

산업 전력 데이터의 이상치 탐지 성능 향상 솔루션 구축

산업 전력 소비량의 이상치 검출 및 최적화를 통한 에너지 소비량 감축

EcoGA - 권우현(팀장), 박범도(팀원), 박준성(팀원), 정민성(팀원)

목차

1 팀원 소개_3P

2 프로젝트 개요_4P

- 주제 선정 배경
- 프로젝트 수행 배경 및 필요성

3 수행 과정_6P

4 프로젝트 내용_7P

- 프로젝트 구성
- 주요 기능
- 주요 기술

5 프로젝트 성과_11P

- 프로젝트 결과물

6 결론_14P

- 기대효과
- 활용방안

팀원 소개 (멘토: 이상금 교수님)



권우현

: 프로젝트 총괄 및
딥러닝 기반 모델 구축



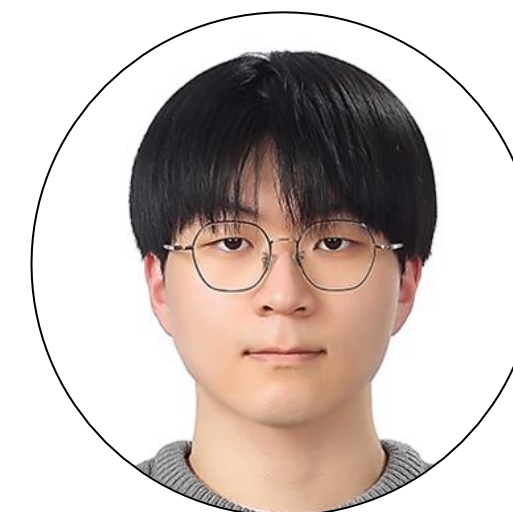
박범도

: FFT 기반
이상치 검출 모델 설계



박준성

: 데이터 수집 및
데이터 분석



정민성

: 데이터 분석 및
데이터 전처리

프로젝트 개요

주제 선정 배경

산업 전력 시계열 데이터의 이상 탐지 필요성

- > 산업 현장의 장비 고장, 성능 저하로 인한 **연간 손실 비용이 증가 추세**(연간 약 20조원 규모)
- > 전력 데이터는 시간당 수백만 건 생성되는 대용량 시계열 데이터로, 육안 검수 불가능
- > 실시간 감시 및 조기대응이 비용 절감과 안정적 생산에 필수적

스마트 팩토리 관점에서의 의의

- > 기존 시스템의 높은 오탐률(40-60%)로 현장 신뢰도 저하
- > 복잡한 **주기성과 계절성을 고려한 개선된 이상 탐지 알고리즘 필요**
- > 전력 데이터 기반 예측적 유지 보수(Predictive Maintenance) 체계 구축 가능성

프로젝트 개요

프로젝트 수행 배경 및 필요성

기존 이상 탐지 기법의 한계

통계 기반 방법

- > 평균, 표준편차 기반 단순 임계값 사용
- > 복잡한 시계열 패턴 감지 어려움

거리 기반 방법

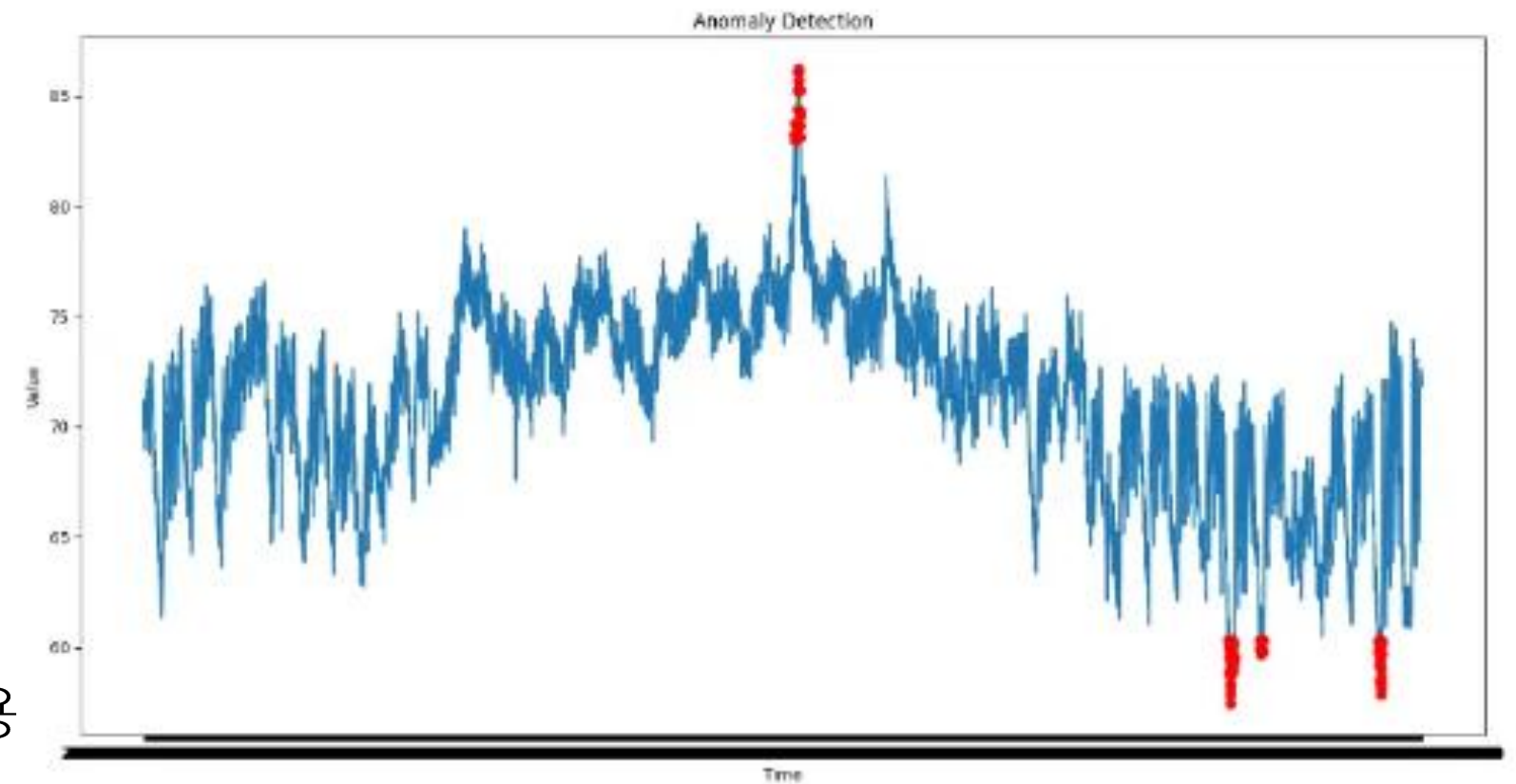
- > 유클리드 거리 등 단순 거리 측정
- > 시간 왜곡에 취약함

밀도 기반 방법

- > 데이터 밀집도에 기반한 군집화
- > 시간적 연속성 고려 부족

고정 윈도우 기반 방법

- > 임의로 설정된 고정 윈도우 크기 사용
- > 데이터의 내재된 주기성 무시
- > 주기 내 미묘한 이상치 탐지 어려움



산업 전력 시계열 데이터의 이상치 탐지 예시

수행 과정

데이터셋 구성

수집 기간: **3년** (산업 전력 데이터)

