

XAI 기반 발전설비 고장 기록 데이터 품질 향상 시스템 개발

김유림* · 박정인* · 박동현* · 강성우*†

* 인하대학교 산업경영공학과

Development of System for Enhancing the Quality of Power Generation Facilities Failure History Data Based on Explainable AI (XAI)

Kim Yu Rim* · Park Jeong In* · Park Dong Hyun* · Kang Sung Woo*†

* Department of Industrial Engineering, Inha University

ABSTRACT

Purpose: The deterioration in the quality of failure history data due to differences in interpretation of failures among workers at power plants and the lack of consistency in the way failures are recorded negatively impacts the efficient operation of power plants. The purpose of this study is to propose a system that classifies power generation facilities failures consistently based on the failure history text data created by the workers.

Methods: This study utilizes data collected from three coal unloaders operated by Korea Midland Power Co., LTD, from 2012 to 2023. It classifies failures based on the results of Soft Voting, which incorporates the prediction probabilities derived from applying the predict_proba technique to four machine learning models: Random Forest, Logistic Regression, XGBoost, and SVM, along with scores obtained by constructing word dictionaries for each type of failure using LIME, one of the XAI (Explainable Artificial Intelligence) methods. Through this, failure classification system is proposed to improve the quality of power generation facilities failure history data.

Results: The results of this study are as follows. When the power generation facilities failure classification system was applied to the failure history data of Continuous Ship Unloader, XGBoost showed the best performance with a Macro_F1 Score of 93%. When the system proposed in this study was applied, there was an increase of up to 0.17 in the Macro_F1 Score for Logistic Regression compared to when the model was applied alone. All four models used in this study, when the system was applied, showed equal or higher values in Accuracy and Macro_F1 Score than the single model alone.

● Received 12 June 2024, 1st revised 10 July 2024, accepted 23 July 2024

† Corresponding Author(kangsungwoo@inha.ac.kr)

© 2024, The Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

* 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 지역지능화혁신인재양성사업임 (IITP-2024-RS-2023-00259678).

Conclusion: This study propose a failure classification system for power generation facilities to improve the quality of failure history data. This will contribute to cost reduction and stability of power generation facilities, as well as further improvement of power plant operation efficiency and stability.

Key Words: Power Generation Facilities, Quality Improvement, Natural Language Processing, XAI, LIME

1. 서 론

발전소는 각종 산업과 일상생활에 필수적인 전력을 생산 및 공급하는 시설로, 발전소의 안정적인 운영은 현대 사회의 지속 가능한 발전 및 국민의 삶의 질 향상에 중요한 요소이다(Batelić et al., 2021). 발전설비의 고장은 전력 공급 중단 및 안전사고와 같은 심각한 위험을 초래할 수 있기에, 이를 방지하기 위해서는 고장 기록 데이터의 체계적인 분석과 관리가 필수적이다(Hwang et al., 2023; Kim et al., 2017; Kim et al., 2021).

그러나 작업자 간의 고장에 대한 해석 차이 및 고장 기록 방식에 대한 일관성 부재로 인한 고장 기록 데이터의 품질 저하에 대한 문제가 있다(Jardine, 2006; Lim and Han, 2018; Liu, 2022). 이는 곧 발전설비 고장 분석에 있어 신뢰성을 저하시키고 유지보수 전략 개발에 지장을 주며, 더 나아가 발전소의 효율적인 운영에 악영향을 끼치는 실정이다.

본 연구에서는 작업자가 명확한 기준 없이 임의로 고장을 분류하는 기존의 고장 기록 데이터 분류 방법의 한계를 극복하고자, 대상 장비인 석탄 하역기에 대해 작업자들이 작성한 고장 기록 텍스트 데이터를 바탕으로 머신러닝과 XAI(eXplainable Artificial Intelligence) 기법 중 하나인 LIME(Local Interpretable Model-agnostic Explanations)을 이용하여 고장을 분류하는 방법론을 제안하고자 한다. LIME을 통해 도출한 예측 결과에 대한 해석을 분류 과정에 결합하여, 분류 과정의 일관성 및 성능을 향상시킬 수 있다. 이를 통해 고장 기록 데이터 표준화를 통한 데이터 품질 향상 및 고장 분류 자동화의 초석을 다지고, 최종적으로 발전소 운영 효율성 향상에 기여할 것으로 기대된다.

본 논문은 총 5개의 장으로 구성되어 있다. 제 1장은 연구의 동기와 목적 및 논문의 전체적 구성을 설명하는 서론으로 이루어진다. 제 2장에서는 관련 연구에 대해 기술한다. 제 3장에서는 XAI 기반 발전설비 고장 기록 데이터 분류에 관한 방법론을 제안한다. 제 4장에서는 실험 결과에 대해 논한다. 제 5장에서는 결론 및 향후 연구에 대해 논한다.

2. 관련 연구

고장 원인 및 부품 정보에 대해 서술된 데이터는 고장 패턴 파악을 통한 설비 보전을 용이하게 한다(Hong, 2022). 또한 설비에서 발생한 고장, 혹은 수행된 작업에 대해 기록하는 것은 설비 성능 평가 및 개선에 큰 도움이 된다. 하지만 고장 기록 데이터는 일반적으로 작업자의 수동 입력에 의존하기에 오류에 취약하다(Jardine, 2006). 따라서 고장 기록 데이터를 효과적으로 관리하는 다양한 방법론이 제안되어 왔다. 설비 데이터 단어 간의 의미 관계를 고려하기 위해 심층 학습 모델을 통해 설비 텍스트 데이터를 분류하는 연구가 진행되었다(Liu, 2022; Meng et al, 2023; Tian, 2021). 하지만 이러한 Black-box 모델은 실제로 어떻게 작동하는지 이해하기 쉽지 않으며, 높은 수준의 책임성을 요구하는 전력 계통 분야에서는 분류 결과에 대해 신뢰하기 어렵다(Machlev, 2022). 또한 고장 빈도가 낮은 설비의 경우, 적은 데이터로 인하여 분석의 신뢰성이 떨어질 수 있다. 이를 기반으로 적절하지 않은 보전이 시행된다면 큰 경제적 손실을 야기할 수 있다. 따라서 적은 데이터로도 충분한 성능을 낼 수 있는 머신러닝 모델이 선호되며,

예측 결과에 대한 설명을 제공할 수 있는 XAI를 적용한 연구들이 진행되고 있다.

