

# 코로나19와 대중의 교통수단 선택 변화에 대한 인공지능 기반 예측 모델 연구

김주희\*, 김현정\*\*, 김규석\*\*\*

## A Study on the Artificial Intelligence-based Predictive Model for COVID-19 and Changes in Public Transportation Choice

Juhee Kim\*, Hyunjung Kim\*\*, and Kyuseok Kim\*\*\*

### 요 약

코로나19의 장기화로 인하여 우리 생활에 많은 변화가 생겼다. 사람들은 외출 전에 항상 마스크를 준비하며, 기업들은 대면 활동을 줄이기 위해 재택근무를 병행하고 있다. 그러나 코로나19 확진자가 처음 발생한 2020년 초와 그 이후의 상황은 조금씩 변해오고 있다. 2020년 초에는 코로나19의 일일 확진자 수가 100명 이상이면 매우 위급한 상황으로 각종 활동들이 비대면으로 전환되었지만, 2023년 현재에는 코로나19의 확진자 수가 100명이면 상대적으로 적은 것으로 느껴진다. 본 연구에서는 코로나19 발생에 따라, 대중들이 어떤 교통수단을 선택하는지에 대한 시점별 분석을 수행한다. 또한, 이 연구 모형을 활용하여 DNN 기반의 교통수단 선택 예측 모형을 제안한다. 본 연구모형의 MAPE는 평균 15.76421, 표준편차 1.10335로 회귀분석의 결정계수 0.5 ~ 0.7 보다도 높은 수준이었다. 본 연구결과를 통해 코로나19 확진자 수에 따른 교통계획 등의 정책에 도움을 줄 수 있을 것으로 기대한다.

### Abstract

COVID-19 has brought about many changes in our lives. People always prepare a mask before going out, and companies are working from home to reduce face-to-face activities. In early 2020, when the number of daily confirmed cases of COVID-19 exceeded 100, it was a very urgent situation and various activities were switched to non-face-to-face. This paper analyzes the choice of transportation between private cars and public transportation such as buses and subways according to the outbreak of COVID-19 at each point in time. In addition, we propose a DNN-based transportation choice prediction model based on this research model. The average MAPE of the research model was average 15.76421 and standard deviation of it was 1.10335, which were higher than 0.5 ~ 0.7 of the Adj. R2 in linear regression analysis. It is expected that the results of this paper would help with the policies such as transportation plans according to the number of confirmed COVID-19 cases.

### Keywords

COVID-19, traffic volume, transportation, seoul city

\* 한국폴리텍대학 대전캠퍼스 스마트소프트웨어과 교수

- ORCID: <https://orcid.org/0000-0002-6343-6587>

\*\* 한동대학교 창의융합교육원 조교수

- ORCID: <https://orcid.org/0000-0003-4894-6906>

\*\*\* 한국폴리텍대학 분당융합기술교육원 데이터융합SW과 조교수,

서울대학교 환경대학원 도시계획학 박사과정(교신저자)

- ORCID: <https://orcid.org/0000-0001-6613-5125>

• Received: Feb. 02, 2023, Revised: Feb. 28, 2023, Accepted: Mar. 03, 2023

• Corresponding Author: Kyuseok Kim

Dept. of Data Convergence Software, Korea Polytechnics

Tel.: +82-31-696-8832, Email: kyuseokkim@kopo.ac.kr

## 1. 서론

2020년은 신종 코로나바이러스 감염증(코로나19) 상황이 발생한 해로서 감염이 점차 확산되면서 우리 생활에 많은 변화를 가져왔다.

첫째, 우리 생활에 언택트(Un-tact, 비대면)라는 단어를 새롭게 가져왔다. 언택트 문화는 소비시장에 전반적인 변화를 가져왔다. 사람들은 마트에 가서 직접 구매를 하기보다는 온라인으로 주문하는 경우가 늘어 2020년 3분기에는 전년도 같은 기간보다 24.6% 증가한 42조 411억원을 기록하였다(통계청, 2020)[1]. 그에 따라 택배 시장도 증가하여 2020년 2월 ~ 7월 기간 동안 전년 대비 24% 증가하였다[2]. 그러나 이러한 언택트 성향은 소비 분야에서만 생긴 것이 아니다. 기업들도 재택근무를 활성화하여 2020년에는 공공기관의 재택근무 인원이 전년도 대비 114.9배 증가하였다[3]. 이렇듯 언택트 문화는 우리 생활의 전반을 바꿔놓고 있다.

둘째, 안전에 대한 사회적 우려가 크게 증가했다. 기존에는 봄철 미세먼지로 인해 마스크를 쓰는 사람들의 수는 일부였지만, 지금은 길거리에서도 마스크를 쓴 사람보다 안 쓴 사람을 찾기 어려울 정도이다. 실제로, 버스나 지하철과 같은 대중교통 탑승 시 마스크 착용은 의무이기도 해서 실외에서 마스크 미착용자는 찾아보기 보기 어려운 상황이다. 마스크 미착용으로 인한 대중교통에서의 충돌 상황도 흔히 않게 보도되는 실정이다[4][5].

코로나19로 인하여 언택트 문화의 확산으로 교통수단의 선택이나 교통량의 변화도 있었다. 신규 코로나 확진자 수가 늘어 사회적 거리두기 단계가 상승하면 재택근무가 더 활성화되고 그에 따라 이동량이 줄어 교통량이 줄어들 수밖에 없다. 또한, 코로나19로부터의 안전을 이유로 대중교통보다 자가용 이용의 비율이 높아질 것이라고 예측할 수 있다.

