능 개선폭에 대한 통계적 유의성을 검증하기 위해 Diebold-Mariano 검정을 수행하였다. Q1 군집에서는 Baseline 대비 RMSE 개선이 유의미($p < 0.05$)한 것으로 나타났으나, Q2 군집에서는 미미하였다. 또한 전통적 시계열 모델(ARIMA, 지수평활법)과 비교하였을 때, Optimized 모델은 전통 모델 대비 RMSE 기준 평균 10~15% 낮은 오차를 기록하여 트리 기반 접근의 상대적 우위를 입증하였다.

Q3 군집은 높은 변동성과 불규칙한 수요 패턴이 혼재하는 품목군이다. Baseline 모델(Model 0)은 RMSE 48.43으로 가장 낮았지만, MAPE가 109.23으로 매우 높게 나타나 절대 오차 대비 비율 오차가 심각하게 증가했다. Optimized 모델(Model 1)은 RMSE 81.13, MAPE 84.52로 RMSE 성능은 저하되었으나 MAPE는 개선되었으며, Simplified 모델 (Model 2)은 RMSE 44.53으로 가장 낮은 값을 기록

했지만 MAPE 123.79로 비율 오차는 가장 컸다. 이는 Q3 군집이 소규모·저수요 품목을 포함하고 있어, 소량 수요 변동이 MAPE를 크게 악화시키는 구조적 한계를 지님을 시사한다.

Q4 군집은 극단적 수요 변동성과 예측 난이도가 높은 품목군이다. Optimized 모델(Model 1)이 RMSE 38.72로 가장 우수했으나, MAPE는 140.55로 Baseline(86.35)이나 Simplified(85.48)보다 현저히 높았다. 이는 Optimized 모델이 절대 오차 기준으로는 개선 효과를 보이지만, 비율 오차 측면에서는 소량 판매일의 예측 편차가 크게 반영된 결과이다. Baseline과 Simplified 모델은 RMSE 83.05, 80.06으로 절대 오차는 높았지만, MAPE는 유사하게 나타났다.

이러한 결과는 군집별 특성에 따라 RMSE와 MAPE의 변동 방향이 상이하며, 단일 지표만으로 모델 성능을 파악하기 어렵다는 점을 보여준다. 특히 Q3·Q4와 같이 저수요·고변동 군집에서는 MAPE가 과대 반응할 수 있어, 보조 지표와 함께 해석하는 것이 필요하다.

표 1. 군집별 모델 예측 성능 비교

Table 1. Comparison of model prediction performance

Cluster	Model type	RMSE	MAPE
Q1	Baseline	66.06	27.90
	Optimized	62.56	27.58
	Simplified	63.01	28.10
Q2	Baseline	52.77	22.62
	Optimized	52.67	22.67
	Simplified	52.81	22.73
Q3	Baseline	48.43	109.23
	Optimized	81.13	84.52
	Simplified	44.53	123.79
Q4	Baseline	83.05	86.35
	Optimized	38.72	140.55
	Simplified	80.06	85.48

4.4 핵심 시사점 요약

실험 결과를 종합적으로 분석한 결과, 군집별 수요 특성에 따라 예측 전략이 상이하게 작동함을 확인할 수 있었다[15]. 먼저, Q1 군집의 경우 수요의 변동성이 크고 외생 변수의 영향을 많이 받는

특성을 갖기 때문에, 단순한 시계열 정보만으로 수요를 정확하게 예측하는 데 한계가 있었다[15]. 그러나 외생 변수와 Smoothing feature 전략을 함께 통합한 확장형 모델을 적용하였을 때, 예측 성능이 유의미하게 향상되었다. 이는 고변동 품목군에서는 외부 환경 정보와 단기 흐름을 반영한 파생 변수가 예측 성능 향상에 중요한 역할을 한다는 것을 시사한다[2].

