전력 사용량 예측 AI 모델 개발

1. 개요

안정적이고 효율적인 에너지 공급을 위해 전력 사용량에 대한 정확한 예측이 필요하다. 전력 사용량 예측 시뮬레이션을 통한 효율적인 인공지능 알고리즘 발굴을 목표로 한다.

목표: 건물 정보와 시공간 정보를 활용하여 특정 시점의 전력 사용량을 예측하는 모델 개발

2. 데이터 정보

- 훈련 데이터: 100개 건물들의 2022년 6월 1일부터 2022년 8월 24일 까지의 데이터. 일시별 기온, 강수량, 풍속, 습도, 일조, 일사, 전력사용량(kWh) 포함.

- 빌딩 정보: 100개 건물에 대한 정보 (건물 번호, 건물 유형, 연면적, 냉방 면적, 태양광 용량, ESS 저장 용량, PCS 용량)

- 검증 데이터: 2022년 8월 25일부터 2022년 8월 31일까지의 데이터. 일시별 기온, 강수량, 풍속, 습도의 예보 정보

3. 데이터 전처리

(1) 초기 데이터 준비

datetime 형식 변환: datetime 열을 날짜-시간 형식으로 변환.

정렬 및 인덱스 재설정: 데이터의 정합성을 위해 building\_type 기준으로 그룹화하고 시간 순으로 정렬. 기존 인덱스는 제거.

(2) 요일 및 주말 여부 추가

weekday: 요일 정보 추가 (0: 월요일 ~ 6: 일요일).

is\_weekend: 주말 여부를 판단하여 True(주말) 또는 False(평일)로 표시.

변수 생성 이후 그래프를 그려서 요일별 전력 사용량을 확인하고 주말의 전력 사용량이 감소하지 않는 건물 종류에 대한 세부 분석을 수행함.

텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 라인, 폰트이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 라인, 폰트이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

자동 생성된 설명

(3) 건물 유형별 휴일 조건 설정 : 할인마트와 데이터 센터는 휴무일이 일반적인 주말과 다을 확인. 이를 반영하여 새로운 변수 holiday를 제작함.

Discount Mart: 매월 2주차와 4주차 일요일을 휴무일로 설정.

주차 계산: 날짜에서 주차(week\_of\_month)를 파생.

해당 조건에 맞는 일요일을 holiday = 1로 설정.

Data Center: 데이터 센터는 항상 운영되므로 모든 holiday 값을 0으로 설정.

기타 건물 유형: 주말과 평일의 전력 소비량(power\_consumption\_kWh) 평균 비교:

주말 평균 소비량이 평일 평균 소비량보다 작으면 주말을 holiday = 1로 설정, 그렇지 않으면 주말을 포함해 모든 holiday를 0으로 설정.

표1) 건물 별 holiday 수

|  |  |  |
| --- | --- | --- |
| **building\_type** | **holiday** | |
| **0** | **1** |
| Apartment | 16320 | 0 |
| Commercial | 11712 | 4608 |
| Data Center | 10200 | 0 |
| Department Store & Outlet | 16320 | 0 |
| Discount Mart | 15360 | 960 |
| Hospital | 11712 | 4608 |
| Hotel & Resort | 16320 | 0 |
| Knowledge Industry Center | 11712 | 4608 |
| Other Building | 30600 | 0 |
| Public | 11712 | 4608 |
| Research Center | 11712 | 4608 |
| University | 11712 | 4608 |

3. 모델 적용

1. 데이터 전처리 및 분리

훈련 데이터를 바탕으로, 전력 소비량 예측을 위해 필요한 피처(특징)와 타겟(목표) 변수를 분리했습니다.

피처 변수(X): 전력 소비량에 영향을 줄 것으로 예상되는 다양한 변수(기온, 풍속, 습도, 총 면적, 냉방 면적, 주차, 월, 시간대, 요일, 휴일 여부)로 구성되었습니다.

타겟 변수(y): 각 시간대의 전력 소비량을 나타내는 power\_consumption\_kWh 변수입니다.

훈련 데이터를 학습용(80%)과 검증용(20%)으로 분리하여 모델이 새로운 데이터에 대한 예측 능력을 평가할 수 있도록 준비했습니다. 이 과정은 train\_test\_split 함수를 사용해 실행되었으며, 재현성을 위해 random\_state를 설정했습니다.

