JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 24, No. 7, pp. 1567-1574, Jul. 2023

Check for updates

CPS 컨텍스트 분석을 활용한 신재생에너지 설비 비지도학습 기반 이상탐지 모델 연구

유 진 도*
*고려대학교 정보보호대학원 석사과정

Anomaly Detection System for Renewable Energy Facilities based on Context Analysis using Cross-sectional Data of CPS System

Jin-do Yu*
*Master's Course, Division of Information Security, Korea University, Seoul 02841, Korea

[요 약]

발전산업망은 전형적인 폐쇄망 구조로 안정적인 시스템 운영이 필수적인 시스템이다. 국내 산업망은 에너지원 다각화를 위해 신재생에너지 비율 확대를 지속적으로 추진하고 있다. 현재 발전회사가 운영중인 신재생에너지 발전단지는 기존 대규모 발전설비와는 달리 설비 자체가 외부에 노출되어 있고, 개별 발전 용량이 적은 한계로 인해 폐쇄망 정책을 적용하기 어려운 문제가 있어서 현실적으로 차별화된 접근이 필요하다.

풍력 발전망에서 취득 가능한 각종 센서값, 관리시설에서 얻을 수 있는 발전설비 상태데이터, 전력거래소로부터 수신 받는 신재생에너지 출력값 등 물리 데이터와 통신 데이터를 비교·분석하여 종합적인 맥락을 분석하고, 이상발생시 이를 해당 발전시스템에 알람 메시지를 통보하여 상대적으로 소규모로 오지에 분산되어 있는 신재생에너지 운영시스템의 효율적 운영체계의 연구를 통해 사이버 물리 시스템 (CPS: Cyber Physical System)에서 종단와 횡단의 이종 영역간(cross-sectional) 데이터를 머신러닝 비지도학습의 비교 분석으로 이상여부 감지와 데이터 교정으로 개방망에서 상호감시를 통해 데이터의 무결성을 검증하여 이상여부를 판단하고, 사이버 위협 측면의 속성 강화 여부를 판단하는 탐지모델 연구로 마이크로그리드의 일종인 신재생 발전단지의 통합 운영에 적합한 이상탐지 모델로 확장 적용하고자 한다.

[Abstract]

The power generation industrial network is a typical closed network structure that requires stable system operation. The domestic industrial network continuously promotes the expansion of renewable energy to diversify energy sources. Unlike existing large-scale power generation facilities, renewable energy power complexes operated by power generation companies are exposed to the outside world. It is difficult to apply a closed network policy due to the limitations of small individual power generation capacity, therefore, a differentiated approach is needed in reality.

색인어 : 이상 탐지, 사이버 물리보안, 기계학습, 신재생에너지

Keyword : Anomaly Detection, CPS, Machine Learning, Renewable Energy Facilities

http://dx.doi.org/10.9728/dcs.2023.24.7.1567

Received 05 June 2023; **Revised** 22 June 2023
Accepted 26 June 2023

***Corresponding Author; Jin-do Yu**
Tel: ※ 개인정보 표시제한
E-mail: jinn05@naver.com

Ⅰ. 서 론

전력시스템은 국가에 의해 보안성이 관리되는 시설로 중요 인프라(Critical Infrastructure)에 속한다. 최근 재생에너지 기반의 분산전원 확산에 따라 국가적 관리 범위 밖에 위치하는 시스템(분산전원단지, 마이크로그리드, 커뮤니티 에너지시스템 등)이 증가하고 있다. 이러한 시스템은 향후 지속적으로 증가할 것이며, 이는 새로운 전력시스템의 잠재적 위협요소가 될 수 있다. 우리나라에서 전통적인 전력시스템은 전용 통신망에 기반한 폐쇄망으로 운영되어 왔으나, 분산형 재생에너지 시스템의 경우 필연적으로 개방형 통신망에 연계될 수 밖에 없으며, 기존 화력발전소와는 달리 출력제어가 아닌 예측을 하여야 하며, 이와 함께 보급되는 에너지저장장치(ESS: Energy Storage System, ESS)와 마이크로그리드는 운영에 있어 자율성을 가지게 되므로, 전력계통 운영의 패러다임이 완전히 변하게 되고, 더불어 보안정책에 있어서도 기존 전력시스템에 적용되던 방식과는 완전히 다른 접근이 요구된다[1].

이러한 배경에서 재생에너지 단지에서 이상을 탐지하기 위한 방법론과 프로토타입 프로그램 개발에 대한 연구로서 시계열 분석과 단면데이터 분석을 결합한 패널 데이터 분석 기반의 알고리즘을 제안한다. 최근 예측·분석 툴의 경향은 기존 알고리즘 및 연산엔진 중심에서 데이터 중심으로 전환되고 있으며, 모델링 자체가 데이터 특성에 의존적인 성향이 강해지고 있다. 본 연구에서는 머신러닝 앙상블 기법을 통한 출력예측을 통해 모델의 성능을 검증하고자 한다. 앙상블은 다수의 머신러닝 모델을 이용해 최적의 답을 찾아내는 기법으로 다수 모델을 이용하여 데이터를 학습하고, 모든 모델의 예측결과 중에서 정확도가 가장 높은 모델을 선택하는 모델이기에, 본 연구에서는 풍력 데이터 머신러닝 모델링 프로세스 방법론 중심으로 수립된 풍력발전 ADS(Anomaly Detection System) 모델이 데이터에 맞는 모델링을 취사 선택할 수 있는 형태에 적합한 접근 여부에 대해 성산풍력발전의 실데이터를 기반으로 학습을 선행하고, 스플렁크에 성산풍력발전의 실데이터를 업로드하여 탐지모델에 대한 결과를 비교 검증을 통해 모델 검증을 수행하고자 한다[12].

