

# 호텔 수요 예측을 위한 전역/지역 모델을 선택적으로 활용하는 시계열 예측 모델

## A Time Series Forecasting Model with the Option to Choose between Global and Clustered Local Models for Hotel Demand Forecasting

박기현 · 정경호 · 안현철<sup>†</sup>

국민대학교 비즈니스IT전문대학원

### 요약

인공지능 기술의 발전으로 인해 여행 및 호텔 산업에서도 다양한 목적의 인공지능과 기계학습 기법이 활용되고 있다. 특히 관광 산업에서는 수요 예측이 매우 중요한 요소로 인식되는데, 이는 서비스 효율성과 수익 극대화에 직접적인 영향을 미치기 때문이다. 수요 예측 시 시간에 따라 변화하는 데이터 흐름을 고려해야 하며, 이를 위해 통계적 기법과 기계학습 모델이 사용된다. 최근에는 수요 예측 데이터의 다양성과 현실의 복잡성을 반영하고자 기존 모델의 변형과 통합 연구가 진행되고 있으며, 그 결과 불확실성과 변동성에 대한 예측 성능이 향상되었음이 보고되고 있다. 본 연구에서는 기존 호텔수요 예측 연구에서 시도되지 않았던 다양한 기계학습 접근법을 통합하여 호텔 판매 수요 예측 정확도를 높이는 새로운 모델을 제안한다. 구체적으로 DTW K-means 클러스터링을 통해 지역모형을 구축하고, 전체 데이터를 활용한 전역모델과 선택적으로 결합하는 XGBoost 기반 시계열 예측 모델을 제시한다. 제안 모델은 지역과 전역 모델의 장점을 살려 호텔 수요 예측 성능을 제고할 것으로 기대된다. 이는 호텔 및 여행 산업 성장에 기여할 뿐만 아니라, 향후 다른 경영 분야 예측에도 확장 적용될 수 있을 것이다.

■ 중심어 : 수요 예측, 시계열 분석, 전역/지역모델 선택, DTW K-means, XGBoost

### Abstract

With the advancement of artificial intelligence, the travel and hospitality industry is also adopting AI and machine learning technologies for various purposes. In the tourism industry, demand forecasting is recognized as a very important factor, as it directly impacts service efficiency and revenue maximization. Demand forecasting requires the consideration of time-varying data flows, which is why statistical techniques and machine learning models are used. In recent years, variations and integration of existing models have been studied to account for the diversity of demand forecasting data and the complexity of the natural world, which have been reported to improve forecasting performance concerning uncertainty and variability. This study also proposes a new model that integrates various machine-learning approaches to improve the accuracy of hotel sales demand forecasting. Specifically, this study proposes a new time series forecasting model based on XGBoost that selectively utilizes a local model by clustering with DTW K-means and a global model using the entire data to improve forecasting performance. The hotel demand forecasting model that selectively utilizes global and regional models proposed in this study is expected to impact the growth of the hotel and travel industry positively and can be applied to forecasting in other business fields in the future.

■ Keyword : Demand forecasting, Time series analysis, Global/local model selection, DTW K-means, XGBoost

## I. 서론

호텔과 여행 산업은 전 세계적으로 주요한 핵심 산업 분야이다. 코로나-19로 인한 여행 제한이 해제되면서, 전 세계적으로 여행 트렌드가 급격히 증가하고 있다. 이 과정에서 해당 산업의 효율성과 수익관리의 중요성이 강조되고 있으며, 호텔 여행 산업에서의 수요 예측은 개별 호텔 사업자의 사업 성과와 밀접하게 연관된 중요한 의사결정문제가 되고 있다.

기업의 목표는 수익 창출이며, 매출 증대와 비용 절감 등을 통해 수익 향상을 도모하고 있다. 다양한 기업 형태 중 호텔과 같은 객실 판매 서비스 산업은 신선식품 판매업과 유사하게 상품의 가치를 유지하는 데 제한 시간이 있다. 호텔 여행 산업 역시 시간이라는 제약 변수가 존재하며, 오늘이라는 시간이 지나면 다시는 오늘의 객실을 판매할 수 없는 제약이 있다. 또한, 호텔 객실은 장소 이동을 통해 판매할 수 없는 특수한 성격을 가진다(김연화, 1995).

일반적인 수요 예측은 시간의 흐름이 반영되는 시계열 특성을 가지며, 효과적인 수요 예측은 생산, 물류, 재고 관리 효율성 향상 등 기업에 직접적인 영향을 미친다. 또한, 수요 예측은 마케팅 상품 구성과 판매 정책에도 영향을 미치며(배선경, 2005), 판매 수요 예측의 정확성은 객실 판매가 책정 및 가중치가 반영된 마케팅 재고 설정 등에 대한 결정요인을 제공함으로써, 여행 산업에서의 수익 극대화와 효율성 향상을 위한 기준을 제공할 수 있다.

지금까지 호텔과 같은 여행 산업의 수요 예측은 안정적인 시계열 모델과 국내총생산(GDP), 유가 등의 경제학적 변수를 주로 이용하였다(박득희 등, 2020). 최근에는 기계학습 모델 또한 활발히 시도되고 있다(Huang & Zheng, 2021). 이러한 연구들은 주로 예측 기법 자체에 초점을 맞추는 경우가 많았으며, 다양한 상황에 대한

대응에 한계가 있을 수 있다. 또한, 온라인 확산, 리드타임의 감소, 판매 상품 관련 정보의 실시간 공유 등으로 인해 실제 관광 수요 예측은 일관성이 없는 비선형의 특성을 가지게 되었다. 이러한 이유로, 통계학적 시계열 모델이나 기계학습 모델만으로는 수요를 정확히 예측하기 어려운 상황이 나타나고 있다. 이러한 어려움을 해결하기 위해 다양한 융합 접근법에 대한 연구가 이루어졌으며, 클러스터링 등의 전처리 접근법이 예측 모델의 성능 향상에 기여할 수 있음이 확인되었다(Kim & Ahn, 2008). 하지만 호텔 수요 예측 분야에 해당 접근법을 도입한 사례는 거의 없다.

호텔 데이터분석을 위해서는 다양한 변수가 필요하다. 총객실수, 판매 가능 객실수, 판매수, 숙박일수, 인원, 가격, ADR(평균판매가), RevPAR(가용객실기준 평균판매가), 휴일 여부, 행사, 취소율 등이 이에 해당한다. 기존 연구에서는 호텔 수익관리를 위한 다양한 마케팅 정책의 적용, 사용자 관점의 접근, 그리고 기술을 이용한 예측 모델 등이 제안되었다. 그러나 이러한 연구는 대부분 전통적인 시계열 모델을 이용한 프레임워크를 구축하는 것이었다(Lee & Zhang, 2023). 최근 다양한 기계학습 모델이 연구되고 있으며(Dowlut, 2023), 이제는 다양한 데이터 대상의 일반화 예측 모델이 필요한 시점이다. 정확한 판매 수요예측은 객실 판매가 책정에 기준을 제공할 수 있으며, 이를 통해 호텔 판매 수익 극대화를 위한 합리적 가격 제안 등의 기반이 될 수 있다. 그러나 기존의 여행 또는 객실 관련 연구는 조사 범위의 제한이나 설문조사 대상 호텔의 제한으로 인해 특정 지역 대상 또는 인지도 있는 특정 호텔에 대한 연구가 대부분이었다(Phumchusri, 2021). 이는 전국 단위의 거시적 관점이나 특정 지역에 대한 지나치게 미시적인 관점으로 조사의 범위를 한정된 것을 의미한다. 또한, 조사 범위 및 조사 방법은 개별 호텔 직원

대상의 설문을 통한 조사 또는 매년 정기적으로 취합되는 연간 점유율, 매출액 등의 거시적 데이터를 기반으로 하였으므로, 개별 호텔에서 시간의 흐름을 통해 발생하는 미시적 데이터 기반의 정확한 통계 모델의 생성에는 한계를 가질 수 있다.