코로나19가 이처럼 생활의 많은 영역에 영향을 끼치기 때문에 그동안 코로나19 바이러스 확산에 따른 학교 교육의 변화, 소비 패턴의 변화 등의 다양한 현상 분석 및 예측 등의 연구가 수행되었다.

본 연구에서는 특히 코로나19의 확산에 따른 교통수단 선택에 초점을 맞추어 분석을 수행한다. 연구 데이터로는 2019년 1월 1일부터 2022년 11월 30

일까지의 서울시의 일별 버스 이용자, 일별 지하철 이용자, 일별 교통량 기준으로 한다. 연구 모형은 연차별로 데이터를 나누어 진행한 선형 회귀분석과 DNN 기반의 인공지능 모형을 활용한다.

본 논문의 2장에서는 기존 연구에 대한 고찰을 진행하며, 3장에서는 본 연구의 방법에 대하여 설명한다. 4장에서는 연구 결과에 대해 기술하며, 5장에서는 결론과 추후 연구에 대하여 설명한다.

## II. 관련 연구

### 2.1 교통수단 선택 연구

Ma et al.(2009)는 교통약자인 고령자를 대상으로 교통수단 선택에 관한 연구를 하였다[6]. 이를 위하여 고령자의 통행특성과 통행 행위에 관한 기본 정보를 파악하였다. 연구 데이터로는 수도권 가구통행실태조사의 자료이며, 연구 방법으로는 임의절편모형을 생성하여 분석을 진행하였다. 분석 결과, 수도권 경우 고령자들은 자동차와 버스보다는 지하철과 도보의 이용 비중이 크다는 점과 지역과 토지이용 상황에 따라 고령자의 통행수단 선택에 큰 편차가 있음을 알 수 있었다.

Sung et al.(2014)은 장거리 통행의 수단선택에 관한 연구를 하였다[7]. 연구 데이터로는 전국 가구통행실태조사 중장거리 통행 원시자료이며, 연구 방법으로는 이항로짓모형을 적용하여 분석을 진행하였다. 장거리 통행에 있어 승용차의 대안인 고속철도의 활용 가능성에 초점을 두어 연구를 진행하였고, 통행목적별 그 요인이 차별적이라는 것을 밝혔다. 이를 통해 중장거리 통행에 있어 고속철도와 승용차의 경쟁관계에 대한 결정요인을 파악할 수 있는 연구였다.

Lee et al.(2019)는 머신러닝을 활용한 개인의 교통수단 선택 예측모형을 구축하는 연구를 하였다[8]. 연구 데이터로는 서울시 가구통행실태조사 자료이며, 연구 방법으로는 다항로짓모형, 의사결정나무, 서포트 벡터 머신의 세 가지 머신러닝 기법을 사용하여 분석을 진행하였다. 이를 통해, 4차 산업혁명 시대에 인공지능 기술을 통해 개인의 교통수단 선택 예측을 할 수 있다는 점을 알 수 있었다.

## 2.2 코로나19와 교통 관련 연구

Jang(2020)은 코로나가 통행행동에 미치는 영향에 관하여 네덜란드를 대상으로 연구하였다[9]. 네덜란드의 사람들은 코로나19의 위기로 인해 통행횟수와 총 통행거리가 약 29%, 37%가 줄었다고 하였다. 이러한 부분들은 E-bike의 성장에도 영향을 부분적으로 받았을 수 있으나 기차와 같은 대중교통 이용이 축소된 점이 근거가 될 수 있다. 따라서 이 연구는 코로나19로 인해 통행수단 선택, 거리 등 통행행동의 변화가 있었음을 분석한 연구이다.

Parker et al.(2020)은 Southern California 지역의 코로나19의 확산을 줄이기 위해 시행한 2020년 3월 19일의 “Stay-At-Home” 행정 명령에 따른 교통량의 변화에 관하여 연구하였다[5]. 이 명령으로 인해 해당 기간 동안의 교통량은 과거 5년에 비해 최대 50%가 감소하였으며, 이산화질소 등의 감소로 공기 질의 개선도 있었음을 분석하였다.

Lee et al.(2020)은 코로나19로 인한 교통의 영향에 대한 연구를 수행하였다[10]. 코로나 확산 이후 대중교통보다는 개인 교통을 선호하는 경향이 있음을 기술통계를 통해 확인하고, 화물 교통에 대한 영향도 정리하였다.

Lee et al.(2020)은 코로나 확산에 따른 2020년 초 3개월간의 교통량 변화에 관하여 연구하였다[10]. 연구결과 코로나의 확진자 수 증가와 교통량의 증가는 반비례함을 밝혔다.

Hong et al.(2020)은 코로나19 확산에 따른 서울시 지하철 이용 변화를 연구하였다[11]. 연구 데이터는 서울 열린데이터광장의 지하철 역별 일별 시간대별 승하차 인원 자료이며, 연구 방법은 군집분석을 이용하여 진행하였다. 이 연구결과를 통해 코로나19 이후 승하차 인원의 감소가 32.4% ~ 46.9%까지 시간대별, 지역별로 다양하게 나타남을 알 수 있었다.