2.1 머신러닝을 통한 설비 데이터 분류

설비별 고장 및 유지보수 관련 데이터의 양은 설비 종류에 따라 큰 차이를 보인다. 머신러닝을 통한 분류는 심층 학습 모델에 비해 상대적으로 적은 양의 학습 데이터에서도 높은 성능을 보인다는 장점이 있다(Sarker, 2021). 따라서 머신러닝을 통하여 설비 텍스트 데이터를 분류하는 연구가 이루어지고 있다. 전력 설비 결함 로그 데이터를 중요도에 따라 분류하기 위해 데이터를 구조화된 형태로 변환하고 SVM을 사용하는 연구가 진행되었다(Wang et al., 2021). 고등 교육 건축물의 HVAC 시스템 관련 유지보수 데이터에 대한 문제 코드를 분류하기 위해 SVM, Random Forest, Naive Bayes 등을 활용한 연구가 진행되었다(Hong, 2022). Decision Tree, Random Forest를 통해 풍력 터빈의 실패 상태를 분류하여 터빈의 유지보수 비용 절감을 목표로 하는 연구가 진행되었다(Blanco et al., 2019). 유지보수 작업의 효율성을 높이고자 SVM과 Naive Bayes를 사용하여 방글라데시 제조 업체의 유지보수 기록 데이터로부터 고장을 자동으로 식별하는 연구가 진행되었다(Rahman, 2023). 이러한 연구들은 수동으로 데이터를 기록하는 방식으로 인한 시간 소모, 일관성 부재 및 오류와 같은 문제를 텍스트 분류 모델을 통해 극복하는 시도를 보였다. 하지만 복잡한 텍스트 구조를 가지는 데이터에 대해서는 단어 간의 의미 관계를 고려하는 데 어려움이 있어 단일 모델로서는 성능이 다소 낮아진다는 단점이 존재한다.

2.2 설명 가능한 인공지능(XAI)

발전소와 같은 대규모 설비의 운영 및 유지보수는 매우 중요하기에, 모델의 예측 결과에 대해 작업자가 신뢰하고 이해할 수 있어야 한다. 이에 따라 다양한 전력 시스템에 대해 설명 가능한 인공지능(XAI)을 도입하는 시도가 진행되고 있다. 전력 시스템에서 발생하는 다양한 사고에 대한 과도 안정성 상태에 대해 분석하는 연구가 진행되었다(Chen et al., 2019). 이를 위해 XGBoost를 사용하여 과도 안정성과 관련된 주요 지표들을 예측하였고, LIME 기법을 사용하여 예측에 대한 설명을 제공함으로써 시스템 운영자의 의사 결정을 지원하고자 하였다. 전력 발전량 데이터를 바탕으로 시간대별 전력 수요를 예측하기 위해 XGBoost를 사용하고, XAI의 일종인 PDP(Partial Dependence Plot)를 통하여 입력 특성의 중요도와 발전량 간 상관관계를 설명하는 연구가 진행되었다(Lee et al., 2020). 해당 연구에서는 중요도가 낮은 입력 특징을 제외하여 적은 데이터에서도 효율적인 예측이 가능하도록 하였다. 핵 발전소에서 비정상적 운영 상황이 발생할 경우를 대비하여, 비정상 시나리오를 예측할 LightGBM 모델과 설명 가능성을 제공하기 위한 SHAP(SHapley Additive exPlanations) 기법을 통합하여 사용자 친화 인터페이스를 만듦으로써 비정상적 상황에 대한 진단 정보를 제공하고자 하는 연구가 진행되었다(Park et al., 2022). 태양광 패널에서 초기 결함을 탐지하고 진단할 수 있는 설명 가능한 결함 감지 및 진단 시스템 구축을 위하여 XGBoost 기반 분류기와 LIME 기법을 통해 태양광 패널의 신뢰성을 확보하는 연구가 진행되었다(Sairam et al., 2020). 이러한 연구들은 XAI가 작업자에게 모델의 예측에 대한 설명을 제공하며, 나아가 유지보수와 같은 시스템 보전에 있어 의사 결정을 지원한다는 것을 보여준다. 본 연구에서는 이에 더하여 설명 가능한 인공지능(XAI)을 통해 도출된 해석을 고장 기록 데이터의 분류 과정에 직접적으로 결합하여 분류 성능을 향상시키는 시스템을 제안한다.

3. 방법론

본 연구에서 제안하는 방법론은 Figure 1과 같다.

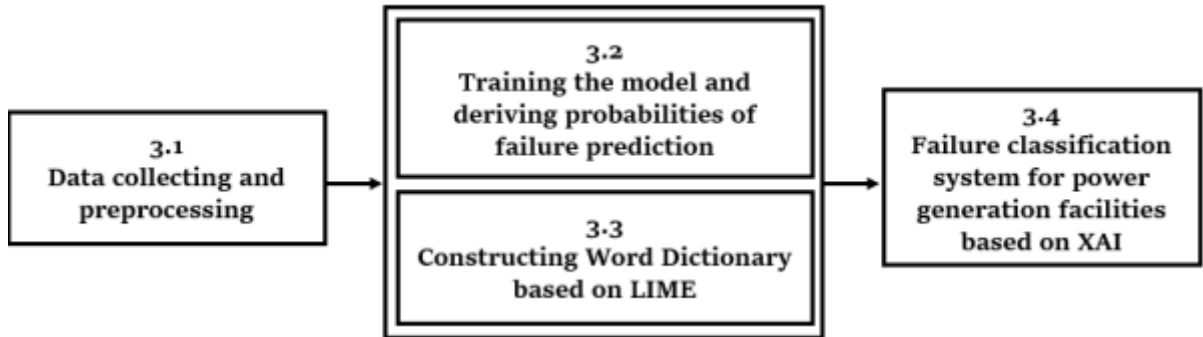


Figure 1. Flow chart for failure classification system

고장 기록 데이터 수집 및 전처리를 거친 후, 머신러닝 모델 학습을 통해 도출한 고장 증상별 예측 확률과 LIME를 통해 구축한 고장 증상별 단어사전을 통해 얻은 LIME Score를 결합한다. 이를 기반으로 고장을 분류하는 XAI 기반 발전설비 고장 분류 시스템을 구축한다.

3.1 데이터 수집 및 전처리

국내 총 발전량에서 석탄화력발전이 차지하는 비율은 2022년 기준 39.7%로, 국내에서는 원자력 다음으로 전력 생산에 가장 많이 사용되는 에너지원이다(Korea Energy Economics Institute, 2023). 석탄화력발전은 한국중부발전의 주력 발전소로써, 본 연구는 한국중부발전에서 운영 중인 연속식 석탄 하역기(CSU : Continuous Ship Unloader) 중 CSU-1B, CSU-2A, CSU-3B를 분석 대상 장비로 선정한다.