시간 간격: **15분**

데이터 포인트: 105,120개

산업 전력 사용 패턴 포함

데이터 전처리

- 1차 차분 적용
- **FFT 분석**
- KNN 보간법

윈도우 구성

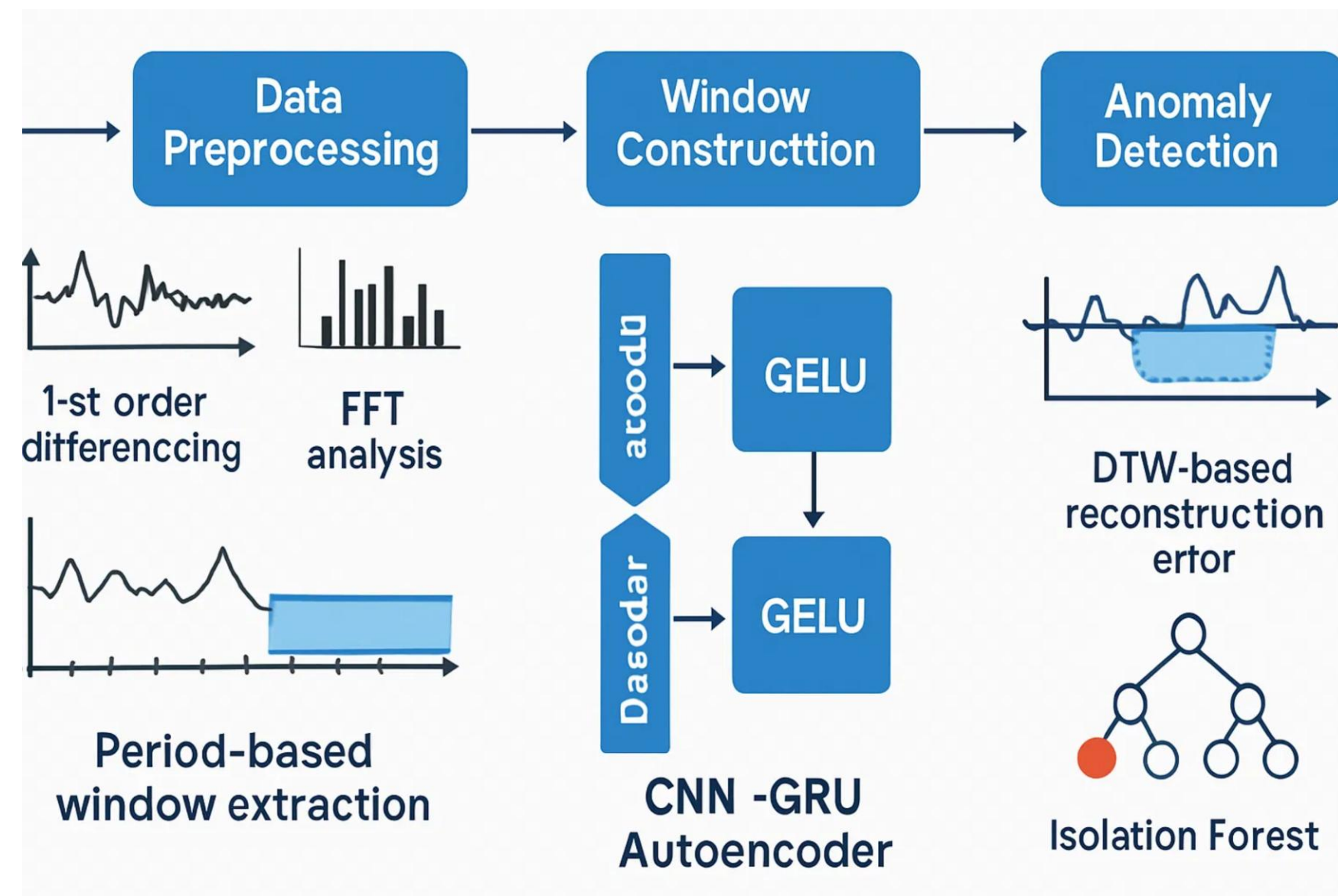
- **주기 기반** 윈도우
- 시계열 데이터 슬라이딩
- 모델 입력 변환

