

# COVID-19 사례를 통한 도시 내 비정상적 수요 예측을 위한 시계열 모형 파이프라인 개발 연구

## Time Series Modeling Pipeline for Urban Behavioral Demand Prediction under Uncertainty

진 민 수\* · 이 동 우\*\* · 김 영 록\*\*\* · 이 현 수\*\*\*\*

\* 주저자 : 한국건설기술연구원 지역협력진흥실 전임연구원  
 \*\* 교신저자 : 인천대학교 도시행정학과 교수  
 \*\*\* 공저자 : 한국건설기술연구원 도로교통연구본부 연구위원  
 \*\*\*\* 공저자 : 인천대학교 도시행정학과 학생연구원(학부과정생)

Minsoo Jin\* · Dongwoo Lee\*\* · Youngrok Kim\*\*\* · Hyunsoo Lee\*\*\*\*

\* Regional Cooperation & Promotion Division, Korea Institute of Civil Eng. and building Technology  
 \*\* Dept. of Urban Planning, Univ. of Incheon National  
 \*\*\* Dept. of Highway & Transportation Research, Korea Institute of Civil Eng. and building Technology  
 \*\*\*\* Dept. of Urban Planning, Univ. of Incheon National

† Corresponding author : Dongwoo Lee, dlee@inu.ac.kr

Vol. 22 No.2(2023)  
 April, 2023  
 pp.80~92

pISSN 1738-0774  
 eISSN 2384-1729  
<https://doi.org/10.12815/kits.2023.22.2.80>

Received 15 March 2023  
 Revised 23 March 2023  
 Accepted 26 April 2023

© 2023. The Korea Institute of  
 Intelligent Transport Systems. All  
 rights reserved.

### 요 약

도시에 많은 사람들이 밀집하여 살아가면서 기존에 예측하지 못했던 범죄, 사고, 감염병 등의 비정상 이벤트가 발생은 도시 내 이용자 수요에 영향을 미치게 된다. 이러한 불확실성(uncertainty)이 내포된 정보를 기반으로 도시 내 이용자 수요에 대한 시계열적 예측을 수행한다면 신뢰성 있는 결과 도출이 불가능하다. 특히, 2020년 초 발발한 COVID-19는 비정상적인 이동통행패턴의 변화를 불러 일으키며 시계열 수요예측을 어렵게 만들었기에 이러한 변화를 감지하고 이를 반영하여 정확한 수요를 예측 수행할 수 있는 방법론의 필요성이 대두되고 있다. 이에 본 연구는 COVID-19로 인한 비정상적 이벤트를 자동으로 감지하고 예측하는 모형 파이프라인을 구축하였다. 이는 도시 내 다양한 분야에서의 불규칙적이고 비정상적인 이벤트로 인한 수요변화가 일어나는 상황에 폭넓게 활용될 수 있을 것으로 생각된다.

핵심어 : 수요예측, 시계열예측, 이상치 감지, 불규칙한 시계열 이벤트, 실시간 예측

### ABSTRACT

As cities are becoming densely populated, previously unexpected events such as crimes, accidents, and infectious diseases are bound to affect user demands. With a time-series prediction of demand using information with uncertainty, it is impossible to derive reliable results. In particular, the COVID-19 outbreak in early 2020 caused changes in abnormal travel patterns and made it difficult to predict demand for time series. A methodology that accurately predicts demand by detecting and reflecting these changes is, therefore, required. The current study suggests a time series modeling pipeline that automatically detects and predicts abnormal events caused by COVID-19. We expect its wide application in various situations where there is a change in demand due to irregular and abnormal events.

Key words : Demand prediction, Time series prediction, Outlier detection, Abnormal time series events, Real-time prediction

## I. 서론

현대의 사람들은 밀집하여 도시를 형성하고 그 도시 속에서 살아가고 있다. 이러한 도시의 밀집된 인구가 각자의 경제생활을 원활하게 할 수 있도록 큰 역할을 하는 것이 바로 도시 인프라이다. 대표적인 도시 인프라는 도로, 철도, 지하철, 공항, 전기, 통신, 상·하수도 등이 있다. 이러한 도시 인프라는 도시의 근간이 되는 기본적인 시설로써 기본적인 생산과 생활을 위해 필요한 각종 시설이다. 도시 인프라는 도시의 건설이 완료된 후 새로운 도시 인프라를 추가하는 것이 매우 어렵기 때문에 도시 건설에 있어서 설계 시부터 반영하여 구축하여야 한다. 또한 도시 인프라가 그 수요에 비해 공급이 턱없이 부족하여 그 기능을 제대로 하지 못하게 된다면 도시 속 많은 인구의 경제활동과 생활이 원활하게 이루어지지 않을 수 있으며, 수요에 비해 너무 많은 공급이 이루어지면 자원의 비효율적인 사용으로 경제성장에 악영향을 미친다. 그렇기 때문에 도시 인프라의 수요에 대한 정확도 있는 예측은 도시에서 생활하는 사람들의 삶의 질을 향상시키고, 효율성 있는 경제활동을 영위할 수 있도록 하며, 도시를 성장시키기 위해서 중요한 부분이다.

그러나 도시에 많은 사람들이 밀집하여 살아가면서 기존에 예측하지 못했던 범죄, 사고, 감염병 등의 비정상 이벤트가 발생은 도시 내 이용자 수요에 영향을 미치게 된다. 이러한 불확실성(uncertainty)이 내포된 정보를 기반으로 도시 내 이용자 수요에 대한 시계열적 예측을 수행한다면 신뢰성 있는 결과 도출이 불가능하다. 최근 도시의 불규칙한 변화가 발생한 대표적인 케이스가 COVID-19 이다. COVID-19는 2020년 초 발발해 전 세계적으로 유행하고 있는 전염성이 강한 질병으로, 우리나라는 COVID-19 확산 초기부터 일일 확진자 수와 확진자 동선관리를 바탕으로 사회적 거리두기를 실시하고, 집합금지, 대중교통 감축운행 방역패스 등의 감염을 예방하고 확산을 방지하기 위한 정부정책을 통해 대응하고 있다. 이로 인해 국민들의 생활은 큰 폭으로 변화하였으며, 이동통행패턴 또한 큰 변화를 보였다.(Lim, 2020) COVID-19가 발발한 직후인 우리나라의 버스, 지하철과 같은 대중교통수단은 2021년 3월의 통행량이 2019년 3월 대비 47.0%, 42.7% 감소한 것으로 나타났다(KOTSA and MOLIT, 2021). 이처럼, COVID-19로 인해 이동통행패턴이 기존과 큰 폭으로 변화함에 따라 도시 인프라에 대한 정확도 있는 수요예측이 어려워졌고, 이러한 비정상적인 수송수요의 변동 속에서도 정확도 있는 예측을 할 수 있는 새로운 방법의 필요성이 대두되고 있다.