Ⅱ. 본 론

2-1 관련연구

이상탐지(Anomaly Detection)는 비정상 데이터를 탐지하는 기법으로 이상 데이터의 샘플을 보유하고 있다면 측정결과의 정확도를 높일 수 있다. 그러나, 이상 데이터에 대한 샘플을 확보하는 것은 쉽지 않고, 특히 발전시설의 사이버 보안 측면이라면 더욱 힘들 개연성이 크다. 이상탐지는 데이터 셋의 레이블 가용도에 따라 모델이 구분된다. 학습 데이터 셋에 비정상적인 샘플이 포함되는지, 각 샘플에 레이블이 존재하는지, 비정상적인 샘플의 성격이 정상 샘플과 어떻게 다른지, 정상 샘플의 클래스가 단일 클래스 인지 멀티 클래스 인지 등에 따라 다른 용어를 사용한다. 본 절은 데이터 샘플 여부 확보에 따른 이상탐지 방법론의 분류기법으로 지도식 이상탐지, 반지도식 이상탐지, 비지도식 이상탐지로 구분할 수 있고, 구성항목은 그림 1과 같다[2],[3],[8]-[11].

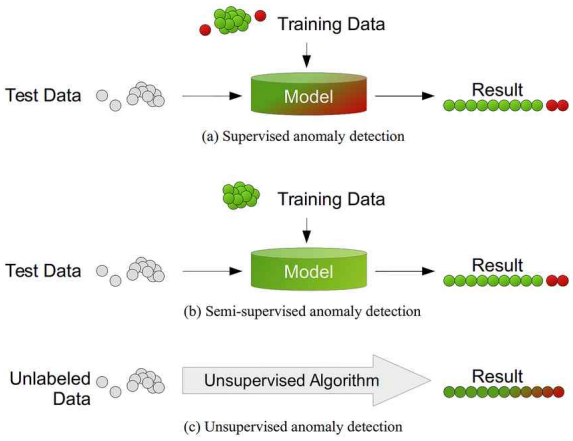

그림 1. 이상탐지 모델
Fig. 1. Anomaly detection model

이상에는 특정 값이 나머지 값들 대비 이상하다는 점 이상과, 하나의 값이 아닌 전체적인 맥락에서의 이상 여부를 의미하는 맥락적 이상으로 구분되며, 맥락 관련 변수는 다음과 같이 정의된다.

가. 맥락적 속성: 맥락 변수(contextual attribute)로 표현되며, 시공간 상에서 관계성을 가지는 변수들을 의미한다. 예를 들면, 재생에너지 발전량 변수와 기상데이터 변수간 관계성이 맥락 변수를 의미한다.

나. 행동적 속성: 행동 변수(behavioral attribute)라고도 하며 맥락적이지 않은 특성을 나타낸다. 예를 들면, 기상 데이터 중 평균 풍속에 대한 데이터가 있을 때, 각 위치에서의 데이터는 상호관계에 관계가 없는 행동 데이터에 해당한다.

이를 통계적 관점에서 봤을 때, 맥락 변수는 종속변수, 행동 변수는 독립변수가 된다. 이 경우 이상 여부는 특정한 맥락에서 행동변수의 값으로 판단하게 되는데, 맥락상 정상이라고 판단되는 일정 범주 내에 행동변수가 존재하지 않는다면, 그것은 모종의 이상이 있다고 판단할 수 있는 것이다.

제어망 이상징후 탐지방법은 크게 두가지 형태로 분류할 수 있다. 첫째 비정상행위 탐지기법(Anomaly based detection)과 시그니처 기반탐지 기법(Signature based detection, misuse detection)이다. 비정상행위 탐지기법은 정상적인 범주를 벗어나는 비정상 트래픽을 감시할 수 있으며, 새로운 공격 탐지가 가능하다. 그러나, 시스템 구현에 어려움이 있고 오탐으로 인한 오경보 가능성이 높다. 시그니처

기반탐지 기법은 시스템 및 프로그램의 알려진 취약점을 기반으로 탐지하므로 탐지율이 우수한 편이나, 새로운 공격에 대한 탐지률이 반영되기전까지는 탐지가 불가능하다는 문제점이 있다. 전력제어망 이상징후 탐지기법의 한계를 극복하기 위해 산업현장의 실제Contents를 취득 및 데이터의 흐름을 분석하고, 정상적인 데이터의 모델을 정립하고 이에 따라 이상징후를 판단할 수 있다[4].

제어시스템 환경에서 발생 할 수 있는 공격 및 오동작 유형을 15개 그룹으로 분류 후, Network계층, Protocol Specification, Control, Statistic로 이루어진 4계층으로 이루어진 화이트리스트를 제시하였다. 그러나, 화이트리스트(Whitelist) 기반 이상 행위 탐지 기법은 Protocol Specification과 Control Message 등 특정 Protocol 중심으로 제시가 되어있어 실제 비표준 프로토콜 기반의 폐쇄적인 특성을 가진 발전제어망에 적용하기 에는 한계가 있는 상황이다[5]-[7].

