



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

바람자원평가 및 경제성 분석을 위한
기계학습 적용 방안

제주대학교 대학원

풍 력 공 학 부

문 경 록

2021年 12月

바람자원평가 및 경제성 분석을 위한
기계학습 적용 방안

指導教授 高 炘 男

文 璟 綠

이 論文을 風力工學部 碩士學位 論文으로 提出함

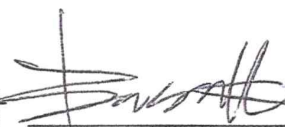
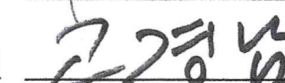




2021년 12월

文璟綠의 風力工學部 碩士學位 論文을 認准함

심사위원장

위 원

위 원

제주대학교 대학원 풍력공학부

2021년 12월

Machine learning application for wind resource and economic viability assessment

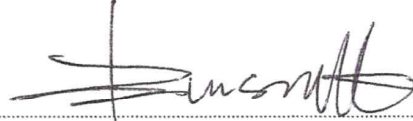
Kyeong-Rok Mun

(Supervised by professor Kyung-Nam Ko)

A thesis submitted in partial fulfillment of the requirement for the
degree of Master of Science

2021. 12.

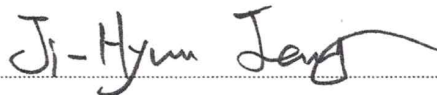
This thesis has been examined and approved.



Thesis director, Bum-Suk Kim, Prof. of Faculty of wind energy engineering



Thesis director, Kyung-Nam Ko, Prof. of Faculty of wind energy engineering



Thesis director, Ji-Hyun Jeong, Prof. Major of Mechanical engineering

.....
Date

Faculty of Wind Energy Engineering
Graduate School
JEJU NATIONAL UNIVERSITY

목 차

List of Figures	ii
List of Tables	iv
Abstract	v
I. 서 론	01
1. 연구배경	01
2. 연구동향	03
3. 연구목적	05
II. 연구방법	06
1. 분석대상 선정	06
2. 연구환경	09
3. 기계학습	09
III. 연구결과	16
1. 데이터 품질 검사(Data quality check)	16
2. 타워 쉐도우 영향 보정(Tower shadow calibration)	26
3. MCP(Measure-correlate-predict)	37
4. 전력계통한계가격(SMP) 장기 추세 분석	46
IV. 결 론	57
V. 참고문헌	59

List of Figures

Fig. 1 The 9th long-term electricity supply and demand plan	01
Fig. 2 Positions of wind resource measurement sites including their coordinate system	06
Fig. 3 Wind sensors and component with measurement period of data	07
Fig. 4 Work flow of machine learning method	09
Fig. 5 Classification method of KNN	12
Fig. 6 Level wise tree method	13
Fig. 7 Leaf wise tree method	14
Fig. 8 Linear classification of SVM	14
Fig. 9 Nonlinear classification of SVM	15
Fig. 10 Work flow of machine learning for data quality check	17
Fig. 11 Wind roses at Sangmyeong met mast	19
Fig. 12 Wind rose at Seogwang AWS	20
Fig. 13 Progress of machine learning for data quality check	21
Fig. 14 Effects of tower shadow in accordance with tower types	26
Fig. 15 Work flow for tower shadow calibration	27
Fig. 16 Azimuth angles of anemometers A and B	28
Fig. 17 Effects of tower shadow	29
Fig. 18 Detection of tower shadow at anemometer A	31
Fig. 19 Detection of tower shadow at anemometer B	32
Fig. 20 Azimuth angle for tower shadow calibration	33
Fig. 21 Progress for tower shadow effect calibration	34
Fig. 22 Graphical comparison with before and after tower shadow calibration ..	35
Fig. 23 Work flow for machine learning for MCP application	38
Fig. 24 Result of correlation analysis	40

Fig. 25 Progress of machine learning for MCP	41
Fig. 26 Work flow for forecast of long term trend of SMP	46
Fig. 27 SMP variation with yearly before and after data filtering	47
Fig. 28 Number of yearly SMP deciding	48
Fig. 29 Correlation analysis of factors for SMP deciding	49
Fig. 30 Progress of machine learning	50
Fig. 31 SMP forecasting for year 2019	51
Fig. 32 Linear regression analysis of world crude oil and WTI	52
Fig. 33 Adjustment of lead-lag correlation between Korea and Japan LNG prices	54
Fig. 34 Yearly main land SMP Forecast	55

List of Table

Table 1 Met sensor and measurement period of Sangmyeong onshore met mast	08
Table 2 Information on ERA-5 data	09
Table 3 Heights of wind sensors and measurement periods at M#1 and M#2	17
Table 4 data quality check result for wind speeds at Sangmyeong met mast	18
Table 5 data quality check result for wind directions at Sangmyeong met mast	20
Table 6 Train and test data set for machine learning	21
Table 7 Train accuracy of every machine leaning algorithm	22
Table 8 Comparison of manual and machine learning results for wind speed data quality check	23
Table 9 Height of wind sensors and measurement period of Daejeong offshore met mast	28
Table 10 Number of data points for machine learning	33
Table 11 Train accuracy of every machine leaning algorithm	34
Table 12 Comparison of before and after Tower shadow calibration using RMSE and R^2	36
Table 13 Information wind data of the offshore met mast and ERA-5 of collected data set	38
Table 14 Quality of Reference data	39
Table 15 Train and test data set periods for MCP	41
Table 16 Self prediction result after MCP application	42
Table 17 Comparison between machine learning and WindPRO wind speed prediction	43
Table 18 Correlation coefficients of train and test data at every ERA-5 data p	

ositions	43
Table 19 Frequency and period of collected data set	49
Table 20 Train and test periods for forecasting of SMP	50
Table 21 Train result of forecasting SMP model	52

Abstract

In order to apply machine learning method to wind resource assessment and economic viability analysis, an investigation was carried out in terms of wind data quality check, met tower shadow correction, Measure-Correlate-Predict (MCP) method and system marginal price (SMP). K-Nearest Neighbor (KNN), Light Gradient Boosting Machine (LGBM) and Support Vector Machine (SVM) were used as algorithm for supervised learning. The accuracy was evaluated using root mean square error (RMSE) and the coefficient of determination (R^2) for numerical data, while using the accuracy ratio of the number of data points classified correctly to the total number of data points for categorical data.

As for wind data quality check, machine learning was performed using Sangmyeong onshore met mast data and LGBM was found to be the best among the three algorithms for data quality check. The LGBM was applied to Daejeong offshore met mast wind data for the quality check, which led to the accuracy ratio of 93.51%.

The wind flow distortion due to tower shadow effect was detected using statistical method for the offshore wind data. Then the normal data points within an azimuth angle of no-shadow effect were used as train data points, which resulted that the best algorithm was SVM among the three algorithms. The SVM algorithm was applied to the data points distorted by tower shadow at 82.5 m height. When the data points without shadow effect were a reference, the RMSE decreased from 1.009 m/s to 0.2611 m/s and R^2 increased from 0.8910 to 0.9926. That is, the tower shadow effect could be corrected using machine learning technique.

For long-term wind data correction, MCP method has been utilized, which was adopted in WindPRO software. The reanalysis wind data at four positions

around the offshore met mast were collected to use them as reference data for MCP application. The above three algorithms at Machine learning method were applied to predict the long-term wind data from Oct. 1st 2013 to Jan. 31st 2015, after being applied to train using the wind data from Jul. 1st 2015 to Jun. 30th 2016. It was found that the LGBM had the highest accuracy among the three algorithms, while the neural network did the highest accuracy among the three algorithms of WindPRO software. The predictions from the LGBM were compared with those from the other three algorithms of WindPRO that were the regression, the matrix and the neural network. The result was that the LGBM had the RMSE of 0.12 m/s and the R^2 of 0.99, while the RMSE and the R^2 were 0.20 m/s and 0.98, respectively, at the neural network algorithm. In other words, the machine learning could improve MCP prediction accuracy more than the traditional MCP algorithms at WindPRO.

For the purpose of estimating long-term SMP from 2020 to 2030 in the mainland of South Korea, the machine learning technique were also applied in this study. Liquefied Natural Gas (LNG), West Texas Intermediate Crude Oil (WTI) and FOB Kalimantan price data were collected from public data portal sites since those were considered main factors deciding the mainland SMP. As a result of correlation analysis, LNG and WTI were determined as the factors being used for the machine learning. The most accurate algorithm was LGBM, which was employed for forecasting the long-term SMPs. Using Japanese LNG price estimates and world crude oil price estimates, the long-term SMPs were estimated, which had a decreasing trend from 2020 to 2022 and then maintain 72 KRW/kWh until 2030.

I. 서 론

1. 연구배경

인류는 급격한 경제 성장에 따른 많은 탄소배출로 인해 지구 온난화가 가속화 되어 산불, 태풍, 홍수 등의 다양한 이상 기후 현상이 세계 곳곳에서 문제 되고 있다. 이러한 문제를 해결하기 위해 기후변화에 관한 정부 간 협의체(Intergovernmental panel on climate change, IPCC)에서는 지구의 평균 온도 상승폭을 2100 년까지 1.5°이내로 제한하기 위해 세계적으로 이산화탄소 배출량을 2030년까지 최소 45%이상 감축하고 2050년에는 탄소 중립(Net-zero)을 달성해야 한다는 목표를 제시하였다 [1]. 이에 따라 우리나라 정부에서는 ‘2050 탄소 중립 전략’을 수립하였고 이를 실천하기 위한 7가지 부문별 비전 및 전략을 발표하였다 [2]. 2017 년 기준 우리나라의 국내 온실가스 전체 배출량 중 약 36%를 에너지 공급 부문에서 발생하였고 이를 줄이는 것은 2050 탄소 중립 전략을 달성하는데 핵심적이다. 이를 달성하기 위해서는 해결하기 위해 석탄 발전을 단계적으로 액화천연가스(Liquefied natural gas, LNG) 발전으로 전환하거나 폐쇄하고 태양광, 풍력 등의 친환경 발전을 통해 생산된 에너지 활용해야 한다. 이에 따라 기존의 ‘신재생에너지 3020 계획’에 ‘2050 탄소 중립 전략’을 반영하여 산업통상자원부에서는 9 차 전력수급기본계획을 발표하였다. Fig. 1은 9차 전력수급기본계획에 따른 발전 설비용량 비중의 전망을 보여 준다 [3].

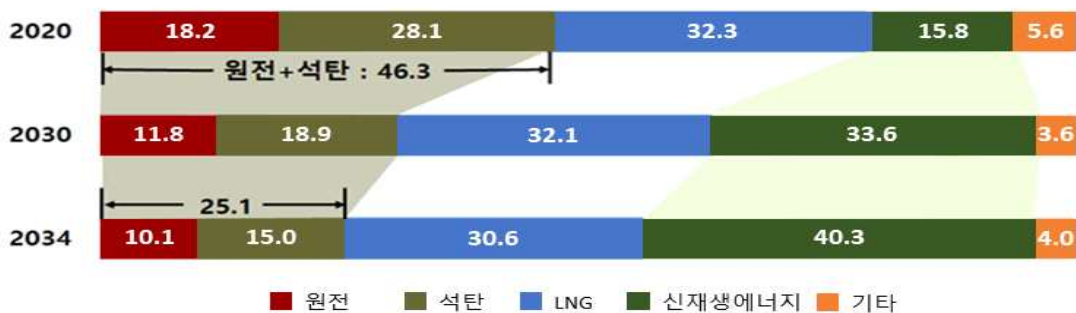


Fig. 1 The 9th long-term electricity supply and demand plan [3]

2020년 현재 태양광, 풍력 등의 신재생 에너지의 설비용량은 20.1GW이고 전체 발전 설비의 15.8%를 차지하고 있다. 산업통상자원부에서는 2034년까지 신재생 에너지 설비용량 57.6GW를 증설하여 전체 발전 설비용량의 40.3%까지 확대하는 계획을 세우고 있어 향후 대규모 풍력발전단지 건설은 필연적이다. 이에 따른 풍력발전단지에 대한 사전 바람자원평가 및 경제성 분석의 중요성은 커질 것이다. 바람자원평가 및 경제성 분석에 대한 다양한 경험, 전문 지식 등에 따라 분석의 결과가 크게 달라질 수 있으며 비전문가는 다양한 분석 수행에 어려움을 겪을 수 있다. 최근 차세대 산업 혁명인 4차 혁명시대의 도래로 빅데이터(Big data)와 인공지능(Artificial intelligence) 등의 기술이 각광 받고 있다 [4]. 사회의 많은 분야에 적용·활용되고 있으며 다양한 경험, 전문 지식을 컴퓨터가 데이터를 기반으로 학습하여 인간에게 많은 도움을 주고 있다. 따라서 풍력발전단지 바람자원평가 및 경제성 분석을 수행함에 있어 인공지능을 적용하는 연구가 필요하다.

2. 연구동향

전 세계적으로 풍력 산업 전반에 인공지능의 한 분야인 기계학습(Machine learning)을 적용하는 연구가 다양하게 수행되어왔다.

국내에서는 기계학습을 활용하여 발전량 예측, 블레이드 고장 유무 확인, 풍력 발전기의 하중을 예측하는 등의 연구가 수행되었다. 허진 등[5]은 기계학습의 모델인 SVM(Support vector machine)을 기반으로 풍력발전단지의 단기 발전량 예측을 수행하였다. 풍력발전단지의 발전량과 풍속의 과거 1주일 데이터를 활용하여 기계학습을 통해 앞으로의 24시간에 대한 풍력발전단지의 출력을 예측하였다. 박수현 등[6]은 블레이드의 고장 진단에 기계학습과 드론을 적용한 연구를 수행하였다. 블레이드의 파단 길이에 따른 검사용 드론 기체 내부에 전달되는 진동 데이터를 수집하고 다양한 모델을 활용하여 기계학습을 수행하였다. 기계학습 모델을 활용하여 블레이드의 고장 유무를 판단하였고 95%의 정확도로 고장 유무 분류하였다. 최정철 등[7]은 풍력발전기의 하중을 예측에 기계학습 방법을 적용하는 연구를 수행하였다. 기계하중과 계측신호 데이터 간의 상관관계를 분석하였고 타워 상단부 앞·뒤 방향의 모멘트의 증가하중 값이 기계하중과의 상관관계가 높음을 확인하였다. 이를 기계학습의 인공신경망(Artificial neural network)의 입력 변수(Input data)로 활용하여 풍력발전기의 하중을 예측하였다. 이들은 풍력발전기의 피로하중 계산에 사용되는 계측신호 데이터 중 상관성 높은 일부의 계측신호만을 활용하여 정확도가 높은 하중 예측하는 방법을 제안하였다.

해외에서는 기계학습을 활용한 풍속 예측, 블레이드 고장 검사, 기상 예측을 통한 풍력 발전기의 손실 예측, MCP(Measure-correlate-predict) 등의 연구가 수행되었다. Halil demolli 등은[8] 일일 풍속 데이터와 기계학습 알고리즘을 기반으로 풍력발전단지의 장기간 발전량을 예측하는 연구를 수행하였다. 터키의 Nigde-nigde지역에서 측정된 풍속 데이터를 활용하여 기계학습을 수행하고 생성된 모델을 활용하여 터키의 다른 4개의 지역에 대해 풍력발전기의 1년간 발전량을 예측하고 이를 실제 발전량을 비교한 결과 유사함을 확인하였다. Alfredo arcos jiménez 등[9]은 블레이드 유지보수에 기계학습을 적용하는 연구를 수행하였다. 블레이드의 적층되어 있는 섬유층의 층간 분리의 전·후 소리 파장을 수집하여 기계학습을 수행

하였고, 학습된 모델을 실제 섬유의 층간 분리가 발생한 블레이드에 적용하여 고장 유무 검출을 수행한 결과 정확도가 90%이상의 분류 정확도를 보였으며, 이를 바탕으로 블레이드 유지보수에 기계학습의 적용 방법을 제안하였다. M.Majidi nezhad 등[10]은 사르데냐 제도 주변에 대한 재해석 데이터와 기계학습을 활용하여 바람자원의 잠재력을 평가하였다. 40년간의 재해석 데이터를 수집하고 이를 기반으로 풍속 예측을 위한 기계학습을 수행하였다. 10개의 풍력발전단지에 대해 예측을 수행하고 GIS 프로그램을 활용하여 사르데냐 제도 주변에 대한 바람자원의 잠재력을 지도에 표현하였다. Michael negnevitsky 등[11]은 전력 시스템의 단기 부하, 전력판매가격, 풍력발전단지의 단기 발전량 예측에 기계학습 방법을 적용하는 연구를 수행하였고 전력 시스템에 대해 다양한 접근법 방법을 통한 기계학습의 적용 방안을 제안하였다. Sebastian scher 등[12]은 기계학습을 기반으로 결빙률에 따른 풍력발전기의 손실을 예측하는 연구를 수행 하였다. 기계학습에 29주의 기상 예보 데이터를 활용하였고 이를 통해 42시간 이후의 풍력발전기의 결빙 예측을 수행하였다. Velazquez 등[13]은 카나리아 제도의 풍력발전단지 후보지의 풍력발전기에 대한 발전량 예측에 활용되는 MCP수행에 인공신경망과 선형회귀 방법을 비교하는 연구를 수행하였다. 이들은 적절하게 인공신경망을 설계하고 참조 데이터와의 상관관계가 높을 경우 기존의 선형회귀 방법을 포함한 일반적으로 사용되는 MCP방법보다 예측 성능이 높다는 것을 확인하였다.

3. 연구목적

풍력발전단지 개발 시 기상탑을 설치하고 1년 이상의 기간 동안 바람 자원을 측정하고 취득한 데이터를 기반으로 바람자원평가를 수행한다. 바람자원분석은 취득 데이터의 정확성뿐만 아니라 분석자의 경험과 지식에 기반하여 이루어지기 때문에 일관성이 떨어진다. 객관적이고 일관성 있는 분석 수행에 인공지능 적용이 도움이 될 수 있으며 이는 비전문가가 분석을 수행함에도 도움이 될 수 있다. 최근 인공지능이 발달함에 따라서 풍력산업 분야에도 인공지능이 적용된 다양한 연구가 수행되고 있으나 바람자원평가에 적용한 사례는 많이 찾아볼 수 없다. 또한 경제성 분석시 CAPEX(Capital expenditures), OPEX(Operating expenditure), 전력계통한계가격(System marginal prices, SMP), 신재생에너지 공급 인증서(Renewable energy certificate, REC) 등의 장기 추정은 매우 중요하다. SMP의 경우 발전사업자의 수익에 직접적인 영향을 주는 변수임에도 이전 수행된 연구의 대부분이 1년 이하의 단기 예측[15,16,17,18]에 집중되어 있다. 풍력발전기의 수명이 20년 이상임을 감안하면 단기 예측된 SMP는 경제성 분석에 사용함에 있어 부족함이 있다. 본 논문은 바람자원평가와 경제성 분석에 필요한 데이터 품질 검사, 타워쉐도우(Tower shadow) 영향 보정, MCP, SMP의 장기 추세 분석에 인공지능의 기계학습을 적용하는 연구를 수행하였으며 연구 방법은 다음과 같다. 1) 실제 측정된 데이터에 따른 데이터 품질 검사를 수행하여 바람자원평가 결과의 정확성 저하시킬 수 있는 요소를 기계학습을 통해 검출해낸다. 2) 윈드쉬어(Wind shear)를 활용한 풍력발전기 허브 높이의 바람 추정에 정확성을 저하시킬 수 있는 타워쉐도우 구간을 통계적 방법을 통해 검출하고 기계학습 방법을 활용하여 보정을 수행한다. 3) 풍력발전단지의 장기간 바람 보정에 기계학습을 활용하여 장기간 바람데이터를 생성한다. 4) 다양한 변수 분석의 결과를 활용하여 기계학습 방법을 기반으로 장기간 SMP 추세 예측을 수행한다.