이러한 한계를 극복하고자 본 연구에서는 특정기간 동안 다수의 호텔에서 수집된 운영 데이터를 이용하여, 클러스터링 전처리를 통해 호텔 수요 시계열 데이터들을 그룹화하고 각 그룹별 예측 모델을 구축하여 성능 향상을 도모한다. 특히 본 연구에서는 전처리 기법으로 시계열 데이터 군집화에 널리 사용되는 DTW K-means를 적용하여 각 군집별 지역 모델들을 구축한 뒤, 학습 단계에서 검증(validation) 데이터셋을 대상으로 각 군집별 지역 모델 예측 성과와 전체 데이터를 활용한 전역 모델 예측 성과를 비교하여 선택적으로 우수한 성능을 가지는 모델을 적용하는 프레임워크 활용을 통해 예측 성능을 높이는 새로운 아이디어를 함께 제안한다.

## II. 관련 연구

호텔 사업을 운영하기 위해서는 일반 기업과 동일하게 매우 다양한 업무 기능이 필요하다. 구체적으로 사업 기획, 자금 조달, 조직 관리, 운영 매뉴얼, 고객 및 협력사와의 다양한 거래, 인사 및 회계 관리 등이 있다. 사업장마다 항목에 따른 중요도는 다를 수 있으나 모두 필요한 기능들이다. 이러한 기능의 효율성을 높이는 것이 호텔정보시스템(HIS: Hotel Information System)의 역할이다. 호텔 산업에서 호텔정보시스템은 자산관리시스템(PMS: Property Management System)으로 알려져 있으며, 호텔 내의 인적, 물적 자원을 포함한 모든 자산을 운영하는 기능을 담당한다. PMS는 특정 고객에 대한 예약이나 투숙 정보 관리와 정산 등의 일부 기능에 국한

되지 않고, 사업 운영에 필요한 모든 기능을 종합적으로 관리하는 다양한 기능을 제공하는 것을 주요 목표로 한다. 그러나 많은 사업자는 PMS의 활용도 향상과 중요성 인식이 부족하며, 이는 상호운용성, 교육 부족, 기술적 특성으로 인한 IT 기능으로의 인식 때문이라고 할 수 있다(Pucciani, 2011). 자산관리시스템은 호텔의 디지털 백본 역할을 하며, 다양한 운영 업무를 처리하고, AI 알고리즘과 병행하여 방대한 양의 데이터 처리를 통한 수익 최적화를 구현한다(Bouchareb, 2023). 본 연구는 호텔에서 운영 중인 PMS를 통해 취합된 데이터를 기반으로 수요 예측을 진행한다.

PMS의 중요성과 함께, 매출 증대를 위한 가격 책정은 매우 민감한 요소이다. 이에 관련하여 Cost-Volume-Profit 분석 등을 통해 점유율을 포함한 변수들과 책정 가격과의 상관관계가 연구되었다(정경훈, 1985). 또한 객실 가격 책정에 영향을 미치는 항목들에 대한 연구를 통해 객실 크기, 체인 호텔, 지하철 연결, F&B 업장 수, Wifi 서비스 등이 주요 변수로 확인되었다(정은별, 2017). 이러한 연구 대부분은 특정 대상 범위와 기간에 대한 통계 데이터와 설문조사 자료를 이용하여 분석되었으며, 일반화 모델 구축과는 다른 방향으로 진행되었다.

### 2.1 수요 예측 관련 연구

많은 산업 분야에서는 수요 예측을 위한 다양한 연구가 진행되어왔다. 효율적인 운영 관리를 위한 수요 예측을 통해 생산량과 보유량의 효율성을 높이고, 적정 재고량 관리, 노후화 방지, 납품 리드타임 최적화, 효율적인 리소스 관리, 수익 향상 등이 가능해진다(유현선, 2017). 기업에서의 수요 예측 이슈는 노동력 확보 및 유지, 생산 시설 보유, 자재 수급 및 배송, 수익관리 등 다양한 관리 항목과 직결된다. 특히, 물리적인 상품 제작 기업의 경우, 자재의 수급과 생산 공

정 최적화, 판매 시점까지의 리드타임은 수익에 직접적인 영향을 준다. 그러나 호텔 객실과 같이 물리적인 서비스가 제공되지만 시간이 지나면 소멸되는 논리적 한계시간 형태의 상품을 다룰 때에는 다른 시각에서의 접근이 필요하며, 주택 가격의 수요 예측이나 가격 예측 모형과 같은 일반적인 시계열 특성을 고려해야 한다.

상품재고관리를 기반으로 하는 분야에서도 다양한 연구가 이루어졌다. 예를 들어, ARIMA를 이용한 수요 예측(이민상, 심완섭, 2002), 주문량 모형을 기반으로 한 수요 예측 모델(김정아 등, 2018), 고객 정보 패턴을 기반으로 한 시계열 분석(김민교, 강대전, 2015), 그리고 LSTM을 활용한 항공기 수리 부품품 수요 예측(김진섭 등, 2020) 등 회귀 분석과 기계학습을 활용한 다양한 연구가 진행되었다. 각 산업 분야나 서비스 형태에 적합한 예측 모델이 개발되었고, 특히 전자상거래 분야에서 많은 수요 예측 모델 연구가 진행되었다(Bandara, 2019). 딥러닝을 통한 수요 예측은 데이터 노이즈 처리와 랜덤성으로 인해 신뢰성에 문제가 있었으나, 시계열 데이터의 딥러닝 모델 적합 여부를 측정하는 프레임워크가 개발되었다(Livieiris, 2020). 또한, ‘블랙박스’라고 불리는 딥러닝의 해석과 사용자의 확신이 어려운 문제점들이 지속적으로 지적되면서(Castelvecchi, 2016), 인공지능의 신뢰성 향상을 위해 딥러닝 모델의 의사결정과정에 관여한 구성 요소를 기반으로 한 희소선형결합법이 제안되기도 하였다(최재식, 2019). 이처럼 수요 예측 결과의 불확실성을 줄이기 위한 다양한 시도가 제안되고 있으며, 생성된 모델의 신뢰성에 대한 지속적인 검증이 이루어지고 있는 상황이다.

## 2.2 시계열 예측 기법

시계열 분석은 데이터를 시간 순서대로 나열하여 의미 있는 정보와 통계를 추출하는 연구이다. 이는 과거 행동을 진단하고 예측하는 데 사

용된다. 의학, 경제, 기상, 자연과학 등 많은 분야에서 광범위하게 사용되며, 추세성(trend), 주기성(cycle), 계절성(seasonality)의 특징을 가진다. 통계적 기법은 장기적인 관점에서 변화가 적은 데이터에 대한 예측에 좋은 결과를 보이며, 이에 따라 다양한 분야에서 통계적 기법을 기반으로 한 시계열 모델이 많이 사용되고 있다. 시계열 데이터의 예측을 위해 주로 사용되는 통계 모델에는 지수평활(Exponential smoothing)과 ARIMA 모델이 있다. Box-Jenkins 모델로 알려진 ARIMA (Auto Regressive Integrated Moving Average, 자동 회귀 통합 이동 평균)는 1900년대 중반에 제안되었지만 여전히 높은 성능을 보여주는 모델이다. ARIMA는 자동 회귀(AR), 차분(Integrated), 이동평균(MA)의 세 가지 요소로 구성되며, AR 과정과 MA 과정을 혼합하여, 차분 절차를 거쳐 분석 및 예측을 진행한다. 과거 데이터의 추세와 패턴을 반영하여 시계열 예측을 진행하는 것이 전통적인 시계열 분석과 차이를 가지며, 보다 정확하고 효과적인 예측을 가능하게 한다.

호텔, 여행 산업에서도 관광객 수요 예측, 매출 예측, 객실 수요 예측 등의 분야에서 시계열 분석이 진행되었다. 호텔 비용 예측 연구에서는 ARIMA를 이용하여 6.62%의 MAPE를 나타내는 좋은 결과를 확인되었고(Syahromi, 2019), 일본인 관광객 수요 예측 연구에서는 지수평활 분석 방법의 하나인 ‘윈터스 승법 모델(Winter’s Multiplicative Model)’이 ARIMA 대비 더 좋은 성능을 보여주기도 하였다(박득희 등, 2020). 지구의 공전, 자전주기와 유관한 계절성과 관련된 시계열 분석에는, 계절성이 강화된 Seasonal-ARIMA (SARIMA) 모델이 두각을 나타내었다. Phumchusri (2021)는 객실 점유율 예측을 위해 작년 같은 날 비교, Holt윈터스, SARIMA, TBATS 통합 모델 등 4가지를 비교하여 계절성이 포함된 SARIMA 모델의 우수성을 확인하였고, 데이터 평활화와 결합하면 더 높은 예측 성능을 가질 수 있음을 보

여주었다. GM(회색모델)과 SARIMA를 결합하여 골삭기 수요 예측에서 MAPE를 4.84% 향상시킨 사례도 있다(Zhao, 2019).