Ⅲ. 풍력발전 이상징후 탐지 모델

3-1 풍력발전 이상탐지 (ADS:Anomaly Detection System) 모델

이상징후를 탐지하고 표현하는 방법은 유효 데이터 여부를 검증하고, 유효 데이터에 대한 상관행렬을 산출한다. 1초 기반의 풍력 단지 분석 결과에 따르면 1개월치의 데이터에 있어서 상관계수 값의 변화는 있었으나, 발전소의 3개월 동안 1분 단위 데이터를 대상으로 이상탐지 모델을 설계하여 이상여부 판단을 수행한다. 모델 학습을 위해 인버터 데이터, 환경센서 데이터, 기상청 데이터를 활용하였으며, 세부 유형별 항목은 표 1에 제시하였다. 데이터의 기본 정보를 바탕으로 데이터 구조, 변수명, 숫자형, 문자형, 논리형의 변수별 데이터 유형, 결측값, 이상치, 데이터의 산점도와 분포 모양을 탐색하였다.

표 1. 모델 학습 데이터
Table 1. Model training data

Classification	Type
Inverter Data	Generation Amount, Cumulative Module Current
Sensor Data	Blade Speed, Blade Angle
Weather Data	Temperatures, Wind speed, Wind direction

수집한 풍력 발전소 현장/기상 정보 수집, 단면 데이터 판정, 데이터 분석/검증, 이상탐지의 절차를 거쳐, 예측값과 이상값을을 예측하여 실측값과 비교와 차이의 학습을 통해 탐지성능을 지속적으로 향상 시키기 위한 모델을 구현하고자 한다.

이러한 개념적 특성에 기반하여 이상탐지시스템의 전체적 구성은 그림 2과 같이 데이터 수집, 데이터 전처리, 데이터 분석, 모델링, 검증, 결과 분석의 단계로 되어 있다.

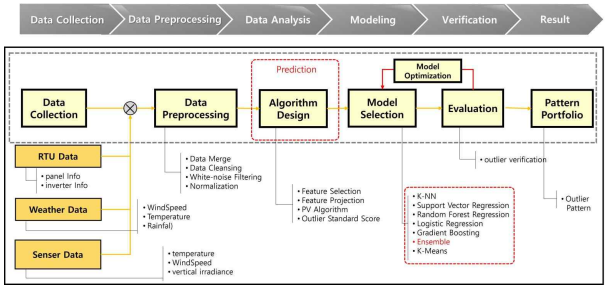

그림 2. 처리 절차
Fig. 2. Process

풍력발전 이상탐지 모델을 실제 시스템에서 운영할 수 있도록 서빙 시스템을 구축하여 신규 데이터 입력 시, 해당 모델을 통해 실시간으로 이상탐지 확인이 가능하다. 또한, 모델을 버전별 관리를 통해 발전소의 운영 특성에 맞추어 계통 안정성, 수익 최대화 등의 활용 목적에 따라 선별적으로 모델을 활용할 수 있고, 풍력발전 이상탐지 모델을 실제 시스템에서 운영할 수 있도록 서빙 시스템을 그림 3와 같이 구축하여 신규 데이터 입력 시, 해당 모델을 통해 실시간으로 이상탐지 확인 한다.

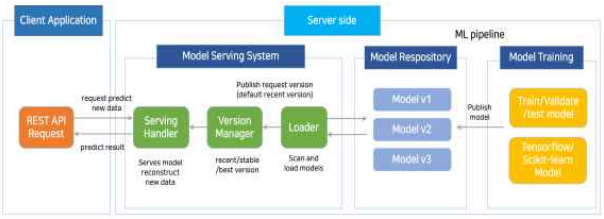

그림 3. 이상탐지 모델 서빙 아키텍처
Fig. 3. Anomaly detection model serving architecture

3-2 풍력발전 ADS(Anomaly Detection System) 방법론

세부 방법론은 CSV 기반의 수집 데이터 파일을 업로드하면, 유효 데이터 여부를 검증하고, 유효 데이터에 대한 상관행렬을 산출한다. 1초 기반의 성산풍력 단지 분석 결과에 따르면 1개월치에 데이터에 있어서 상관계수 값의 변화는 있었으나, 상대적인 순위는 변화가 거의 없었고 이 경우 상관계수 순위가 바뀌는 것은 이상치가 있음을 암시한다고 볼 수 있다. 이는 데이터 정상/이상을 판단하는 중요한 기준의 하나로 설정된다.

풍력발전 ADS의 수행절차와 구성항목간의 구성요소는 그림 4와 같이 수집된 원천 데이터를 제안 모델을 통한 이상징후 판단과 통지 수행의 과정으로 처리된다.

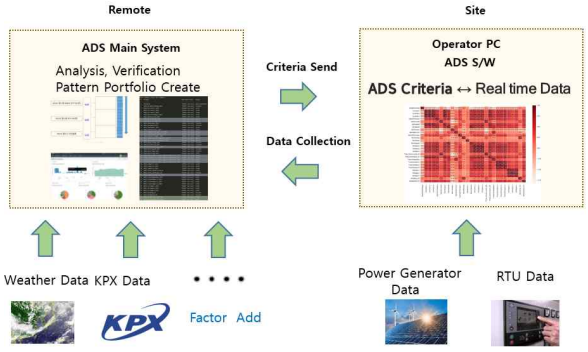

그림 4. 풍력 발전 이상탐지를 위한 데이터 수집
Fig. 4. Data collection for wind power generation anomaly detection

그림 5는 이상탐지를 위한 점수 표현 방법론으로 풍력발전 CSV파일을 통해 유효 데이터 필드 선정과 데이터 기본 분석을 통해 상관행렬의 순서 비교를 통해 이상탐지를 위한 점수 표현을 위한 절차이다.

