

# 코로나19 확산에 따른 공공 자전거 이용자 수 예측을 위한 인공지능 모형의 제안

김규석\*, 서예진\*\*

## A Proposal for an Artificial Intelligence Model for Predicting the Number of Public Rental Bike in the COVID-19 Era

Kyuseok Kim\*, Yejin Seo\*\*

### 요 약

코로나19 바이러스의 빠르고 강한 전파력으로 인해 대면 활동의 규모와 빈도를 줄이고 비대면으로 활동하는 기회가 늘어나고 있다. 학교나 직장에서는 비대면과 대면 활동을 혼합한 교육, 근무를 수행하기도 한다. 또한, 사람들이 밀접해 있는 대중교통 대신 자가용을 활용하는 수가 늘고 있다. 이에 따라, 개인형 교통수단의 이용자 수 변화로 “따릉이” 공급 과·부족 현상을 겪고 있다. 본 연구에서는 코로나19 시대에 공공 자전거 이용자 수를 예측하기 위한 인공지능 모형을 제안한다. 연구 결과, 다중 선형 회귀분석의 설명력인 Adj. R2는 0.676으로 코로나19 관련 독립변수를 포함하지 않고 수행한 결과인 0.600 보다 약 12.67% 상승하였음을 알 수 있었다. 또한, 이 변수들을 기반으로 한 인공지능 모형의 공공 자전거 이용자 수의 예측 오차 MAPE(Mean Absolute Percentage Error)는 약 0.43998815 이었다. 추후 본 연구 모형을 통해 코로나19로 인하여 변화된 각종 공공 재화의 이용자 수를 예측하는데 도움이 될 것으로 기대한다.

### Abstract

Due to the fast and strong spread of the COVID-19, the scale and frequency of face-to-face activities are reduced, and the number of opportunities of non-face-to-face activities has been increasing. In schools and workplaces, education and works are also conducted with both non-face-to-face and face-to-face activities. In addition, the number of people using private vehicles instead of public transportation is increasing. Accordingly, we are experiencing the oversupply and shortage of the rental bikes due to the change in the number of the personal mobility users. In this paper, an artificial intelligence model to predict the number of public rental bike users in the COVID-19 era is proposed. As a result, the explanatory power of multiple linear regression analysis, Adj. R2 is 0.676, which is about 12.67% higher than 0.600, which is the result of not including the COVID-19 related independent variables. Moreover, the prediction error Mean Absolute Percentage Error(MAPE) of the number of public rental bike users from the AI model based on these variables was about 0.43998815. In the future, it is expected that the proposed model of this paper will be helpful in predicting the number of users of various public goods that have changed due to the COVID-19.

### Keywords

COVID-19, public rental bike, artificial intelligence, linear regression analysis

---

\* 한국폴리텍대학 분당융합기술교육원 데이터융합SW과/ · Received: Jul. 20, 2021, Revised: Aug. 17, 2021, Accepted: Aug. 20, 2021  
서울대학교 환경대학원 환경계획학과 (교신저자) · Corresponding Author: Kyuseok Kim  
- ORCID: <https://orcid.org/0000-0001-6613-5125>  
Dept. of Data Convergence Software, Korea Polytechnics, Gyeonggi, Korea  
\*\* 한국폴리텍대학 구미캠퍼스 전기과  
Tel.: +82-31-696-8832, Email: kyuseokkim@kopo.ac.kr  
- ORCID: <https://orcid.org/0000-0003-4823-5047>

## I. 서 론

코로나19 바이러스의 확산에 따라 우리 생활에 큰 변화가 생겼다. 코로나19 확진자가 많이 발생하여 사회적 거리두기 단계가 높아지면 학교에서 공부하는 학생들은 집에서 비대면으로 온라인 학습을 하고 회사에서 근무하는 직장인들은 재택근무를 하는 기회가 많아졌다[1].

코로나19 바이러스는 강하고 빠른 확산 속도와 전파력 때문에 우리의 일상생활에 있어 대면 활동을 감소시키고 있다. 이에 따라 사람들의 교통수단도 많은 사람들이 동시에 이용하는 버스나 지하철과 같은 대중교통보다 자가용이나 공공 자전거, 킥보드 등의 개인형 교통수단의 이용자가 증가하고 있는 상황이다[2][3]. 실제로 서울시 발표에 따르면, 코로나19의 첫 대규모 확산이 있었던 2020년 2~4월인 3개월간의 “따릉이” 이용자 수는 전년 동기 대비 59.1%가 증가했으며 특히 3월에는 74.4%가 증가하였다[4].