호텔 객실 매출 예측에서는 ARIMA 모델 연구가 주목받았으며, 국내에서는 부산 지역 호텔 객실 매출에 대한 SARIMA 분석을 통해 특정 환경, 특히 계절적 시계열에서 상대적으로 높은 예측 성능을 보여주는 것을 확인하였다(이현찬 등, 2023).

### 2.2.1 기계학습

기계학습(Machine Learning)은 데이터를 활용하여 최적의 결과를 도출하기 위해 컴퓨터가 스스로 학습하는 인공지능의 한 분야이다. 기계학습은 다양한 알고리즘을 사용해 학습 데이터를 바탕으로 패턴을 인식하고 이를 기반으로 새로운 데이터에 대한 예측을 수행한다. 기계학습 모델은 크게 지도학습(Supervised Learning)과 비지도학습(Unsupervised Learning)으로 분류된다. 지도학습은 정답이 포함된 데이터를 사용하여 학습하는 방식으로 주로 분류(Classification)와 회귀(Regression)에 사용된다. 비지도학습은 정답이 없는 데이터를 사용하여 학습하는 방식으로 주로 군집화(Clustering)와 차원 축소(Dimensionality Reduction)에 사용된다.

딥러닝(Deep Learning)은 기계학습의 일종으로 인공신경망(Artificial Neural Network)을 기반으로 한다. 딥러닝은 대규모 데이터셋, 고성능 하드웨어, 그리고 긴 학습 시간을 요구하지만, 자동으로 특징(feature)을 추출하여 높은 예측 성능을 발휘한다. 주요 딥러닝 알고리즘으로는 CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), MLP(Multi-Layer Perceptron), LSTM(Long Short-Term Memory), GRU(Gated Recurrent Unit) 등이 있다. 김재성, 양여진(2020)의 연구에서는 ARIMA, LSTM, 1D-CNN 알고리즘을 비교 분석하여 1D-CNN과 LSTM이

ARIMA보다 높은 예측 성능을 보였으며, 특히 CNN 모델에 외생 변수를 반영한 연구에서는 코로나 기간 동안의 한국 내 입국자 수 예측 성능이 높게 나타났다.

### 2.2.2 기계학습 기반의 시계열 분석

시계열분석(Time Series Analysis)은 데이터를 시간의 흐름에 따라 분석하고 패턴을 모델링하여 미래값을 예측하거나 이상치를 탐지하는 분석방법이다. 이는 주식 가격 예측, 기상 예보뿐만 아니라 미래값 예측, 패턴 인지, 수요 예측 등 다양한 분야에서 활용된다. 시계열 분석은 데이터수집, 탐색 및 시각화, 전처리, 모델링, 훈련 및 평가, 결과분석 및 미래 예측 그리고 모델 개선과 같은 여러 단계를 거쳐 진행된다. 전통적인 시계열 분석 모델로는 ARIMA, 지수평활법(Exponential Smoothing), SARIMA 등이 널리 사용된다. 그러나 이러한 모델들은 계절성 적용이 어렵고 비선형 패턴 처리의 한계가 있어, 최근에는 LSTM과 같은 딥러닝 모델이 시계열 데이터분석에 많이 활용되고 있다.

기계학습 기반의 시계열 분석에서는 RNN, LSTM을 포함한 신경망 기반 모델, 서포트 벡터 머신(SVM)으로 대표되는 커널 기반 모델, 데이터의 구조적인 특징 파악이 용이한 심볼릭 회귀(Symbolic Regression), 여러 개의 모델을 조합하는 앙상블 모델(Ensemble Models) 등 비전통적인 시계열 분석 방법이 많이 사용되고 있다. 또한, 페이스북의 Prophet과 Azure에서 제공하는 모듈과 같은 상용화된 모델은 편리성과 높은 예측력을 제공하며, 다양한 모델의 적용이 가능하다.

국내에서도 기계학습은 데이터 기반의 모형 개발과 결과 예측에 널리 활용되고 있다. 예를 들어, 이인지와 윤현식(2020)은 지역 축제 방문객 예측에서 RF(Random Forest)를 도입하여 선형회귀 대비 25% 성능이 향상되는 결과를 보였다. 또한, 기계학습 해석의 어려움에 대한 대안

으로 설명 가능한 인공지능(XAI, Explainable AI) 개념이 중요시되면서, 강보람과 안현철(2022)은 리조트 교차판매 모형 개발에 SHAP(SHapley Additive exPlanations) 모형을 도입하여 기계학습 모델의 예측 결과를 해석가능한 형태로 제공함으로써 모델의 결정 과정을 이해하기 쉽게 만들었다.

### 2.2.3 여행, 호텔 분야에서의 시계열 분석

여행 및 호텔 산업 분야에서도 시장 예측을 통한 능동적 대응이 매우 중요하다. 특히 코로나 이후 급변하는 환경에 대한 예측 모델의 성능이 중요시되었다. 이를 위해 RNN, LSTM, GRU 등의 인공지능망 모델이 수요 예측에 도입되기 시작했다. 그러나 예측 근거가 필요한 의사결정 지원시스템을 위한 수요 예측에는 일반적으로 ARIMA 모델이 많이 사용되었다(장스위 등, 2022). 최근에는 향상된 신경망 모델 개발과 앙상블 모델의 출현으로 많은 분야에서 인공지능망 모델에 대한 관심이 증가하였다. 외부 회귀 변수를 사용한 SARIMAX 모델과 ANN-MLP (Multi-Layer Perceptron) 비교 연구가 진행되는 등 다양한 모델이 적용되고 있다(Ampountolas, 2021).

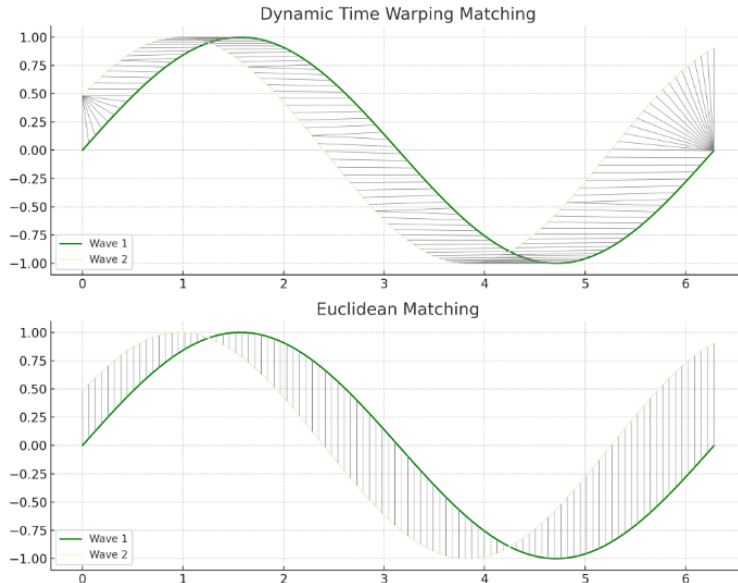
Pereira and Cerqueira(2022)는 18종류의 수요 예측 모델을 비교하면서, 기계학습 모델을 활용하여 기존의 지수평활법과 비교하여 1일 예측 기간에서는 54%, 14일 예측 기간에서는 45%까지 성능이 향상됨을 측정하였다. 이 연구에서는 기계학습이 기존 방법보다 낮은 SMAPE(Symmetric Mean Absolute Percentage Error)를 가져온다는 것을 확인했으며, 모델별 7일 단기 예측의 경우 0.12~2.65 사이, 14일 장기 예측의 경우 0.18~4.0 사이의 SMAPE를 보였다. 단기 수요 예측에서는 앙상블 모델이 우수하다는 결과를 도출했으며, 장기 수요 예측에서는 개별 기계학습 모델이 더 적합하다는 것을 입증하였다. Dowlut(2023)은 2017~2022년 동안 하드웨어와 소프트웨어 발전

으로 인해 기계학습 연구가 폭발적으로 증가한 것을 고려하여 호텔 여행 관련 기계학습 연구 50 개를 비교 분석하여, CNN-LSTM 하이브리드 모델의 우수성을 확인하였다.