## 2.3 선행연구와의 차이점

본 연구는 교통수단 선택의 변화에 대하여 코로나19의 발생 전후를 비교하여 면밀한 수치 분석을 한다는 점에서 큰 차이가 있다.

첫째, 코로나19로 인한 교통량의 변화나 사회현상의 정량적 변화에 대해 분석한 연구들은 기존에도 존재한다. 그러나 본 연구는 교통량 및 대중교통 이용자 데이터를 기반으로 교통수단 선택의 변화에 대하여 연구를 한다는 점에서 차별성을 갖는다.

둘째, 교통수단 선택에 있어 요일, 날씨, 계절 등의 요인도 포함을 하는 상세한 연구이다. 따라서 코로나19의 확진자 수만을 변수로 두는 것이 아니라 다양한 요인들을 포함시켜 그 인과관계를 분석하는데 정확도를 높이고자 한 연구이다.

따라서 본 연구는 교통수단 선택에 영향을 줄 수 있는 다양한 요인을 검토하고 활용하여 코로나19로 인한 그 변화를 분석하는데 주안점을 둔 연구이다.

## III. 연구 방법

### 3.1 연구 데이터

본 연구에서는 코로나19로 인한 서울시민의 교통수단 선택의 변화를 분석하였다. 표 1과 같이 코로나 발생 전, 후를 비교할 수 있도록 2019년 1년치와 2020년부터 2022년 11월까지 약 3년치를 약 1년 단위로 구분하였다. 공간적 범위는 서울시내 전역이며, 데이터의 종류는 일별 버스 이용자 수, 지하철 이용자 수, 교통량, 날씨, 계절 변수이다. 교통수단의 선택을 연구 주제로 잡았기 때문에 교통수단 간 비율을 종속변수로 선택하였다.

표 1. 연구 데이터  
Table 1. Research data

| Category         | Description   |
|------------------|---|
| Temporal range   | 2019.01.01. ~ 2022.11.30.   |
| Spatial range    | Seoul city  |
| COVID-19         | Number of COVID-19 confirmed patients   |
| Bus passenger    | All the bus stops in Seoul city   |
| Subway passenger | All the subway links in Seoul city  |
| Traffic volume   | All the links in Seoul city   |
| Weather          | Temperature and rainfall  |
| Season           | Winter : December ~ February<br>Spring : March ~ May<br>Summer : June ~ August<br>Fall : September ~ November |

코로나19의 신규 확진자 수는 KOSIS 국가통계포털의 일별 통계자료를 활용하였다[1]. 그리고 일별 버스 이용객 수와 지하철 이용객 수는 서울 열린데이터광장의 일별 시간대별 역별 승하차 인원 자료를 활용하였다[12]. 또한, 교통량은 서울시 교통정보시스템의 교통량 정보를 활용하였으며[13], 기온, 강수량 등의 날씨 정보는 기상청의 과거 날씨 일별 자료를 활용하였다[14].

### 3.2 변수 설정

본 연구에서의 변수는 표 2와 같다. 종속변수는 해당 일의 교통량 대비 대중교통 이용객인  $RPT_i$ 로 한다. 이 값은 해당 일의 버스 이용객인  $NBD_i$ 와 지하철 이용객인  $NSD_i$ 의 합을 일별 교통량인  $TVD_i$ 로 나눈 값으로 식 (1)과 같이 계산된다.

$$RPT_i = \frac{NBD_i + NSD_i}{TVD_i} \quad (1)$$

표 2. 변수  
Table 2. Variables

| Variable | Abbr.(Description)  |
|----------|---|
| RPT      | Ratio of public transportation passengers per traffic volume) |
| NBD      | Number of bus passengers per day                              |
| NSD      | Number of subway passengers per day                           |
| TVD      | Traffic volume per day  |
| COV      | NCP Number of COVID-19 patients                               |
|          | ED Elapsed days of COVID-19                                   |
| TEM      | DS Dummy variable for Saturday                                |
|          | DH Dummy variable for holiday                                 |
|          | DW Dummy variable for Winter                                  |
|          | DSP Dummy variable for Spring                                 |
|          | DSM Dummy variable for Summer                                 |
| WT       | AT Average temperature  |
|          | HT Highest temperature  |
|          | LT Lowest temperature   |
|          | ACC Average cloud cover                                       |
|          | RF Volume of rainfall   |

서울 시내에서의 대중교통은 버스와 지하철이며, 자가용의 이용은 따로 추산되지 않으나 교통량의 변화에 크게 의존적이다. 그리고 도로 교통에서의

버스의 비중은 상대적으로 적기 때문에 도로에서의 교통량을 자가용의 이용량으로 대표한다.

독립변수는 크게 3가지로 나뉜다. 첫째, 코로나19 관련 변수인  $COV(COVID-19)$  변수 중  $NCP$ 는 일일 코로나19 확진자 발생 수이며,  $ED$ 는 코로나19 확진자가 처음 발생한 2020년 1월 19일로부터의 경과일을 의미한다. 2020년 1월 19일 이전은 모두 0이며 그 이후는 1부터 하루 당 1씩 증가한다. 둘째, 요일과 계절 관련 더미변수인  $TEM(Temporal)$  변수 중  $DS$ 는 토요일 여부,  $DH$ 는 휴일 여부,  $DW$ 는 겨울 여부,  $DSP$ 는 봄 여부,  $DSM$ 은 여름 여부를 나타내는 더미변수들이다. 마지막으로 날씨 관련 변수인  $WT(Weather)$  변수 중  $AT$ 는 평균 기온,  $HT$ 는 최고 기온,  $LT$ 는 최저 기온이며,  $ACC$ 는 평균 구름의 양,  $RF$ 는 강수량을 의미한다.