이러한 불확실성을 가지는 데이터를 보정하기는 방법은 과거 자료를 바탕으로 한 단순한 통계기법인 선형 보간법이나 Factor법을 여전히 활용하고 있는 실정이다. 데이터가 가지는 불확실성의 문제는 장래 도입되는 차세대 도시 인프라 관리 시스템에서도 피할 수 없는 문제가 될 것으로 보인다. 스마트 시티 내의 센서 기반으로 수집되는 시계열적 데이터의 경우 실시간으로 방대한 규모의 자료가 축적되지만, 이러한 첨단 장비를 통해 수집된 데이터도 이상 및 결측값은 항상 존재하게 된다. 따라서 이러한 변화로 인한 불확실성을 감지하여 예측하고, 또한 이러한 예측 시스템이 자체적으로 지속적으로 업데이트될 수 있는 모형 파이프라인 구축이 매우 중요한 시점이다.

본 연구는 COVID-19라는 비정상적인 시계열 이벤트와 같은 불확실성을 인한 수요 변화를 자동으로 감지하고 정확한 예측값을 도출해 내기 위한 모형 파이프라인을 구축하고자 하였다. 특히, 예측력이 보장된 머신 기반의 모형을 개발하여 데이터 수집 및 모니터링, 이상치 검지, 특성 공학(시계열 변수 추출), 예측 주기 설정, 모형 구축, 예측력 검증의 전 과정을 반복적으로 수행할 수 있는 아키텍처를 제시하고자 한다. 이를 검증하기 위해 2017년 6월부터 2022년 6월까지의 일 단위 교통량 데이터를 기반으로 모형을 구축하고 COVID-19로 인한 수요의 급격한 감소세를 시계열 예측 모형에 반영하여 불확실성에 대한 반영이 가능한 시계열 예측 모형을 구축하였다. 특히, 시계열적 수요예측모형의 범용성을 높이기 위하여 이상치를 검지하는 알고리즘을 결합하고 모형 파이프라인을 개발 및 적용하였다.

## II. 기존 문헌 고찰

기존의 수송수요예측은 전통적 4단계 수요추정모형과, 중력모형에 근거한 직접수요모형, 지수 평활법등 여러 방법이 사용되어 왔지만 (Kim and Kim, 2019) 이는 단순한 통계모형을 사용하고 있고, 과거의 일정 시점을 기초로 하여 구한 자료로 모형화를 하기 때문에 데이터가 많이 필요하며, 비정상적인 이벤트의 반영이 힘들기 때문에 장래를 예측하는 것에 경직성이 있다. 또한 모형구축에도 시간이 오래걸린다는 단점을 가지고 있기 때문에 대표적으로 많이 사용되어온 모형은 계절 개입형 시계열 모형(seasonal autoregressive integrated moving average, 이하 SARIMA)이다. SARIMA모형은 주기성 및 계절성이 강한 데이터를 시계열 예측에 사용하기 위해 자기상관 모형(Auto Regression, AR)과 오차항(E)으로 인한 영향을 반영한 이동평균모형인 MA(Moving Average, MA)가 결합된 모형에 자기상관 모형인 SAR(Seasonal Auto Regressive)과 계절 이동평균 모형인 SMA(Seasonal Moving Average)가 결합된 분석기법이다. Kim and Kim(2022)는 SARIMA 모형을 통해 COVID-19의 확진자 수를 예측하는 모형을 구축하였고, Noh and Do(2015)는 SARIMA 모형을 통해 새마을-무궁화 열차의 주요 5개노선(경부선, 호남선, 전라선, 강항선, 중앙선)의 수송수요를 예측하였다. 그러나 위 연구에서는 수송수요가 크게 벌어지는 주중과 주말의 예측모형을 따로 구축하며, 이는 COVID-19와 같은 비정상적인 시계열 이벤트 또한 제대로 반영하여 정확한 예측값을 도출해 내는 것에 어려움이 있기 때문이라고 생각된다. 이러한 비정상적 이벤트를 반영하여 정확도 높은 예측값을 도출해 내기 위해 컴퓨터 기술의 발달과 빅데이터의 활용이 가능해진 최근에는 머신러닝 및 딥러닝을 통한 방법에 대한 연구 또한 활발히 이루어 지고 있다. Sin et al.(2022)의 연구에서는 기존 시계열 모형과 LSTM(Long Short-Term Memory), GRU(Gated Recurrent Unit)의 딥러닝 모델을 통해 코로나 확진자수를 예측하였고, 또한 백신 접종률을 변수로 하여 정확한 예측을 통해 예상 확진자수를 구하고 정책 수립에 도움을 주는 SARIMA모형에 외생변수를 추가시킨 SARIMAX를 활용한 모형을 개발하였다. 그 결과 외생변수로 백신 접종률을 투입한 SARIMAX 모형이 딥러닝 모형인 LSTM, GRU에 비해 예측 수준이 실제값에 더 근접하였으며, 부스팅을 통해 딥러닝 모형의 예측 정확도와 신뢰도를 향상시켰다. Bajari et al.(2015)의 연구에서는 머신러닝을 통한 수요예측 방법들에 대해서 설명하고, 예측력을 분석하였으며, Han et al.(2022)의 연구에서는 딥러닝 기법의 RNN(Recurrent Neural Network, 순환신경망)을 통해 대중교통 이용량을 예측하고, LSTM, ARIMA, HA의 기법과 비교하였다. 그 결과 단기예측에서 RNN의 예측이 가장 정확한 것으로 확인되었고 이는 COVID-19로 인한 대중교통 이용에 대한 사람들의 인식이 단기적임을 확인하였다.