### 2.3 시계열 데이터를 위한 클러스터링

데이터의 다양성과 과적합 방지, 일반화를 통한 예측 성능 향상을 위해 여러 기존 연구에서 다양한 전처리 기법과 과적합 방지 기법, 일반화 기법이 적용되었다. 정규화는 데이터의 범위를 조정하여 서로 다른 규모의 데이터에 대한 일반화 성능을 높여주는 기법이다. 정규화 기법에는 값의 범위를 통일시키는 최대-최소 정규화, 평균값을 중심으로 분포도를 조정하는 Z-Score 정규화 등이 있다. 권보성 등(2018)은 단기 전력 수요 예측 연구에서 정규화를 적용하여 MAPE 27% 향상을 확인하였고, RNN, LSTM을 이용한 주가 예측 모델에서는 데이터 크기가 작을 경우 딥러닝의 특성을 고려한 정규화를 통해 DNN 대비 15%의 예측 정확도 향상을 확인하였다(신동하 등, 2017).

비계층적 군집분석에 이용되는 K-means 클러스터링은 미리 설정한 군집의 수를 정한 뒤 그 수만큼의 임의 초기점을 생성하고, 유클리디안 거리 계산을 통해 군집을 생성한 뒤, 새로운 군집 중심점 계산 과정을 반복하여 유사 그룹을 분류하는 기법이다. 시계열 분석에 이용되는 DTW (Dynamic Time Warping, 동적 시간 왜곡)는 두 개의 시계열이 상호 간 얼마나 유사한가를 측정하는 기법 중 하나이며, 시간 또는 속도가 다른 두 개의 데이터에 대해 데이터의 길이를 왜곡(warping)하는 방식을 적용하여 음성 인식, 동작 인식 등의 분야에서 유사성 측정에 적용되고 있다. K-means Clustering에 DTW를 적용하는 경우, 서로 다른 길이를 가지는 데이터 그룹의 분류에 좋은 성능을 나타내는 것이 확인되었다. 서보원



〈그림 1〉 DTW 매칭(上)과 일반적인 유클리드 매칭(下)의 비교

등(2021)은 제스처 데이터에 DTW와 K-means를 적용하여 음성 인식 및 제스처 인식을 향상을 확인하였고, Oviedo et al.(2023)은 DTW K-means 클러스터링 비지도 접근 방식을 통해 태양광 모듈의 결함감지 분류에서 성능 향상을 확인하였다.

### III. 제안모델

전술했듯이 전통적인 여행 및 호텔 산업에서의 수요 예측에는 주로 ARIMA, SARIMA 등의 통계기법이 사용되었다. 최근 몇몇 연구에서는 기계학습, 특히 딥러닝을 이용한 수요 예측과 매출 예측이 진행되고 있지만, 특정 지역이나 특정 호텔에 한정된 데이터를 기반으로 수행되기 때문에 일반화된 모델 구축에는 한계가 있다. 본 연구에서는 호텔 운영시스템(PMS)에서 직접 생산되는 데이터를 기반으로 다수 지역과 다수 호텔에 대한 일반화된 수요 예측 모델을 구축하고자 한다.

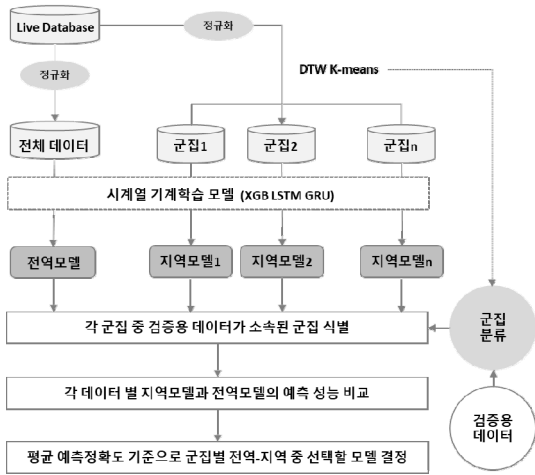
이를 위해, 모든 호텔들의 수요 패턴을 동일한 조건으로 설정하기 위해 정규화를 적용했으며,

데이터가 수집된 호텔들의 운영 기간 차이로 인한 왜곡을 줄이기 위해 동적 시간 왜곡(DTW)을 적용하였다.

모델 및 방법론 관점에서, 본 연구는 클러스터링 전처리를 통해 수요 예측 대상 호텔들을 유사한 수요 변화를 갖는 여러 개의 집단으로 세분화하였다. 이후, 각 군집별 시계열 예측 모델을 구축하여 예측에 활용하는 ‘지역모델(local model)’을 제안하였다. 이렇게 지역모델을 이용해 예측할 경우, 동질적인 학습 데이터를 사용함으로써 예측 정확도가 향상되는 효과를 기대할 수 있다. 그러나 군집 내 동질성이 떨어질 경우, 전체 데이터를 사용한 모델(전역모델)보다 예측 정확도가 감소하는 부작용이 발생할 수 있다.

따라서 본 연구에서는 검증(validation)용 데이터셋을 기반으로 전역모델과 지역모델의 성능을 비교한 뒤, 지역모델의 성능이 더 우수할 때만 지역모델을 적용하고, 그렇지 않은 경우에는 전역모델을 적용하는 ‘선택적 모델’을 제안한다.

<그림 2>는 본 연구에서 제안하는 모델의 구체적인 작동 과정을 설명하고 있다. 이 그림에



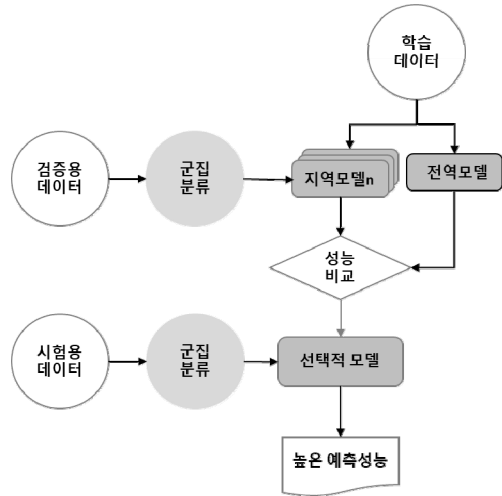
〈그림 2〉 제안모델의 작동 과정

서 볼 수 있듯이, 호텔의 운영시스템(PMS)에서 실시간으로 취합되는 운영 데이터베이스(Live Database)는 먼저 정규화를 통해 동일한 스케일로 변환된다. 이후 시계열에 특화된 클러스터링 기법인 DTW K-means를 사용하여 n개의 군집으로 세분화된다. 각 군집별로 기계학습 기법들(XGB, LSTM, GRU)이 적용되어 모델1부터 모델n까지의 지역모델이 학습된다. 동시에, DTW K-means를 적용하지 않고 전체 데이터를 사용하여 학습하는 전역모델이 동일한 시계열 기계학습 기법들을 적용해 학습된다.

전역모델과 지역모델이 모두 학습되면, 각 군집별로 검증용 데이터셋을 사용하여 전역모델의 예측 성능과 지역모델의 예측 성능을 비교한다. 이를 통해 어느 모델이 더 우수한지 결정하며, 그 결과에 따라 적절한 모델을 선택적으로 사용한다.

모델 학습이 완료되면, 새로운 대상 데이터(또는 시험용 데이터)가 입력되었을 때, 해당 데이터가 어느 군집에 속하는지를 판단한다. 그리고 각 군집에서는 전역모델과 지역모델 중 어느 모델이 더 우수한지에 따라 적합한 모델을 선택하여 새로운 대상 데이터의 시계열 예측값을 산출하게 된다. 다음의 <그림 3>은 전역 모델과 지

역 모델이 어떻게 선택적으로 적용되는지를 도식화하여 표현하고 있다.



