

K-means 클러스터링을 이용한 지역별 전동 모빌리티 수요 예측

박세준*, 변영철**

Electric Mobility Demand Prediction by Region using K-means Clustering

Se-Joon Park*, Yung-Cheol Byun**

이 논문은 2021학년도 제주대학교 교육·연구 및 학생지도비 지원에 의해서 연구되었음

요 약

경쟁에서 우위를 선점하고 고객에게 양질의 서비스를 제공하기 위해서는 고객이 원하는 시간과 장소에 장치를 적절히 배치하는 것이 매우 중요하다. 이는 해외 A사의 퀵 배송, 국내 C사의 로켓배송 등에서도 알 수 있다. 따라서 본 연구는 과거의 지역별 고객의 전동 모빌리티 수요 데이터를 기반으로 미래에 있을 수 있는 수요를 보다 정확히 예측하기 위한 것이다. 예측의 정확도가 낮으면 적재적소에 고객들이 서비스를 이용하지 못하는 문제점이 발생할 수 있다. 이를 해결하기 위해 지역에 따라 전동 모빌리티 수요를 예측하고 지역을 어떻게 나눌지에 대한 연구가 필요하다. 본 연구에서는 K-means 클러스터링 알고리즘을 이용하여 데이터를 나누어 학습함으로써 그렇지 않은 경우에 비해 나은 성능을 얻을 수 있음을 알 수 있었다. 실험결과, 지역을 나누기 전에 약 70%의 정확도를 보였고, 클러스터링 방법으로 지역을 나눌 경우 약 90%의 정확도로 성능이 개선됨을 알 수 있었다.

Abstract

Proper placement of equipment at the time and place desired by the customer is very important in order to stay ahead of the competition and provide quality service to the customer. This can also be seen in Amazon's express delivery and Coupang's rocket delivery. Therefore, this study is to more accurately predict potential future demand based on past regional customer electric mobility demand data. If the accuracy of prediction is low, there may be a problem that customers cannot use the service in the right place. To solve this problem, it is necessary to predict electric mobility demand by region and study how to divide the region. In this research, we use the K-means clustering algorithm to divide and learn the data. It was possible to know that superior performance was obtained compared to the case where it was not. As a result of the experiment, it was found that the accuracy was about 70% before the regions were divided, and the performance was improved with the accuracy of about 90% when the area was divided by the clustering method.

Keywords

machine learning, time series data, pattern recognition, regression, forecasting, clustering, data smoothing

* 제주대학교 컴퓨터공학과 석사과정
- ORCID: <https://orcid.org/0000-0003-2220-9318>
** 제주대학교 컴퓨터공학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0002-1579-5323>

· Received: Jul. 31, 2021, Revised: Oct. 18, 2021, Accepted: Oct. 21, 2021
· Corresponding Author: Yung-Cheol Byun
Dept. of Computer Engineering, Jeju National University, Jejudaehakro 102,
Jeju, jeju Special Self-Governing Province, Korea
Tel.: +82-64-754-3657, Email: ppakse6@gmail.com

I. 서 론

4차 산업혁명이 발전함에 따라 기업들은 고객의 니즈를 충족시키기 위해 이전의 데이터를 모아놓은 빅 데이터를 어떻게 사용해야할지 연구하기 시작하였다. 이러한 빅 데이터는 사업을 계속 이어오던 기업에게 해당될 뿐, 새롭게 사업을 시작하는 스타트업들에게는 데이터의 수가 적고 수입의 형태가 불안정한 데이터가 대부분이다[1]. 질적, 양적으로 부족한 데이터를 가지고 고객의 니즈를 맞추는 것은 쉽지 않다.

본 연구는 제주도 내의 전동 모빌리티 서비스를 제공하는 기업의 데이터를 사용하였다. 전동 모빌리티를 이용하고 싶은 고객입장에서는 이용을 원할 때 시간과 장소의 제약을 받지 않고 서비스를 이용하는 것에 만족을 느낀다. 하지만 유한한 전동 모빌리티로 인하여 이용하지 못하게 된다면 고객은 불편함을 느낄 수 있으며 기업의 입장에서는 이러한 경우가 지속되면 수요가 줄게 되어 매출에도 영향을 줄 수 있다[2]. 그렇기 때문에 기업은 고객이 원하는 장소와 시간에 배치하는 것이 중요하다.