II 장에서는 연구에 활용된 기상탑 및 참조 데이터의 정보와 기계학습 모델의 종류와 특성에 대해 소개한다. III 장에서는 기계학습을 활용하여 데이터 품질 검사, 타워쉐도우 보정, MCP 분석과 경제성 분석에 필요한 SMP의 장기 추세를 분석한다. IV 장에서는 III 장에서의 연구 결과를 최종 정리한다.

II. 연구 방법

1. 분석 대상 선정

본 논문에서는 대정해상 기상탑(M#1) 데이터를 중심으로 바람자원평가 및 기계 학습 방법을 적용하였다. 또한 상명목장지구 기상탑(M#2) 데이터와 재해석 데이터인 ERA-5(E#1,2,3,4) 4개의 데이터를 활용하였다. Fig. 2는 본 연구를 위해 활용된 데이터의 위치를 나타낸다.

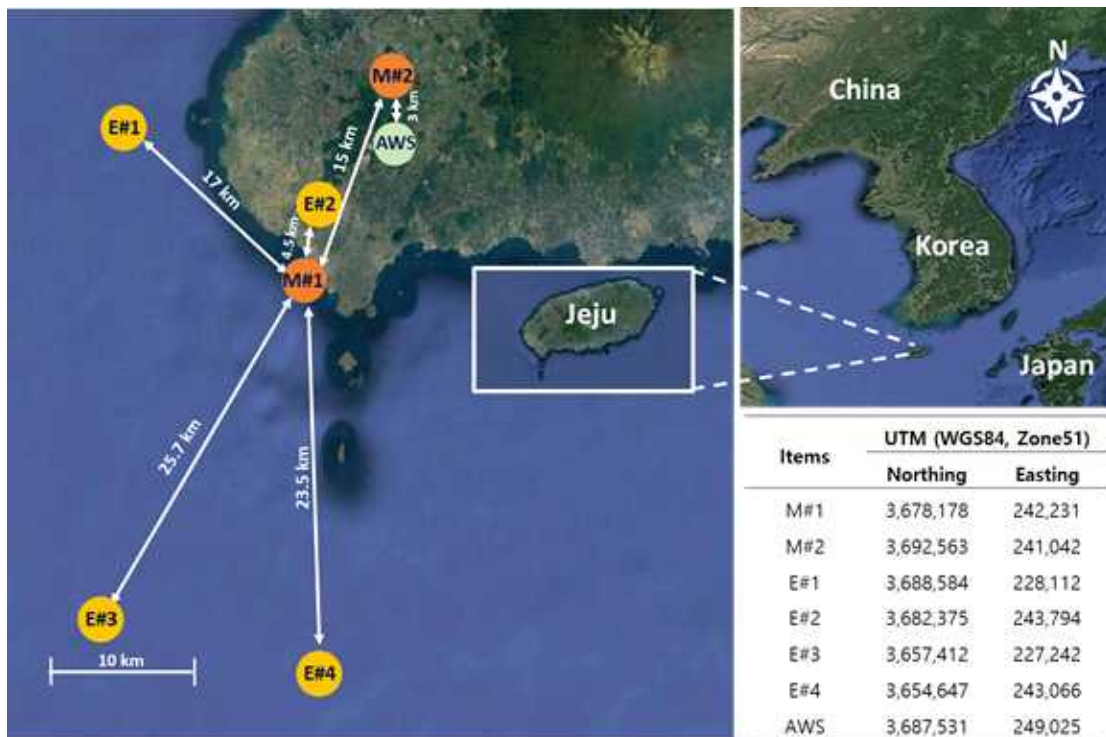


Fig. 2 Positions of wind resource measurement sites including their coordinate system

대정해상 기상탑을 기준으로 상명목장지구 기상탑은 북동쪽으로 약 15km 떨어져 있으며 ERA-5 데이터 중 대정해상 기상탑과 가장 가까운 위치는 E#2로 북쪽으로 4.5km에 위치해 있으며 E#3은 남서방향으로 25.7km로 대정해상 기상탑으로부터 가장 멀리 위치해 있다.

1.1 대정해상 기상탑

Fig. 3은 대정해상 기상탑의 측정 장비 구성요소 및 바람자원을 측정 한 기간을 보여준다. 대정해상 기상탑은 평균 해수면 높이(Above mean sea level: AMSL)로부터 111.5m의 높이로 해상에 설치되어 2013년 8월 12일부터 2016년 10월 28일까지 10분 평균 데이터의 바람자원 데이터를 수집하였다. 평균 해수면 높이는 12.5m이다.

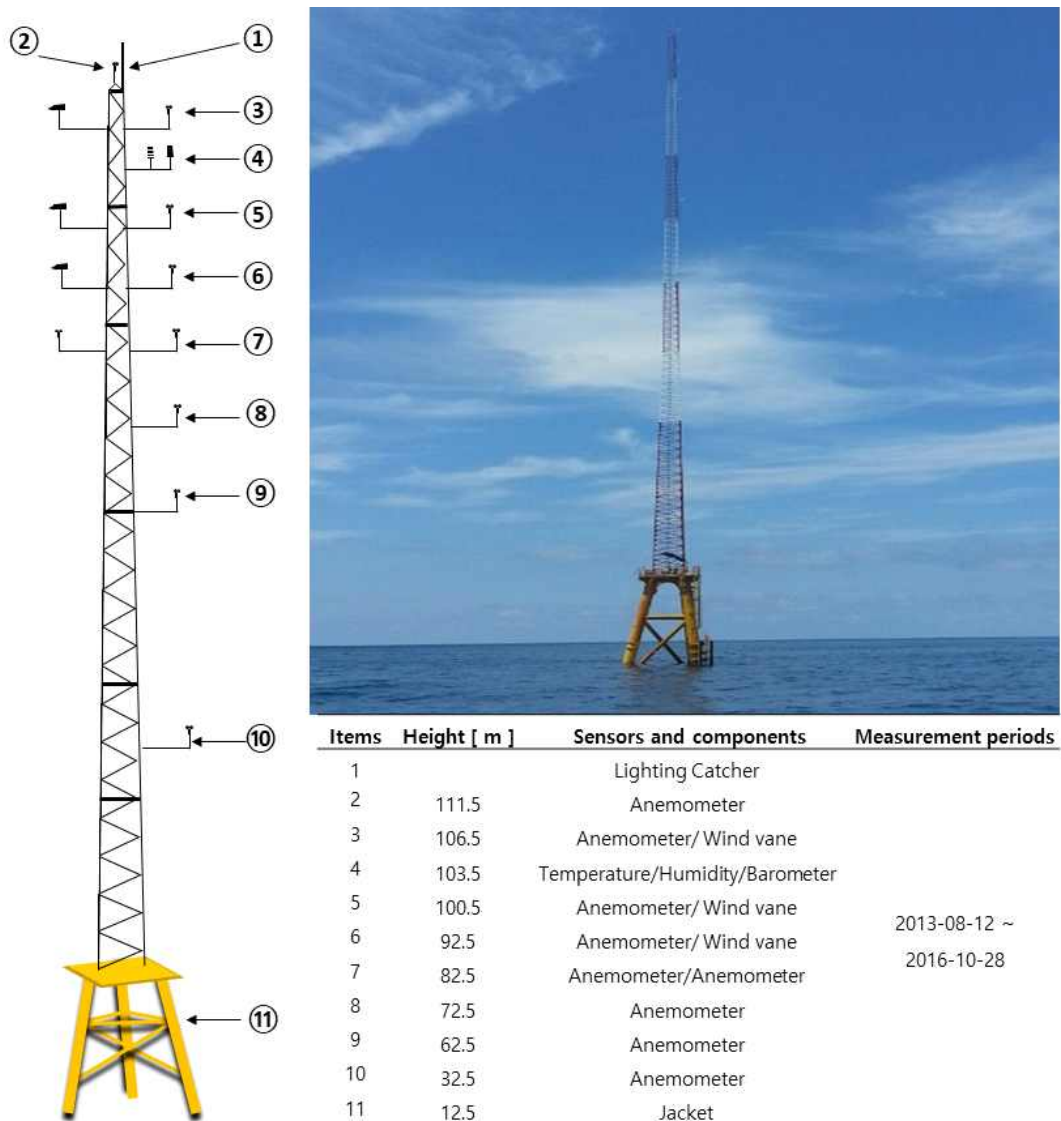


Fig. 3 Wind sensors and component with measurement period of data

본 연구를 위해 111.5m, 100.5m, 82.5m의 풍속계와 106.5m의 풍향계가 활용되었다. 또한 82.5m에는 두 개의 풍속계가 설치되어 있는데 이를 구분하기 위해 A와 B로 분류하였다.

1.2 상명목장지구 기상탑

Table 1은 상명목장지구 기상탑의 높이별 측정 장비의 구성요소 및 측정 기간을 나타낸다. 상명목장지구 기상탑은 40m 높이로 육상에 설치되어 2006년 1월 1일부터 12월 31일까지 1년간 바람자원을 측정하였다. 본 연구를 위해 20m, 30m, 38m, 40m 풍속계와 30m, 40m 풍향계를 활용하였다.

Table 1 Met sensor and measurement period of Sangmyeong onshore met mast

Heights [m]	Met sensor	Measurement periods
40	Anemometer / Wind vane	
38	Anemometer	2006.01.01. ~
30	Anemometer / Wind vane	2006.12.31.
20	Anemometer / Temperature	

1.3 ERA-5(ECMWF Reanalysis v5)

Table 2는 대정해상 기상탑을 주변에서 수집된 ERA-5의 기간 및 높이를 나타낸다. ERA-5의 4개 지점에서 2000년 1월 1일부터 2019년 12월 31일까지의 데이터를 수집하였으며 본 연구에서는 100m 높이의 풍속과 풍향을 활용하였다.

Table 2 Information on ERA-5 data

Items	Heights [m]	Data components	Data periods
E#1			
E#2	100, 10	Wind speed,	2000.01.01. ~
E#3		Wind direction	2019.12.31.
E#4			

2. 연구 환경

본 연구를 위해 사용된 실험 환경은 Intel(R) Xeon(R) CPU E5-2630 v2 @ 2.6 0GHz, 64GB RAM, Window 10인 PC를 활용하였으며 분석 프로그램으로는 풍력 발전단지 설계 프로그램인 WindPRO 3.4와 Pyhton 3.8.3.을 활용하였다.

3. 기계학습

3.1 기계학습 데이터 셋

Fig. 4는 본 연구에서 사용된 기계학습 방법의 순서도를 나타낸다. 전체 데이터는 기계학습을 위해 크게 두 가지 데이터 셋(Data set)으로 나누는데 학습에 참여하는 데이터인 학습 데이터 셋(Train data set)과 시험 데이터 셋(Test data set)으로 분류하여 기계학습에 활용한다.

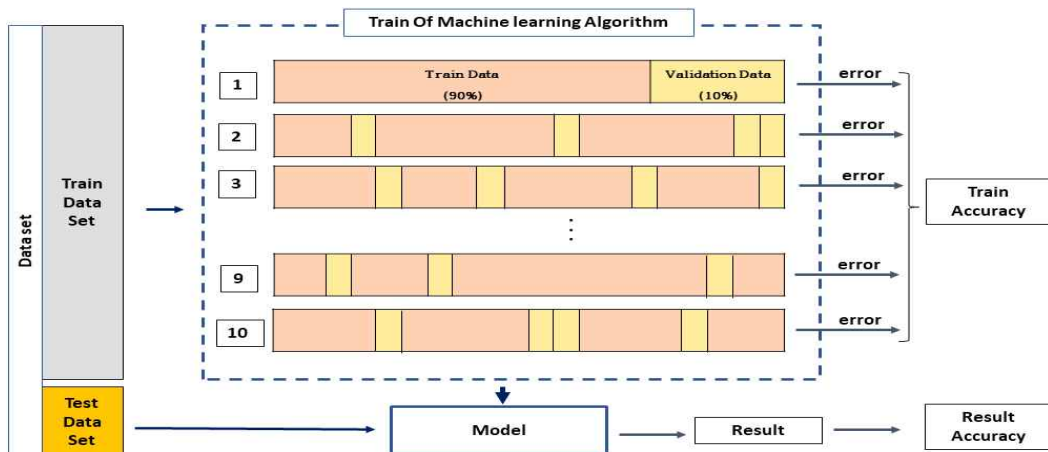


Fig. 4 Work flow of machine learning method

기계학습 모델의 데이터 학습 과정에 학습 데이터 셋을 투입하며 이 과정에서 매개변수(Parameter) 최적화 및 적합(Fitting)의 정도인 학습 정확도를 확인하기 위해 학습 데이터 셋의 일부를 추출하여 검증 데이터(Validation data)로 활용한다. 검증 데이터 추출에는 다양한 방법이 있는데 랜덤 추출 방법은 적은 데이터로 다양한 상황에 따른 학습을 수행할 수 있다는 장점이 있다. 본 논문에서는 학습 데이터 셋의 10%를 검증 데이터로 랜덤으로 추출하여 교차검증(Cross validation)을 통해 데이터 학습하여 모델(Model)을 생성하였다. 이러한 과정으로 생성된 모델에 시험 데이터 셋을 투입하여 모델의 예측 정확도를 확인하였다.

3.2 기계학습 정확도 평가

기계학습을 통해 산출된 예측, 분류 등의 결과에 대한 정확도를 평가 방법에는 평균제곱근오차(Root mean square error, RMSE), 결정계수(R^2), 분류 정확도 점수 등의 다양한 방법이 있다. 수치형 데이터(Numerical data)는 식(1)의 RMSE와 식(2)의 R^2 을 활용하였다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (1)$$

$$R^2 = 1 - \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \quad (2)$$

여기서 y_i 는 실제 데이터의 i 번째 값, \hat{y}_i 는 기계학습에 따른 산출된 i 번째 추정 값, \bar{y} 는 실제 데이터의 평균이다.

수치형 데이터를 활용한 기계학습 모델 평가시 RMSE를 통해 산출된 값이 0에 근사할수록 높은 정확도를 갖는다. 또한 R^2 은 0부터 1사이 값을 갖는데 0에 근사할수록 정확도가 낮다가 평가할 수 있으며 1에 근사할수록 정확도가 높다고 평가할 수 있다. 범주형 데이터(Categorical data)는 (3)의 분류 정확도(Accuracy classification score)를 활용하여 평가하였다.

$$Accuracy = \frac{Number\ of\ data\ classified\ correctly}{Number\ of\ data} \quad (3)$$

여기서 Number of data는 전체 데이터의 개수, Number of data classified correctly는 정확히 분류한 개수이다.

분류 정확도는 0~1의 수치를 갖는다. 0에 가까울수록 분류 정확도가 낮고 1에 가까울수록 모델의 정확도가 높음을 의미한다.

3.3 기계학습 모델

기계학습은 지도 학습(Supervised learning)과 비지도 학습(Unsupervised learning), 강화학습(Reinforcement learning) 세 가지로 분류할 수 있다. [8]. 지도 학습은 분석자가 각각의 데이터에 태그 또는 해답을 통해 학습하는 방법이다. 지도 학습의 종류는 데이터에 따라 범주형 데이터는 분류(Classification)와 수치형 데이터는 회귀(Regression)로 나뉜다. 비지도 학습은 지도 학습과 달리 태그 또는 해답을 통한 학습이 아닌 특정 데이터의 패턴과 특징(Feature)을 기반으로 데이터를 군집을 나누는 기법이다. 강화학습은 특정 정해진 환경 또는 규칙 속에서 현재의 상태를 인식하고 행동을 통해 보상을 최대화하는 학습 방식이다. 본 연구에서는 데이터 품질 검사에 따른 측정 오류 데이터 구간의 분류에 따른 범주형 데이터 분석과 타워체도우 영향을 받은 풍속 데이터의 보정과 장기간 바람 데이터를 활용한 MCP 수행, 장기간 SMP 예측에 따른 수치형 데이터 분석으로 인해 지도학습이 적합하다고 판단되어 이를 활용하여 다양한 기계학습 분석을 수행하였다. 기계학습의 지도학습 알고리즘의 종류는 다양하게 존재하는데 본 연구에서는 최근접 이웃 모델(K-nearest neighbor, KNN), Light gradient boost machine (LGBM), Support vector machine(SVM) 3가지 모델을 선정하여 분석을 수행하였다.

3.3.1 K-nearest neighbor(KNN)

Fig. 5는 KNN이 데이터 학습에 따른 분류 방법을 도식화하여 보여준다. KNN은 분류하려는 데이터를 기준으로 가장 가까운 거리의 K개의 데이터를 통해 다수결 투표를 통해 데이터를 분류하는 알고리즘이다. 그러므로 KNN은 거리를 측정 방법에 따라 결과가 크게 달라지는 알고리즘이다 [18].

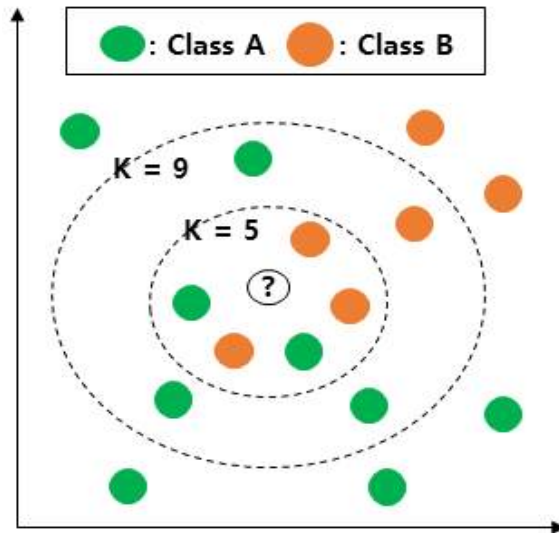


Fig. 5 Classification method of KNN

참조 데이터와 거리를 측정하는 방법에는 유클리드 거리(Euclidean distance), 맨하탄 거리(Manhattan distance), 민코프스키 거리(Minkowski distance) 등이 사용되며 민코프스키 거리는 유클리드 거리와 맨하튼 거리를 일반화한 것으로 식(4)와 같다. 식(4)를 통해 p가 1일 때 유클리드 거리이고, p가 2일 때 맨하튼 거리이다.

$$d(x^{(i)}, x^{(j)}) = \sqrt[p]{\sum_k |x_k^{(i)} - x_k^{(j)}|^p} \quad (4)$$

또한 KNN은 메모리를 기반으로 학습이 수행되므로 수집된 새로운 데이터에

대한 적용이 용이하며 파라미터(Parameter)의 수가 많지 않아 구현하기 쉽다는 장점이 있다.

3.3.2 Light gradient boosting machine(LGBM)

LGBM은 GBM(Gradient boosting machine) 알고리즘에서 발전된 알고리즘이다. Fig. 6은 GBM의 학습 방법을 도식화하여 나타낸다. GBM은 병렬적으로 트리(Tree)와 리프(Leaf)의 개수를 확장하는 균형 트리 분할(Leaf wise tree)방식을 채택하여 오차를 줄여나간다 [19].