반면, Q2 군집은 수요가 비교적 안정적이며 반복적인 특성을 지니고 있어, 기본적인 시계열 변수만을 활용한 단순 모델에서도 충분히 높은 예측 성능을 확보할 수 있었다[15]. 확장형 모델을 적용하였을 때 일부 성능 향상이 있었으나 차이는 미미하였으며, 오히려 과도한 변수 확장이 불필요하거나 성능 저하로 이어질 수 있다[3]. 이는 안정적인 품목군의 경우, 과도한 복잡도를 지닌 모델보다는 구조가 단순하면서 반복 패턴을 충실히 반영할 수 있는 모델이 더욱 적합하다는 점을 보여준다[6].

Q3 군집은 불규칙한 수요 패턴과 소규모 판매량이 혼재하는 특성으로 인해, MAPE 변동 폭이 크게 나타났다. 일부 모델은 RMSE는 낮았지만 MAPE가 급격히 악화되었으며, 이는 저수요 품목에서 소량 오차가 비율 지표에 과대 반영된 결과다. 따라서 Q3와 같은 군집에서는 RMSE·MAE 등 절대 오차 지표를 병행하여 평가하는 것이 바람직하다.

Q4 군집은 극단적 수요 변동성을 가진 고난도 예측군으로, Optimized 모델이 RMSE에서는 가장 우수했으나, MAPE가 현저히 악화되는 양상이 나타났다. 이는 소량 판매일의 예측 편차가 비율 오차에 크게 반영된 구조적 한계를 보여준다. 따라서 이 군집에서는 성능 지표 선택과 해석에 주의가 필요하며, 예측 안정성을 높이기 위한 사전 수요 필터링이나 변동성 완화 기법 적용이 요구된다.

또한, 본 연구에서 도입한 Simplified 모델은 주요 변수만을 선별하여 경량화된 구조로 설계되었음에도 불구하고, 전체 모델 성능에서 크게 뒤쳐지지 않는 결과를 보였다[3]. 이는 SHAP 기반 분석을 통해 기여도가 높은 핵심 변수만을 활용해도 일정 수준 이상의 예측력을 유지할 수 있음을 의미하며, 설명력과 계산 효율성을 동시에 고려한 전략적 모델 설

계가 가능함을 실증적으로 입증하였다[13]. 결과적으로 Simplified 모델은 실무 환경에서 유용하게 활용될 수 있는 방안을 제시한 것으로 평가된다.

종합하면, Q1·Q2는 예측 가능성이 상대적으로 높아 전략적 변수 선택과 복잡도 조절로 성능 최적화가 가능했으나, Q3·Q4는 데이터 특성상 MAPE 단독 해석이 왜곡을 초래할 수 있음을 확인하였다. 특히 Q3·Q4에서는 판매량 규모, 변동성, 데이터 불균형 등을 반영한 보정 전략이 필요하며, 실무 적용 시 모델별 지표 해석 가이드를 병행하는 것이 바람직하다.

V. 결론 및 향후 과제

본 연구는 유통 산업 내 수요 예측의 정확성과 실무 활용성을 동시에 확보하기 위한 방안으로, 품목별 수요 특성 정량화하여 설명 가능한 수요 예측 프레임워크를 제안하였다. 특히, 품목별 판매 시계열 데이터를 ADI와 CV²로 요약하고, 이를 기반으로 사분면 분류 체계를 구축하여 수요 유형 간 구조적 차이를 명확히 구분하였다. 이러한 군집화를 바탕으로 각 군집의 특성에 최적화된 트리 기반 머신러닝 모델을 적용하였다.

실험 결과, Q1·Q2 군집은 외생 변수와 시계열 파생 변수 결합이 성능 향상에 기여했으며, 특히 Q1에서는 Optimized 모델이 RMSE와 MAPE 모두에서 유의미한 개선 효과를 보였다. 반면, Q3 군집에서는 일부 모델이 RMSE는 낮지만 MAPE가 급격히 상승하여 소규모 판매량 변동이 비율 지표에 과도하게 반영되었다. Q4 군집은 Optimized 모델이 RMSE에서 가장 우수한 성능을 보였으나, MAPE는 오히려 큰 폭으로 악화되었다. 이러한 결과는 저수요·고변동 품목군에서 단일 지표로 성능을 해석할 경우 예측 품질을 과대 또는 과소평가할 위험이 있음을 시사한다. 이는 모델 평가 시 단일 지표 의존의 한계를 보여주는 동시에, 군집별 특성을 고려한 지표 선택의 중요성을 강조한다.