Oh et al.(2020)의 연구에서는 머신러닝 기반의 Prophet을 통한 마늘 가격의 장기예측과 함께, SARIMA 모형과 LSTM 모형과의 예측력 차이또한 살펴보았다. 이때 Prophet모형은 장기예측에 있어서 오차가 다른 모형들에 비해 작은 것으로 나타났으며, 불안정한 마늘가격의 변동 속에서도 우수한 예측 성능을 보여주고 있다는 것을 알 수 있다. 따라서 본 연구는 COVID-19라는 비정상적인 시계열 이벤트가 발생한 상황속에서 보다 편리하고 단기 의사결정에 있어서 도움이 될 수 있는 정확도 높은 예측값을 도출해 내기 위한 모형을 구축하는 것을 목적으로 하기 때문에 외부환경효과를 모두 반영하기 힘들고, 시계열마다 외생변수를 모두 반영해 주어야 하는 SARIMAX를 사용하기 어렵다고 판단하였다. 또한 비정상적인 시계열 데이터가 충분히 확보되지 않았고, 모형구축에 시간과 비용이 많이 들며 단기 예측에서 Prophet과 SARIMA보다 예측력이 떨어지는 딥러닝 기법 또한 본 연구에 사용하기 적합하지 않다고 판단하였다.

### Ⅲ. 연구방법론

#### 1. 분석방법론

앞서 말했듯이, 기존 시계열적인 교통수요 예측에서는 주로 Winter(1960)의 계절지수 평활법을 기반으로 한 계절형 ARIMA모형, 계절 개입형 시계열 모형인 SARIMA모형 등이 사용되어왔다.(Han et al., 2021). 그러나 ARIMA 모형에는 시계열이 일정한 주기성(정상성, stationary state)을 가져야 안정적인 결과값을 도출해 낼 수 있다는 제약이 존재하는데, 최근 COVID-19로 인해 수요의 주기성이 불규칙하게 변화하였으며, 수요예측에서 비정기적인 이벤트(holiday)의 영향이 커졌다.

시계열 예측에는 다양한 알고리즘을 적용할 수 있는데 AutoML의 시계열 예측용 라이브러리인 AutoTS를 활용하여 사전분석을 수행하여 다양한 시계열 알고리즘의 성능을 검증하였다. LSTM과 Prophet의 설명력이 가장 높게 나타났지만, 시스템과의 결합을 고려하여 컴퓨터 연산이 상대적으로 가볍고 직관적인 Prophet을 선정하고 이를 기반으로 하여 모형 알고리즘을 일부 변경 및 개발하는 과정을 수행하였다. 분석 시 모든 분석을 python을 활용하여 수행되었다.

Prophet 모형은 페이스북(現, 메타)에서 개발한 비즈니스 모형으로, ARIMA와 달리 시간에 대한 영향력을 고려하지 않기 때문에 불규칙한 주기성이 발생하는 구간에 대한 설명력이 월등히 높으며, Prophet 알고리즘은 시계열의 추세 및 계절성 변화 시점을 자동으로 파악하여 이를 반영하기 때문에 기존 통계모형에 비해 큰 장점으로 볼 수 있다. 또한 이상치 및 결측값의 영향을 비교적 적게 받으며, 시계열의 급격한 변화에도 안정적으로 예측값을 도출해낼 수 있다. 일반적으로 Prophet 모형은 일반화 가법 모형(Generalized Additive Model, GAM)을 기반으로 추세성(trend), 계절성(seasonality) 및 비정기적인 사건을 나타내는 휴일(holiday) 효과를 반영하여 예측값을 도출하지만, 시계열의 잔폭이 변화하는 수준에 따라 승법 모형 또한 적용하여 예측값을 도출해 낼 수 있다.

Prophet 모형의 수식은 다음과 같으며, 여기서  $g(t)$ 는 시계열의 비주기적인 변화를 모형화한 추세(trend) 함수이고,  $s(t)$ 는 주간 및 연간 계절성(seasonality)을 나타내며,  $h(t)$ 는 휴일(holiday)효과,  $\epsilon_t$ 는 오차항으로 모형에 반영되지 않는 모든 특이한 변화를 나타낸다.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \dots\dots\dots (1)$$

- 여기서  $g(t)$  : 시계열의 비주기적인 변화를 모형화한 추세(trend) 함수
- $s(t)$  : 주간 및 연간 계절성(seasonality)
- $h(t)$  : 휴일(holiday) 효과
- $\epsilon_t$  : 오차항, 즉 모형에 반영되지 않는 모든 특이한 변화

Prophet 모형은 통계 관련 지식 없이도 쉽게 이해하고 적용하여 예측값을 도출해 낼 수 있는 직관적인 파라미터를 가지고 있어 사용자가 쉽게 예측 방법을 조정하고 예측 성능을 개선할 수 있다는 장점을 지닌다.

본 연구에서 활용한 수요 데이터는 시계열 수준에 따라 잔폭이 달라지므로 가법 모형이 아닌 승법 모형을 적용해야 할 것으로 판단하였고 이는 다음과 같다.

$$y(t) = g(t) \times s(t) \times h(t) \times \epsilon_t \dots\dots\dots (2)$$

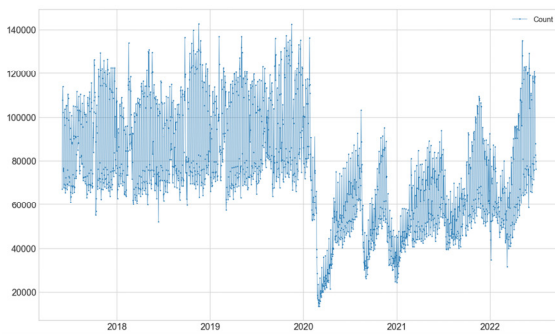
## 2. 분석모형구축

일반적인 Prophet 모형의 경우 시계열이 가진 주기성 및 계절성이 제대로 반영이 불가능하다. 또한, COVID-19 라는 비정상적인 이벤트로 인한 급격한 수요감소 때문에 모형 적합지 제대로 이루어지지 않으므로 예측의 정확도가 떨어진다. 이에 예측 모형의 정확도를 높이기 위해서 휴일 및 계절성을 반영하고, COVID-19의 비정상적인 이벤트를 반영하여 예측의 정확도를 향상시켰다.