모델 학습

- **CNN-GRU** 구조
- GELU 활성화 함수
- 특징 학습 및 재구성

이상 탐지

- **DTW 기반** 오차
- Isolation Forest
- 시간 왜곡 대응



프로젝트 내용

프로젝트 구성

[1단계] 데이터 전처리

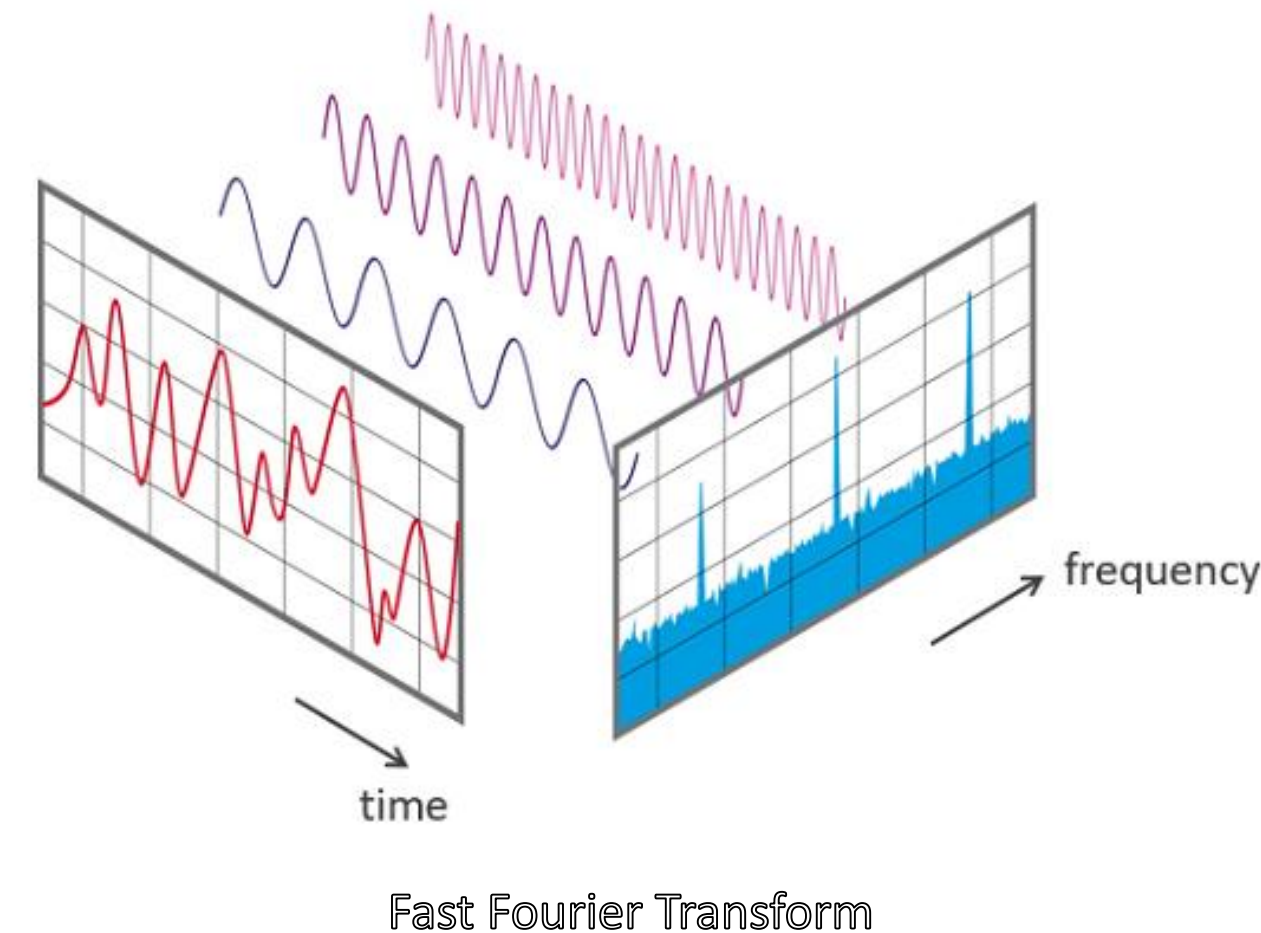
- > 3년치 산업 전력 데이터를 수집하여 0값을 결측치로 처리
- > KNN Imputer를 활용하여 결측 구간 보간
- > 시계열의 정상성 확보를 위해 1차 차분(differencing) 수행

[2단계] 주기 추출 및 윈도우 구성

- > FFT(Fast Fourier Transform)를 적용하여 상위 5개 주기를 추출
- > 주기를 기반으로 다양한 크기의 입력 윈도우(예: 11, 16, 24, 32, 96)를 구성
- > 각 윈도우에 대해 표준화(normalization) 진행

[3단계] 딥러닝 기반 모델 학습

- > 각 주기별로 독립된 CNN-GRU 기반 오토인코더 모델을 학습
- > 입력 데이터를 CNN을 통해 지역적 특징 추출 → GRU를 통해 시계열 데이터의 시간적 순서 정보 학습

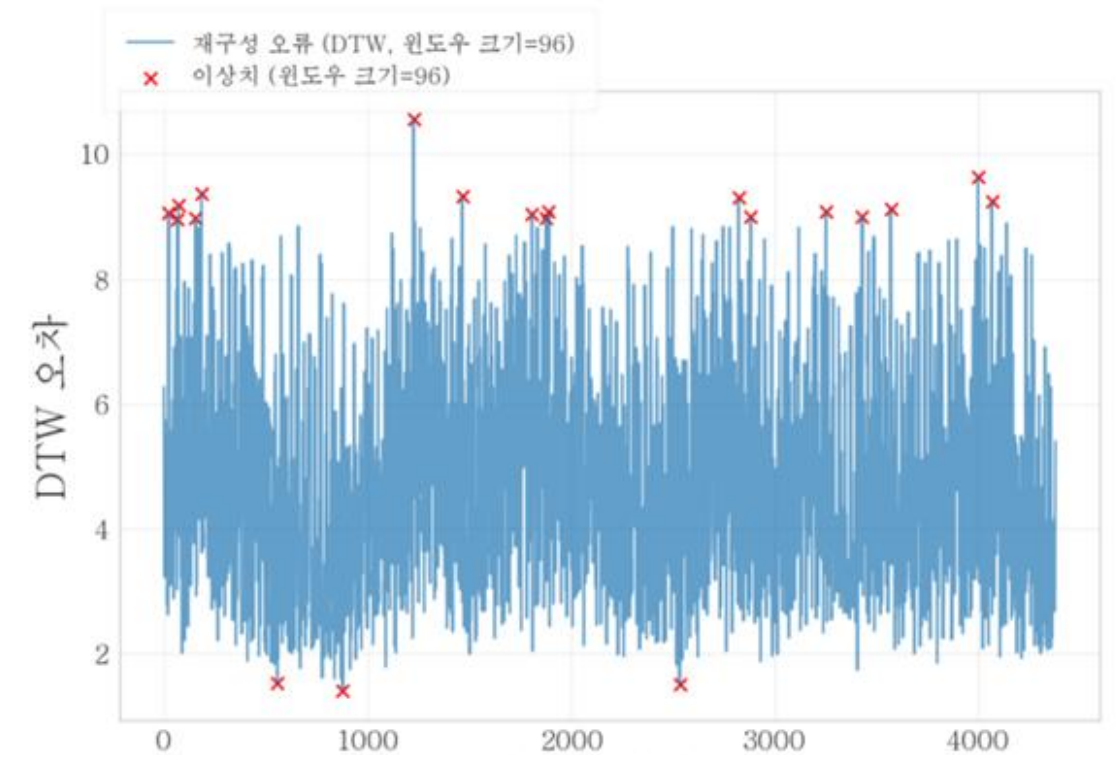


프로젝트 내용

프로젝트 구성

[4단계] 이상치 탐지

- > 입력과 재구성된 시퀀스 간의 DTW(Dynamic Time Warping) 거리 계산
- > DTW 오차 기반 이상 점수를 Isolation Forest로 분류하여 이상 윈도우 판별
- > 이상 윈도우 내에서 RMSE 기반 threshold로 개별 이상 시점 확정



