

공유자전거 따릉이 재배치를 위한 실시간 수요예측 모델 연구

Demand Forecasting Model for Bike Relocation of Sharing Stations

김 유 신¹
Yoosin Kim

요 약

서울 도심 내 교통량 감축과 탄소배출을 줄이기 위해 2015년 도입된 공공자전거 따릉이는 이용자가 해마다 배 이상 증가하여 2023년 기준 2700여 대여소에서 4만 3천여 대가 운영 중이며 누적 가입자 4백만 명을 넘어서는 서울시민이 뽑은 가장 성공적인 공공 정책으로 자리매김하였다. 그러나 따릉이 이용이 급속도로 증가됨에 따라 자전거 수요·공급 불일치로 인한 자전거 부족 민원도 급증하여 효율적인 자전거 재배치가 강하게 요구되었다. 이에 본 연구는 공유자전거의 대여·반납 이력 데이터, 기상데이터, 공휴일 정보, 따릉이 대여소 정보 등을 기반으로 따릉이 이용 패턴과 특성을 분석하고, 기계학습 알고리즘을 활용해 대여소별 따릉이 대여·반납 예측 모델을 개발하였다. 이를 이용하여 대여소별 안전재고를 확보할 수 있는 따릉이 재배치 수량을 도출하고 이를 서울시설공단 따릉이 관리App에 시범서비스 하였다. 따릉이의 수요를 실시간으로 예측하고 현재 거치 중인 재고량과 비교하여 적절한 수량의 자전거를 재배치한다면 자전거 부족으로 인한 시민들의 불편 해소에 기여할 수 있을 것으로 기대된다.

☞ 주제어 : 공유자전거, 따릉이, 빅데이터, 수요예측, 안전재고, 기계학습

ABSTRACT

The public bicycle of Seoul, Ttareungyi, was launched at October 2015 to reduce traffic and carbon emissions in downtown Seoul and now, 2023 Oct, the cumulative number of user is upto 4 million and the number of bike is about 43,000 with about 2700 stations. However, super growth of Ttareungyi has caused the several problems, especially demand/supply mismatch, and thus the Seoul citizen has been complained about out of stock. In this point, this study conducted a real time demand forecasting model to prevent stock out bike at stations. To develop the model, the research team gathered the rental·return transaction data of 20,000 bikes in whole 1600 stations for 2019 year and then analyzed bike usage, user behavior, bike stations, and so on. The forecasting model using machine learning is developed to predict the amount of rental/return on each bike station every hour through daily learning with the recent 90 days data with the weather information. The model is validated with MAE and RMSE of bike stations, and tested as a prototype service on the Seoul Bike Management System(Mobile App) for the relocation team of Seoul City.

☞ keyword : Bike Sharing, Relocation, Demand Forecasting, Machine Learning

1. 서 론

도시화의 급속한 진전으로 교통체증, 환경훼손 등이 심각해짐에 따라 이를 해결하고자하는 스마트시티 정책들이 점차 확산되고 있다[1, 2]. 네덜란드 암스테르담에서 시작된 공공자전거도 이의 일환으로 파리와 같이 세계 유명 도시들에서 도심 내 교통량 감축과 탄소배출을 줄이기 위해 널리 활용되고 있다. 서울시 공유자전거 따릉이는 자전거가 거치된 대여소에 이용자가 일정한 비용을 지불하고 거치된 자전거를 빌려타고 반납하는 자전거 공유서비스이다. 따릉이는 2015년 10월 도입 이후 도심

내 틈새 교통 수단으로서 시민 이용률이 급증하였고, 정책 만족도 역시 매우 높아 2017년 이후 3년 연속 서울시민이 뽑은 1위 정책에 선정되었다[16]. 따릉이는 2015년 10월 약 2천여 대의 자전거로 성수, 상암, 여의도, 사대문안, 신촌 다섯 시범 지역에서 첫 선을 보인 후 2023년 현재 2700여 대여소에서 4만 3천여 대가 운영 중이며 누적 가입자 4백만 명을 돌파하여 서울시민 10명중 4명이 이용하고 있다[14, 15]. 또한 하루 평균 이용 건수의 50.6%는 출퇴근 시간대에 활용되는 등 생활밀착형 교통수단으로 자리매김하였다[14]. 이와 같이 공유자전거는 초기 구축 비용이 적고, 탄소 배출을 하지 않음으로 친환경적이며 교통량 감축에도 도움이 되는 등 다양한 측면에서 큰 장점을 가지고 있다. 그러나 공공자전거 서비스의 규모가 커지고 이용이 급속도로 증가됨에 따라 이용자의 다양한 민원 대응과 따릉이의 효율적 운용 관리 시스템에

¹ Data Analytics, AirDeep Inc, Suwon, 16506, Korea

* Corresponding author (yoosin@airdeep.co.kr)

[Received 7 October 2023, Reviewed 10 October 2023, Accepted 15 October 2023]

대한 필요성도 점차 부각되고 있다[1, 2, 3, 4, 5, 6].

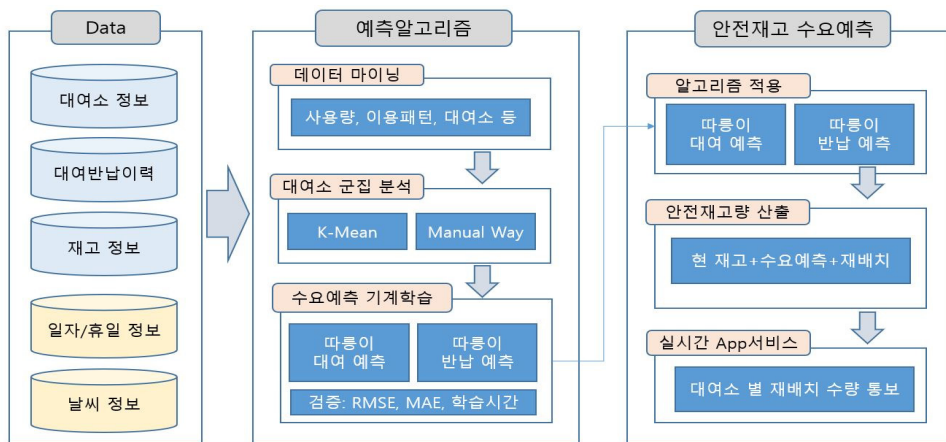
야외에서 운행되는 따릉이의 특성상 계절적 요인에 크게 영향을 받는데, 추운겨울에는 이용율이 급감했다가 봄부터 이용율이 점차 증가하고 야외활동에 적합한 초여름부터 가을까지 사용량이 급증하는 성수기를 맞는다. 날씨와 같은 외부 변수는 따릉이 대여량 전체에 영향을 미칠 수 있으며, 대여소의 위치와 주변 환경 또는 따릉이를 이용하는 목적에 따라 다양한 이용 패턴이 나타날 수 있다. 따릉이는 자전거가 거치된 대여소를 기반으로 운영되는 시스템으로, 어느 대여소에서나 빌리고 다른 대여소에 반납할 수 있는 **One-Way** 방식이다. 때문에 사용자의 자전거 이용이 증가하면 대여소에 거치되어 있는 자전거가 부족하여 대여가 불가능한 경우가 발생한다. 이러한 자전거의 수요·공급 불일치는 사용자의 불만을 야기하게 되므로 운영자는 이러한 불균형을 해소하기 위해 대여소의 자전거 거치량을 적정수준으로 조정하는 자전거 재배치를 하고 있다. 이러한 재배치 활동은 배송팀의 규모, 대여소의 지리적 위치와 이동 거리, 이동 당시의 교통량 등 여러가지 제약이 존재하므로 대여소를 빈번하게 오가며 재배치하는 것은 현실적으로 쉽지 않다.

서울시는 이를 위해 강북에 5개, 강남에 6개, 총 11개 배송센터에서 136명의 배송 직원들이 대여소에 따릉이를 재배치를 하고 있다[15]. 자전거 재배치 방식은 배송팀이 서울시설공단의 대여소현황 App에서 제공되는 대여소별 실시간 자전거 거치 현황을 보고 특정 대여소의 자전거가 일정수준 이상으로 많거나 적을 경우 해당 대여소 자

전거를 회수 혹은 재배치한다. 그럼에도 불구하고 따릉이 이용자가 급증하면서 대여소별로 자전거 부족, 과잉 등 이용률 편차가 발생하고 있는데, 이를 해소할 수 있는 효율적 자전거 재배치 방안이 요구되고 있다.