석탄 하역기는 60,000 DWT~135,000DWT 크기의 대양 횡단 선박에서 발전소용 석탄을 하역하기 위한 장비로, 구조 및 하역 경로는 Figure 2와 같다.

석탄 하역기는 하역 설비로 이루어진 상부 구조물(Super Structure)과 동력 설비가 포함된 하부 구조물(Sub Structure)로 구성된다. 상부 구조물은 선박의 화물칸으로부터 버킷 엘리베이터(Bucket/Elevator)를 통해 하역할 석탄을 퍼 올린 후, 스파이어럴 슈트(Spiral chute)로 옮겨 붐 컨베이어(Boom conveyor)로 이송한다. 이송된 석탄은 하부 구조물의 갠트리 컨베이어(Gantry conveyor)로 옮겨지고, 슈트 시스템(Chute system)에 의해 부두로 이송된다. 주요 구동부는 유압 구동 방식으로 이루어져 있다.

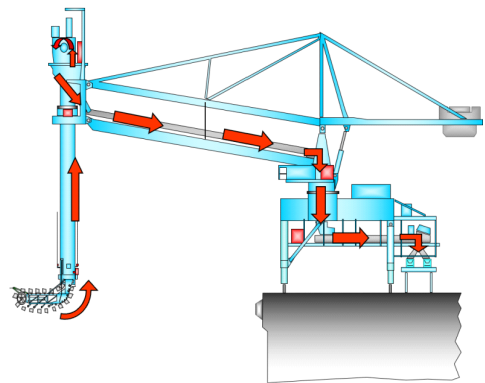


Figure 2. Continuous ship unloader structure and unloading route

Table 1. Continuous ship unloader failure history data

| TM 제목 | 발행일자 | ... | 증상 |
|-------------------------------------|------------|-----|-------|
| [CSU#1B BOOM BELT 낙탄회수설비 체인이탈] | 2021/12/17 | ... | 누설 |
| [csu#1B gantry belt 낙탄회수설비 체인이탈] | 2021/11/25 | ... | 변형, 휨 |
| [낙탄회수설비] csu#1B gantry 낙탄회수설비 체인 이탈 | 2021/06/20 | ... | 이완 |
| csu#1b gantry belt 낙탄회수설비 체인 이탈 | 2021/06/05 | ... | 이완 |

Table 1과 같이, 작업자 간의 고장에 대한 해석 차이 및 분류 일관성 부재로 인하여 유사한 내용의 TM(Trouble Memo) 제목임에도 분류 일관성이 낮음을 확인할 수 있다. 이는 모델 학습에 부정적인 영향을 미칠 수 있으므로, 2012년 7월부터 2023년 4월까지의 고장 기록 데이터에 대하여 한국중부발전과의 논의를 거쳐 전문가에 의해 연속식 석탄 하역기의 고장 기록에 대한 일관된 해석을 도출하였다.

Table 2. Selected failure classification criteria

| 고장 증상 | | 고장 증상 | | 고장 증상 | |
|-------|--------|-------|-----------------|-------|--------|
| 1 | 고착, 간섭 | 6 | 변형, 휨 | 11 | 오염, 변질 |
| 2 | 누설 | 7 | 부식, 침식 | 12 | 오지시 |
| 3 | 단선, 단락 | 8 | 부족 | 13 | 이완 |
| 4 | 마모 | 9 | 소손, 손실 | 14 | 이음, 소음 |
| 5 | 막힘 | 10 | 오동작, 동작불량, 한계초과 | 15 | 파손, 파열 |

선정된 주요 고장 분류 기준은 Table 2와 같으며, 이에 상응하는 고장으로 라벨링된 데이터를 사용하였다.

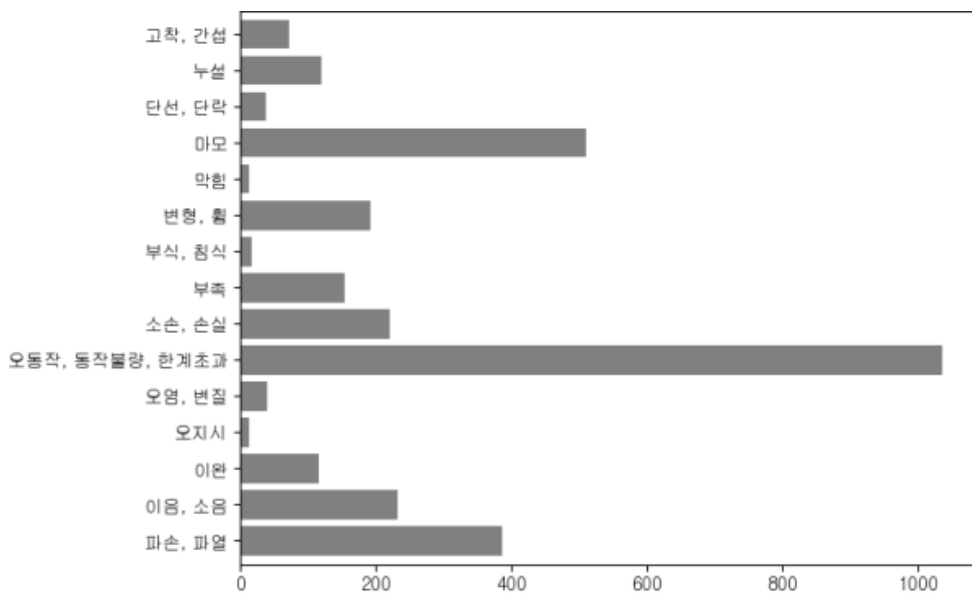


Figure 3. Distribution of data by failure type

본 연구는 3,150개의 데이터를 실험에 사용하였으며, 고장 증상별 데이터 분포는 Figure 3와 같다. '오동작, 동작 불량, 한계초과'에 속하는 데이터가 1035개로 다른 고장 증상에 비해 많은 비중을 차지하고 있으며, '오지시'의 경우 12개로 가장 적었다.

Word Tokenization 진행 후, 소문자로의 변환을 거치고 기호, 특수문자, '번', '개', 'ea'와 같은 분석에 영향을 끼치지 않는 불용어를 제거한다. 또한 중복 문자열이 존재할 경우, 이를 제거한다. 고장 기록 데이터를 머신러닝 모델에 사용하기 위해 TF-IDF Vectorizer를 사용하여 텍스트 데이터를 숫자 벡터로 변환한다. 고장 기록 데이터의 특성상 설비 관련 용어가 많고, 본 연구에서 사용된 데이터의 경우 문맥이 중요하지 않은 짧은 텍스트로 구성되어 있다. BERT와 같이 사전 학습된 벡터화 모델의 경우, 보편적인 언어 사용 패턴을 기반으로 학습되어 있기에 본 연구와 같이 특정 도메인에 특화된 용어 사용이 빈번한 경우 적합하지 않다(Han et al., 2022). 또한 벡터화 과정에서 상대적으로 많은 계산 리소스를 필요로 하기에, 현장에서의 적용을 고려하여 TF-IDF Vectorizer를 사용하여 벡터화를 수행함으로써 이러한 문제를 해결하였다.