2. 모델 학습 : 랜덤 포레스트, XGBoost 모델 사용

(1) 랜덤 포레스트 회귀(Random Forest Regressor)

(2) XGBoost 회귀

두가지 모델을 사용하여 전력 소비량 예측 모델을 학습시켰습니다. 훈련 데이터를 바탕으로 모델을 학습(fit 메서드)하여 각 시간대 전력 소비량을 예측하는 규칙을 학습했습니다.

3. 모델 평가

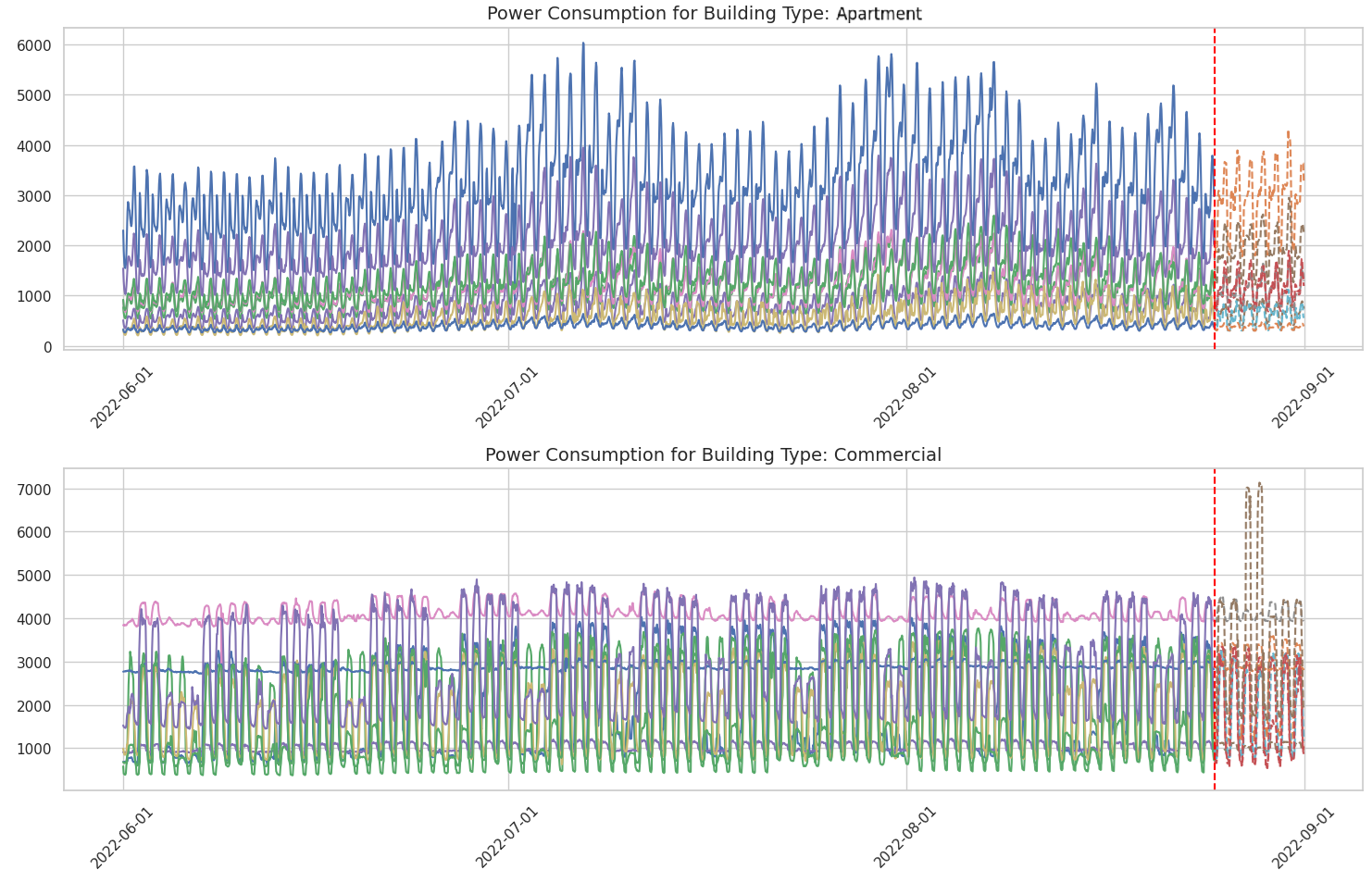
모델 성능은 검증 데이터(X\_val)를 활용해 예측값(y\_pred)을 계산한 후, 실제값(y\_val)과 비교하여 평가했습니다. 사용한 평가 지표는 Mean Squared Error(MSE): 예측값과 실제값 간의 평균 제곱 오차를 계산하여, 예측 정확도를 수치화하고 결정계수를 사용하였습니다.