그러므로 재배치 활동을 최적화하기 위해서는 이동 시점의 대여소별 자전거 거치 현황뿐만 아니라 앞으로 발생할 자전거 대여와 반납을 미리 예측하여 적절한 수준의 재배치 수량을 파악하고 이를 위해 필요한 자전거 배치 또는 회수 활동이 요구된다. 만약 자전거 수요를 실시간으로 예측하고 현재 거치 중인 재고량과 비교하여 적절한 수량의 자전거를 재배치한다면 자전거 부족으로 인한 시민들의 불편을 해결할 수 있을 것이다. 이러한 관점에서 본 연구는 데이터 분석과 기계학습 알고리즘을 활용해 따릉이 수요예측 모델을 개발하여 각 대여소별 안전재고를 확보할 수 있는 자전거 재배치 수량을 도출하고자 한다. 먼저 따릉이와 대여소의 특성 및 이용 패턴 분석 위해 주관기관인 서울시시설관리공단으로부터 따릉이 대여와 반납 이력 데이터와 대여소별 자전거 거치 현황을 수집하고, 공공데이터 포털을 통해 따릉이 대여소 현황, 서울시 각 지역별 날씨 데이터 등을 수집하여 활용하였다. 따릉이 이용 특성과 대여소의 자전거 거치현황을 이용해 대여소 유형을 군집화하고, 각 대여소 별 자전거 거치율, 피크타임 재고소진 시간 등을 파생 변수로 계량화하여 여타 데이터들과 결합분석을 시도하였다.

따릉이 수요예측 알고리즘은 예측 모형의 최신성을 유지하고 예측의 정확도와 활용성을 높이기 위해 매일 자



(그림 1) 따릉이 실시간 수요예측 모델 구조

(Figure 1) Demand forecasting model structure for relocating the sharing bikes

정 예측 알고리즘을 재학습하여 업데이트하고 이를 하루 동안 실시간 예측에 사용되도록 하였다(Daily Learning - Hourly Prediction). 이를 위해 예측 알고리즘은 최근 90일 동안의 따릉이 이력 데이터를 이용하여 자정을 기준으로 예측 알고리즘을 학습하고 예측 모델은 매 시간 따릉이 수요를 예측하도록 설계하였다. 실제 매시간 수요예측에서는 예측 1시간 전 실제고량을 반영하여 1시간 단위로 실행되도록 설계되었다. 따릉이 대여/반납 이력 데이터와 대여소의 거처 자전거수를 반영하고, 매시간 업데이트되는 기상청의 날씨데이터 결합하여 예측 시점으로부터 3~5시간 후까지 발생가능한 따릉이 수요를 예측하였다. 이렇게 개발된 수요예측 알고리즘의 성능평가를 위해 예측값과 실제값 사이의 편차를 비교하는 RMSE를 이용하였다. 마지막으로 안전재고 수요예측 모델은 서울시설관리공단 따릉이 대여소관리 App에 시범서비스로 구현하였다. 여기서 따릉이 수요예측 알고리즘은 향후 3~5시간 동안에 발생할 대여/반납량을 예측한 후 대여소별 적정재고량과 재배치 필요수량을 산정하고 이를 대여소현황 App에 제공하였다.

2. 선행 연구

근래들어 차량 공유 또는 공공 자전거와 같이 교통수단을 공유하는 모빌리티 서비스가 확산됨에 따라 공유 모빌리티 수요예측 및 재배치 모델에 관한 연구들 또한 점차 증가하고 있다. 먼저 공유 모빌리티 시스템에 대한 Laport et al (2015)의 서베이 연구에서는 대여소 위치(Station location), 배송차량 규모(fleet dimensioning), 대여소의 거처대 수(station inventory)와 같은 장기 계획과 재배치 인센티브(rebalancing incentives)와 같은 정책 수단, 그리고 운송수단 재배치(vehicle repositioning)와 같은 단기적 조치에 따라 모빌리티 재배치의 의사결정을 전략, 전술, 운영 레벨로 정의하였다[7]. 또한 운송수단의 재배치는 대여/반납이 빈번한 낮 시간대의 동적 재배치와 이용이 저조한 밤 시간대의 정적 재배치로 구분하여 제시하고 있다. 대여소 자전거의 적정 재고 수준 관리를 연구한 Raviv와 Kolka(2013)는 이용자의 수요를 충족시키기 위해 사용자 불만족 함수(User Dissatisfaction Function) 모델을 제안하였다[8]. 이는 자전거 대여소 관리에 적합한 재고 모델과 이를 해소하기 위한 수리 모형으로 거처대가 비어있거나 가득 차 있는 시간을 최소화하는 것을 목적으로 하고 있다.

공유자전거의 이용 행태나 서비스 유형에 따른 자전거 수요예측과 대여소 규모에 대한 연구도 이루어지고 있다.

Tang와 Golab(2018)은 공유자전거가 출퇴근 이용자에게 대중교통 체계로의 환승을 위한 라스트마일 또는 퍼스트마일 수단으로 활용되고 있음을 주시하였다[9]. 그들은 대중교통 정거장과 가까운 곳에 있어야 할 자전거의 거처대 규모(Pool size)를 최적화하기 위해 하루 동안의 공유 자전거 도착과 출발 사이의 과도상태 분석과 가장 분주한 기간 동안의 안정상태 분석을 활용하였다. 이렇게 산출된 최적화 솔루션을 몬테카를로 시뮬레이션으로 검증한 결과, 기존 방식으로 산정된 거처대 규모를 39~75% 줄일 수 있음을 보여주었다. 그러나 공유 자전거를 출퇴근용으로 한정하고 자전거 재배치를 이용자가 많지 않은 밤이나 새벽으로 가정한 한적 환경에서의 거처대 규모나 최적 거처량 예측은 실용적이지 않을 수 있다. 실제 공유 자전거의 이용행태를 분석해보면 출퇴근 이외의 용도, 가령 공원 산책이나 운동 목적 등 또한 상당하며, 서울시와 같이 공유자전거를 주간에도 적극적으로 재배치하는 경우도 많기 때문이다.

Huang et al(2019)의 연구에서는 공유 자전거의 대여소 재고량을 최적화하기 위해 자전거 거처대수를 예측하는데 초점을 두었다[10]. 연구자들은 자전거 거처대수 예측을 위해 이형가우시안 이질 포아송(Bimodal Gaussian Inhomogeneous Poisson) 알고리즘을 제안하고 이를 샌프라시스코의 Bay Area 공유자전거 시스템(70개 대여소, 2014년)과 보스턴의 Hubway 시스템(92개 대여소, 2012년)의 데이터셋을 이용하여 실험하였다. 두 시스템은 대여소의 상태를 분단위로 기록하고 있으며 여기에는 가용자전거수(number of available bikes)와 거처대 정보가 포함되어있다. BGIP 알고리즘은 실험 데이터셋의 가용자전거수를 예측하였고 성능은 RMSE를 사용하였다. RMSE 결과는 대체로 2~3 사이로 나타났으며, 또한 알고리즘의 가동 시간도 1초를 넘지 않았으므로 빠른 속도로 예측이 가능함을 보여주었다.

Bao et al(2019)은 무거처대 자전거 공유 시스템(FFBS: Free-Floating Bike System)에 주목하였다[11]. 최근의 이러한 프리플로팅을 모델로 한 라스트마일 모빌리티 서비스가 자전거뿐만 아니라 전동킥보드 등으로도 나타나고 있다. 프리플로팅 수요 예측을 위해 Mobike가 제공하는 Shanghai Trip 데이터와 날씨 데이터, 공기질 데이터를 기반으로 LSTM과 Conv-LSTM을 결합한 딥러닝 프레임워크 구축하였다. 연구자들이 제안한 예측 알고리즘은 기본적으로 15분-20분-30분과 같이 단시간의 수요예측을 하고 있으며, 예측성능은 MSE, MAE, MAPE를 사용하였다. 다양한 데이터 소스의 효과를 보여주기 위해 요인을 Trip으로만 제한한 것과 날씨-공기질 데이터를 결합한 경우

를 비교하기도 하고, 이용자가 많은 시간대와 그렇지 않은 시간대를 견주어 보기도 했다. 또한 제안 프레임워크의 성능을 CNN, ANN, GBRT, ARIMA 등과 같은 다른 알고리즘과 비교하여 성능을 검증하였다.