### 3.3 연구 모형 및 방법론

본 연구에서는 대중의 교통수단 선택에 변화를 줄 수 있는 요인을 분석하는 것을 목적으로 한다. 우선 종속변수로는 대중교통과 개인교통의 상대적인 비율을 살펴볼 수 있는  $RPT$ 이다. 그리고 독립변수인  $COV$ ,  $TEM$ ,  $WT$  변수들을 포함하여 모형은 식 (2)와 같이 나타낼 수 있다.

$$RPT_i = \beta_0 + \beta_1 COV_i + \beta_2 TEM_i + \beta_3 WT_i + \epsilon \quad (2)$$

본 연구에서는 위 연구 모형을 활용하여 다중회귀분석을 수행하고,  $DNN(Deep\ neural\ network)$  기반의 인공지능 모형의  $MAPE(Mean\ absolute\ percentage\ error)$ 를 계산하여 예측의 정확도를 산출한다.

$DNN$  예측에 활용되는 모형의 Dense 레이어의  $Input\ shape$ 는 12개부터 시작하여 절반씩 줄여나가 총 4개의 레이어로 구성된다. 그리고 이 모형에 사용되는 변수도 다중 선형 회귀분석과 같이 식 (2)의 변수들로 구성된다. 또한 학습 데이터의 사이즈(batch size)는 1, 학습 횟수는 100(epochs)으로 설정하며, 10회 반복 수행한다. 수행 시마다 학습 데이터와 검증 데이터를 7:3으로 임의 분할하여 연산된  $MAPE$  값을 통해 이 연구모형의 정확도를 측정한다.

#### IV. 연구 결과

##### 4.1 기술 통계

본 논문에서 활용한 코로나19 확진자와 대중교통 이용객과 교통량 관련 기술 통계량은 표 3과 같다.

연도별 코로나19 확진자 수의 평균은 2019년부터 2022년 11월 말까지 일평균 각 0, 167.2, 1,607.1, 79,356.3명이다. 2021년의 이 수치는 2020년 대비 약 9.61배, 2022년에는 2021년 대비 약 49.38배, 2020년 대비 약 4,74.619배로 기하급수적으로 증가했음을 알 수 있다.

일평균 버스 이용객(NBD)은 2019년 약 10,206,355.6명이었으나 코로나19 발생 첫 연도인 2020년에는 약 7,136,146.6명으로 약 30.1% 감소하였다. 일평균 지하철 이용객(NSD)은 2019년 약 14,856,349.1명이었으나 코로나19 발생 첫 연도인 2020년에는 약 10,738,585.1명으로 약 27.7% 감소하였다. 일평균 교통량(TVD)은 2019년 약 9,187,764.6건이었으나 코로나19 발생 첫 연도인 2020년에는 약 8,736,219.3으로 약 4.9% 감소하였다.

표 3. 연도별 기술 통계

Table 3. Descriptive statistics by year

|                      | 2019   | 2020   | 2021   | 2022   |
|----------------------|--|--|--|--|
|                      | Average(Minimum/ Maximum/ STD.DEV)                         |  |  |  |
| NCP<br>[pers<br>ons] | 0(0/0/0)   | 167.2<br>(0/1,241/<br>254.1)                               | 1,607.1<br>(288/22,448/<br>1,908.4)                    | 79,356.3<br>(3,024/<br>621,328/<br>97,140.6)           |
| NBD<br>[pers<br>ons] | 10,206,355.6<br>(3,672,125/<br>12,303,233/<br>2,095,880)   | 7,136,146.6<br>(2,363,799/<br>11,328,678<br>2,382,604.7)   | 7,734,515.1<br>2,560,592<br>9,823,727<br>1,869,410.9   | 8,412,409.1<br>2,956,180<br>10,822,542<br>1,938,395.9  |
| NSD<br>[pers<br>ons] | 14,856,349.1<br>(5,652,814/<br>18,425,509/<br>2,981,073.9) | 10,738,585.1<br>(3,559,972/<br>16,956,830/<br>3,137,583.9) | 10,982,559.5<br>3,149,873<br>14,250,817<br>2,864,650.1 | 12,162,712.3<br>3,766,887<br>16,103,816<br>2,951,006.4 |
| TVD<br>[case]        | 9,187,764.6<br>(3,540,383<br>10,221,439<br>837,708.9)      | 8,736,219.3<br>(5,546,294<br>102,922.0<br>968,051.9)       | 8,551,258<br>4,235,912<br>10,116,188<br>971,560.1      | 8,660,199.1<br>6,018,401<br>10,320,593<br>860,355.6    |
| Misc.                | Jan. 1st ~<br>Dec. 31st,<br>2019                           | Jan. 1st ~<br>Dec. 31st,<br>2020                           | Jan. 1st ~<br>Dec. 31st,<br>2021                       | Jan. 1st ~<br>Nov. 30th,<br>2022                       |

표 3에 따르면 NBD, NSD와 TVD 값은 코로나19 발생 첫 연도인 2020년에 전년대비 급감하였다가 2020년 이후인 2021년, 2022년에는 NBD와 NSD 값은 소폭 증가하였고, TVD는 2021년도에는 감소하였다가 2022년도에는 다시 증가하는 양상을 보인다. 그러나 이 세 변수들의 값은 2022년에는 2021년 대비 모두 증가하였음을 알 수 있다. 하지만 NBD, NSD, TVD 모두 2022년의 값은 코로나19 발생 이전인 2019년의 값에 도달하지 못했음을 알 수 있다. 이는 코로나19 발생 약 3년이 지난 후에도 이전만큼의 대면 활동은 이루어지지 않는 것으로 미루어 볼 수 있다.