수요 예측을 통해 배치하는 것에는 오차가 생길 수 있다. 전동 모빌리티가 부족하게 배치가 되어 있다면 수요가 남는 곳에서 가져와 채울 수 있다. 하지만 두 지역 간의 거리가 멀다면 고객에게 서비스를 제공하는 시간이 길어지게 된다. 예를 들어 국내 C사의 온라인 물에는 로켓배송이라는 시스템이 있다. 미리 수요를 예측하여 해당 지역의 허브에 물품을 배치하여 먼 지역이라도 당일 배송이 가능하도록 한다. 이러한 아이디어에서 본 연구 또한 제주도 내 지역을 분할하여 각 지역의 중심지에 수요가 남는 전동 모빌리티를 배치하여 오차가 생기는 곳에 빠르게 재배치를 함으로써 서비스를 이어나갈 수 있다[3].

본 논문은 제주도 내 지역을 여러 곳으로 분할할 때 각 지역의 중심지를 기준으로 가까운 거리의 대여소를 묶고 전동 모빌리티 수요 예측을 진행하였다. 각 지역 중심지에 전동 모빌리티를 저장하고, 지역별 예측을 통해 수요가 예측보다 많은 지역이 생기면 해당 지역 중심지에 있는 전동 모빌리티를 가져와 배치하는 것으로 빠른 서비스를 제공할 수

있다. 또한, 지역을 분할하여 학습을 진행하였을 때의 결과를 지역 분할 전의 결과와 비교하였을 때, 예측 정확도가 증가한 것을 확인하였다.

II. 관련 연구 및 기술

2.1 회귀 모델을 이용한 시계열 데이터 예측

지도학습은 주어진 데이터 변수들의 패턴을 파악하고 그것을 기반으로 예측을 한다. 지도학습은 크게 분류와 회귀로 나뉜다. 분류는 예측하고자 하는 타겟이 클래스일 경우, 학습한 데이터를 토대로 해당 클래스를 예측하는 것이다. 회귀를 분류와는 다르게 예측하고자 하는 타겟이 수치일 경우를 회귀라고 한다[4][5]. 시계열 데이터에서 예측은 시간의 흐름에 따라 변수들이 변하는 패턴을 찾아내는 것이다. 회귀 모델을 이용한 시계열 데이터 예측은 시간이 지나며 달라지는 변수들의 패턴을 파악하여 타겟 데이터의 수치를 예측하는 것이다[6]. 본 연구는 일별 전동 모빌리티의 수요를 예측하는 연구로서, 일별 날씨정보, 주말과 주중을 나눈 정보, 대여 수 등의 하루마다 변하는 정보들을 기반으로 타겟 데이터인 일별 대여 수를 예측하였다.

2.2 데이터 스무딩을 통한 데이터 불안정 해소

데이터 스무딩은 패턴 파악에 어려움을 주는 격차가 큰 데이터 간의 차이를 줄여주는 데에 도움을 준다. 지도학습의 데이터는 일정한 패턴의 형태를 찾고 그것을 기반으로 예측을 하게 되는데, 데이터 간의 격차가 크게 되면 패턴을 인식하지 못하게 된다. 지도학습이 인식할 수 있는 패턴을 형성시키기 위해 데이터 스무딩은 전체적인 패턴은 유지하되 지도학습이 패턴인식을 할 수 있도록 데이터 간의 격차를 최소화한다[7].

2.3 클러스터링을 이용한 예측

클러스터링은 데이터들의 특징을 기반으로 집단을 구성하고 해당 집단의 중심점을 기준으로 군집화 시킨다. 비슷한 특징을 가진 데이터들을 묶음

로써 패턴 인식에 도움을 줄 수 있다[8]. 비슷한 유형의 데이터들 간의 차이를 줄여주고 반대로 다른 유형의 데이터들과는 차이를 둔다. 본 논문에서는 클러스터링 기법 중 대표적인 K-means 알고리즘을 이용하였다.