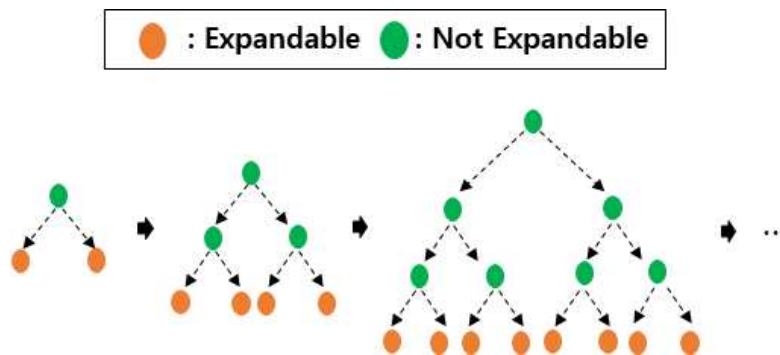


Fig. 6 Level wise tree method

이는 데이터의 숫자가 많지 않은 경우 빠른 학습 속도와 높은 정확도를 보이지만 데이터의 숫자가 많아지는 경우 트리의 개수도 증가하므로 학습의 속도가 느려진다는 단점이 있다. 이러한 단점을 개선하기 위해 LGBM은 학습과정에서 트리를 병렬적으로 개수를 늘리는 것이 아닌 트리의 리프 중심으로 분할(Leaf wise tree)하는 방법을 채택하였다. Fig. 7은 LGBM의 데이터 학습 방법인 리프 중심 분할(Leaf wise tree)을 도식화하여 보여준다. 리프 중심 분할은 데이터의 숫자가 많은 경우에도 학습 속도가 빠르다는 장점이 있다. 또한 전세계적으로 유명한 예측 모델 및 분석 대회인 케글(Kaggle)에서 대회 우승 실적이 가장 많은 모델이라고 알려져 있다.

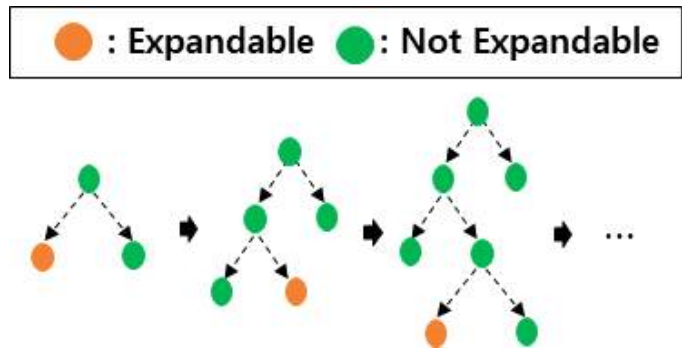


Fig. 7 Leaf wise tree method

3.3.3 Support vector machine(SVM)

SVM(Support vector machine)은 선형적 SVM과 비선형적 SVM으로 나뉜다. Fig. 8은 SVM의 선형 분류를 도식화하여 보여준다. 선형적 SVM은 선형적으로 분리할 수 있는 데이터에 대해서 마진(Margin)을 최대화하는 분리경계면(Separating hyperplane)을 식(4)를 활용하여 데이터를 분류한다 [20].

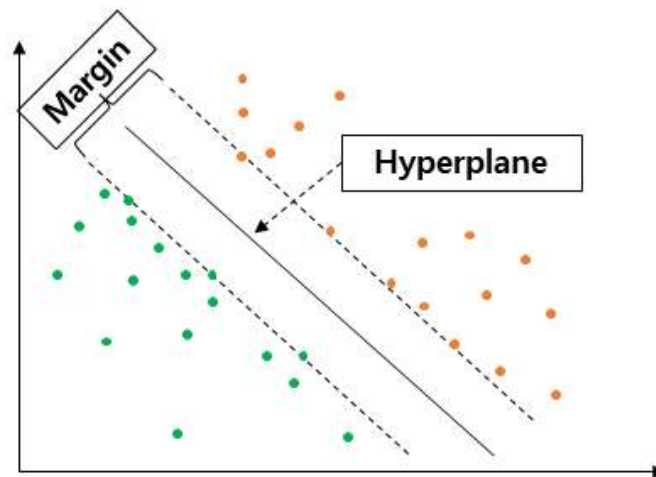


Fig. 8 Linear Classification of SVM

$$y = \mathbf{w} \cdot \mathbf{x} + b \quad (4)$$

여기서 \mathbf{x} 는 입력 값, \mathbf{w} 는 가중치, y 는 +1과 -1의 값을 갖는 출력 값, b 는 편향 값이다.

이때 분리경계면에서 가장 가까운 데이터를 Support vector라고 하는데 마진은 Support vector와 분리경계면까지의 거리를 의미하므로 식(5)와 같이 나타낼 수 있다.

$$Margin = \frac{2}{\|\mathbf{w}\|} \quad (5)$$

Fig. 9는 SVM의 비선형 분류 방법을 도식화하여 보여준다. SVM은 선형적으로 데이터를 분류할 수 없는 경우에 사상(Mapping)함수 ϕ 를 활용하여 고차원 공간으로 변환한다.

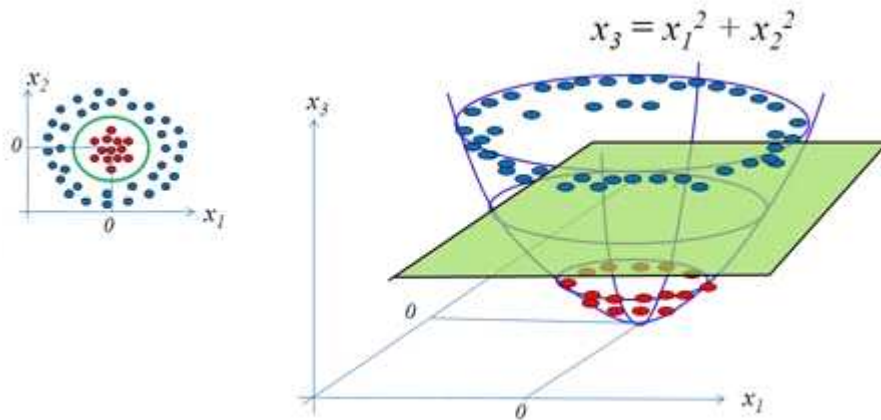


Fig. 9 Nonlinear classification of SVM [21]

이를 위해 아래 식(6)의 커널함수를 활용하여 비선형 결정함수를 도출하고 최적의 초평면을 구축하여 다양한 비선형 패턴을 분류한다 [20].

$$K(x^{(i)}, x^{(j)}) = \phi(x^{(i)})^T \phi(x^{(j)}) \quad (6)$$

Ⅲ. 연구 결과

1. 데이터 품질 검사(Data quality check)

기상탑은 설치하고 보통 1년이 넘는 시간동안 바람 데이터 측정을 수행한다. 하지만 설치된 측정 장비의 일시적 고장과 전압 이상, 환경적 요인 등의 다양한 원인으로 인해 측정되지 않거나 잘못된 값이 측정될 수 있다. 이러한 잘못된 값이 포함된 데이터를 활용하여 바람자원분석 등을 수행하는 경우 정확한 분석을 할 수 없으므로 Measenet(Measuring network of wind energy institutes)에서는 데이터 품질 검사 결과 데이터 복구율(Data recovery)이 90%이상인 데이터를 활용하여 분석을 수행하는 것을 권고한다 [22]. 그러므로 기상탑 데이터를 분석하기 전 데이터 품질 검사를 반드시 먼저 수행하고 이 절차를 통과한 고품질의 데이터를 활용하여 바람 분석을 수행해야 한다. 하지만 기상탑은 1년 이상의 바람을 측정하므로 많은 데이터가 축적되기 때문에 각각의 측정 장비의 측정 오류 구간을 자세히 파악·분석하기에는 매우 많은 시간이 소요되고 분석자의 많은 데이터 분석 경험을 필요로 한다. 이 절에서는 데이터 분석에 소요되는 시간을 단축하고 비전문가가 데이터 품질 검사를 쉽게 접근·수행할 수 있도록 데이터 품질 검사에 기계학습 방법을 적용하여 분석자에게 도움을 주고자 한다.

Fig. 10은 기계학습을 활용한 데이터 품질 검사 방법의 흐름도를 나타내며 그 절차는 다음과 같다. 첫째, 투입 변수 선정을 위해 상명목장지구 기상탑에서 수집된 데이터의 데이터 품질 검사를 수행한다. 둘째, 데이터 품질 검사에 따라 정상 구간과 오류 구간의 데이터를 분류한다. 셋째, 데이터 품질 검사 결과에 따라 분류된 데이터를 투입 변수로 선정하여 기계학습을 수행하고 데이터 품질 검사 모델을 생성한다. 넷째, 생성된 모델 중 학습 정확도가 높은 모델에 데이터 품질 검사를 수행하지 않은 대정해상 기상탑 데이터를 투입하여 정확한 데이터 품질 검사 모델의 정확도를 검증한다.

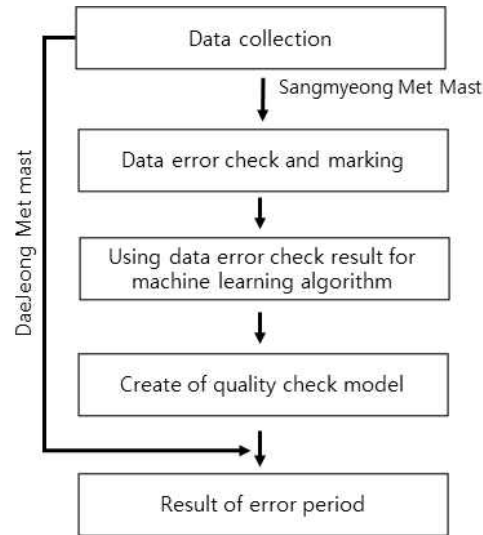


Fig. 10 Work flow of machine learning for data quality check

1.1 분석 대상 선정

Table 3은 1절에서 사용된 대정해상 기상탑(M#1)과 상명목장지구 기상탑(M#2) 데이터의 정보를 보여준다. 상명목장지구 기상탑은 2006년 1월 1일부터 2006년 12월 31일까지 약 1년간 측정되었고 측정기간 중 데이터 손실(Data loss)은 발견되지 않았다. 대정해상 기상탑은 2013년 8월 12일부터 2016년 10월 28일까지 약 3년 2개월 측정되었으며 2013년 8월 21일부터 2013년 9월 20일까지의 구간과 2015년 2월 14일부터 2015년 6월 16일까지 구간에서 데이터 손실이 발생하였다.

Table 3 Heights of wind sensors and measurement periods at M#1 and M#2

Items	Heights [m]	Components	Measurement periods
M#1	82.5A, 82.5B, 100.5, 111.5	Anemometer, Wind vane	2013.08.12. ~ 2016.10.28.
M#2	20, 30, 38, 40	Anemometer, Wind vane	2006.01.01. ~ 2006.12.31.

1.2 투입 변수 선정

모든 측정 장비는 6시간 이상 같은 오류가 지속적으로 발생하는 경우 오류 데이터로 분류하였으며 본 연구에서는 기계학습을 위한 오류 데이터의 기준을 다음과 같이 정의하였다.

- 연속해서 오프 셋(Off-set) 값이 측정되거나 또는 측정값이 지속적으로 같은 값을 측정되는 경우
- 서로 다른 높이의 풍속계에서 측정된 풍속이 역전 현상이 지속적으로 일어나는 경우
- 두 개의 인접한 높이의 풍향계에서 측정된 값이 $\pm 10^\circ$ 이상 차이가 발생하는 경우

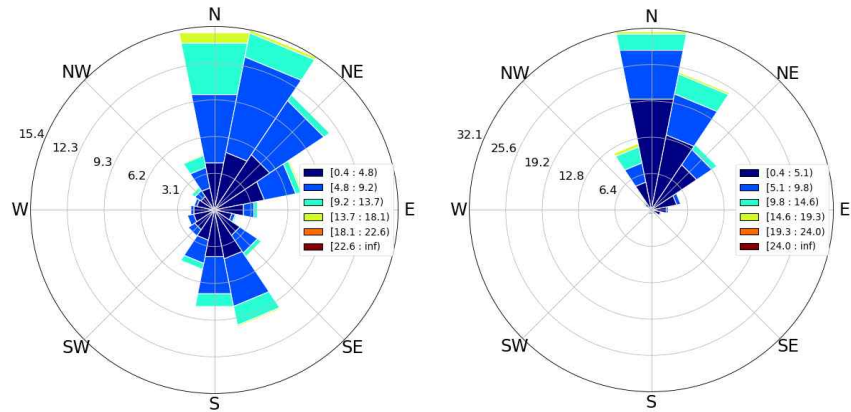
Table 4는 상명목장지구 기상탑의 1년 풍속계 데이터의 데이터 품질 검사 결과를 나타낸다. 투입 변수 선정을 위해 앞에서 정의한 데이터 품질 검사 기준에 따라 상명목장지구 기상탑 데이터를 대상으로 데이터 품질 검사를 수행하였다.

Table 4 data quality check result for wind speeds at Sangmyeong met mast

Anemometer				
Start	End	Height	Description	
2006.08.03. 01:00	2006.08.03. 21:50	40m	Sensing Error	
2006.08.07. 07:00	2006.08.07. 13:00	40m	Sensing Error	
2006.08.08. 18:20	2006.08.09. 03:20	40m	Sensing Error	
2006.08.12. 21:50	2006.08.14. 12:20	40m	Sensing Error	
2006.08.14. 08:00	2006.12.31. 21.50	40m	Sensing Error	
Number of error data		20,484	Recovery rate	61.02%

풍속계 데이터의 데이터 품질 검사 결과, 40m의 풍속계는 총 52,560개 중 20,484개의 데이터가 측정 오류로 분류되었고 데이터 복구율은 61.02%로 확인되었다.

다른 풍속계에서는 측정 오류가 발견되지 않았으며 모든 풍속계의 데이터 손실은 없었다. Fig. 11은 상명목장지구 기상탑 데이터의 데이터 품질 검사 전 30m와 40m 풍향계의 바람장미(Wind rose)를 보여준다.



(a) Wind rose at 30m height (b) Wind rose at 40m height
Fig. 11 Wind roses at Sangmyeong met mast

30m와 40m 풍향계의 높이 차이가 10m임을 감안한다면 바람장미의 형태가 다르게 나타났다. 40m 바람장미는 거의 북풍 위주의 바람이 측정되었지만 30m 바람장미는 북풍뿐만 아니라 남풍 계열의 빈도도 높았다. 30m와 40m의 풍향의 빈도가 다름이 확인되었으므로 두 풍향계를 신뢰하기 어렵다고 판단하였다. 이때, 측정 장비의 오류 등을 확인하기 위해 동기간의 주변 기상청에서 측정한 AWS (Automatic weather system), 재해석 데이터 등을 참고 자료로 활용할 수 있다. 따라서 상명목장지구 기상탑으로부터 약 3km 거리의 기상청 데이터인 서광 AWS를 참조하였다. Fig. 12는 상명목장지구의 기상탑 근처에 있는 서광 AWS의 위치와 바람장미를 보여준다.

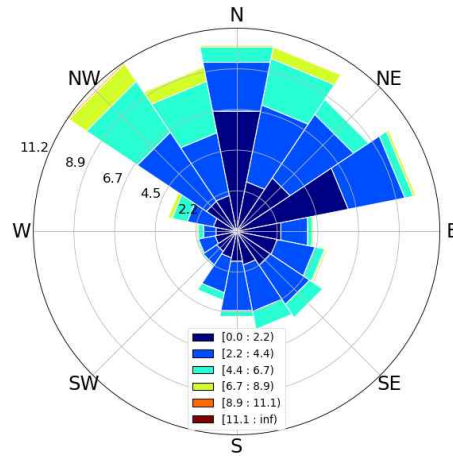


Fig. 12 Wind rose at Seogwang AWS

서광 AWS의 바람장미와 상명목장지구의 바람장미를 비교한 결과, 30m의 풍향계가 40m 풍향계보다 서광AWS와 비슷하였으므로 30m 풍향계가 비교적 신뢰성이 있는 것으로 판단된다. 따라서 30m에 설치된 풍향계를 기준으로 데이터 품질 검사를 오류 데이터 분류 기준에 따라 수행하였고 Table 5는 상명목장지구 기상탑의 풍향계에 대한 데이터 품질 검사 결과를 나타낸다.

Table 5 data quality check result for wind directions at Sangmyeong met mast

Wind vane			
Start	End	Height	Description
2006.01.10. 17:10	2006.01.13. 14:10	40m	Sensing Error
2006.01.28. 18:20	2006.01.31. 21:40	40m	Sensing Error
2006.02.06. 17:30	2006.02.07. 14:50	40m	Sensing Error
2006.02.10. 12:40	2006.02.11. 10:40	40m	Sensing Error
2006.02.12. 18:50	2006.02.14. 13:20	40m	Sensing Error
2006.02.19. 15:30	2006.02.22. 03:50	40m	Sensing Error
2006.03.04. 08:00	2006.03.06. 07:40	40m	Sensing Error
2006.03.06. 21:30	2006.03.11. 18:30	40m	Sensing Error
2006.03.13. 13:20	2006.10.24. 10:50	40m	Sensing Error
Number of error data	35,269	Recovery rate	32.9%

풍향계 데이터의 데이터 품질 검사 결과, 총 52,560개 중 35,269개의 데이터가 측정 오류로 분류되었고 데이터 복구율은 32.9%로 확인되었다. 또한 상명목장지구 기상탑의 기온계에 대해 데이터 품질 검사를 수행하였다. 분석 결과 기온계가 1년동안 영하의 기온을 측정하고 있는 것으로 나타나 기온계에서 측정된 데이터를 신뢰할 수 없었으므로 본 연구에서는 사용하지 않았다.

1.3 기계학습 적용을 위한 데이터 셋

Table 6은 데이터 품질 검사에 기계학습을 적용하기 위해 사용된 학습 데이터와 시험 데이터를 나타낸다. 상명목장지구 기상탑의 데이터 품질 결과를 활용하여 기계학습을 수행하고, 그 결과로 생성된 모델의 오류 데이터 분류 정확도 확인을 위해 대정해상 기상탑의 데이터를 활용하였다.

Table 6 Train and test data set for machine learning

Item	Train data (Sangmyeong met mast)	Test data (Daejeong met mast)
Collected period	2015.1.1. ~ 2019.12.31.	2013.08.12. ~ 2016.10.28.

Fig. 13는 기계학습을 활용한 데이터 품질 검사에 따른 입·출력 데이터 및 알고리즘을 나타낸다.



Fig. 13 Progress of machine learning for data quality check

풍속계 또는 풍향계의 정상데이터와 오류 데이터를 입력 변수로 투입하여 데이터 품질 검사 기준을 기반으로 기계학습을 수행하여 측정 오류 구간을 분류하여 출력하도록 기계학습을 수행하였다.

1.4 데이터 품질 검사의 기계학습 결과

Table 7은 상명목장지구 기상탑 데이터를 활용한 기계학습 모델의 학습 정확도를 보여준다. 학습 정확도가 가장 높았던 모델은 LGBM으로 정확도가 98.80%이었고 SVM은 학습 정확도가 98.07%로 3가지 모델 중 가장 낮았다. 따라서 대정해상 기상탑 데이터 품질 검사에 학습 정확도가 높았던 LGBM 모델을 적용하여 데이터 품질 검사를 수행하였다.

