

2024년도 공공기관 용역과제
AI개발 수행내역서

과제명	기상정보를 활용한 공공자전거 수요분석 및 예측
담당자	정지용

2025년 01월 07일

1. 사업과제 : 기상정보를 활용한 공공자전거 수요분석 및 예측

2. 개요 및 현황

2.1 추진배경 및 목적

- 장기간 축적된 데이터베이스를 기반으로하여 인공지능 기반의 예측모델에 대한 수요가 점진적 증대 예상
- 공공자전거 이용의 증가와 함께, 기상정보를 활용한 수요 예측의 필요성이 점차 커지고 있음.
특히, 기후 변화와 계절적 요인이 자전거 이용 패턴에 미치는 영향이 크기 때문에, 이를 분석하는 것이 중요함.
- 기상정보는 자전거 이용에 직접적인 영향을 미치는 요소로, 기온, 강수량, 일사량 등 다양한 기상 변수들이 자전거 이용자의 선택에 영향을 줄 수 있음. 따라서, 이러한 데이터를 기반으로 한 예측 모델 개발이 필요함.
- 공공자전거 시스템의 효율성을 높이기 위해, 기상정보를 활용하여 수요를 예측하고, 자전거 대여소의 운영 및 관리에 대한 전략을 수립하고자 함. 이를 통해 자전거 이용자들에게 더 나은 서비스를 제공할 수 있을 것으로 기대됨.
- 기상정보를 활용한 수요분석을 통해, 자전거 이용의 활성화 및 대중교통과의 연계성을 강화하고, 도시 내 친환경 교통수단으로서의 공공자전거의 역할을 증대시키고자 함.
- 향후, 기상정보를 활용한 수요 예측 모델을 기반으로, 자전거 이용 외에도 대중교통, 보행자 이동 등 다양한 교통수단에 대한 분석으로 확대할 계획임. 이를 통해 도시의 교통 체계 전반에 대한 통합적인 접근이 가능해질 것으로 기대됨.

2.2 과제 범위

과제 구분		내용
시각화	실시간 환경계측정보 연계 및 시각화	환경계측정보 실시간 연계 모형 구현
		상황관리 대시보드 등 시각화 구현
		예측모델 시각화
		테스트

2.3 과제 추진 방법

1) 구축 대상 선정 기준

- 데이터 접근성 및 활용성
 - 데이터 수집 및 관리의 용이성
 - 정부 및 공공기관에서 이미 구축된 데이터베이스 활용 여부
 - 종속변수에 영향을 미치는 다양한 독립변수에 대한 정보 포함여부를 통한 모델학습의 유용성
- 예측모델 개발 효율성
 - 모델 학습 및 평가 과정 간소화를 위한 다른 환경 기초데이터에 비해 변수가 상대적으로 단순한 구조 여부
 - 개발된 모델을 통해 다른 데이터 수요 예측에 적용 가능 여부
- 환경문제 해결 기여도 및 경제성(변경)
 - 예측모델을 통해 환경관리에 상대적 기여도가 높은지 여부(ex. 오염도 저감, 에너지 절감 등)
 - 운영 효율성을 높여 비용 절감 효과 여부(자원 최적화, 유지보수 비용 절감)
 - 환경문제 해결을 통한 사회적 비용감소 효과 여부(대기 질 개선, 교통 혼잡 완화)

2) AI 예측 분석모델 적용 대상

환경관리 기능	수집 데이터	예측모델인자(독립변수)	AI예측 분석 대상
공공 자전거 수요	- 2023년 서울시 일별 공공자전거 이용 현황 데이터 - 2023년 일별 기상관측 데이터	- 기상 변수 : 평균기온, 상대습도, 강수량, 일사량, 적설량	- 공공자전거 수요 예측 - 평균기온, 상대습도, 일사량 등 상관관계 분석

3) AI 분석모델 구축 프로세스

- 데이터 수집
 - 서울열린데이터광장 서울시 공공자전거 이용정보(일별)
 - 기상자료 개방포털 서울시 기상관측 데이터(일별)
- 데이터 전처리
 - 데이터 표준화 및 전처리(결측치, 이상치 처리)
 - 변수를 표준화 및 정규화하여 모델학습 효율성을 향상.
 - 상관관계 분석을 통해 독립변수 간의 관계를 확인.
 - 특성 중요도를 통해 중요도가 낮은 독립변수 제거.
- 데이터 모델링
 - 모델 성능 평가를 위해 MSE, r-squared 등의 지표를 사용.
 - 모델 비교 및 최적 모델 도출 (XGBoost 채택)
- 데이터 예측
 - 다양한 기상요인에 따른 공공자전거 수요 예측
- 결과 시각화 및 분석
 - 모델의 정확도 평가를 위해 실제값과 예측값을 비교
 - 예측결과에 대한 분석 결과를 산점도로 시각화.
 - 실제값과 예측값 간의 차이를 바탕으로 잔차 분석 진행.
 - 기상요인에 따른 공공자전거 수요 예측을 통한 운영 효율성 향상 및 환경적 기여

