



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

마이크로 모빌리티 수요 예측을 위한
양상블 모델

고지영

제주대학교 대학원

컴퓨터공학과

2024년 2월



마이크로 모빌리티 수요 예측을 위한 양상블 모델

이 논문을 공학 석사 학위논문으로 제출함

고 지 영

제주대학교 대학원

컴퓨터공학과

지도교수 변 영 철

고지영의 공학 석사 학위논문을 인준함

2023년 11월

심사위원장 이 상 권 이

위 원 김 호 영

위 원 신 영 철

목차

I. 서론	1
1.1 연구 동기 및 목적	1
1.1.1 연구 동기	1
1.1.2 연구 목적	3
1.1.3 연구의 기여점	3
1.2 논문 구성	4
II. 이론적 배경	5
2.1 배송 서비스 전략	5
2.2 마이크로 모빌리티	6
2.3 전동킥보드 공유서비스	6
2.4 시계열 데이터 분석	7
2.5 머신러닝	7
2.6 머신러닝 모델	8
2.6.1 Random Forest Regressor	9
2.6.2 Extra Trees Regressor	9
2.6.3 LightGBM Regressor	9
2.6.4 CatBoost Regressor	10
2.7 관련 연구	10
III. 제안하는 방법	12
3.1 실험 데이터	12
3.2 시스템 구성도	14
3.3 데이터 전처리	16
3.3 양상블 기법	17
3.3.1 보팅(Voting)	18

IV. 실험 결과	19
4.1 시스템 환경	19
4.2 성능평가 지표	19
4.2.1 평균제곱오차	20
4.2.2 평균절대오차	20
4.2.3 평균 제공근 오차	20
4.2.4 결정계수	20
4.3 데이터 분석 및 시각화	21
4.4 학습 방법	24
4.5 학습 및 테스트 결과	24
4.6 특징 중요도	25
4.7 활용 사례	26
4.7.1 미래의 대여 수 예측	26
4.7.2 기상청 단기예보	27
4.7.3 기상청 중기예보	29
4.7.4 예측 결과	31
4.7.5 서비스 적용	32
V. 결론 및 토의	33
참고문헌	35

그림 목차

[그림 1-1] 개인형 모빌리티 시장 규모, 점유율 및 추세 분석	2
[그림 3-1] 마이크로 모빌리티 대여 데이터 예시	13
[그림 3-2] 전체 흐름도	15
[그림 3-3] Voting Regressor	18
[그림 4-1] 전체기간 동안 대여 수	21
[그림 4-2] 월별 대여 수의 합	22
[그림 4-3] 요일별/주말 여부에 따른 대여 수의 평균	22
[그림 4-4] 전체기간에 대한 히트맵	23
[그림 4-5] 훈련/테스트 방법	24
[그림 4-6] 특징 중요도	26
[그림 4-7] 미래 시점의 대여 수 예측 흐름도	27
[그림 4-8] 단기예보 크롤링 예시	28
[그림 4-9] 중기예보 크롤링 예시	30
[그림 4-10] 완성된 테스트 데이터 예시	30
[그림 4-11] 미래 대여 수 예측 결과	31
[그림 4-12] 실제 서비스 화면	32

표 목차

<표 3-1> 마이크로 모빌리티 유형	12
<표 3-2> 마이크로 모빌리티 데이터의 속성	13
<표 3-3> 날씨 데이터	14
<표 3-4> 강수량의 범주화	16
<표 3-5> 완성된 데이터 셋	17
<표 4-1> 시스템 환경	19
<표 4-2> 여름 / 겨울에 따른 상관관계 표	24
<표 4-3> 머신러닝 결과	25
<표 4-4> 단기예보가 제공하는 날씨정보	28
<표 4-5> 중기예보 지역 코드	29
<표 4-6> 미래 대여 수 예측 성능평가	31

마이크로 모빌리티 수요 예측을 위한 앙상블 모델

고 지 영

제주대학교 대학원 컴퓨터공학과

요약

최근 유통시장에서 기업들은 경쟁에서 우위를 점하기 위해 소비자의 만족도를 높일 수 있는 배송시스템에 대하여 상당한 관심을 보이고 있다. 단거리를 빠르게 이동할 수 있는 교통수단으로 공유 모빌리티가 주목받기 시작하면서 다양한 대여 업체들이 생겨났다. 이렇게 생겨난 업체들은 소비자가 어디에서나 쉽게 대여할 수 있도록 공유 모빌리티를 여러 장소에 분산시켜 놓는다. 즉, 여러 장소마다 공유 모빌리티가 미리 적절하게 배치되어야 한다. 업체들은 어느 시기에 어떤 장소에서 대여 수가 어느 정도 발생할지 미리 알지 못하면 공유 모빌리티 준비에 빠르게 대처하기가 어렵다는 문제가 있다. 본 연구를 통해 공유 전동 모빌리티 대여 수에 미치는 요소를 분석하고 대여 수를 예측함으로써 전동 모빌리티를 분산시키고 고객이 제때 서비스를 받을 수 있도록 한다.

본 연구에서는 머신러닝 모델을 통하여 전동 모빌리티의 수요 예측을 실험했다. Random Forest, ExtraTrees, CatBoost, LightGBM 모델을 사용하여 전동 모빌리티의 수요량을 예측하고 모델의 성능을 비교하였다. 또한, 앙상블 기법의 하나인 Voting 방법을 이용하여 오차를 줄이는 방법을 연구하였다. 수요에 영향을 미치는 여러 특징의 상관관계를 분석한 결과 습도, 강수량, 일사량 요소가 수요에 상관관계가 큰 것을 보여주었다. 성능을 평가해 봤을 때, Voting Regressor를 사용했을 때 R2 score가 0.7629로 단일 모델보다 나은 성능을 보였다. 본 연구 결과를 활용함으로써 사전에 전동 모빌리티를 적절하게 배치할 수 있고 사용자의 만족도가 올라갈 것으로 기대한다. 업체들은 장비 관리나 인적 관리를 효율적으로 하여 더 나은 공유 전동 모빌리티 대여 서비스를 제공할 수 있을 것으로 기대된다.

I. 서론

1.1 연구 동기 및 목적

1.1.1 연구 동기

코로나19로 인해 온라인 쇼핑에 대한 선호도가 급격히 높아졌으며, 배송 업계는 빠르고 안정적인 배송시스템을 구축하기 위해 노력하고 있다[1]. 예를 들어 쿠팡은 '로켓배송'을 넘어 보다 빠르고 효율적인 배송서비스를 제공하기 위해 자체 물류기지 구축에 많은 투자를 하고 있다. 쿠팡은 익일 배송을 위해 자체 물류센터와 자체 유통 시스템을 구축했고, 로켓배송 서비스가 소비자를 만족시키며 쿠팡은 급속도로 성장하기 시작했다. 로켓와우 서비스는 출시 2개월 만에 회원 수 100만 명을 돌파했으며, 빠른 배송으로 인한 높은 소비자 만족도 향상으로 플랫폼 간 속도 전쟁은 더욱 치열해질 것으로 예상된다[2]. 쿠팡은 인공지능 알고리즘을 활용해 상품 재고, 상품 위치, 최적의 배송 경로 등을 예측해 소비자 수요에 빠르게 대비한다. 쿠팡에서는 물류센터 공간을 효율적으로 관리 할 뿐만 아니라 빠른 배송을 위해 인공지능을 활용하여 재고와 고객 수요를 분석한다. 쿠팡의 진열대는 여러 종류의 상품이 무질서하게 진열되어 있는데 이는 빅데이터 분석으로 얻은 결과이다. 쿠팡의 인공지능 알고리즘은 상품의 판매량, 판매 시기, 같이 주문할 확률이 높은 상품군 등을 고려해 상품의 위치를 결정한다[3]. 이는 도서 지역의 소비자도 빠른 배송을 받을 수 있게 하고, 소비자의 만족도도 높인다. 소비자의 만족은 소비자가 기업의 서비스를 지속해서 사용하게 한다. 이는 기업의 성장에 큰 영향을 미친다.

정보통신의 발전으로 다양한 유형의 모빌리티 서비스들이 발전했다. 여기서 카셰어링, 자전거 셰어링, 전기 킥 스쿠터와 같은 새로운 이동성 서비스가 등장하였다[4]. 여러 나라에서 교통 분야에서 탄소배출 문제에 대해 많은 관심을 기울임에 따라 전기를 이용하는 친환경 모빌리티 산업에 관한 관심도 높아지고 있다.

마이크로 모빌리티는 걸어서 이동할 때는 긴 시간이 걸리는 거리를 빠르게 이동하기에 편리하여 젊은 층 사용자들에게 인기가 많다. 또한, 자동차에 비해 주

차에서 강점을 보이면서 미래 교통수단으로 주목받고 있다. 실제 PM의 수요는 전 세계적으로 증가하고 있다. Global Research News[5]에서 전 세계 개인 이동 장치 시장 규모는 2022년 1억 2,450만 달러로 평가되었으며, 예측 기간 동안 연평균 복합 성장률이 7.26%로 확장되어 2028년에는 1억 5,414만 달러에 이를 것으로 예상된다. <그림 1-1>은 Global Research News 개인형 모빌리티 시장 규모, 점유율 및 추세 분석을 예측한 결과를 나타낸다. 한국에서도 연평균 20% 이상 성장세를 보인다. 한국교통연구원은 2022년 한국의 퍼스널 모빌리티 시장은 약 6천억 원에 달할 것으로 전망한다. 그중에서도 전동킥보드의 경우는 코로나19의 장기화로 인해 많이 늘어나고 있는 교통수단이다. 이륜차는 코로나19 이전에도 레저, 취미, 출퇴근용 교통수단으로 애용됐지만 본격적으로 수가 늘어나기 시작한 것은 코로나19가 등장하면서부터이다[6]. 앞서 설명한 이유로, 마이크로 모빌리티와 관련된 시장도 빠르게 성장하면서 다양한 업체들이 생겨났다. 이러한 업체들이 시장에서 살아남으려면 소비자가 지속해서 이용하도록 만들어야 한다.

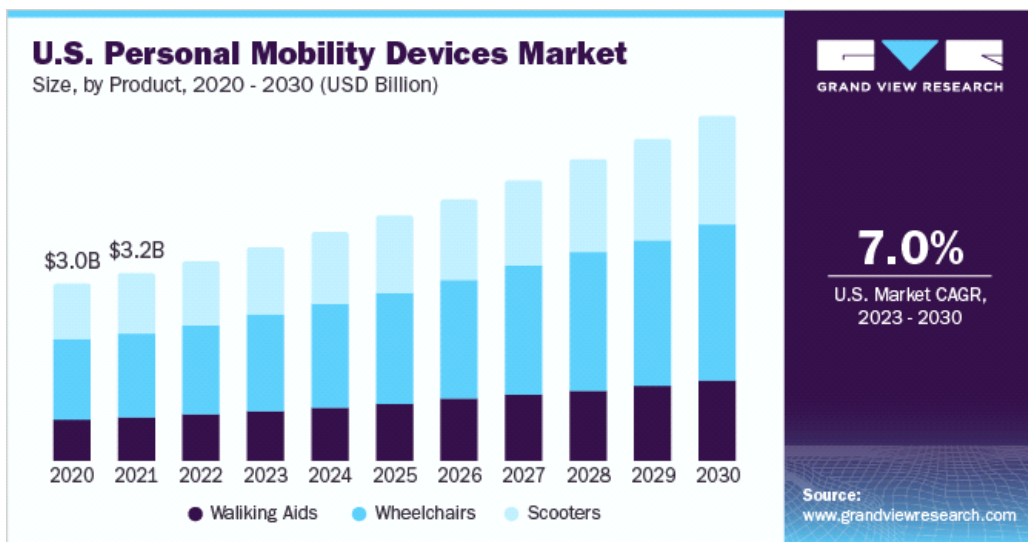


그림1-1. 개인형 모빌리티 시장 규모, 점유율 및 추세 분석(Global Research News)

1.1.2 연구 목적

때때로 대여소마다 반납량이 많아서 정류소에 반납 자리가 없는 상태가 발생하거나, 원하는 장소에 마이크로 모빌리티가 부족해서 대여할 수 없는 상황이 발생할 수 있다. 이러한 소비자의 불편을 해소하기 위해서는 적절한 수요량을 예측하여 미리 마이크로 모빌리티를 확보하고 필요한 곳에 배치해야 한다.