III. K-means 클러스터링을 이용한 지역별 수요 예측 방법

본 연구는 회귀 모델과 데이터 스무딩, 클러스터링을 이용하여 소규모 시계열 데이터를 지역별로 나눠 예측 정확도를 개선하는 방법에 대해 소개한다. 그림 1은 제안하는 방법을 전체 순서도로 표현하였다.

연구에 사용한 데이터는 2019년 4월부터 제주도 내에서 서비스를 시작한 전동 모빌리티 서비스 기업의 수요 데이터를 사용하였다. 또한, 기상청의 날씨 데이터를 가져와 일평균 기온과 강수 여부, 주말과 주중을 나누는 등 전동 모빌리티를 이용할 때 영향을 미치는 외적인 요소들을 추가하였다[9].

표 1은 연구 데이터 정보에 대한 표이다. 데이터는 제주도 내에서 전동 모빌리티 서비스를 제공하고 있는 기업인 E사의 데이터를 사용하였다.

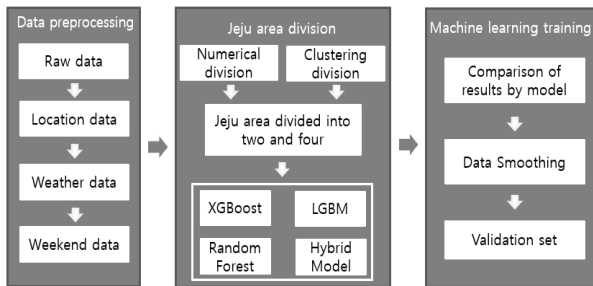


그림 1. 전체 순서도
Fig. 1. Flow diagram of the study

표 1. 데이터 정보
Table 1. Information of data

| | |
|-------------------------|------------|
| Data source | EVPASS |
| Total number of records | 11,971 |
| Start date of data | 2019-04-16 |
| End date of data | 2021-06-11 |

총 데이터 인스턴스 수는 데이터 수집 기간 동안 서비스 이용자들의 이용 건수이며 수집 기간은 2019년 4월 16일부터 2021년 6월11일까지, 총 717일 간의 데이터이다. 본 연구는 일별로 전동 모빌리티 수요량을 묶어 일별 수요량을 예측하였다.

그림 2는 전동 모빌리티 일일 수요량을 보여준다. X축은 날짜, Y축은 해당 날짜의 하루 총 수요량을 나타낸다. 서비스 초기 적었던 수요가 어느 정도 시간이 지난 후 늘어났다. 이는 일반적으로 스타트업 업체에게 나타나는 수요 현상으로 생각할 수 있다. 또한, 수요가 일정하지 않고 날마다 큰 차이를 보였다. 본 연구는 이러한 불안정한 데이터를 스무딩을 통해 안정화시켜 연구를 진행하였다. 데이터 스무딩의 윈도우 크기는 11로 설정했는데, 윈도우 크기는 해당 날짜 그 전후 날들을 묶어 데이터를 표준화한다. 윈도우 크기 11은 중심 날짜 전후 5일씩을 묶은 것을 의미하며 그림 3에서 데이터 스무딩의 결과를 보여준다.

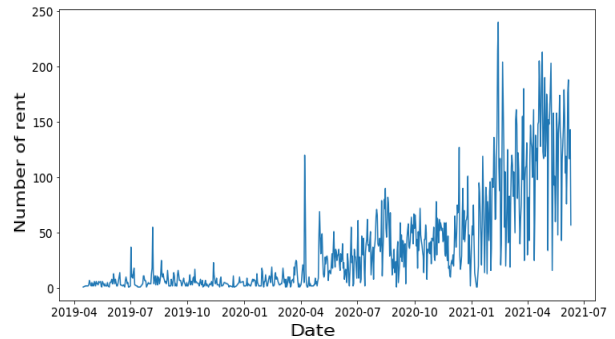


그림 2. 전동 모빌리티 일일 수요량
Fig. 2. Daily demand for electric mobility

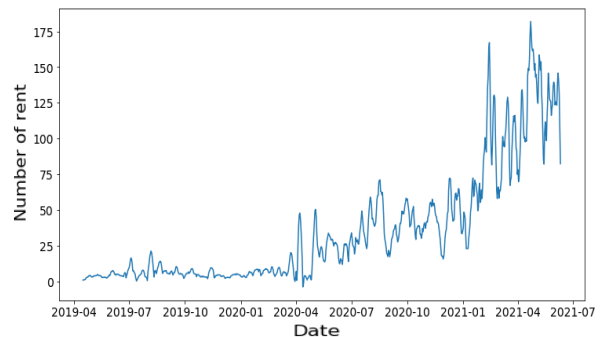


그림 3. 데이터 스무딩 이후 전동 모빌리티 일일 수요량
Fig. 3. Daily demand for electric mobility after data smoothing