〈그림 3〉 전역, 지역모델의 선택적 적용 흐름도

#### IV. 실증 분석

본 연구의 실증 분석 절차는 다음과 같이 진행한다.

- ① 국내 호텔들을 대상으로 서비스 중인 A사 시스템에 기록된 데이터 중, 코로나 기간을 제외하고 최근 5년간 일판매 레코드가 180개(6개월) 이상 존재하는 데이터를 추출한다.
- ② 수집된 데이터에 대해 최소-최대 정규화를 적용한다.
- ③ 수집된 데이터를 학습 데이터, 검증 데이터, 시험 데이터로 분할한다.
- ④ 학습 데이터는 DTW K-means 클러스터링을 통해 분류한다.
- ⑤ 전체 데이터와 클러스터링된 데이터 군집별로 기계학습을 통해 전역 및 지역모델을 생성하고, 검증 데이터를 통해 각 군집별로 선택할 모델을 결정한다.
- ⑥ 시험 데이터에 학습된 제안 모델을 적용



하여 최종적인 예측 성능을 확인하고 비교한다.

보다 상세한 실증 분석의 설계와 분석 결과는 다음과 같다.

#### 4.1 실험 데이터

이전의 유사 연구들은 동일한 규모와 마케팅 수준을 가진 집단 내 상관관계를 도출하기 위해, 특정 지역 또는 자료의 확보가 가능한 특급호텔 등으로 조사 집단을 한정하여 분석하는 것이 일반적이었다. 이로 인해 편향이나 특정 고객군 대상의 객실 상품 판매에 대한 획일적인 관계성 추출 등 제한적인 분석 결과만을 얻게 될 위험이 있었다. 또한, 학습 데이터의 과도한 제한으로 인해 학습된 모델을 시험 데이터에 적용할 때 성능이 크게 저하되는 경우가 발생하였다. 이는 데이터 유형이나 대상이 변화할 때 성능이 저하되는 일반화 모델의 한계를 나타낸다.

이러한 위험을 피하고자 본 연구에서는 철저하게 표본을 무작위로 추출하여 다양한 규모와 유형의 호텔들이 자연스럽게 데이터셋에 포함되도록 하였다. 데이터 추출은 2017년 1월에서 2023년 12월까지의 기간 중 코로나 팬데믹 기간(2020년 1월~2021년 12월)을 제외하고 진행하였으며, 서울, 부산, 제주 등에 위치한 국내 호텔들을 대상으로 하였다. 이상치 영향을 미칠 수 있는 점유율 180% 이상과 비정상 데이터 일부를 제거하였고, 해당 기간 중 판매 데이터 레코드 수가 180개 이상(즉, 6개월 이상의 일별 판매 데이터)을 보유한 호텔 119개를 선정하여 범위에 포함되는 예약 데이터 정보 119,546건을 수집하였다. 시설 정보에는 객실 수, 호텔 등급, 지역 정보가 포함되었고, 예약 데이터 정보에는 일자, 판매 수, 개인 판매 수, 그룹 판매 수, 성인 수, 아동 수, 객실 판매가, 투숙 건수, 퇴실 건수, 워크인 수, 고장 룸 수, 직원 이용 룸 수, 대실 룸 수, 노쇼 건수, 예약 접수 건수, 취소 건수 등

이 포함되었다.

본 연구에서는 시계열 예측 모델을 고도화하는 제안모델을 실제 데이터로 검증하는 것이 목적이었기 때문에 ‘판매 수’ 단일 변수만을 활용하여 시계열 수요 예측 모델을 학습하였다. 구체적으로, t-30일부터 t-1일까지의 판매 수를 활용하여 t+14일의 판매 수를 예측하는 모델을 실험 대상으로 설정하였다. 예측의 대상이 되는 종속 변수를 t+14일로 설정한 이유는, 호텔 산업에서는 일반적으로 2주 후의 수요를 예측해야 사전에 대응 준비를 효과적으로 할 수 있기 때문이다(Pereira and Cerqueira, 2022).

전체 데이터는 시간 순서를 반영하여 학습, 검증, 시험용 데이터로 구분하였다. 데이터 구분 비율과 포함된 데이터 수는 <표 1>과 같다.

<표 1> 데이터 분할

데이터 구분	비율	데이터수
학습 데이터	70%	83,682
검증 데이터	15%	17,932
시험 데이터	15%	17,932

#### 4.2 실험설계

모델 성능 비교를 위한 실험에는 Python version 3.10.1이 사용되었다. 실험에 사용된 소프트웨어와 하드웨어는 다음 <표 2>에 자세히 기술되어 있다.

전역/지역 모델을 선택적으로 활용하는 시계열 예측 모델(제안 모델)을 적용하기 위해서는 그 근간이 될 기계학습 예측(regression) 기법을 먼저 결정해야 한다. 이를 위해 본 연구에서는 학습용 데이터와 검증용 데이터를 활용해 먼저 시계열 기반 기법 2개(LSTM, GRU)와 비시계열 기반 기법 3개(XGBoost, Decision Tree, Random Forest)를 적용해 보고, 이 중 가장 우수한 성능을 보이는 기법을 제안 모델의 기본모델(base

<표 2> 실험 환경

구분	데이터
하드웨어	CPU : Ryzen 9 7950x3D GPU : RTX 4090 M/M : 64GB
소프트웨어	OS : Ubuntu 22.04.4 LTS Python 3.10.12 GPU-Driver : nvidia-driver-535 Cuda : 12.2 Pytorch : 2.1.2 Tslern : 0.6.3 Scikit-learn : 1.2.2 Numpy : 1.23.5 Pandas : 1.5.3

model)로 활용하고자 하였다. 기본모델 학습을 위해 각 기법별로 적용된 파라미터는 다음의 <표 3>에 정리되어 있다.

한편 수요 예측 모델의 평가지표로는 일반적으로 MAE, MSE, RMSE, MAPE, SMAPE 등이 사용될 수 있다. 이 중, MAE (Mean Absolute Error)는 실제값과 예측값의 차이의 절대값의 평균을 나타내어 오차를 측정한다. MSE(Mean Square Error)는 실제값과 예측값 차이의 제곱의 평균으로 오차의 크기를 계산한다. RMSE(Root Mean Squared Error)는 MSE의 제곱근으로, 예측값과 실제값 사이의 차이를 제곱하여 평균한 뒤, 제곱근을 반영한 값이며, 작을수록 성능이 높다. MAPE(Mean Absolute Percentage Error)는 예측값과 실제값 사이의 백분율 오차의 평균으로, 작을수록 성능이 높다(Syahromi, 2019). SMAPE (Symmetric Mean Absolute Percentage Error)는 대칭 평균 절대 백분율 오차이며, 실제값과 예측값의 차이의 절대값을 실제값과 예측값 합의 절반으로 나눈 값의 평균으로 평가하며, MAPE의 변형으로, 0으로 나누는 문제를 방지하고 상대적 오차에 대한 공정성이 높다. 이 중에서 본 연구에서는 각 호텔마다 수요의 편차가 크게 나타날 수 있다는 점을 고려하여 MAPE를

<표 3> 기본모델 학습에 적용된 각 기법별 파라미터

기법	파라미터	값	비고
XGBoost	objective	Squarederror	
	Max_depth	5	
	Min_child_weight	5	
	subsample	0.5	
	Tree_method	hist	
LSTM	N_estimators	30	GRU 동일
	Hidden_size	32	
	Num_layers	2	
	dropout	0	
	Bidirectional	X	
	Optimizer	Adam	
Decision Tree	Learning_rate	0.001	Random Forest 동일
	Batch_size	2000	
	Window_length	9	
	criterion	Friedman_mse	
	Max_length	3	
	Max_features	Log2	
	Min_impurity_decrease	0.3573	
Min_sample_leaf	2		
Min_samples_split	4		

기준으로 최종 모델별 성능을 평가하고 비교하였다.

$$MAE = \frac{1}{n} \sum \left( \left| \frac{a-f}{a} \right| \right)$$

$$MSE = \frac{1}{n} \sum (a-f)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum ((a-f)^2)}$$

$$MAPE = \frac{1}{n} \sum \left( \left| \frac{a-f}{a} \right| \right) \times 100$$

$$SMAPE = \frac{100}{n} \sum \left( \frac{(|a-f|)}{((|a|+|f|)/2)} \right)$$

\* a : 실제값, f : 예측값

### 4.3 실험결과

전술한 실험 설계에 따라 수집된 데이터를 처리하고, 우선적으로 기본 모델을 설정하기 위해 각 기계학습 기법의 예측 성능을 측정하였다. <표 4>는 LSTM, GRU, XGBoost, Random Forest, Decision Tree 기법의 성능 및 학습 시간을 나타내고 있다. 이 표를 통해 알 수 있듯이, 14일 후의 수요를 가장 정확하게 예측하고 가장 적은 시간으로 학습할 수 있는 기법은 XGBoost로 확인되었다. 이에 전역-지역 모델의 선택적 적용을 위한 기본 모델로 XGBoost를 선정하고 이후 실험을 진행하였다.