이에 따라, 대여소별로 대여 가능 공공 자전거의 과·부족 현상이 발생하고 있음을 모바일 또는 웹 기반의 서울시 공공 자전거 서비스를 통해 실시간으로 확인할 수 있다[5]. 그래서 이용자들은 공공 자전거 추가 배치, 대여소 개설 등을 요청하고 있으며 이러한 수요·공급 불균형을 조절하기 위해 “따릉이” 이용자 수를 예측할 필요성이 커지고 있다.

본 논문에서는 2020년 1월 이후, 코로나19 바이러스의 확산으로 변화된 개인형 교통수단 중 서울시 공공 자전거 서비스인 “따릉이”의 이용자 수를 예측할 수 있는 인공지능 모형을 제안한다.

인공지능 모형을 제안하기에 앞서 다중 선형 회귀분석을 통해 종속변수인 일별 서울시 공공 자전거 이용자 수와 원인이 되는 독립변수 간의 설명력과 유의수준을 확보하여 인과 관계를 검증한다. 그리고 해당 모형을 파이썬(Python) 기반의 텐서플로우(TensorFlow) 모듈을 통해 예측 값을 분석한다. 본 인공지능 모형에 대한 정확도 검증은 실제 값과 예측 값의 차이의 평균을 나타내는 MAPE(Mean Absolute Percentage Error)를 활용한다.

본 논문의 구성은 다음과 같다. 2장에서는 공공 자전거 서비스인 “따릉이”의 이용자 수와 관련된

기존의 회귀분석, 인공지능 기반의 문헌을 고찰한다. 3장에서는 본 연구에서 활용하는 데이터와 변수를 설정하며 4장에서는 해당 데이터를 활용한 연구 방법에 대하여 설명한다. 5장에서는 연구 결과에 대하여 분석하며 마지막 6장에서는 결론과 함께 향후 연구 방향에 대하여 제안한다.

## II. 문헌 고찰

### 2.1 서울시 공공자전거 수요 예측 연구

민소아 외(2021)는 다양한 기법을 활용한 시계열 데이터 모델링을 통해 서울시 공공 자전거의 수요 예측을 위한 정확도 비교 연구를 하였다[6]. 해당 논문에서는 시계열 데이터 군집 분석(Time series clustering), 서포트 벡터 회귀(Support vector regression), LSTM 네트워크(Long Short-Term Memory networks)의 분석 방법론을 통해 서울시 공공 자전거 수요 예측을 하였다. LSTM은 학습 데이터의 지속성을 갖는 RNN(Recurrent Neural Network) 기법의 단점인 입력과 출력 사이의 거리가 멀수록 연관 관계가 적어지는 장기 의존성 문제를 개선한 기법이다. 연구 결과, LSTM 모형은 다른 모형들에 비해 상대적으로 정확도가 높아 서울시 공공 자전거 이용자 수 예측을 위해 적합한 것으로 판단하였다.

김예슬 외(2020)는 서울시 공공 자전거 신규 대여소를 위한 수요량 예측 분석을 하였다[7]. 서울시 공공 자전거의 새로운 입지 선정에 한 후 랜덤 포레스트 방법론을 활용하여 신규 대여소의 이용량을 예측하였다.

이근혁 외(2019)는 서울시 공공 자전거 서비스에 대한 이용자 수 예측을 위하여 기상 인자와 시간을 활용하였다[8]. 연구 방법론으로는 자전거 이용률에 영향을 미치는 기상요소와 시간 및 행정구별 간의 상관관계 및 인과관계를 분석한 뒤, 인공지능망을 적용해 기상예보를 토대로 시간 및 지역에 따른 대여율을 예측하였다. 연구 데이터로는 273,632개의 학습 데이터로서 배치 사이즈(Batch size) 1000, 에포크(Epoch)는 540으로 하여 약 35%의 예측 정확도를 보였다.

## 2.2 머신러닝을 활용한 현상 예측 연구

고은경(2020)은 한국 초등학교 3학년 아동의 행복감 예측을 연구하였다[9]. 연구 방법론으로는 로지스틱 회귀 분석 모형, 라쏘 로지스틱 회귀 분석 모형, 랜덤 포레스트 모형, 그레디언트 - 부스팅 의사결정나무 모형 등 4가지의 모형을 기반으로 한 머신러닝을 수행하였다. 독립변수로는 자아 존중감, 학교 선호도, 또래애착 신뢰, 교사 선호도, 학교생활 느낌, 시간 만족도, 수학 선호도를 활용하였으며 그 예측 정확도는 모두 90% 이상을 보였음을 알 수 있었다.