그림 3은 데이터 스무딩 이후 전동 모빌리티 일일 수요량을 나타낸다. 그림 2와 같이 일일 수요량의 격차가 큰 불안정한 데이터에서 전체적인 수요의 형태를 유지하면서 데이터 간의 격차를 해소하는 역할을 한다.

현재 연구에 사용하는 row 데이터의 수는 717개로, row 데이터의 수를 추가하기 위해 제주도내의 수요를 지역별로 나눴다. 크게 제주도내 지역을 전체, 2분할, 4분할하여 각 분할별 수요 예측 결과를 비교한다. 지역을 분할하는 데 있어서 위도와 적도를 기준으로 나눈 결과와 K-means 알고리즘을 이용한 클러스터링 분할의 결과 또한 비교한다.

그림 4는 경도를 기준으로 동과 서로 제주도 지역을 2분할 한 것이다. 데이터 수집 기간 동안 제주도 내에 각 대여소 지점별 총 대여 수에 따라 점의 크기 차이를 나타낸 그림이다. 제주 공항 근처와 애월, 성산 등 유명 관광지의 전동 모빌리티 대여 수가 대체적으로 많은 것을 확인할 수 있다.

그림 5는 위도와 경도를 기준으로 제주시와 서귀포시를 나누고 제주시를 추가로 3분할한다. 서귀포시의 전동 모빌리티 수요량이 제주시에 비해 현저히 적어 먼저 분할하고 제주시를 동, 서, 중앙으로 분할한다.

그림 6은 클러스터링의 대표적인 기법 중 하나인 K-means 알고리즘을 이용하여 제주도를 2지역 분할한 것이다. 클러스터링 결과의 중심점은 빨간색 점이며, 해당 점을 기준으로 가까운 지역끼리 묶는 것이 클러스터링의 특징이다. 경도를 기준으로 나눈 2분할과 비슷하게 크게 동, 서로 나뉜다.

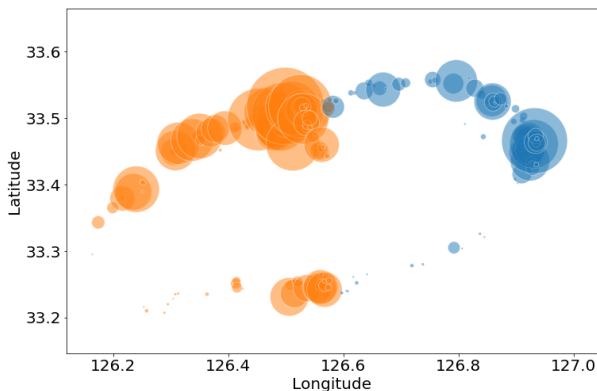


그림 4. 경도를 기준으로 한 2지역 분할
Fig. 4. Two regional division based on longitude

그림 7 또한 K-means 알고리즘을 이용하여 제주도를 4지역 분할한 것이다. 4개의 빨간색 점은 각 지역의 중심점을 나타내며 2분할과 마찬가지로 위도와 경도를 기준으로 나누었을 때와 비슷하게 나뉜다.

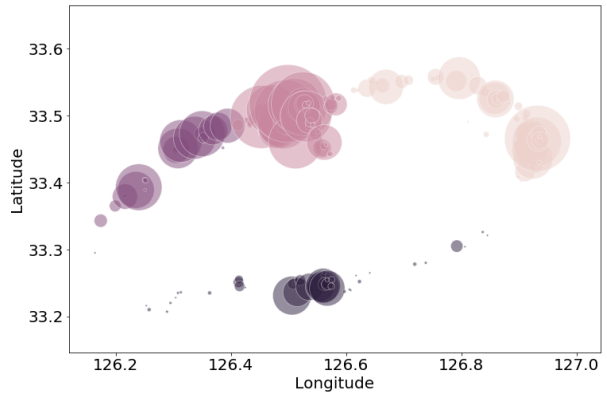


그림 5. 위도와 경도를 기준으로 한 4지역 분할
Fig. 5. Four regional division based on latitude and longitude