DTW 재구성 오차(윈도우 크기 96)

[5단계] 시각화 및 해석

- > 이상 탐지 결과를 시간대별로 시각화 하여 실제 이상 이벤트와 매칭
- > (기대효과)산업 운영자에게 실시간 모니터링 및 조기 대응 정보 제공



실시간 조기 경보 제공 예시

프로젝트 내용

주요 기술

CNN-GRU Autoencoder 구조 설명

CNN (Convolutional Neural Network)

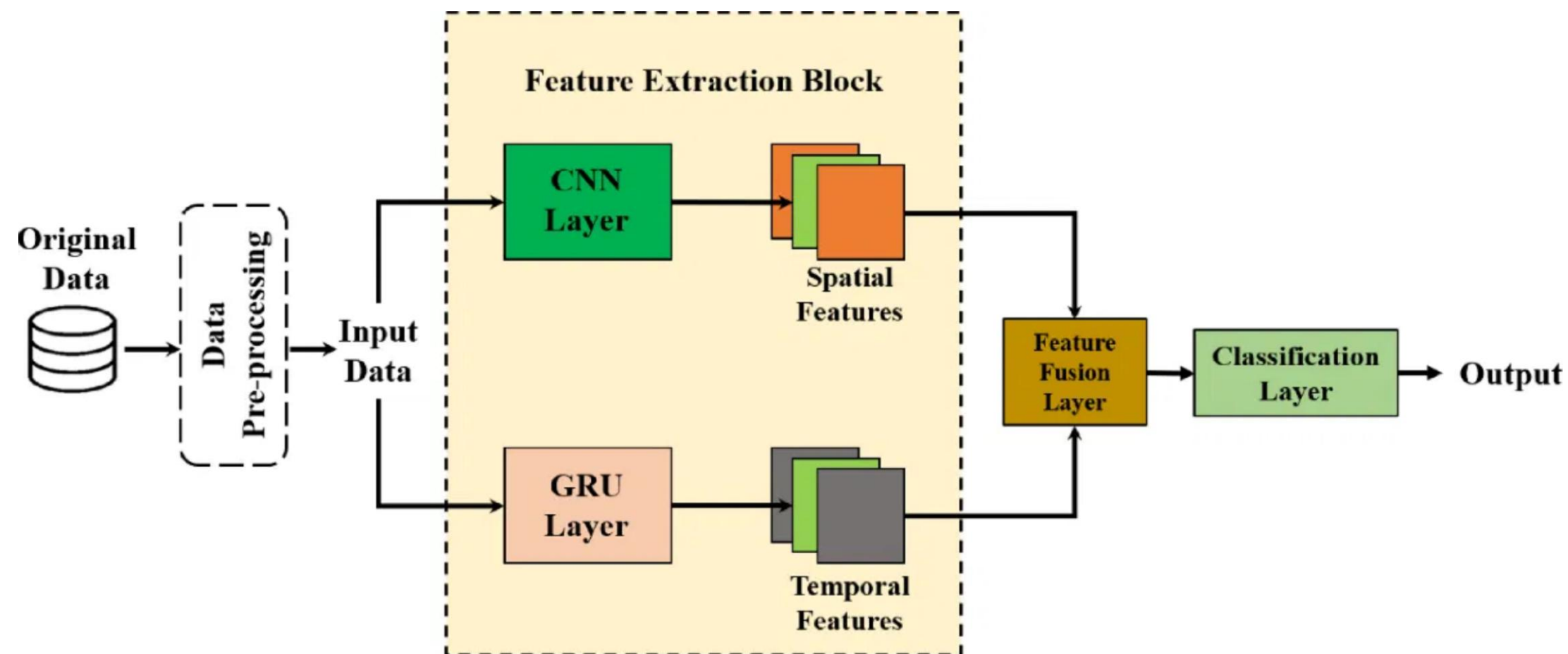
시계열 데이터의 지역적 특징 추출
필터(커널)를 사용하여 데이터 스캔 및 특징 맵 생성
단기적 패턴 및 변화 감지에 효과적

GRU (Gated Recurrent Unit)

시계열 데이터의 장기적 의존성 학습
LSTM보다 적은 게이트로 계산 효율성 향상
리셋 게이트와 업데이트 게이트로 정보 제어

GELU 활성화 함수

Gaussian Error Linear Unit
ReLU보다 부드러운 비선형성 제공
학습 안정성 및 성능 향상에 기여



CNN-GRU Autoencoder 특징

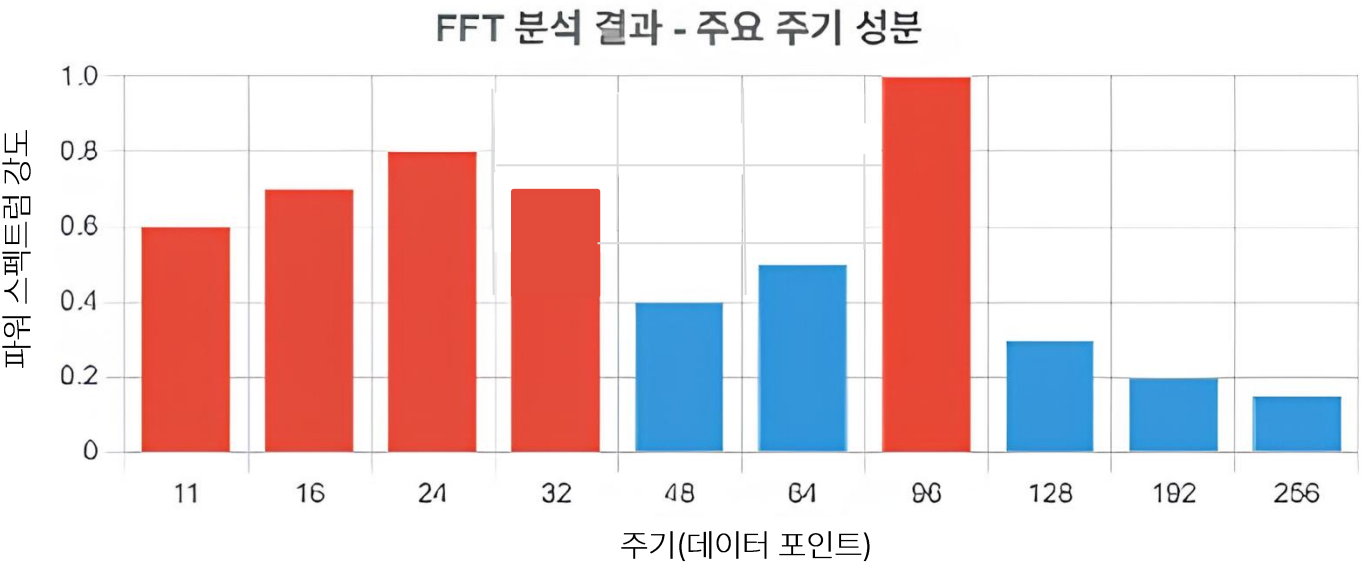
- ✓ CNN으로 시계열의 지역적 패턴 포착
- ✓ GRU로 시간적 흐름에 따른 장기 패턴 학습
- ✓ 오토인코더 구조로 정상 패턴 학습 및 재구성
- ✓ GELU 활성화로 학습 안정성 및 성능 향상