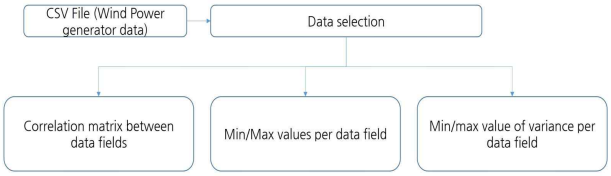

그림 5. 이상탐지를 위한 점수 표현 방법
Fig. 5. Score expression methodology for anomaly detection

2번째 범주로는 그림 6과 같이 데이터 필드별 실측 데이터의 변동 범위로 설정하였는데, 과거 데이터의 변동 범위에 대한 최소, 최대값을 설정하고 특정 데이터 필드가 해당 범위를 벗어나면(아웃라이어로 간주) 감점하는 형태로 신뢰도를 낮출 수 있다.

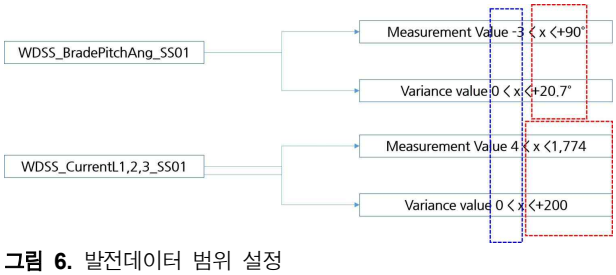

그림 6. 발전데이터 범위 설정
Fig. 6. Data range measurement

개별 변수의 제약 조건은 1초(데이터 최소 주기)당 변량, 특성값의 최대-최소 범위로 제한되는데, 이를 기반으로 시계열 분석을 적용할 수 있다. 단면 데이터 분석을 수행하여 변수간 상관관계 분석을 통한 정량 매트릭스(상관행렬)를 생성할 수 있고, 2가지 분석 항목들에 대한 가중치 적용하여 트러스트(신뢰도) 기반의 점수화를 시도하는 것이 기본적 접근 개념이다. 유의미한 특성 변수에 대해, 현재 측정값과 변량에 대한 범위 내 여부를 검토하는 것이다. 예를 들면, 데이터 필드가 10개 존재하고 이중 1개의 데이터 필드가 범위를 벗어나게 되면 -10점을 감점하는 방식이다. 데이터 필드가 100개가 되면 -1점을 감소시키는 형태로 개별 데이터의 단위 감점 수준이 떨어지게 된다. 또한, 일정 시간 주기동안에 이러한 개념을 적용한다면 시간대별로 점수를 다시 할당하거나 중복 감점을 적용할 수 있다. 이를 일반화하면 다음 그림 5와 같은 형태로 3가지 범주에 대한 100점 만점의 점수를 할당하고 이를 기반으로 점수가 차감될수록 이상으로 분류될 확률이 높아지는 개념으로 이해할 수 있다.

Ⅳ. 실험 결과

제안 모델의 성능을 검증하기 위해, 실제 풍력발전 데이터를 활용하여 구성하였다. 실험에 활용한 풍력발전 데이터는 성산풍력 1호기 데이터로 1초 주기의 데이터를 1주일, 1개월 구간의 자료로 분류하였다. 해당 데이터에 대한 기본적 분석을 수행하여 데이터의 가용성과 시계열적 특성을 분석하고, 변수간 상관관계를 분석하는 절차를 통해 풍력발전을 비롯한 국내 발전망은 국가의 규제하에 운영되기에 사설 발전의 경우라도 활용 가능한 포괄 모델을 검증하였다.

풍력발전 로그를 받아서 시각화 해주는 서버와 함께, 스플렁크는 빅데이터 플랫폼으로 여러가지 다양한 로그를 분석할 수 있게 해준다. 풍력로그를 스플렁크 시스템을 이용하여 분석을 진행하고 시각화 대시보드를 제작하여 검증하였다.

데이터를 스플렁크에 insert 하여, 아래와 같이 WDSS_HydroOilTemper_SS01에 대해서 시간에 따른 평균, 시간에 따른 최대값/최소값, 상위 값, 시간별 상위 값, 회귀 값 등에 대한 분석을 진행하였으며, 그림 7과 같이 풍력로그를 1차원 데이터 통계를 이용하여 선 그래프, 파이차트로 구성하였다. 또한 4일 간격으로 해당 데이터가 어떤 추이를 그리는지 그래프의 연관성을 통해 최대, 최소 그리고 변량값을 확인할 수 있도록 구성하였고, 날개각도와 발전기전류 A, B, C는 서로 그래프의 상관관계를 분석하여 추이를 확인하였다.