##### 4.2 다중선형분석 결과

다중선형회귀 분석을 수행하기 위하여 데이터를 총 5가지로 구분하였다. 첫 번째는 2019년 1월 1일부터 2020년 1월 19일까지 코로나19 확진자가 0명이었던 시기이다. 두 번째는 2020년 1월 20일부터 2020년 12월 31일까지의 코로나19 발생 첫해이다. 세 번째는 2021년 1월 1일부터 2021년 12월 31일까지의 코로나19 발생 2년차이며, 네 번째는 2022년 1월 1일부터 2022년 11월 30일까지 코로나19 발생 3년차이다. 마지막은 2019년 1월 1일부터 2022년 11월 30일까지의 전체 데이터이다. 첫 번째 데이터에서 코로나19 관련 변수인 NCP와 ED의 값은 모두 0이므로 해당 변수를 제외한 후 분석을 수행하며, 다른 데이터 그룹은 이 변수들을 모두 포함시킨다.

코로나19 발생 이전의 데이터를 기반으로 다중선형회귀 분석을 수행한 결과는 표 4와 같다. 분석 결과, DS, DH, DSM 변수만이 종속변수에 영향을 주는 유의한 변수로 분석되었다. 그리고 이 변수들의 계수는 모두 음수로 토요일, 휴일 그리고 여름이면 대중교통 승객의 비율이 줄어드는 것을 의미한다. 그러나 AT, HT, LT, ACC, RF 등의 수치 변수들은 종속변수에 영향을 주지 못하고, 통제변수로 사용된 더미변수들만이 영향을 끼친다는 것을 알 수 있다. 그러나 이 데이터들을 통한 분석은 오직 더미변수들만이 유의하기 때문에 분석 자체의 유의성은 부족함을 알 수 있다.

표 4. 코로나19 발생 이전의 데이터 분석  
Table 4. Analysis of pre-COVID-19 data

| Variable  | Coefficient | t-value           |
|-----------|-------------|-------------------|
| Intercept | 3.0689***   | 28.207            |
| DS        | -0.5358***  | -10.846           |
| DH        | -0.9706***  | -21.698           |
| DW        | -0.1292     | -1.845            |
| DSP       | -0.0535     | -1.053            |
| DSM       | -0.1816**   | -3.036            |
| AT        | -0.0261     | -0.765            |
| HT        | 0.0061      | 0.315             |
| LT        | 0.0223      | 1.238             |
| ACC       | 0.0056      | 0.729             |
| RF        | -0.0022     | -0.940            |
| R2 0.590  |             | Adjusted R2 0.579 |

코로나19 발생 1년차의 데이터를 기반으로 다중 선형회귀 분석을 수행한 결과는 표 5와 같다. 분석 결과, NCP, ED, DS, DH, DW, DSP, DSM의 변수가 모두 종속변수에 영향을 주는 유의한 변수로 분석되었다. 그리고 이 변수들의 계수는 모두 음수로서 코로나19 확진자 수가 증가할수록, 코로나19 확진자 발생 후 경과일이 증가할수록, 토요일, 휴일, 겨울, 봄, 여름이면 대중교통 승객의 비율이 줄어드는 것으로 분석되었다. 이는 코로나19 발생 1년차에는 코로나19 확진자의 발생 인원과 경과일에 따라 대중교통 이용보다는 자가용 이용을 고려했음을 알 수 있다.

표 5. 코로나19 발생 1년차의 데이터 분석  
Table 5. Analysis of data in the 1st year of COVID-19

| Variable  | Coefficient | t-value           |
|-----------|-------------|-------------------|
| Intercept | 3.2166***   | 37.645            |
| NCP       | -0.0000564  | -0.577            |
| ED        | -0.0028***  | -15.691           |
| DS        | -0.5460***  | -14.071           |
| DH        | -0.8950***  | -25.559           |
| DW        | -0.3254***  | -5.677            |
| DSP       | -0.5016***  | -12.080           |
| DSM       | -0.2450***  | -5.089            |
| AT        | 0.0138      | 0.495             |
| HT        | -0.0127     | -0.822            |
| LT        | -0.0053     | -0.361            |
| ACC       | -0.0023     | -0.387            |
| RF        | 0.0009      | 0.697             |
| R2 0.646  |             | Adjusted R2 0.640 |

코로나19 발생 2년차의 데이터를 기반으로 다중 선형회귀 분석을 수행한 결과는 표 6과 같다. 분석 결과는, 코로나19 발생 1년차의 분석결과와 동일하게 NCP, ED, DS, DH, DW, DSP, DSM의 변수가 모두 종속변수에 영향을 주는 유의한 변수로 분석되었다. 그리고 앞서와 같이 이 변수들의 계수도 모두 음수로서 코로나19 확진자 수가 증가할수록, 코로나19 확진자 발생 후 경과일이 증가할수록, 토요일, 휴일, 겨울, 봄, 여름이면 대중교통 승객의 비율이 줄어드는 것으로 분석되었다. 그러나 코로나19 확진자 수와 확진자 발생 후 경과일인 NCP와 ED 변수의 계수 절댓값이 줄어들었음을 알 수 있다. 이는 코로나19 발생 2년차에는 1년차 보다 코로나19 발생자 수에 따라 덜 민감했음을 시사하는 것으로 볼 수 있다.