안성준 외(2020)는 저유소 휘발유 출하량 예측을 위한 머신러닝 모형의 성능을 비교하였다[10]. 독립 변수로는 기온 강수 여부 등의 18가지 정보를 활용하였다. 이 데이터들을 활용하여 GLM(Generalized Linear Model), 랜덤 포레스트(Random forest), GBM (Gradient Boosting Machine), 딥러닝, 앙상블 학습 모델을 포함한 7가지 모델에 대한 성능을 비교 분석하였다. 분석 결과, GBM과 랜덤 포레스트 모델을 활용한 Ensemble 3 모델을 통한 예측 값들의 MAPE는 29.39%로 전문가들의 예측 수준인 약 45.0%보다 15.61% 가량 향상된 수준을 보임을 알 수 있었다.

## 2.3 본 연구의 차별점

본 연구는 기존의 서울시 공공 자전거 이용자 수 예측을 위한 연구들과 다음과 같은 차별점이 존재한다.

첫째, 코로나19 팬데믹 상황에 따라 증가하게 된 PM(Personal Mobility) 수요에 맞춰 강우량 및 계절, 요일을 활용하였을 뿐만 아니라 일별 코로나19 확진자 수를 활용하였다. 따라서 국내 코로나19 첫 확진자가 발생하기 전인 2019년도와 발생 후인 2020년도의 공공 자전거 이용자 수를 비교해 볼 수 있는 연구이다.

둘째, 특정 공공 자전거 대여소나 지역구에 국한되지 않고 서울시 전역을 기준으로 하여 일별 공공 자전거 이용자 수 예측을 위한 모형을 제안하는 연구이다.

셋째, 본 연구는 다중 선형 회귀분석과 머신러닝을 모두 수행한 연구이다. 기존 고은경(2020)의 한국 초등학교 3학년 아동의 행복감 예측 연구[9]와 같이 머신러닝을 수행함에 있어 변수들의 계수, 유의수준 등의 회귀분석을 통해 각 변수들의 영향도를 확인하였다. 또한, 회귀분석 외삽(Extrapolation)의 한계를 텐서플로우를 통해 보완한 연구로서 그 의미를 갖는다[11].

따라서 본 연구는 코로나19의 확산과 함께 서울시 전 지역구를 기준으로 공공 자전거 서비스인 “따릉이”의 이용자 수를 합리적으로 예측하는데 의의를 둔다.

## III. 연구 데이터

본 연구에서 활용하는 데이터는 표 1과 같다.

종속변수로는 서울 열린 데이터 광장[12]에서 제공하는 2019년 1월 1일부터 2020년 12월 31일까지의 “따릉이”의 이용 17,324,057건을 일별로 정리한 것이다(NUM)[4]. 이 기간 중 2019년 2월 22일과 2019년 9월 7일은 데이터에 포함되어 있지 않아 독립변수들에 대해서도 해당 일자는 모두 제외하였다.

독립변수로는 질병관리청에서 제공하는 일별 코로나19 확진자 수(COV)를 활용하여 2020년 1월 20일 국내 첫 코로나19 확진자 발생일 이전은 0, 이후는 1로 하는 더미변수(TREAT)와 2020년 1월 20일로부터의 경과일(EDAY)을 활용하였다[13].

표 1. 변수  
Table 1. Variables

| Var.   | Data   |
|--|--|
| DV   | Number of Seoul bike usage per day(NUM)                        |
| IV   | Whether after the COVID-19 outbreak(TREAT)                     |
|  | Number of COVID-19 confirmed cases per day(COV)                |
|  | Number of days since the first case of COVID-19 in korea(EDAY) |
|  | Rainfall(RNFL)   |
|  | Dummy Var. for winter(WNTR)                                    |
|  | Dummy Var. for spring(SPRG)                                    |
|  | Dummy Var. for summer(SUMM)                                    |
| Dummy Var. for holiday(WKND)   |  |
| Jan. 1, 2019 ~ Dec. 31, 2020<br>(Except for Feb. 22, 2019 and Sep. 07, 2019) |  |

계절 더미변수는 한국민족문화대백과사전에 정의된 것과 같이 3~5월은 봄(SPRG), 6~8월은 여름(SUMM), 9~11월은 가을, 12월~2월은 겨울(WNTR)로 설정하였다[14]. 강우량 변수(RNFL)는 기상청에서 제공하는 정보를 활용하였고 휴일 여부 더미변수는 국경일, 명절, 토요일, 일요일 등을 활용하였다[15].