국내에서도 공유자전거 시스템의 자전거 재배치를 위한 다양한 방법론과 예측 알고리즘에 관한 연구들이 몇몇 진행되었다. 가장 최근의 연구에서는 창원시의 누비자 공공자전거를 사례로 자전거 이용의 쏠림현상으로 인한 불편을 해소하고자 동적 재배치 라우팅 모델을 제안하고 각 터미널별 수요에 맞도록 개선할 수 있는 방안을 제시하고자 하였다[1]. 김진식과 이철웅의 연구에서도 창원시 누비자 공공자전거에 적용하여 운영비용을 최소화할 수 있는 대여소 간 최적 자전거 재고수준을 제안하는 시뮬레이션 모델을 제시하였고, 그 결과 자전거 거치수준을 대여소의 60~70% 정도로 유지할 때 시스템 운영비가 최소화되고 거치율이 약 30% 정도가 되었을 때 자전거를 보충해주는 것이 가장 효과적인 것으로 나타났다[2]. 예를 들면 거치대가 20인 대여소는 초기 14대를 보유하고 7대가 되었을 때 보충하는 것을 의미한다. 다만 본 연구가 창원시 진해항 중심의 30개 대여소를 대상으로 시뮬레이션한 결과로서 자전거의 거치 비율이 대여소 사이즈에 따라 일률적으로 정해짐으로써 대여소마다의 수요/공급의 특성이 제대로 반영되지 못한 한계를 가지고 있다. 이상복 외(2018)의 연구에서는 따릉이가 이미 배송차량과 배치전담인력에 의해 재배치되고 있음을 감안하여 관리구역 설정을 최적화하는 모델을 개발하였다[3]. 이는 운송차량의 이동 범위와 관리 구역의 대여 반납의 균형을 맞출 수 있도록 대여소를 최적으로 균집화하여 각 군집간의 Workload의 차이를 최소화하고자 하였다. 그러나 실제 따릉이를 운영하는 기관의 입장에서는 구단위 또는 지역센터와 보관소의 위치 등에 따라 관리구역이 설정되어 있고, 또한 자전거를 재배치하는 과정에서 여러가지 환경요인, 가령 교통 체증이나 대여소의 지리적 특징 등도 따릉이 재배치에 많은 영향을 미친다.

앞서 두 연구가 재배치 관리구역을 최적화하거나 자전거 거치 수량을 도출하였던데 반해 임희중, 정광현(2019)의 연구에서는 과거 따릉이 사용자 수요 데이터의 분석 결과에 기반하여 따릉이의 수요를 예측하고자 하였다[4]. 이를 위해 따릉이 이용의 수준, 추세, 두개의 계절성의 네 가지 요소를 갖는 Double Seasonal Holt-Winters 기법을 사용하였고 성과지표로는 MSE를 이용하였다. 그리고 마포구와 영등포구의 따릉이 수요 예측을 통해 제시한 예측 모형의 유효성을 검증하고자 하였는데 그 결과로 마포구는 183,358.312, 영등포구는 161,750.415의 평균제곱오차

를 보였다. 이러한 수요예측은 해당 지역에서 하루 동안 대여된 자전거의 수를 모두 합한 일간 수요로 진행되어 지역구 전체의 일간 수요예측 값을 자전거 재배치에 활용하기에는 한계가 있다.

3. 연구 설계 및 기초자료 분석

3.1 연구 설계 및 분석 데이터

본 연구는 서울시 공유자전거 따릉이의 이용 패턴 및 특성을 파악하고, 대여소별로 필요한 따릉이 실시간 수요 예측 모델 개발을 위해 빅데이터 분석과 기계학습 알고리즘을 개발하고자 한다. 이를 위해 따릉이 관리전담 기관인 서울시설관리공단으로부터 2019년 1월부터 12월까지 1년 동안의 따릉이 대여·반납 이력 데이터와 대여소 현황 데이터 등을 제공받았고, 공공데이터 포털을 통해 서울 지역의 날씨 데이터와 휴일이 포함된 달력 데이터 등을 오픈API로 추가 수집하였다[17]. 이러한 데이터를 결합하여 따릉이 이용 행태, 대여소의 특성, 대여·반납 패턴 등을 살펴보았다. 분석에 사용된 수집 데이터는 아래 표 1과 같다.

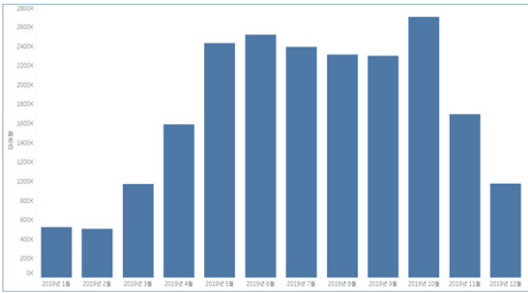
(표 1) 데이터 수집 현황
(Table 1) Collected data for analysis

DB Table	No. of Row	Columns
2019년 대여/반납 이력	20,993,336	대여일자, 대여한 대여소, 반납일자, 반납한 대여소, 사용시간, 사용거리, 이용권 타입
대여소정보	1,540	대여소_구, 대여소ID, 대여소명, 대여소주소, 위도, 경도, 기준시작일자, 거치대수
기상정보	8,760	지역, 측정일시, 기온, 강수량, 풍속, 습도
공휴일	85	공휴일 이름, 날짜

3.2 따릉이 이용 현황 EDA

• 따릉이 월별 이용 현황

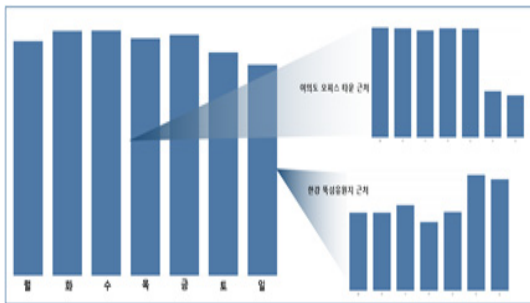
공유자전거 따릉이의 월별 대여 건수는 날씨가 포근해지는 3월부터 급격히 늘기 시작해 5월부터 10월까지 200만 건 이상의 대여 수준을 유지하다 기온이 떨어지는 11월부터 이용 건수가 현저히 줄어든다(그림 2). 이용자가 가장 많았던 10월은 2,701,762건의 자전거 대여가 이루어졌고, 가장 작았던 2월에는 502,687건에 그쳐 그 차이가 약 6배에 달했다.



(그림 2) 월별 따릉이 이용 현황
(Figure 2) Monthly usage of the sharing bikes

• 따릉이 요일별 이용 현황

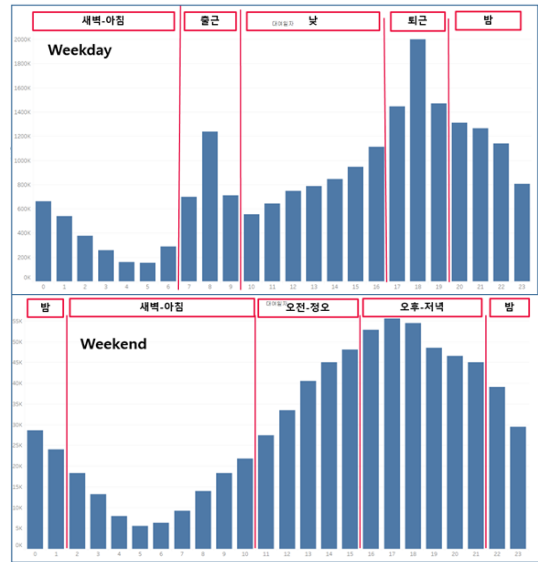
전체 대여소를 합산 평균한 따릉이의 요일별 이용현황은 그림 3처럼 월요일부터 일요일까지 약 260만대에서 300만대로 비교적 균등하게 나타나고 있다. 그러나 개별 대여소의 요일별 현황을 살펴보면 대여소의 위치에 따라 시민들의 이용 패턴이 달라지는 것을 확인 할 수 있다. 그림 3의 오른쪽 위 그래프는 여의도 오피스 타운으로 출퇴근 목적의 주중 이용자가 많았으며, 아래는 한강 뚝섬유원지와 같은 공원 근처 대여소는 주중보다는 주말에 여가를 즐기러 온 이용자가 많은 것을 보여준다.



(그림 3) 요일별 따릉이 이용 현황
(Figure 3) Weekly pattern of usage

• 따릉이 시간대별 이용 현황

공공자전거 대여의 목적이 주중과 주말에 다를 수 있듯, 이용 시간대별 분석에서도 사용 목적이 비교적 분명히 드러난다. 그림 4에서 보여지는 것처럼 주중에는 아침 7시~9시 출근 시간과 오후 5시~7시 퇴근 시간대에 따릉이 이용이 빈번한 반면 주말에는 오후 2시부터 저녁 9시까지 이용건수가 계속 높게 나타나고 있다.

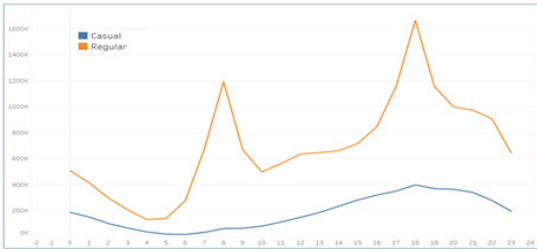


(그림 4) 주중-주말 따릉이 이용 패턴 비교
(Figure 4) Comparison of use between weekday and weekend

• 따릉이 사용권별 이용 현황

따릉이를 대여하기 위해 구매하는 사용권의 종류에 따라서도 이용 목적과 패턴이 다르게 나타난다. 따릉이 사용권은 정기권(Regular)과 일일권(Casual) 두가지 종류가 기본 이용시간에 따라 1시간권(일반권)과 2시간권(프리미엄권)으로 구분되어 있다. 정기권은 7일, 30일, 180일, 365일권을 구매할 수 있으며, 정기권/일일권 모두 기본 시간(1시간 또는 2시간)을 초과할 경우에 5분당 일정한 추가요금이 부과된다.