본 연구는 실제 마이크로 모빌리티 데이터를 다양한 그래프로 시각화하여 속성 간의 상관관계와 대여 수에 상대적으로 큰 영향을 미치는 속성에 대하여 분석하였다. 그리고 기온, 습도, 일사량, 강수량 등의 날씨 정보와 평일, 주말 정보를 이용하여 머신러닝 모델을 학습시키고 마이크로 모빌리티의 수요량을 예측한다. 또한, MAE, MSE, RMSE, R2 SCORE의 성능 지표로 여러 가지 모델의 성능을 비교한 뒤 상위 모델 3개를 선정하고, 이 모델들로 앙상블 모델을 구현하여 예측 성능을 개선한다. 이를 바탕으로 정류소에 배치할 수 있는 마이크로 모빌리티의 수를 예측하여 실제 서비스에 적용할 수 있는 것을 목표로 한다.

1.1.3 연구의 기여점

본 연구의 기여점은 다음과 같다.

본 연구에서는 마이크로 모빌리티 대여에 미치는 요인을 분석하고 머신러닝 모델을 이용하여 대여 수를 예측한다. 이를 통해 실제 업체가 현재 제공하는 대여 서비스의 질을 개선할 수 있도록 한다.

본 연구에서는 다양한 모델의 성능을 비교하고 성능이 더 나은 모델을 제시한다. 다양한 모델 중에서 성능이 상대적으로 뛰어난 모델을 선정하여 성능을 향상하는 앙상블 모델을 제안한다.

본 연구에서 제안하는 시스템을 구현하여 마이크로 모빌리티 서비스 회사는 사용자들이 필요로 하는 마이크로 모빌리티를 사전에 적절하게 분산배치 하여 사용자들의 수요에 신속하게 대응 및 지원하여 불편함을 덜 수 있다. 예측 수가 적

은 날에는 대여되지 않은 모빌리티와 관리 인력을 효율적으로 배치할 수 있다.

1.2 논문 구성

본 논문은 서론을 포함하여 총 5장으로 구성되어 있다. I 장 서론에서는 본 연구의 배경 및 목적에 관해서 설명한다. II 장에서는 이론적 배경에 대한 설명으로 배송 서비스 전략과 마이크로 모빌리티, 시계열 데이터와 머신러닝에 관해 설명한다. 그리고 본 연구와 관련된 연구들에 관해서 기술한다. III 장에서는 제안하는 방법으로 데이터와 데이터 전처리 방법에 관해 설명한다. IV장은 실험 결과 챕터로 실험환경과 실험 방법에 관해 기술하며 시각화, 분석 결과, 실험 결과를 설명한다. 그 결과에 대한 모델의 성능평가가 이루어진다. 또한, 실제 서버에 적용하는 과정으로 Flask에 대해 언급하고 실제 업체가 사용하고 있는 서비스를 보여준다. V 장에서는 연구의 결론과 토의에 대하여 설명한다.

II. 이론적 배경

2.1 배송 서비스 전략

전자 상거래의 발전은 물류 업계에도 혁신을 불러오고 있다. 배송 서비스는 인터넷 기반 전자 상거래 발전의 원동력으로 작용하며 배송서비스 품질에 관한 연구는 매우 중요하게 떠오르고 있다. 인터넷 쇼핑몰의 경쟁우위 고객 확보는 고객 만족도를 통한 고객과의 장기적 관계 구축에 달려있다[7].

쿠팡은 2010년 설립된 한국의 전자상거래 및 온라인 쇼핑몰 중 하나이다. 쿠팡은 2015년에 1조 1천 337억 5천만원의 매출을 달성하였고, 국내의 소셜 커머스 시장에서 1위를 차지하였다. 쿠팡은 소비자의 만족도를 위해 빠른 배송, 빠른 보상, 빠른 반품 서비스를 제공하고 오픈 시간을 9시로 변경하여 소비자가 기다려야 하는 불편함을 줄였다[8]. 이후 2014년 로켓배송이라는 서비스를 시작하면서 급성장하게 되었다. 이러한 로켓배송이 가능하게 된 데에는 쿠팡의 자체 물류센터 구축과 자체 택배기사 고용이 있었다[9]. 또한 물류센터에 수요 예측 머신러닝을 투입하였다. 쿠팡의 AI 머신러닝은 소비자의 기전 주문데이터를 분석하여 상품 주문량을 예측하는 것이다. 이것을 통해서 전국의 쿠팡 물류센터에 위치별로 나눠서 미리 상품을 구비 할 수 있다. 이뿐만 아니라 물류센터에 입고된 상품을 빨리 출고하기 위한 방법으로 상품의 진열 위치, 상품을 꺼내는 동선, 어떤 배송 트럭에 실을지, 배송트럭의 어느 자리에 놓을 지등의 모든 배송 과정을 모두 인공지능을 통해 결정한다[10]. 머신러닝을 이용해 급격한 수요 증가를 예측하고, 상품이 필요하기도 전에 물류와 배송 네트워크에 선제적으로 배치한다 [11].

2.2 마이크로 모빌리티

마이크로 모빌리티는 비교적 짧은 거리를 이동하는 2인 이하의 소형 개인 이동 수단으로 정의할 수 있다. 마이크로 모빌리티는 짧은 거리를 이동하는 1~2인승의 소형 개인 이동 수단이다. 여기에는 전동킥보드, 공유자전거, 전기자전거, 전동휠, 세그웨이 등이 포함된다. 여기서 마이크로란 가볍고 차지하는 공간이 작으며 보통 짧은 거리를 모두 포함한다. 거리로는 0.5km~4km, 시간으로는 5분~45분 정도의 거리를 이동하는 데 적합한 이동 수단이다[12]. 전동킥보드는 마이크로 모빌리티 중에서 전동 스쿠터는 핸들 바, 데크 및 바퀴가 달린 전기로 동작하는 차량으로 약 25km/h의 비교적 낮은 속도로 운전자 한 명만 탑승할 수 있다. 혼잡한 도로에서 이동 시간을 단축하고 단거리 이동 속도를 높일 수 있다. 이러한 마이크로 모빌리티는 경제적이며 환경적 영향이 적은 지속 가능한 이동 수단으로 주로 여행용으로 많이 이용된다. 최근에는 전동킥보드를 활용한 공유서비스가 활발하게 보급되고 있다[13].

2.3 전동킥보드 공유서비스

전동킥보드 공유서비스란 전동킥보드를 단기 임대 목적으로 이용할 수 있는 서비스를 말한다[14]. 공유 마이크로 모빌리티 서비스는 사람들에게 디지털 플랫폼을 통해 마이크로 차량을 대여할 기회를 제공한다. 이러한 서비스는 사용자가 필요에 따라 단기간에 교통수단을 이용할 수 있도록 하는 혁신적인 교통 전략이다. 사용자는 다양한 위치에서 다양한 종류의 마이크로 모빌리티 대여 업체로부터 서비스를 받을 수 있다[15]. 전동킥보드 공유서비스는 캘리포니아 산타모니카에 본사를 둔 Bird는 2017년 9월에 처음으로 등장하였다. Bird는 도시의 길모퉁이에 킥보드를 배치하고, 사용자가 스마트폰 애플리케이션을 사용하여 전동킥보드를 쉽게 찾을 수 있고, 요금을 내면 잠금이 해제되고 전동킥보드를 사용할 수 있게 하였다[16]. 전동킥보드는 접근이 편리하고, 비용이 저렴하여 커뮤니티에서 전동킥보드 이용이 빠르게 확산되었다. 이후로 다양한 공유 전동킥보드 대

여 서비스 업체들이 등장하여 뒤를 따르기 시작했다.

2.4 시계열 데이터 분석

시계열 데이터는 일정 기간동안 수집된 데이터 세트를 의미한다. 시계열 데이터는 시간이 지남에 따라 지속적으로 발생하며 연속한 관측값은 상관관계를 갖고 있다는 특징이 있다. 정렬된 시계열 데이터에서 의미 있는 정보를 추출하는 것을 시계열 분석이라고 한다. 시계열 분석의 목적은 과거가 미래에 미치는 영향을 알아내는 것이다. 대량의 데이터에서 유용하고 유효한 정보를 추출하는 과정을 데이터 마이닝이라고 한다. 데이터의 형식은 텍스트 데이터, 오디오 데이터, 비디오 데이터, 그래프 데이터, 스트리밍 데이터 등의 다양한 형식으로 존재한다. 이러한 많은 데이터 중에서 비즈니스나 시장 관련 분야에서 활용가능한 대량의 데이터를 빅데이터라고 한다[17]. 데이터는 유용한 정보로 변환되어야만 사용할 수 있다. 데이터 마이닝을 통해 유용한 정보와 패턴을 추출하고, 이를 분석해 미래를 예측할 수 있다. 이러한 방식으로 도출된 예측은 회사나 조직의 발전을 위한 유용한 결정을 내리는 데 도움이 될 수 있다[18]. 시계열 분석은 데이터 마이닝의 일부이다. 일정한 시간 간격 내의 많은 수의 데이터 값을 수집하는 것을 시계열 데이터라고 한다[19]. 시간은 연, 월, 일 등으로 나타낼 수 있으며 시계열을 분석하여 주어진 데이터 내에서 일어나는 변화를 예측할 수 있다.

2.5 머신러닝

머신러닝은 규칙성과 패턴을 유도하는 데 사용되는 인공지능(AI)의 한 분야로, 낮은 계산 비용으로 쉽게 구현할 수 있을 뿐만 아니라 빠른 학습, 검증, 테스트 및 평가를 할 수 있으며, 물리적 모델에 비해 높은 성능과 상대적으로 적은 복잡성을 제공한다. 머신러닝을 사용하는 데이터 기반 예측 모델은 최소한의 입력으로 더 빠르게 개발할 수 있다[20]. 머신러닝은 자동화된 컴퓨터 알고리즘과 통계 방법을 결합하여 데이터 세트에 숨겨진 패턴을 학습하는 분야이다[21]. 머신

러닝은 크게 지도학습과 비지도 학습으로 나눌 수 있다. 지도학습은 레이블이 지정된 학습 데이터를 사용하는 방식으로 입력 데이터를 원하는 출력으로 추정하거나 매핑한다. 반면에 비지도 학습은 레이블이 제공되지 않는 데이터를 사용하며 학습 과정 중에 출력에 대한 개념이 없다. 패턴을 찾거나 입력 데이터의 그룹을 발견하는 것은 학습 모델에 따라 다르다. 분류는 데이터를 유한한 클래스 집합으로 분류하는 프로세스를 말한다. 회귀 문제는 학습 함수가 데이터를 실제 값 변수로 매핑하여 새로운 샘플마다 이 프로세스를 기반으로 예측 변수의 값을 추정하는 프로세스이다. 클러스터링은 데이터 항목을 설명하기 위해 범주 또는 클러스터를 찾는 일반적인 비지도 작업이다. 이 프로세스를 기반으로 새로운 샘플들의 공유되는 유사한 특성에 따라 클러스터 중 하나에 할당하는 방식이다[22].