제안 모델을 구현하기 위해서는 우선 지역 모델을 구분하기 위한 클러스터링을 적용하여야 한다. 본 연구에서는 시계열 데이터에 적합한 클러스터링 기법인 DTW K-means 클러스터링을 적용하였다. <표 5>는 본 연구에서 적용된 DTW K-means의 각종 파라미터 값을 나타내고 있다. 이 표에 제시되어 있듯이, 본 연구에서는 전체 호텔을 총 20개의 군집으로 구분하여 유사한 집단별로 분류하고자 하였다.

<표 4> 기본모델을 위한 각 기법별 성능 및 학습시간

기계학습 기법	MAPE	학습시간
LSTM	40.651	2h+
GRU	40.414	2h+
XGBoost	38.960	1m-
Random Forest	39.951	10m+
Decision Tree	47.749	5m+

<표 5> DTW K-means 파라미터 설정값

파라미터	값	비고
K	20	전체 군집의 수
init	20	초기 중심점 설정 횟수
iter	100	최대반복횟수
state	42	난수생성시드

DTW K-means를 통해 클러스터링 모델이 구축되면, 이 모델을 기준으로 학습 데이터를 구분한 뒤 동일한 군집으로 분류된 학습 데이터만 활용하여 각 군집별 지역 모델(local model)을 앞서 선택된 XGBoost 기법으로 학습하였다. 이렇게 학습된 각 지역 모델에 검증(validation) 데이터를 적용했을 때 얻은 예측 결과가 <표 6>에 정리되어 있다.

<표 6>에 제시된 바와 같이 각 군집별 측정 결과를 살펴보면, 전체 데이터를 이용해 학습한 전역 모델의 MAPE 38.961 대비 총 9개 군집에서 지역 모델의 MAPE가 상대적으로 더 우수한 것으로 확인되었다. 특히 Cluster 5와 Cluster 9를 기준으로 학습된 지역 모델 2개는 전역 모델 대비 매우 우수한 성능을 보여주었다. 종합적으로 볼 때 전체 20개의 군집 중에서는 MAPE 기준으로 평가할 때 ‘우수(10% 미만)’한 성능을 보이는 2개의 군집과 ‘중음(10% 이상 30% 미만)’ 수준으로 평가되는 4개의 군집이 있다(Syahromi, 2019). 본 연구에서는 군집마다 예측력의 차이를 초래하는 원인을 해석하기는 어렵지만, 검증용 데이터에 대해 우수한 성능을 보인 군집 정보를 활용하는 것이 모형 개선에 도움이 될 것으로 판단하였다. 이에 따라 상위 9개 군집(Cluster 5, 9, 2, 16, 3, 18, 1, 6, 12)에 대해서는 지역 모델이 전역 모델보다 우수한 것으로 나타나 지역 모델을 선택하도록 하였고, 나머지 11개 군집에 대해서는 전역 모델을 그대로 적용하는 것으로 제안 모델을 구성하였다.

끝으로 제안 모델의 최종적인 성능을 점검하

〈표 6〉 군집별 성능 및 전역/지역모델 상대평가표

순번	군집구분 (지역모델)	MAE	MSE	RMSE	SMAPE	MAPE	선택 MAPE	MAPE 개선%p
1	Cluster 5	4.274	53.504	7.315	4.848	5.395	5.395	622.129
2	Cluster 9	11.116	186.720	13.665	8.705	8.984	8.984	333.695
3	Cluster 2	9.705	157.310	12.542	12.401	13.258	13.258	193.868
4	Cluster 16	10.307	203.660	14.271	16.044	17.084	17.084	128.055
5	Cluster 3	21.200	529.170	23.003	18.255	20.581	20.581	89.306
6	Cluster 18	14.283	347.970	18.654	19.779	24.052	24.052	61.987
7	Cluster 1	13.449	345.070	18.576	39.729	32.114	32.114	21.321
8	Cluster 6	16.737	416.410	20.406	27.333	34.161	34.161	14.051
9	Cluster 12	17.192	458.590	21.415	33.946	36.611	36.611	6.419
10	Cluster 13	18.164	468.420	21.643	28.147	39.013	38.961	0.133
11	Cluster 17	19.035	536.320	23.159	32.624	48.251	38.961	19.253
12	Cluster 0	21.345	753.935	27.457	40.450	48.559	38.961	19.766
13	Cluster 15	17.374	470.980	21.702	38.135	50.985	38.961	23.583
14	Cluster 8	21.131	675.120	25.983	40.672	55.406	38.961	29.681
15	Cluster 4	18.066	535.320	23.137	45.961	63.665	38.961	38.803
16	Cluster 11	17.258	440.120	20.979	41.232	66.670	38.961	41.561
17	Cluster 10	20.886	662.510	25.739	50.166	72.661	38.961	46.380
18	Cluster 14	21.466	690.870	26.284	85.697	77.526	38.961	49.745
19	Cluster 19	21.121	682.450	26.124	105.346	94.554	38.961	58.795
20	Cluster 7	10.299	410.220	20.254	123.490	108.640	38.961	64.138
	지역모델 평균	13.449	345.07	18.576	32.114	39.729	-	-
-	전역모델	13.231	329.25	18.145	31.748	38.961		

기 위해 시험 데이터를 기준으로 제안 모델의 성능을 확인해 보았다. <표 7>은 시험 데이터를 기준으로 적용했을 때 도출된 제안 모델과 비교 모델의 성능을 나타내고 있다. 비교 모델로는 100% 전역 모델을 적용했을 때의 실험 결과와 100% 지역 모델을 적용했을 때의 실험 결과를

사용하였다. 이를 통해 본 연구에서 제안하는 전역과 지역 모델의 선택적 적용이 실제로 효과적임을 확인하기 위함이다.

DTW K-means 클러스터링 전처리 없이 전체 학습 데이터를 모두 사용한 전역 수요 예측 모델의 성능은 MAPE 38.960으로 나타났다. 이는 검증 데이터 기준 전역 모델 성능인 MAPE 38.961과 거의 동일한 값이다. 반면, DTW K-means 클러스터링 결과에 따라 각 군집별 지역 모델만을 사용해 예측할 경우 성능은 MAPE 39.729로 나타나 오히려 전역 모델을 사용하는 것보다 성능이 하락하는 것으로 나타났다.

앞서 검증 데이터를 통해 확인한 바와 같이,

〈표 7〉 제안모델과 비교모델의 성능 비교

모델		MAPE	제안모델 적용시 개선 %p
비교 모델	전역모델	38.960	↓3.80
	지역모델	39.729	↓5.66
제안모델		37.481	

전역 모델 성능을 능가한 지역 모델은 9개인 반면, 전역 모델 대비 성능이 하락한 지역 모델은 11개로 나타났다. 시험 데이터를 기준으로 해도 이와 유사한 패턴의 결과가 나타났다.

검증 데이터를 기준으로 성과에 따라 선택된 전역 모델 또는 지역 모델을 선별적으로 적용하는 본 연구의 제안 모델에서 획득된 MAPE는 37.481로 나타났다. 이는 전역 모델을 적용했을 때 대비하여 3.80% 성능 향상을 나타내며, 모두 지역 모델을 선택한 경우에 비해서는 5.66%의 성능이 향상된 것이다. 이를 통해 본 연구의 제안 모델이 유효함을 확인할 수 있다.