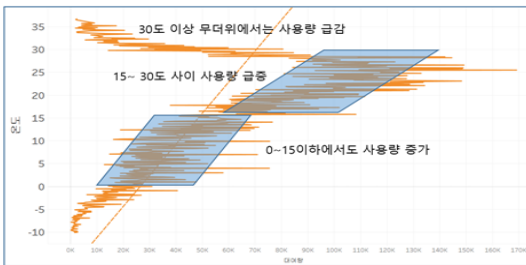
그림 5는 정기권과 일일권 이용자들의 따릉이 대여건수를 하루 24시 기준으로 비교한 것 그래프이다. 전체적으로 정기권 사용이 일일권보다 월등히 높았고, 출퇴근 시간에는 급격히 증가하였다. 이는 따릉이가 출퇴근 시 대중교통의 틈새 이동수단(Last Mile Mobility)으로서 역할을 수행하고 있으며, 일일권은 오후와 저녁 시간대에 근접거리 이동, 관광, 운동, 여가 등 비정기적 활동에 이용된 것으로 보인다. 정기권과 일일권의 월별 대여량도 계절적 요인에 영향을 받는데, 다만 정기권의 경우 5월부터 10월까지의 대여건수가 급증한데 비해 일일권은 상승폭이 대체로 완만하였다.



(그림 5) 따릉이 사용권별 이용 패턴
(Figure 5) Usage pattern by ticket types

• 날씨에 따른 따릉이 이용 현황

야외에서 운행되는 자전거의 특성상 따릉이는 날씨에도 많은 영향을 받는다. 계절적으로 기온이 매우 낮은 겨울철에는 사용량이 극감하고 점차 기온이 올라가는 봄부터 이용자가 늘기 시작해서 5월 초여름부터 10월 가을까지를 성수기라고 볼 수 있다(그림 6). 그러나 기온이 30도에 육박하면 오히려 자전거 이용이 급격히 줄었다.



(그림 6) 기온에 따른 이용 패턴
(Figure 6) Usage pattern by temperature

기상 요인 중 비가 내리는 것도 자전거 이용이 매우 큰 영향을 미친다. 강수량이 4mm 이상일 때 따릉이 대여량은 10분이 1 이하로 급감하였다.



(그림 7) 강수량에 따른 이용 패턴
(Figure 7) Usage pattern by precipitation

그 외 바람의 경우는 따릉이 이용에 크게 영향을 미치지 않는 것으로 분석되었다.

3.3 따릉이 대여소 현황 EDA

• 대여소별 따릉이 이용 현황

따릉이 대여소는 2019년 12월 10일 기준 1,540개소가 서울 전역에 설치되어있다. 각 구별로는 유동인구가 많은 강남구와 송파구가 99 개소로 가장 많고 도심 외곽의 거주 중심 지역인 도봉구는 33개소로 가장 적었다. 따릉이 이용이 가장 많았던 대여소는 광진구의 ‘뚝섬유원지역 1번 출구’ 대여소로 2019년 1년 동안 총 115,898대가 대여 되었고 일 평균 318대에 이른다. 다음은 ‘여의나루역 1번 출구’ 대여소가 114,670대로 비슷한 수준이었다. 상위 10개 대여소들은 일 평균 200대 전후의 따릉이 대여가 이루어지고 출퇴근 시 이용과 저녁 시간의 여가용 이용도 빈번하였다.

이와 달리 이용률이 저조한 하위 10개 대여소는 자전거 대여 패턴도 일정한 모습을 보이지 않으며 하루평균 대여량이 2회에 미치지 못했다. 이렇듯 대여소에 따라 이용량이 폭증하는 곳도 있고 방치되다시피 한 곳도 있는데, 이용률이 높은 곳은 자전거가 없어서 빌리지 못하는 경우도 빈번히 발생한다. 실제, 따릉이 민원 사이트에서는 대여소에 자전거가 부족하다는 민원이 수시로 접수되는데 자전거의 수량이 적지 않고 재배치활동을 함에도 불구하고 자전거 수요가 많은 시점에 일시적으로 자전거 부족한 수요공급 불균형이 발생하기 때문으로 보인다.



(그림 8) 상위 10개 대여소의 따릉이 이용 패턴
(Figure 8) Usage pattern of Top 10 stations

• 대여소별 따릉이 거치 현황

이러한 불균형 상태를 파악하기 위해 대여소에 자전거가 어느정도 거치되어 있는지를 따릉이 이용이 활발한

5월~10월 성수기를 대상으로 계산해보았다. 거치율은 개별 대여소별로 산정되었으며 대여소의 거치대 수에서 시간대별 자전거 재고 수량을 조사하여 하루 24시간을 기준으로 따릉이가 대여소에 거치되어있는 비율을 평균한 값이다. 거치율 분석결과 1537개 대여소 전체 평균은 60.42%, 상위 30개 대여소는 65.48%, 하위 30개소는 51.12%로 나타났다(표 2).

(표 2) 대여소 따릉이 거치율(5~10월사이)
(Table 2) Stock ratio of bike stations from May to Oct

대여소	하루 평균	출퇴근	휴일 오후
1537개	60.42%	58.23%	54.56%
상위 30개	65.48%	62.97%	70.49%
하위 30개	51.12%	50.80%	50.84%

※ 거치율 = $\frac{1}{n} \sum_{i=1}^n \frac{stock_i}{rackTot_i}$ (시간대별 재고량 / 총 거치대 수의 평균)

자전거를 빈번히 이용하는 피크타임을 추려서 살펴본 경우에도 상위 대여소의 거치율은 평일 출퇴근 시간 62.97% 휴일 오후시간에는 70.49%로 높게 나타났다. 반면 하위 30개소는 계속 50% 수준에 머무르는데 이는 하루 24시간 중 거치대의 절반 정도는 자전거가 비어 있었음을 의미한다.

한편 전체 대여소 거치율 평균이 50%이상을 상회하고 있는데 이정도 수준만 유지해도 대여소에 거치된 따릉이가 부족한 경우는 없을 것이다. 그러나 이는 평균적인 거치율이 그렇다는 것이지 실제 자전거 재고가 하나도 없는 경우가 발생하기 때문에 실제 대여를 못하는 상황이 빈번히 나타나고 있다. 그러므로 수급 불균형 상태를 파악하기 위해서는 대여소에 거치된 따릉이가 하나도 없는 무재고 현상을 좀 더 분석해볼 필요가 있다.

• 따릉이 수급 불균형 현황

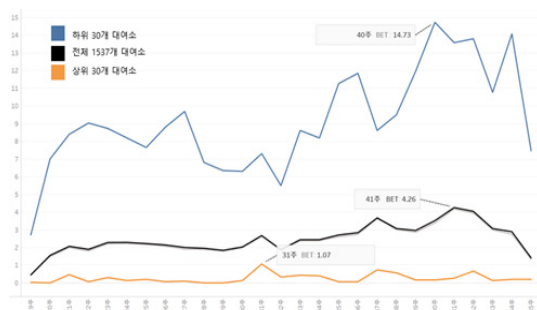
따릉이 대여소의 수급 불균형 상태를 파악하기 위해 대여소에 거치된 따릉이가 전혀 없는 시간, 즉 자전거 재고가 Zero(0) 상태가 1주일 중 몇 시간 정도인지를 “무재고시간(Bike Stock-out Time / 1 Week peak time)”으로 정의하고 산출하였다. BST는 따릉이 성수기인 5월부터 10월 사이의 자전거 이용 피크타임인 평일 출퇴근 오전 7시~9시와 오후 5시~7시(8h * 5 working-day), 그리고 휴일 오후 4시~10시(8h * 2 holiday-day)사이의 총 42시간(1 Week Peak Time) 중 대여소의 자전거가 모두 동나서 모든 거치대가 비어있던 시간을 측정한다(표 3). 대

소 1537개소의 BST 평균은 2.45 시간으로 주간 피크타임 42시간 중 약 3시간 정도는 대여소 거치대가 비어 있었던 것으로 나타났다. 한편 상위 30개소와 하위 30개소를 비교한 결과는 상위 30개소의 경우 BST 평균이 0.24 시간이고 최대값도 1.25 시간이어서 대체로 자전거 수급이 원활하게 이루어지고 있는 것으로 보인다. 반면 하위 30개 대여소는 BST 평균이 8.82로 약 9시간 정도 거치대가 비어있고 가장 높았던 경우(Max)는 거의 30시간 동안 한대도 없었던 것으로 파악되었다.