연구개발 주요 결과물

1. 데이터 수집

- 서울시 공공자전거 일별 이용 현황 데이터(엑셀) : 2023년
- 서울시 기상관측 일별 기상 데이터(엑셀) : 2023년

지점	A	B	C	D	E	F	G	H
지점명	일시	평균기온(°C)	일강수량(mm)	평균 상대습도(%)	합계 일사량(MJ/m2)	일 최상적설(cm)		
108 서울	2023-01-01	-0.2		54.5	10.81			
108 서울	2023-01-02	-4.5		45.9	11.63			
108 서울	2023-01-03	-5		49	11.77			
108 서울	2023-01-04	-1.8		51.4	10.89			
108 서울	2023-01-05	-1.6		58.1	6.09			
108 서울	2023-01-06	0.6	3.9	71.9	8.78	3.6		
108 서울	2023-01-07	1.5	0.1	80.9	4.78	3.6		
108 서울	2023-01-08	1.3		69.3	9.37			
108 서울	2023-01-09	3.1		60.1	10.53			
108 서울	2023-01-10	1.6		59.9	11.33			
108 서울	2023-01-11	2.6		61.9	10.86			
108 서울	2023-01-12	5.9	0	46.6	10.88			
108 서울	2023-01-13	8.3	37.3	99.1	1.11			
108 서울	2023-01-14	6.6	1.6	92.9	1.84			
108 서울	2023-01-15	-0.2	3.2	93.9	1.32	0.6		
108 서울	2023-01-16	-4.6	0	63.1	12.95	0.3		
108 서울	2023-01-17	-3.7	0	85.1	5.27			
108 서울	2023-01-18	-1.9	0	60.5	12.71			
108 서울	2023-01-19	-0.4	1	69.3	7.63			
108 서울	2023-01-20	-3.3		52.4	13.49			
108 서울	2023-01-21	-4.5		55.4	12.37			
108 서울	2023-01-22	-0.6	0	62.8	4.41			
108 서울	2023-01-23	-0.8		66.6	10.15			
108 서울	2023-01-24	-14.7	0	51.9	14.08			
108 서울	2023-01-25	-11.3	0	45.1	14.08			
108 서울	2023-01-26	-6	0.8	80.1	4.51	4.5		
108 서울	2023-01-27	-7.3		56.5	14.1	1.7		
108 서울	2023-01-28	-6	0	54.9	12.58			
108 서울	2023-01-29	-1.8		58.3	9.79			
108 서울	2023-01-30	-0.9		55.9	14.17			
108 서울	2023-01-31	2.4		69.1	6.5			
108 서울	2023-02-01	0.9		68.4	13.24			
108 서울	2023-02-02	-2.4		50.3	10.94			
108 서울	2023-02-03	-1.6		47.4	13.54			
108 서울	2023-02-04	-0.2		49.8	13.23			
108 서울	2023-02-05	1.7		71	11.35			
108 서울	2023-02-06	3		72.3	10.69			
108 서울	2023-02-07	3.5		72.9	10.12			

대여일시	대여건수
2023-01-01	38,037
2023-01-02	56,609
2023-01-03	61,252
2023-01-04	67,721
2023-01-05	68,711
2023-01-06	63,125
2023-01-07	42,492
2023-01-08	44,620
2023-01-09	74,134
2023-01-10	80,604
2023-01-11	81,988
2023-01-12	87,595
2023-01-13	20,792
2023-01-14	25,896
2023-01-15	21,415
2023-01-16	61,311
2023-01-17	67,337
2023-01-18	71,361
2023-01-19	68,359
2023-01-20	51,398
2023-01-21	30,226
2023-01-22	22,752
2023-01-23	30,880
2023-01-24	15,265
2023-01-25	38,329
2023-01-26	24,340
2023-01-27	40,221
2023-01-28	35,890
2023-01-29	38,347
2023-01-30	65,611
2023-01-31	74,828
2023-02-01	72,476
2023-02-02	70,755
2023-02-03	72,693
2023-02-04	57,939
2023-02-05	53,806
2023-02-06	81,306
2023-02-07	86,528