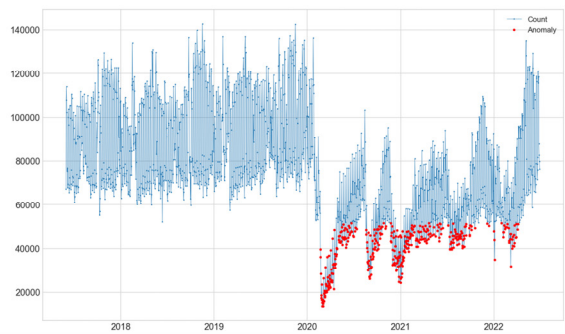
이러한 시계열 예측에서 중요한 이상치의 경우 기준값이 존재하지 않는 데이터의 경우 시스템 상에서 판단이 거의 불가능한 경우가 많다. 따라서 과거 정상적 시계열 정보를 학습하여 이상치를 검지하는 알고리즘을 Prophet 모형과 결합하여, 이상치를 자동적으로 검지하고 예측할 수 있는 모형 파이프라인을 구축하였다.

### 1) 연구 데이터 및 이상치 검지

본 연구에서는 2017년 6월부터 2022년 6월까지의 일별 교통량 데이터를 활용하여 일변량 시계열 예측 모형을 구축하였다. 시계열 데이터를 살펴본 결과 COVID-19가 발발한 시기인 2020년 초 교통량이 급격히 감소한 것을 확인할 수 있다.



<Fig. 1> Daily Traffic count

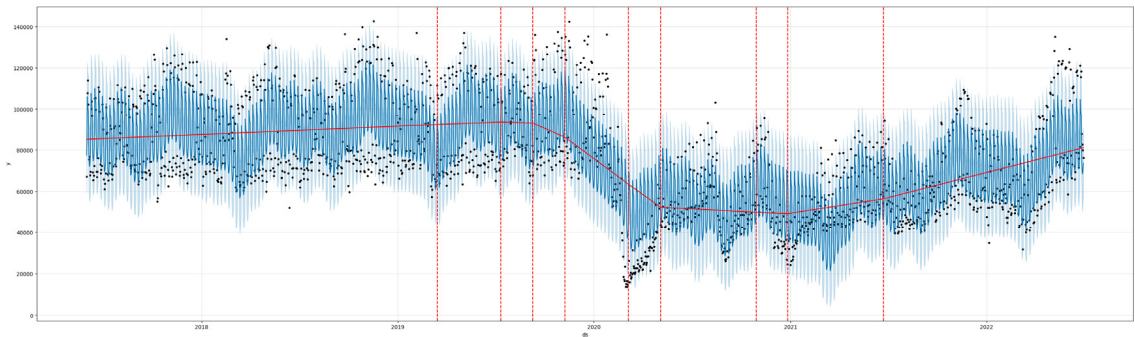


<Fig. 2> Detection of anomaly data

시계열 예측을 위한 일반화된 아키텍처 개발을 위하여 이상 및 결측치를 어떠한 방식으로 처리 및 보정하는가에 따라 교통량과 같은 시계열 자료의 장래 예측 결과에 따라 합리적인 보정방식을 결정해야만 한다. 반영하는 과정에 있어서 이상치를 주관적인 기준에 의해 판단하는 것이 아닌 과거의 실제값을 바탕으로 패턴을 학습하여 이상치를 검지하는 알고리즘을 적용하여 이상치를 선별하고 반영하였다. 이상치 검지는 Stepwise 방식으로 과거 데이터를 순차적으로 학습하여 과거의 패턴에서 벗어나는 수치를 이상치로 간주하는 시스템을 의미한다. 여기서 실제값과 비교를 통해 91%로 검지 성능을 가진다는 의미는 1) 기존 시스템에서 이상치로 간주하여 보정을 수행한 날짜의 교통량 (실제값)과 2) stepwise를 통해 이상치로 예측된 날짜의 교통량을 비교했을 때의 검지 성능을 나타낸다. 적용된 알고리즘은 seasonal decomposition, Quantile AD, aggregation 기법을 결합하여 활용하였으며, 실제값과 비교를 통해 91% 이상의 검지 성능을 나타내는 것을 확인하였다. 알고리즘을 통한 검지 결과 COVID-19로 인해 교통량 급격히 감소한 부분이 이상치로 나타났으며, 이를 참고하여 모형의 예측 정확도를 향상시키기 위한 조정에 반영하였다.

2) 조정요소

수요에 영향을 주는 개입으로 연도별 명절 및 공휴일과 같은 휴일효과와 계절성 등을 반영하였다. 명절 및 공휴일은 평일과 달리 철도 수요가 급증하기 때문에, 이를 반영하기 위하여 휴일 효과를 적용하여 모형의 적합도를 상승시켰다. 또한 본 연구의 목적인 비정상적인 시계열을 반영한 수요예측을 위해 COVID-19라는 변수를 반영하였다. 이는 앞서 살펴보았듯이 COVID-19가 발생하고 재유행하는 시기마다 수송수요가 급격한 감소를 보였고, 이러한 비정상적인 시계열을 통해서 기존의 수요예측 방법으로는 정확도 있는 예측이 불가능했기 때문이다. 이에 이상치 검지 알고리즘을 통한 검지와 Prophet 자체적인 알고리즘을 통한 변곡점들과 COVID-19 대유행 시점을 비교하여 5번의 COVID-19 대유행 구간을 이상치로 판단하여 휴일로 설정하고 수요예측에 미치는 영향을 최소화시키며 모형의 예측 정확도를 향상시켰다.