본 연구에서 활용하는 데이터들의 기술통계량은 표 2와 같다. 전체 기간에 대한 일수는 729일이며, 일평균 “따릉이” 이용 건(NUM)은 58,606.84건이었다. 일평균 코로나19 확진자 수(COV)는 83.95명이었으며 강우량(RNFL)의 평균값은 3.48이었다.

표 2. 기술통계량  
Table 2. Descriptive statistics(n = 729)

| Variable | Mean      | SD        | Min.  | Max.    |
|----------|-----------|-----------|-------|---------|
| NUM      | 58,606.84 | 29,847.53 | 1,920 | 118,637 |
| COV      | 83.95     | 198.67    | 0     | 1,241   |
| EDAY     | 82.82     | 111.11    | 0     | 347     |
| RNFL     | 3.48      | 11.89     | 0     | 103.1   |
| WNTR     | 0.25      | 0.43      | 0     | 1       |
| SPRG     | 0.25      | 0.43      | 0     | 1       |
| SUMM     | 0.25      | 0.43      | 0     | 1       |
| WKND     | 0.32      | 0.47      | 0     | 1       |

#### IV. 연구 방법

본 연구에서는 1차적으로 다중 선형 회귀분석을 통해 표 1에서 정의한 변수들 간의 인과 관계를 파악한다. 연구에서 활용한 다중 선형 회귀분석 모형은 식 (1)과 같다. 또한, 코로나19에 따른 “따릉이”의 이용자 수의 변화를 파악하기 위하여 관련 독립 변수인 TREAT, COV, EDAY를 포함하였을 때와 포함하지 않았을 때의 다중 선형 회귀분석을 수행한다. 이 변수들을 포함하였을 때의 모형은 식 (1)과 같다.

$$\begin{aligned}
 NUM_i = & \quad (1) \\
 & \beta_0 + \beta_1 TREAT_i + \beta_2 COV_i + \beta_3 WKND_i \\
 & + \beta_4 WNTR_i + \beta_5 SPRG_i + \beta_6 SUMM_i \\
 & + \beta_7 RNFL_i + \beta_8 EDAY_i + \epsilon
 \end{aligned}$$

다중 선형 회귀분석으로 통해 변수들 간의 인과 관계가 파악되면 파이썬 기반의 텐서플로우 모듈에

적용하여 MAPE 값을 구함으로써 해당 머신러닝 모형의 예측 정확도를 측정한다[16]-[18]. MAPE는 식 (2)와 같이 실제 값과 예측 값의 차이를 실제 값으로 나눈 것의 평균으로 예측 값의 오차율을 구할 수 있다[17][18].

$$MAPE_i = \frac{100}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \quad (2)$$

$A_i$  : 실제 값  
 $F_i$  : 예측 값

파이썬 기반의 텐서플로우 모듈에 적용한 환경설정 값은 표 3과 같다. 총 에포크와 배치 사이즈에 따라 총 15회를 수행한다.

배치 사이즈가 1일 때에는 그의 1,000,000배인 1,000,000까지 에포크를 10배씩 증가시켜서 수행하고, 배치 사이즈가 10일 때에는 그의 1,000,000배인 10,000,000까지 에포크를 10배씩 증가시켜서 수행한다.

공통 설정으로 모형의 은닉층(Hidden layer)은 설정하지 않았으며 Output node는 1개로 단일화 하였다.

표 3. 텐서플로우 환경 설정값  
Table 3. Setting values for tensorflow

| No. | Epoch      | Batch size |
|-----|------------|------------|
| 1   | 1          | 1          |
| 2   | 10         | 1          |
| 3   | 100        | 1          |
| 4   | 1,000      | 1          |
| 5   | 10,000     | 1          |
| 6   | 100,000    | 1          |
| 7   | 1,000,000  | 1          |
| 8   | 1          | 10         |
| 9   | 10         | 10         |
| 10  | 100        | 10         |
| 11  | 1,000      | 10         |
| 12  | 10,000     | 10         |
| 13  | 100,000    | 10         |
| 14  | 1,000,000  | 10         |
| 15  | 10,000,000 | 10         |

· No hidden layer  
· Number of output node : 1

## V. 연구 결과

### 5.1 다중 선형 회귀분석 결과

코로나19 관련 독립변수인 TREAT, COV, EDAY 변수를 제외한 경우와 포함한 경우의 다중 선형 회귀분석의 결과는 각각 표 4, 5와 같다.