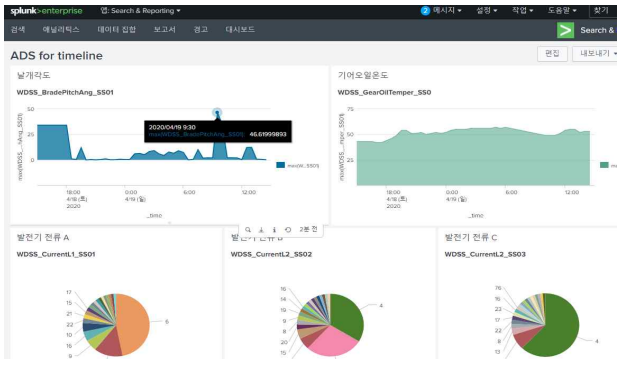

그림 7. 모델 서빙을 통한 예측 결과
Fig. 7. Predicted results through model serving

4-1 데이터 추이

풍속은 발전출력에 가장 큰 영향을 미치는 변수이므로 출력과의 관계를 살펴야 한다. 그림 8과 그림 9와 같이 출력과 풍속 그래프를 보면 두 변수는 선형 관계를 보이며, 풍속 3m/s 이상 구간에서 대부분의 발전 출력을 보인다. 만일, 일정 수준 이상의 풍속에도 출력이 매우 낮거나 없는 경우에는 해당 시점에서의 다른 변수의 값 및 변량 등을 통해 이상 여부를 확인해야 한다.

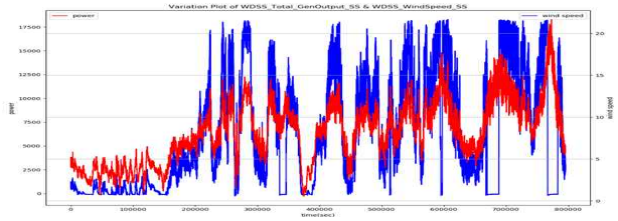

그림 8. 풍속에 따른 발전출력 변화
Fig. 8. Change in power output due to wind speed

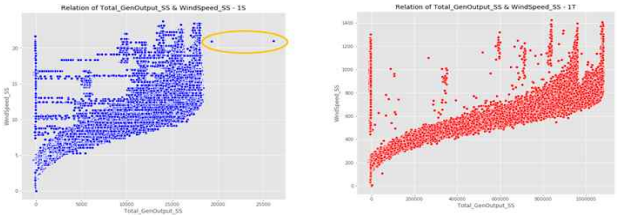

그림 9. 1초 및 1분 단위 발전출력과 풍속 관계
Fig. 9. Relationship between wind speed and power generation of 1sec, 1min

데이터의 전처리로 수집 가능한 필드를 모두 활용하였으며, 수집된 데이터 중 결측이 발생한 경우 선형보간법으로 대치 하여 보완하였다.

발전기 전류는 발전기의 기어오일온도, 발전출력, 발전기 속도 등과 연계성을 가지게 된다. 예를 들어 발전출력이 지속적으로 높으면 기어오일온도는 상승할 확률이 높으나 급격한 변화를 보이지는 않는다. 이러한 특성을 기어오일온도 측정값과 변량에 적용하여 그림 10과 같이 이상여부를 탐지하는 재료 주요 데이터로 검증하였다.