본 프레임워크는 사전 실험으로 활용한 Walmart M5 데이터셋과 국내 유통 기업 GS25의 실제 판매 데이터에 모두 적용되어 실증적으로 검증되었다. 특

히, Q1과 Q2와 같은 주요 수요 유형에서는 외생 변수와 시계열 파생 변수의 통합이 예측 정밀도 향상에 효과적이었으며, SHAP 분석을 통해 도출된 주요 변수 기반의 경량화 모델 또한 안정적인 성능을 유지하며 실무 적용 가능성을 입증하였다. 이는 기존의 일괄 적용식 단일 모델링 접근법의 한계를 넘어, 품목별 수요 특성에 따라 전략적으로 모델을 분기 적용하는 방식이 효과적임을 실증적으로 입증한 결과이다.

또한 본 연구는 단순히 모델의 성능 향상에 그치지 않고, 설명 가능한 AI를 실무에 도입할 수 있는 구체적 방법론을 제시함으로써, 예측 결과에 대한 신뢰도와 의사결정 기반 마련이라는 측면에서 유통 물류 산업 현장의 실질적 기여 가능성을 높였다.

향후 연구에서는 본 연구의 한계를 보완하고 예측 프레임워크의 실무 적용 가능성을 높이기 위한 후속 접근이 필요하다. Q3·Q4 중심의 심층 분석을 통해 저수요·고변동 품목군의 MAPE 과대 반응 문제를 완화할 수 있는 방안을 마련할 필요가 있다. 이를 위해 보정된 비율 오차 지표인 sMAPE(Symmetric Mean Absolute Percentage Error)와 WAPE(Weighted Absolute Percentage Error) 또는 변동성 완화 기법(예: Box-Cox 변환)의 적용 가능성을 검토할 수 있다. 또한 가격, 프로모션, 기온 등 기존 외생 변수 외에도 계절성, 소비자 심리, 소셜 트렌드 등 비정형 데이터를 반영한 변수 확장이 예측 성능 향상에 기여할 수 있다. 분석 단위 역시 품목 단위에서 점포 단위로 확대하여 지역별 수요 특성과 입지 요인을 고려한 Geo-AI 기반 예측 전략이 요구된다. 아울러, 시계열 정보와 상품 속성 등 비시계열 정보를 통합한 멀티모달 예측 구조의 개발도 중요한 방향이다.

이러한 개선 방향은 재고 운영, 프로모션 기획, 물류 배차 등 유통 현장의 핵심 의사결정을 정밀하게 지원하는 기반을 마련할 것으로 기대된다. 본 연구는 설명력과 실용성을 갖춘 수요 예측 프레임워크를 제안하였으며, 이러한 확장 방향을 반영함으로써 유통 현장의 정밀한 의사결정 체계 구축에 실질적 기여를 할 것으로 기대된다.