(표 3) 대여소 따릉이 무재고시간(BST)
(Table 3) Bike stock-out time of stations

BST	Mean	Max	Sd
1537개	2.45h/42h (5.83%)	29.14 (69.38%)	3.12 h (7.43%)
상위 30개	0.24 h (0.57%)	1.25 h (2.98%)	0.27 h (0.64%)
하위 30개	8.82 h (21.00%)	29.14 h (69.38%)	9.05 h (21.55%)

아래 그림 9는 따릉이 이용 성수기인 5월부터 10월 사이의 무재고시간 BST의 시계열 분석 결과로 대여소별 격차가 지속적이고 상당히 큼을 보여주고 있다. 상위 30개 대여소는 BST가 낮게 잘 관리되고 있는 반면, 하위 30개소는 BST가 매우 높고 변동폭도 커서 대여소의 자전거 재배치가 잘 이루어지지 않고 있음을 알 수 있다.



(그림 9) 무재고시간 BST 시계열 분석
(Figure 9) Time series pattern of bike stock-out

결과적으로 대여소의 피크타임 거치율과 BST 분석을 통해 상위 30개 대여소는 피크타임 거치율이 65.48% (평일 출퇴근 시간 62.97% 주말 오후·저녁 70.49%)로 70%에 미치지 못하지만 실제 재고부족시간은 0.24로 매우 낮게 나타나 대여소에 따릉이 재고가 없는 상태로 방치되는 경우는 높지 않음을 알 수 있다. 결국 적정 대수

의 따릉이 자전거가 유지되지 못하고 있는 중하위 대여소의 자전거 재배치와 적정 수준의 거치를 확보가 시급하다고 볼 수 있다.

4. 따릉이 수요예측 모델 개발

따릉이 대여소의 자전거 거치수준을 적정 수준이상으로 유지하여 수급불균형을 예방하기 위해서는 따릉이 이용 패턴과 대여소별 특성을 감안하여 따릉이 재고가 소진되지 않도록 효과적인 재배치 활동이 필요하다. 그러나 2019년 말 기준 서울시에는 1540여개소의 따릉이 대여소에서 2만여대의 자전거가 하루에도 수없이 많은 대여·반납을 통해 공유되고 있으며 대여소의 지리적 위치와 주변 환경, 거치대 수, 고객 특성 등에 따라 따릉이 이용 패턴은 너무나도 복잡하고 다양해서 이를 분류하고 예측하기가 쉽지않다. 게다가 따릉이 이용은 한두시간 이내로 짧은 편이나 주위 환경이나 고객 상황에 따라 이용 행태가 급변하고 자전거 재배치 활동 역시 단시간의 주위 요건을 반영하여 능동적으로 대처해야 한다. 이는 곧 수요공급 예측이 최근 상황을 반영하여 실시간으로 진행되어야만 유효할 수 있음을 의미한다.

그러므로 본 연구에서는 먼저 1500여개의 대여소들을 비슷한 유형과 패턴으로 군집화하고, 다음으로 각 군집에 속한 대여소들을 대상으로 과거의 따릉이 대여량과 반납량을 기계학습한 후 마지막으로 이를 이용해 앞으로 발생할 대여량과 반납량을 예측하는 실시간 따릉이 수요 예측 알고리즘을 개발하고자 한다.

4.1 대여소 유형별 군집화

따릉이 대여소 군집 분석을 위해 2019년 1월부터 10월까지의 대여소 정보와 따릉이 대여반납 이력을 기본 데이터로 활용하여 k-Means 군집분석을 수행하였다. k-Means는 주어진 데이터를 k개의 클러스터로 묶는 알고리즘으로, 각 군집내 데이터들 간의 거리 차이 분산을 최소화하는 방식이다. 먼저 군집분석을 위한 데이터 처리과정에서 데이터 변환, 이산화, 파생변수 생성 등이 이루어졌고, 결과적으로 대여소의 거치대수, 평일(새벽아침, 출근, 낮, 퇴근, 밤)과 휴일(밤, 새벽아침, 오전정오, 오후저녁) 9개의 시간대로 구분한 대여량, 반납량, 사용량, 일일권사용량, 정기권 사용량의 평균치 등 총 29개 항목을 추출하여 k-Mean 변수로 이용하였다.

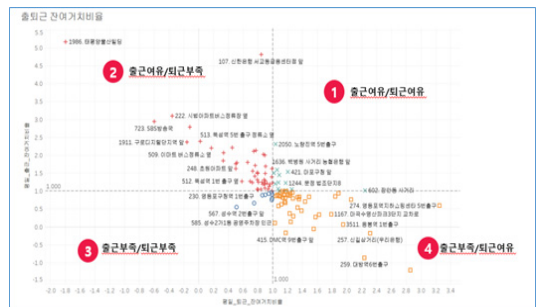
실제 군집화 과정에서는 Python 라이브러리 yellowbrick을 이용하여 대여소 1500여개를 임의의 군집 k로 정의하고

군집 결과의 타당성 평가와 파라미터 조정 등의 작업을 반복하여 최적 군집 k를 선정하였다. 군집 분석 결과의 타당성 평가는 Elbow곡선과 Silhouette score를 이용하였다. 이러한 과정을 통해 Elbow곡선 변동점이 급격하게 작아지는 최적 군집수 k는 8이었고 이때 Silhouette score는 0.409로 나타났다.

(표 4) 대여소 k-Mean 군집분석 결과
(Table 4) K-mean clusters of stations

k=8	대여소 n=1540	평균 거치대 수	시간당 평균 대여/반납 건수
1	497	9.7	0.9 / 0.9
2	167	11.3	3.8 / 3.9
3	272	16.9	1.2 / 1.0
4	6	22.5	11.5 / 12.2
5	69	15.2	6.3 / 6.5
6	402	10.8	2.3 / 2.3
7	2	33.5	16.9 / 17.7
8	93	22.2	3.4 / 3.3

표 4에서 보는 바와 같이 대여소 군집은 대여소의 크기(보유 자전거 수)와 따릉이 대여/반납 빈도가 군집화에 큰 영향을 미치는 것으로 보인다. 그러나 따릉이 대여소 1500여개를 8개의 군집으로만 구분하였을 때 같은 군집에 속한다고 보기에는 이질적 특징을 보이는 대여소가 존재하였다. 가령 동일한 군집내 대여소에서 출퇴근 시간에 상반된 대여/반납 패턴을 보이는 경우도 있었는데 이를 동일 군집으로 묶는 것은 타당하지 않다고 판단되었다. 이를 해소하고자 출퇴근 시간대의 대여/반납 차이 잔여거치비율(대여량-반납량)을 이용해 대여소 특징을 4가지 유형으로 구분한 결과 1) 출근여유/퇴근여유, 2) 출근여유/퇴근부족, 3) 출근부족/퇴근부족, 4) 출근부족/퇴근여유로 나눌 수 있었다(그림 10).



(그림 10) 출퇴근 거치비율에 따른 군집세분화
(Figure 10) Segmentation of clusters

예를 들어 4) ‘출근부족/퇴근여유’에 속한 대방역6번 출구 대여소는 출근시간에는 따릉이 수요가 많아 자전거 거치비율이 낮고 퇴근시에는 적은 수요로 인해 따릉이 거치율이 높게 나타났다. 이렇게 k-Means 군집분석(8개 군집)과 출퇴근 잔여거치비율에 따른 군집 세분화(4개 유형)를 결합하여 총 32개 군집이 도출되었다. 이 중 해당 대여소가 존재하지 않는 3개 군집을 제외한 29개 군집에 1540개의 대여소가 각각 분류되었다(표 5 참조).

(표 5) 대여소 군집분석 세분화 결과
(Table 5) Segmentation result of stations

k=8	대여소 1540	출근 부족/ 퇴근 부족	출근부족/ 퇴근여유	출근 여유/ 퇴근 부족	출근 여유/ 퇴근 여유
1	497	120	223	130	24
2	167	11	77	52	27
3	272	64	11	89	8
4	6	-	1	3	2
5	69	6	24	28	11
6	402	38	199	132	33
7	2	1	-	-	1
8	93	18	31	41	3

군집분석 결과는 대여소의 거치대 수와 따릉이 대여/반납 규모 외에도 출퇴근시 자전거 이용 패턴에 따라 서로 세분화 될 수 있음을 보여준다. 군집 7의 경우 잔여거치비율에 따라 대여소의 자전거 거치대 수가 많고 따릉이 대여와 반납도 매우 빈번하지만 출퇴근 상황에 따른 다른 유형의 군집으로 나눌 수 있음을 확인하였다.

4.2 수요예측 알고리즘 개발

약 2만여대의 자전거가 1540여개 대여소에서 끊임없이 대여/반납을 통해 공유되고 있으며 대여소 위치와 주변환경, 거치대 수, 고객 유형에 따라 따릉이 이용 패턴은 너무나도 복잡하고 다양하다. 게다가 따릉이 수요 예측은 개별 대여소를 대상으로 실시간으로 이루어져야 재배치가 가능하다. 이러한 특성을 반영할 수 있는 실시간 따릉이 수요예측 알고리즘은 대량의 트랜잭션을 단시간에 처리하면서도 즉각적으로 예측 결과를 도출할 수 있어야 하고, 원인에 대한 이해보다는 예측결과의 정확도가 중요하다.