2.6 머신러닝 모델

2.6.1 Random Forest

Random Forest[23]는 의사결정 나무 모델의 단점을 보완하기 위해 개발된 모델로 의사 결정 나무 여러 개를 훈련해 그 결과를 종합해 예측하는 앙상블 알고리즘이다. Random Forest는 배깅(Bagging) 방식을 사용한다. 이 방식은 속성을 랜덤하게 뽑아서 결정 나무를 훈련시키고 테스트 데이터에 대해 여러 개의 결정 나무를 모아서 다수결 또는 평균을 사용하여 예측한다. 기본적으로 100개의 의사결정 나무를 갖는다. Random Forest는 시계열 데이터에서 시간에 따라 변화하는 여러 특성을 고려하여 패턴을 학습하기에 적절하다. 또한 이상치에 강하기 때문에 시계열 데이터에서 모델의 안정성을 높일 수 있다.

2.6.2 Extra Trees Regressor

Extra Trees Regressor(Extremely Randomized Trees Regressor)[24]는 여러 개의 결정 트리로 구성된 앙상블 알고리즘이다. Random Forest와 비슷하

게 동작하지만, 랜덤 포레스트와 다르게 중복된 훈련 샘플을 사용하지 않고 결정 트리를 만들 때 전체 훈련 세트를 사용한다는 특징이 있다. 또한, 특성들을 무작위로 선정하여 과대적합을 방지한다. 랜덤 포레스트보다 무작위성이 증가하며 과적합의 위험이 적고, 연산량이 적어 속도가 훨씬 빠르다는 장점이 있다. 이 알고리즘은 더 다양한 특성을 고려하고 더 높은 유연성을 제공하면서도 안정적인 예측을 할 수 있다. 이러한 특성으로 인해 Extra Trees Regressor는 시계열 데이터에서의 노이즈와 이상치에 강하고, 효과적인 예측 성능을 제공한다.

2.6.3 LightGBM Regressor

LightGBM[25]은 여러 개의 결정 트리로 구성된 앙상블 알고리즘이다. 그래디언트 부스팅(Gradient Boost Machine, GBM) 알고리즘의 일종으로 트리가 서로 독립적으로 생성된 Extra Trees와 다르게, Light GBM은 트리가 순차적으로 생성되면서 이전 트리가 틀리게 예측한 자료에 대해 가중치를 부여하여 학습한다. 순차적인 학습을 통해 오차를 보완해 가는 알고리즘을 부스팅이라고 하며, 가중치 업데이트에 경사 하강법을 사용하는 것이 GBM 알고리즘이다. LightGBM 알고리즘은 트리가 수직적으로 확장하는 leaf-wise 방식으로 기존의 부스팅 알고리즘보다 손실을 더 줄일 수 있다. 또한, 속도가 매우 빠르다는 장점이 있다. 이 알고리즘은 빠른 학습과 예측 속도를 제공하는데, 대용량 데이터 세트나 빠른 응답이 필요한 실시간 예측에 적합하다. 또한 시계열 데이터에서 범주형 변수가 있는 경우에 유용하게 활용될 수 있다.

2.6.4 CatBoost Regressor

CatBoost[26] 알고리즘은 기존의 부스팅 알고리즘의 과적합 문제를 개선한 Ordered Boosting 기법이다. 그래디언트 부스팅은 과적합 문제가 치명적인 단점인데, 이 과적합 문제를 해결하면서 동시에 기존 GBM 계열의 알고리즘인 XGboost, LightGBM 알고리즘보다 학습 속도가 개선된 모델이 CatBoost이다. 또한, Ordered Boosting 방식으로 그래디언트 부스팅 알고리즘에서의 타겟 누수(target

t leakage)를 개선했다. 다음 스텝의 트리를 만들 때 이전에 사용했던 데이터를 다시 재사용하여 과적합이 쉽게 발생하는 문제를 해결할 수 있다. 또한, 범주형 데이터 처리에 대한 자동화된 방식을 제공하며 이는 시계열 데이터에서 발생하는 범주형 특성을 효과적으로 다룰 수 있다. 또한, 데이터의 전처리 과정을 간편화하며, 적은 수의 하이퍼파라미터 튜닝으로 높은 성능을 보여준다.

2.7 기존의 관련연구

전 세계적으로 탄소 배출과 교통 혼잡으로 인한 환경 문제를 해결하기 위해 차량이나 자전거 공유를 활성화하는 서비스가 활성화되고 있다[29]. 이와 관련하여 탄소를 배출하지 않는 교통수단에 관한 연구가 활발히 진행되고 있다[30]. 예를 들어, 전기 모빌리티 기기의 배터리 수명과 관련된 연구가 있다. Jafari(2022)는 전기 모빌리티 기기 배터리를 위한 디지털 트윈 기반의 배터리 구조를 제안하는 연구를 진행하였다[31]. 자전거 또는 마이크로 모빌리티 수요에 미치는 영향을 분석한 연구도 있다[32-34]. El-Assi, W.(2017) 등은 역과 역 사이의 주요 도로와 교차로의 수가 자전거 공유에 부정적인 영향을 미친다는 것을 알아냈고 전체 노선 길이 대비 자전거 인프라의 비율이 높을수록 사용자의 수가 증가한다는 것을 알아냈다. Mathew, J.K.(2019) 등은 통계적 모델링을 사용하여 시간별 날씨요인을 시간대별로 분석하여 날씨가 전동킥보드 이용에 미치는 영향을 분석하였다. Tin Tin, S.(2012) 등은 뉴질랜드에서 자전거 이용량의 시간적, 계절적 변동성과 날씨와의 연관성을 조사하였다. Brandenburg, C.(2007) 등은 기상 조건이 자전거 활동에 미치는 영향을 분석하였다. 이 연구에서는 통근 및 레크리에이션 자전거 타기의 날씨 의존성을 분석했다. 한 연구에서는 뉴욕과 워싱턴 DC의 자전거 대여 데이터를 사용하여 동적 클러스터 기반 프레임워크를 사용하여 초과 수요 확률을 예측하였다[36]. 여러 연구에서도 통계적 예측 방법과 선형 회귀분석을 활용하여 자전거 공유 사용량과 날씨 간의 상관관계를 분석하여 추운 날씨와 비, 높은 습도는 자전거 이용가능성과 여행시간을 모두 감소시킨다는 것을 알아내었다. 대여소 근처에 지하철역이 있을 경우에 강수량의 영향

을 많이 받으며 시간에 따른 영향이 있다는 것을 알아냈다[37]. Yang, Z.(2016) 등은 에서 동력 이동 수단과 마찬가지로 자전거의 고르지 않은 분포로 인해 어려움을 겪는 자전거 공유시스템의 효율적인 재배치 전략을 위해 시공간적 자전거 이동성 모델을 공식화하였다. 이전 연구에서는 하이브리드 CNN-LSTM 모델을 사용하여 15분 전 자전거 수요를 예측하는 것을 목표로 하는 연구가 진행되었다. Chatterjee(2023)은 전기 스쿠터 수요 예측의 오차를 줄이기 위해 생성적 적대 신경망(GAN)을 사용하는 방법을 제안했다. 이를 통해 합성 시계열 데이터 생성이 개선되어 정확도가 향상되었다[40, 41]. 공유형 자율주행 전기자동차에서도 유사한 재배치 문제가 연구된 바 있으며, 택시 빅데이터를 활용한 승객 수요 예측 모델을 제안한 연구가 있다[42]. 재배치 문제 관련 연구로 전기차 충전소의 충전 최적화나 충전 인프라 입지를 위해 전기차 수요를 예측하는 연구도 있다[43]. 랜덤 포레스트를 이용하여 독일 두 도시의 여행 출발지 특성을 파악하는 연구가 진행되었다[44]. 일부 연구에서는 머신러닝의 정확도를 높이기 위해 투표 방법과 같은 앙상블 방법을 사용하기도 한다[45]. 터키의 의료 폐기물 예측 연구[46], 단기 에너지 예측 연구[47]에서 앙상블 투표 모델을 활용했다. 국내에서도 자전거 수요에 영향을 미치는 요인을 분석하고 서울시, 세종시, 대전광역시 등의 지역을 대상으로 자전거 수요 예측을 수행한 연구가 다수 진행되었다[48-53].

Ⅲ. 제안하는 방법

3.1 실험 데이터

다음은 본 연구에 사용된 데이터에 관한 설명이다. 마이크로 모빌리티의 대여 데이터는 E사로부터 수집한 실제 전동킥보드의 대여 수 데이터로 총 16,042개의 데이터 세트를 수집하였다. 수집한 데이터의 전체기간은 2019년 4월 16일부터 2022년 8월11일까지 이다. 본 연구에서는 전동킥보드의 유형으로는 <표 3-1>에서처럼 4개 유형의 마이크로 모빌리티 데이터를 사용하였다. 이 데이터에는 대여 정류장, 대여 수, 대여 날짜, 대여 위치가 포함된다. 4종류의 장비마다 포함하는 속성이 다른데 이것을 <표 3-2>에서 나타내었다. x pos와 lat은 대여 장소의 위도를 나타내며 y pos와 lon은 대여 장소의 경도를 나타낸다. IoT Station 장비의 경우는 rent station이라는 E사에서 지정한 대여소 속성을 포함하지만, non-IoT Delivery 장비의 경우에는 배달 주소인 od_addr라는 속성을 갖는다. <그림 3-2>는 마이크로 모빌리티 대여 데이터 예시를 나타낸 것이다.

장비 유형	설명
Station based IoT	GPS가 부착되어있고, station에서 빌리는 유형
Station based non-IoT	GPS가 부착되어있지 않고, station에서 빌리는 유형
Delivery based IoT	GPS가 부착되어있고, 사용자의 위치로 배달하는 유형
Delivery based non-IoT	GPS가 부착되어있지 않고, 사용자의 위치로 배달하는 유형

표 3-1. 모빌리티 유형

feature	IoT Station	IoT Delivery	non-IoT Station	non-IoT Delivery
rent date	○	○	○	○
year	○	○	○	○
month	○	○	○	○
day	○	○	○	○
day name	○	○	○	○
rent num(count)	○	○	○	○
rent station	○			
x pos(lat)	○	○	○	○
y pos(lon)	○	○	○	○
address		○	○	
od_addr				○

표 3-2. 마이크로 모빌리티 데이터의 속성

	rent_date	day_name	year	month	day	rent_num	x_pos	y_pos
0	2021-06-04	Friday	2021	6	4	114	1605.035540	6075.615122
1	2021-06-05	Saturday	2021	6	5	178	1806.785835	6837.263958
2	2021-06-06	Sunday	2021	6	6	195	1739.039766	6581.752018
3	2021-06-07	Monday	2021	6	7	120	1705.686691	6457.083521
4	2021-06-08	Tuesday	2021	6	8	118	1371.547256	5189.086991

그림 3-1. 마이크로 모빌리티 대여 데이터 예시

<표 3-3>은 날씨 데이터를 설명한다. 날씨 데이터는 기상청 기상자료개방 포털에서 제공하는 지상관측 자료를 사용하였다. 제주, 고산, 성산, 서귀포 총 4개의 관측소에서 관측된 자료로 평균기온(℃), 일 강수량(mm), 평균 상대습도(%), 합계 일사량(MJ/m²)의 데이터를 사용하였다.

날씨 데이터 속성	설명	기간
지점 코드	지역을 숫자 코드로 표시한 것	2019.04.16.~ 2022.08.11
지점명	지점 이름	
일시	관측 날짜	
평균기온(° C)	하루의 평균기온	
일강수량(mm)	하루 동안 내린 강수량	
평균 상대습도(%)	하루 동안 상대습도의 평균	
합계 일사량(MJ/m2)	하루 동안 일사량의 총합	

표 3-3. 날씨 데이터

3.2 시스템 구성도

<그림 3-1>은 본 연구의 시스템 구성도이다. 마이크로 모빌리티 대여 데이터와 날씨 데이터를 각각 수집한 후 전처리 과정을 거친다. 전처리가 완료된 데이터를 통하여 다양한 그래프로 시각화하여 데이터를 분석한다. 그 후 훈련 세트와 테스트 세트로 분리하여 다양한 머신러닝 알고리즘을 통해 대여 수를 예측한다. 그 결과를 통해 모델을 비교한 후에 여러 성능 지표가 상위인 모델을 선정하여 앙상블 모델을 만든다. 앙상블 모델과 단일 모델의 성능을 비교한 후 더 좋은 모델을 사용하여 날씨 예보 데이터와 함께 미래 시점의 마이크로 모빌리티 대여 수를 예측한다. 실제 모빌리티 대여 업체에서 사용할 수 있도록 서버에서 작동하도록 하였다.