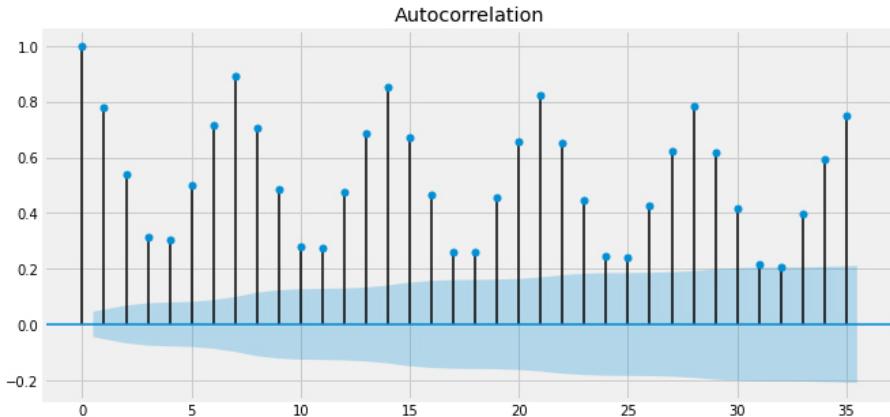
<Fig. 3> Demand Wave Shifting Points

<Table 1> Adjustment factors for increasing model fit

Category		Type
Seasonality effect		weekly seasonality
		monthly seasonality
		yearly seasonality
holiday effect	Country holidays	New Year's Day, Samiljeol, Lunar New Year's Day, Buddha's Birthday, Children's Day, Labor Day, Memorial Day, Liberation Day, Chuseok Holiday, Hangeul Day, Christmas, other alternative holidays
	COVID-19	2020.02 ~ 2020.03 pandemic
		2020.08 ~ 2020.09 pandemic
		2020.12 ~ 2021.01 pandemic
		2021.08 ~ 2021.09 pandemic
2022.01 ~ 2022.03 pandemic		

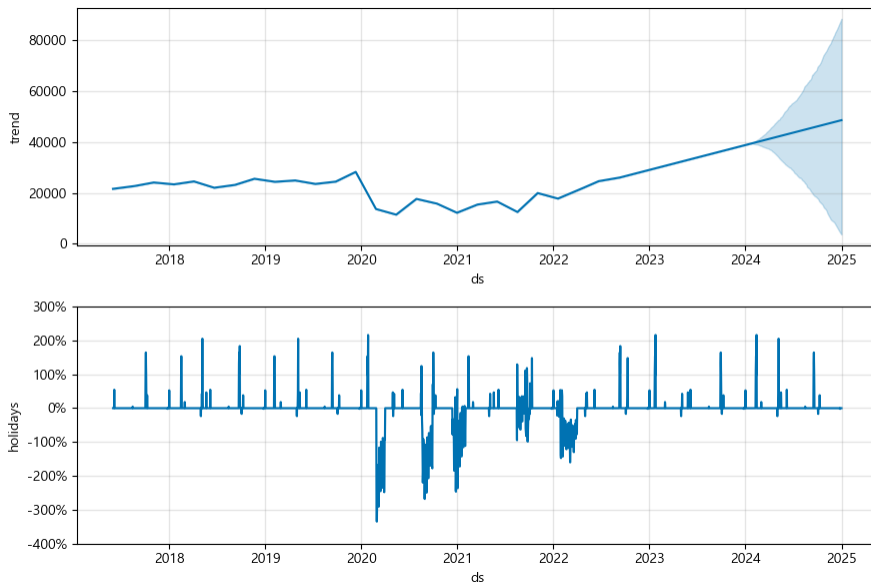
조정을 통해 구축한 모형의 구성요소는 그림 5와 같다. 먼저 모형의 장기적인 트렌드는 우상향 하는 모습을 보였고 예측값 또한 트렌드를 따라가는 것을 확인할 수 있다. 주간 계절성은 주중에 수요량이 계속 감소하다가 주말에 수요량이 급증하는 주기성을 보이고 있다. 이는 본 연구에 사용된 시계열 데이터가 가지고 있는 자기상관성(autocorrelation)이 반영된 결과라고 생각된다. 자기상관성은 어떠한 확률변수가 주어졌을 때, 서로 다른 두 시점에서의 관측치 사이에 나타나는 상관성을 뜻하며, 본 연구에 사용된 데이터의 경우 7일을

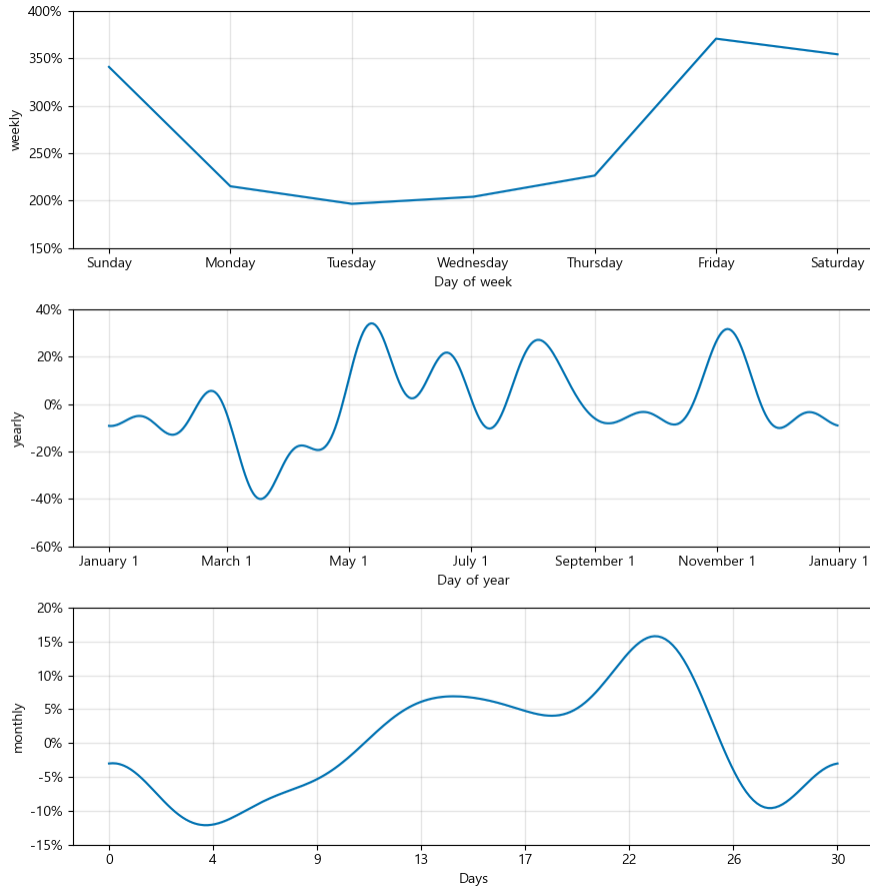
주기로 자기상관성을 가지고 있는 것을 알 수 있다. 모형의 과적합 및 과소적합을 검증하기 위해 훈련용(train), 검증용(validation), 테스트(test) 데이터로 나뉘었으며 비율은 70%, 20%, 10%로 설정하였다. 검증 및 테스트를 위한 데이터는 학습된 모형을 검증하고 예측력을 시험하는 용도이므로 무작위로 다양한 시점 구간을 선택하여 검증할 수 있도록 교차검증방법도 적용하였다. 또한, 머신기반의 모형은 입력 데이터의 스케일이 학습에 큰 영향을 미치며, 이는 편향된 추정을 발생시킬 수 있다. 따라서 본 연구에서 활용된 모든 데이터셋을 스케일링 하였고, 머신기반 모형에서 입력값의 상대적 크기에 따른 편향된 추정을 방지하고자 하였다.