이에 본 연구진은 알고리즘 개발 초기 단계에서 Random Forest, SVM, Gradient Boosting, LSTM 등을 시험한 결과 범주형 변수(요일, 월, 시간대 등)와 이진변수(주말 여부, 휴일 여부), 연속형 변수(기온, 풍속, 강수량)등

을 함께 다루는데 용이한 Random Forest를 사용하기로 하였다. 알고리즘 개발에 사용된 데이터는 군집분석 결과를 반영한 대여소 정보와 대여소별 따릉이 대여/반납 이력, 이를 가공한 시간당 대여량/반납량, 주중/주말/공유일 정보와 시간대별 기상정보(기온, 강수량, 풍속, 습도) 등이 포함되었다.

수요예측 학습 데이터 주기에 관해서도 여러번의 테스트를 거쳐 90일 데이터를 이용한 일일 학습(Daily Learning) 방식을 선정하였다. 이는 예를 들면 10월 1일에 적용할 수요예측 알고리즘은 최근 90일(7월 1일~ 9월 30일) 데이터로 학습한 결과이다(표 6 참조). 너무 오랜 기간 학습할 경우 알고리즘 학습 시간이 길어지고, 예측 주기가 길수록 최신 데이터를 반영하지 못해 정확도가 떨어지는 문제가 발생하였기 때문이다.

(표 6) 알고리즘 학습 주기
(Table 6) Training and test data day by day.

수요예측 D-Day	학습데이터 D-1 ~ D-90
10월 1일	9월 30일~7월 1일
10월 2일	10월 1일~7월 2일
10월 3일	10월 2일~7월 3일
...	...
10월 31일	10월 30일~ 8월 1일

마지막으로 따릉이 수요예측 알고리즘은 모든 대여소에 대하여 대여량과 반납량을 1시간 단위로 예측하도록 개발되었다. 대여소의 따릉이 거치수준을 적정 수준 이상으로 유지하면서 수급불균형을 방지하기 위해서는 개별 대여소의 따릉이 재배치 활동이 가능하도록 수요예측 모델이 작동해야한다. 그러므로 수요예측 알고리즘은 최근의 재고상황과 대여반납 현황, 기상정보 등을 반영하여 실시간으로 진행되어야하는데 이러한 정보들은 대체로 시간단위로 업데이트되고 있다. 예를 들어 기상정보는 매 시간 예보가 이루어지며 3~5시간 후의 기상예측 정보를 제공한다. 이러한 요소들을 반영하여 따릉이 수요예측 알고리즘은 개별 대여소의 수요예측을 매시각 실행(Hourly Prediction)하고 향후 3~5시간의 대여와 반납 대수를 예측한다. 본 연구에서는 Python의 sklearn라이브러리의 Random Forest 알고리즘을 수요예측 모델에 활용하였다.

알고리즘 예측 성능은 2019년 10월 한달 동안의 예측값과 실제값을 비교하여 평가하였다. 따릉이 데이터는 2019년 1년 동안의 데이터가 모두 수집되었으나 10월 한달 만을 알고리즘 검증에 사용하였다. 이는 10월이 이용

피크 기간이고, 11월부터는 추위로 따릉이 사용량이 급감하기 때문이다. 앞서 설명한 바와 같이 따릉이 수요예측 알고리즘은 모든 대여소에 대하여 대여량과 반납량을 1시간 단위로 예측하도록 개발되었다. 그러므로 한시간 단위의 대여·반납 예측값이 실제 동일시간에 발생한 대여·반납량과 어느정도 차이를 보이는지를 비교함으로써 수요예측 알고리즘의 성능을 확인할 수 있다. 이러한 차이를 비교하는 지표로 본 연구에서는 RMSE(Root Mean Square Error)와 MAE(Mean Absolute Error)를 사용하였다. 두 지표는 예측값과 실제값의 차이를 비교하여 예측 알고리즘의 성능을 검증하는 지표이므로 기본적으로 결과값이 낮을수록 예측값과 실제값에 근접함을 의미한다.

$$RMSE(\text{평균제곱근오차}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{실제값} - \text{예측값})^2}$$

2019년 10월 한달 동안의 따릉이 수요예측 성능평가 RMSE 결과는 전체 1540개 대여소의 대여예측 RMSE는 2.336이었고, 반납예측 RMSE는 2.343으로 실제 대여/반납량과 예측의 편차가 약 2~3대로 나타났다. 이는 Random Forest 알고리즘을 이용해 1540개의 대여소의 따릉이 대여와 반납량을 기계학습 후 1시간 단위로 예측하고 이를 실제 발생한 값과 비교하는 RMSE와 MAE를 산출하여 평균한 값이다. 그러나 각 군집에 따라 보다 세부적인 예측성능을 살펴보면, 군집별 RMSE의 편차가 상당하였다.

먼저 대여예측 알고리즘 성능평가 결과대부분의 군집에서 RMSE가 1~3 사이의 값을 보이는데 반해, 4와 7번 군집에서는 편차가 매우 크게 나타나고 있다(표 7). 이

(표 7) 대여예측 성능평가
(Table 7) Rent forecasting performance

k	10월평균 RMSE	출근부족/퇴근부족	출근부족/퇴근여유	출근여유/퇴근부족	출근여유/퇴근여유
1	497	1.18 0.83	1.27 0.91	1.46 0.94	1.27 0.90
2	167	2.89 1.92	2.9 2.05	3.17 2.16	2.99 2.14
3	272	1.42 0.95	1.54 1.08	1.74 1.10	1.44 1.03
4	6	-	8.11 4.81	8.11 5.14	6.23 4.45
5	69	4.15 2.89	4.29 2.93	5.58 3.56	4.70 3.13
6	402	1.96 1.4	1.99 1.46	2.36 1.56	2.03 1.48
7	2	14.52 8.48	-	-	14.42 8.20
8	93	3.01 2.03	2.80 2.00	3.66 2.25	3.08 2.29

는 해당 군집의 대여소가 여의나루역 1번출구, 뚝섬유원지역1번출구, 마포구민체육센터 등과 같이 거치대수도 많고 따릉이 수요도 매우 큰 곳으로 대여와 반납도 빈번하기 때문에 예측값과 실제의 편차도 큰 것으로 보인다.

따릉이 반납예측 성능평가표(표 8)에서도 대여예측 결과와 같이 거대 대여소가 포함된 군집에서의 RMSE, 즉 예측과 실제값의 편차가 유사하게 높게 나타났다.

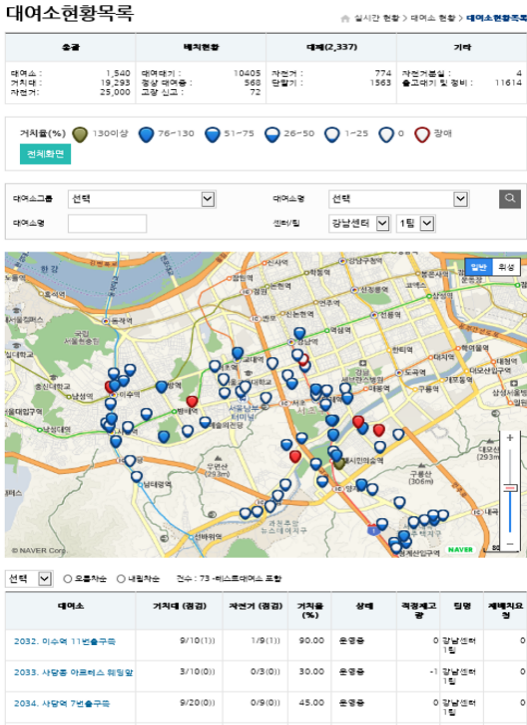
(표 8) 반납예측 성능평가
(Table 8) Return forecasting performance

k	10월평균 RMSE	출근부족/퇴근부족	출근부족/퇴근여유	출근여유/퇴근부족	출근여유/퇴근여유
1	497	1.03 0.72	1.24 0.88	1.45 0.94	1.27 0.90
2	167	2.71 1.95	2.92 2.02	3.32 2.22	2.98 2.13
3	272	1.11 0.76	1.41 0.99	1.66 1.05	1.38 0.99
4	6	-	8.23 4.83	9.35 5.94	6.15 4.38
5	69	3.97 2.76	4.46 2.92	5.44 3.58	4.99 3.29
6	402	1.84 1.33	1.95 1.40	2.48 1.59	2.06 1.48
7	2	16.51 9.02	-	-	19.54 9.15
8	93	2.39 1.67	2.56 1.82	3.82 2.32	3.05 2.22

한편 수요예측 알고리즘을 안전재고 예측 서비스에 적용하기 위한 품질 요소로서 학습과 예측에 소요되는 가동시간(Running Time)은 Daily 알고리즘 학습은 30분 이내, Hourly 예측에는 10분 이내 구현이 가능하도록 데이터 수집부터 전처리, 학습과 예측 모듈을 분리하여 함수화하고 Python 코드를 통해 최적화하였다.