프로젝트 내용

주요 기술

FFT(고속 푸리에 변환) 개념

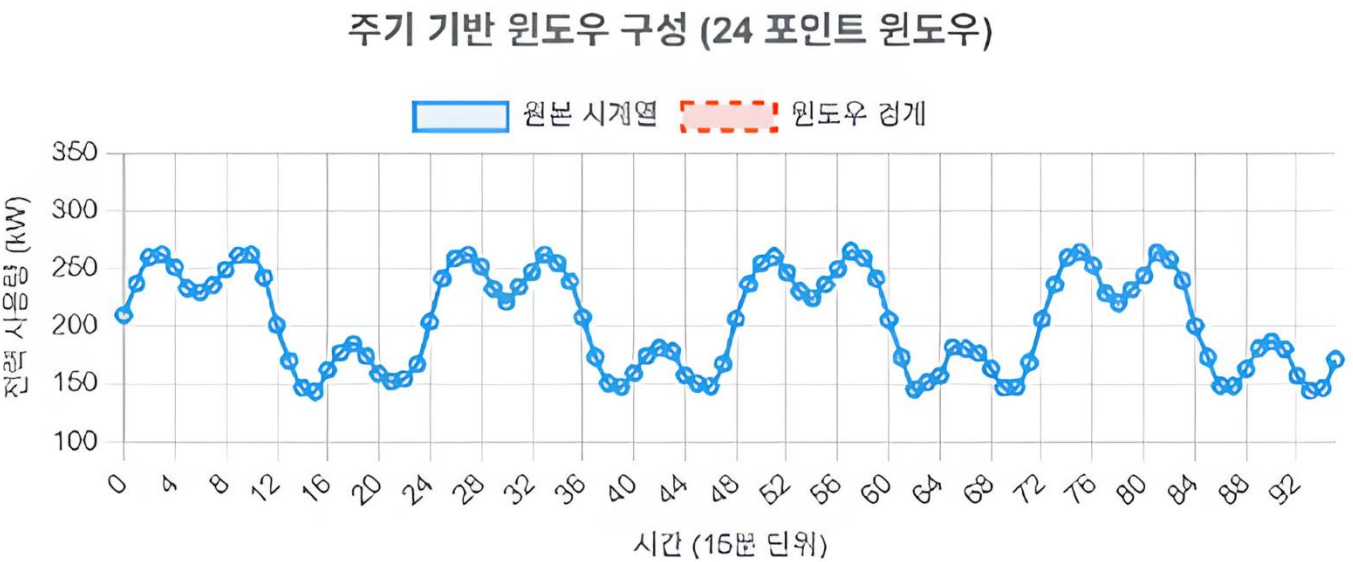
- 시간 영역의 신호를 주파수 영역으로 변환하는 알고리즘
- 시계열 데이터에 숨겨진 주기적 패턴을 찾아내는데 활용
- 계산복잡도 $O(n \log n)$ 으로 효율적인 주기 분석 가능



순위	주기 (데이터 포인트)	실제 시간	의미
1	96	24시간	일간 주기
2	32	8시간	근무 교대
3	24	6시간	생산 주기
4	16	4시간	설비 가동
5	11	2.75시간	공정 주기

주기 기반 윈도우 구성

- FFT로 추출한 주요 주기를 윈도우 크기로 활용
ex. 24 주기 -> 24개 데이터 포인트를 하나의 윈도우로 구성
- 주기 기반 윈도우로 데이터의 내재된 주기성을 모델이 학습
- 주기 내 발생하는 미묘한 이상 패턴까지 감지 가능



윈도우 구성의 장점

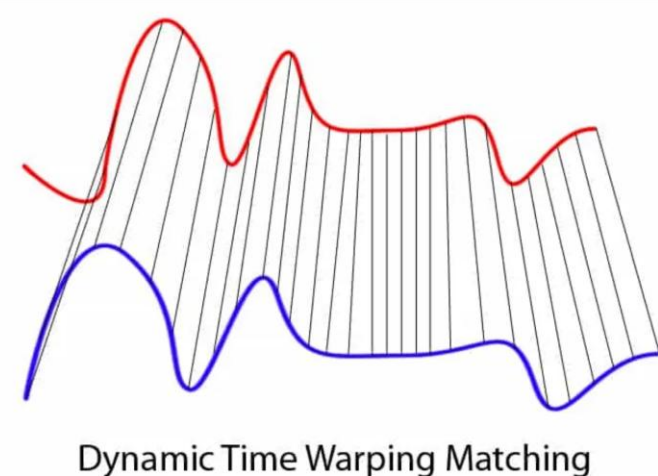
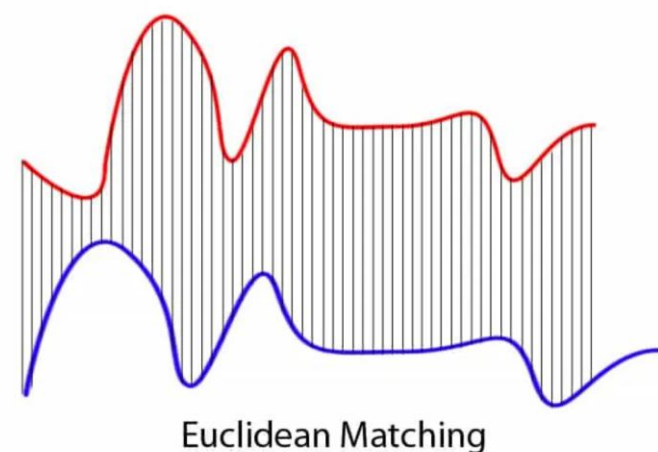
- 데이터의 자연적 주기를 반영한 최적 윈도우 크기 결정
- 고정 윈도우 대비 패턴 인식 성능 향상
- 다양한 주기를 조합하여 다중 스케일 분석 가능