Table 7 Train accuracy of every machine learning algorithm

Item	KNN	LGBM	SVM
Accuracy [%]	98.72	98.80	98.07

우선, 기계학습 된 LGBM의 실제 분류 정확도 확인을 위해 1.2절에서 정의한 오류 데이터 기준을 기반으로 분석자가 대정해상 기상탑의 데이터 품질 검사를 수행하였다. Table 8은 대정해상 기상탑을 대상으로 분석자와 기계학습 모델인 LGBM의 데이터 품질 검사 결과를 비교하여 보여준다. 대정해상 기상탑에 대한 데이터 품질 검사 결과 총 20개의 구간에서 측정 오류가 발생한 것으로 판단되었다. 분석자가 판단한 하나의 실제 측정 오류 구간에서 LGBM은 여러 구간으로 나누어 분류하였다. 이는 측정 오류 구간에서 풍속계가 일정기간 잘못된 측정값을 측정하다가 일시적으로 정상 값을 회복하는 경우 기계학습 모델이 측정 오류 구간을 분류하는데 어려움을 겪는 것이라 판단된다. 분석자가 데이터 품질 검사를 수행한 결과를 기준으로 LGBM의 오류구간에 대한 분류 정확도는 93.51%로 산출되었다.

Table 8 Comparison of manual and machine learning results for wind speed data quality check

No	Manual analysis result		Machine learning result	
	Time stamp		Time stamp	
	Start	End	Start	End
1	2014-12-16 16:10	2015-02-14 09:00	2014-12-16 16:30	2014-12-30 20:40
			2014-12-30 21:20	2015-02-14 09:00
2	2015-07-18 17:20	2015-07-21 23:10	2015-07-19 18:00	2015-07-21 14:20
			2015-07-21 17:00	2015-07-21 23:10
3	2015-11-07 11:40	2015-11-19 09:20	2015-11-07 11:40	2015-11-17 21:10
			2015-11-17 21:30	2015-11-18 07:10
			2015-11-18 07:30	2015-11-19 09:10
4	2015-11-26 06:20	2015-11-27 02:30	2015-11-26 06:20	2015-11-26 15:40
			2015-11-26 16:10	2015-11-27 02:30
5	2016-01-04 15:40	2016-01-07 08:10	2016-01-04 15:40	2016-01-06 18:30
6	2016-01-18 17:40	2016-01-24 21:00	2016-01-18 17:40	2016-01-24 21:00
7	2016-02-07 16:10	2016-02-11 18:40	2016-02-07 16:10	2016-02-11 18:40
			2016-02-13 13:30	2016-02-19 13:30
			2016-02-19 13:50	2016-02-26 1:40
			2016-02-26 2:00	2016-03-15 12:20
			2016-03-15 12:40	2016-03-16 07:30
9	2016-03-16 19:20	2016-03-17 04:20	2016-03-16 19:30	2016-03-17 04:20
10	2016-03-18 17:30	2016-03-18 23:20	2016-03-18 17:30	2016-03-18 23:20
			2016-03-20 6:00	2016-03-21 02:10
			2016-03-21 4:20	2016-03-21 21:10
			2016-03-22 2:00	2016-03-22 08:10
			2016-03-22 11:10	2016-03-23 01:00
			2016-03-23 13:10	2016-03-23 21:50
			2016-03-24 19:10	2016-03-25 01:10
			2016-03-25 06:50	2016-03-25 16:30
			2016-03-25 18:50	2016-03-26 01:30
			2016-03-27 16:10	2016-03-30 06:10
11	2016-03-20 06:00	2016-03-30 06:10	2016-03-20 06:00	2016-03-30 06:10
			2016-03-21 16:50	2016-04-03 21:20
			2016-04-03 21:50	2016-04-04 05:10
			2016-04-04 22:10	2016-04-05 05:50
			2016-04-06 00:50	2016-04-08 22:20
			2016-04-08 23:00	2016-04-09 18:20
			2016-04-10 00:20	2016-04-10 08:50
			2016-04-10 10:30	2016-04-10 17:40
			2016-04-10 18:10	2016-04-11 8:10
			2016-04-16 06:20	2016-04-16 21:30
12	2016-04-01 16:50	2016-04-11 18:00	2016-04-17 03:20	2016-04-17 11:40
			2016-04-17 20:30	2016-04-18 2:40
			2016-04-18 09:20	2016-04-19 18:40
			2016-04-21 21:00	2016-04-23 3:20
			2016-04-23 10:00	2016-04-24 18:50
13	2016-04-15 08:00	2016-04-19 18:50	2016-04-25 00:00	2016-04-25 13:10
			2016-04-26 15:20	2016-04-26 21:20
			2016-04-27 04:10	2016-04-28 8:20
			2016-04-29 10:40	2016-04-29 19:30
14	2016-04-21 21:00	2016-04-28 08:20	2016-04-29 10:40	2016-04-29 19:30
15	2016-04-29 10:40	2016-04-29 19:30	2016-04-29 10:40	2016-04-29 19:30

No	Manual analysis result		Machine learning result	
	Time stamp		Time stamp	
	Start	End	Start	End
16	2016-05-04 15:20	2016-05-12 22:00	2016-05-04 15:20	2016-05-06 11:20
			2016-05-06 12:00	2016-05-12 4:20
			2016-05-12 04:40	2016-05-12 20:30
17	2016-05-13 22:40	2016-05-14 21:40	2016-05-13 22:40	2016-05-14 21:30
18	2016-05-17 21:00	2016-05-31 07:50	2016-05-17 21:00	2016-05-24 21:50
			2016-05-24 22:10	2016-05-27 18:50
			2016-05-27 19:10	2016-05-29 9:50
19	2016-06-22 02:20	2016-06-23 17:40	2016-06-22 02:20	2016-06-23 17:40
			2016-06-28 01:10	2016-07-09 8:30
			2016-07-09 10:00	2016-07-09 23:10
20	2016-06-28 01:10	2016-07-15 17:50	2016-07-09 23:30	2016-07-10 9:40
			2016-07-10 10:50	2016-07-14 16:00
			2016-07-14 17:00	2016-07-15 17:40

풍향계에 대해서 분석자가 데이터 품질 검사를 수행하였고 측정 오류는 발견되지 않았다. LGBM을 활용한 풍향계의 데이터 품질 검사를 수행한 결과 또한 측정 오류가 발견되지 않았다.

1.5 분석 결과 및 고찰

1절에서는 바람자원평가의 한 부분인 데이터 품질 검사에 대해 기계학습을 적용하였고 그 결과는 아래와 같다.

1) 상명목장지구 기상탑 데이터를 대상으로 데이터 품질 검사를 수행한 결과 총 52,560개의 데이터 중 20,484개의 데이터가 측정 오류로 분류되었고 데이터 복구율은 61.02%였다. 풍향계 데이터의 경우 52,560개의 데이터 중 35,269개의 측정 오류가 발생하였다. 기온계 데이터의 경우 영하의 기온을 지속적으로 측정하고 있어 신뢰할 수 없으므로 분석에 사용하지 않았다.

2) 데이터 품질 검사를 통해 분류된 상명목장지구 기상탑 데이터를 활용하여 기계학습을 수행하였다. 기계학습 알고리즘을 활용한 데이터 학습 결과는 LGBM이 98.80%로 학습 정확도가 가장 높았고 SVM이 98.07로 학습 정확도가 3개의 모델 중 가장 낮았다. 학습 정확도가 높았던 LGBM을 활용하여 대정해상 기상탑의 데이터 품질 검사를 수행하였다.

3) 대정해상 기상탑에 대해 분석자가 수행한 데이터 품질 검사 결과 풍속계는 총 20개의 구간에서 측정 오류가 발생한 것으로 확인되었다. LGBM을 활용한 대정해상 기상탑에 대한 데이터 품질 검사 결과는 분석자가 분류한 측정 오류 구간과 유사했지만 분석자가 데이터 품질 검사 결과를 기준으로 LGBM의 분류 정확도는 93.51%였다. 풍속계가 일정기간 잘못된 측정값을 측정하다가 일시적으로 정상 값을 회복하는 경우 기계학습 모델은 정상구간과 측정 오류 구간을 분류하였지만 분석자는 측정 오류 구간으로 분류하여 분석자와 기계학습간의 결과에 대한 차이가 발생하였다.

Ⅲ장의 1절에서는 상명기상탑의 풍속계와 풍향계 데이터를 활용하여 기계학습을 수행하였고 기온계 데이터는 신뢰할 수 없어 학습 데이터로 활용하지 않았다. 추후 연구에서 풍속계와 풍향계 데이터뿐만 아니라 기온계 및 습도계 등의 다양한 측정 장비 데이터를 포함하여 기계학습을 수행하는 경우 전압 이상, 결빙 등으로 상세한 분류를 수행하는 연구를 할 수 있을 것으로 기대된다.

2. 타워 섀도우 영향 보정(Tower shadow calibration)

풍력발전단지 설계에서 AEP와 CF를 추정하기 위해서는 풍력발전기의 허브 높이의 풍속을 필요로 한다. 이를 측정하기 위해서는 풍력발전기의 허브 높이에 해당하는 기상탑을 설치하여 풍속을 측정해야하지만 최근 풍력발전기의 대형화와 기상탑의 높이가 높아짐에 따라 많은 비용이 발생하는 등의 다양한 요인으로 허브 높이와 동일한 기상탑을 설치하는데 어려움을 겪고 있다. 그러므로 기상탑과 풍력발전기의 허브 높이가 다른 경우 허브 높이의 바람을 추정하기 위한 다양한 방법이 고안되었고 대표적으로 멱법칙(Power law)과 로그법칙(Log law)을 활용하여 윈드시어(Wind Shear) 지수를 산출하고 외삽(Extrapolation) 또는 내삽(Interpolation)을 통하여 허브 높이의 풍속을 추정한다 [23]. 이러한 추정 과정에서 기상탑의 다양한 높이에 설치되어 있는 풍속계 데이터를 활용하는데 기상탑의 제일 상단에 설치된 풍속계는 자유 풍속을 측정하지만 그 외의 높이에서 측정된 풍속계는 설치 위치와 방향 등의 다양한 원인으로 인해 특정 방위각에서 기상탑 타워의 후류 영향을 받는 타워섀도우(Tower Shadow) 영향을 받을 수 있다. 이러한 구간에서 풍속계는 실제 자유 풍속보다 증가 또는 감소된 풍속을 측정한다. Fig. 14는 기상탑의 종류에 따른 타워섀도우 영향에 의한 풍속의 변화를 나타낸다.

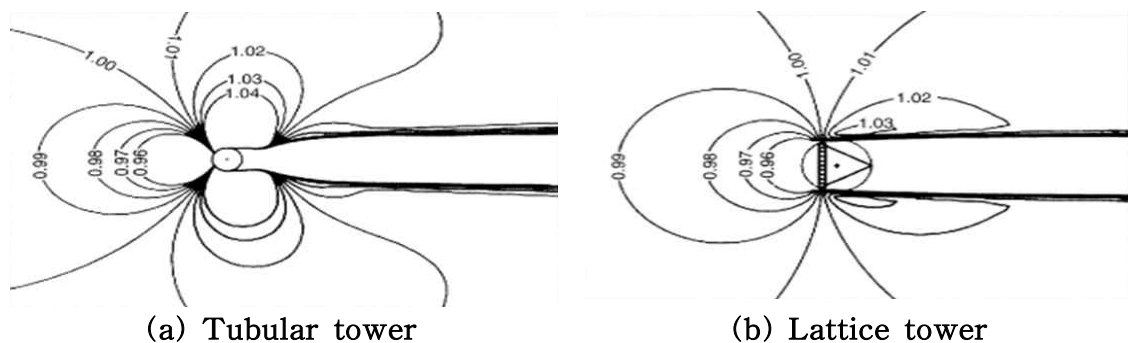


Fig. 14 Effects of tower shadow in accordance with tower types [23]

특정 방위각에서 타워섀도우 영향으로 왜곡된 풍속이 포함된 데이터로 허브 높이의 풍속 추정하는 경우, 실제 허브 높이의 풍속과 오차가 크게 발생할 수 있다.

이러한 타워쉐도우 영향을 받는 구간의 데이터를 보정하는 방법에는 CFD(Computational fluid dynamics)를 활용하는 방법[24]가 연구되었지만 이는 소프트웨어에 대한 지식과 분석 경험이 부족한 분석자의 경우 보정에 어려움 겪을 수 있다. 따라서 2절에서는 통계적 기법을 활용하여 풍속계 데이터의 타워쉐도우 영향 발생 구간을 검출하고 기계학습을 활용하여 타워쉐도우 영향을 받은 풍속계 데이터를 보정하는 방법을 분석하고자 한다. Fig. 15는 기계학습 방법을 활용하여 타워쉐도우 영향을 보정하는 순서를 보여주며 그 절차는 다음과 같다. 첫째, 통계적 기법을 활용하여 타워쉐도우 영향이 발생한 구간을 검출한다. 둘째, 타워쉐도우 영향이 발생하지 않은 정상 구간의 데이터를 활용하여 기계학습을 수행한다. 셋째, 기계학습된 모델에 타워쉐도우 영향이 발생한 구간의 데이터를 투입하여 보정을 수행한다. 넷째, 보정된 데이터를 같은 높이의 다른 풍속계의 데이터를 활용하여 검증을 수행한다.

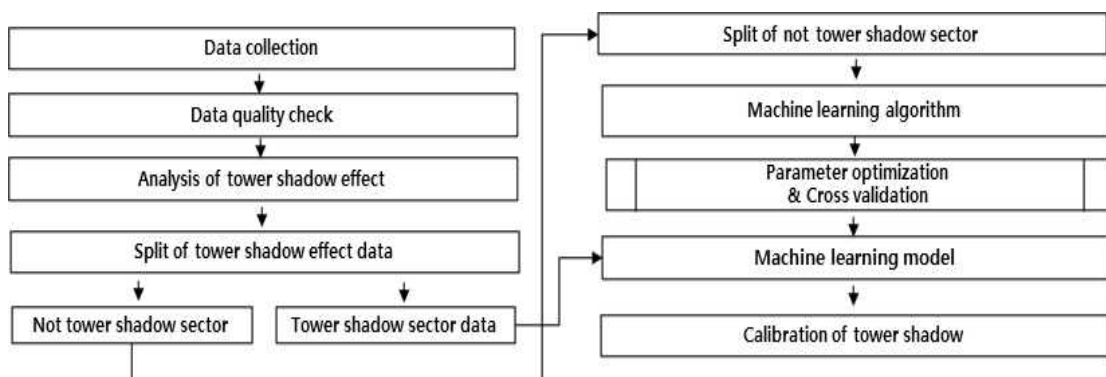


Fig. 15 Work flow for tower shadow calibration

2.1 타워쉐도우 영향 보정 방법에 따른 분석대상 선정

Table 9는 2절의 연구를 위해 사용된 데이터를 나타낸다. 대정해상 기상탑의 11.5m 풍속계와 82.5m 풍속계 A, B, 106.5m 풍향계가 활용되었다.

Table 9 Height of wind sensors and measurement period of Daejeong offshore met mast

Wind sensors	Heights [m]	Measurement periods
Anemometer	111.5, 82.5A, 82.5B	2013.08.12.
Wind vane	106.5	~ 2016.10.28.

Fig. 16은 대정해상 기상탑 82.5m 높이에 설치되어 있는 A와 B 풍속계에 설치된 방향을 보여준다. 82.5m 높이에 설치되어 있는 A풍속계의 설치 방향은 진북을 기준으로 60°방향에 설치되어 있으며 B풍속계의 설치 방향은 240°방향으로 설치되어 있다.

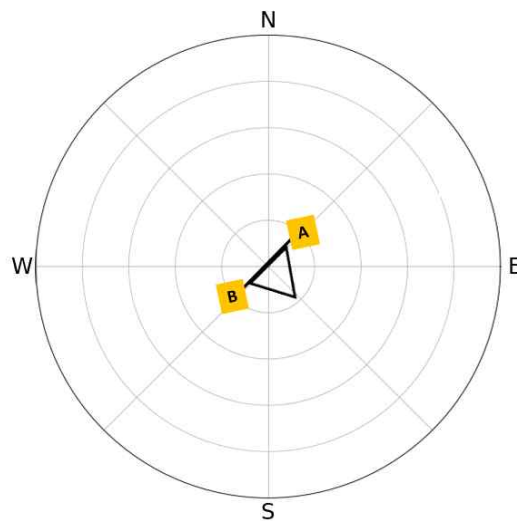


Fig. 16 Azimuth angles of anemometers A and B

2.2 타워쉐도우 영향 구간 검출

서로 다른 풍속계의 풍속의 차이를 확인하기 위해 풍속비(Wind Speed Ratio)를 활용하였고 이를 산출하기 위한 식은 다음과 같다. 풍속비가 1에 근사할수록 서로 다른 풍속계의 풍속 차이가 낮음을 의미한다.

$$\text{Wind Speed Ratio} = \text{Anemometer}_1(V_i) / \text{Anemometer}_2(V_i) \quad (7)$$

여기서 $\text{Anemometer}_1(V_i)$: 풍속계의 i 번째의 풍속, $\text{Anemometer}_2(V_i)$: 풍속계의 i 번째의 풍속

Fig. 17은 풍속비 산포도와 5°빈(Bin)으로 분류한 결과를 보여준다. 풍속비는 식(7)을 활용하여 111.5m / 82.5m A 풍속계와 111.5m / 82.5m B 풍속계에 대하여 산출한 결과이다. 타워체도우 영향으로 인해 82.5m 높이의 두 개의 풍속계에서 왜곡된 풍속이 특정 방위각에서 측정된 것을 확인할 수 있다.

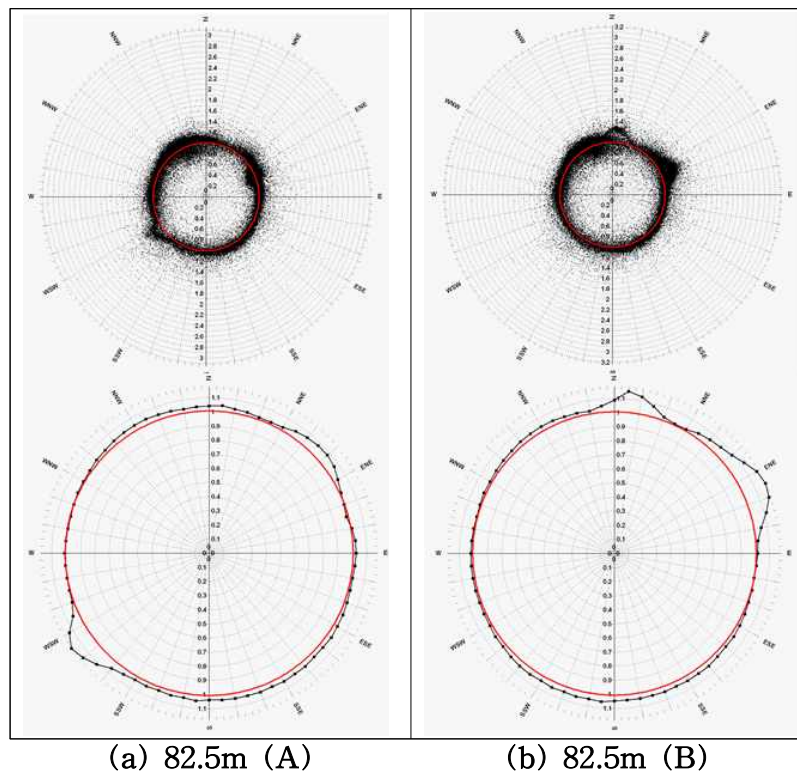


Fig. 17 Effects of tower shadow

본 연구에서는 신뢰성 있는 타워체도우 구간을 검출하기 위해 다섯 숫자 요약 (five-number summary)의 통계적 기법을 활용하였다. 풍속 데이터를 5°빈으로

분류하여 각 빈의 풍속비에 대한 중앙값(Median)을 산출하였다. 또한 중앙값이 전체 구간의 풍속비의 사분위수(Quartile) 25% (Q_{25}) 이상이거나 75% (Q_{75}) 이하인 경우를 타워쉐도우 영향을 받지 않는 정상적인 구간으로 정의하였고 이를 다음 식과 같이 나타내었다.

$$M_{normal} = \begin{cases} \sum_{i=1}^n (Q_{25} < Median_i) \\ \sum_{i=1}^n (Q_{75} > Median_i) \end{cases} \quad (8)$$

여기서 M_{normal} 은 타워쉐도우 영향을 받지 않은 정상 구간의 중앙값이고 Q_{25} 은 사분위수 25%이고 Q_{75} 은 사분위수 75%이다.