분석 결과, TREAT, COV, EDAY 변수를 포함한 경우의 Adj. R2는 0.676으로 67.6% 정도의 설명력을 갖췄으며 모든 독립변수들의 P-value는 0.001 미만으로 99.9% 이상의 신뢰수준을 갖췄음을 알 수 있다. 이는 TREAT, COV, EDAY 변수를 제외한 다중 선형 회귀분석의 설명력인 Adj. R2의 0.600 보다 약 12.67% 상승한 값이다. 따라서 코로나19 관련 독립변수인 TREAT, COV, EDAY 변수를 포함하면 모형의 설명력이 더 높아지는 것을 확인할 수 있다.

표 4. 코로나19 관련 독립변수 없이 수행한 회귀 분석 결과

Table 4. Result of regression analysis without the COVID-19 related independent variables(n = 729)

| Variable                     | Coef.          | t-test  | VIF      |
|------------------------------|----------------|---------|----------|
| Constant                     | 80,820***      | 55.885  |          |
| RNFL                         | -968.8574***   | -15.935 | 1.067945 |
| WNTR                         | -52,630***     | -26.436 | 1.509136 |
| SPRG                         | -19,070***     | -9.639  | 1.512186 |
| SUMM                         | 3,221.8528     | 1.610   | 1.546836 |
| WKND                         | -5,782.8183*** | -3.852  | 1.001409 |
| R2 : 0.603 / Adj. R2 : 0.600 |                |         |          |

표 5. 코로나19 관련 독립변수를 포함하여 수행한 회귀 분석 결과

Table 5. Result of regression analysis with the COVID-19 related independent variables(n = 729)

| Variable                     | Coef.           | t-test  | VIF      |
|------------------------------|-----------------|---------|----------|
| Constant                     | 69,980***       | 42.841  |          |
| TREAT                        | 7,901.3391****  | 3.784   | 2.749291 |
| COV                          | -27.9275****    | -5.932  | 2.204074 |
| EDAY                         | 68.2255****     | 5.486   | 4.810115 |
| RNFL                         | -1,045.7996**** | -18.982 | 1.080379 |
| WNTR                         | -45,360****     | -20.918 | 2.205904 |
| SPRG                         | -13,770****     | -6.655  | 2.038709 |
| SUMM                         | 5,894.1852****  | 3.150   | 1.666426 |
| WKND                         | -5,689.3239**** | -4.206  | 1.001545 |
| R2 : 0.679 / Adj. R2 : 0.676 |                 |         |          |

또한, 코로나19 관련 모든 독립변수들을 포함하여 다중 선형 회귀분석을 수행한 표 5와 같이 VIF (Variation Inflation Factor) 값은 모두 10미만의 약 1.0에서 4.8 사이로 독립 변수 개별적으로 상관도를 나타내는 다중 공선성을 가진 변수는 없는 것으로 판단된다.

따라서 표 5의 결과를 해석하면 다음과 같다. 종속변수인 “따릉이”의 일별 이용 건을 증가시키는 요인은 2020년 1월 20일 이후인 코로나19 확진자 발생일 이후 인지 여부, 계절이 여름인지 여부, 국내 첫 코로나19 확진자 발생 후 경과일이었다. 이와 반대로, “따릉이”의 일별 이용 건을 감소시키는 요인은 코로나 확진자 수의 증가, 주말 여부, 겨울 여부, 봄 여부, 강우량이었다.

예를 들어, 국내 첫 코로나19 확진자가 발생한 2020년 1월 20일 이후(TREAT)일 때에는 이전에 비해 약 7,901.3명의 이용자 수가 늘었으며 코로나19 확진자 수(COV)가 1명이 늘어남에 따라 약 27.9명의 이용자 수가 줄어들지만 국내 첫 코로나19 확진자 발생 경과일(EDAY)이 하루 지날 때마다 약 68.2명의 이용자 수가 늘어남을 알 수 있었다.

그리고 눈 또는 비가 오면 이용자 수는 약 1,045.8명 줄어들고 겨울인 12월에서 2월까지의 약 45,360명, 봄인 3월에서 5월까지의 약 13,770명의 이용자 수가 줄어들고 여름인 6월부터 8월까지의 약 5894.1명 늘어나는 것으로 분석됐다. 마지막으로 공휴일, 주말 등의 휴일에 경우에는 평일보다 약 5,689명의 이용자 수가 줄어드는 것으로 분석이 됨을 알 수 있었다.

### 5.2 머신러닝 결과

표 5와 식 (1)과 같이 0.676의 설명력과 모든 독립 변수가 99.9% 이상의 신뢰수준을 갖춘 다중 선형 회귀분석 모형을 표 3과 같은 설정의 파이썬의 텐서플로우 모듈에 대입하여 수행한 결과는 표 6과 같았다. 수행 결과, 동일 배치 사이즈일 때, 에포크가 커질수록 MAPE 값이 작아지는 것을 알 수 있었다.