프로젝트 내용

주요 기능

DTW(Dynamic Time Warping) 개념

- 두 시계열 데이터 간의 유사도를 측정하는 알고리즘
- 시간 축의 늘림이나 줄임을 허용하여 시간적 왜곡에 강인
- 두 시퀀스의 각 지점을 최적으로 연결하여 최소 비용으로 매칭



재구성 오차 계산

- 오토인코더가 입력 시계열을 재구성
- 시원본과 재구성 시계열 간 DTW 거리 계산
- 재구성 오차가 클수록 이상치일 가능성 높음

DTW vs MSE 비교

DTW (동적 시간 왜곡)

- 유연한 시간 정렬
- 시간 왜곡에 강인
- 패턴 유사성 중시

MSE (평균 제곱 오차)

- 시점별 1:1 대응 비교
- 시간 왜곡에 취약
- 시간 지연에 민감

DTW 기반 재구성 오차의 장점

시간 지연 패턴 탐지

형태적 유사성 기반

속도 변화에 강인

다양한 이상 유형 탐지

DTW 적용 효과

- MSE 대비 F1-Score 12.4% 향상
- 시간 왜곡 이상 패턴 탐지 18.7% 증가
- 오탐지(False Positive) 8.3% 감소

프로젝트 내용

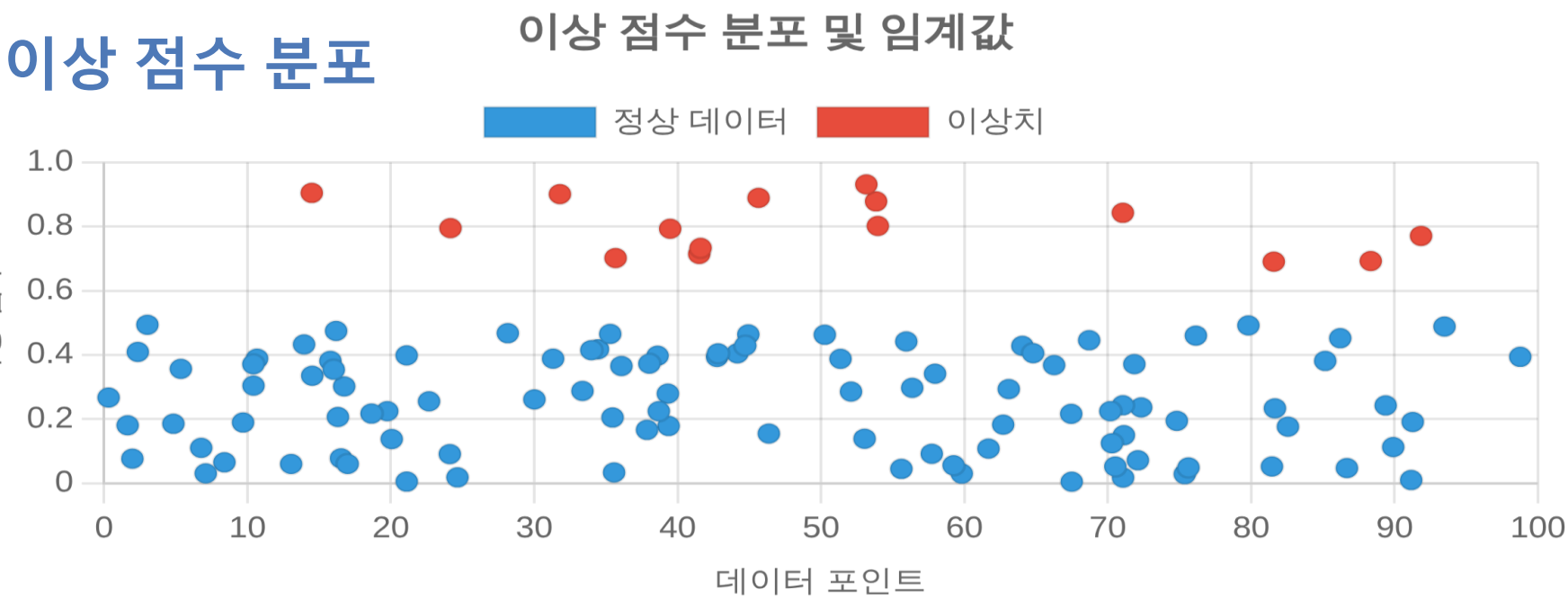
주요 기능

Isolation Forest 알고리즘

- 데이터를 **재귀적으로 분할**하여 이상치 식별
- 이상치는 정상 데이터보다 **더 빨리 고립** 됨
- 트리 구조에서 **짧은 경로**로 도달하는 데이터가 이상치

이상치 판별 과정

- 1 DTW 기반 재구성 오차 계산
- 2 재구성 오차를 Isolation Forest 입력으로 사용
- 3 이상 점수(Anomaly Score) 계산
- 4 임계값 기반 이상치 최종 판별