본 연구는 전처리가 완료된 데이터의 80%를 6:4로 분할하여 60%는 모델 학습에, 40%는 단어사전 구축에 활용한다. 나머지 20%의 데이터는 평가를 위한 검증 데이터로 사용한다.

3.2 모델 학습 및 고장별 예측 확률 도출

모델 학습시 predict_proba 기법을 적용하여, TM 제목에 대한 고장 증상별 예측 확률을 도출한다. predict_proba는 머신러닝의 분류 모델에서 사용되는 기법 중 하나로, 특정 데이터가 각 클래스에 속할 예측 확률을 반환한다. 단순히 가장 예측 확률이 높은 클래스를 선택하는 것을 넘어, 예측에 대한 추가적인 정보를 확보할 수 있다는 이점이 있다.

Table 3. Prediction probability by failure

| 0 | TM 제목 | 고착, 간섭 | 누설 | 단선, 단락 | ... | 이완 | 이음, 소음 | 파손, 파열 |
|-----|-------------------------------------|--------|------|--------|-----|------|--------|--------|
| 1 | two way chute 하부 이음발생 | 0.07 | 0.00 | 0.00 | ... | 0.01 | 0.63 | 0.08 |
| 2 | hpp oil 누유 | 0.00 | 1.00 | 0.00 | ... | 0.00 | 0.00 | 0.00 |
| 3 | 이동용 운전실 전원 cable 단선 | 0.00 | 0.00 | 0.68 | ... | 0.00 | 0.01 | 0.17 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ... | ⋮ | ⋮ | ⋮ |
| 630 | boom luffing chain overtension trip | 0.01 | 0.00 | 0.00 | ... | 0.05 | 0.00 | 0.15 |

TM 제목에 대해 각 고장 증상별 예측 확률을 도출한 예시는 Table 3와 같다.

본 연구에서는 4가지 머신러닝 모델을 사용한다. Tree 기반 모델인 Random Forest, Regression Classifier인 Logistic Regression Classifier, Boosting 기반 모델인 XGBoost, Max-Margin Classifier인 SVM 모델을 사용한다. Random Forest는 여러 결정 트리의 앙상블을 사용하여 복잡한 데이터셋에서 모델의 일반화 성능을 향상시킬 수 있다는 이점이 있어 선정하였다(Lee and Kim, 2022). 각 결정 트리는 학습한 패턴을 기반으로 독립적으로 입력 샘플에 대해 예측을 수행하며, 모든 트리의 예측을 종합하여 predict_proba 기법을 통해 각 트리의 예측 결과로부터 각 클래스에 속할 확률을 추정한다. Logistic Regression은 벡터화된 텍스트 데이터와 같은 희소 행렬을 처리하는

데 효과적이며, 고차원 데이터에서도 효율적으로 작동하기에 선정하였다. 다중 분류의 경우, 각 클래스에 대해 모델이 학습하는 weights와 bias를 사용하여 정의되는 입력 특성의 선형 결합으로 이루어진 함수를 학습하고, 이를 Softmax 함수에 적용하여 확률을 추정한다. predict_proba 기법을 통해 데이터가 각각의 클래스에 속할 확률을 제공하여, 가장 높은 확률을 가진 클래스를 최종 예측으로 선택한다. XGBoost는 고성능 부스팅 알고리즘으로, 복잡한 데이터셋에서도 높은 정확도를 제공하며 대규모 데이터셋 처리에 효과적이기에 선정하였다. Weak Learners를 순차적으로 학습시켜 예측을 결합함으로써 Strong Learner를 생성하는 boosting 방식을 사용하기에, 각 트리는 이전 트리들의 오류를 보완해 나가는 형태로 학습된다. 각 결정 트리는 입력 데이터에 대한 예측을 수행하며, 이를 모두 합산하여 구해진 최종 예측값을 log-odds 형태로 계산하고, predict_proba 기법을 통해 Softmax 변환을 거쳐 데이터가 각 클래스에 속할 확률을 출력한다. SVM은 커널 트릭을 바탕으로 복잡한 데이터 구조에서도 유의미한 패턴과 관계를 발견할 수 있게 해주고, 일반화 성능이 우수하기에 선정하였다. 모델에서 predict_proba를 사용하기 위해서는 확률적 결과를 직접 제공하지 않는 머신러닝 모델의 출력을 확률로 변환하는 방법인 Platt scaling을 통해 각 결정의 확률 추정치를 얻을 수 있다. 다중 분류의 경우 OvA 전략을 사용하여 각 클래스에 대한 확률을 추정할 수 있으며, 이 확률값들을 조합하여 각 클래스에 대한 최종 확률을 도출한다.

3.3 LIME 기반 단어사전 구축

Black-box 모델은 내부 작동 방식이 불투명하거나 사람이 쉽게 이해할 수 없는 복잡한 계산 모델을 의미한다. Black-box 모델이 보유하고 있는 불투명성은 작업자가 시스템을 신뢰하기 어렵게 만든다(Zednik, 2021). LIME은 Black-box 모델을 포함한 다양한 형태의 분류기에 대하여 국소적으로 신뢰할 수 있는 해석 가능한 표현을 통해 모델을 해석 가능하도록 정의하는 방법이다(Ribeiro, 2016). LIME을 통해 생성되는 특정 데이터 포인트(x)에 대한 설명은 <Algorithm 1>과 같다.