표 6. 인공지능 학습 수행 결과

Table 6. Result of artificial intelligence learning

| No. | Epoch      | Batch size | MAPE       |
|-----|------------|------------|------------|
| 1   | 1          | 1          | 0.99683614 |
| 2   | 10         | 1          | 0.98295114 |
| 3   | 100        | 1          | 0.90024521 |
| 4   | 1,000      | 1          | 0.89815548 |
| 5   | 10,000     | 1          | 0.7293443  |
| 6   | 100,000    | 1          | 0.5684107  |
| 7   | 1,000,000  | 1          | 0.4453619  |
| 8   | 1          | 10         | 0.99754048 |
| 9   | 10         | 10         | 0.99502472 |
| 10  | 100        | 10         | 0.96986218 |
| 11  | 1,000      | 10         | 0.89239082 |
| 12  | 10,000     | 10         | 0.82960073 |
| 13  | 100,000    | 10         | 0.65265572 |
| 14  | 1,000,000  | 10         | 0.46742352 |
| 15  | 10,000,000 | 10         | 0.43998815 |

그림 1은 본 연구에서 제안하는 인공지능 모형의 배치 사이즈가 1일 때의 MAPE 결과물을 나타낸 것으로 에포크가 1,000,000 일 때, 0.4453619까지 낮아짐을 알 수 있었다.

그림 2는 본 연구에서 제안하는 인공지능 모형의 배치 사이즈가 10일 때의 MAPE 결과물을 나타낸 것으로 에포크가 10,000,000 일 때, 0.43998815까지 낮아짐을 알 수 있었다.

따라서 본 연구에서 1차적으로 다중 선형 회귀분석을 통한 Adj. R2는 0.676으로 약 67.6%의 설명력을 갖추었음을 알 수 있었다. 또한, 이 회귀분석 모형을 기반으로 텐서플로우를 수행한 결과, 동일한 배치 사이즈를 기준으로 에포크를 늘림에 따라 MAPE 값이 낮아짐으로써 머신러닝의 예측 정확도가 더 높아지는 것을 알 수 있었다.

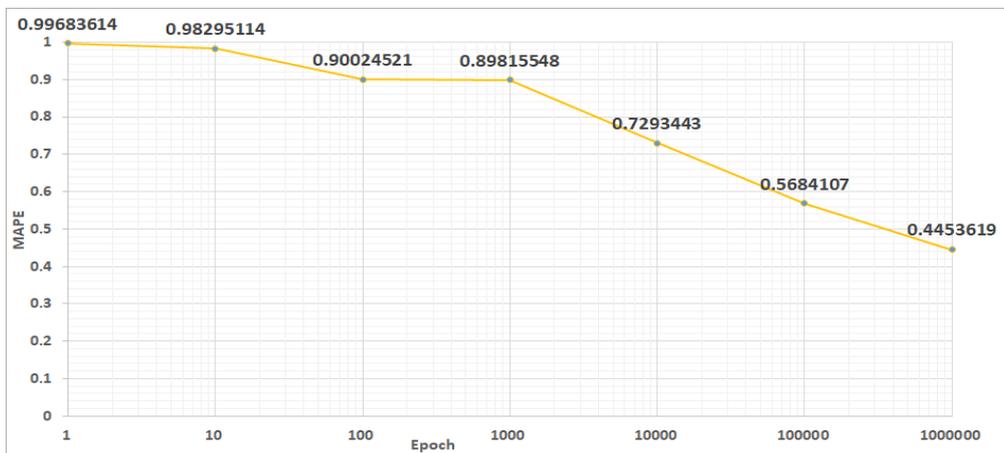


그림 1. 배치 사이즈 1일 때 인공지능 모형의 MAPE 결과물  
Fig. 1. MAPE results for the artificial intelligence model with the batch size of 1