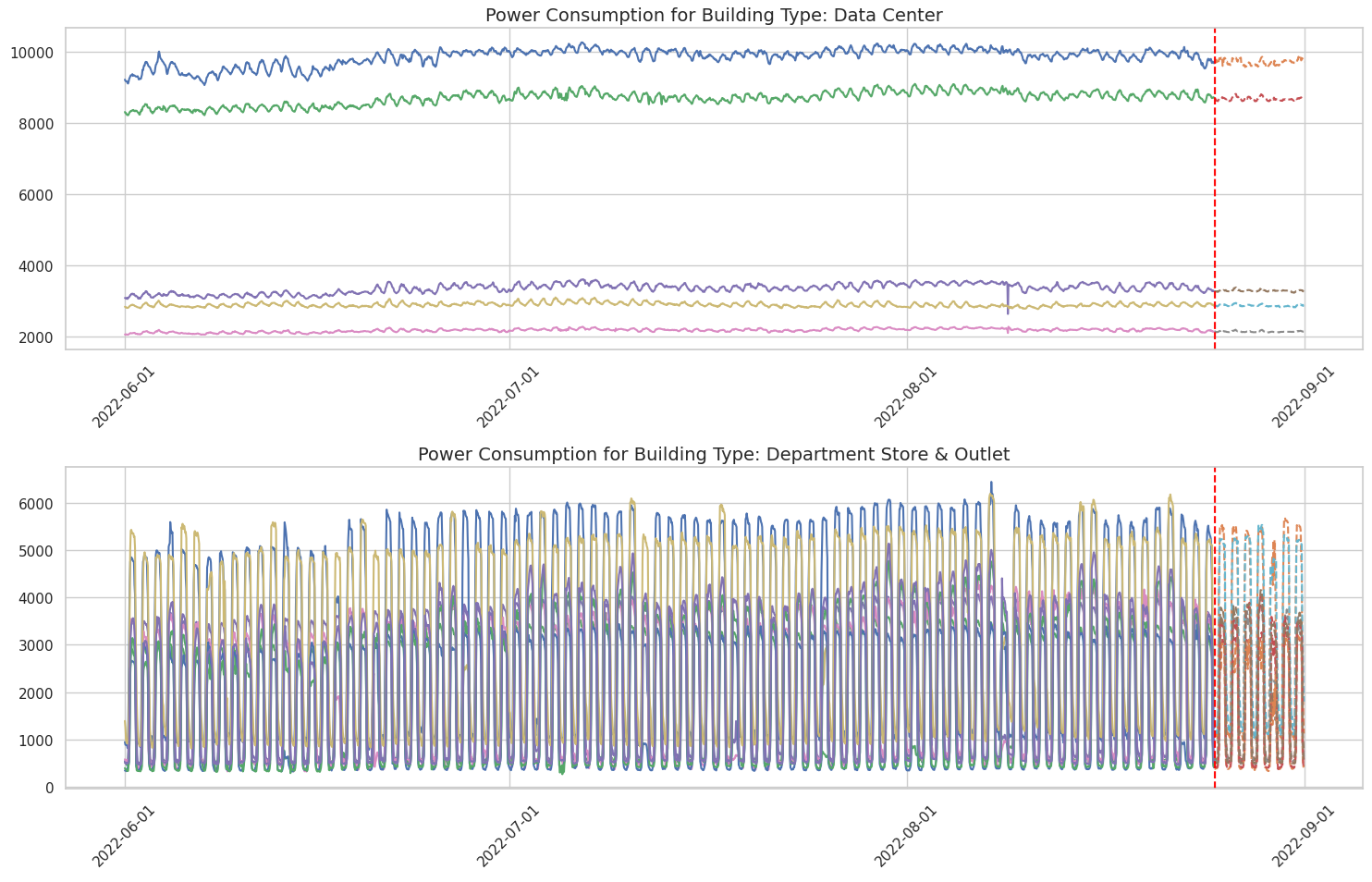
평과 결과 : 랜덤 포레스트가 더 좋은 결과를 보여 최종 모델로 랜덤 포레스트를 선택하였습니다.