Table 4. Explaining failure data prediction using LIME

| <i>Algorithm 1.</i> | |
|--|---|
| <i>Explaining failure data prediction using LIME</i> | |
| Input: | f : The Predictive Model to Interpret VD : Vectorized Failure Data Points to Interpret VD' : A Version that Simplified to be Interpretable N : Number of Data Points to Generate near VD |
| Output: | EXP : Explanation of $f(VD)$ |
| Procedure: | $SET = \{ \}$ For i in $\{1, 2, 3, \dots, N\}$ Do Sample PVD'_i by perturbing VD'_i Reconstruct PVD_i from PVD'_i Compute $\pi_{VD}(PVD_i)$ by measuring Similarity between VD and PVD_i $SET = SET \cup \langle PVD'_i, f(PVD_i), \pi_{VD}(PVD_i) \rangle$ End For Train an interpretable model q using SET toward minimizing $\xi(VD)$ return EXP through q |

고장 증상별 예측 확률 값 $f(VD)$ 는 해석 가능한 모델 q 를 통해 설명 가능하며, 모델 q 를 학습시키기 위한 목적 함수 $\xi(VD)$ 는 수식 (1)로 나타낼 수 있다. 모델의 해석 가능성과 국소적인 신뢰도를 보장하기 위해 손실 함수 $\mathcal{L}(f, q, \pi_{VD})$ 를 최소화하면서 모델의 복잡도 $\Omega(q)$ 를 조절해야 한다. 해석 가능한 모델 q 는 해석 가능한 모델의 집합 Q 의 원소이며, 선형 모델, 의사 결정 트리 등을 포함한다. $\Omega(q)$ 는 선형 모델의 경우 0이 아닌 계수의 수, 트리 모델의 경우 나무의 깊이에 해당한다. $\mathcal{L}(f, q, \pi_{VD})$ 은 모델 q 가 π_{VD} 으로부터 정의된 공간에서 설명하고자 하는 모델 f 와 얼마나 다른지 측정하는 함수로, 수식 (2)로 표현한다(Ribeiro, 2016). 손실함수 \mathcal{L} 과 복잡도 $\Omega(q)$ 은 trade-off 관계로써, 이 두 값의 합이 최소가 되도록 모델 q 를 학습한다. 학습한 모델을 통하여 고장 증상에 대한 모든 Keyword 별 가중치를 얻을 수 있다.

$$\xi(VD) = \arg \min_{q \in Q} \mathcal{L}(f, q, \pi_{VD}) + \Omega(q) \tag{1}$$

$$\mathcal{L}(f, q, \pi_{VD}) = \sum \pi_{VD}(PVD_i)(f(PVD_i) - q(PVD_i))^2 \tag{2}$$

15개의 고장 증상에 대해 고장 기록 데이터의 TM 제목에 대한 LIME 분석을 진행한다.

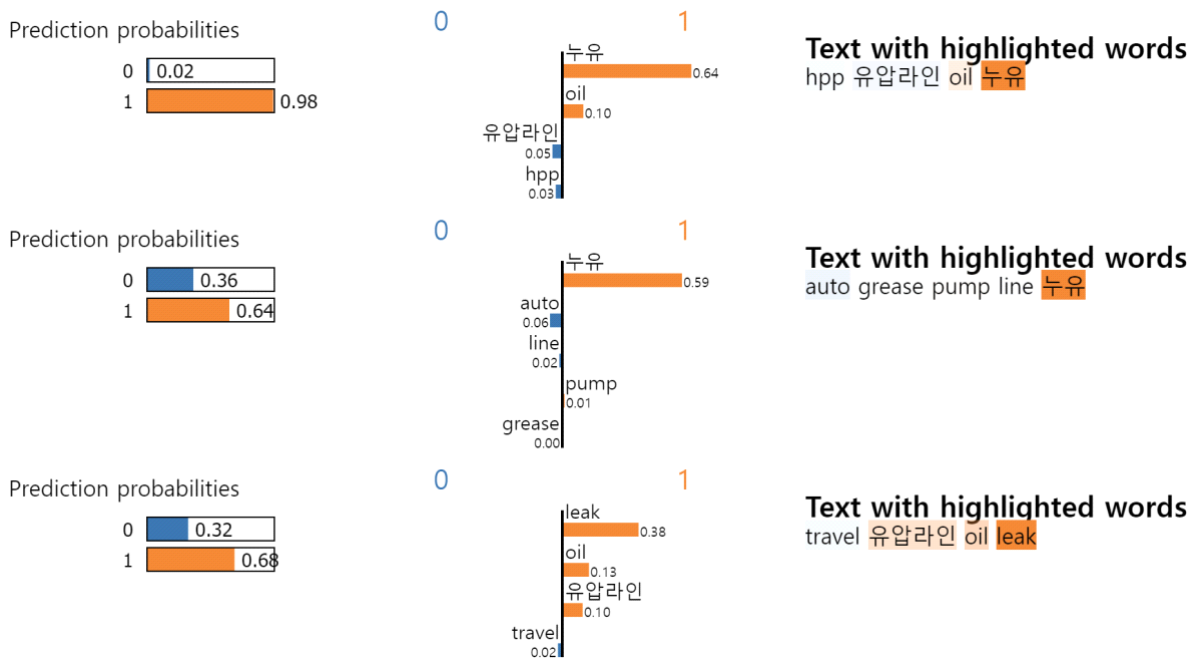


Figure 4. Example of LIME output

모델의 예측에 어떤 특징들이 영향을 미쳤는지에 대한 LIME의 설명을 나타낸 예시는 Figure 4와 같다. 이를 토대로 특정 고장증상의 예측에 긍정 및 부정적 영향을 끼친 단어를 시각적으로 확인할 수 있다.

LIME 단어사전 구축 절차는 <Algorithm 2>와 같다.

Table 5. System for constructing word dictionary

| <i>Algorithm 2.</i> | |
|--|---|
| <i>System for constructing word dictionary</i> | |
| Input: | AK : All keywords constituting the specific failure data W_i : Weight of keyword obtained through LIME analysis N_i : Number of appearances of keyword |
| Output: | WD : Word dictionary with keywords and average weight |
| Procedure: | $WD = \{ \}$ For each keyword K in AK : $AW_k = \left(\sum_{i=1}^{N_k} W_i \right) / N_k$ If, $AW_k \geq 0$: Add K and AW_k to the Word Dictionary WD return WD |

각 고장 증상별로 데이터를 구성하는 모든 Keyword 별 가중치 합이 평균값을 취득한 후, Keyword 별 가중치 합이 평균값이 0 이상인 단어들로 고장 증상별 단어사전을 구축한다.