표 6. 코로나19 발생 2년차의 데이터 분석  
Table 6. Analysis of data in the 2nd year of COVID-19

| Variable  | Coefficient  | t-value           |
|-----------|--------------|-------------------|
| Intercept | 3.1603***    | 50.203            |
| NCP       | 0.0000434*** | 4.949             |
| ED        | -0.0011***   | -23.393           |
| DS        | -0.5643***   | -19.595           |
| DH        | -0.9163***   | -34.925           |
| DW        | -0.2523***   | -6.155            |
| DSP       | -0.0862**    | -2.880            |
| DSM       | -0.1169***   | -3.348            |
| AT        | -0.0285      | -1.444            |
| HT        | 0.0077       | 0.689             |
| LT        | 0.0145       | 1.415             |
| ACC       | 0.0038       | 0.830             |
| RF        | -0.0020      | -1.733            |
| R2 0.752  |              | Adjusted R2 0.748 |

코로나19 발생 3년차의 데이터를 기반으로 다중 선형회귀 분석을 수행한 결과는 표 7과 같다. 분석 결과는, 코로나19 발생 1, 2년차의 분석결과와 동일하게 NCP, ED, DS, DH, DW, DSP, DSM의 변수가 모두 종속변수에 영향을 주는 유의한 변수로 분석되었다. 그리고 코로나19 발생 1, 2년차와 같이 계수도 모두 음수이다. 그러나 2년차에서의 분석과 같이 3년차의 NCP와 ED의 계수 절댓값은 1, 2년차 대비 더 줄어들었다. 이는 코로나19 발생 경과일이 지날수록, 코로나19 확진자 수가 증가할수록, 코로나19

확진자 발생 후 경과일이 증가할수록, 토요일, 휴일, 겨울, 봄, 여름이면 대중교통 승객의 비율이 줄어드는 것으로 분석되었다. 그러나 코로나19 확진자 수와 확진자 발생 후 경과일인 NCP와 ED 변수의 계수 절댓값이 줄어들었음을 알 수 있다. 이는 코로나19 발생 2년차에는 1년차 보다 코로나19 발생자 수에 덜 민감했음을 시사하는 것으로 볼 수 있다.

표 7. 코로나19 발생 3년차의 데이터 분석  
Table 7. Analysis of data in the 3rd year of COVID-19

| Variable             | Coefficient   | t-value                       |
|----------------------|---------------|-------------------------------|
| Intercept            | 3.1703***     | 48.143                        |
| NCP                  | -0.0000104*** | -6.385                        |
| ED                   | -0.0003***    | -11.609                       |
| DS                   | -0.5605***    | -19.479                       |
| DH                   | -0.9522***    | -36.556                       |
| DW                   | -0.2983***    | -6.825                        |
| DSP                  | -0.1961***    | -6.425                        |
| DSM                  | -0.2009***    | -5.673                        |
| AT                   | -0.0133       | -0.646                        |
| HT                   | 0.0014        | 0.125                         |
| LT                   | 0.0118        | 1.094                         |
| ACC                  | 0.0014        | 0.302                         |
| RF                   | 0.0007        | 0.943                         |
| R <sup>2</sup> 0.736 |               | Adjusted R <sup>2</sup> 0.731 |

마지막으로, 2019년 1월 1일부터 2022년 11월 30일까지 약 4년간의 데이터를 기반으로 다중선형회귀 분석을 수행한 결과는 표 8과 같다.

표 8. 전체 데이터의 분석  
Table 8. Analysis of overall data

| Variable             | Coefficient   | t-value                       |
|----------------------|---------------|-------------------------------|
| Intercept            | 2.7328***     | 45.043                        |
| NCP                  | 0.0000003967* | 2.010                         |
| ED                   | -0.0002***    | -7.096                        |
| DS                   | -0.5670***    | -20.249                       |
| DH                   | -0.8935***    | -35.283                       |
| DW                   | -0.1953***    | -4.781                        |
| DSP                  | -0.2667***    | -9.134                        |
| DSM                  | -0.1507***    | -4.406                        |
| AT                   | -0.0321       | -1.621                        |
| HT                   | 0.0222*       | 2.004                         |
| LT                   | 0.0068        | 0.659                         |
| ACC                  | 0.0033        | 0.741                         |
| RF                   | 0.0009        | 1.132                         |
| R <sup>2</sup> 0.531 |               | Adjusted R <sup>2</sup> 0.527 |

분석결과, 앞에서 언급된 NCP, ED, DS, DH, DW, DSP, DSM의 변수 외에도 HT 변수도 종속변수에 영향을 주는 유의한 변수로 분석되었다. 그리고 일별 코로나 확진자 발생 수인 NCP와 최고 기온 변수인 HT 변수의 계수가 양수였다.