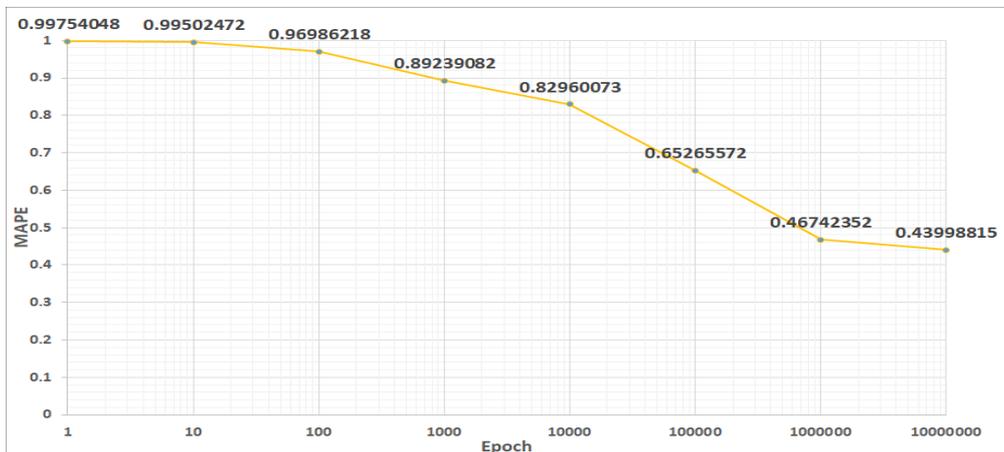


그림 2. 배치 사이즈 10일 때 인공지능 모형의 MAPE 결과물  
Fig. 2. MAPE results for the artificial intelligence model with the batch size of 10

## VI. 결 론

코로나19 바이러스의 확산은 우리의 생활방식을 비대면화, 온라인화 하면서 개인의 교통수단 선택에도 영향을 주고 있다. 이로 인해 버스, 지하철과 같은 대중교통보다 PM의 한 종류인 서울시 공공 자전거인 “따릉이”의 일별 이용 건도 증가하고 있다.

본 연구에서는 코로나19 바이러스 확산에 따른 “따릉이” 이용자 수를 예측할 수 있는 인공지능 모형을 제안하였다. 1차적으로 해당 모형의 설명력과 신뢰도를 확보하기 위하여 다중 선형 회귀분석을 수행하였다. 이 모형을 기반으로 파이썬 텐서플로우 모듈에 데이터를 학습시키고 해당 모형의 신뢰성을 MAPE를 통해 검증하였다.

그 결과, 다중 선형 회귀분석의 설명력인 Adj. R2는 0.676이었으며 각 독립변수들의 유의수준은 모두 0.001 미만으로 신뢰도는 99.9% 이상이였다. 이 모형을 텐서플로우 모듈에 대입하여 데이터를 학습시킨 후 모형의 신뢰성을 검증하였다. 동일한 배치 사이즈를 기준으로 에포크를 증가시킬수록 MAPE가 작아져 그 예측 정확도가 높아지는 것을 알 수 있었다. 또한, 코로나19 관련 독립변수인 TREAT, COV, EDAY 변수를 포함한 경우의 Adj. R2는 0.676으로 해당 변수들을 포함하지 않았을 때의 0.600 보다 약 12.67% 높아지는 것을 확인하였다. 따라서 2020년 이후 데이터에 대해서는 TREAT, COV, EDAY 변수를 포함해야 그 설명력이 더 높아지는 것을 알 수 있으므로 코로나19의 영향을 알 수 있었다.

추후 연구에서는 코로나19 바이러스 확산에 따른 거리두기 단계, 재택근무, 비대면 교육 여부, “따릉이”의 배치 장소, 사용가능 “따릉이”의 대수 등의 환경변수를 추가함으로써 MAPE 값을 더 줄여 정확도가 높은 예측 모형을 구현할 수 있을 것이다. 본 모형을 통해 서울시 공공 자전거인 “따릉이” 뿐만 아니라 킥보드나 렌터카 등의 공유 모빌리티의 이용 건 예측에도 활용될 수 있을 것으로 기대한다.

## Acknowledgements

본 논문은 2021년 한국정보기술학회 종합학술대

회에서 발표한 논문(코로나19 확산에 따른 공공 자전거 이용자 수 예측을 위한 인공지능 모형의 제안)을 확장한 것임.