그림 1

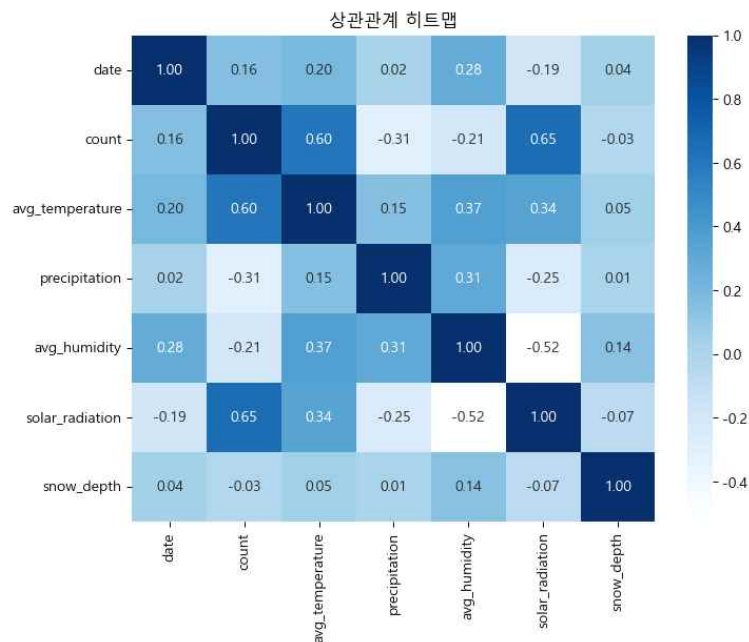
서울 열린데이터광장 공공데이터

그림 2

기상자료 개방포털
기상관측 데이터

2. 데이터 분석

2.1 데이터 상관관계(Heatmap)



2.2 탐색적 데이터 분석

○ 결측치 및 중복값 통계

Overview

Alerts8

Reproduction

Dataset statistics

Number of variables	6
Number of observations	365
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	38.3 KiB
Average record size in memory	107.4 B

Variable types

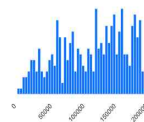
Numeric	5
DateTime	1

○ 주요 변수별 데이터 분포(Histogram)

count
Real number (R)

High correlation Unique

Distinct	365	Minimum	4992
Distinct (%)	100.0%	Maximum	211000
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	123028.2	Memory size	3.0 KiB

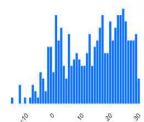


More details

avg_temperature
Real number (R)

High correlation

Distinct	225	Minimum	-14.7
Distinct (%)	61.6%	Maximum	30.9
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	41
Infinite (%)	0.0%	Negative (%)	11.2%
Mean	14.109041	Memory size	3.0 KiB

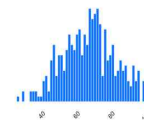


More details

avg_humidity
Real number (R)

High correlation

Distinct	250	Minimum	25.35
Distinct (%)	68.5%	Maximum	99.1
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	66.688356	Memory size	3.0 KiB

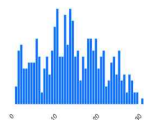


More details

solar_radiation
Real number (R)

High correlation

Distinct	337	Minimum	0.69
Distinct (%)	92.3%	Maximum	31.19
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	14.233562	Memory size	3.0 KiB



More details

○ 데이터 전처리

First rows

Last rows

Unnamed: 0						
		date	count	avg_temperature	avg_humidity	solar_radiation
0	0	2023-01-01	38037	-0.2	54.5	10.81
1	1	2023-01-02	56609	-4.5	45.9	11.63
2	2	2023-01-03	61252	-5.0	49.0	11.77
3	3	2023-01-04	67721	-1.8	51.4	10.89
4	4	2023-01-05	68711	-1.6	58.1	6.09
5	5	2023-01-06	63125	0.6	71.9	8.78
6	6	2023-01-07	42492	1.5	80.9	4.78
7	7	2023-01-08	44620	1.3	69.3	9.37
8	8	2023-01-09	74134	3.1	60.1	10.53
9	9	2023-01-10	80604	1.6	59.9	11.33