## V. 결론

호텔 객실 수요 예측 모델은 최근 온라인 다양성 강화와 확산 등으로 인해 비정형 패턴을 포함하게 되었다. 이러한 비정형성을 반영하기 위해 기계학습을 이용한 수요 예측 모델이 통계적 시계열 분석보다 효과적임이 다수의 연구를 통해 입증되었다. 본 연구 역시 기계학습 모델 기반의 호텔 객실 수요 예측 모델을 제안하고 있다. 본 연구에서는 다양한 지역, 형태, 특성을 가진 호텔의 수요 데이터에 정규화를 적용하여 일반화된 모델을 구축하고, ‘전역/지역모델의 선택적 활용’을 제안하여 성능을 개선하고자 하였다. 실증분석 결과, 제안 모델은 전적으로 전역 모델이나 지역 모델에 의존하는 방식보다 더 나은 예측 정확도를 보였다.

본 연구는 통계자료나 특정 장소, 시설기반의 제한적 데이터를 활용하여 개발된 기존 수요 예측 모델의 일반화 성능을 향상시키기 위해 호텔 관광 분야의 다수 지역, 다양한 특징, 다양한 규모의 데이터를 결합한 통합 수요 예측 모델을 제안하였다. 또한, DTW K-means 클러스터링을 통해 지역 모델을 활용하는 방안을 제시하여 학습 시간 단축 및 예측 정확도 향상을 도모하였

다. 하지만, 실증분석 과정에서 DTW K-means 클러스터링을 통한 전처리는 정보 감소 (information loss)의 부작용을 유발하여 전역 모델 대비 예측 성능을 감소시킬 수 있는 위험이 있었다. 이를 극복하기 위해 본 연구에서는 전역 모델과 지역 모델을 선택적으로 활용할 수 있는 새로운 접근법을 제안하였고, 이 접근법을 통해 전역 모델이나 지역 모델을 단독으로 사용할 때보다 MAPE 기준으로 3.8~5.66%의 개선 효과를 확인하였다.

이러한 본 연구가 갖는 학술적 의의를 정리해 보면 다음과 같다. 첫째, 기존 연구들은 수요 예측을 위한 기계학습 알고리즘 자체를 비교하거나 개선하는 것에 초점을 맞추었지만, 본 연구는 클러스터링 전처리를 통해 성능 향상을 도모했다는 점에서 일반적인 기존 수요 예측 연구들과 차별점을 갖는다. 둘째, Kim & Ahn(2008) 등 클러스터링을 전처리로 적용한 예측 모델을 제안한 기존 연구들과 달리, 본 연구는 각 군집별 지역모델과 전체 데이터 대상의 전역모델을 비교하여 선택적으로 사용하는 방법을 제안함으로써, 다양한 모델을 효과적으로 적용하는 보다 강건한 접근법을 제안하였다는 점에서 학술적 의의가 있다. 셋째, 본 연구에서는 제안모델을 개념적인 수준에서만 제안하지 않고, 실제 호텔 수요 데이터에 적용하여 그 성능을 확인하였다. 이렇게 실증분석을 통해 제안모델의 유용성을 검증한 점 역시 본 연구의 학술적 의의라고 할 수 있다.

한편 본 연구의 실무적 의의는 다음과 같다. 첫째, 여행 및 호텔산업 분야에서 현재 인공지능과 기계학습의 적용이 두드러지고 있으며, 특히 수요 예측은 중요한 분야로 부상하고 있다. 이는 온라인 확산, 다양한 매체 증가, 강제 디지털화 등으로 인해 해당 산업에서의 경쟁은 더욱 치열해지고 있고, 수익성 확보가 생존에 매우 중요한 역할을 하고 있으며, 정확한 수요 예측

을 통해 수익확보를 위한 최적의 가격 설정과 운영 효율화가 가능하기 때문이다. 본 연구에서 제안하는 수요 예측 방법론은 이러한 요구를 충족시키고 있기 때문에 실무적으로 유용하게 활용될 수 있다는 의의가 있다.

둘째 제안모델은 전역모델과 지역모델을 함께 사용하는 접근법을 제안하고 있어, 효율적인 모델 관리와 운영이 가능하다. 예를 들어, 전역 모델은 자주 재학습하기 어려울 수 있지만, 상대적으로 가벼운 지역모델은 재학습을 통한 성능 개선이 가능할 것으로 기대된다. 또한 적용 단계에서도 지역모델이 경량화 될 수 있으므로, 더 빠른 대응이 가능할 것으로 예상된다.

끝으로 본 연구에서 제안하는 접근법은 호텔 수요 예측뿐만 아니라 시계열 데이터를 활용한 다양한 예측에 적용 가능한 범용성을 가지고 있다. 따라서, 타 분야로의 응용도 추후 시도해 볼 수 있을 것이다.

이처럼 다양한 의의가 있지만 본 연구는 다음과 같은 한계 역시 갖고 있다. 우선 기본적으로 수요량이라는 단일 변수만 활용하고 있기 때문에 실증분석 과정에서 전체적인 예측의 성능이 낮게 나타났다. 후속 연구에서는 보다 다양한 독립변수들을 활용하여, 성능이 개선된 예측 모델을 개발하는 노력이 필요할 것으로 보인다.

둘째, 최근 학계에서는 TFT(Temporal Fusion Transformer) 등 최신 트랜스포머에 기반한 보다 고도화된 딥러닝 기반의 시계열 예측 기법이 제안되고 있는데, 본 연구에서는 이런 최신의 기법들이 적용되지 않았다. 향후 연구에서는 분석 대상이 되는 기계학습 예측 알고리즘을 더욱 확대하여 다양한 형태의 비교 분석이 요구된다.

셋째, 전술한 바와 같이, 본 제안모델은 호텔 수요 예측 외에도 모든 경영 분야의 시계열성 예측에 적용 가능하다. 따라서, 농산물 가격예측, 전력 수요예측 등 시간의 흐름에 따라 변화하는 다른 예측 문제에도 적용을 해 보고 새로

운 도메인에서도 과연 유효하게 성능 개선이 이루어지는지 후속 연구를 통해 검증할 필요가 있다.

이러한 다양한 의의와 한계를 통해 본 연구는 호텔 산업뿐만 아니라 다양한 시계열 예측 문제에 적용 가능한 유용한 예측 모델을 제안하였다. 향후 연구를 통해 더욱 발전된 모델을 개발하고, 다양한 분야에 적용해 나갈 필요가 있다.

## 참 고 문 헌

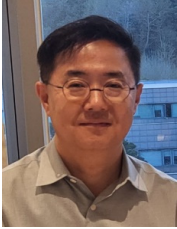
- [1] Ampountolas, A. (2021). Modeling and forecasting daily hotel demand: A comparison based on SARIMAX, neural networks, and GARCH models. *Forecasting*, 3(3), 580-595.
- [2] Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., & Seaman, B. (2019). Sales demand forecast in e-commerce using a long short-term memory neural network methodology. *Computer Science Statistics*. Retrieved from <http://arxiv.org/abs/1901.04028>
- [3] Bouchareb, N. (2023). The role of artificial intelligence in improving hotels property management systems (PMS). *International Journal of Economic Performance*, 6(2), 554-567.
- [4] Castelvechi, D. (2016). Can we open the black box of AI? *Nature*, 538(7623), 20-23.
- [5] Dowlut, N., & Gobin-Rahimbux, B. (2023). Forecasting resort hotel tourism demand using deep learning techniques – A systematic literature review. *Heliyon*, 9(7), e18385.
- [6] Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H. S., & Schmidt-Thieme, L. (2021). Do we really need deep learning models for time series forecasting? *Computer Science Statistics*, Department of Computer Science, University of Hildesheim, Germany. Retrieved from <http://arxiv.org/abs/2105.08885>