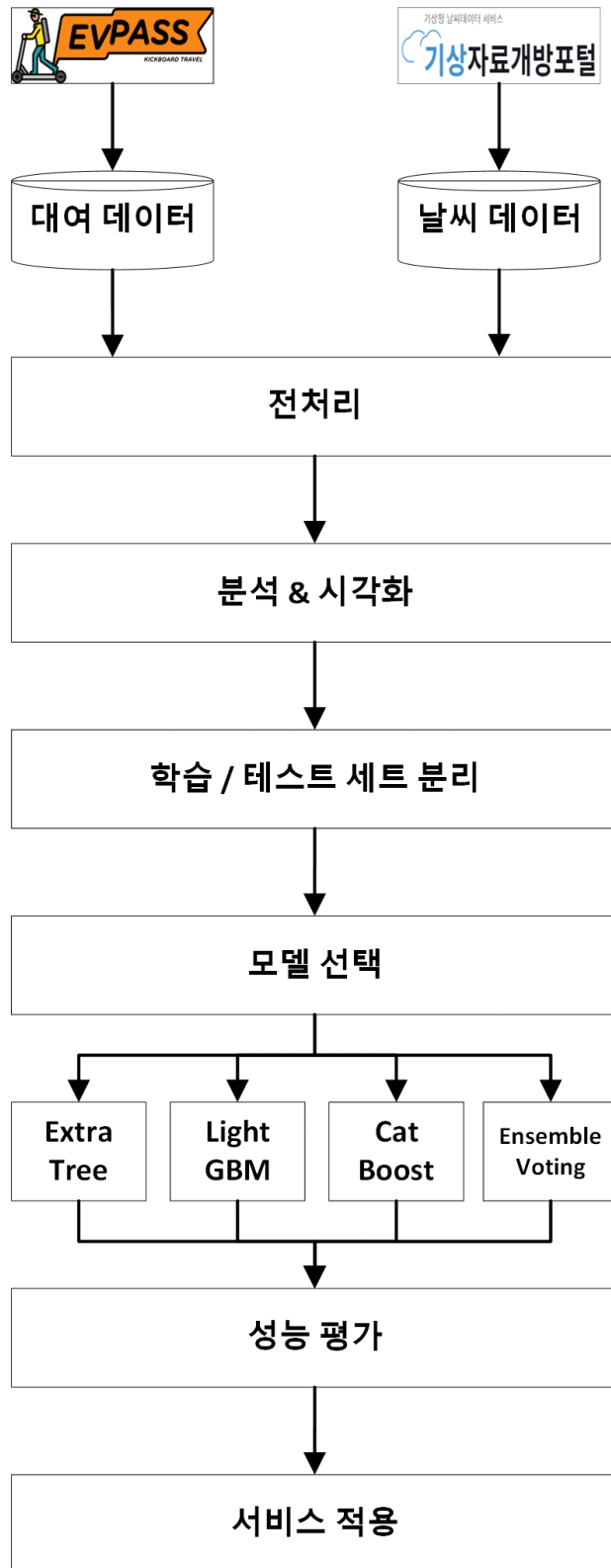


그림 3-2. 전체 흐름도

3.3 데이터 전처리

앞서 <표 3-1>에서 설명한 마이크로 모빌리티 데이터와 <표 3-3>의 날씨 데이터를 전처리하고 합쳐 전체 데이터를 구성한다. 전처리 과정에서는 먼저 4개로 나누어져 있는 모빌리티 데이터와 날씨 데이터를 날짜별로 그룹화한다. 인코딩 과정에서는 평일을 0, 주말을 1로 인코딩하였다. 강수량 데이터는 최소 0에서 최대 174까지의 수치를 가지며 이를 0~6단계까지 7단계로 범주화하였다. 강수량의 범주화는 <표 3-4>와 같다. 그런 다음, 두 개의 데이터 세트를 통합하여 하나의 전체 데이터 세트를 구성한다. 전처리 과정을 통하여 2019년 4월 16일부터 2022년 8월 11일까지의 데이터를 가공하였고, 여기에는 10개의 컬럼으로 구성된 총 1,141일의 데이터 인스턴스가 포함된다. 그 결과 <표 3-5>와 같은 형태의 실험 데이터 셋이 만들어진다.

강수량	Level
0(강수량없음)	0
0~1 mm	1
1~5 mm	2
5~10 mm	3
10~15 mm	4
15~20 mm	5
20 mm ~	6

표 3-4. 강수량의 범주화

기간	속성
2019.04.16.~ 2022.08.11	rent date
	rent num
	year
	month
	day
	temperature
	humidity
	insolation
	rain
	weekend

표 3-5. 완성된 데이터 셋

3.3 앙상블 기법

다음은 앙상블 방식에 대한 설명이다. 앙상블은 여러 개의 모델을 생성하고 그 예측을 결합하여, 더욱더 정확한 예측을 도출하는 기법을 말한다. 앙상블 모델은 단일 모델의 약점을 다수의 모델을 결합하여 보완한다. 앙상블 기법 유형은 크게 부스팅(Boosting) [26], 배깅(Bagging) [27], 보팅(Voting) [53] 방식이 있다.

부스팅은 약한 학습기를 여러 개 결합하여 강한 학습기를 만드는 기법이다. 부스팅은 이전 모델이 잘못 예측한 샘플에 높은 가중치를 주어 다음 모델이 그 부분에 대해 더 학습하도록 만든다. 점차 예측 오류를 줄어든도록 학습하는 방식이다. 배깅은 bootstrap aggregating의 줄임말이다. 훈련데이터 셋에서 중복을 허용하여 샘플을 랜덤하게 선택하는 부트스트래핑(Bootstrapping) 분할 방식을 사용한다. 이러한 방식으로 샘플링한 데이터 세트에 대해서 여러 개의 기본모델을 훈련한다. 각각의 모델을 독립적으로 학습시킨 후 예측 결과를 결합한다. 배깅 방식을 사용하면 분산을 줄이고 과적합을 방지할 수 있다. Voting 방식은 다른 알고리즘이 낸 결과물에 대하여 투표를 통해 결정하는 방식이다. Voting regression는 다른종류의 회귀 모델들을 결합하여 평균 예측값을 반환하는 앙상블 방법이다. 배깅과는 달리, Voting regression의 각 베이스 모델은 전체 데이터 셋에 맞춰 학습되고, 이러한 예측의 결합에 따라 최종 예측이 결정된다[39]. 본 연구에서는 다양한 모델 조합을 만들 수 있는 보팅(Voting)기법을 채택하였다.

3.3.1 보팅 (Voting)

Voting 유형에는 Hard Voting과 Soft Voting 두 가지의 방식이 존재한다. Hard Voting 방식은 각 알고리즘이 낸 결과에서 가장 투표를 많이 받은 결과를 선택하는 방식이고, Soft voting 방식은 각 알고리즘의 결과의 확률을 구하고 이것을 더해서 평균을 산출한다. 평균 중에 가장 높은 결과를 선택하는 방식이다. 본 논문에서는 Soft Voting 방식을 채택하였다. 본 연구에서는 PyCaret 라이브러리를 통해 다양한 모델 간의 성능을 비교해본 결과 MAE, MSE, RMSE, RMSLE, MAPE의 성능평가 점수의 상위 4개의 모델인 Random Forest, ExtraTrees, LightGBM, CatBoost를 선택했다. 이 세종류의 모델을 이용하여 Voting regressor[27] 모델을 구성하였다.

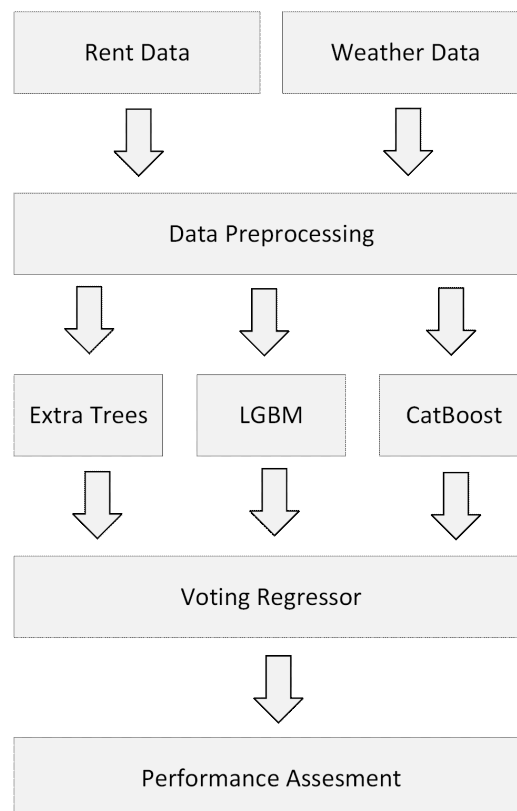


그림 3-3. Voting Regressor

IV 실험 결과

4.1 시스템 환경

본 연구는 <표 4-1>에서 설명하는 시스템 환경에서 진행되었다. Intel i7-8700, 3.20GHz CPU와 RAM은 16GB, 운영체제는 Windows 10 Pro 환경에서 진행되었다. 파이썬의 버전은 3.8.13을 사용하였다.

CPU	Intel i7-8700 3.GHz
RAM	16GB
OS	Windows 10 Pro
Language	Python 3.8.13

표 4-1. 시스템 환경

4.2 성능평가 지표

다음은 머신러닝 모델의 성능평가에 관한 설명이다. 마이크로 모빌리티 수요 예측의 평가 지표로는 평균 제곱 오차, 평균절대오차, 평균 제곱근 오차 그리고 결정계수(R² score)를 사용하였다.

4.2.1 평균제곱오차 (Mean Squared Error, MSE)

평균제곱오차는 알고리즘이 예측한 값과 실제 정답과의 차이를 제곱한 값의 평균값이다. 주로 회귀 문제에서 모델의 성능을 평가하는 데 사용되는 지표로 제곱을 하기 때문에 특이치에 민감하다는 특징이 있다. 각각 모델이 예측한 값과 실제 관측값(타겟 값) 사이의 오차를 계산하고 이를 평균화하여 측정한다. MSE의 값이 작을수록 모델의 예측이 실제 값과 가깝다고 볼 수 있다. 수식은 다음과 같다.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

4.2.2 평균절대오차 (Mean Absolute Error, MAE)

평균절대오차는 알고리즘이 예측한 값과 실제 값의 차이를 측정하는 데 사용되는 성능평가 지표로 각각 예측값과 실제값의 오차 절댓값을 적용한 후 이를 평균한 값이다. MSE와는 다르게 오차를 제곱하지 않아 특이값의 영향을 적게 받는다. MAE가 낮을수록 모델의 성능이 좋다고 평가할 수 있다. MAE의 수식은 다음과 같다.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

4.2.3 평균 제곱근 오차 (Root Mean Squared Error, RMSE)

평균 제곱 오차(MSE)의 제곱근을 계산한 값이다. 예측값과 실제값의 오차 제곱의 평균에 다시 제곱근을 취하기 때문에 실제 값과 같은 단위를 갖는다. 모델의 평균적인 오차를 측정하고 낮을수록 성능이 좋다고 볼 수 있다. RMSE의 수식은 다음과 같다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

4.2.4 결정계수 (R2 score)

결정계수는 회귀 문제에서 사용되는 통계적 측정 지표로, 데이터를 얼마나 잘 설명하는지를 나타낸다. 모델이 종속 변수의 변동성을 얼마나 잘 설명하는지를 나타내며 예측의 적합도를 평가하는 데 사용된다. 결정계수는 0과 1 사이의 값

으로 존재하며 1에 가까울수록 모델이 잘 예측했다고 볼 수 있다[57]. 다음은 결정계수의 식에 대한 설명으로 (1)-(4)에 의해서 정의될 수 있다.

$$(1) SSR = \sum_{i=1}^n (y_i - \bar{y}_i)^2$$

$$(2) SST = SSE + SSR$$

$$(3) SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$(4) R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

4.3 데이터 분석 및 시각화

전체 데이터는 1,141개의 인스턴스와 10개의 칼럼(전체 날짜, 년, 월, 일, 대여 수, 온도, 강수량, 습도, 일사량, 주말 여부)으로 구성된다. 다음은 전처리 된 데이터를 시각화하고 분석한 것이다. <그림4-1>은 데이터의 전체기간 2019년 4월 ~ 2022년 8월까지 대여 수를 그래프로 나타낸 것이다. 행의 개수는 1,141개이고 평균이 52~53대, 최솟값은 1, 최댓값은 251대이다. x축은 날짜를 나타내고, y 축은 대여 수를 나타낸다.