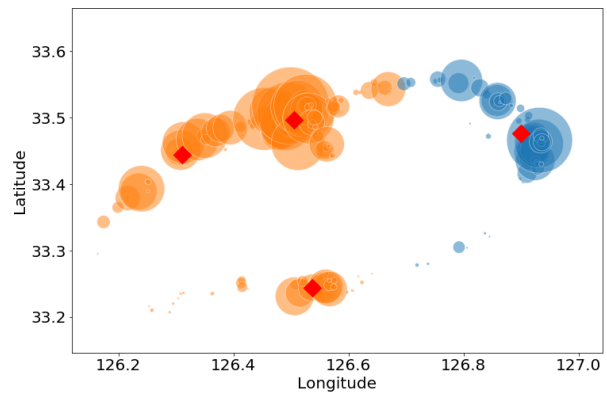


그림 6. 클러스터링을 이용한 2지역 분할
Fig. 6. Two regional division using clustering

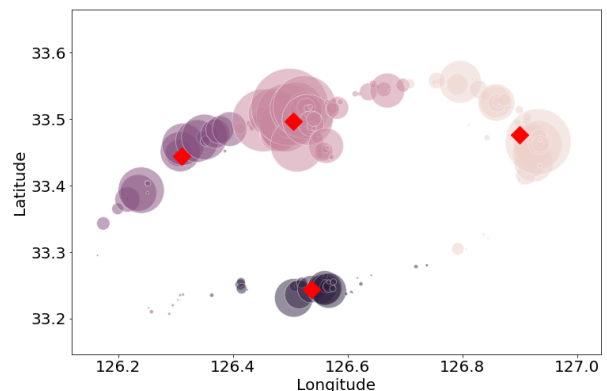


그림 7. 클러스터링을 이용한 4지역 분할
Fig. 7. Four regional division using clustering

4.2장 성능 평가에서는 임의적으로 위도와 경도로 나눈 결과와 K-means 클러스터링 알고리즘을 이용하여 나눈 결과를 비교한다.

IV. 연구 환경 및 성능 평가

4.1 연구 환경

본 연구는 프로그래밍 언어 Python 3.7.6 버전을 이용했고 Google chrome 브라우저와 Jupyter notebook에서 연구를 진행하였다. 표 3은 본 논문의 연구 환경에 대한 정보를 담고 있다.

4.2 성능 평가

성능 평가를 위해 XGBoost(eXtreme Gradient Boost)와 LGBM(Light Gradient Boost Model), Random Forest, 그리고 두 개의 모델이 합쳐진 하이브리드 모델(XGBoost+Random Forest)의 회귀 모델을 이용하여 결과를 비교한다. XGBoost와 LGBM은 부스팅(Boosting) 모델로 한 결정트리에서 학습하여 나온 결과의 오류에 가중치를 조절하여 계속해서 모델이 오차를 줄여주는 방식이다. 본 연구에서는 부스팅 기반 모델을 여러 번 사용했을 때 결과의 차이가 크지 않아 안정적 모습을 보였고 예측 정확도 또한 높았기 때문에 채택하였다.

Random Forest는 여러 개의 결정트리 이용하기 때문에 결과에 차이가 크다는 단점이 있지만, 부스팅과는 다른 방식으로 학습이 진행되고 학습 속도가 빠르다는 장점이 있기 때문에 다른 부스팅 모델들과 결과를 비교하였다[10]. 평가 지표는 R2 score, RMSE를 이용했으며, R2 score는 실제 값과 예측 값의 차이를 퍼센티지로 나타낸 정확도이다. RMSE (Root Mean Square Error)는 평균 제곱근 오차로써 결과를 평가하는 지표이다. 또한, 학습 데이터와 테스트 데이터를 각 80%, 20%로 분할하였다.

표 4는 제주도를 지역 분할하기 전 데이터를 이용하여 수요 예측한 결과이다. LGBM의 결과는 R2 score 0.75, RMSE 22.54로 가장 높았고, Random Forest의 결과가 R2 score 0.68, RMSE 25.73으로 가장 낮았다. 평균 70%의 정확도를 보였다.