- org/abs/2101.02118
- [7] Gulmez, M., Ajanovic, E., & Karayun, I. (2015). Cloud-based vs desktop-based property management systems in hotel. *USV Annals of Economics & Public Administration*, 15(1), 160-168.
- [8] Huang, L., & Zheng, W. (2021). Novel deep learning approach for forecasting daily hotel demand with agglomeration effect. *International Journal of Hospitality Management*, 98.
- [9] Kim, D., Shyn, S. K., Kim, D., Jang, S., & Kim, K. (2021). A daily tourism demand prediction framework based on multi-head attention CNN: The case of the foreign entrant in South Korea. *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*.
- [10] Kim, K., & Ahn, H. (2008). A recommender system using GA K-means clustering in an online shopping market. *Expert Systems with Applications*, 34(2), 1200-1209.
- [11] Laaroussi, H., Guerouate, F., & Sbihi, M. (2020). Deep learning framework for forecasting tourism demand. *IEEE International Conference on Technology Management*.
- [12] Livieris, I., Stavroyiannis, S., Pintelas, E., & Pintelas, P. (2020). A novel validation framework to enhance deep learning models in time-series forecasting. *Neural Computing and Applications*, 32(23), 17149-17167.
- [13] Oviedo, E. H. S., Travé-Massuyès, L., Subias, A., Pavlov, M., & Alonso, C. (2023). DTW K-means clustering for fault detection in photovoltaic modules. XI Congreso Internacional de Ingeniería Mecánica, Mecatrónica y Automatización, Carthagene, Colombia. Retrieved from <http://arxiv.org/abs/2306.08003>
- [14] Pereira, L. N., & Cerqueira, V. (2022). Forecasting hotel demand for revenue management using machine learning regression methods. *Current Issues in Tourism*, 25(17), 2733-2750.
- [15] Phumchusri, N. (2021). Forecasting hotel daily room demand with transformed data using time series methods. *Journal of Revenue and Pricing Management*, 22(1), 44-56.
- [16] Pucciani, K. K., & Murphy, H. C. (2011). An investigation of data management and property management systems in hotels. *Tourism & Hospitality Management*, 17(1), 101-114.
- [17] Rusman, J. A., Chunady, K., Makmud, S. T., Setiawan, K. E., & Hasani, M. F. (2023). Crude oil price forecasting: A comparative analysis of ARIMA, GRU, and LSTM models. *IEEE 9th International Conference*.
- [18] Syahromi, T., & Sumitra, I. D. (2019). Forecasting hotel expenses using the ARIMA method. *IOP Conference Series: Materials Science and Engineering*, 662(2).
- [19] Zhang, Q., Qiu, L., Wu, H., Wang, J., & Luo, H. (2019). Deep learning based dynamic pricing model for hotel revenue management. *2019 International Conference on Data Mining Workshops*, 370-375.
- [20] Zhao, J., Wang, Z., Zhang, Z., & Han, Y. (2019). A combined model based on GM and SARIMA: An example of excavator demand forecasting. *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 232-236.
- [21] 강보람, & 안현철. (2022). 리조트 교차판매 예측모형 개발 및 SHAP을 이용한 해석. *한국빅데이터학회지*, 7(2), 195-204.
- [22] 권보성, 박래준, & 송경빈. (2018). 다양한 정규화 방법에 따른 평일 단기 전력수요예측 정확도 분석. *조명·전기설비학회논문지*, 32(6), 30-33.
- [23] 김민교, & 강대진. (2015). 온라인 쇼핑 사이트

- 의 추천 시스템을 위한 고객패턴 분석 모델. 한국IT서비스학회 학술대회 논문집, 467-470.
- [24] 김연화. (1995). 판매수익 극대화를 위한 전략적 객실가격 수립에 관한 연구. 사회과학연구, 3, 119-156.
- [25] 김재성, 양여진, 오민지, 이성웅, 권순동, & 조완섭. (2020). 딥러닝 기법을 활용한 가구 부자재 주문 수요예측. 한국빅데이터학회지, 5(2), 111-120.
- [26] 김정아, 정종필, 이태현, & 배상민. (2018). 중소기업 제조공장의 수요예측 기반 재고관리 모델의 효용성 평가. 한국인터넷방송통신학회논문지, 18(2), 197-207.
- [27] 김진섭, 황재성, & 정재우. (2020). 시계열 분해 데이터를 이용한 LSTM 기법 기반 항공기 수리 부속 수요예측 방안 연구. 경영과학, 37(2), 1-18.
- [28] 박득희, 강상훈, & 이계희. (2020). 지속가능한 관광성장을 위한 방한 일본인 관광객 수요예측: 시계열 계량 모형 적용. 관광연구저널, 34(3), 47-60.
- [29] 배선경. (2005). 마케팅관점에서 본 수익률 관리 전략의 중요 요소. 관광연구저널, 19(1), 249-259.
- [30] 서보원, 이선덕, 곽용진, & 장호림. (2021). k-means clustering을 활용한 음성 인식 성능 개선. 한국정보과학회 학술발표논문집, 6, 940-942.
- [31] 신동하, 최광호, & 김창복. (2017). RNN과 LSTM을 이용한 주가 예측을 향상을 위한 딥러닝 모델. 한국정보기술학회논문지, 15(10), 9-16.
- [32] 유현선, & 유선중. (2017). 패널 연립방정식 모형을 이용한 호텔 객실의 단기 수요 및 가격 결정요인 분석. 부동산연구, 27(2), 95-109.
- [33] 이민상, & 심완섭. (2002). 인터넷쇼핑몰 거래 상품의 수요예측에 관한 연구. 유통연구, 7(2), 15-42.
- [34] 이승운, & 곽기영. (2022). GRU 기법을 활용한 서울시 공공자전거 수요예측 모델 개발. 지능정보연구, 28(4), 1-25.
- [35] 이인지, & 윤현식. (2020). 머신러닝을 활용한 지역축제 방문객 수 예측모형 개발. 정보시스템연구, 29(3), 35-52.
- [36] 이현찬, 장스위, 임택, & 양위주. (2023). SARIMA를 이용한 국내 호텔 객실매출에 관한 시계열 분석 연구. 관광연구, 38(1), 47-64.
- [37] 장스위, 임택, & 양위주. (2022). 시계열 분석을 통한 관광목적지 수요예측 - 감천문화마을 중심으로. 산업혁신연구, 38(3), 113-120.
- [38] 정경훈. (1985). 관광호텔 객실가격 결정방법에 관한 연구. 관광학연구, 9(1), 239-256.
- [39] 정은별. (2017). 서울소재 특급호텔 객실 요금에 영향을 주는 호텔 속성에 대한 연구. 관광학연구, 41(3), 131-149.
- [40] 정호철, & 선영규. (2019). 에너지인터넷에서 1D-CNN과 양방향 LSTM을 이용한 에너지 수요예측. 전기학회논문지, 68(1), 134-142.
- [41] 조성주, 최우석, & 최상현. (2023). 앙상블 기반의 악취 농도 다지역 통합 예측 모델 개발. 지능정보연구, 29(3), 383-400.
- [42] 주일택, & 최승호. (2018). 양방향 LSTM 순환신경망 기반 주가예측모델. 한국정보전자통신기술학회논문지, 11(2), 204-208.
- [43] 최재식. (2019). 설명가능 인공지능 연구동향. 정보과학회지, 37(7), 8-14.



## 저 자 소 개



### 박 기 현(Keehyun Park)

- 1993년 3월 : 아주대학교 (공학사)
- 2018년 3월 : 연세대학교 경영전문대학원 (경영학석사)
- 2022년 3월~현재 : 국민대학교 비즈니스IT전문대학원 비즈니스IT전공 (박사과정)

<관심분야> 프로세스 인텔리전스 및 인공지능 기반 고객경험 강화, 수요예측을 통한 수익관리



### 정 경 호(Gyeongho Jung)

- 2023년 2월 : 한라대학교 스마트자동차공학 (공학사)
  - 2023년 3월~현재 : 국민대학교 비즈니스IT전문대학원 비즈니스IT전공 (석사과정)
- <관심분야> 기계학습과 딥러닝, 추천시스템, 강화학습



### 안 현 철(Hyunchul Ahn)

- 1999년 2월 : KAIST 산업경영학과 (공학사)
  - 2002년 8월 : KAIST 테크노경영대학원 경영공학 (석사)
  - 2006년 8월 : KAIST 테크노경영대학원 경영공학 (박사)
  - 2009년 3월~현재 : 국민대학교 비즈니스IT전문대학원 교수
- <관심분야> 지능형 의사결정지원시스템, 고객관계관리