다중선형회귀 분석을 통한 결과의 특징은 3가지로 요약될 수 있다.

첫째, 기간별로 나누어 분석한 결과인 표 5~7에서의 ED 변수의 계수는 모두 음수였고, 그 절댓값은 감소하였다. 전체 데이터를 기준으로 수행한 표 8에서도 음수이다. 그리고 표 8에서 NCP의 계수의 절댓값은 표 4-6 대비 1/10 이하로 미미함을 알 수 있다. 이는 기간별로 나누어서 보면 코로나19 확진자의 수가 대중교통 이용률에 영향을 주지만 전체 기간으로 보면 그 영향도는 매우 작다는 것을 의미하는 것이다.

둘째, 코로나19 확진자 발생 후 경과일을 의미하는 ED 변수도 앞에서의 NCP 변수와 동일하게 그 절댓값이 시간이 지나면서 줄어드는 것으로 나타났다. 이도 코로나19 확진자 발생일이 경과 할수록 코로나19 확진자 발생 초기의 영향보다는 적음을 의미하는 것으로 볼 수 있다.

셋째, 코로나19 발생 이전의 데이터와 전체 기간의 데이터를 기반으로 분석한 표 4와 표 8의 결과를 비교하면 DS, DH, DSM은 종속변수에 영향을 주는 공통의 독립변수로서 토요일, 휴일, 여름에는 대중교통의 이용 비율이 줄어드는 것을 알 수 있다.

#### 4.3 DNN 수행 결과

식 (2)의 변수를 기반으로 DNN을 10회 수행한 MAPE 결과는 표 9와 같다. MAPE의 최솟값은 13.67476, 최댓값은 17.92479였으며, 평균값은 15.76421, 표준편차는 1.10335로 매회 학습 데이터와 검증 데이터를 7:3으로 임의 분할하였음에도 안정적인 결과를 보였음을 알 수 있다.

이는 선형 회귀분석과 DNN을 단순 비교하기는 어렵지만, 전체 데이터를 기준으로 수행한 선형 회귀분석의 설명력인 Adj. R<sup>2</sup>는 0.527이었으나 DNN 기반의 인공지능 모형을 수행한 결과의 MAPE는

평균 15.76421, 표준편차 1.10335였다. 이는 DNN 기반의 인공지능 모형을 수행한 결과가 상대적으로 우수함을 나타낸다고 할 수 있다.

표 9. DNN 수행 결과

Table 9. DNN analysis result

| Category | Value    | Misc.               |
|----------|----------|---------------------|
| Minimum  | 13.67476 | 10 times of repeats |
| Maximum  | 17.92479 |                     |
| Average  | 15.76421 |                     |
| STD.DEV  | 1.10335  |                     |

## V. 결론 및 향후 과제

2019년 겨울, 전 세계로 급속히 확산되기 시작한 코로나19 바이러스로 인해 우리의 생활이 전반적으로 바뀌고 있다[15]. 그중에서도 매일 등하교나 출퇴근을 하는데 활용되는 교통수단에 있어서도 변화가 생겼다. 기존에는 통행특성이나 이용자 특성을 기반으로 한 교통량 변화, 교통수단 선택의 변화에 관한 연구들이 진행되었다. 또한, 코로나19의 확산으로 그에 따른 교통량 변화나 지하철과 같은 단일 교통수단의 변화를 분석하는 연구들은 있었다. 하지만, 본 연구에서는 2020년부터 약 3년간의 코로나19 확산으로 인한 교통수단 선택의 변화에 관하여 실측 데이터를 기반으로 연구를 했다는 점에서 기존 연구와 차별화된 장점을 가진다.

2019년 1월 1일부터 2022년 11월 30일까지 약 4년간의 데이터를 기반으로 다중선형회귀 분석을 수행한 결과, 코로나19 팬데믹 초기와 비교하면 코로나19 확진자 수, 경과일과 같은 관련 독립변수가 대중교통 이용에 미치는 영향도가 줄어드는 것을 알 수 있었다. 또한 단순 비교하기는 어렵지만 DNN 기반의 인공지능 모형의 MAPE 값이 선형 회귀분석의 설명력인 Adj. R2 보다 상대적으로 우수함을 알 수 있었다.

본 연구 결과는 코로나19와 유사한 감염성 질환과 또는 집적의 불이익의 높아지는 사회적 상황에서 대중들의 교통수단 선택에 대한 예측을 통해 보다 선제적인 대응을 할 수 있는 자료로 활용될 수 있을 것이다. 특히, 시계열 분석을 통해 대중들의

초기의 대응 방식과 일정한 시간을 통해 변화되는 양상을 모두 파악할 수 있기 때문에 정부 기관뿐만 아니라 통근을 관리해야 하는 기업, 시민들의 입장에서 이를 참고하여 선제적인 의사결정을 할 수 있을 것이라고 판단한다.

추후 연구에서는 서울시뿐만 아니라 전국을 대상으로 공간적 범위를 넓히고, 코로나19 확진자 수뿐만 아니라 코로나19 바이러스 확산에 따라 영향을 준 다양한 사회적 현상에 대해 분석을 할 수 있을 것이다. 또한, 코로나19의 변이 바이러스가 발생하고 있는 만큼 그에 따른 정책, 계획 수립 등을 선제적으로 할 수 있는 연구의 기초가 될 수 있을 것이다.