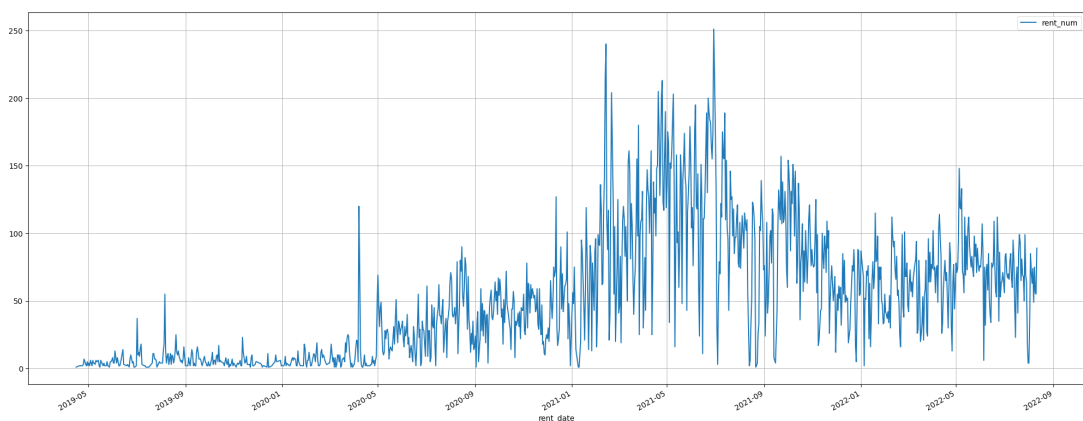


그림4-1. 전체기간 동안 대여 수

<그림 4-2>는 월별로 대여 수의 합을 막대그래프로 나타낸 것이다. 월별 대여 수를 살펴보면 6월과 7월의 대여 수가 가장 많은 것을 볼 수 있다.

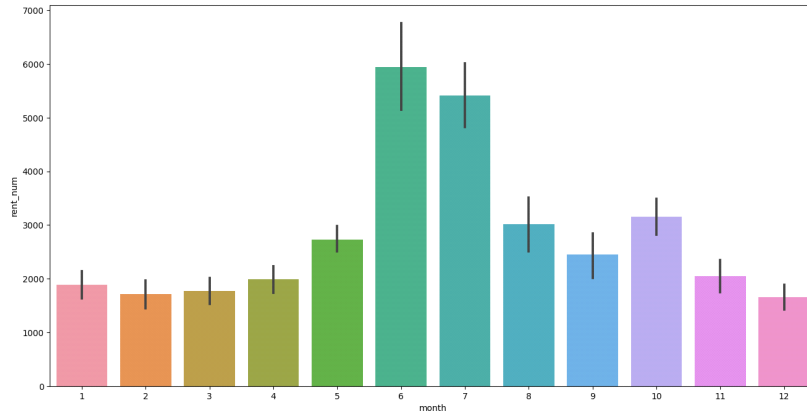


그림 4-2. 월별 대여 수의 합

<그림 4-3>은 요일별로 대여 수의 평균을 막대그래프로 나타내어 분석한 것이다. <그림 4-3>의 왼쪽 그래프(A)에서 평일에는 목요일이 가장 대여량이 많았고 주말 중에는 토요일의 대여 수가 더 많았다. 오른쪽 그래프(B)는 주말 여부에 따른 대여 수의 평균을 막대그래프로 나타낸 것이다. 평일은 0, 주말은 1로 나타낸 것이다. 평일에 대여 수의 평균이 약간 더 많은 것을 볼 수 있다.

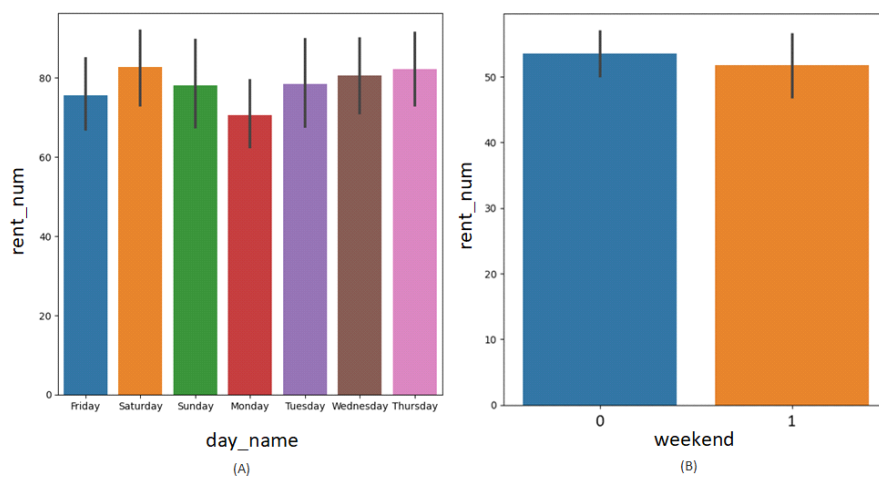


그림 4-3. 요일별/ 주말 여부에 따른 대여 수의 평균

<그림 4-4>는 특징들 사이의 상관관계를 분석한 히트맵이다. 히트맵을 통해 전체기간의 마이크로 모빌리티 대여량과 나머지 독립변수들 상관관계를 볼 수 있다. 마이크로 모빌리티의 대여 수는 일사량, 강수량, 온도와 상관관계가 큰 것을 알 수 있다. 일사량, 강수량, 온도가 미치는 영향이 큰 것을 알 수 있다. <표 4-2>는 여름과 겨울을 따로 분리해서 상관관계를 분석한 후 상관관계가 높은 특징 상위 4개를 표로 만든 것이다. 계절별로 대여 수에 영향을 미치는 특징을 분석해 보았을 때 여름과 겨울에 대여 수에 영향을 미치는 중요한 요소가 다른 것을 볼 수 있다.

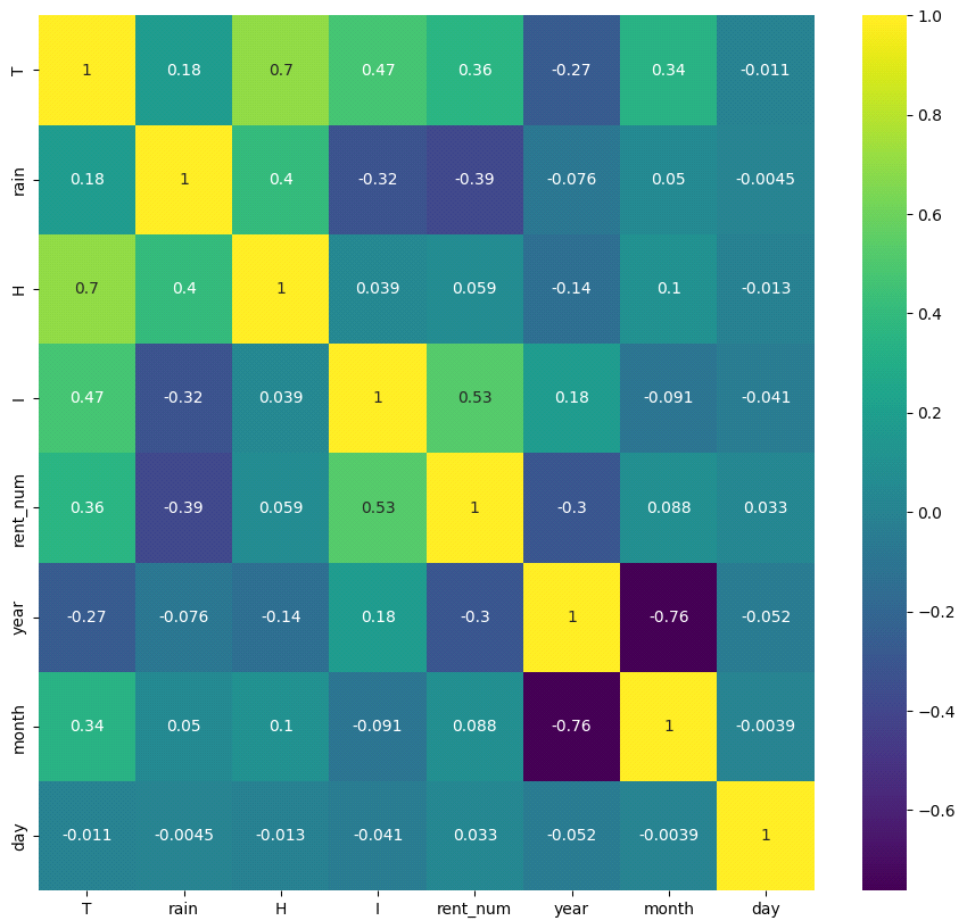


그림 4-4. 전체기간에 대한 히트맵

feature	여름(7월, 8월, 9월)	겨울(1월, 2월)
일사량	0.48	0.41
강수량	-0.63	-0.27
습도	-0.4	-0.23
온도	-0.08	0.39

표 4-2. 여름 / 겨울에 따른 상관관계 표

4.4 학습 방법

본 연구에서는 총 1,141의 데이터 중에서 임의로 과거의 200일 번째 데이터부터 시작하여 테스트를 진행하였다. 학습 방식은 데일리 트레이닝 방식으로, 전날까지의 데이터를 전부 테스트 데이터로 사용해서 다음 날의 대여 수 또는 이틀 후, 10일 후까지 예측하는 방식이다. <그림 4-5>는 데일리 트레이닝 방식을 그림으로 표현한 것이다. 예를 들어 2022년 8월 11일의 대여 수를 예측할 때 2019년 4월 16일부터 2022년 8월 10일까지의 데이터를 테스트 세트로 사용하여 8월 11일의 대여 수를 예측한다. 그림 8은 데일리 트레이닝 방식을 그림으로 나타낸 것이다. 대여 날짜(연도, 월, 일), 주말 여부, 기온, 강수량, 일사량, 습도가 Input 데이터로 사용된다.