난류, 대기안정도, 풍속계의 폐쇄 영향(Blockage effect) 등의 다양한 변동으로 인한 풍속의 변화는 기상탑에 의한 영향을 받지 않는 요인임에도 불구하고 타워쉐도우 영향을 받는 구간으로 검출될 수 있다. 그러므로 이러한 영향을 분류하기 위해 정상 구간의 속해있는 전체 중앙값들의 평균을 산출하고 이를 활용하여 정상구간에 대한 중앙값의 편차를 산출한다. 그 결과 산출된 편차의 절대값을 평균화하여 도출된 값을 가중치로 정의하였고 이는 식(9)로 나타내었다.

$$w = \frac{\sum_{i=1}^n \left(\left| nM_{normal}(i) - \sum_{i=1}^n M_{normal}(i) \right| \right)}{n^2} \quad (9)$$

식(10)은 사분위수에 가중치를 적용한 타워쉐도우 영향 구간 검출 기준을 나타낸다.

$$T = \begin{cases} \sum_{i=1}^n ((Q_{25} - w) > Median_i) \\ \sum_{i=1}^n ((Q_{75} + w) < Median_i) \end{cases} \quad (10)$$

여기서 T 는 가중치를 적용한 타워쉐도우 영향을 받는 구간이다.

Fig. 18은 식(10)을 활용하여 검출된 82.5m 높이의 A풍속계에 대한 타워쉐도우 영향을 받는 구간을 나타내며 225° ~ 240°빈 구간이 타워쉐도우 영향을 받는 구간으로 검출되었다.

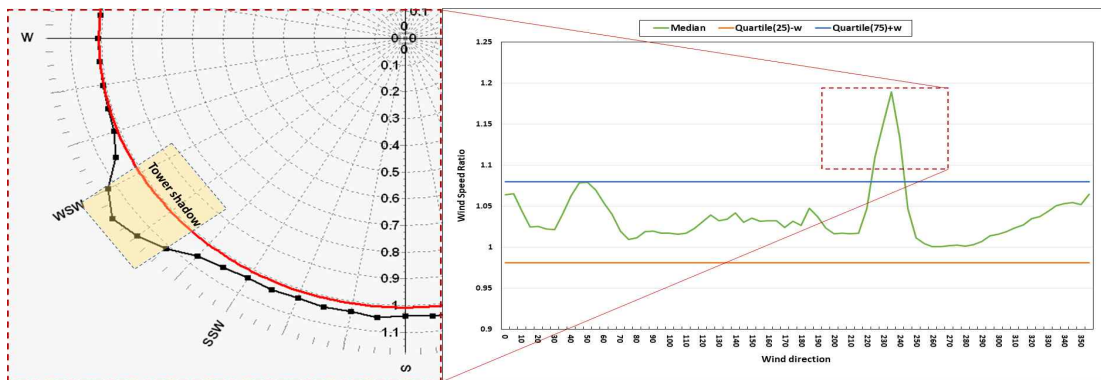


Fig. 18 Detection of tower shadow at anemometer A

Fig. 19은 식(10)을 활용하여 검출된 82.5m 높이의 B풍속계에 대한 타워쉐도우 영향을 받는 구간을 보여준다. 82.5m의 B풍속계는 5° ~ 15°빈구간과 55° ~ 75° 빈 구간이 타워쉐도우 영향을 받는 구간으로 검출되었고 그 외의 구간에서는 타워쉐도우 영향을 받지 않는 것으로 확인되었다. A풍속계와 다르게 두 개의 구간에서 타워쉐도우 영향을 받는 것으로 나타났다. 이를 확인한 결과 B풍속계 근처에 측정 장비에 전원 공급 케이블이 제대로 결박되어 있지 않아 바람에 의한 휘날림이 발생하여 5° ~ 15°빈 구간에서 왜곡이 발생된 것으로 판단된다.

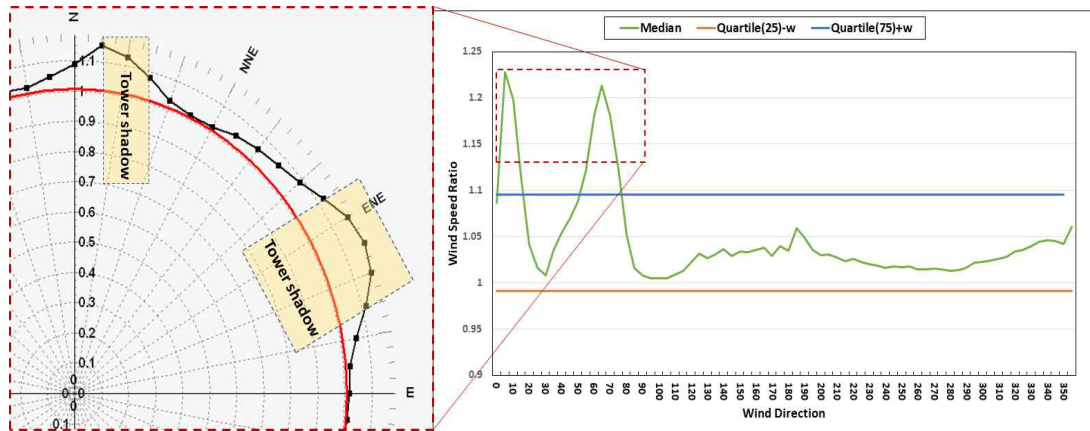


Fig. 19 Detection of tower shadow at anemometer B

타워쉐도우 영향은 지형, 기상탑 돛대 설치 각도 등의 다양한 원인으로 발생하는 형태가 상이하게 나타날 수 있다. 분석자 판단에 따라 식(10)에 추가적인 타워쉐도우 영향을 받는 구간의 빈을 확장 또는 축소를 고려할 수 있으며 이는 식(11)에 나타내었다.

$$TS = T \pm \alpha \quad (11)$$

여기서 TS 는 추가적인 고려된 빈의 크기가 적용된 타워쉐도우 구간, T 는 타워쉐도우 영향을 받는 구간, α 는 타워쉐도우 영향을 받는 구간에 대해 추가적으로 고려하고자하는 빈의 크기이다.

본 연구에서는 82.5m의 A풍속계의 타워쉐도우 구간에 대해 5°빈을 추가적으로 고려하였고 그 결과 타워쉐도우 보정을 필요로 하는 구간은 220° ~ 245°빈으로 확인되었다. Fig. 20은 식(11)을 활용하여 82.5m A풍속계의 타워쉐도우 영향 구간을 검출한 결과를 보여준다. 기계학습을 활용하여 타워쉐도우 영향 구간에 대한 보정을 수행하였으며 동일한 높이의 B 풍속계를 활용하여 비교·검증을 수행하였다.

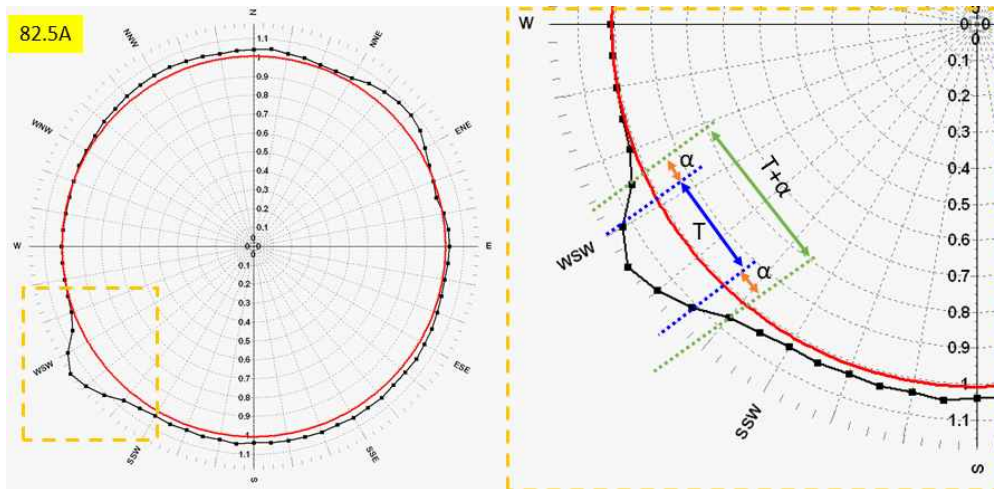


Fig. 20 Azimuth angle for tower shadow calibration

타워쉐도우 영향을 받지 않는 정상 구간을 학습 데이터로 활용하였으며 풍력발전기의 시동풍속(Cut-in wind speed)보다 낮은 풍속은 연간발전량 산출에 많은 영향을 주지 않으므로 다양한 분석에서 고려가 되지 않고 작은 변화에도 풍속비의 변동이 크게 나타나므로 본 연구에서는 시동풍속을 4m/s이라고 가정하고 그 이하의 풍속은 기계학습과 보정 대상에서 제외하였다.

2.3 기계학습 데이터 셋

Table 10은 타워쉐도우 영향을 받는 구간의 보정을 위해 활용된 학습 데이터와 시험 데이터를 보여준다. 학습 데이터는 타워쉐도우 영향을 받지 않는 정상 구간의 데이터 85,487개를 활용하여 기계학습을 수행하였다. 학습된 모델을 활용하여 타워쉐도우 보정이 필요한 1,243개의 데이터에 대한 보정을 수행하였다.

Table 10 Number of data points for machine learning

Collected period	Train data	Test data
2013.08.12. ~ 2016.10.28.	85,487	1,243

Fig. 21은 타워쉐도우 보정에 따른 입·출력 변수를 보여준다. 타워쉐도우 영향을 받는 220° ~ 245°빈에 해당하는 111.5m의 풍속계 데이터를 입력변수로 기계학습 모델에 투입하면 타워쉐도우 영향에 대한 보정 값이 출력되도록 기계학습을 수행하였다.



Fig. 21 Progress for tower shadow effect calibration

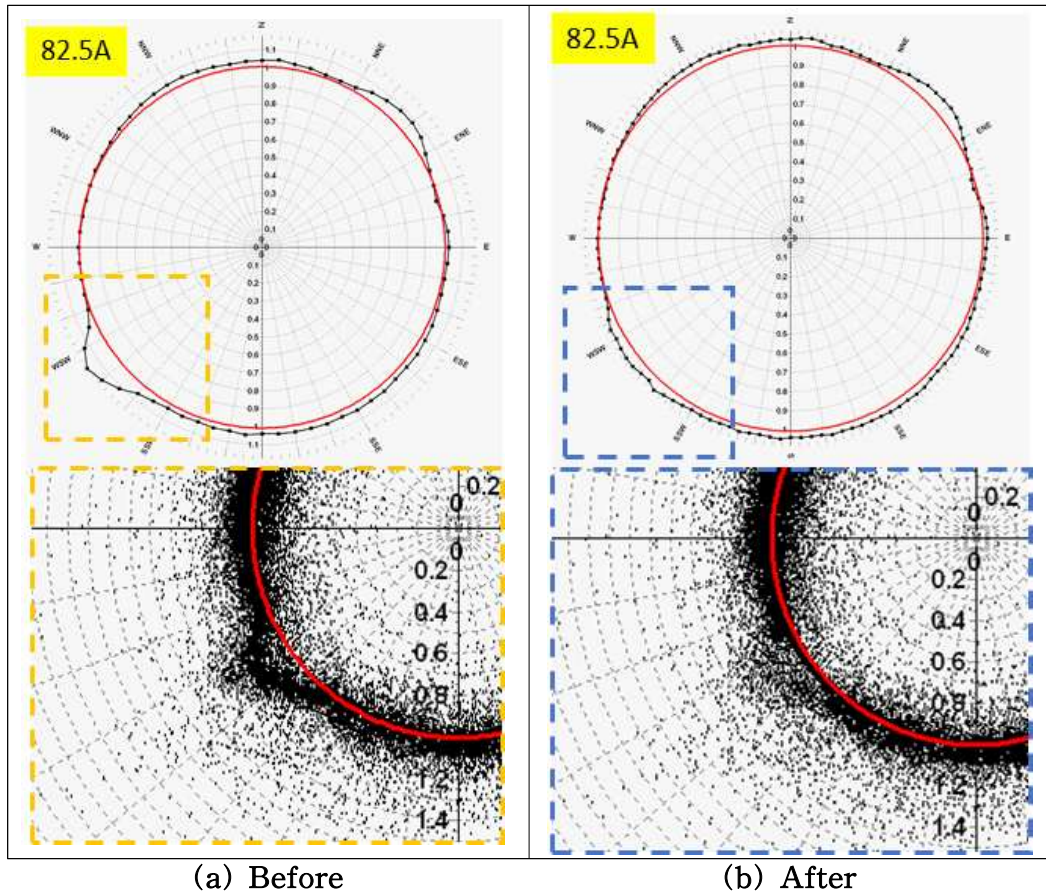
2.4 타워쉐도우 영향 보정을 위한 기계학습 결과

Table 11은 기계학습 모델이 타워쉐도우 영향을 받지 않은 정상 구간에 대한 데이터를 학습하는 과정의 학습 정확도를 보여준다. 3개의 기계학습 모델 모두 학습 정확도가 높았는데 그 중에서도 SVM은 RMSE와 R²이 각각 0.4753, 0.9817으로 학습 정확도가 가장 높았다. 따라서 학습 정확도가 높았던 SVM을 활용하여 타워쉐도우 영향을 받는 구간에 대한 보정을 수행하였다.

Table 11 Train accuracy of every machine leaning algorithm

Item	KNN	LGBM	SVM
RMSE [m/s]	0.4928	0.4883	0.4753
R ² [-]	0.9815	0.9813	0.9817

Fig. 22는 SVM을 활용한 82.5m A 풍속계의 타워쉐도우 영향을 받는 구간의 보정 전·후를 보여준다. 대체로 타워쉐도우 영향을 받는 구간인 220° ~ 245°빈 구간이 잘 보정한 것으로 보인다.



(a) Before (b) After
 Fig. 22 Graphical comparison with before and after tower shadow calibration

타워쉐도우 영향 보정 결과에 대해 82.5m의 B 풍속계 데이터를 참 값으로 활용하여 $220^{\circ} \sim 245^{\circ}$ 빈 구간에 대한 교차검증을 수행하였다. Table 12는 $220^{\circ} \sim 245^{\circ}$ 빈 구간에 대해 RMSE와 R^2 을 활용하여 보정 전·후를 비교한 결과를 보여준다. 타워쉐도우 영향을 보정하기 전 82.5m의 A와 B 풍속계의 RMSE는 1.009m/s 이고 R^2 은 0.891으로 산출되었다. 기계학습의 SVM을 활용하여 타워쉐도우 영향 구간에 대해 보정을 수행한 결과, 82.5m의 A와 B 풍속계의 RMSE는 0.2611m/s 산출되어 보정 전과 비교했을 때 0.7479m/s 개선되었다. 보정 후 R^2 은 0.9926으로 산출되었고 보정 전과 비교하였을 때 0.1016 개선되었다.

Table 12 Comparison of before and after tower shadow calibration using RMSE and R²

Item	Before	After	Difference
RMSE [m/s]	1.0090	0.2611	- 0.7479
R ² [-]	0.8910	0.9926	+ 0.1016

2.5 분석 결과 및 고찰

본 연구는 통계적 기법을 활용하여 타워쉐도우 영향을 받는 구간을 검출하고 기계학습 방법을 활용하여 검출된 구간에 대해 보정을 수행하였으며 그 결과는 아래와 같다.

1) 82.5m 높이의 A풍속계에 대해 다양한 요인을 고려하여 5°빈을 추가 적용한 결과 220° ~ 245°빈 구간이 타워쉐도우 영향을 받는 구간으로 검출되었고 기계학습 방법을 활용하여 보정을 수행하였다. 또한 보정에 따른 비교·검증을 위해 동일한 높이의 B풍속계를 활용하였다.

2) 타워쉐도우 영향을 받지 않는 정상 구간의 데이터를 활용하여 기계학습을 수행하였고 3개의 기계학습 모델 중 SVM의 RMSE와 R²이 각각 0.4753, 0.9817으로 학습 정확도가 가장 높았다. 따라서 타워쉐도우 영향을 받는 구간에 대해 SVM을 활용하여 보정을 수행하였다.

3) 82.5m의 A와 B풍속계 220° ~ 245°빈 구간에 대한 RMSE와 R²은 1.009m/s, 0.891으로 산출되었다. SVM을 활용하여 82.5m의 A의 타워쉐도우 영향을 받는 구간에 대해 보정을 수행한 결과, 82.5m의 A와 B 풍속계의 RMSE는 0.2611m/s 산출되어 보정하기 전과 비교하였을 때 0.7479m/s 개선되었다. 보정 후 R²은 0.9926으로 산출되었고 보정하기 전과 비교하였을 때 0.1016 개선되었다.

III장의 2절은 타워쉐도우 영향을 받는 구간을 통계적인 기법을 활용하여 검출된 구간에 대해 기계학습을 활용하여 보정을 수행하였다. 추후 연구에서는 CFD 등의 유동 해석 결과를 추가적으로 결합하는 연구를 수행한다면 기계학습을 통해 보정 결과의 정확성을 높일 수 있을 것으로 기대된다.

3. MCP(Measure-correlate-predict)

바람은 매년 그 세기와 방향이 조금씩 변동하는 자원이다. 풍력발전기의 수명이 약 20년 이상임을 감안하면 풍력발전단지 사업 타당성 검토 시 단기 바람 데이터만을 이용하여 다양한 분석을 수행하는 경우, 연간 바람의 변동성 및 계절적 영향을 반영하기 어렵다. 기상탑을 설치하고 장기간동안 운용하며 높은 품질의 바람 데이터를 수집하는 방법이 가장 좋은 방법이지만 많은 자본과 시간이 투자되므로 현실적인 한계가 있다. 그러므로 MCP방법을 통해 풍력발전단지 예정 사이트의 단기 측정 데이터와 주변의 참조할 수 있는 기상청 데이터, 재해석 데이터 등의 참조 데이터(Reference data)의 상관관계를 이용하여 해당 사이트의 장기적인 바람의 특성을 추정한다. MCP방법[26,27]에는 대표적으로 Regression, Matrix, Neural Network 등이 있으며 풍력발전단지 개발 소프트웨어인 WindPRO에서 제공되는 방법이다. 분석자들 사이에서 신뢰성 있는 MCP 수행 방법이라고 알려져 있다. 그러므로 본 연구에서는 기존 MCP방법과 기계학습을 활용한 MCP방법을 비교하여 MCP 분석 수행시 분석자가 다양한 MCP 분석을 수행할 수 있도록 돕고자 한다. Fig. 23은 기계학습 방법을 활용하여 MCP 수행 순서를 보여주며 그 절차는 다음과 같다. 첫째, 기상탑에서 수집된 데이터와 주변의 장기간 데이터 간의 상관성 분석을 수행한다. 둘째, 상관성이 높은 데이터를 활용하여 기계학습을 수행한다. 셋째, 기계학습을 통해 생성된 모델과 기존의 MCP방법을 실제 기상탑 데이터와 비교·분석한다.