<Fig. 4> Autocorrelation in the time series

또한 연간 계절성 및 주간 계절성을 살펴보았을 때 본 연구에 사용된 데이터와 유사한 주기성을 가지고 있는 것으로 나타났으며, 이로 모형의 적합이 잘 이루어 졌음을 알 수 있다.





<Fig. 5> Example of adjusted model's components

#### IV. 연구 결과 분석

COVID-19를 반영한 Prophet 모형의 예측력 결과는 <Table 2>와 같다. 2017년 6월부터 2022년 3월까지의 데이터를 모형의 학습에 사용하였고 2022년 4월부터 2022년 6월까지를 검증에 사용하였으며, 평가지표는 Mean Absolute Percentage Error(MAPE)와 결정계수 R<sup>2</sup>을 통해 모형의 예측력을 검증하였으며 이는 실제값과 예측값을 비교하는 지표라고 볼 수 있다.

$$MAPE = \frac{100}{n} \sum_{i=1}^n abs\left(\frac{y_i - \hat{f}(x_i)}{y_i}\right) \dots\dots\dots (3)$$

- 여기서  $\hat{f}(x_i)$  : predicted value
- $y_i$  : actual value
- $x_i$  : independent variable (time)
- $n$  : number of observation



구축된 모형의 검증 결과는 MAPE는 0.08(8%), R<sup>2</sup>계수는 0.85로 유의미한 값을 가지고 있다.

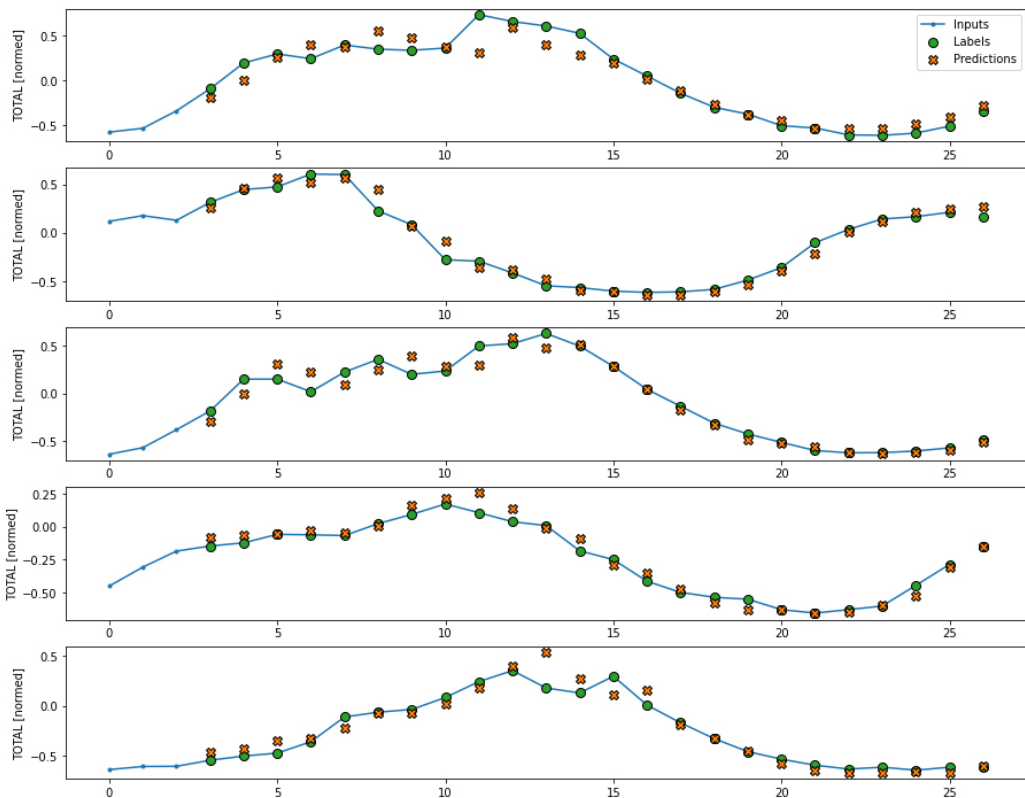
<Table 2> Verification of transport demand forecast results

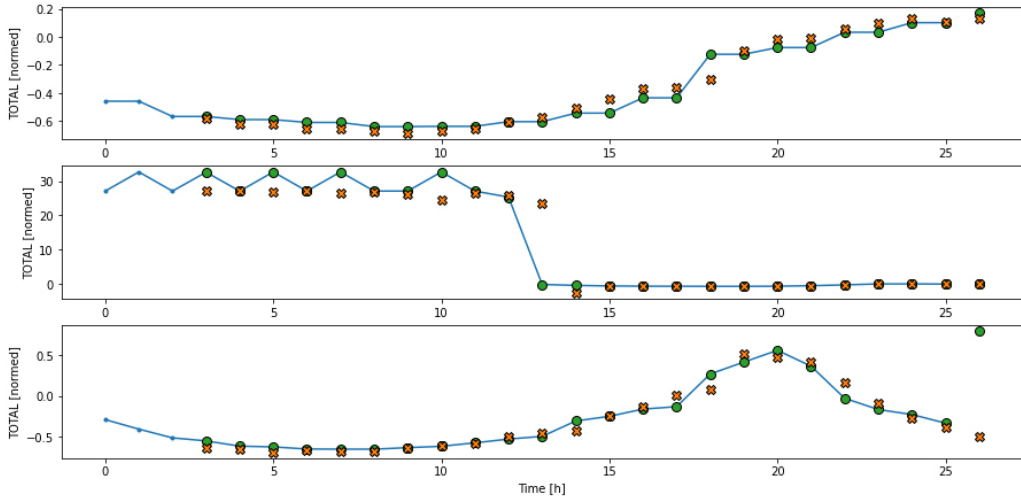
Category	Predictive performance of model			
	RMSE	MAE	MAPE	R <sup>2</sup>
22.04~22.06 Traffic count	9250.75	7081.97	0.08	0.85

22년 4월부터 22년 6월까지의 실제값과 예측값 주중과 주말로 나누어 비교해본 결과, 평균절대오차율은 주중 6.15% 주말 4.09%로 나타났으며 의미있는 결과값을 가진다.