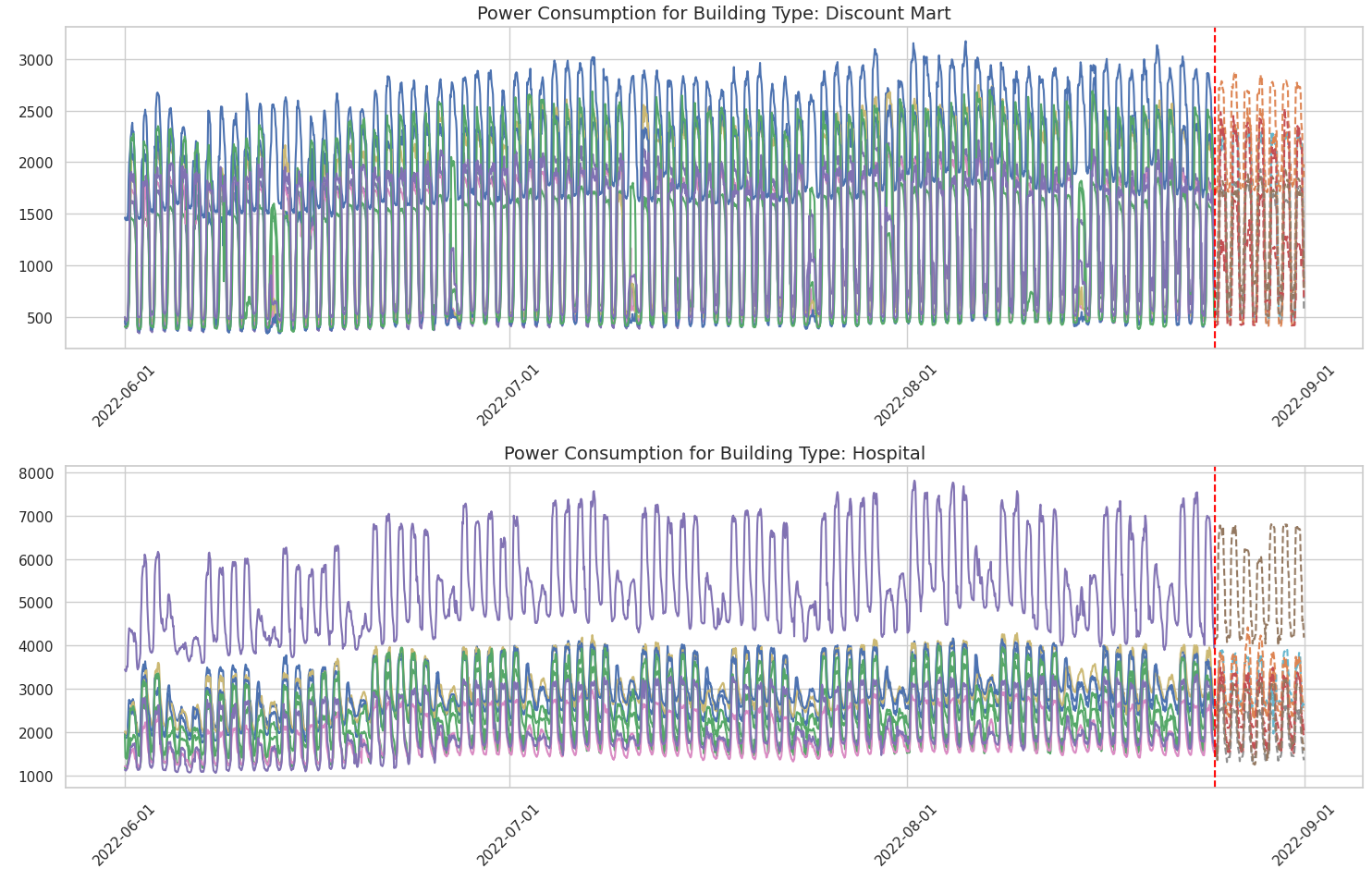
|  |  |  |
| --- | --- | --- |
| Model | MSE | R^2 |
| Random Forest | 30290.97 | 0.995062 |
| XGBosst | 150683 | 0.975434 |