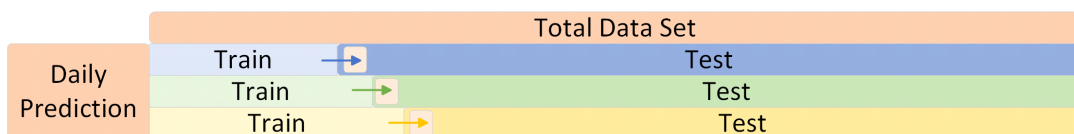


그림 4-5. 훈련/테스트 방법

4.5 머신러닝 학습 및 테스트 결과

<표 4-3>은 머신러닝을 통해 마이크로 모빌리티 대여 수를 예측한 결과에 대

한 성능평가이다. 표를 보면 단일 모델을 사용할 때 ExtraTrees Regressor, CatBoost Regressor, LightGBM Regressor, RandomForest Regressor 순서로 성능이 좋은 것을 알 수 있다. 서로 다른 알고리즘이 도출해 낸 결과물에 대하여 최종 투표하는 방식을 통해 최종 결과를 선택하는 앙상블 기법의 하나인 Voting 을 이용한 Voting Regressor 모델을 만들어 학습시켰다. 제안하는 모델에 대해서 MAE는 16.3187, MSE는 501.4068, RMSE는 22.3921, R2 score는 0.7629로 성능이 향상된 것을 보여주었다. 이는 성능 가장 성능이 좋은 단일 모델에 비해 모든 성능평가 지표에 대해 각각 1.14%, 1.76%, 3.26%, 1.64% 개선된 것을 알 수 있다.

Model	Performance Evaluation				
	R2	MAE	MSE	RMSE	Time(sec)
Random Forest	0.6726	19.2933	701.1863	26.4799	222.1845
Extra Trees	0.7543	16.6109	522.7416	22.8635	145.7737
LightGBM	0.7150	17.5575	606.3860	24.6249	56.9202
Cat Boost	0.7564	16.6631	518.2903	22.7659	1079.8587
Voting	0.7629	16.3187	501.4086	22.3921	1332.3511

표 4-3. 머신러닝 결과

4.6 특징 중요도

특징 중요도(Feature Importance)는 머신러닝 모델에서 각 특징이 결과에 얼마나 큰 영향을 미치는지를 나타내는 척도로 이는 모델이 예측을 수행할 때 특정 특징이 결과에 얼마나 기여 하는지를 파악하는 데 사용된다. 각 특징이 예측을 수행할 때 기여한 정도를 추정할 수 있다. 다음 <그림 4-6>은 ExtraTrees Regressor의 특징 중요도를 추출하여 시각화한 것이다. 연도와 일사량이 각각 51% 와 14%로 결과에 큰 영향을 미친다는 것을 알 수 있었다.

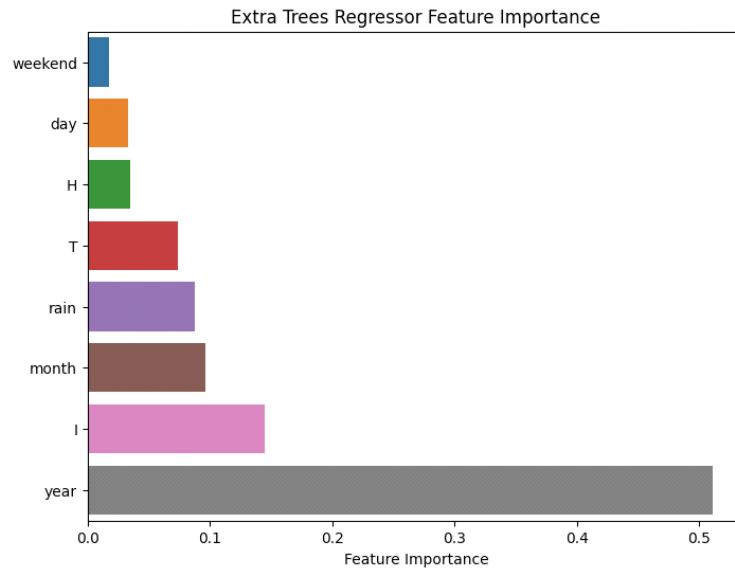


그림 4-6. 특징 중요도

4.7 활용 사례

4.7.1 미래의 대여 수 예측

본 연구에서는 기상청으로부터 날씨 예보 정보를 가져와서 대여 수를 예측하는 연구를 진행하였다. 우선 날씨 예보 데이터는 기상청 Open API를 통해 크롤링 할 수 있다. 본 연구의 테스트 기간은 2022년 10월 19일부터 2023년 4월 7일 까지이다. 데이터 세트를 매일매일 업데이트하는 과정은 다음과 같다. <그림 4-7>은 이 과정을 그림으로 나타낸 것이다.

- 1단계 - 기상청 지상 일 자료 API를 이용하여 어제의 날씨 정보를 가져온다.
- 2단계 - 기상청의 단기예보 API를 이용하여 오늘 ~ 2일 후까지의 예보 정보를 가져온다.
- 3단계 - 기상청 중기예보 API를 이용하여 3일 후 ~ 10일 후까지의 예보 정보를 가져온다.
- 4단계 - 예보 데이터를 결합한 후 전처리 과정을 거쳐 테스트 데이터 셋을

완성한다.

- 5단계 - Voting 모델을 이용하여 어제까지의 데이터를 이용해 학습시킨 후 새로 만든 테스트 데이터 셋에 대하여 예측한다.

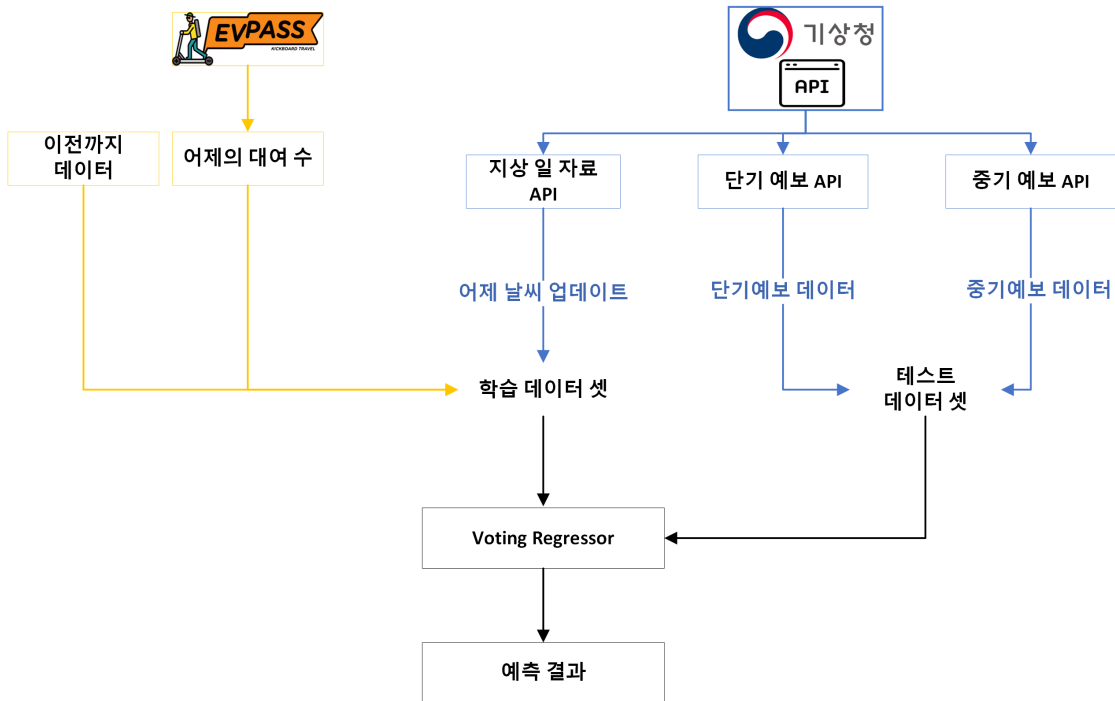


그림 4-7. 미래 시점의 대여 수 예측 흐름도

4.7.2 기상청 단기예보

단기예보는 예보 기간과 구역을 시·공간적으로 세분화하여 발표하는 예보로, 지역별 시간별 차이를 최소화하기 위해 전국을 5km * 5km 간격의 격자로 나누어 읍, 면, 동 단위의 행정구역 중심으로 상세한 날씨를 제공한다. 단기예보는 예보 기간을 글피까지 확장 및 예보 단위를 상세화하여 시공간적으로 세분화한 예보를 제공한다. 단기예보 API가 제공하는 항목은 <표 4-4>에서 설명하는 것과 같다. 단기예보 API를 통해서 제주도의 애월읍과 성산읍의 일 최저기온, 최고기온, 습도를 수집하였다. 단기예보는 한 시간마다의 습도를 제공하기 때문에 하루

의 평균을 구해서 사용하였다. 두 지역 기상정보의 평균을 사용하였다. <그림 4-8>은 단기예보 크롤링 결과의 예시를 보여준다.

예보구분	항목값	항목명	단위
단기예보	POP	강수확률	%
	PTY	강수형태	코드값
	PCP	1시간 강수량	범주(1 mm)
	REH	습도	%
	SNO	1시간 신적설	범주(1 cm)
	SKY	하늘상태	코드값
	TMP	1시간 기온	°C
	TMN	일 최저기온	°C
	TMX	일 최고기온	°C
	UUU	풍속(동서성분)	m/s
	VVV	풍속(남북성분)	m/s
	WAV	파고	M
	VEC	풍향	deg
	WSD	풍속	m/s

표 4-4. 단기예보가 제공하는 날씨정보

	baseDate	baseTime	category	fcstDate	fcstTime	fcstValue	nx	ny
0	20221104	0500	TMP	20221104	0600	6	33	126
1	20221104	0500	UUU	20221104	0600	2.9	33	126
2	20221104	0500	VVV	20221104	0600	-11.5	33	126
3	20221104	0500	VEC	20221104	0600	346	33	126
4	20221104	0500	WSD	20221104	0600	11.9	33	126
5	20221104	0500	SKY	20221104	0600	1	33	126
6	20221104	0500	PTY	20221104	0600	0	33	126
7	20221104	0500	POP	20221104	0600	0	33	126
8	20221104	0500	WAV	20221104	0600	1.5	33	126
9	20221104	0500	PCP	20221104	0600	강수없음	33	126
10	20221104	0500	REH	20221104	0600	65	33	126
11	20221104	0500	SNO	20221104	0600	적설없음	33	126
12	20221104	0500	TMP	20221104	0700	6	33	126
13	20221104	0500	UUU	20221104	0700	2.5	33	126
14	20221104	0500	VVV	20221104	0700	-11.3	33	126
15	20221104	0500	VEC	20221104	0700	348	33	126
16	20221104	0500	WSD	20221104	0700	11.6	33	126
17	20221104	0500	SKY	20221104	0700	1	33	126

그림 4-8. 단기예보 크롤링 예시

4.7.3 기상청 중기예보

중기예보란 중기 전망, 중기 육상 예보, 중기기온, 중기 해상 예보 정보를 조회하는 서비스이다. 중기예보는 단기예보와는 다르게 <표 4-5>에서 보는 것처럼 지역별로 3일 후부터 10일 후까지의 육상, 해상 날씨, 최고, 최저기온, 최고, 최저 파고를 제공한다. 다음 <그림 4-9>는 중기예보의 크롤링 결과 예시로 총 8 일간의 하루의 평균기온을 구할 수 있다. 본 연구에서는 두 지역으로부터 얻은 기온의 평균을 사용하였다.