3. 데이터 학습 및 모델정의

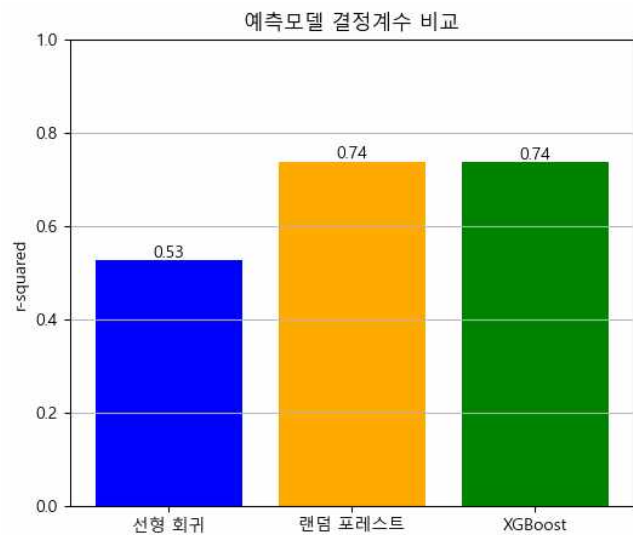
3.1 예측 모델 선정

- 결정계수 비교 : Ensemble 기법 중 하나인 XGBoost 모델 채택

```
# 모델 정의
models = {
    '선형 회귀': LinearRegression(),
    '랜덤 포레스트': RandomForestRegressor(random_state=42),
    'XGBoost': XGBRegressor(random_state=42)
}

# R-squared 점수 저장
r2_scores = {}

# 각 모델 학습 및 R-squared 점수 계산
for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    r2 = r2_score(y_test, y_pred)
    r2_scores[name] = r2
```



3.2 모델학습 및 학습 시각화

- 모델 학습

```
# 데이터 전처리
X = df[['avg_temperature', 'avg_humidity', 'solar_radiation']]
y = df['count']

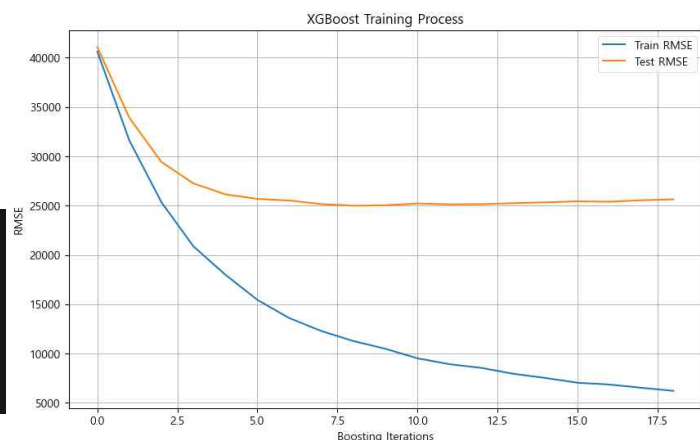
# 데이터 스케일링
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# 데이터 분할
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.3, random_state=42)

# XGBoost 모델 생성 및 학습
model = XGBRegressor(random_state=42)
model.fit(X_train, y_train)
```

- 학습과정 시각화

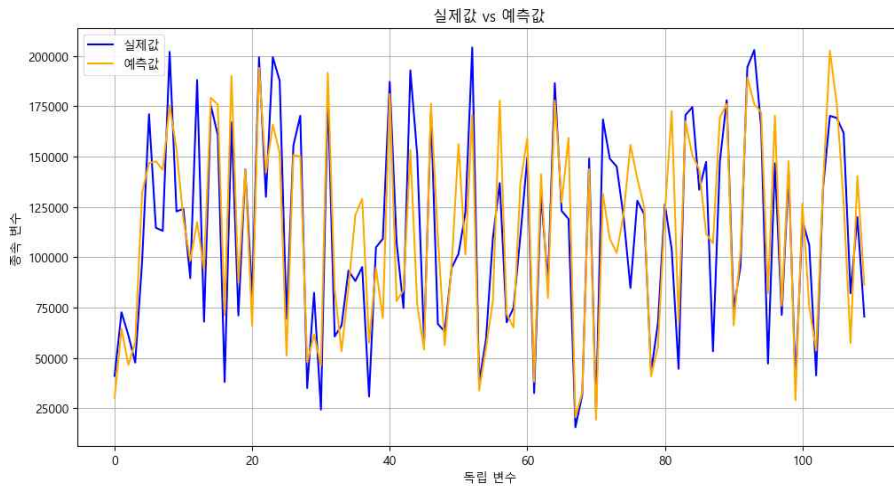
```
# 학습 곡선 시각화
plt.figure(figsize=(10, 6))
plt.plot(evals_result['train']['rmse'], label='Train RMSE')
plt.plot(evals_result['test']['rmse'], label='Test RMSE')
plt.xlabel('Boosting Iterations')
plt.ylabel('RMSE')
plt.title('XGBoost Training Process')
plt.legend()
plt.grid()
plt.show()
```