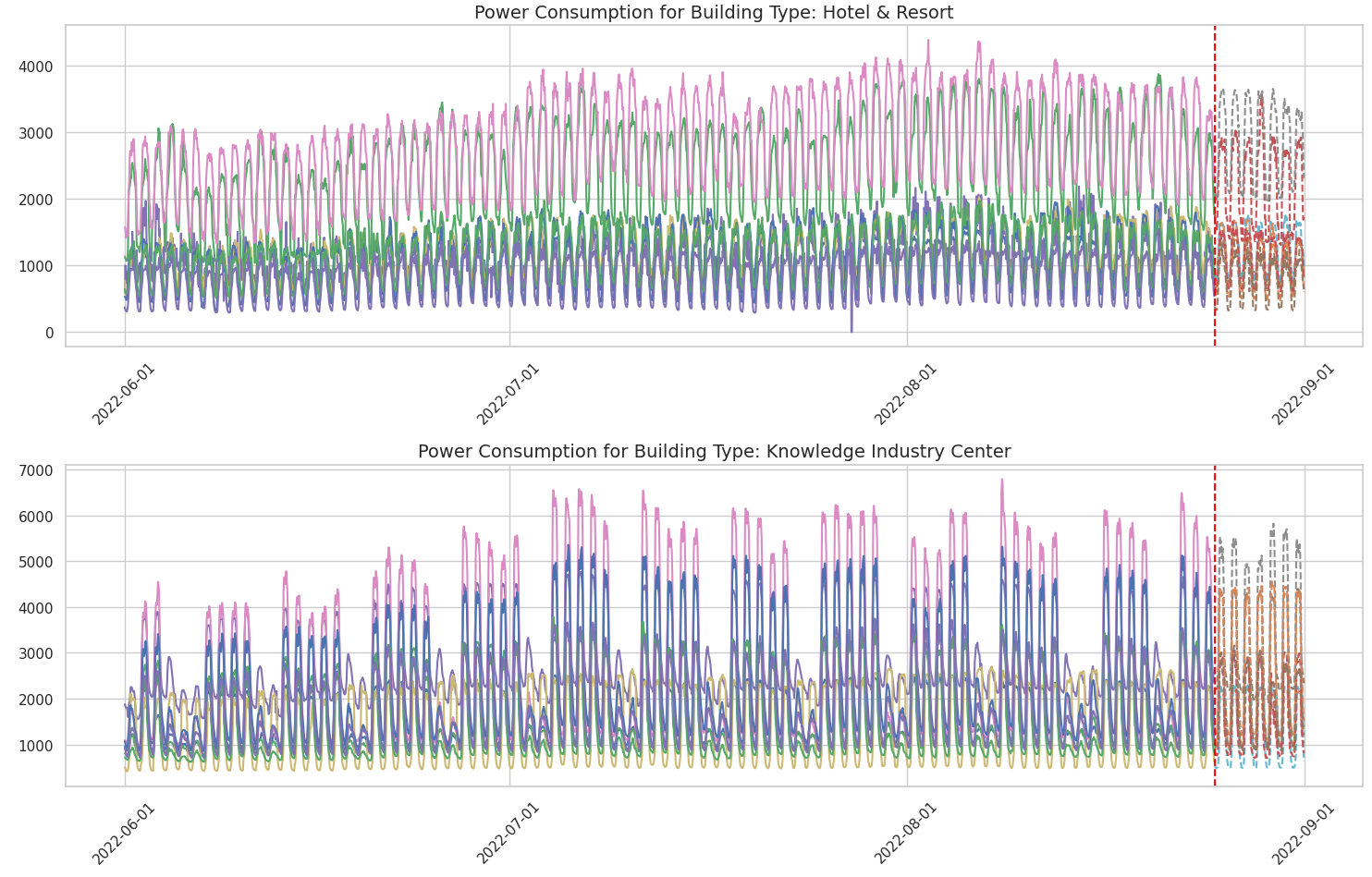
4. 모델 예측

테스트 데이터에 선택된 랜덤 포레스트 모델을 적합하여 결과를 출력하였습니다. 빨간 점선 이후의 값이 모델을 활용해 예측한 값입니다.