예보구역코드	구역	예보구역코드	구역
11B10101	서울	11G00401	서귀포
11B20201	인천	11F20501	광주
11B20601	수원	21F20801	목포
11B20305	파주	11F20401	여수
11D10301	춘천	11F10201	전주
11D10401	원주	21F10501	군산
11D20501	강릉	11H20201	부산
11C20401	대전	11H20101	울산
11C20101	서산	11H20301	창원
11C20404	세종	11H10701	대구
11C10301	청주	11H10501	안동
11G00201	제주	11H10201	포항

표 4-5. 중기예보에서 지역 코드

날짜 기온			날짜 기온		
0	2023-11-03	21.0	0	2023-11-04	21.5
1	2023-11-04	21.5	1	2023-11-05	22.0
2	2023-11-05	19.5	2	2023-11-06	20.0
3	2023-11-06	15.0	3	2023-11-07	15.5
4	2023-11-07	16.0	4	2023-11-08	16.0
5	2023-11-08	16.0	5	2023-11-09	17.0
6	2023-11-09	17.0	6	2023-11-10	18.0
7	2023-11-10	15.0	7	2023-11-11	17.0
(A)제주시			(B)서귀포시		

그림 4-9. 중기예보 크롤링 예시

중기예보는 기온밖에 제공하지 않아서 습도는 단기예보로부터 수집한 습도의 평균을 사용하였고, 강수량과 일사량은 이전 7일의 평균을 사용하였다. 수집한 두 데이터를 결합하고 날짜를 이용하여 요일을 구한 다음 데이터 프레임으로 만든다. 그 결과는 다음 <그림 4-10>과 같다.

	rent_date	T	H	year	month	day	weekend	I	rain	rent_num
0	2023-01-26	6.00	75.00	2023	1	26	0	2.442143	2	0
1	2023-01-27	3.75	75.10	2023	1	27	0	2.442143	0	0
2	2023-01-28	2.25	72.29	2023	1	28	1	2.442143	0	0
3	2023-01-29	6.00	74.13	2023	1	29	1	2.442143	0	0
4	2023-01-30	6.00	74.13	2023	1	30	0	2.442143	0	0
5	2023-01-31	6.50	74.13	2023	1	31	0	2.442143	0	0
6	2023-02-01	9.50	74.13	2023	2	1	0	2.442143	0	0
7	2023-02-02	3.75	74.13	2023	2	2	0	2.442143	2	0
8	2023-02-03	4.50	74.13	2023	2	3	0	2.442143	1	0
9	2023-02-04	6.25	74.13	2023	2	4	1	2.442143	2	0
10	2023-02-05	6.25	74.13	2023	2	5	1	2.442143	1	0

그림 4-10. 완성된 테스트 데이터 예시

4.7.4 예측 결과

<그림 4-11>은 10월 18일부터 4월 7일까지의 테스트 결과를 그래프로 나타낸 것이다. 위에서 설명한 Voting Regressor 모델을 사용하였고 모델의 성능평가 결과는 <표 4-6>와 같다. <그림 4-11>에서 푸른색 선은 매일 예측한 하루의 예측 대여 수이고 노란색 선은 테스트 마지막 날인 3월 22일에 예측한 10일 후까지의 예측 대여 수이다. 빨간색 선이 나타내는 것은 실제 대여 수로 성능을 평가할 때 사용한다. 2023년 1월 20일~ 25일까지는 폭설로 인해 날씨가 매우

좋지 않아서 대여 수가 없다. 더불어 1월에 한 달 동안 13일 동안 비가 내려 안 좋은 날씨가 지속되었으며, 이에 따라 대여 수가 불안정한 양상을 보였다.

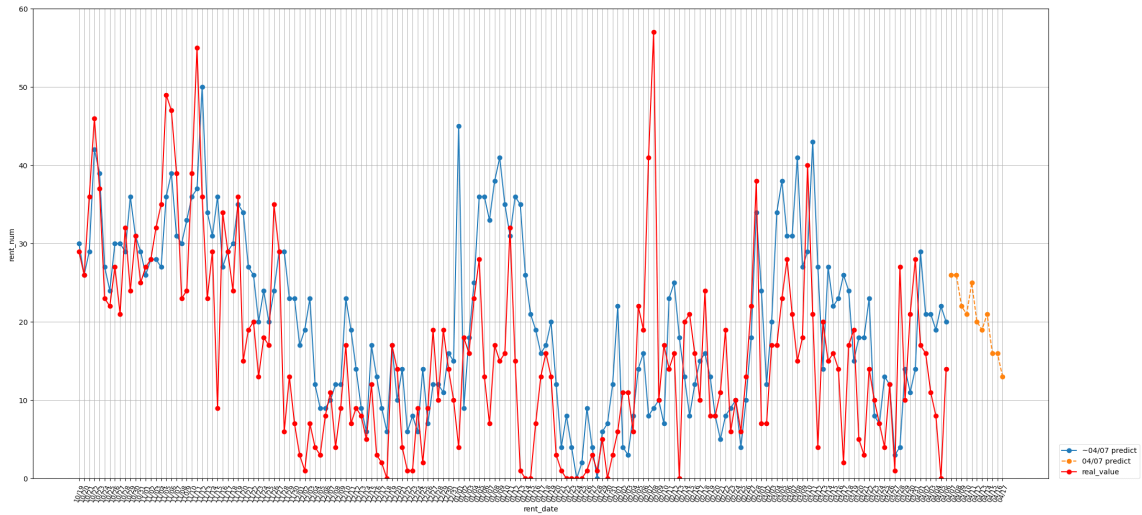


그림 4-11. 미래 대여 수 예측 결과

날짜	R2 Score	MAE	MSE	RMSE
1/10	0.1212	8.8222	145.8222	12.0756
1/26	0.2092	8.39	134.21	11.5849
2/21	0.1328	8.5476	145.7380	12.0722
4/6	0.006	8.8235	144.8705	12.03

표 4-6. 미래 대여 수 예측 성능평가

4.7.5 서비스 적용

본 연구를 기반으로 업체에서 실제로 적용할 수 있도록 플라스크를 이용하여 웹에서 볼 수 있도록 구현하였다. 플라스크 프레임워크는 파이썬 기반의 웹프레임워크이다. 플라스크는 모든 것을 처음부터 다 만들지 않아도 웹 사이트를 구축하는 데 사용할 수 있는 라이브러리를 제공한다. 플라스크는 가볍고 간단하여 외부 라이브러리에 의존하지 않는다[58]. 다음 <그림4-12>는 실제 업체에서 사용하고 있는 웹 사이트 화면을 보여준 것이다.

오늘어디가

10월 예측 조회

기간별 예측 조회

오늘어디가 조회

☛ 오늘어디가 > 조회하기

예측 기준일 2023-06-16

날짜	수량(대)		매출(원)	
	예측	실제	예측	실제
2023-06-16	22	26	60,045	111,797
2023-06-17	15	21	42,207	44,599
2023-06-18	16	23	17,436	65,599
2023-06-19	17	18	29,392	66,000
2023-06-20	16	11	28,347	69,600
2023-06-21	12	8	19,395	26,000
2023-06-22	13	7	9,603	53,400
2023-06-23	12	14	15,041	66,000
2023-06-24	12	21	42,521	85,900
2023-06-25	14	3	35,129	0
2023-06-26	6	9	0	36,400
합계	155	161	299,116	625,295
평균	14	15	27,192	56,845

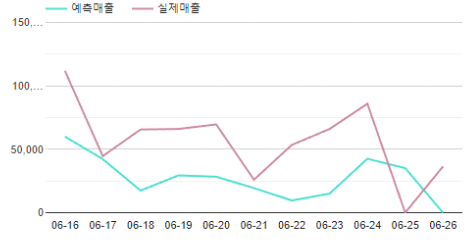
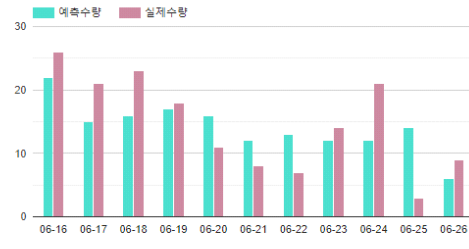


그림 4-12. 실제 서비스 화면

V. 결론 및 토의

본 연구에서는 제주 전 지역의 마이크로 모빌리티의 실제 대여자료를 이용하여 데이터를 분석하였다. 겨울철보다 여름철에 상대적으로 대여량이 더 많았으며, 여름에는 일사량, 강수량, 습도의 영향을 더 받았다. 반면, 겨울철에는 여름철보다 온도의 영향을 많이 받았다. 비가 오지 않은 맑은 날에 15℃~25℃ 사이에 가장 많은 대여 수를 보여주었다. 주말 여부에 따른 분석 결과 평균 대여량은 각각 54대와 52대로 큰 차이를 보이지는 않았다. 이를 바탕으로 대여 수 예측에 미치는 특징을 찾아내었고, 본 연구에서는 여러 가지 모델을 테스트한 결과 상위 모델인 Extra Trees Regressor, LGBM Regressor, CatBoost Regressor 모델을 사용해 마이크로 모빌리티의 대여 수를 예측해 보았다. 또한, 이 모델들을 결합한 Voting Regressor 모델을 이용하여 대여 수를 예측하는 방법을 제안했다. R2 score를 봤을 때 Voting Regressor, CatBoost, ExtraTrees, LightGBM 순서로 성능이 좋았고, 단일 모델 중 가장 성능이 높았던 CatBoost Regressor 모델과 비교해서 Voting Regressor 모델이 0.86% 더 좋은 성능을 보였다. Voting Regressor 모델이 단일 모델을 사용했을 때보다 나은 성능을 보여준 것을 알 수 있었다.

현재는 제주도 전체지역에 대한 마이크로 모빌리티의 수요를 예측한 것이다. 지역을 나눠서 어떤 곳에 수요가 더 많은지도 비교하고 분석하는 것이 필요할 것으로 판단된다. 향후 미래의 모빌리티 수요량을 예측할 때 폭우나 폭설과 같은 환경 정보를 수집할 수 있다면 더욱 정확한 예측을 할 수 있을 것으로 판단된다. 월드컵과 같은 이벤트 일정 또는 선거일이나 대체 공휴일과 같은 특수일 정보를 확인하여 예측에 사용할 수도 있다. 현재 사용하고 있는 데이터의 기간은 사업 초기기간을 포함한 약 3년간의 데이터이다. 또한, 사업 중간 기간에도 전체 유통되는 마이크로 모빌리티의 수가 일정하지 않았고, 이것은 대여 수에 영향을 미치는 것으로 보인다. 실제 대여 수도 매일매일 편차가 심한 것으로 보아 비즈니스 환경이 안정화된다면 더욱 신뢰할 수 있는 예측을 할 수 있을 것으로 판단된다.

향후 다른 모델들을 조합하여 최적의 조합을 찾기 위한 모델 연구도 필요하다.

현재는 제주도 전체지역에 대한 수요를 예측했지만, 이후의 연구에서 지역을 세부적으로 나누어 지역별로 예측하는 연구를 할 수 있다. 또한 10일 후뿐만 아니라 2주 후 또는 한 달 후까지 예측하는 연구를 진행할 수 있다. 이에 따라 마이크로 모빌리티를 적절하게 분산, 배분한다면 사용자들의 만족도 제고와 판매량 증가를 기대할 수 있다.