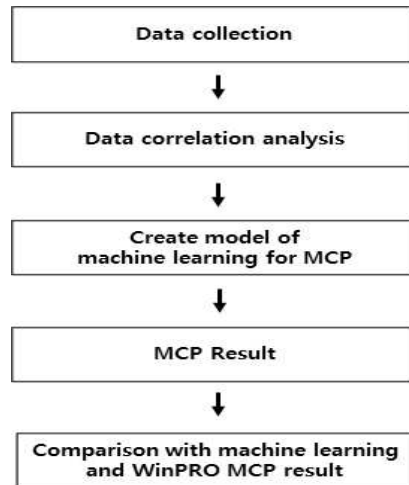


Fig. 23 Work flow for machine learning for MCP application

3.1 MCP 분석대상 선정

Table 13은 3절의 연구에 사용된 대정해상 기상탑과 참조 데이터인 ERA-5의 데이터를 보여준다. MCP 수행을 위해 두 가지 구간의 데이터를 활용하였으며 2015년 7월 1일부터 2016년 6월 30일까지의 1년간 데이터와 2013년 10월 1일부터 2015년 1월 31일까지 약 1년 4개월간의 데이터를 활용하였다.

Table 13 Information wind data of the offshore met mast and ERA-5 of collected data set

Items	Heights [m]	Data Components	Data Periods
M#1	100.5		
E#1	100.0	Wind Speed,	2013.10.01. ~ 2015.01.31.
E#2	100.0	Wind Direction	/ 2015.07.01. ~ 2016.06.30.
E#3	100.0		
E#4	100.0		

3.2 MCP를 위한 투입 변수 선정

MCP방법을 활용한 풍속의 정확한 추정을 위해 1년간의 기상탑 데이터와 동기 간의 참조 데이터 중 상관계수(R)가 높은 참조 데이터를 활용하여야 한다. Table 14는 두 데이터 간의 풍속의 상관관계에 따른 참조 데이터의 품질의 나누는 기준 [22]를 나타낸다. 상관관계가 높을수록 장기간 데이터 추정에 따른 품질이 올라가며, 낮을수록 떨어진다. 일반적으로 MCP방법 수행 시에 상관계수가 0.7 이상인 참조 데이터를 활용하여 MCP를 수행하는 것을 추천한다 [22].

Table 14 Quality of reference data [22]

Correlation	Quality
0.90 ~ 1.00	Very good
0.80 ~ 0.90	Good
0.70 ~ 0.80	Moderate
0.60 ~ 0.70	Poor
0.50 ~ 0.60	Very poor

Fig. 24는 기상탑과 4개의 참조 데이터인 ERA-5 데이터 E#1, E#2, E#3, E#4 간의 상관관계를 보여준다. 기상탑과 수집된 모든 참조 데이터의 상관계수는 0.8 이상으로 산출되어 상관관계가 높음을 알 수 있다. 기상탑 데이터에 대해 E#4는 상관계수가 0.86로 모든 참조 데이터 중 가장 높은 상관관계를 갖는 것으로 확인되었으며 E#1, #2의 상관계수는 모두 0.82, E#3에 대한 상관계수는 0.81로 산출되었다.

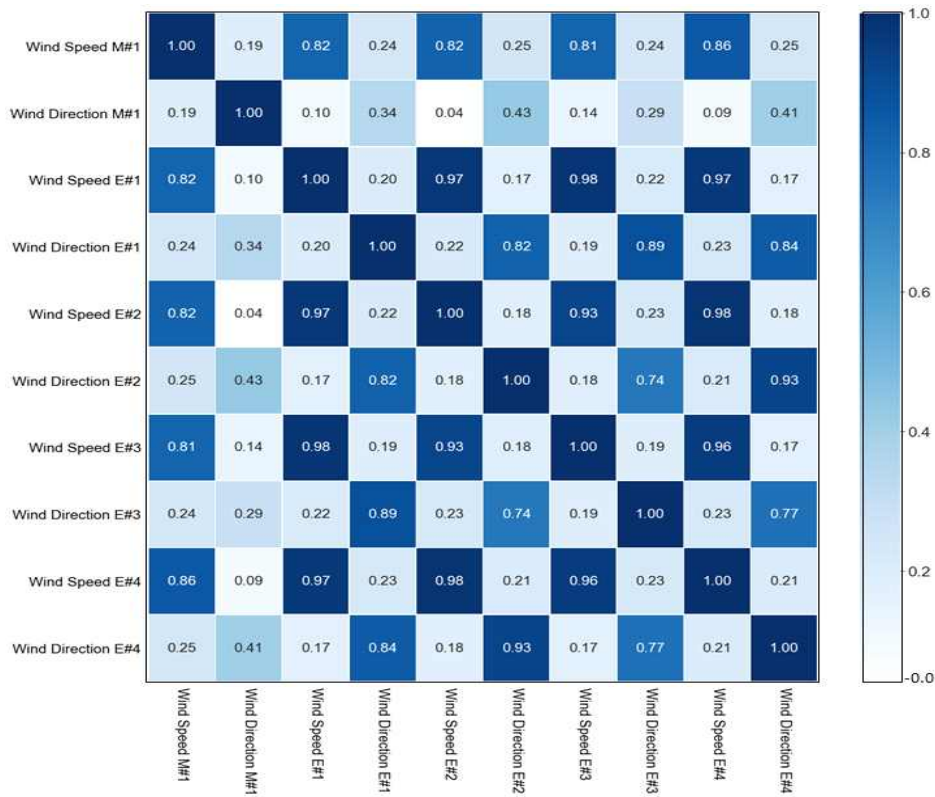


Fig. 24 Result of correlation analysis

WindPRO에서 제공하는 MCP모듈은 기상탑 데이터와 수집된 참조 데이터 중 상관관계수가 가장 높은 1개의 데이터를 선정하여 MCP를 수행한다. 그러므로 참조 데이터 중 상관관계가 가장 높았던 E#4를 참조 데이터로 활용하여 MCP를 수행하였다. 본 연구의 기계학습 방법은 기존의 MCP방법과 다르게 상관관계가 높은 4개의 참조 데이터의 다양한 변동성을 적용하기 위해 모든 참조 데이터를 활용하여 MCP를 수행하였다.

3.3 MCP를 위한 기계학습 데이터 셋

Table 15는 기계학습을 활용하여 MCP를 수행하기 위한 학습 데이터와 시험 데이터를 보여준다. 학습 데이터는 2015년 7월 1일부터 2016년 6월 30일까지 1년의 데이터를 사용하였으며 MCP 수행 결과를 비교·검증하기 위한 시험 데이터는 2013년 10월 1일부터 2015년 1월 31일까지의 데이터 기간을 활용하였다.

Table 15 Train and test data set periods for MCP

Collected period	Train data	Test data
2013.08.12. ~ 2016.06.30.	2015.07.01. ~ 2016.06.30.	2013.10.01. ~ 2015.01.31.

Fig. 25는 기계학습 방법을 활용한 MCP 수행에 따른 입·출력 데이터 및 모델을 나타낸다. WindPRO의 Regression, Matrix, Neural Network 방법과 동일하게 입력 변수로 ERA-5 데이터를 투입하면 기계학습 모델을 통해 MCP 결과가 출력된다.

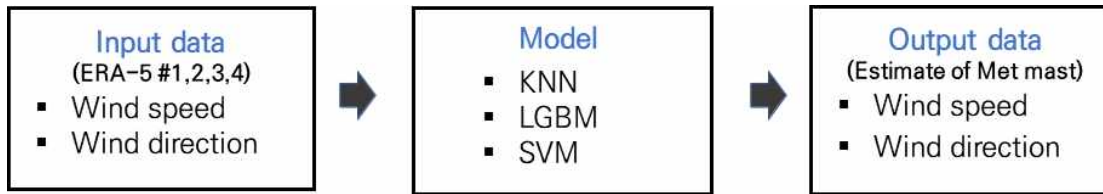


Fig. 25 Progress of machine learning for MCP

3.4 기계학습을 활용한 MCP 결과

WindPRO에서 제공하는 모델과 기계학습 모델의 학습 정확도 확인 방법을 통일하기 위해 기존의 기계학습의 검증 과정인 학습데이터의 일부를 랜덤으로 추출하여 교차 검증을 수행하는 것이 아닌 학습에 따른 검증 방법인 학습에 사용된 데이터 전체를 활용한 자기 검증(Self validation)방법[25]로 학습 정확도를 산출하였다. Table 16은 WindPRO에서 지원하는 MCP 모델과 기계학습 모델의 데이터 학습에 따른 자기 검증을 수행한 결과이다. 모든 모델 중 실제 기상탑의 평균 풍속과 가장 큰 차이를 보이는 모델은 Matrix 모델이고 낮은 차이를 보이는 모델은 LGBM과 SVM, Regression 모델로 확인되었다.

Table 16 Self prediction result after MCP application

Items	Models	Wind speed [m/s]	RMSE [m/s]	R ² [-]
Met mast	-	7.61	-	-
Machine learning	KNN	7.65	0.14	0.99
	LGBM	7.60	0.10	0.99
	SVM	7.62	0.30	0.95
WindPRO	Neural Network	7.56	0.20	0.98
	Matrix	7.53	0.20	0.98
	Regression	7.61	0.19	0.98

모델의 학습 정확도 확인을 위해 RMSE와 R²을 활용하였다. 기계학습 모델중 학습 정확도가 높은 모델은 기계학습의 LGBM으로 RMSE는 0.10m/s, R²은 0.99로 확인되어 모든 모델 중에 자기 검증에 따른 학습 정확도 결과가 높았다. 기계 학습 모델 중 학습 정확도가 낮은 모델은 SVM으로 RMSE와 R²은 각각 0.30m/s, 0.95로 확인되어 3개의 기계학습 모델 중에 자기 검증에 따른 학습 정확도가 가장 낮았다. WindPRO에서 제공하는 Regression방법으로 학습을 수행한 경우 RMSE와 R²은 각각 0.19m/s, 0.98로 산출되어 WindPRO에서 제공하는 모델 중 가장 학습 정확도가 높았다. Table 17은 WindPRO의 MCP 모델과 기계학습 모델을 활용하여 2013년 10월 1일부터 2015년 1월 31일까지의 데이터 기간에 대해 MCP를 수행한 결과를 보여준다. 기계학습의 KNN은 실제 기상탑의 풍속에 근접한 값을 추정하였고 그 차이는 0.01m/s였다. 예측된 풍속과 기상탑의 풍속의 차이가 가장 큰 모델은 SVM으로 0.07m/s 나타났는데 이는 학습 정확도는 높지만 예측 정확도가 낮은 결과를 보이는 과적합(Over fitting)으로 인한 결과로 판단된다.

Table 17 Comparison between machine learning and WindPRO wind speed prediction

Items	Models	Wind speed	RMSE	R ²
		[m/s]	[m/s]	[-]
Met mast	-	8.14	-	-
Machine learning	KNN	8.13	0.15	0.98
	LGBM	8.16	0.12	0.99
	SVM	8.21	0.34	0.94
WindPRO	Neural Network	8.08	0.20	0.98
	Matrix	8.18	0.23	0.97
	Regression	8.05	0.21	0.97

RMSE와 R²을 활용하여 예측 정확도 평가를 수행하였으며 WindPRO에서 제공하는 모델과 기계학습의 모델 중 LGBM은 RMSE는 0.12m/s로 산출되었고 R²은 0.99로 확인되어 모든 모델 중에 성능이 가장 높았다. WindPRO에서 제공하는 모델 중 Regression방법은 RMSE와 R²이 각각 0.21m/s, 0.97로 산출되어 모든 WindPRO에서 제공하는 모델 중 가장 성능이 높았다. WindPRO의 방법과 기계학습을 활용한 MCP 수행 결과의 차이를 보이는 이유는 기상탑과 참조 데이터 간의 상관성이 기간에 따라 변화할 수 있는데 이러한 변동성을 모델의 학습에 반영하는지 여부에 따라 결과가 달라질 수 있다고 생각된다. Table 18은 학습 및 시험 데이터 기간에 대한 기상탑과 참조 데이터 간의 상관관계를 분석한 결과를 보여준다.

Table 18 Correlation coefficients of train and test data at every ERA-5 data positions

Correlation		E#1	E#2	E#3	E#4
Met Mast	Train data	0.82	0.82	0.81	0.86
	Test data	0.79	0.68	0.80	0.71

학습 데이터 기간인 2015년 7월 1일부터 2016년 6월 30일까지의 기상탑 데이터

와 상관성이 높은 데이터는 E#4이고 가장 낮은 데이터는 E#3이다. 하지만 시험 데이터의 기간인 2013년 10월 1일부터 2015년 1월 31일까지의 기상탑에 대한 상관성이 높은 데이터는 E#3이고 상관관계가 가장 낮은 데이터는 E#2이다. 그러므로 WindPRO의 전통적인 방법에서는 1가지 참조 데이터를 활용하였으므로 이러한 변동을 반영하지 못하여 기계학습 모델과의 결과 차이가 나타난 것으로 판단된다. 또한 WindPRO의 모델 중 Neural Network은 기계학습 방법 중 하나이지만 WindPRO에서 제공하는 Neural Network의 경우 단순한 신경망으로 구성되어 있으며[25] WindPRO 소프트웨어에서 분석자가 다양한 파라미터를 조작할 수 없으므로 기계학습의 LGBM과 KNN보다 낮은 성능을 보인 것으로 판단된다.

3.5 분석 결과 및 고찰

본 연구는 기계학습 방법을 활용하여 MCP를 수행한 결과와 WindPRO에서 제공하는 MCP방법의 결과를 비교·분석하였으며 그 결과는 아래와 같다.

1) MCP 수행을 위해 대정해상 기상탑 주변의 4개의 참조 데이터에 대해 상관관계 분석을 수행하였고 E#4는 상관계수가 0.86으로 모든 참조 데이터 중 가장 높은 상관관계를 갖는 것으로 확인되었으며 E#1, E#2의 상관계수는 모두 0.82로 나타났으며 E#3에 대한 상관계수는 0.81로 확인되었다.

2) WindPRO에서 제공하는 MCP모듈은 기상탑 데이터 1개와 수집된 참조 데이터 중 상관계수가 가장 높은 1개의 데이터를 선정하여 MCP를 수행한다. 그러므로 WindPRO에서 제공하는 모델을 활용한 MCP 수행에 참조 데이터 중 상관관계가 가장 높았던 E#4를 참조 데이터로 활용하여 MCP를 수행하였다. 본 연구의 기계학습 방법에는 기존의 전통적인 MCP방법과 다르게 상관관계가 높은 4개의 참조 데이터의 다양한 변동성을 반영하기 위해 모든 참조 데이터를 활용하여 MCP를 수행하였다.

3) 기계학습의 모델 중 LGBM은 RMSE는 0.12m/s로 산출되었고 R^2 은 0.99로 확인되어 모든 모델 중에 성능이 가장 높았다. WindPRO에서 제공하는 모델중 Regression방법은 RMSE와 R^2 이 각각 0.21m/s, 0.97로 산출되어 WindPRO에서 제공하는 모든 모델 중 가장 성능이 높았다.

4) 기상탑 데이터와 모든 참조 데이터는 시간에 따라 서로의 상관성이 변화하는데 1개의 참조 데이터를 활용하는 전통적인 WindPRO의 MCP방법은 이러한 상관성 변동에 취약한 것으로 나타났으며, 다양한 참조 데이터를 활용한 기계학습 방법은 이러한 변동성을 반영함으로써 높은 성능을 보인 것으로 판단된다.

3절에서는 MCP 수행에 기계학습 방법을 적용하였고 다양한 바람의 변동성 적용을 위해 참조 데이터를 활용하여 분석을 수행하였다. 추후 연구에서는 바람뿐만 아니라 기압, 기온 등의 참조 데이터를 추가함으로써 더욱 다양한 변동성을 고려하여 정밀한 MCP를 수행하는 연구를 할 수 있을 것으로 기대된다.

4. 전력계통한계가격(SMP) 장기 추세 분석

바람자원평가지 AEP와 CF를 산출하고 경제성 분석을 수행한다. 이때 경제성 분석을 위해 SMP, REC, CAPEX, OPEX 등의 다양한 변수가 고려된다. 풍력발전기의 수명이 20년 이상임을 고려하여 이러한 변수들의 장기 추정값을 활용하여 경제성을 분석을 수행해야 한다. 이 변수들 중 SMP의 다양한 변동성은 발전사업자의 직접적인 수익에 영향을 주므로 신뢰성 있는 장기간의 SMP 추정은 중요하다. 하지만 경제성 분석을 수행하는 과정에서 고정 SMP나 단순회귀 분석 등을 활용하여 분석하는 경우 SMP의 장기적인 추세의 변동을 반영하지 못하는 한계가 있다. SMP를 결정하는 요인에는 기온, 계절, 환율, 석유 가격, LNG 가격, 석탄 가격 등이 있다 [27]. 이러한 다양한 요인들의 변동성으로 인해 SMP를 추정하는데 어려움이 있다. 그러므로 신뢰성 있는 장기 SMP 추정을 위해 변동성 반영과 결정 요인 선정이 중요하다. 본 연구에서는 기계학습을 활용하여 신뢰성 있는 SMP의 장기 추세 방법을 제공하여 발전사업자의 경제성 분석을 수행하는 경우 다양한 판단을 할 수 있도록 도움을 주고자 한다. 이를 위하여 2010년부터 2019년까지의 SMP를 포함한 다양한 데이터와 세계은행(World bank)에서 발간한 2019년 보고서[28]을 기반으로 육지 SMP의 장기 추세 분석을 수행하였다. Fig. 26은 SMP의 장기 추세 분석을 위한 순서도를 보여주며 육지 SMP의 장기 추세 분석을 위한 절차는 다음과 같다. 첫째, 육지 SMP와 관련된 데이터를 수집하고 상관계수를 활용하여 상관성 분석을 수행한다. 둘째, 상관성 분석 결과를 통해 SMP와 상관관계가 높은 데이터를 입력 변수로 선정하여 기계학습을 수행한다. 넷째, 기계학습을 수행하고 예측 결과가 높은 모델을 활용하여 장기 추세 분석을 수행한다.

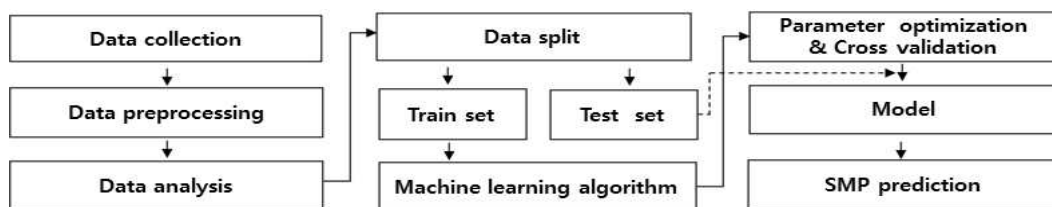


Fig. 26 Work flow for forecast of long term trend of SMP

4.1 SMP 장기 추세 분석을 위한 분석대상 선정

2010년 이전에는 전국 통합으로 SMP가 집계되다가 2010년부터 지역별 가격제도의 도입으로 인해 제주 SMP(Jeju SMP)와 육지 SMP(Main land SMP)로 분리되어 집계되기 시작하였다 [29]. 제주지역의 경우 신재생 에너지 발전을 제외하면 화력발전과 LNG 발전에 의존하므로 제주 SMP는 육지 SMP보다 높은 가격과 변동성을 갖는다 [30]. 따라서 변동성이 낮은 육지 SMP를 활용하여 SMP 장기 추세 분석을 수행하였다.