Table 6. Example of word dictionary by model (Shortage)

| Random Forest | | | Logistic Regression | | |
|---------------|-----------|--------|---------------------|---------|--------|
| | Keyword | Weight | | Keyword | Weight |
| 1 | level | 0.4201 | 1 | level | 0.4486 |
| 2 | empty | 0.3370 | 2 | empty | 0.2913 |
| 3 | grease | 0.3116 | 3 | grease | 0.2666 |
| 4 | low | 0.2668 | 4 | 부족 | 0.2065 |
| 5 | 부족 | 0.2296 | 5 | low | 0.2057 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 20 | starboard | 0.0026 | 15 | cable | 0.0005 |

| XGBoost | | | SVM | | |
|---------|---------|--------|-----|---------|--------|
| | Keyword | Weight | | Keyword | Weight |
| 1 | level | 0.6693 | 1 | empty | 0.6799 |
| 2 | grease | 0.5474 | 2 | 부족 | 0.6546 |
| 3 | empty | 0.3396 | 3 | level | 0.5533 |
| 4 | 부족 | 0.3351 | 4 | grease | 0.2804 |
| 5 | oil | 0.2845 | 5 | low | 0.2700 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 14 | side | 0.0158 | 13 | 보충 | 0.0001 |

선정 모델별로 구축된 단어사전의 예시는 Table 6와 같다. 구축된 단어사전을 토대로, 다양한 작업자들이 공통적으로 주요하게 여기는 키워드를 파악할 수 있다 [Appendix A].

3.4 XAI 기반 발전설비 고장 분류 시스템

본 연구에서는 3.2절에서의 고장별 예측 확률과 3.3절에서 구축한 LIME 기반 단어사전을 통합하여 XAI 기반 발전설비 고장 분류 시스템을 제안한다.

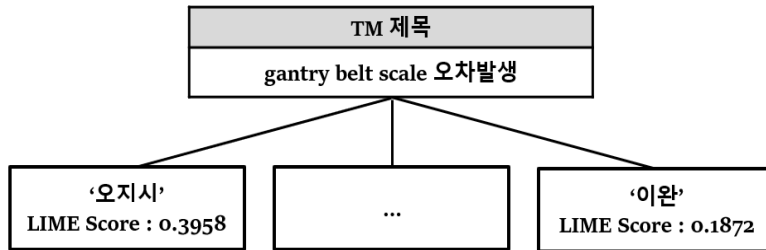


Figure 5. Example of LIME Score derivation process

TM 제목에 각 고장 증상별 단어사전에 해당하는 단어가 포함되어 있을 경우, 해당 단어의 가중치를 합산하여 Figure 5와 같이 고장 증상별로 LIME Score를 도출한다.

XAI 기반 발전설비 고장 분류 시스템 구축 절차는 <Algorithm 3>와 같다.

Table 7. System for XAI-based failure classification of power generation facilities

| <i>Algorithm 3.</i> | |
|---|--|
| <i>System for XAI-based failure classification of power generation facilities</i> | |
| Input : | <i>TM titles</i> : Trouble memo data for the facilities <i>WD_i</i> : Failure specific word dictionary (15 types of failure) <i>Failure</i> : Type of failure <i>predict proba</i> : Prediction Probability Function for Classification |
| Output : | <i>FC</i> : Classification of each TM title based on the highest total score |
| Procedure : | <ol style="list-style-type: none"> 1. Process of assigning LIME Scores based on word dictionary <ul style="list-style-type: none"> For each <i>title</i> in <i>TM titles</i>: For each <i>WD_i</i> (For <i>i</i> from 1 to 15) : Set <i>LIME Score[title]</i> = 0 For each <i>word</i> in <i>title</i> : If <i>word</i> in <i>WD_i</i> : <li style="padding-left: 20px;"><i>LIME Score[title][i]</i> += <i>WD_i[word]</i> 2. Scaling of TM data with assigned LIME Scores <ul style="list-style-type: none"> For each <i>Failure_i</i> (For <i>i</i> from 1 to 15) : For each <i>title</i> in <i>TM titles</i> : |

Algorithm 3.*System for XAI-based failure classification of power generation facilities*Scale $LIME\ Score[title][i]$ between 0 and 1 using Min-Max Scaling**3. Final failure classification through soft voting**Get $proba$ for each $Failure$ using $predict\ proba$ For each $Failure_i$ (For i from 1 to 15)

$$\text{Calculate } Total\ Score[title][i] = \frac{Scaled\ LIME\ Score[title][i] + proba[title][i]}{2}$$

For each $title$ in $TM\ titles$:Assign $FC[title] = Failure$ with the highest $Total\ Score$ return FC

각 TM 제목의 고장 증상별로 도출한 LIME Score와 3.2절에서 predict_proba 기법을 통해 구한 고장 증상별 예측 확률을 Soft Voting 하기 위해 LIME Score를 0~1 사이의 값으로 Scaling 한다. predict_proba 기법을 통해 얻은 고장 증상별 예측 확률은 0~1 사이의 값을 가진다. 이를 다른 지표와 함께 결합하여 Soft Voting을 진행하기 위해서는 예측의 일관성 및 정확성을 보장하기 위해 모든 지표가 동일한 스케일을 가지고 있어야 한다. 따라서 본 연구에서는 Min-Max Scaler를 통해 LIME Score를 Scaling 한다. Soft Voting을 진행하여 도출된 값을 Total Score로 산정하고, TM 제목에 대한 고장 증상별 Total Score 중 가장 높은 값을 가지는 고장 증상으로 최종 분류한다.

4. 실험 결과

본 실험에서 사용한 하드웨어는 NVIDIA RTX 4090 GPU, RAM 32GB이며, 모든 모델 학습은 동일한 환경에서 진행한다. 본 연구는 앞서 분리한 데이터로 모델을 학습시키고 단어사전을 구축한 다음, 검증 데이터에 대해 설비 고장 증상 예측 및 분류를 진행한다. 고장 증상별 데이터가 불균형하게 분포되어 있기에, Accuracy만을 사용하여 성능을 도출할 경우 결과가 편향되어 모델의 성능을 정확하게 반영하기 어렵다(Chawla, 2002; He, 2009; Hong et al., 2024). 이에 Precision과 Recall 값의 조화평균인 F1 Score를 통해 클래스 간 데이터가 불균형해도 더욱 정확한 성능 평가를 가능하게 하였다. 또한 모든 고장 종류에 대한 동등한 성능 평가를 위하여, 데이터 수가 적은 클래스의 가중치를 높일 수 있는 Macro_F1(Macro average F1 Score)를 사용하였다(Vong, 2020).

Table 8. Confusion matrix

| Actual \ Predicted | True | False |
|--------------------|------|-------|
| | True | TP |
| False | FN | TN |

Table 8은 다중 분류기에서 주로 사용되는 분류기의 성능 검증 방법론인 혼동 행렬을 의미한다(Lee and Hwang, 2015). 본 실험에서는 이를 통해 Accuracy, F1 Score, Macro_F1을 도출하여 모델의 성능을 평가한다.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1\ Score = 2 \times \frac{precision \times recall}{precision + recall} \tag{6}$$

$$Macro\ F1 = \frac{\sum_{i=1}^N F1\ score\ of\ i\ th\ class}{N} \quad (N : Total\ number\ of\ classes) \tag{7}$$

해당 성능 평가 지표들은 (3)~(7)과 같이 표현할 수 있으며(Lee et al., 2021), 실험 결과는 Table 9와 같다.