<Table 3> Comparison of predictive values and actual values

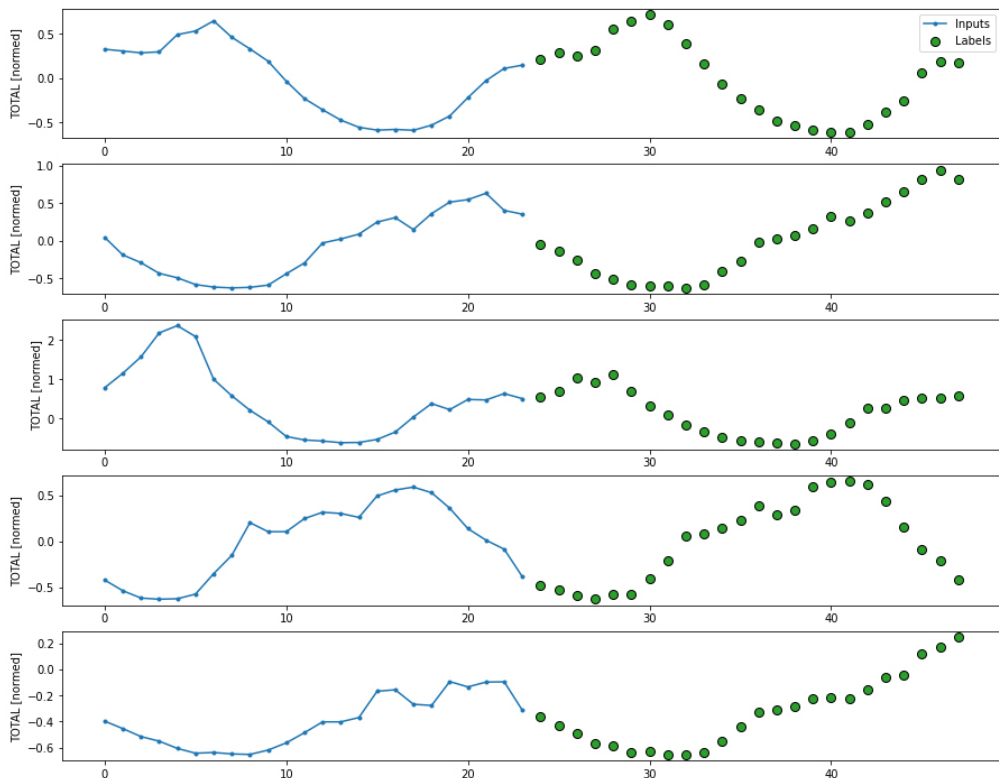
Category		WEEKDAY				WEEKEND			
		Predicted Value	Actual Value	Error	Error Rate	Predicted Value	Actual Value	Error	Error Rate
2022	04	63,846	57,202	6,644	11.61%	95,432	95,125	307	0.32%
	05	76,339	74,029	2,310	3.12%	110,115	121,602	-11,487	-9.45%
	06	79,952	77,092	2,860	3.71%	113,148	116,066	-2,918	-2.51%
MAE / MAPE				3938	6.15%			4904	4.09%

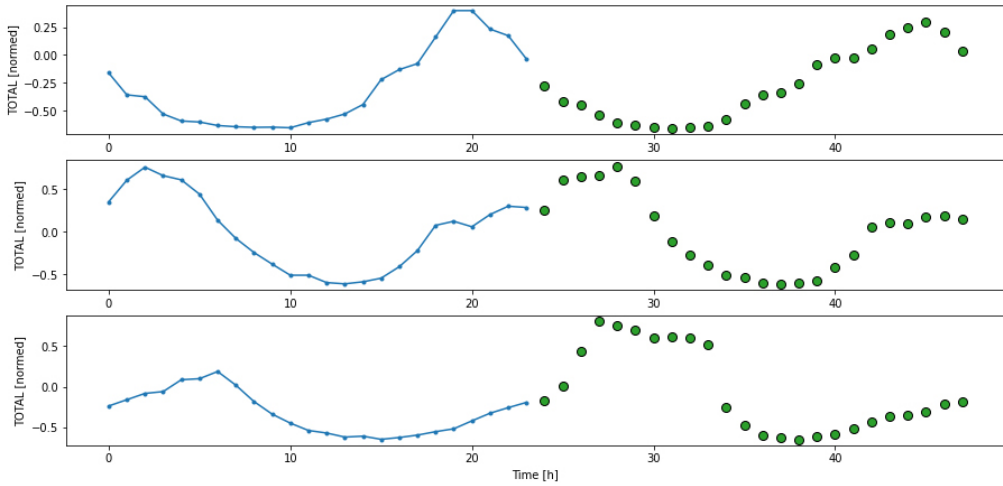




<Fig. 6> Validate predictive model

<Fig. 6>의 결과는 입력값(input)은 시간(time-step)에 따른 실제 교통량(ground truth)을 의미하고, 데이터 라벨(Label)은 예측하고 하는 시점을 나타낸다. 라벨을 기준으로 예측한 결과는 예측값(Predictions)으로 나타나고 있다.





<Fig. 7> Future Prediction based on Labeled Input

## V. 결 론

본 연구는 COVID-19라는 비정상적인 시계열 이벤트와 같은 불확실성을 인한 수요 변화를 자동으로 감지하고 정확한 예측값을 도출해 내기 위한 모형 파이프라인을 구축하였다. 스마트 시티 내의 센서 기반으로 수집되는 시계열 데이터가 가지는 불확실성의 문제를 해결하고, 양질 데이터를 수집 및 관리하기 용이한 시스템을 구축하는데 중요한 역할을 할 것이라 기대된다. 특히, 실시간으로 수집되는 데이터의 이상치를 자동으로 감지하여 예측하는 알고리즘인, Seasonal AD와 Prophet을 결합하여 모형 파이프 라인을 구축하였다.

