



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

크리깅 보간법을 활용한 풍력
발전량 예측에 관한 연구

제주대학교대학원

에너지응용시스템학부 기계공학전공

김 성 지

2023년 2월

크리깅 보간법을 활용한 풍력 발전량 예측에 관한 연구

지도교수 김 남 진

김 성 지

이 논문을 기계공학 석사학위 논문으로 제출함

2022년 12월

김성지의 기계공학 석사학위 논문을 인준함

심사위원장 _____ 강 창 남

위 원 _____ 고 진 환

위 원 _____ 박 윤 철



제주대학교 대학원

2022년 12월

A Study on the Prediction of Wind Power Generation
Using Kriging Interpolation Method

Sung-Ji Kim

(Supervised by professor Nam-Jin Kim)

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

2022. 12.

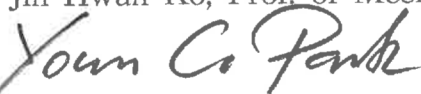
This thesis has been examined and approved.



.....
Thesis director, Chang Nam Kang, Prof. of Mechanical Engineering



.....
Thesis director, Jin Hwan Ko, Prof. of Mechanical Engineering



.....
Thesis director, Youn Cheol Park, Prof. of Mechanical Engineering

2022. 12.

.....
Date

Department of Mechanical Engineering
GRADUATE SCHOOL
JEJU NATIONAL UNIVERSITY

목 차

LIST OF FIGURES	iii
LIST OF TABLES	v
NOMENCLATURE	vi
SUMMARY	vii
I. 서론	1
1-1. 연구의 배경	1
1-2. 연구의 목적	3
II. 이론적 배경	4
2-1. 풍력발전기(Wind Turbine)	4
2-2. 기상기상관측장비	6
2-2-1. 자동기상관측장치(Automatic Weather System)	6
2-2-2. 종간기상관측장치(Automatic Synoptic Observing System)	6
2-3. 공간 보간법(spatial interpolation)	7
2-4. 기계학습(Machine Learning)	12
2-5. 변수 및 모델 평가 지표	15
2-5-1. 피어슨 상관계수(Pearson Correlation Coefficient)	15
2-5-2. 결정계수(Coefficient of Determination, R-Square)	15
2-5-3. 평균 제곱근 오차(Root Mean Square Error)	16

III. 풍력 발전량 예측 모델 개발	17
3-1. 연구 대상 선정 및 분석	18
3-2. 풍력 발전량 예측을 위한 기상정보 추정	23
3-2-1. 연구 대상 인근 기상정보	23
3-2-2. 정규 크리깅을 이용한 기상정보	25
3-3. 예측 모델링을 위한 변수 탐색 및 선정	28
3-3-1. 기상 파생 변수 탐색 및 추가	28
3-3-2. 시계열 파생 변수 탐색 및 추가	30
3-3-3. 최종 입력 변수 결정	31
3-4. 발전량 예측 모델 학습 알고리즘 선정	32
3-4-1. 모델 예측 평가 선정 및 평가	32
3-5. 매개변수 최적화를 통한 예측 모델 고도화	33
IV. 발전량 예측 결과 분석	36
V. 결론	38
참고문헌	40
감사의 글	43

LIST OF FIGURES

Fig. 1. Main components of wind turbines	4
Fig. 2 Types of machine learning	12
Fig. 3. Design and Validation Process for Wind Power prediction Model	17
Fig. 4. Yeonggwang-Baeksu wind power plant layout	18
Fig. 5. Wind turbine(Unison U113 model) power curve	19
Fig. 6. Hourly power generation chart	20
Fig. 7. Monthly power generation chart	20
Fig. 8. Monthly box plot of wind speed values	20
Fig. 9. Power curve before outlier removal	22
Fig. 10. Time frequency chart with generation range of 1 to 1.2 MW with wind speed higher than 9 m/s	22
Fig 11. Time frequency chart with zero power generation in the driving wind speed interval (3-20 m/s)	22
Fig. 12. Power curve after outlier removal	22
Fig. 13. Weather observation equipment closest to the study target	23
Fig. 14. Pearson's correlation heat map for YeomsanAWS(left)dataset and Yenggwanggun ASOS(Right) dataset	24