주요 파라미터

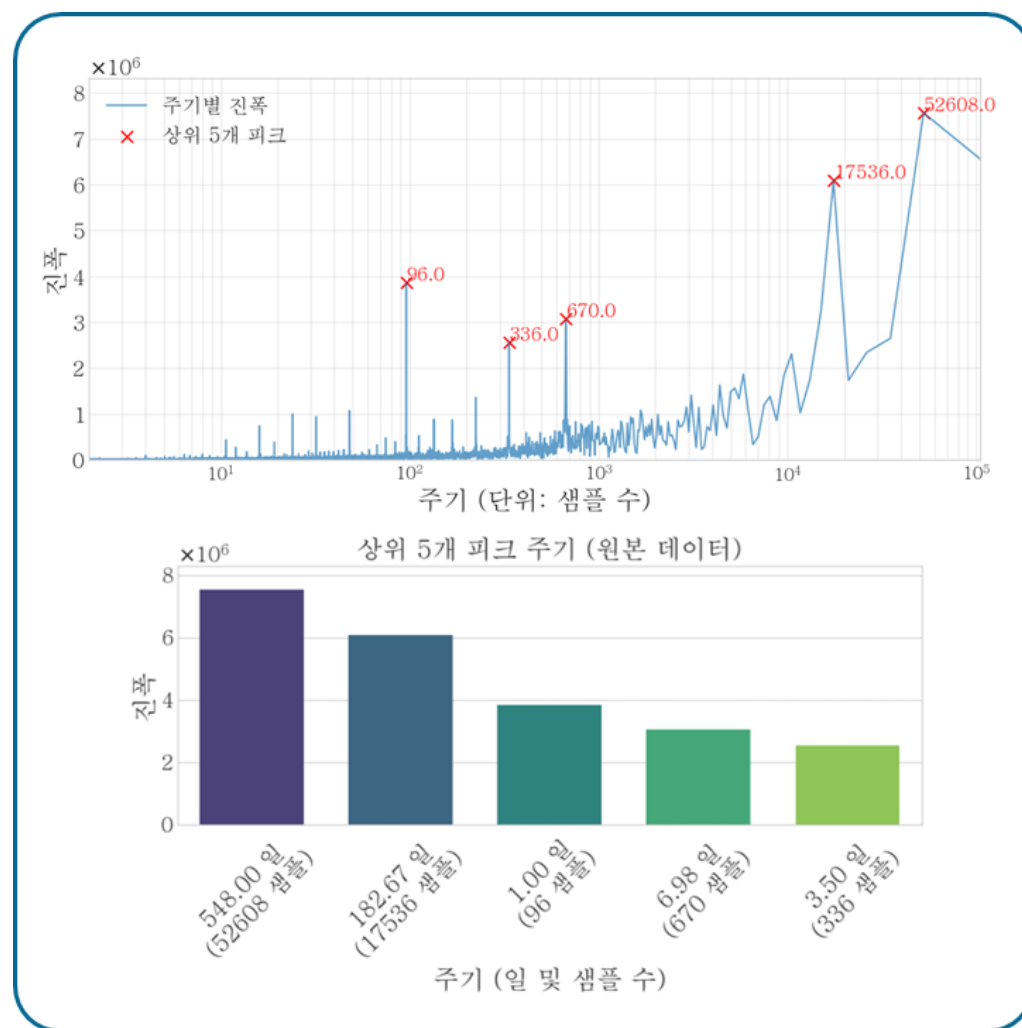
- **트리 개수**: 100개 (양상블 효과)
- **샘플링 크기**: 256 (계산 효율성)
- **오염도(contamination)**: 0.05 (데이터의 5%가 이상치)
- **임계값**: 0.65 (이상 점수 기준)

Isolation Forest 장점

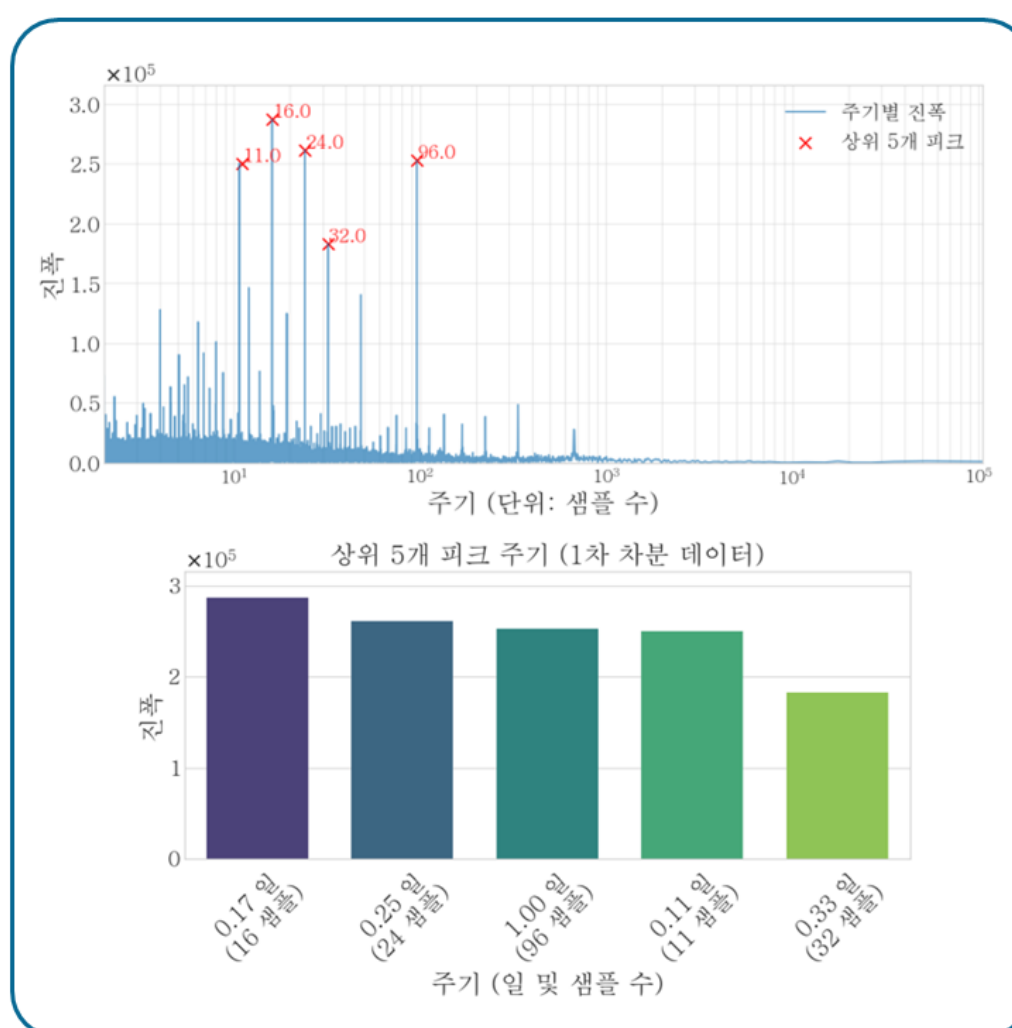
- ✓ **계산 효율성** 우수
- ✓ **확장성** 높음
- ✓ **사전 가정 불필요**
- ✓ **해석 용이성** 제공