참 고 문 헌

- [1] Kang, S.; Hyun, B. A study on Key Factors of Delivery Service Quality Affecting Online Shopping Mall Loyalty: Focusing on Coupang' s Rocket Delivery. *Korean Mark. Rev.* 2021, 36, 49-77.
- [2] [Delivery War of Money] Coupang Shakes the Plate . . . Dreaming of a Korean Version of Amazon. 2020. Available online: [https://www.inews24.com/view/1313451\(2020.11.04\)](https://www.inews24.com/view/1313451(2020.11.04)).
- [3] 쿠팡 뉴스룸[웹사이트]. (2023.11.22.).URL: <https://news.coupang.com/archives/29293/>
- [4] Shibayama, T.; Emberger, G. New mobility services: Taxonomy, innovation and the role of ICTs. *Transp. Policy* 2020, doi.org/10.1016/j.tranpol.2020.05.024.
- [5] Global research News. (2023). Personal Mobility Devices Market Size, Share & Trends Analysis Report By Product (Walking Aids, Wheelchairs, Scooters), By Region (North America, Europe, APAC, Latin America, MEA), And Segment Forecasts, 2023 - 2030.<https://www.linkedin.com/pulse/personal-mobility-devices-market-2023-global-industry-rui3c/>
- [6] (2016),I.PersonalMobility...Whatisthehomeworkfor' eco-friendlytransportation' ?<http://www.greenpostkorea.co.kr/news/articleView.html?idxno=129834,2023>.
- [7] 김정흠, 서수석, and 이종호. "인터넷 쇼핑몰의 물류서비스 품질이 고객만족도와 충성도에 미치는 영향." *경영교육연구* 26.1 (2011): 443-461.
- [8] Park, J.; Lee, T. Obstacles that the Development of the Courier Service in Relation to Delivery War of Social Commerce with a Case Study on Rocket Delivery of Coupang. *Korean Manag. Consult. Rev.* 2016, 16, 173-184
- [9] 최유나, and 이도형. "이커머스 (E-commerce) 기업의 물류관리 과정에

- 관한 연구-Fulfillment 중심으로." *한중사회과학연구* 65 (2022): 106-127.
- [10] 김형원 기자, e커머스 물류센터 핵심기술 'AI·머신러닝'<https://it.chosun.com/news/articleView.html?idxno=2021072802118>
- [11] 쿠팡 뉴스룸 테크[웹사이트].(2023.11.22.).URL: <https://news.coupang.com/tech/>
- [12] 최재호(2019). 공유 전동킥보드 중심으로 살펴보는 마이크로모빌리티, K DB 미래전략연구소 산업기술리서치센터, 제 768 호, 37-53
- [13] Bozzi, A. D., & Aguilera, A. (2021). Shared E-scooters: A review of uses, health and environmental impacts, and policy implications of a new micro-mobility service. *Sustainability*, 13(16), 8676.
- [14] 이웅열 and 김승인. (2021). 전동킥보드 공유 서비스의 사용자 경험에 관한 고찰 -킥고잉과 라임을 중심으로-. *디지털융복합연구*, 19(2), 425-431.
- [15] Shaheen, S.; Cohen, A. Shared Micromobility Policy Toolkit: Docked and Dockless Bike and Scooter Sharing. Institute of Transportation Studies, Berkeley. 2019. Available online: <https://escholarship.org/uc/item/00k897b5> (accessed on 10 June 2020).
- [16] Ishmael, Chad R., et al. "An early look at operative orthopaedic injuries associated with electric scooter accidents: bringing high-energy trauma to a wider audience." *JBJS* 102.5 (2020): e18.
- [17] Sebastian, Libina Rose, Sheeba Babu, and Jubilant J. Kizhakkethottam. "Challenges with big data mining: A review." 2015 International Conference on Soft-Computing and Networks Security (ICSNS). IEEE, 2015.
- [18] Mahalakshmi, Ganapathy, S. Sridevi, and Shyamsundar Rajaram. "A survey on forecasting of time series data." 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16). IEEE, 2016.
- [19] Esling, Philippe, and Carlos Agon. "Time-series data mining." *ACM Computing Surveys (CSUR)* 45.1 (2012): 1-34.

- [20] Mekanik, F., et al. "Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes." *Journal of Hydrology* 503 (2013): 11–21.
- [21] Masini, Ricardo P., Marcelo C. Medeiros, and Eduardo F. Mendes. "Machine learning advances for time series forecasting." *Journal of economic surveys* 37.1 (2023): 76–111.
- [22] Kourou, Konstantina, et al. "Machine learning applications in cancer prognosis and prediction." *Computational and structural biotechnology journal* 13 (2015): 8–17.
- [23] Breiman, Leo. "Random forests." *Machine learning* 45 (2001): 5–32.
- [24] Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* 2006, 63, 3-42.
- [25] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* 2017, 30, 3149–3157.
- [26] Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *arXiv* 2018, arXiv:1706.09516.
- [27] Breiman, Leo. "Bagging predictors." *Machine learning* 24 (1996): 123–140.
- [28] Kearns, Michael, and Leslie Valiant. "Cryptographic limitations on learning boolean formulae and finite automata." *Journal of the ACM (JACM)* 41.1 (1994): 67–95.
- [29] Galatoulas, N.F.; Genikomsakis, K.N.; Ioakimidis, C.S. Analysis of potential demand and costs for the business development of an electric vehicle sharing service. *Sustain. Cities Soc.* 2018, 42, 148-161.
- [30] Ai, Y.; Li, Z.; Gan, M.; Zhang, Y.; Yu, D.; Chen, W.; Ju, Y. A deep learning approach on short-term spatiotemporal distribution forecasting

of dockless bike-sharing system. *Neural Comput. Appl.* 2019, 31, 1665-1677.

[31] Jafari, S.; Byun, Y.C. Prediction of the Battery State Using the Digital Twin Framework Based on the Battery Management System. *IEEE Access* 2022, 10, 124685-124696.

[32] El-Assi, W.; Salah Mahmoud, M.; Nurul Habib, K. Effects of built environment and weather on bike sharing demand: A station level analysis of commercial bike sharing in Toronto. *Transportation* 2017, 44, 589-613.

[33] Mathew, J.K.; Liu, M.; Bullock, D.M. Impact of weather on shared electric scooter utilization. In *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, 27-30 October 2019; pp. 4512-4516.

[34] Tin Tin, S.; Woodward, A.; Robinson, E.; Ameratunga, S. Temporal, seasonal and weather effects on cycle volume: An ecological study. *Environ. Health* 2012, 11, 1-9.

[35] Brandenburg, C.; Matzarakis, A.; Arnberger, A. Weather and cycling—a first approach to the effects of weather conditions on cycling. *Meteorol. Appl. J. Forecast. Pract. Appl. Train. Tech. Model.* 2007, 14, 61-67.

[36] Chen, L.; Zhang, D.; Wang, L.; Yang, D.; Ma, X.; Li, S.; Wu, Z.; Pan, G.; Nguyen, T.M.T.; Jakubowicz, J. Dynamic cluster-based over-demand prediction in bike sharing systems. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Heidelberg, Germany, 12-16 September 2016; pp. 841-852.

[37] Gebhart, K.; Noland, R.B. The impact of weather conditions on bike share trips in Washington, DC. *Transportation* 2014, 41, 1205-1225.

[38] Yang, Z.; Hu, J.; Shu, Y.; Cheng, P.; Chen, J.; Moscibroda, T. Mobility modeling and prediction in bike-sharing systems. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Application*

- s, and Services, Singapore, 26-30 June 2016; pp. 165-178.
- [39] Mehdizadeh Dastjerdi, A.; Morency, C. Bike-sharing demand prediction at community level under COVID-19 using deep learning. *Sensors* 2022, 22, 1060.
- [40] Chatterjee, S.; Byun, Y.C. A Synthetic Data Generation Technique for Enhancement of Prediction Accuracy of Electric Vehicles Demand. *Sensors* 2023, 23, 594.
- [41] Chatterjee, S.; Byun, Y.C. Generating Time-Series Data Using Generative Adversarial Networks for Mobility Demand Prediction. *Comput. Mater. Contin.* 2023, doi.org/10.32604/cmc.2023.032843.
- [42] Kim, S.; Lee, U.; Lee, I.; Kang, N. Idle vehicle relocation strategy through deep learning for shared autonomous electric vehicle system optimization. *J. Clean. Prod.* 2022, 333, 130055.
- [43] Yi, Z.; Liu, X.C.; Wei, R. Electric vehicle demand estimation and charging station allocation using urban informatics. *Transp. Res. Part Transp. Environ.* 2022, 106, 103264.
- [44] Zwick, F.; Axhausen, K.W. Ride-pooling demand prediction: A spatiotemporal assessment in Germany. *J. Transp. Geogr.* 2022, 100, 103307.
- [45] Chen, S.; Luc, N.M. RRMSE Voting Regressor: A weighting function based improvement to ensemble regression. *arXiv* 2022, arXiv:2207.04837.
- [46] Erdebilli, B.; Devrim-İltenbaş, B. Ensemble Voting Regression Based on Machine Learning for Predicting Medical Waste: A Case from Turkey. *Mathematics* 2022, 10, 2466.
- [47] Phyo, P.P.; Byun, Y.C.; Park, N. Short-term energy forecasting using machine-learning-based ensemble voting regression. *Symmetry* 2022, 14, 160.
- [48] Lim, H.; Chung, K. Development of demand forecasting model for s

- eoul shared bicycle. J. Korea Contents Assoc. 2019, 19, 132-140.
- [49] Yun, S.; Min, K.; Ko, H. Spatio-Temporal Patterns of a Public Bike Sharing System in Seoul-Focusing on Yeouido District. J. Korean Inst. Landsc. Archit. 2020, 48, 1-14.
- [50] Do, M.; Noh, Y.S. Analysis of the affecting factors on the bike-sharing demand focused on Daejeon City. KSCE J. Civ. Environ. Eng. Res. 2014, 34, 1517-1524.
- [51] Lee, H.; Baek, K.; Chung, J.H.; Kim, J. Factors affecting heterogeneity in willingness to use e-scooter sharing services. Transp. Res. Part D Transp. Environ. 2021, 92, 102751.
- [52] Min, J.; Mun, H.; Lee, Y. Demand forecast for public bicycles ("Tashu") in Daejeon using random forest, Kor. Infor. Sci. Soc. 2017, Vol. 2017 No.06, 969-971.
- [53] Ham, S.W.; Cho, J.H.; Park, S.; Kim, D.K. Spatiotemporal demand prediction model for e-scooter sharing services with latent feature and deep learning. Transp. Res. Rec. 2021, 2675, 34-43.
- [54] Ren, Y.; Zhang, L.; Suganthan, P.N. Ensemble classification and regression-recent developments, applications and future directions. IEEE Comput. Intell. Mag. 2016, 11, 41-53.
- [55] 위영민, 송경빈, and 주성관. "결정계수 기반의 데이터 마이닝을 이용한 특수일 최대 전력 수요 예측." 대한전기학회 학술대회 논문집 (2008): 552-553.
- [56] R. Mufid, A. Basofi, M. U. H. Al Rasyid, I. F. Rochimansyah and A. rokhim, "Design an MVC Model using Python for Flask Framework Development," 2019 International Electronics Symposium (IES), Surabaya, Indonesia, 2019, pp. 214-219, doi: 10.1109/ELECSYM.2019.8901656.

Ensemble Model for Micro Mobility Demand Prediction

JiYoung Ko

Department of Computer Engineering
The Graduate School
Jeju National University

Abstract

In recent years, companies in the retail market have shown considerable interest in delivery systems that can increase consumer satisfaction in order to stay ahead of the competition. The rise of shared mobility as a fast way to travel short distances has given rise to a variety of rental companies. To make it easier for consumers to rent from anywhere, these companies distribute their shared mobility across multiple locations, which means that the shared mobility must be pre-positioned at each location.

The problem is that it is difficult for companies to quickly prepare for shared mobility if they do not know in advance how many rentals will occur at any given time and in any given location. This study analyzes the factors that affect the number of shared e-mobility rentals and forecasts the number of rentals, so that e-mobility can be distributed and customers can be served in a timely manner. In this study, we use real-world data of electric mobility provided by Company E.

We experimented with demand prediction of electric mobility through machine learning models. We used the Extra Trees Regressor, Cat Boost Regressor, and LightGBM models to predict the demand of electric mobility and compared the performance of the models. We also studied how t

o reduce the error by using the voting method, which is an ensemble technique. We analyzed the correlation of various features that affect demand and found that humidity, precipitation, and insolation are highly correlated with demand. When evaluating the performance, the R^2 score was 0.7629 when using the Voting Regressor, which is better than the single model. By utilizing the results of this study, it is expected that electric mobility can be appropriately deployed in advance and user satisfaction will increase. It is expected that companies will be able to provide better shared electric mobility rental services by efficiently managing equipment and human resources.