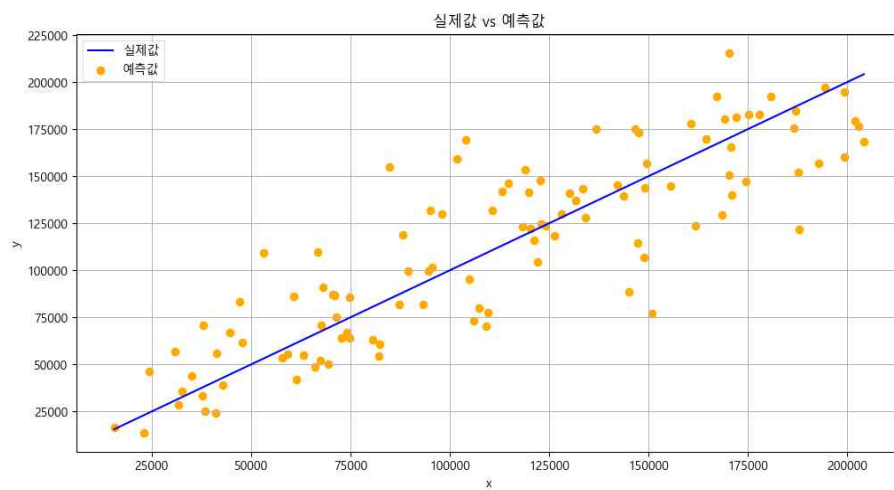
3.3 모델 예측

○ 예측값 vs 실제값 비교

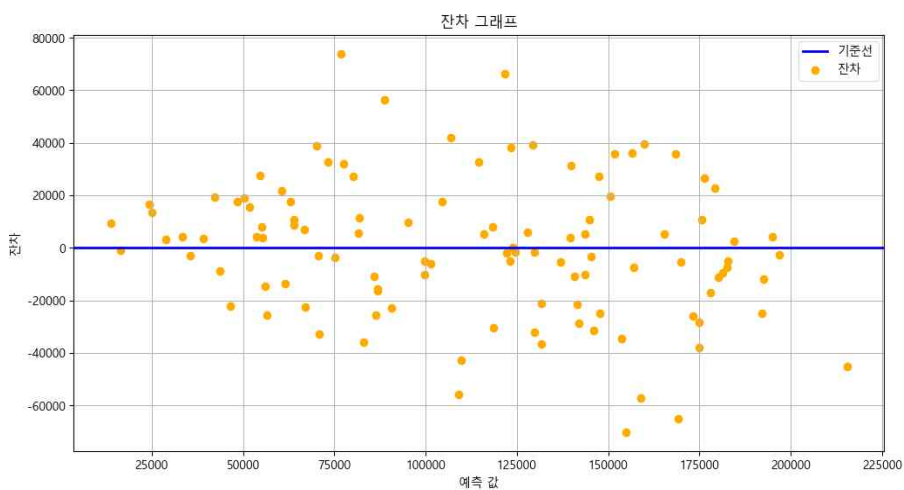
■ 선 그래프 비교



■ 산점도 분석



■ 잔차 분석



4. 프로토타이핑(화면)

4.1 모델 예측

- 기상요인에 따른 공공자전거 이용건수 예측

기상정보에 따른 이용건수 예측하기

평균 기온 (°C)

0.00

- +

평균 습도 (%)

0.00

- +

일사량 (MJ/m2)

0.00

- +

예측하기

4.2 예측결과

- 기상요인에 따른 공공자전거 이용건수 예측

날짜	이용건수	평균기온	상대습도	일사량
2023-01-13	20792	8.3	99.1	1.11

기상정보에 따른 이용건수 예측하기

평균 기온 (°C)

8.30

- +

평균 습도 (%)

99.10

- +

일사량 (MJ/m2)

1.11

- +

예측하기

예상되는 공공자전거 이용 건수: 20799건