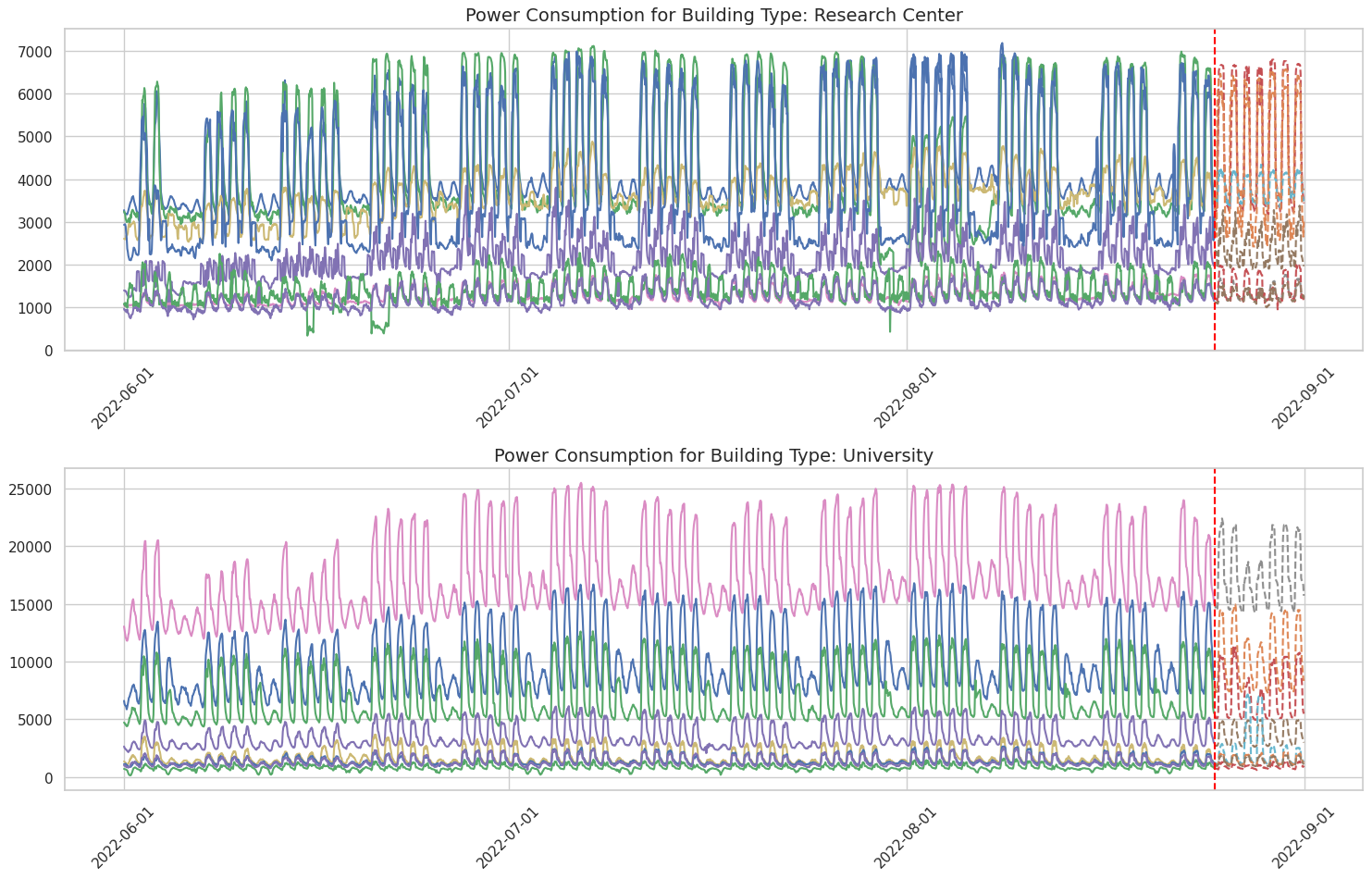












5. 결론

랜덤 포레스트 모델은 대부분의 건물 유형에서 전력 소비량을 잘 예측하며, 특히 계절성과 시간대와 같은 주요 요인을 잘 반영하였습니다. 다만, 일부 특이한 데이터(공휴일, 특정 건물 유형 등)에서는 추가적인 피처 엔지니어링이나 모델 개선이 필요합니다. 이러한 결과는 전력 사용량 관리나 에너지 최적화 전략에 실질적인 도움을 줄 수 있을 것으로 기대됩니다.

훈련 및 테스트 데이터 패턴:

건물 유형별로 전력 소비량 패턴이 크게 달랐으며, 특히 특정 시간대나 요일에 따라 소비량이 급격히 변화하는 경향이 관찰되었습니다. 또한 각 유형별로 휴무일 여부와 주기를 파악하는 것이 올바른 모델을 구축하는데 영향을 미치는 것을 확인할 수 있었습니다.