선행연구에서는 주말과 휴일의 큰 변동성을 필터링(Data filtering)하여 SMP의 예측 정확도를 향상시켰다 [15,16,17]. 따라서 이러한 필터링 방법이 적합하다고 판단되어 기계학습을 수행하기 전에 주말과 휴일을 제거하는 필터링을 수행하였다. Fig. 27은 육지 SMP의 전체 데이터와 주말과 휴일을 제거한 시계열 데이터를 보여주며 필터링을 수행했을 때 이분산성이 제거된 것을 확인할 수 있다.

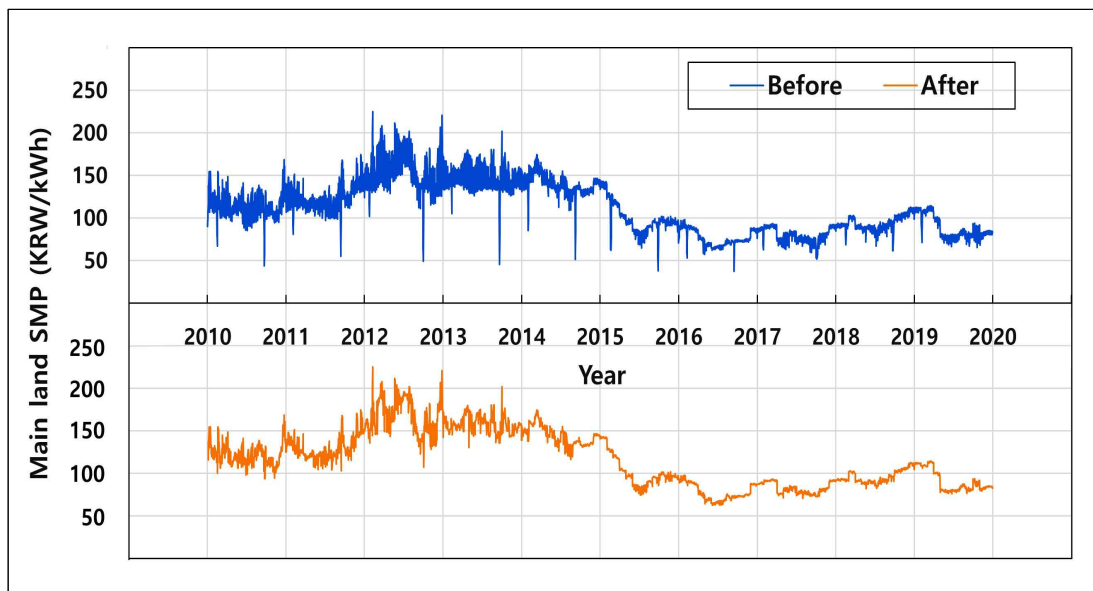


Fig. 27 SMP variation with yearly before and after data filtering

육지 SMP의 시계열 데이터는 2015년을 기준으로 전·후의 변동성이 다르게 나타난다 [31]. 2015년 이전에는 제한된 발전공급능력과 전원구성의 침투부하 발전

원인 LNG와 유류에 대한 높은 가격 변동성과 발전 비중이 높아지는 등의 영향으로 인해 이러한 변동성이 나타난 것으로 판단되며 2015년 이후부터 발전공급능력이 개선됨에 따라 변동성이 줄어들었다.

Fig. 28은 2010년부터 2019년까지의 발전원료별 SMP 결정횟수를 나타낸다. 전력거래소에서는 전력수요에 대해 전날 익일의 예측을 수행하고 발전사업자가 예측된 전력수요에 대해 공급 용량 입찰에 참여하게 된다. 또한 공급과 수요가 일치하는 시점까지 저렴한 발전원부터 값 비싼 발전원이 투입되어 공급과 수요가 일치하는 시점에서 SMP가 결정된다. 1년이 8,760시간임을 감안하면 LNG는 연평균 7,680회로 가장 많이 SMP 결정하는 것으로 나타났다. 기저 부하 발전원인 원자력의 경우 SMP 결정횟수는 0회로 나타났다.

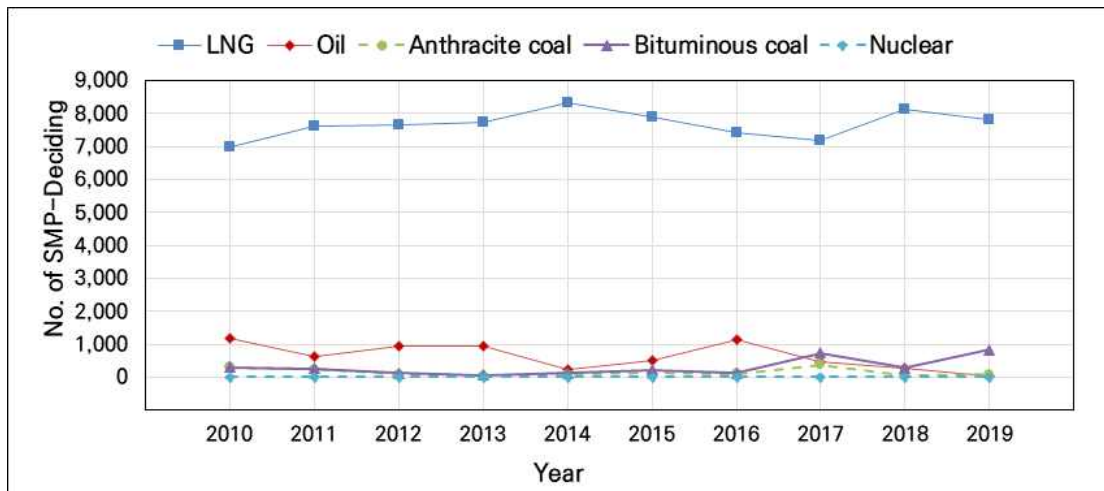


Fig. 28 Number of yearly SMP deciding

Table 19는 본 연구를 위해 육지 SMP와 발전원별 원료 가격 데이터의 기간과 제공주기를 보여준다. 일평균 육지 SMP와 전력 공급과 수요의 변동을 나타내는 일평균 공급용량(Capacity), 최대전력(Max power), 공급예비력(Supply reserve)의 데이터를 수집하였다. 또한 직접적인 LNG 발전에 사용되는 발전용 LNG가격 데이터와 국제원유시장에서의 가격을 대표하는 유종 중 하나인 서부 텍사스 중질유(West Texas intermediate crude oil, WTI) 선물가격, 유연탄 가격 중 하나인 FOB Kalimantan 가격 데이터를 공공데이터 포털[32]를 통하여 수집하였다. 주기가

다른 데이터를 분석에 활용하는 경우 일정한 값을 활용하여 주기가 같아지도록 변환하여 사용하였다. 발전원 데이터의 경우 발전원에 대한 수입 대금 정산 시점으로 인하여 육지 SMP와 선·후행 시차가 발생하는데 일관성을 유지하기 위하여 선행, 후행 지수는 반영하지 않았다.

Table 19 Frequency and period of collected data set [32]

Items	SMP	Supply reserve	Capacity	Max power	LNG	WTI	FOB Kalimantan
Frequency	Daily	Daily	Daily	Daily	Monthly	Daily	Weekly
Period	2010 ~ 2019 (10 years)						

4.2 SMP 장기 추세 분석을 위한 투입 변수 선정

Fig. 29는 변수 선정을 위하여 SMP와 다양한 변수에 대한 피어슨 상관관계수 분석 결과를 보여준다. 실제 발전에 사용되는 데이터인 발전용 LNG가격은 상관관계수가 0.91로 다양한 변수 중 상관관계가 가장 높았다. 또한 WTI 선물가격은 국제 원유시장에서의 가격을 대표하는 유종 중 하나로 0.80으로 상관관계수가 산출되어 발전용 LNG 다음으로 상관관계가 높았다. 따라서 다양한 변수 중 상관관계가 높은 발전용 LNG가격과 WTI선물가격을 투입 변수로 선정하여 SMP 장기 추세 분석을 수행하였다.



Fig. 29 Correlation analysis of factors for SMP deciding

Fig. 30은 기계학습 모델과 입·출력 데이터를 보여준다. 발전용 LNG가격과 WTI 선물가격을 입력 변수로 활용하여 육지 SMP의 예측값을 출력하도록 기계학

습을 수행하였다.

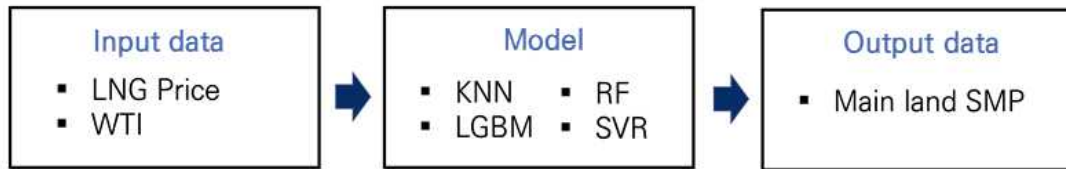


Fig. 30 Progress of machine learning

4.3 SMP의 장기 추세 분석을 위한 기계학습 데이터 셋

Table 20은 기계학습에 활용된 학습 및 시험 데이터 기간을 나타낸다. 학습 데이터로 2010년 1월부터 2018년 12월까지의 데이터를 선정하여 기계학습을 수행하였다. 기계학습 모델의 실제 예측 정확도 확인을 위해 2019년 1월 1일부터 2019년 12월 31일까지의 데이터를 시험 데이터로 선정하여 예측 정확도를 확인하였다. 또한 기계학습의 정확도 평가 지표로 RMSE와 R²값을 활용하였다.

Table 20 Train and test periods for forecasting of SMP

Collected Period	Train data	Test data
2010.1.1. ~ 2019.12.31.	2010.1.1. ~ 2018.12.31.	2019.1.1. ~ 2019.12.31.

4.4 SMP의 장기 추세 분석을 위한 기계학습 알고리즘의 학습 결과

Table 21은 학습 데이터를 활용한 기계학습 수행에 따른 모델의 학습 정확도 확인을 위한 교차검증 결과를 보여준다. 3가지 모델 중 KNN은 RMSE는 4.187KRW/kWh로 산출되었고 R²은 0.977로 학습 정확도가 높았다. SVM의 RMSE와 R²은 각각 7.136KRW/kWh, 0.932로 가장 낮았다.

Table 21 Train result of forecasting SMP model

Items	KNN	LGBM	SVM
RMSE[kW/h]	4.187	5.342	7.136
R ² [-]	0.977	0.962	0.932

Fig. 31은 학습된 3개의 기계학습 모델을 활용하여 2019년의 육지 SMP를 예측한 결과를 보여준다. 기계학습에 활용된 KNN, LGBM, SVM 모두 2019년 육지 SMP의 추세를 비교적 잘 예측하는 것으로 보인다. 2010년부터 2018년까지의 데이터를 활용하여 학습된 3개의 모델에 대한 예측 정확도 확인을 위해 RMSE와 R²을 활용하여 정확도를 평가하였다. LGBM의 RMSE는 2.770KRW/kWh, R²은 0.953으로 산출되어 3개의 모델 중 예측정확도가 높았다. 기계학습 모델이 산출한 결과가 그림과 같은 패턴을 나타내는 이유는 기계학습 모델의 특성과 Table 19의 데이터의 제공 주기로 인한 결과로 판단된다. 즉, LGBM과 SVM은 상관성이 높은 발전용 LNG가격의 가중치가 반영되어 월별로 제공되는 발전용 LNG 가격의 데이터 특성이 크게 반영되었으며, KNN의 특성은 가까운 거리의 데이터를 참조하므로 발전용 LNG가격과 WTI 선물가격의 데이터 제공 주기에 따른 특성이 추가적으로 반영된 것으로 판단된다. 육지 SMP의 장기 추세 분석을 위해 예측 정확도가 높았던 LGBM을 활용하였다.

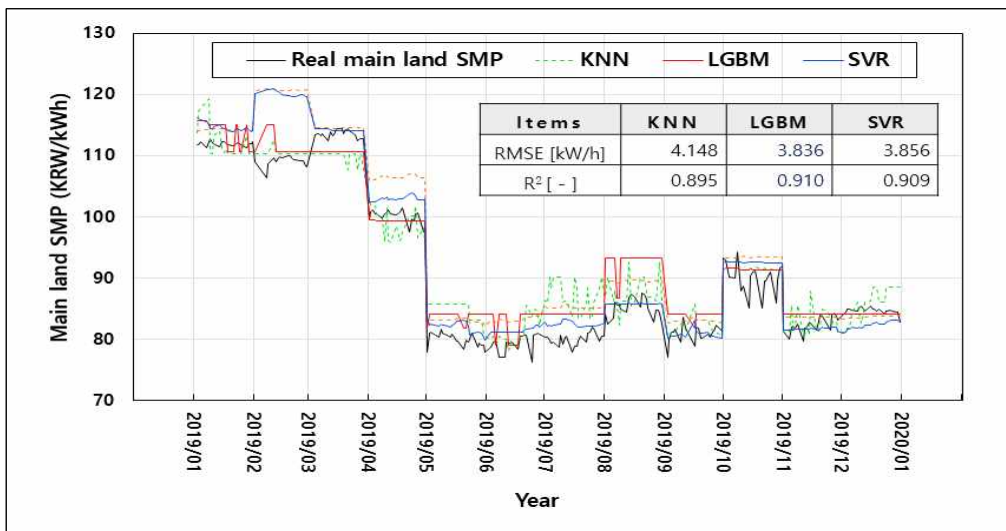


Fig. 31 SMP forecasting for year 2019

4.5 SMP의 장기 추세 분석 결과

발전용 LNG가격과 WTI 선물가격을 입력변수로 육지 SMP 장기 추세 분석 및 예측하기 위해서는 두 변수의 장기 전망치가 필요하다. 세계은행에서는 매년 장기 전망치를 보고[28]하는데 본 연구에서는 2019년 세계은행에서 보고한 2020년부터 2030년까지의 브렌트(Brent)유, 두바이(Dubai)유, WTI의 평균가격인 국제원유(World Crude Oil)의 가격과 일본 LNG가격의 장기 전망치를 다양한 분석을 통해 각각 WTI선물가격과 한국 발전용 LNG 가격을 추정값을 도출하여 육지 SMP의 장기 추세 분석에 활용하였다. Fig. 32 1999년부터 2018년까지 국제 원유와 WTI 선물가격에 대한 선형회귀 분석 결과를 나타낸다. 국제 원유와 WTI 선물가격의 결정계수는 0.9868로 분석되었으며, 회귀식을 활용하여 2020년부터 2030년까지의 WTI 선물가격을 추정값을 산출하였다.

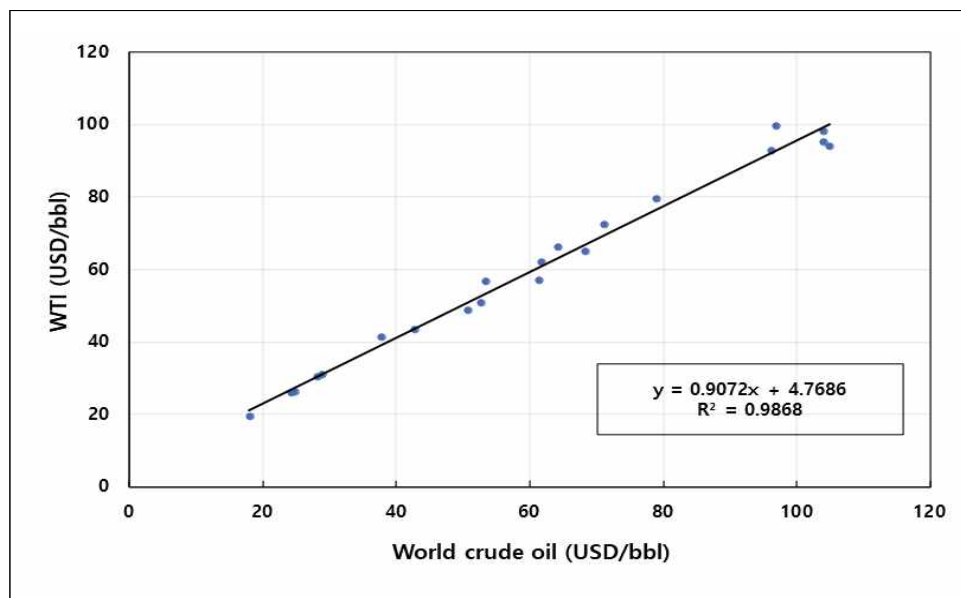


Fig. 32 Linear regression analysis of world crude oil and WTI

일본의 원유수입가격(Japan Crude Cocktail)은 한국의 중장기의 수입되는 LNG 가격의 기준이 되는 경우가 많으므로 일본 LNG 가격과 한국의 발전용 LNG가격에 대해 분석하였다. 일본의 LNG 가격과 한국의 발전용 LNG 가격의 단위는 각

각 USD/mmbtu와 KRW/GJ를 사용하므로 단위를 통일하기 위하여 일본의 LNG 가격 단위를 USD/mmbtu에서 KRW/GJ로 변환하였으며 이 과정에서 환율(KRW/USD)은 연간 환율을 활용하여 적용하여 단위를 변환하였다.

Fig. 33은 2010년부터 2019년까지의 한국의 발전용 LNG 가격 및 일본의 LNG 가격의 시점 보정 전·후에 대한 추세를 보여준다. 원유수입가격(Japan Crudeoil Cocktail)은 동북아시아의 LNG 수입 가격에 약 4개월 이상 선행한다고 알려져 있다. 가격 선행 이외에도 자연재해, 예상치 못한 수요 변화, 환율의 적용시점 등 외부의 다양한 요인을 고려했을 때 일본 LNG 가격은 한국의 발전용 LNG 가격에 대해 약 1년정도 선행하는 것으로 분석되었다. 그러므로 일본 LNG 가격에 대한 시점 보정을 수행하였고 일본 LNG 가격의 시점 보정 전·후와 한국 발전용 LNG 가격에 대한 상관성 분석을 수행하였다. 한국 발전용 LNG가격과 시점 보정을 수행하기 전 일본 LNG 가격의 상관계수는 0.53으로 산출되었다. 그러나 시점 보정을 수행한 후 일본 LNG 가격의 상관계수는 0.96으로 약 1년의 시점 보정을 수행하였을 때 상관관계가 크게 개선되는 것을 확인할 수 있었다. 따라서 보정된 일본 LNG 가격에 대해 연간환율을 적용하여 한국 발전용 LNG 가격을 추정하였다. 연간환율은 2020년의 연간환율인 1,179.90원을 적용하였고, 2021년 2월 2일 매매 기준 환율은 1,101.64원으로 이 환율을 2021년부터 2030년까지의 연간환율을 일괄 적용하였다. 추정된 한국 발전용 LNG가격과 WTI 선물가격을 활용하여 육지 SMP 장기 추세 분석을 수행하였다. 전망치에 대한 오차 발생할 수 있는데 이는 유가 변동, 환율의 변동 등의 다양한 요인으로 인해 발생할 수 있을 것이라고 생각된다.