References

- [1] O. R. Amosu, P. Kumar, Y. M. Ogunsuji, S. Oni, and O. Faworaja, "AI-driven demand forecasting: Enhancing inventory management and customer satisfaction", *World Journal of Advanced Research and Reviews*, Vol. 23, No. 2, pp. 100-110, Feb. 2024. <https://dx.doi.org/10.30574/wjarr.2024.23.2.2394>.
- [2] R. S. Ahmed, M. Hasnain, M. H. Mahmood, and M. A. Mehmood, "Comparison of deep learning algorithms for retail sales forecasting", *ICCK Transactions on Intelligent Systematics*, Vol. 1, No. 3, pp. 112-126, Mar. 2024. <http://dx.doi.org/10.62762/TIS.2024.300700>.
- [3] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 competition: Results, findings, conclusion and way forward", *International Journal of Forecasting*, Vol. 36, No. 1, pp. 54-74, Jan. 2020. <https://doi.org/10.1016/j.ijforecast.2018.06.001>.
- [4] A. A. Syntetos, J. E. Boylan, and J. D. Croston, "On the categorization of demand patterns", *Journal of the Operational Research Society*, Vol. 56, No. 5, pp. 495-503, May 2005. <https://doi.org/10.1057/palgrave.jors.2601841>.
- [5] A. P. Wellens, R. N. Boute, and M. Udenio, "Simplifying tree-based methods for retail sales forecasting with explanatory variables", *European Journal of Operational Research*, Vol. 314, No. 2, pp. 523-539, Jan. 2024. <https://doi.org/10.1016/j.ejor.2023.10.039>.
- [6] M. Mansur, M. K. Hossen, and A. Khatun, "Sales forecasting for retail stores using hybrid neural networks", *PeerJ Computer Science*, Vol. 11, pp. e3058, Feb. 2025. <https://doi.org/10.7717/peerj-cs.3058>.
- [7] R. Caetano, J. M. Oliveira, and P. Ramos, "Transformer-based models for probabilistic time series forecasting with explanatory variables", *Mathematics*, Vol. 13, No. 5, pp. 1-29, Mar. 2025. <https://doi.org/10.3390/math13050814>.
- [8] J. M. Oliveira and P. Ramos, "Evaluating the effectiveness of time series transformers for demand forecasting in retail", *Mathematics*, Vol. 12, No. 17, pp. 2728, Sep. 2024. <https://doi.org/10.3390/math12172728>.
- [9] A. K. Singh, J. B. Simha, and R. Agarwal, "Prediction of intermittent demand occurrence using machine learning", *EAI Endorsed Transactions on Internet of Things*, Vol. 10, pp. 1-7, Mar. 2024. <https://doi.org/10.4108/eetiot.5381>.
- [10] D. Fredén and H. Larsson, "Forecasting daily supermarket sales with machine learning", Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, Jun. 2020. <https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-276483>. [accessed: Feb. 01, 2025].
- [11] S. Steinker, K. Hoberg, and U. W. Thonemann, "The value of weather information for e-commerce operations", *Production and Operations Management*, Vol. 26, No. 10, pp. 1854-1874, Oct. 2017. <https://doi.org/10.1111/poms.12721>.
- [12] M. Teixeira, J. M. Oliveira, and P. Ramos, "Enhancing hierarchical sales forecasting with promotional data: A comparative study using ARIMA and deep neural networks", *Machine Learning & Knowledge Extraction*, Vol. 6, No. 4, pp. 2659-2687, Dec. 2024. <https://doi.org/10.3390/make6040128>.
- [13] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier", *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, pp. 1135-1144, Aug. 2016. <https://doi.org/10.1145/2939672.2939778>.
- [14] Walmart, "M5 forecasting - accuracy", Kaggle, <https://www.kaggle.com/competitions/m5-forecasting-accuracy/overview>. [accessed: Feb. 01, 2025].
- [15] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks",

International Journal of Forecasting, Vol. 36, No. 3, pp. 1181-1191, Jul. 2020. <https://doi.org/10.1016/j.ijforecast.2019.07.001>.

저자소개

강 지 호 (Jiho Kang)



2022년 3월 ~ 현재 :
한동대학교 경영경제학 학부과정
(경영학 & AI융합학 전공)
관심분야 : 디지털 마케팅, 데이터
기반 의사결정,
설명가능한 인공지능(XAI),
자연어처리,
소비자 행동 연구, 기업경영전략지원

김 진 영 (Jinyoung Kim)



2020년 2월 ~ 현재 :
한동대학교 경영경제학 학부과정
(경영학 & AI융합학 전공)
관심분야 : Data-driven Marketing,
Consumer Behavior Analytics,
Content Planning & Branding

옥 준 용 (Junyong Ok)



2020년 3월 ~ 현재 :
한동대학교 ICT창업학 학부과정
2025년 7월 ~ 현재 :
주)마이메타 연구원
관심분야 : 머신러닝, 수요예측,
스타트업 생태계

오 주 희 (Joohee Oh)



2004년 3월 :
서울대학교 경제학부(석사)
2011년 8월 : University of
Southern California (경영학
박사)
2021년 3월 ~ 현재 :
한동대학교 부교수

관심분야 : 디지털 경제, 정보시스템,
비즈니스 애널리틱스, 금융데이터마이닝