표 5는 제주도 2지역을 나눌 때 사용한 두 개의 분할 방법과 모델별 결과를 나타낸 표이다. XGBoost의 성능이 다른 모델들의 비해 좋았고, Random Forest를 제외한 다른 결과들은 90% 이상의 정확도를 보였다. 클러스터링을 이용했을 때 결과와 경도, 위도로 나눈 결과보다 1% 정도 정확도의 증가를 보였다. 또한, 분할하기 전 수요 예측 정확도보다 평균 20% 높은 정확도를 보였다.

표 6은 제주도를 4지역으로 나눌 때 사용한 두 개의 분할 방법과 모델별 결과를 나타낸 표이다. 전체적으로 2지역 분할보다 결과는 떨어지지만 XGBoost에서는 90% 이상의 정확도를 보였다. 그림 8에서 분할, 모델별 R2 score 결과에 대한 전체적인 그래프를 보여준다.

표 3. 연구 환경

Table 3. Research environment

| | |
|-----------------------|------------------------------------|
| Programming language | Python 3.7.6 |
| Operating system | Window 10 pro 64bit |
| Browser | Google chrome |
| Library and framework | Jupyter notebook |
| CPU | Intel(R) Core(TM) i5-9600k@3.70GHz |
| Memory | 16GB |

표 4. 분할 전 모델별 수요 예측 결과

Table 4. Demand forecast results by model before partitioning

| Regression model | Evaluation metrics | |
|------------------|--------------------|-------|
| | R2 Score | RMSE |
| XGBoost | 0.71 | 24.91 |
| LGBM | 0.75 | 22.54 |
| Random forest | 0.68 | 25.73 |
| XGB + RF | 0.68 | 25.49 |

표 5. 제주도 2지역의 분할 방법 및 모델별 결과

Table 5. Partitioning method for Jeju Island two regions and results by model

| Regression model | As of latitude and longitude | | Clustering | |
|------------------|------------------------------|-------|------------|-------|
| | R2 score | RMSE | R2 score | RMSE |
| XGBoost | 0.94 | 9.50 | 0.94 | 9.38 |
| LGBM | 0.9 | 11.46 | 0.91 | 10.92 |
| Random forest | 0.87 | 13.21 | 0.89 | 11.88 |
| XGB + RF | 0.9 | 11.37 | 0.9 | 10.97 |

표 6. 제주도 4지역의 분할 방법 및 모델별 결과
Table 6. Partitioning method for Jeju Island four regions and results by model

| Regression model | As of latitude and longitude | | Clustering | |
|------------------|------------------------------|-------|------------|-------|
| | R2 score | RMSE | R2 score | RMSE |
| XGBoost | 0.91 | 10.68 | 0.9 | 11.05 |
| LGBM | 0.87 | 13.15 | 0.89 | 11.63 |
| Random forest | 0.88 | 12.84 | 0.89 | 11.45 |
| XGB + RF | 0.85 | 14.52 | 0.86 | 13.47 |

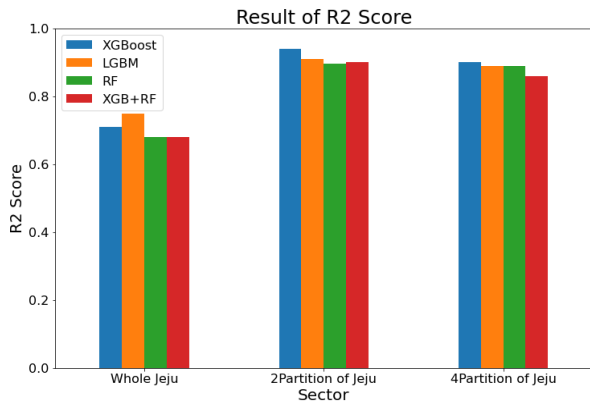


그림 8. 분할, 모델별 R2 score 결과
Fig. 8. R2 score result by model and partition

4.3 검증 세트

지도학습은 학습 데이터와 그것을 기반으로 예측한 결과인 테스트 데이터로 구성되어있다. 그러나 학습 데이터가 테스트 데이터에만 초점을 두고 학습을 하게 되면 테스트 데이터에 대해서는 오차가 줄어들지만 실제 데이터나 다른 데이터를 예측하려고 하면 오차가 증가하게 되는 경우가 있다. 이러한 경우를 지도학습의 과적합이라고 하는데, 학습 데이터와 테스트 데이터 사이에 검증 세트를 추가하여 예측을 더하여 결과를 비교하는 것으로 과적합을 예방할 수 있다[11].