Table 9. Experimental results

| | Random Forest | | Logistic Regression | | XGBoost | | SVM | |
|--------------------|---------------|-------|---------------------|-------|----------|-------|----------|-------|
| | Before | After | Before | After | Before | After | Before | After |
| | F1 Score | | F1 Score | | F1 Score | | F1 Score | |
| 고착, 간섭 | 0.94 | 0.94 | 0.91 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| 누설 | 0.94 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 1.00 | 1.00 |
| 단선, 단락 | 0.82 | 0.88 | 0.71 | 0.88 | 0.88 | 0.88 | 0.80 | 0.88 |
| 마모 | 0.97 | 0.96 | 0.96 | 0.97 | 0.96 | 0.97 | 0.98 | 0.98 |
| 막힘 | 0.00 | 0.50 | 0.00 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| 변형, 휨 | 0.89 | 0.91 | 0.90 | 0.90 | 0.92 | 0.91 | 0.91 | 0.91 |
| 부식, 침식 | 0.91 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 부족 | 0.93 | 0.94 | 0.89 | 0.94 | 0.91 | 0.94 | 0.93 | 0.91 |
| 소손, 손실 | 0.89 | 0.89 | 0.84 | 0.83 | 0.92 | 0.92 | 0.91 | 0.89 |
| 오동작, 동작불량, 한계초과 | 1.00 | 1.00 | 0.89 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 |
| 오염, 변질 | 0.87 | 0.87 | 0.56 | 0.87 | 0.87 | 0.87 | 0.82 | 0.92 |
| 오지시 | 0.00 | 0.40 | 0.00 | 0.40 | 0.40 | 0.89 | 0.40 | 0.67 |
| 이완 | 0.82 | 0.81 | 0.65 | 0.70 | 0.82 | 0.78 | 0.77 | 0.78 |
| 이음, 소음 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 파손, 파열 | 0.88 | 0.88 | 0.89 | 0.89 | 0.89 | 0.89 | 0.90 | 0.89 |
| Macro_F1 | 0.79 | 0.86 | 0.68 | 0.85 | 0.86 | 0.89 | 0.85 | 0.88 |
| Accuracy | 0.92 | 0.94 | 0.89 | 0.92 | 0.93 | 0.93 | 0.93 | 0.94 |

단일 모델만으로 고장을 분류하였을 때(Before)와 XAI 기반 발전설비 고장 분류 시스템을 적용한 이후(After)의 결과를 비교하였을 때, 적용 이후의 Macro_F1과 Accuracy는 본 연구의 시스템을 적용하기 전보다 같거나 높은 성능을 보였다. Macro_F1을 기준으로, XAI 기반 발전설비 고장 분류 시스템을 적용한 모델 중 XGBoost가 89%로 가장 좋은 성능을 보였으며, Accuracy 기준으로는 Random Forest, SVM이 94%의 성능을 보였다. 또한 XAI 기반 발전설비 고장 분류 시스템을 적용하였을 때, Logistic Regression 모델에서 Macro_F1이 0.17 향상되어 다른 모델들에 비해 가장 높은 성능 향상을 보였다. 고장 증상별 F1 score를 비교하였을 때, 데이터가 많은 고장 증상에 대한 분류 성능은 유지하면서, ‘오지시’와 같이 상대적으로 데이터가 적은 고장 증상에 대한 분류 성능이 최대 2.23배 상승하였다.

5. 결론

본 연구는 발전소에 석탄을 공급하는 설비인 연속식 석탄 하역기에 대한 고장 기록 데이터를 바탕으로 XAI 기반 발전설비 고장 분류 시스템을 제안한다.

4가지 머신러닝 모델인 Random Forest, Logistic Regression, XGBoost, SVM을 통해 고장 증상별 예측 확률을 도출한다. 15개의 고장 증상별 텍스트 데이터에 대해 XAI 기법 중 하나인 LIME을 이용하여, Keyword 별 가중치 합인 평균값이 0 이상인 주요 Keyword 들로 고장 증상별 단어사전을 구축한다. 이를 기반으로 얻은 고장 증상별 LIME Score와 예측 확률을 Soft Voting을 통해 결합하여 도출한 Total Score를 토대로 최종 분류한다. 그 결과, XGBoost 모델에서 Macro_F1 기준 89%로 가장 높은 성능을 보였다. 또한 4가지 모델 중 Accuracy는 최대 0.03, Macro_F1은 0.17의 상승을 보였다.

고장 기록 데이터는 설비 및 시스템의 운영과 유지보수에 있어 근본적인 역할을 수행한다. 그러나 기존 고장 증상 분류 과정에는 분류 가이드라인이 부재한 상태에서 작업자가 경험적 지식에 의존하여 직접 고장 기록을 작성 및 분류하였다. 이로 인해 데이터 품질 저하에 대한 문제가 유발되었기에, 품질 개선의 필요성이 대두되었다. 본 연구에서 제안한 XAI 기반 발전설비 고장 분류 시스템을 적용할 경우, 데이터 품질을 개선하여 설비 고장 분석 및 유지보수와 같은 연구의 전반적인 신뢰성을 향상시킬 수 있을 것이다. 이는 곧 비용 절감 및 발전설비의 안정성과 수명 연장뿐만 아니라, 더 나아가 발전소 운영 효율성 및 안정성 향상에 기여할 수 있을 것이다.

본 연구는 작업자에 의해 수기로 기록되는 고장 기록 데이터의 특성상, 데이터의 양이 적은 고장 증상의 경우 고장 증상의 특성을 충분히 반영하지 못할 수 있다는 한계가 있다. 이로 인해 새로운 데이터에 대한 유연성이 떨어질 수 있으므로, 향후 작업자와의 논의를 거쳐 데이터 증강을 고려해볼 수 있을 것이다. 또한 본 연구에서는 연속식 석탄 하역기만을 대상으로 분석을 진행하였으나, 향후 연구에서는 연속식 석탄 하역기 이외에도 더 많은 발전설비에 XAI 기반 발전설비 고장 분류 시스템을 적용해 볼 수 있을 것이다. 이를 토대로, 특정 설비에 국한되지 않은 전반적인 발전소 운영 효율성 및 신뢰성을 향상시킬 수 있을 것이다. 또한 본 연구에서 제안하는 시스템을 바탕으로, 현장에서 고장 기록 입력 시 고장 증상이 자동 분류되는 프로세스를 구축할 수 있을 것이다.