그림 10. 풍력발전 주요 데이터
Fig. 10. Key data for wind power plants

4-2 결과분석

풍력 데이터는 그림 11과 같이 발전기 출력을 포함하여 44개의 필드로 구성되어 있으며, 각 필드는 1초 단위로 2월 한 달(02.01~02.29) 데이터와 4월 한 주(04.14~04.23) 데이터로 필드별로 3,301,205개의 데이터를 갖는다. 필드 유형에 따라 데이터는 Numerical Type과 Categorical Type으로 구분되며, 13개의 Categorical Type이 존재한다. 이 중 12개 필드는 Pt_Created를 나타내는 미사용 Tag이며, 마지막 44번째 필드인 WDSS_WindStatus_SS01는 풍력설비 1호기 운전상태(Run, Stop, Emergency)를 의미한다. 하기 그림은 풍력 데이터 종류 및 유형은 나타낸다.

```
Data columns (total 45 columns):
 #   Column                                    Non-Null Count   Dtype
---  ------                                    --------------   -----
 0   dateInfo                                  795601 non-null  datetime64[ns]
 1   WDSS_BradePitchAng_SS01                   795601 non-null  float64
 2   WDSS_CurrentL1_SS01                       795601 non-null  int64
 3   WDSS_CurrentL2_SS01                       795601 non-null  int64
 4   WDSS_CurrentL3_SS01                       795601 non-null  int64
 5   WDSS_GearOilTemper_SS01                   795601 non-null  int64
 6   WDSS_GenAccured_SS01                      795601 non-null  int64
 7   WDSS_GeneratorDE_SS01                     795601 non-null  object
 8   WDSS_GeneratorNDE_SS01                    795601 non-null  object
 9   WDSS_GenOutput_SS01                       795601 non-null  float64
 10  WDSS_GenSpeed_SS01                        795601 non-null  float64
 11  WDSS_GenTemper_SS01                       795601 non-null  int64
 12  WDSS_GenWaterCool_SS01                    795601 non-null  int64
 13  WDSS_HighSpeedStageFront_SS01             795601 non-null  object
 14  WDSS_HighSpeedStageRear_SS01              795601 non-null  object
 15  WDSS_HydroOilTemper_SS01                  795601 non-null  int64
 16  WDSS_IntermediateStageFront_SS01          795601 non-null  object
 17  WDSS_IntermediateStageRear_SS01           795601 non-null  object
 18  WDSS_NacelleAng_SS01                      795601 non-null  float64
 19  WDSS_PitchPress_SS01                      795601 non-null  float64
 20  WDSS_PlanetaryStage_SS01                  795601 non-null  object
 21  WDSS_RotorSpeed_SS01                      795601 non-null  float64
 22  WDSS_Spot1Temp_SS01                       795601 non-null  float64
 23  WDSS_Spot2Temp_SS01                       795601 non-null  int64
 24  WDSS_StatusOfDioxideExtinguishingSystem1_SS01  795601 non-null  object
 25  WDSS_StatusOfDioxideExtinguishingSystem2_SS01  795601 non-null  object
 26  WDSS_StatusOfDioxideExtinguishingSystem3_SS01  795601 non-null  object
 27  WDSS_StatusOfDioxideExtinguishingSystem4_SS01  795601 non-null  object
 28  WDSS_TemperL1_SS01                        795601 non-null  int64
 29  WDSS_TemperL2_SS01                        795601 non-null  int64
 30  WDSS_TemperL3_SS01                        795601 non-null  int64
 31  WDSS_Total_GenAccured_SS                  795601 non-null  int64
 32  WDSS_Total_GenOutput_SS                   795601 non-null  float64
 33  WDSS_TowerWithNacelle_SS01                795601 non-null  object
 34  WDSS_TransTemperBus_SS01                  795601 non-null  int64
 35  WDSS_TransTemperL1_SS01                   795601 non-null  int64
 36  WDSS_TransTemperL2_SS01                   795601 non-null  int64
 37  WDSS_TransTemperL3_SS01                   795601 non-null  int64
 38  WDSS_VoltageL1_SS01                       795601 non-null  int64
 39  WDSS_VoltageL2_SS01                       795601 non-null  int64
 40  WDSS_VoltageL3_SS01                       795601 non-null  int64
 41  WDSS_WindAng_SS                           795601 non-null  float64
 42  WDSS_WindSpeed_SS                         795601 non-null  float64
 43  WDSS_WindStatus_SS01_Run                  795601 non-null  uint8
 44  WDSS_WindStatus_SS01_Stop                 795601 non-null  uint8
```

그림 11. 풍력발전 데이터 형태
Fig. 11. Wind power generator data types

풍력발전 데이터는 44개 필드로 구성되며, 각 필드는 795,601 데이터를 가지고 있다. 풍력발전 Tag 중 12개는 미사용 데이터이며, Pt_Created 값으로 되어 있고 아래의 회색 음영부분으로 표기되었다. 43, 44번째 필드는 WindStatus을 나타내며, Run or Stop 값을 갖는 범주형 데이터로 녹색 음영 부분이다.

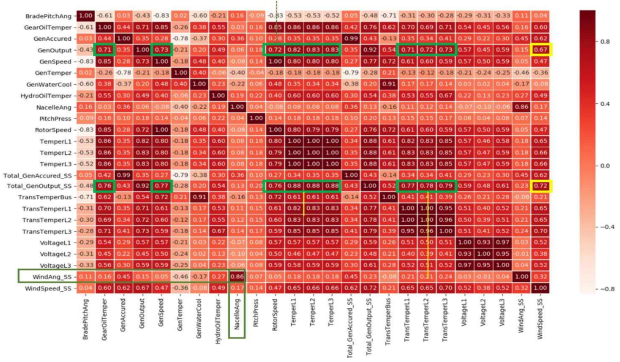

그림 12. 상관관계 분석
Fig. 12. Attribute variable correlation

특성 데이터 간의 상관관계 분석을 통해 단면(cross-sectional) 데이터 측면에서 상관관계가 높은 데이터 쌍을 가려내고, 상관관계의 우선순위를 정하여 신규 데이터가 유입될 경우 상관관계의 상대적 순위가 유지되는지 변화는지 여부에 따라 이상/정상 여부를 판별하였고, 그림 12와 같이 데이터 필드간의 상관관계 분석을 수행하였다.

예측모델 성능은 그림 13와 같이 MAPE: 6.4%, RMSE: 5.4kW로 매우 높은 정확성을 보이며, 구간별 발전오차는 11kW 미만에서 98.99%의 비율을 차지하고 있다. 이는 발전소 설비용량의 10%에 해당하는 수치로 매우 작다고 볼 수 있다.

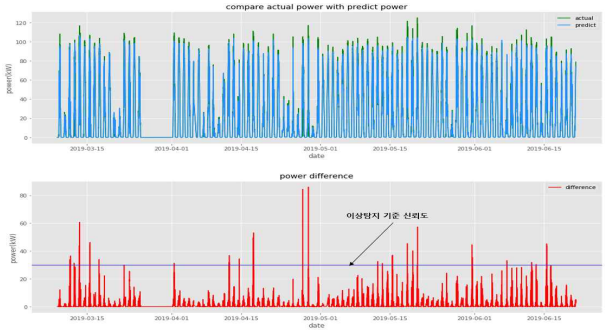

그림 13. 모델 예측 결과 및 발전 오차 비교
Fig. 13. Comparison of model prediction results and power generation errors

그림 14는 이상탐지 기준 신뢰도 초과 발전전력 추이를 보여주며, 모델이 제시한 범주내에서 정상 작동하는 것을 확인할 수 있다.