본 연구는 검증세트를 이용한 예측에서 60%의 학습 데이터와 20%의 검증세트, 20%의 테스트 데이터를 이용하였다. 검증 세트의 결과를 표 7에서 보여준다.

4.4 특징 중요도

특징 중요도는 보통 Feature importance라고도 불

린다. 지도학습이 학습된 데이터를 통해 예측을 할 때 각각의 특징들이 결과에 미치는 영향을 수치로 표현한 것이다[12]. 일별로 정렬된 시계열 데이터에서 특징 중요도는 매일 하루를 나타내는 정보 중 패턴이 뚜렷하고 세분화된 수치를 가지고 있는 특징일수록 특징 중요도가 높게 나타났다.

그림 9는 본 연구의 특징 중요도를 나타낸 그래프이다. 결과에 가장 영향을 많이 끼친 특징은 기온으로 전동 모빌리티를 이용한 고객들은 기온의 영향에 따라 이용률이 뚜렷하다는 것을 알 수 있다. 다음으로 주말과 주중을 나타내는 특징, 강수 여부, 지역에 따른 수요 등의 순서로 결과에 영향을 끼치는 것을 확인하였다[13].

V. 결 론

본 논문은 적은 량의 데이터와 수요의 차이 폭이 큰 불안정한 데이터를 이용하여 지도학습의 회귀 예측을 진행하였다. 데이터들 간의 불안정한 패턴을 데이터 스무딩을 통해 전체적인 수요 패턴은 유지하되 안정화하였다.

표 7. 검증 세트의 R2 score 결과
Table 7. R2 score result of validation set

| Regression model | As of latitude and longitude | | Clustering | |
|------------------|------------------------------|------------|------------|------------|
| | 2 division | 4 division | 2 division | 4 division |
| XGBoost | 0.9 | 0.9 | 0.92 | 0.89 |
| LGBM | 0.92 | 0.87 | 0.91 | 0.89 |
| Random forest | 0.85 | 0.86 | 0.93 | 0.88 |
| XGB + RF | 0.84 | 0.85 | 0.85 | 0.87 |

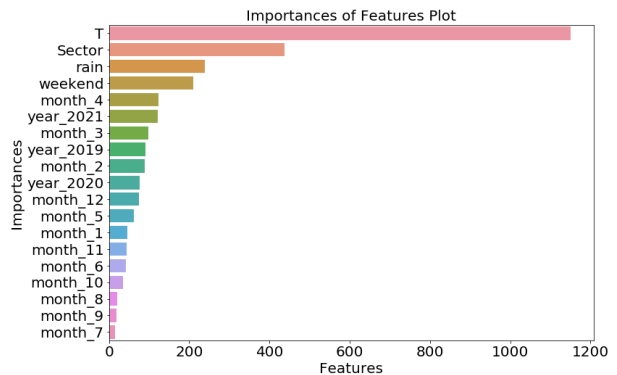


그림 9. 특징 중요도
Fig. 9. Feature importance

클러스터링을 이용하여 비슷한 특징을 가진 데이터들은 묶고, 다른 특징을 가진 데이터를 분리하여 수요 예측을 진행하였다. 이를 통해 예측 결과 또한 큰 폭으로 증가했으며 최소 85% 이상의 수요 예측 정확도를 보였다. 또한, 제주도 지역 분할 이전의 정확도 결과와 비교해보았을 때 평균 20% 높은 정확도의 증가를 보였다.