REFERENCES

- Batelić J, Griparić K, and Matika D. 2021. Impact of Remediation-Based Maintenance on the Reliability of a Coal-Fired Power Plant Using Generalized Stochastic Petri Nets. *Energies* 14(18):5682.
- Blanco, M. A., Martí-Puig, P., Gibert, K., Cusidó, J., and Solé-Casals, J. 2019. A text-mining approach to assess the failure condition of wind turbines using maintenance service history. *Energies* 12(10):1982.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16:321–357.
- Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang, C. H., and Liu, R. 2019. XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system. *IEEE Access* 7:13149–13158.
- Han, M., Kim, Y., and Kim, N. 2022. The Effect of Domain Specificity on the Performance of Domain-Specific Pre-Trained Language Models. *Journal of Intelligence and Information Systems* 28(4):251–273.
- He, H., and Garcia, E. A. 2009. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* 21(9):1263–1284.
- Hong, J., Jung, Y., and Kang, S. 2024. Defect Prediction and Variable Impact Analysis in CNC Machining Process. *Journal of the Korean society for Quality Management* 52(2):185–199.
- Hong, S., Kim, J., and Yang, E. 2022. Automated text classification of maintenance data of higher education buildings using text mining and machine learning techniques. *Journal of Architectural Engineering* 28(1):04021045.
- Hwang, S., Kim, Y., and Kang, S. 2023. Study of Reliability Analysis Based Power Generation Facilities Maintenance System – Focused on Continuous Ship Unloader –. *Journal of the Korean society for Quality Management* 51(2):315–327.
- Jardine, A. K., Lin, D., and Banjevic, D. 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 20(7):1483–1510.
- Kim, K., Park, J., Lee, H., Jeong, H., Kim, H., and Kim, H. 2017. A Study on Fault Diagnosis of Boiler Tube Leakage based on Neural Network using Data Mining Technique in the Thermal Power Plant. *Transactions of The Korean Institute of Electrical Engineers* 66(10):1445–1453.
- Kim, M., Jung, S., Kim, J., Kim, S. 2021. A Study on Fault Detection Method Using Weight-based Local Outlier Factor, *Journal of Korean Institute of Intelligent Systems* 31(4):273–278
- Korea Energy Economics Institute. KEEI Yearbook of Energy Statistics 2022; Korea Energy Economics Institute: Ulsan, Korea, 2023.
- Lee, H., Hong, Y., and Kang, S. 2021. Identifying process capability index for electricity distribution system through thermal image analysis. *Journal of Korean Society for Quality Management* 49(3):327–340.
- Lee, S., and Hwang P. 2015. Industrial and Technological Trends of Personal Mobility in Korea and Abroad. *Journal of the Korean Society of Automotive Engineers* 37(1):36–45.
- Lee, S., and Kim, Y. 2022. A Pre-processing Process Using TadGAN-based Time-series Anomaly Detection. *Journal of the Korean society for Quality Management* 50(3):459–471.
- Lee, Y., Oh, J., and Kim, G. 2020. Interpretation of load forecasting using explainable artificial intelligence techniques. *The Transactions of The Korean Institute of Electrical Engineers* 69(3):480–485.
- Lim, C., and Han, S. 2018. A study on development of power grid fault prediction system based on big data and preceding activities to calculate optimal investment cost. *Journal of the Korean Data And Information Science Society* 29(3):779–794.
- Liu, J., Ma, H., Xie, X., and Cheng, J. 2022. Short text classification for faults information of secondary equipment

- based on convolutional neural networks. *Energies* 15(7):2400.
- Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., and Levron, Y. 2022. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI* 9:100169.
- Meng, Q., Song, Y., Mu, J., Lv, Y., Yang, J., Xu, L., et al. 2023. Electric power audit text classification with multi-grained pre-trained language model. *IEEE Access* 11:13510–13518.
- Park, J., Jo, H., Lee, S., Oh, S., and Na, M. 2022. A reliable intelligent diagnostic assistant for nuclear power plants using explainable artificial intelligence of GRU-AE, LightGBM and SHAP. *Nuclear Engineering and Technology* 54(4):1271–1287.
- Rahman, U., and Mahbub, M. 2023. Application of classification models on maintenance records through text mining approach in industrial environment. *Journal of Quality in Maintenance Engineering* 29(1):203–219.
- Ribeiro, M. T., Singh, S., and Guestrin, C. 2016. “Why should I trust you?” Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
- Sairam, S., Srinivasan, S., Marafioti, G., Subathra, B., Mathisen, G., and Bekiroglu, K. 2020. Explainable incipient fault detection systems for photovoltaic panels. *arXiv preprint arXiv:2011.09843*.
- Sarker, I. H. 2021. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science* 2(6):420.
- Tian, X., Li, C., and Zhao, B. 2021. A novel classification model SA-MPCNN for power equipment defect text. *Transactions on Asian and Low-Resource Language Information Processing* 20(6):1–21.
- Vong, C. M., and Du, J. 2020. Accurate and efficient sequential ensemble learning for highly imbalanced multi-class data. *Neural Networks* 128:268–278.
- Wang, H., Liu, Z., Xu, Y., Wei, X., and Wang, L. 2020. Short text mining framework with specific design for operation and maintenance of power equipment. *CSEE Journal of Power and Energy Systems* 7(6):1267–1277.
- Zednik, C. 2021. Solving the black box problem: A normative framework for explainable artificial intelligence. *Philosophy & Technology* 34(2):265–288.

저자소개

- 김유림** 인하대학교 산업경영공학과 학사를 취득하였다. 현재 인하대학교 산업경영공학과 석사과정에 재학 중이며 주요 관심 분야는 빅데이터 분석, 데이터 사이언스이다.
- 박정인** 인하대학교 산업경영공학과 학사를 취득하였다. 현재 인하대학교 산업경영공학과 석사과정에 재학 중이며 주요 관심 분야는 데이터 사이언스, 머신러닝이다.
- 박동현** 인하대학교 산업경영공학과 교수로 재직 중이다. 연구 분야는 인간공학 및 WMSDs이다.
- 강성우** 교신저자. 인하대학교 산업경영공학과에서 학사를, 펜실베니아 주립대학교 산업제조공학과에서 석사와 박사 학위를 취득하였다. 현재 인하대학교 산업경영공학과 부교수로 재직 중이다. 관심 연구 분야는 빅데이터(3D 이미지, 텍스트 데이터) 프로세싱을 기반한 제품 설계, 공학 설계, 생산 장비 예측 진단 및 관리이다.