이를 위해 2017년부터 2022년까지의 일평균 교통량 데이터를 바탕으로 모형을 구축하고 검증하였다. 기존에 주로 사용하던 계절 개입형 시계열 모형이 사용하기 어렵고 복잡한 과정을 거치며 비정상적인 시계열을 제대로 반영하지 못하며, 이는 COVID-19로 인한 수송수요의 급격한 변화를 제대로 반영하지 못한다는 것일 뿐만 아니라 교통량이 급격히 변하는 주말과 주중의 모형까지 따로 구축해야 했기에 보다 더 정확도 있는 예측값을 도출하면서도 사용하기 편리한 Prophet을 사용하여 새로운 모형을 구축하였다. Prophet모형은 직관적인 파라미터를 가지고 있어 누구나 쉽게 사용이 가능하고, 비정상적인 시계열 이벤트를 모형에 반영하여 모형의 예측 정확도를 향상시킬 수 있다는 장점을 가지고 있다. 또한 주말과 주중 수요를 별도의 모형을 구축하여 예측하던 기존과 달리 그 차이를 구분하여 예측값을 도출하기 때문에 하나의 모형으로 전체 예측값을 도출해 낼 수 있었다.

구축한 모형의 2022년 4월부터 2022년 6월까지 예측값과 실제값을 비교해 MAPE와 R<sup>2</sup>계수를 통한 검증결과 모든 노선에서 유의미한 값을 가지는 것으로 나타났다. 본 연구 결과는 교통량 예측한다는 것에만 국한되는 것이 아닌, 사회의 다른 부문에서의 비정상적인 시계열 상황 속에서 간편하면서도 안정적인 예측값을 도출해 내고 이를 참고하여 향후 의사결정을 내리는데 있어 큰 도움이 될 수 있을 것이라 기대된다.

## ACKNOWLEDGEMENTS

본 연구는 국토교통부 “AI·데이터 기반 스마트시티 통합플랫폼 모델 개발 및 실증연구”의 연구비지원(과제번호: 22AIP-C163095-02)에 의해 수행되었습니다.

## REFERENCES

- Bajari, P., Nekipelov, D., P.Ryan, S. and Yang, M.(2015), “Machine Learning Methods for Demand Estimation”, *American Economic Review*, vol. 105, no. 5, pp.481-485.
- Cha, D. H., Lee, K. H., Son, J. H., Sun, D. W., Lim, H. T., Lee, J. K., Choi, T. J. and Goo, Y. M.(2021), “Prediction of Breeding Animals by Swine Breeding Scale using ARIMA Model in Time Series Analysis”, *Journal of Animal Breeding and Genomics*, vol. 5, no. 2, pp.31-50.
- Han, D. C., Lee, D. W. and Jung, D. Y.(2021), “Estimation of Shared Bicycle Demand Using the SARIMAX Model: Focusing on the COVID-19 Impact of Seoul”, *Journal of the Korea Institute of Intelligent Transportation Systems*, vol. 20, no. 6, pp.1-13.
- Han, D., Lee, D. W. and Jung, D.(2021), “A Study on the Traffic Volume Correction and Prediction Using SARIMA Algorithm”, *The Journal of the Korea Institute of Intelligent Transport Systems*, vol. 20, no. 6, pp.1-13. doi: 10.12815/kits.2021.20.6.1
- Hong, J. Y., Han, E. R., Choi, C. H., Lee, M. S. and Park, D. J.(2021), “Estimation of Shared Bicycle Demand Using the SARIMAX Model: Focusing on the COVID-19 Impact of Seoul”, *Journal of The Korea Institute of Intelligent Transportation Systems*, vol. 20, no. 1, pp.10-21.
- Kim, D. K. and Kim, H. Y.(2019), “Forecasting Urban Railway Demand of Daegy Based on Time Series Analysis”, *Journal of Decision Making Research*, vol. 27, no. 1, pp.1-11.
- Kim, D. K.(2020), “Urban Railway Demand Forecast using Time Series Data”, *Journal of the Korean Data Analysis Society*, vol. 22, no. 2, pp.753-765.
- Kim, J. H. and Kim, J. Y.(2022), “Prediction of Covid-19 confirmed number of cases using SARIMA model”, *Journal of Korea Institute of Information and Communication Engineering*, vol. 26, no. 1, pp.58-63.
- Kim, J. S., Gang, J. H., Kim, S. H. and Youn, J. S.(2020), “A Univariate Time Series Forecasting with Prophet”, *Journal of Korea Institute of Information and Communication Engineering*, vol. 24, no. 1, pp.329-331.
- Korea Transportation Safety Authority and Ministry of Land and Transport(2021), *Investigation of Public Transportation System*.
- Lim, S. H.(2020), “An Analysis of Change in Traffic Demand with Coronavirus Disease 2019”, *Journal of the Korea Institute of Intelligent Transportation Systems*, vol. 19, no. 5, pp.106-118.
- Noh, Y. S. and Do, M. S.(2015), “Short-term Railway Passenger Demand Forecasting by SARIMA Model”, *Journal of the Korea Institute of Intelligent Transportation Systems*, vol. 14, no. 4, pp.18-26.
- Oh, S. W., Lim, N. H., Lee, S. H. and Kim, M. S.(2020), “Long-term Price Prediction and Trend Analysis of Garlic Using Prophet Model”, *Journal of the Korean Data Analysis Society*, vol. 22,

no. 6, pp.2325-2336.

Seo, C. B. and Sung, M. J.(2022), “The Impact of COVID-19 on Transportation Demand-OO Private Expressway”, *Proceedings of the KOR-KST Conference*, pp.269-274.

Shin, D. R., Chae, G. Y. and Park, M. J.(2022), “Prediction of COVID-19 Confirmed Cases by Using Big Data and Time Series Analysis”, *Journal of Applied Reliability*, vol. 22, no. 4, pp.352-362.

Winters, P. R.(1960), “Forecasting sales by exponentially weighted moving averages”, *Management Science*, vol. 6, no. 3, pp.324-342.