References

- [1] Sunwoo Kim and Kangmin Kim, "A Study on the Startup Growth Stage in Korea", Korean Society of Business Venturing, Asia-Pacific Journal of Business Venturing and Entrepreneurship, Vol. 15, No. 2, pp. 127-135, Apr. 2020.
- [2] G. Z. de Rubens, L. Noel, J. Kester, and B. K. Sovacool, "The market case for electric mobility: Investigating electric vehicle business models for mass adoption", Elsevier, Energy, Vol. 194, Mar. 2020. <https://doi.org/10.1016/j.energy.2019.116841>.
- [3] Khan, P.W and Byun, Y.C, "Genetic Algorithm Based Optimized Feature Engineering and Hybrid Machine Learning for Effective Energy Consumption Prediction", IEEE Access, Vol. 8, pp. 196274-196286, Oct. 2020. <https://doi.org/10.1109/ACCESS.2020.3034101>.
- [4] Liaw Andy and Matthew Wiener, "Classification and regression by randomForest", R news, Vol. 2, No. 3, pp, 18-22, Dec. 2002.
- [5] Loh, W.Y, "Classification and regression trees", Wiley Interdiscip. Rev. Data Min. Knowl. Discov, Vol. 1, pp. 14-23, Jan. 2011. <https://doi.org/10.1002/widm.8>.
- [6] WEI William WS, "Time series analysis", In: The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2. 2006. <https://doi.org/10.1093/oxfordhb/9780199934898.013.0022>.
- [7] Guo Diansheng and Xi Zhu, "Origin-destination flow data smoothing and mapping", IEEE, Transactions on Visualization and Computer Graphics, Vol. 20, No. 12, pp. 2043-2052, Nov, 2014. <https://doi.org/10.1109/TVCG.2014.2346271>.
- [8] A. Goia, C. May, and G. Fusai, "Functional clustering and linear regression for peak load forecasting", Elsevier, International Journal of Forecasting, Vol. 26, No. 4, pp. 700-711, Oct, 2010. <https://doi.org/10.1016/j.ijforecast.2009.05.015>.
- [9] Korea Meteorological Administration, <https://www.weather.go.kr>. [accessed: Jun, 05, 2021]
- [10] Park Se-Joon and Yung-Cheol Byun, "Improving Recommendation Accuracy based on Machine Learning using Multi-Dimensional Features of Word2Vec", Journal of KIIT, Vol. 19, No. 3 pp. 9-14, Mar. 2021. <http://dx.doi.org/10.14801/jkiit.2021.19.3.9>.
- [11] S. E. Bleeker, et al., "External validation is necessary in prediction research:: A clinical example", Elsevier, Journal of clinical epidemiology, Vol. 56, No. 9, pp. 826-832, Sep. 2003. [https://doi.org/10.1016/S0895-4356\(03\)00207-5](https://doi.org/10.1016/S0895-4356(03)00207-5).
- [12] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, "Permutation importance: a corrected feature importance measure", Oxford Academic, Bioinformatics, Vol. 26, No. 10, pp. 1340-1347, May 2010. <https://doi.org/10.1093/bioinformatics/btq134>.
- [13] Park Se-Joon, Chul-Ung Kang, and Yung-Cheol Byun, "Extreme Gradient Boosting for Recommendation System by Transforming Product Classification into Regression Based on Multi-Dimensional Word2Vec", Symmetry, Vol. 15, No. 5, 758, Apr. 2021. <https://doi.org/10.3390/sym13050758>.

저자소개

박 세 준 (Se-Joon Park)



2020년 2월 : 제주대학교 컴퓨터
공학과(공학사)
2020년 3월 ~ 현재 : 제주대학교
컴퓨터 공학과 석사과정
관심분야 : 인공지능(머신러닝,
딥러닝), 자연어 처리, 인지과학

변 영 철 (Yung-Cheol Byun)



1993년 2월 : 제주대학교
정보공학과 (공학사)
1995년 8월 : 연세대학교
컴퓨터과학과 (공학석사)
2001년 8월 : 연세대학교
컴퓨터과학과 (공학박사)
1998년 ~ 2001년 : 삼성전자, SDS

전문강사

2001년 ~ 2003년 : 한국전자통신연구원 선임연구원

2012년 ~ 2014년 : University of Florida 방문 교수

2003년 ~ 현재 : 제주대학교 교수

관심분야 : 딥러닝, 패턴인식, 시계열 데이터 처리, 추천
시스템, 지식발견, 딥러닝 기반 신재생에너지 시스템,
블록체인