4.3 수요예측 알고리즘 적용

본 연구팀은 앞서 개발된 따릉이 수요예측 알고리즘을 이용하여 대여소별 따릉이 안전재고량을 산출하여 서울시설공단 따릉이 대여소현황 App에 재배치필요수량을 제공하는 시범서비스를 구축하였다(그림 11). 따릉이를 전담관리하는 서울시설공단은 재배치 전담요원의 업무를 지원하기 위해 대여소현황 App을 통해 전체 대여소 현황과 거치대수, 따릉이 자전거 대수와 함께 배치 현황 등을 보여주고 있으며, 특히 개별 대여소의 안전재고 확보를 위해 각 대여소의 자전거 거치대수와 거치율, 상태, 적정재고량, 재배치필요수량을 서비스하고 있다.



(그림 11) 서울시설공단 따릉이 재배치App (Figure 11) Bike relocation guide map of web-application

안전재고 수요예측 시스템은 따릉이 수요예측 모델링 서버에서 과거데이터의 학습을 통해 예측 알고리즘이 생성되고 이를 이용한 예측값이 서울시설공단에 제공되는 방식이다. 학습에 필요한 데이터는 서울시설공단 따릉이 시스템으로부터 대여이력 데이터를 수집하고, 그 외에도 따릉이 대여소 현황, 공휴일 정보, 외부 날씨 데이터는 공공데이터포털의 오픈 API를 이용하여 수집하였다. 수요예측 알고리즘은 일단위로 재학습되어 최신 경향을 반영하도록 하였는데, 매일 자정 수집된 데이터의 전처리부터 변수생성, 수요예측 알고리즘 학습이 약 30여분에 걸쳐 진행될 수 있도록 최적화되어있다. 학습된 알고리즘은 매 시간 외부 데이터인 날씨정보를 반영하여 예측 시점으로부터 3~6시간 후의 단기 대여/반납 수요를 예측하고 이를 대여소별 안전재고 적치율을 60%로 설정하여 이를 충족하는 적정재고량과 예측 시점의 재고현황을 비교하여 추가로 재배치가 필요한 따릉이 대수를 산출하였다. 이러한 과정을 통해 계산된 결과값이 따릉이 대여소현황 App의 적정재고량과 재배치필요수량으로 제공되도록 시스템을 개선하였다.

5. 결 론

서울 도심 내 교통량 감축과 탄소배출을 줄이기 위해 2015년 10월 도입된 공유자전거 따릉이는 서울시의 가장 성공적인 공공 정책으로 자리매김하였다. 그러나 공공자전거의 규모가 커지고 이용이 급속도로 증가됨에 따라 서비스 이용자의 다양한 민원 대응과 공유자전거의 효율적 운용 관리 시스템의 필요성도 부각되었다. 특히 대여소의 자전거 부족으로 대여를 못한 민원이 지속적으로 발생하고 있는데, 따릉이는 어느 대여소에서나 빌리고 반납할 수 있는 One-Way 방식의 공유자전거시스템이기 때문에 사용자의 자전거 대여/반납량에 따라 대여소에 자전거가 부족하거나 또는 거차 물량이 넘칠 수도 있다.

이러한 자전거 수요/공급 불일치를 해소하기위해 서울시는 공유자전거 전담 배송직원을 통해 대여소간 자전거를 인위적으로 재배치하고 있으며, 서울시설공단의 대여소현황 App에서 제공되는 대여소별 실시간 자전거 거차 현황을 보고 특정 대여소의 자전거가 일정수준 이상으로 많거나 적을 경우 해당 대여소 자전거를 회수 혹은 재배치한다. 그럼에도 불구하고 따릉이 이용자가 급증하면서 대여소별로 자전거 부족, 과잉 등 이용률 편차가 발생하고 있는데, 이를 해소할 수 있는 효율적 자전거 재배치 방안으로서 본 연구에서는 따릉이의 대여·반납에 따른 수요를 예측하고 이에 따라 추가적인 자전거 재배치 수량을 조절하여 안전재고량을 확보할 수 있는 따릉이 안전재고 수요예측 모델을 개발하였다. 이와 관련한 데이터는 서울시설공단으로부터 2019년 따릉이 대여·반납 이력 데이터를 지원받았으며, 서울의 기상데이터와 공휴일 정보, 따릉이 대여소정보는 공공빅데이터 API를 통해 수집 분석하였다.

따릉이 이용 패턴을 분석한 결과 날씨가 포근해지는 3월부터 이용이 급격히 증가하여 5월에서 10월까지 월 200만건 이상의 대여가 지속되는 성수기를 보였으며, 이용자가 가장 많았던 10월에는 270만 건의 대여가 이루어졌다. 이는 야외에서 운행되는 자전거의 특성상 날씨에 영향을 많이 받기 때문으로 자전거를 타기 온화한 기온에는 이용이 늘고 기온이 낮은 겨울철에는 사용량이 극감하는 일반적인 패턴을 보여주고 있는데, 다만 한여름의 기온이 30도를 넘어서면 자전거 이용도 급격히 감소하였다. 비가 내리는 경우도 자전거 이용에 많은 영향을 미쳐서 강우량이 4mm 이상이면 따릉이 이용은 평소보다 10분이 1로 급감하였다. 다만, 풍속은 자전거 이용에 크게 영향을 미치지 않았다.

따릉이의 요일별 이용 패턴은 대여소의 위치에 따라 다소 다르게 나타나는데 출퇴근 이용이 많은 곳은 주중 이용률이 높았고, 공원 유원지의 경우는 주말과 휴일에 이용자가 급증하였다. 이러한 요일별 패턴은 자전거 이용 시각에도 영향을 미쳐서 주중 이용은 출퇴근 시간대에 집중된 반면 주말·휴일 이용은 정오를 지나 오후부터 저녁 시간대에 길고 고르게 대여가 이루어졌다. 따릉이 이용권은 정기권과 일일권으로 나뉘는데 상당수의 이용자들이 정기권을 구매하여 출퇴근 시 틈새이동수단으로 따릉이를 이용하고 있으며 일일권 사용자들은 주로 저녁 시간대에 근접거리 이동, 관광, 운동, 여가 등 비정기적 활동을 위해 따릉이를 대여하는 것으로 나타났다.

따릉이의 이용자가 급증함에 따라 대여소도 2019년 12월 기준 1,540개소에서 2023년 10월 기준 4,700여개로 증가하였다. 대여소 중 따릉이 이용이 가장 많았던 곳은 광진구의 ‘뚝섬유원지역 1번출구’와 영등포구의 ‘여의나루역 1번출구’ 대여소로 이들을 포함한 상위 10개 대여소들은 출퇴근 시간대 이용과 저녁 시간의 여가용 모두 활성화 되고 있었다. 반면 하위 10개 대여소는 연간 이용대수가 1천대에도 미치지 못하고 있어 필요성에 의문이 제시되기도 한다. 이렇듯 다양한 요인에 의해 영향을 받는 따릉이는 수요와 공급간의 불일치가 발생하여 이용고객의 불편을 초래하는 경우가 빈번하게 발생하였다.

따릉이 수요공급 불균형을 파악하기 위해 성수기(5월~10월) 피크타임(주중 출퇴근, 주말 오후·저녁)을 중심으로 대여소의 자전거거치율과 자전거 무재고시간(BST)을 계량화하여 분석하였다. 그 결과 1537개 대여소의 자전거 거치율은 평일 출퇴근 시간 58.23%, 휴일 오후·저녁 시간 54.56%에 그치는데 반해, 상위 30개 소는 평일 출퇴근 시간에도 62.97% 휴일 오후·저녁 시간에는 70.49%로 높게 나타났으며, 하위 30개 대여소는 50% 수준에 머물렀다. 자전거 무재고시간 BST도 전체 평균은 2.45시간이었으나, 상위 30개 대여소는 0.24로 피크타임에도 자전거 수요공급이 원활한 반면 하위 30개 대여소는 8.82로 약 9시간 정도 비어있던 것으로 나타났다.