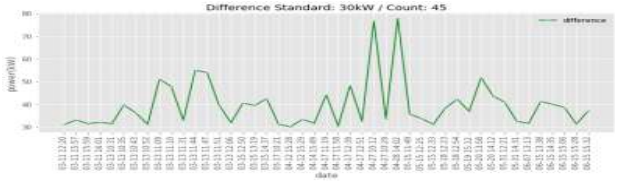

그림 14. 이상탐지 기준 신뢰도 초과 발전전력 추이
Fig. 14. Trend of generated power that exceeds the reliability of the anomaly detection standard

발전출력과 직접적인 관련성을 갖는 발전기 정보(전류, 전압, 온도) 제외하고 발전출력에 대한 피어슨 상관분석 결과 풍속은 0.90, 로터 속도는 0.83으로 매우 강한 양의 상관계수를 갖는다. 그림 15는 발전출력에 대한 피어슨 상관분석 결과를 보여준다. 이는 풍속이 증가할수록 발전출력이 증가함을 의미하며, 로더속도 역시 터빈의 블레이드 회전에 의해 증가한다. 단, 풍속은 실시간으로 변화하게 되지만 블레이드의 회전속도는 기계적 관성이 있기 때문에 풍속과 같이 실시간으로 변하는 것이 불가능하기에 과거 풍속의 변화에 따른 출력의 관계를 잘 파악하는 것이 풍력발전기 출력 예측에 중요한 부분이다.

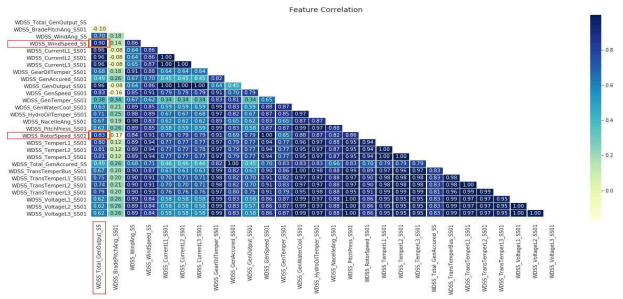

그림 15. 피어슨 상관계수
Fig. 15. Pearson's correlation coefficient

그림 16의 측정 결과를 통해 앙상블 모델 중에서는 RandomForest 알고리즘이 가장 높은 수치를 보였으며, GradientBoosting 알고리즘이 가장 낮은 결과를 보여주었다. 부스팅 모델은 약한 학습기를 순차적으로 학습을 하되, 이전 학습에 대하여 잘못 예측된 데이터에 가중치를 부여하여 오차를 보완해 나가는 방식이지만, 제한적인 과거 데이터 수로 인해 지속적인 오차 갱신이 이루어지지 않은 결과로 볼 수 있다.

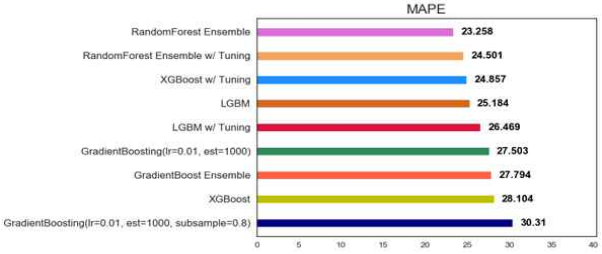

그림 16. MAPE 결과
Fig. 16. MAPE result

모델 서빙은 학습한 모델을 저장하고 내보내어 실시간으로 신규 데이터에 대한 예측 결과를 전달하는 시스템으로 학습 모델에서 예측 결과를 가져오고 Flask API를 통해 최종 사용자에게 해당 결과를 전달하며, 그림 17은 모델 서빙을 통한 예측 결과를 보여준다. 예측치와 실측치의 오차 검증을 통해 모델의 성능이 의미 있음을 알 수 있다.

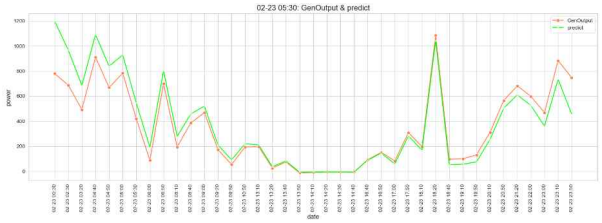

그림 17. 모델 서빙을 통한 예측 결과
Fig. 17. Predicted results through model serving

V. 결 론

풍력재생에너지를 대상으로 사이버 보안 측면에서 국내 풍력설비 운전기준에 대한 권고 사항이나 가이드라인의 새로운 기준으로 제시하였다. 풍력을 포함한 재생에너지 설비 부문에서 사이버보안과 관련한 운전기준을 수립하는 것은 현재 시점에서는 쉽지 않은데, 기존의 기력발전기와는 달리 스케일이나 운영 환경이 제각각이고 아직까지 충분한 데이터가 축적되지 않았기 때문이다.

미국 국립표준기술연구소(NIST: National. Institute of Standards and Technology)에서 전력시스템의 사이버보안 프레임워크에 대한 백서를 발간하였으나 개념적 차원이고, 북미전력안정성회사(NERC: North American Electric Reliability Council)에서도 전력시스템에 대한 사이버보안 측면을 강조하고 있기는 하나 개별 발전사별로 주기적인 위협 분석과 그에 대한 대응 조치 수립보고서를 제출 권고하는 형태로 체계 수립를 수립한 단계이다.