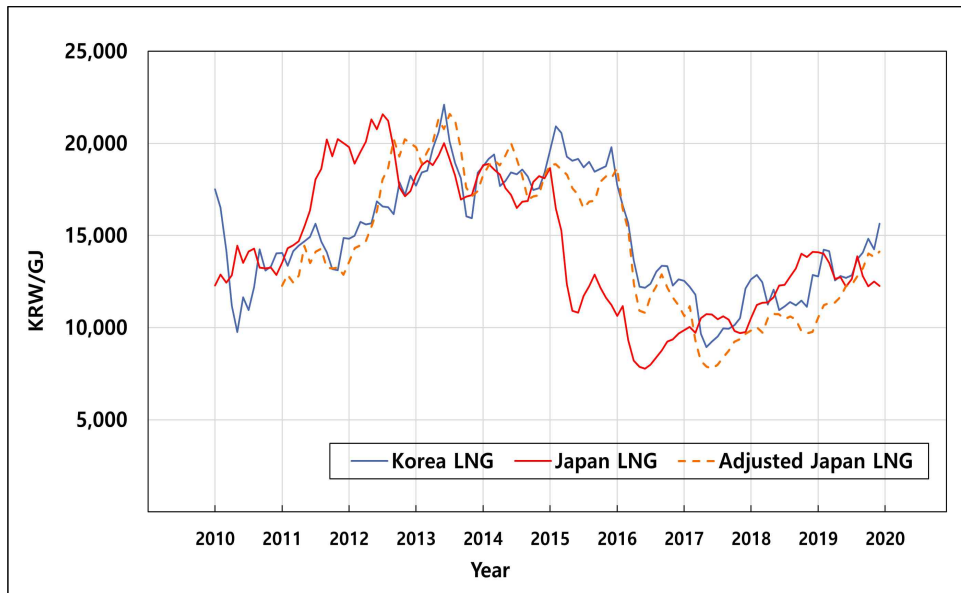


Fig. 33 Adjustment of lead-lag correlation between Korea and Japan LNG prices

Fig. 34는 육지 SMP의 장기 추세 분석을 수행을 위해 추정된 장기 전망치인 한국의 발전용 LNG 가격과 WTI 선물가격을 입력변수로 하여 기계학습 모델인 LGBM에 투입하여 2020년부터 2030년까지 산출된 결과를 보여준다. 육지 SMP는 장기적으로 2020년부터 2030년까지 감소하는 경향을 갖다가 평균 72KRW/kWh 선에서 유지하는 것으로 분석되었다. 한국에너지경제연구원(KEEI)에서 발간된 2015년 보고서[33]의 장기 SMP 예측 시나리오인 S1과 S2의 2016년부터 2024년에 대한 예측은 SMP가 장기적으로 감소하는 경향을 갖는 것으로 전망되었으며 이는 기계학습을 활용한 육지 SMP의 장기 추세 분석 결과와 비슷하였다. COVID-19의 상황과 유가의 큰 변동 등으로 인해 2020년에는 육지 SMP의 실제 값과 전망치의 오차는 약 26KRW/kWh로 크게 발생하였다. 2020년에 비해 2021년 1월부터 5월까지는 유가 변동과 COVID-19로 인한 상황이 조금 안정되었는데 이 기간의 실제 육지 SMP의 평균은 약 78KRW/kWh이었고, 기계학습을 통해 전망된 값은 약 74KRW/kWh로 산출되어 약 4KRW/kWh정도 차이가 있었다. SMP의 장기 추세 분석의 결과는 추후 발전용 LNG 가격과 WTI 선물가격의 변동에 의해 변동될 수 있을 것으로 판단된다.

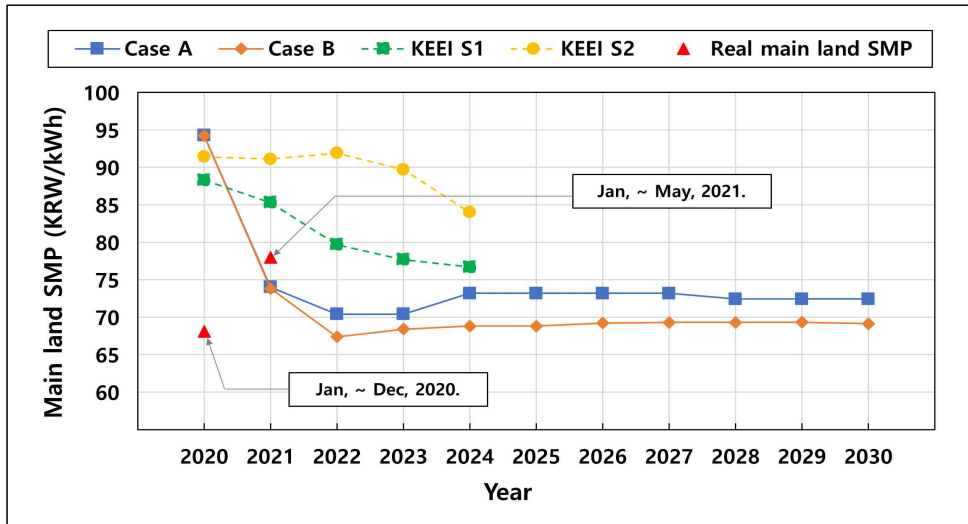


Fig. 34 Yearly main land SMP forecast

4.6 분석 결과 및 고찰

본 연구는 기계학습 방법에 다양한 데이터를 활용하여 육지 SMP의 장기 추세 분석을 수행하였으며 그 결과는 아래와 같다.

1) 육지 SMP의 장기 추세 분석에 따른 입력 변수를 선정하기 위하여 육지 SMP, 공급용량, 최대전력, 공급예비력, WTI 선물가격, 발전용 LNG 등 다양한 변수에 대해 상관성 분석을 수행하였다. 발전용 LNG 가격과 WTI 선물가격은 다양한 변수 중 높은 상관관계를 갖는 것으로 확인되어 이를 육지 SMP의 장기 추세 분석에 활용하였다.

2) 기계학습된 모델에 대해 실제 2019년의 육지 SMP를 예측하였으며 예측된 결과에 대해 평가지표인 RMSE와 R^2 을 활용하여 예측 정확도를 확인하였다. 그 결과 예측 정확도는 3개의 기계학습 모델 중 LGBM이 가장 높은 것으로 확인되어 육지 SMP의 장기 추세 분석에 LGBM을 활용하여 분석을 수행하였다.

3) 장기 WTI 선물가격과 한국의 발전용 LNG 가격을 추정하기 위해 세계은행에서 전망한 국제원유 평균 가격과 일본 LNG 가격을 활용하여 추정치를 산출하였다. 산출된 장기 WTI 선물가격과 한국의 발전용 LNG 가격을 활용한 육지 SMP 장기 추세 분석 결과는 일본 LNG 가격과 국제원유 평균치를 활용하여 한국

의 발전용 LNG와 WTI 선물가격을 예측하였고 이를 활용하여 육지 SMP의 장기 추세를 분석하였다. 그 결과, 장기적으로 육지 SMP는 2020년부터 2030년까지 감소하는 경향을 갖다가 평균 72KRW/kWh 선에서 유지하는 것으로 분석되었다.

4절은 경제성 분석시 기계학습을 기반으로 발전사업자의 직접적인 수익에 영향을 주는 변수인 SMP의 장기 추세 예측 방법을 제공하였다. 이러한 결과와 기계학습 방법과 다양한 변수를 활용하여 장기 REC를 예측하는 연구를 융합하여 연구를 수행한다면 장기 가격계약 종류 후 REC 미지상황에서 사업자의 수익을 전망하는 연구와 장기 고정가격 계약제도의 계약금액 전망 등의 연구를 수행할 수 있을 것으로 기대된다.

IV. 결 론

본 연구는 바람자원평가의 수행 절차 중 데이터 품질 검사, 타워체도우 영향에 따른 보정, MCP와 경제성 분석에서 발전사업자의 수익에 영향을 주는 SMP의 장기 추세 분석에 기계학습을 적용하였고 그 결과는 아래와 같다.

1) 상명목장지구 기상탑 데이터를 활용하여 기계학습을 수행한 결과 LGBM이 3개의 모델 중 학습 정확도가 가장 높았으므로 대정해상 기상탑 데이터의 품질 검사에 LGBM을 활용하였다. 그 결과 대정해상 기상탑에 LGBM을 활용한 데이터 품질 검사 결과의 분류 정확도는 93.51%였다. 이러한 오차가 발생하는 원인으로 기계학습 모델은 풍속계가 일시적으로 정상 값을 회복하는 경우 정상 구간으로 분류하였지만 분석자는 측정 오류 구간으로 분류하여 이로 인한 차이로 판단된다.

2) 통계적 방법을 활용하여 타워체도우 영향을 받는 구간을 검출하였고 정상 구간을 학습데이터로 활용하여 기계학습을 수행한 결과 3개의 기계학습 모델 중 학습정확도가 가장 높은 SVM을 활용하여 82.5m A풍속계에 대한 $220^{\circ} \sim 245^{\circ}$ 빈에 대해 보정을 수행하였다. 그 결과 82.5m의 A와 B 풍속계의 RMSE는 보정하기 전 1.009m/s에서 보정한 후 0.2611m/s로 보정되어 이 둘을 비교하였을 때 0.7479m/s 만큼 개선되었다. R^2 은 보정하기 전 0.891에서 보정한 후 0.9926으로 보정되었고 이 둘을 비교하였을 때 0.1016 개선되었다.

3) 기상탑과 참조 데이터를 활용하여 MCP를 수행한 결과 3개의 기계학습 모델 중 가장 성능이 높았던 LGBM은 RMSE와 R^2 은 각각 0.12m/s와 0.99로 산출되었다. WinPRO에서 제공하는 모델 중에는 Neural Network가 RMSE는 0.20으로 산출되었고 R^2 은 0.98로 산출되었다. 이러한 결과로 참조 데이터의 상관성은 기간에 따라 변화하는데 전통적인 WindPRO의 MCP방법은 1개의 참조 데이터를 참조하므로 상관성 변동에 취약한 것으로 나타났다. 기계학습 모델은 4개의 참조 데이터를 참조하여 MCP를 수행하기 때문에 WindPRO의 MCP방법 보다 상관성 변동에 유리한 것으로 확인되어 기계학습 방법이 상용소프트웨어인 WindPRO보다 M

CP 수행 성능이 높았다.

4) 육지 SMP에 대해 발전용 LNG와 WTI 선물가격의 상관계수가 각각 0.91, 0.80으로 상관관계가 높은 것으로 확인되어 이를 기계학습의 입력 변수로 활용하였다. 세계은행에서 전망한 국제원유 평균 가격과 일본 LNG 가격을 활용하여 장기 WTI 선물가격과 한국의 발전용 LNG 가격의 장기 추정치를 산출하였고 이를 활용하여 육지 SMP의 장기 추세를 분석하였다. 그 결과, 장기적으로 육지 SMP는 2020년부터 2023년까지 감소하는 경향을 갖다가 그 이후는 평균 72KRW/kWh 선에서 유지하는 것으로 분석되었다.

본 논문에서는 바람자원평가 및 경제성 분석에 기계학습을 적용하였다. 국내 풍력발전단지 대부분은 설치되고 운영된 기간이 오래되지 않아 장기간의 데이터를 취득하기 어렵기 때문에 빅데이터를 활용하는 기계학습 특성상 한계가 있다. 그러므로 앞으로 국내 풍력발전단지의 장기 데이터가 수집되고 축적됨에 따라 더 많은 풍력산업 분야에 기계학습을 다양하게 적용하여 분석자에게 편의를 제공할 수 있을 것으로 기대된다.

V. 참고 문헌

- [1] Intergovernmental Panel on Climate Change, Global Warming of 1.5 °C Special Report, Switzerland, IPCC, 2018
- [2] 지속가능한 녹색사회 실현을 위한 대한민국 2050 탄소중립 전략, 대한민국, 2020
- [3] Ministry of Trade, Industry and Energy, The 9th Basic Plan for Long-term Electricity Supply and Demand, MOTIE, 2019
- [4] 4차산업혁명위원회, 4차 산업혁명 대응을 위한 기본 정책방향, 대한민국, 2017
- [5] Hur, J. and Park B., The Development of the Short-Term Wind Power Forecasting System using Support Vector Regression, Journal of the Korean Institute of Illuminating and Electrical Installation Engineers 31(9), 104-110, 2017
- [6] Park, Soo., Do, J., Choi, S. and Hur, J., Fault Diagnosis of Drone Using Machine learning, Journal of the Korean Society of Manufacturing Process Engineers 20(9), pp 28-34, 2021
- [7] Choi, J., Son, E., Lee, G., Kang, M., Lee, Jin., Hwang, S. and Park, Sail., Development of Fatigue Load Prediction Algorithm for Wind Turbines, Journal of the Korean Solar Energy Society, Vol.41, No.4, pp.93-105, 2021
- [8] Wind power forecasting based on daily wind speed data using machine learning algorithms, Energy conversion and Management, Volume 198, 111823, 2019
- [9] Arcos Jiménez, A., Gómez Muñoz, C. Q., & García Márquez, F. P. Machine learning for wind turbine blades maintenance management. Energies, Vol11 (1), 13., 2018
- [10] Nezhad, M. M., Heydari, A., Groppi, D., Cumo, F., & Garcia, D. A., Wind source potential assessment using Sentinel 1 satellite and a new forecasting model based on machine learning: A case study Sardinia islands. Rene

wable Energy, 155, 212–224., 2020

- [11] NEGNEVITSKY, Michael; MANDAL, Paras; SRIVASTAVA, Anurag K. Machine learning applications for load, price and wind power prediction in power systems. In: 2009 15th International Conference on Intelligent System Applications to Power Systems. IEEE, pp. 1–6, 2009
- [12] Scher, S., & Molinder, J. Machine learning-based prediction of icing-related wind power production loss. IEEE Access, 7, 129421–129429. 2019
- [13] Velázquez, S., Carta, J. A., & Matías, J. M. Comparison between ANNs and linear MCP algorithms in the long-term estimation of the cost per kWh produced by a wind turbine at a candidate site: A case study in the Canary Islands. Applied energy, 88(11), 3869–3881., 2011
- [14] Shin, D., Baek, S., Lee, Y., and Kang, S., System marginal price time series data forecasting model development. The Korean Society of Mechanical Engineers, Gangwon Land Convention Center, Jeongseon-gun, Gangwon-do, pp. 3172–3176, 2018.
- [15] Kim, J., Kwon, O., Oh, B., Kwak, H., Lee, S., Jang, J., Yang, S., Kim, S. and Lee, D., LSTM based intelligent forecasting power load and SMP, The Korea Institute of Electrical Engineers(KIEE) Autumn Conference 2019, pp. 115–117, 2019.
- [16] Lee, S., Forecasting Volatility of Weekday System Marginal Prices : A Multi-frequency Approach, Korean Energy Economic Review, Vol. 15, No. 2, pp. 89–119, 2016.
- [17] Noh, J. and Cho, H., Forecasting System Marginal Price Using Multilayer Perceptron and Nonlinear Autoregressive exogenous model, Journal of The Korean Society of Mineral and Energy Resources Engineers, Vol. 57, No. 6, pp. 585–592, 2020.
- [18] Kim, B., and Ko K., Offshore wind resource assessment and wind turbine class determination off the coast of Daejeong, Jeju Island, Jeju National University, 2018.

- [18] Zhang, H.; Berg, A.C.; Maire, M.; Malik, J. SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17 - 22 June 2006; Volume 2, pp. 2126 - 2136.
- [19] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, pp3146-3154., 2017
- [20] S. Raschka and V. Mirjalili, *Python Machine learning*, Birmingham, U.K.:Packt, 2017.
- [21] https://www.bogotobogo.com/python/scikit-learn/scikit_machine_learning_Support_Vector_Machines_SVM_2.php
- [22] Measnet, *Evaluation of Site-specific Wind Conditions Version 2*, 2016
- [23] M. C. Brower, *Wind Resource Assessment: A Practical Guide to Developing a Wind Project*, Wiley
- [24] Kim, T., Rhee, H., Kim, H., *Analysis of the Effect of Met Tower Shadow using Computational Fluid Dynamics*, Korean Society for New and Renewable Energy, Gyeongju Hotel Hyundai, Gyeongju-si, Gyeongsangbuk-do, pp 35
- [25] WindPRO 3.5 user manual, chapter 11, MCP: Measure Correlate - Predit, <http://help.emd.dk/knowledgebase>
- [26] Energi- og MiljøData, EMD International A/S 2021:WindPRO 3.5 HandBook, Denmark : EMD .2021.
- [27] Park, M., Moon, Y. and Park, J., An Analysis on the Causal Relation Among SMP, Base-Load Share, LNG Import Price, and Exchange Rate, *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, Vol. 28, No. 7, pp.97-105, 2014.
- [28] World Bank, *Commodity Markets Outlook*, October 2019.
- [29] Kim Yeong Hwan, Kim Se-Ho, *Increasing effect of the Wind Power Limi*

- t Using Energy Storage System in Jeju Korea, Journal of the Korean Solar Energy Society, Vol.34, No.1, pp. 81-90, 2014
- [30] Kyu-Ho Park, Eun-Bin Oh, Su-Jin Lee, Su-Wan Kim, Ankhzaya Baatarbileg and Gae-Myoung Lee, Analysis of Sale Price of Electrical Energy Generated by a Large-scale Photovoltaic Plant in Jeju Island, Journal of the Korean Institute of Illuminating and Electrical Installation Engineers 31(5): pp.75-82, 2017.
- [31] Shin, D., and Kim, J The Identification of Structural Shocks and Analysis of System Marginal Price Volatility in Korean Electricity Power Market, Korean, Association of Applied Economics, Vol.17, No.2, pp. 121-166, 2015
- [32] <https://www.data.go.kr>, Public data Portal, May, 2021
- [33] Korea Energy Economics Institute, Development and application of Forecasting method for Renewable Energy Certificate price, 2015.

감사의 글

2년의 석사과정동안 많은 것을 배울 수 있었고 자그마한 결실을 맺을 수 있도록 도움을 주신 많은 분들에게 이 자리를 빌려 감사의 인사를 드리고자 합니다.

먼저 아무것도 모르고 풍력에 대한 열정만 가득한 직장인이었던 저에게 풍력이라는 학문을 공부할 수 있는 기회를 열어주시고 제 연구의 방향을 정해주시고 항상 믿어주시고 부족한 저를 이 자리까지 이끌어주신 지도교수 고경남 교수님께 진심으로 은혜에 머리 숙여 진심으로 감사드립니다.

풍력에 대한 이론뿐만 아니라 다양한 정책, 추세 등을 통해 다양한 시각으로 연구에 접근할 수 있도록 도와주시고 따뜻한 격려와 연구에 대한 코멘트를 해주신 김범석 교수님 감사합니다.

또한 언제든 찾아가도 항상 격려와 응원을 해주신 수학과 학부 시절 지도교수이신 유상욱 교수님께도 감사드립니다. 또한 수학전공 관련 서적 및 자료를 아낌없이 제공해주신 강경태 교수님께도 감사드립니다.

더불어 언제든 주말과 낮·밤 가리지 않고 도움을 주시고 면밀하게 도움을 준 우리 연구실의 건우형, 운다르마, 병택이에게 정말 감사합니다. 풍력이라는 분야를 전공하고 있다는 사람이 있다는 것이 얼마나 큰 힘이 되는지 깨닫게 해준 단지설계 선배 및 그 외 연구실의 종화형, 대영이형, 희전이형, 성건, 근석, 효정이 누나, 인섭, 동은이에게 같이 공부하고 연구할 수 있어 행복했고 너무 감사합니다. 이외에도 도움을 준 많은 분들에게 고마움을 전하지 못해 죄송하고 감사합니다.

마지막으로 오늘이 있기까지 저를 믿고 응원해주신 존경하는 아버지, 어머니, 동생에게 감사의 마음을 바칩니다.