## References

- [1] Statistics Korea, <http://kostat.kr> [accessed: Oct. 1, 2022]
- [2] Labortoday, "24% increase in delivery volume due to COVID-19", <https://www.labortoday.co.kr/news/articleView.html?idxno=166785> [accessed: Oct. 1, 2022]
- [3] ALIO, <https://www.alio.go.kr> [accessed: Oct. 1, 2022]
- [4] J. Xiang, E. Austin, T. Gould, T. Larson, J. Shirai, Y. Liu, J. Marshall, and E. Seto, "Impacts of the COVID-19 responses on traffic-related air pollution in a Northwestern US city", *The Science of the Total Environment*, Vol. 747, pp. 141325, Dec. 2020. <https://doi.org/10.1016/j.scitotenv.2020.141325>.
- [5] H. A. Parker, S. Hasheminassab, J. D. Crouse, C. M. Roehl, and P. O. Wennberg, "Impacts of Traffic Reductions Associated With COVID-19 on Southern California Air Quality", *Geophysical research letters*, Vol. 47, No. 23, pp. 1011-1024, Nov. 2020. <https://doi.org/10.1029/2020GL090164>.
- [6] K. R. Ma and Y. H. Yoon, "A Study on the Changes in Travel Mode Choice of the Elderly", *Seoul Studies*, Vol. 10, No. 4, pp. 159-171, Dec. 2009. <http://doi.org/10.23129/seouls.10.4.200912.159>.



- [7] H. G. Sung, M. J. Choi, and S. G. Lee, "Determinants of Transportation Mode Choice for Long-distance Travel in Korea", Journal of Korea Planning Association, Vol. 49, No. 2, pp. 245-257, Apr. 2014. <https://doi.org/10.17208/jkpa.2014.04.49.2.245>.
- [8] Y. H. Lee and S. Y. Hong, "A machine learning approach to the prediction of individual travel mode choices", Journal of the Korean Data Information Science Society, Vol. 30, No. 5, pp. 1011-1024, Sep. 2019. <http://dx.doi.org/10.7465/jkdi.2019.30.5.1011>.
- [9] S. H. Jang, "Impact of COVID-19 on Travel Behavior", Monthly KOTI Magazine on Transport, Vol. 271, Sep. 2020.
- [10] H. C. Lee, S. J. Park, G. R. Lee, J. E. Kim, J. H. Lee, Y. S. Jung, and E. W. Nam, "The relationship between trends in COVID-19 prevalence and traffic levels in South Korea", International journal of infectious diseases, Vol. 96, pp. 399-407, Jul. 2020. <https://doi.org/10.1016/j.ijid.2020.05.031>.
- [11] Y. J. Hong, "Analysis of changes in subway use in Seoul after the spread of COVID-19", Seoul Studies, Jan. 2021.
- [12] Seoul Metropolitan Government, <https://data.seoul.go.kr> [accessed: Oct. 1, 2022]
- [13] Seoul TOPIS, <https://topis.seoul.go.kr> [accessed: Oct. 1, 2022]
- [14] KMA, <https://www.weather.go.kr> [accessed: Oct. 1, 2022]
- [15] K. Kim and Y. Seo, "A proposal for an artificial intelligence model for predicting the number of public rental bike in the covid-19 era", Journal of KIIT, Vol. 19, No. 10, pp. 11-18, Oct. 2021. <https://doi.org/10.14801/jkiit.2021.19.10.11>.

## 저자소개

### 김 주 희 (Juhee Kim)



1990년 2월 : 계명대학교  
컴퓨터공학과(공학사)  
1992년 8월 : 계명대학교  
컴퓨터공학과(공학석사)  
1997년 8월 : 계명대학교  
컴퓨터공학과 박사 수료  
1993년 12월 ~ 현재 :

한국폴리텍대학 대전캠퍼스 스마트소프트웨어과 정교수  
관심분야 : 데이터베이스, 데이터 분석, 머신러닝, 딥러닝

### 김 현 정 (Hyunjung Kim)



2010년 2월 : 한동대학교 (경제학사,  
경영학사)  
도시환경공학(부전공)  
2012년 2월 : 서울대학교  
건설환경공학(공학석사)  
도시계획(세부전공)  
2015년 9월: 도쿄대학 (The

University of Tokyo) 도시공학(공학박사)  
2017년 : Environmental Systems Research Institute  
(ESRI) Korea, Manager(과장)  
2021년 ~ 2022년 : 서울대학교 건설환경종합연구소,  
연구교수  
2022년 ~ 현재 : 한동대학교 창의융합교육원 조교수  
관심분야 : 도시 및 지역계획, 스마트시티, 시공간  
빅데이터 분석, 도시공간분석, GIS, 공간AI

### 김 규 석 (Kyuseok Kim)



2011년 2월 : 한국항공대학교  
정보통신공학(공학사)  
2019년 8월 : 아주대학교  
정보통신공학(공학석사)  
2023년 2월 현재 : 서울대학교  
도시계획학 박사수료  
2019년 6월 : LG전자(주)

선임연구원

2020년 2월 : ㈜LG유플러스 책임연구원  
2020년 ~ 현재 : 한국폴리텍대학 분당융합기술교육원  
데이터융합SW과 조교수  
관심분야 : 데이터 수집, 데이터 분석, 머신러닝, 딥러닝,  
도시계획, 근거리 무선통신, 상황인지