본 연구에서는 이러한 환경에서 가이드라인을 수립하고, 물리 데이터와 사이버 데이터를 혼용하여 위험 수준의 이상이 생길 경우, 이를 탐지하고 운전 상의 이상을 사전에 방지하기 위한 목적으로 접근하여 실제 운영 데이터를 확보하여 데이터 공유와 현장 시스템 환경 공유를 통해 실증적인 결과를 토대로 그림 17과 같이 풍력발전 이종 영역 데이터를 활용한 이상탐지 모델의 방향성에 대해 검증을 하였다.

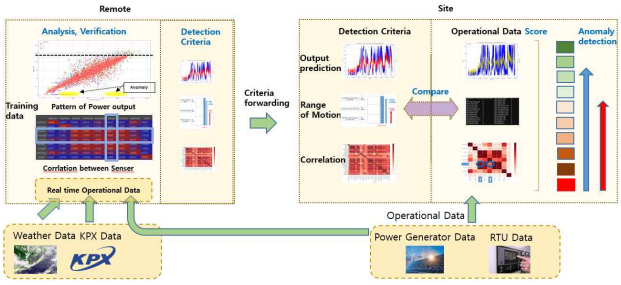

그림 18. 이상 탐지 작동 구조
Fig. 18. Anomaly detection mechanism

모델 검증을 통해 현장 적용이 가능한 방향으로 연구 방향성을 정립하였고, 현재 연구단계에서는 주어진 물리 데이터(풍력 발전량 및 기계전기적 부대 데이터)와 가상의 사이버 데이터를 연계하여 CPS 환경에서 이상탐지를 할 수 있는 프로토타입을 통해 향후 전형적인 마이크로그리드 환경인 신재생에너지 전분야에서 확산이 가능한 모델의 고도화 과정을 지속할 필요가 있다.

신재생에너지 전분야의 표준 모델로서 이상탐지를 위한 마이크로그리드 융합보안 솔루션으로 확대 적용하고, 향후 가상발전소(VPP) 구축 과정에서의 데이터 무결성 확보를 위한 필수 솔루션으로 확산하기 위한 풍량 신재생에너지의 마이크로그리드 확장을 위한 표준모델의 연구가 필요하다. 향후 모델의 고도화를 위한 연구 방향을 논하기 위해 학습용 데이터의 샘플링을 확대하여 지역별 특성에 국한된 모델이 아니라 국내 민간 재생에너지 발전망에서의 안전한 시스템 관리를 위한 이상탐지 모델 확장과 함계 융합 보안 측면의 가이드 제정이 필요하다.

참고문헌

[1] D. Kang, J.-K. Seo, and H. Kim, Research on Anomaly Detection System for Renewable Energy System, pp. 2355-2357, July 2020. http://kiee.or.kr,

[2] USALAB. Research Blog [Internet]. http://research.sualab.com/introduction/review/2020/01/30/anomalydetectionoverview-1.html

[3] M. Goldstein and S. Uchida, “A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data,” *PLoS ONE*, Vol. 11, No. 4, e0152173. https://doi.org/10.1371/journal.pone.0152173

[4] J.-S. Back, Research of Anomaly Detection Method Through Content Analysis of SCADA Protocol, Master's Thesis, Korea University, February 2012.

[5] H. Yoo, J.-H. Yum, and T. Shon, “Whitelist-Based Anomaly Detection for Industrial Control System Security,” *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 38B, No. 8, pp. 641-653. http://dx.doi.org/10.7840/kics.2013.38B.8.641

[6] W.-K. Sung, “Power Generate Control System(DCS) Network Packet Analysis and Whitelist Modularization Implementaion,” Master's Thesis, Korea University, August 2017.

[7] M. Oh, “Research of Whitlist Development for Reinforcing Security in Control System Network,” Master's Thesis, Korea University, August 2017.

[8] D. H. Maulud and A. M. Abdulazeez, “A Review on Linear Regression Comprehensive in Machine Learning,” *Journal of Applied Science and Technology Trends*, Vol. 1, No. 4, pp. 140-147, December 2020. https://doi.org/10.38094/jastt

1457
[9] D. Bohning, “Multinomial Logistic Regression Algorithm,” *Annals of the Institute of Statistical Mathematics*, Vol. 44, pp. 197-200, 1992.
[10] R. Xu and D. Wunsch, “Survey of Clustering Algorithms,” IEEE Transaction on Neural Network, Vol. 16, No. 3, pp. 645-678, May 2005. https://doi.org/10.1109/TNN.2005.845141
[11] F. Hohman, K. Wongsuphasawat, M. B. Kery, and K. Patel, Understanding and Visualizing Data Iteration in Machine Learning, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, 25-30 April 2020, pp. 1-13. https://doi.org/10.1145/3313831.3376177
[12] T. G. Dietterich, Ensemble Methods in Machine Learning. in *MCS 2000: Multiple Classifier Systems*, Berlin, Heidelberg: Springer, pp. 1-15, 2000.

유진도(Jin-do Yu)

1999년 : 명지대학교 컴퓨터공학과 학사
2023년 : 고려대학교 대학원 (석사과정-사이버보안학과)

2014년～현 재: 한국남부발전
2017년～현 재: 고려대학교 정보보호대학원 석사과정
※관심분야 : 정보보호(Personal Information), 인공지능, 통신공학 등