빅데이터 분석 결과를 기반으로 대여소 안전재고 확보를 위한 따릉이 실시간 수요예측 알고리즘을 기계학습으로 개발하였다. 이를 위해 먼저 1540여개소의 따릉이 대여소를 비슷한 유형과 패턴으로 묶는 군집분석을 실시하였다. 군집분석에는 대여소의 거치대수, 평일과 휴일의 따릉이 대여량, 반납량, 사용량, 일일권사용량, 정기권사용량의 평균치 등 총 29개 항목을 추출하여 k-mean 알고리즘을 이용해 8개의 군집을 도출하였다. 여기에 출퇴근

시 잔여거치비율을 결합하여 총 29개의 군집으로 세분화하였다. 다음으로 Random Forest 기계학습 모델을 이용해 각 군집에 속한 대여소들을 대상으로 과거의 따릉이 대여량과 반납량을 학습하고, 향후 발생할 대여량과 반납량을 예측하는 따릉이 수요예측 알고리즘을 개발하였다. 예측 모델은 매일 자정을 기해 최근 90일 동안의 따릉이 대여와 반납 패턴을 학습한 후 예측 알고리즘을 생성하고, 예측 당일 D-Day에는 이를 이용해 매시간 서울시 전체 대여소의 따릉이 대여/반납량을 예측하였다.

이렇게 개발된 수요예측 알고리즘의 성능평가를 위해 예측값과 실제값 사이의 편차를 비교하는 RMSE를 이용하였다. 그 결과 전체 1540개 대여소의 대여예측 RMSE는 2.336 반납예측은 2.343으로 실제 대여/반납량과의 편차는 약 2~3대로 나타났다. 그러나 군집별 RMSE는 군집의 특성에 따라 편차가 존재하며, 특히 거치대수도 많고 수요도 많은 여의나루역 1번출구, 뚝섬유원지역1번출구, 마포구민체육센터 등은 대여/반납의 절대량이 매우 크기 때문에 예측과 실제값 사이의 편차도 높게 나타났다. 마지막으로 안전재고 수요예측 시스템은 서울시시설관리공단 따릉이 대여소현황 App에 시범서비스로 구현되었다. 따릉이 수요예측 알고리즘은 향후 3~5시간 동안에 발생할 대여/반납량을 예측한 후 대여소별 적정재고량과 재배치필요수량을 산정하고 이를 대여소현황 App에 제공하였다. 여기에는 전체 대여소 현황과 거치대수, 따릉이 자전거 대수와 함께 개별 대여소의 자전거 거치대수와 거치율, 상태, 적정재고량, 재배치필요수량이 표시되고 재배치 전담요원들이 이를 참고하여 따릉이 재배치 활동을 수행하고 있다. 이러한 안전재고 수요예측을 통해 자전거 부족으로 인한 시민들의 불편 해소에 기여할 수 있을 것으로 기대된다.

다만, 본 연구가 따릉이의 안전재고량을 확보하긴 위한 수요예측 알고리즘을 개발하고 시범서비스를 구현하였음에도 몇 가지 제약과 한계점이 존재한다. 먼저 수요예측 시스템은 따릉이 재배치를 실무적으로 지원하기 위해 매 시각 개별 대여소의 안전재고량을 산출하여 서비스되고 있으나, 실제 재배치 활동은 전담요원의 지정된 관할구역 내에서만 가능하며 교통상황이나 이동 경로를 고려하지는 못하고 있다. 향후 연구에서는 이러한 실무 효율성을 높일 수 있도록 재배치 우선순위와 최적 경로 추천 등의 추가 서비스가 검토되어야 할 것으로 보인다. 또한 본 연구에서는 대여소의 특성을 파악하고 신규 대여소 발굴에 참고할 수 있는 참조모델로서의 역할을 기대하며 군집분석을 실시하였다. 나아가 수요예측 알고리

즘 학습을 군집별로 실시하여 1540개 대여소의 학습시간을 10여분 내로 수행할 수 있었다. 그러나 군집별 특성 특히 규모가 크고 이용이 빈번한 대여소는 예측과 실제 값 사이의 편차가 적지 않았는데, 향후 연구에서는 개별 대여소를 대상으로 수요예측 알고리즘을 학습하고 예측하여 그 성능을 확인할 필요가 있다.

참고문헌(Reference)

- [1] Hayeong Jeong, Taihun Lee, Jae Kwon Kim, Oe Jung Kim, Chan Young Kim, "A dynamic relocation routing model for mitigating the hot-spot congestion of public bike: A case of Nubija in Changwon city," JKDIS, vol.32, no.1, pp. 75-95, 2021.
<http://dx.doi.org/10.7465/jkdi.2021.32.1.75>
- [2] Jin-Sik Kim, Chul Ung Lee, "Optimal Inventory Level of Bicycle Sharing Service Considering Operation Costs," JKSCI, Vol. 20, No.1, pp. 163-173, 2015.
UCI : G704-001619.2015.20.1.012
- [3] Sangbok Lee, Heejong Lim, Kwanghun Chung, "Optimal Clustering of Stations for the Bike Sharing System in Seoul," KMSR, vol.35, no.1 pp.55-67, 2018.
<https://doi.org/10.7737/KMSR.2018.35.1.055>
- [4] Heejong Lim, Kwanghun Chung, "Development of Demand Forecasting Model for Seoul Shared Bicycle," The Journal of the Korea Contents Association, Vol.19, No.1 pp.132-140, 2019.
<https://doi.org/10.5392/JKCA.2019.19.01.132>
- [5] Jae Min Jang, Tae-Hyoung Tommy Gim, M Lee, "A Study on the Seoul Public Bikes Use Characteristics - A Case of the Districts of Yeouido and Sangam," Seoul Studies, vol.17, no.4 pp.77-91, 2016.
<https://doi.org/10.23129/seouls.17.4.201612.77>
- [6] Inwoong Jung, Hyunseop Uhm, Young Hoon Lee, "Demand Driven Reallocation in Bike Sharing System," JKORMS, vol.43, no.4 pp.17-31, 2018.
<https://doi.org/10.7737/JKORMS.2018.43.4.017>
- [7] Laporte, G., Meunier, F. & Wolfler Calvo, R., "Shared mobility systems", 4OR-Q J Oper Res, Vol. 13, pp. 341-360, 2015.
<https://doi.org/10.1007/s10288-015-0301-z>
- [8] Tal Raviv & Ofer Kolka, "Optimal inventory management of a bikesharing station", IIE Transactions, Vol. 45, No.10, 1077-1093, 2013.
<https://doi.org/10.1080/0740817X.2013.770186>
- [9] Guoming Tang, Lukasz Golagb, "Bikeshare Pool Sizing for Bike-and-Ride Multimodal Transit," IEEE Transactions on Intelligent Transportation Systems, Vol.19, No.7, pp. 2279-2289, 2018.
<https://doi.org/10.1109/TITS.2018.2814583>
- [10] Feihu Huang, Shojie Qiao, Jian Peng, Big Guo, "A Bimodal Gaussian Inhomogeneous Poisson Algorithm for Bike Number Prediction in a Bike-Sharing System," IEEE Transactions on Intelligent Transportation Systems, Vol.20, No.8, pp. 2848-2857, 2019.
<https://doi.org/10.1109/TITS.2018.2868483>
- [11] Jie Bao, Hao Yu, Jiaming Wu, "Short-term FFBS demand prediction with multi-source data in a hybrid deep learning framework," IET Intelligent Transport Systems, Vol.13, No.9, pp. 1340-1347, 2019.
<https://doi.org/10.1049/iet-its.2019.0008>
- [12] Supriyo Ghosh, Pradeep Varakantham, Yossiri Adulyasak, Patrick Jaillet, "Dynamic Repositioning to Reduce Lost Demand in Bike Sharing Systems," Journal of Artificial Intelligence Research, Vol.58, pp. 387-430, 2017.
<https://doi.org/10.1613/jair.5308>
- [13] Federico Chiariotti, Chiara Pielli, Andrea Zanella, Michele Zorzi, "A Dynamic Approach to Rebalancing Bike-Sharing Systems," Sensors, Vol.18, No.2, 2018.
<https://doi.org/10.3390/s18020512>
- [14] KBS, "따릉이 대여 3천만건 넘어... 절반은 출퇴근때 이용", 2019.
<http://news.kbs.co.kr/news/view.do?ncd=4315859&ref=A>
- [15] TBS뉴스, "서울 시민의 발 '따릉이' 어떻게 관리되고 있을까?" 2023.
http://tbs.seoul.kr/news/newsView.do?typ_800=7&idx_800=3507701&seq_800=20500380
- [16] 한겨레, "서울시민이 뽑은 1위 정책은 따릉이," 2019. www.hani.co.kr/arti/area/capital/920256.html
- [17] 서울열린데이터광장, <http://data.seoul.go.kr>

● 저 자 소 개 ●



김 유 신(Yoosin Kim)

2000년 국민대학교 정보관리학부 졸업(학사)

2009년 국민대학교 비즈니스IT전문대학원 졸업(석사)

2013년 국민대학교 비즈니스IT전문대학원 졸업(박사) & 텍사스 주립대 Post-Doctoral Research Fellow

2020년~현재 (주)에어딥 Founder/CEO

관심분야 : AI, IoT, 빅데이터

E-mail : yoosin@airdeep.co.kr