프로젝트 성과

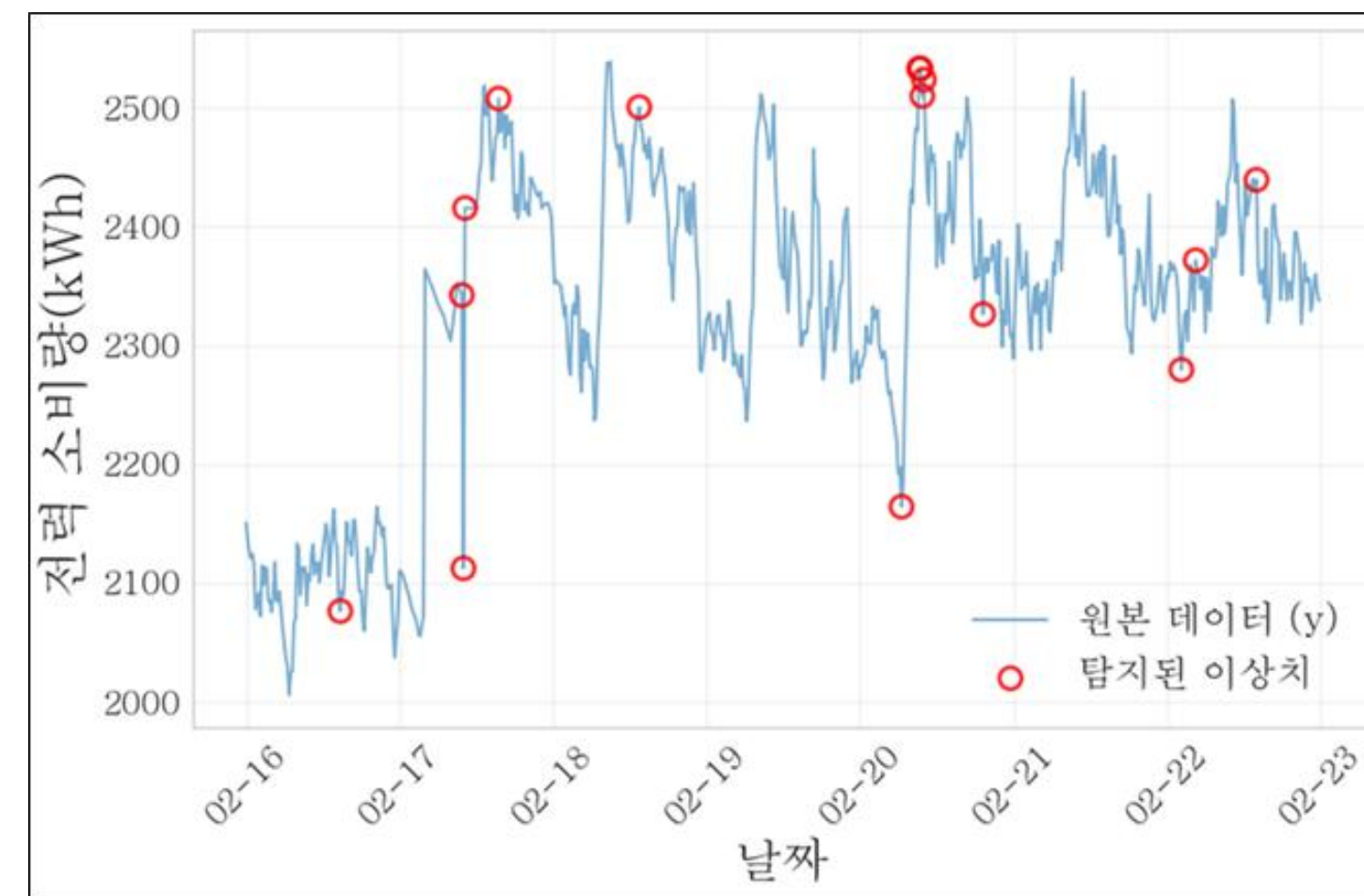
프로젝트 결과물



원본 데이터의 FFT 분석 및 피크 식별 결과



1차 차분 데이터의 FFT 분석 및 피크 식별 결과



(2020년 2월 16일 ~ 2020년 2월 22일) 이상치 검출 결과

결론

기대효과

정밀한 이상 탐지로 산업 설비 안정성 강화

- 다양한 시간 주기에 걸친 이상 신호를 정교하게 탐지함으로써, 산업 설비의 고장 및 효율 저하를 조기에 인지.

다중 주기 기반 탐지로 복잡한 시계열 패턴 대응

- FFT 기반 주기 추출을 통해 여러 시간 스케일에서 나타나는 이상 패턴까지 포착해 보다 정밀한 탐지 가능.

실제 산업 데이터를 활용한 실증적 효과 입증

- 실제 산업 현장의 장기간 전력 데이터를 기반으로 모델 성능을 검증하여, 실질적인 적용 가능성 입증.

학습 안정성과 일반화 성능 향상

- GELU 활성화 함수 및 GRU 기반 AE 구조 채택으로 기존 ReLU 기반보다 부드러운 학습과 높은 재현 능력을 확보.

결론

활용방안

스마트 팩토리의 실시간 이상 감지 시스템에 적용

- 센서 기반 전력 모니터링 시스템에 모델 연동을 통한 실시간 이상 탐지 및 경고 기능 구현으로, 예지 보전 체계의 핵심 기술로 활용 가능.

다양한 산업군의 에너지 관리 시스템에 통합

- 제조업, 데이터 센터 등에서 에너지 소비 이상 감지 및 비용 최적화 시스템의 핵심 알고리즘으로 적용 가능.

기존 SCADA 및 EMS 시스템 고도화에 기여

- 기존의 시스템보다 더 정교하고 유연한 이상 감지 기능으로 운영 시스템의 인공지능 기반 고도화에 활용 가능.

이상 탐지 결과 기반 경보 및 자동 대응 시스템 구축

- 이상치로 탐지된 이벤트를 기반으로 관리자 알림, 전력 차단, 설비 점검 예약 등의 자동화된 대응 체계 설계 가능.

출처

- Henning, S., Hasselbring, W., Burmester, H., Möbius, A., & Wojcieszak, M. (2021). Goals and measures for analyzing power consumption data in manufacturing enterprises. *Journal of Data, Information and Management*, 3(1), 65-82.
- Guo, Y., Liao, W., Wang, Q., Yu, L., Ji, T., & Li, P. (2018, November). Multidimensional time series anomaly detection: A gru-based gaussian mixture variational autoencoder approach. In *Asian conference on machine learning* (pp. 97- 112). PMLR.
- Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- Xu, H., Chen, W., Zhao, N., Li, Z., Bu, J., Li, Z., ... & Qiao, H. (2018, April). Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications. In *Proceedings of the 2018 world wide web conference* (pp. 187-196).
- Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*.
- Sakoe, H., & Chiba, S. (2003). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1), 43-49.
- Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In *2008 eighth ieee international conference on data mining* (pp. 413-422). IEEE.

감사합니다

ECO GA - 산업 전력 소비량의 이상치 검출 및 최적화를 통한 에너지 소비량 감축