## References

- [1] Y. I. Bae and H. R. Shin, "COVID-19, accelerate the untact society", Gyeonggi Research Institute, pp. 1-26, May 2020.
- [2] G. J. Lee, H. S. Kim, S. J. Kim, and S. H. Choo, "Analysis of the impact of the spread of COVID-19 on transportation use using the VAR model", Proceedings of the Korea Institute of Intelligent Transport Systems, Jeju, Korea, pp. 349-352, Nov. 2020.
- [3] S. H. Shin and S. H. Choo, "Exploring the Impact of COVID-19 Pandemic on Change in Use of Sharing Mobility Focused on Residents in New Towns", Proceedings of the Korea Institute of Intelligent Transport Systems, Gangneung, Korea, pp. 560-566, Apr. 2021.
- [4] Seoul Metropolitan Government, <https://news.seoul.go.kr/traffic/archives/504919>. [accessed: May 16, 2021]
- [5] SEOUL BIKE, "<https://bikeseoul.com/>". [accessed: May 16, 2021]
- [6] S. A. Min and Y. S. Jung, "Comparative study of prediction models for public bicycle demand in Seoul", Journal of the Korean Data and Information Science Society, Vol. 32, No. 3, pp. 585-592, May 2021. <http://dx.doi.org/10.7465/jkdi.2021.32.3.585>.
- [7] Y. S. Kim, S. O. Park, and G. W. Park, "Analysis of the Seoul public bikes usage for new rental locations", The Korean Journal of Applied Statistics, Vol. 33, No. 6, pp. 739-751, Dec. 2020. <https://doi.org/10.5351/KJAS.2020.36.6.739>.
- [8] G. H. Lee and H. J. Park, "Analysis of data and forecast of public bicycle demand according to weather factor and public bicycle rental rate", Proceedings of the Korean Information on Science

Society Conference, Jeju, Korea, pp. 960-962, Jun. 2019.

[9] E. K. Goh, "Application of Machine Learning Algorithm to Predict Happiness of Elementary 3rd Graders in Korea", The Journal of Learner-Centered Curriculum and Instruction, Vol. 20, No. 3 pp. 1113-1128, Jul. 2020. <http://dx.doi.org/10.22251/jlcci.2020.20.13.1113>.

[10] S. J. Ahn, D. S. Jang, and S. S. Park, "Comparing Gasoline Shipment Prediction Model Performance of Oil Reservoir using Machine Learning", The Journal of Korean Institute of Information Technology, Vol. 18, No. 11, pp. 17-23, Nov. 2020. <http://dx.doi.org/10.14801/jkiit.2020.18.11.17>.

[11] J. Qin, J. Liang, T. Chen, X. Lei, and A. Kang, "Simulating and Predicting of Hydrological Time Series Based on TensorFlow Deep Learning", Polish Journal of Environmental Studies, Vol. 28, No. 2, pp. 795-802, Sep. 2019. <http://dx.doi.org/10.15244/pjoes/81557>.

[12] Seoul Open Data Plaza, <https://data.seoul.go.kr>. [accessed: May 16, 2021]

[13] Korea Centers for Disease Control and Prevention, <http://ncov.mohw.go.kr/> [accessed: May 16, 2021]

[14] Encyclopedia of Korean Folk Culture, <http://encykorea.aks.ac.kr/Contents/Item/E0003235>. [accessed: May 16, 2021]

[15] Korea Meteorological Administration, <https://www.weather.go.kr>. [accessed: May 16, 2021]

[16] Tensorflow, <https://www.tensorflow.org/> [accessed: May 16, 2021]

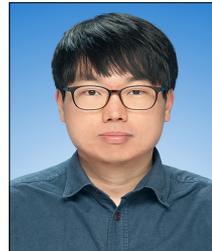
[17] F. Liantoni and A. Agusti, "Forecasting Bitcoin Using Double Exponential Smoothing Method Based on Mean Absolute Percentage Error", International Journal on Informatics Visualization, Vol. 4, No. 2, pp. 91-95, Apr. 2020. <http://dx.doi.org/10.30630/ijov.4.2.335>.

[18] S. W. Kim, "A Comparison of Accuracy among Tourism Demand Forecasting Models :

Jeju-island", Northeast Asia Tourism Research, Vol. 14, No. 2, pp. 219-237, May 2018. <http://dx.doi.org/10.35173/NATR.14.2.11>.

## 저자소개

김 규 석 (Kyuseok Kim)



2011년 2월 : 한국항공대학교

정보통신공학(공학사)

2019년 8월 : 아주대학교

정보통신공학(공학석사)

2019년 6월 : LG전자(주)

선임연구원

2020년 2월 : (주)LG유플러스 책임

2021년 10월 현재 : 서울대학교 환경계획학과 교통학  
전공(박사과정), 한국폴리텍대학 분당융합기술교육원  
데이터융합SW과 조교수

관심분야 : ITS, UTM, UAV, Context-awareness, 데이터  
분석

서 예 진 (Yejin Seo)



2002년 2월 : 울산대학교

전기전자자동화공학부(공학사)

2004년 2월 : 울산대학교

전기전자정보시스템공학부(공학  
석사)

2015년 2월 : 울산대학교

전기전자컴퓨터공학과(공학박사)

2021년 10월 현재 : 한국폴리텍대학 구미캠퍼스 전기과  
조교수

관심분야 : 시스템제어, 신호처리, 머신러닝, 딥러닝