Fig. 15. Study target nearby weather observation equipment	25
Fig. 16. Pearson's correlation heat map for kriging interpolation dataset	27
Fig. 17. Comparison of ordinary kriging interpolation wind speed (10 m) and nacelle wind speed	28
Fig. 18. Comparison of ordinary kriging interpolation wind speed (100 m) and nacelle wind speed	28
Fig. 19. Wind roses by Season	29
Fig. 18. Time series decomposition chart based on additive model of power generation data	30
Fig. 21. Pearson's correlation heat map for the input variables	31
Fig. 22. Hyperparameter optimization history plot	34
Fig. 23 Hyperparameter importance	35
Fig. 24. Evaluation process of the prediction model based on weather data ..	36
Fig. 25. Chart of hourly power generation prediction result based on weather data of Yeomsan AWS	37
Fig.. 26. Chart of hourly power generation prediction result based on weather data of Yenggwanggun ASOS	37
Fig. 27. Chart of hourly power generation prediction result based on ordinary kriging interpolated weather data	37

LIST OF TABLES

Table 1. Application of wind power generation prediction over Time Range ..	2
Table 2. Feature and formulas of semi-variogram	9
Table 3. Wind turbine(Unison U113 model) data-sheet	19
Table 4. Evaluation of Ordinary kriging prediction value by semi-variogram models	26
Table 5. Pearson correlation coefficient between major variables and power generation	27
Table 6. Performance evaluation of learning models	32
Table 7. Types and definitions of hyperparameters of LGBM	33
Table 8. Hyperparameter optimization range and result values	35
Table 9. Performance evaluation of predictive model based on weather data(AWS, ASOS, kriging interpolated data	37

NOMENCLATURES

- P : Power in Watts (출력)
- ρ : Air Density in kg/m^3 [about $1.225kg/m^3$] (공기밀도)
- A : Rotor Swept Area in m^2 (로터 회전 면적)
- V : Wind Speed in m/s (풍속)
- z_i : Observations at the target point i (목표 지점 i 에서의 관측값)
- x_i : The x-coordinate of the target point i (목표 지점 i 의 x좌표)
- y_i : The y-coordinate of the target point i (목표 지점 i 의 y좌표)
- λ_i : Krigging weights at the target point i (목표 지점 i 에서의 크리깅 가중치)
- γ : Semivariogram (반분산도)
- h : Lag distance (분리거리)
- C_0 : Sill (문턱값)
- a : Range (상관거리)
- L : Lagrange function (라그랑지 목적함수)
- Z_a : Wind speed measurement altitude(풍속 측정 고도)
- U : Wind speed at a certain height (특정 고도에서의 풍속)
- α : Wind shear (풍속할증계수)
- Z_g : Geometric mean height (기하평균높이)
- Z_0 : Surface roughness lengths (지면 거칠기 길이)
- \hat{y}_i : Predicted value (예측값)
- \bar{y}_i : Average value (평균값)
- t : Observation time(관측 시점)
- R : Actual power generation(실제 발전량)
- F : Forecasting power generation(예측 발전량)

SUMMARY

When climate change accelerates due to continuous carbon emissions, the Earth's surface temperature rises 1.5°C compared to pre-industrial times, and the frequency of abnormal weather will be doubled. When 2°C rises, 20% to 30% of the world will be deserted and 54% of species will disappear. Accordingly, it is inevitable to increase the power generation ratio from renewable energy to decline carbon emission. In the matter of renewable power, however, the importance of predicting power generation is increasing because of the nature of renewable power generation that depends on uncontrollable weather environments.

In the case of wind power generation by new and renewable energy, it is the most important to predict wind speed and direction depending on topographical conditions. However, it is difficult to secure accurate weather information in an area where meteorological tower is not installed and operated that require expensive cost.

In this study, a wind power prediction model using machine learning techniques based on weather values of AWS and ASOS, which are ground weather observation equipment near the wind power generator, is developed and evaluated.

In order to select weather data for wind power prediction model, the Pearson correlation coefficient, between the wind power generation amount and the wind speed values by the observed region, was calculated. As a result of the calculation, the values were 0.67 for Yeonggwanggun ASOS, 0.72 for Yeomsan AWS, and 0.8 for prediction by Kriging interpolation. The result shows that the wind speed value predicted by Kriging interpolation was most related to the power generation.

In order to improve the accuracy of the model's learning, the meteorological values predicted by the kriging interpolation method were pre-processed. By using the values, four decision tree and ensemble-based learning models (decision tree, random forest, XGBoost, and LGBM) were learned and tested. As a result, the LGBM algorithm showed the highest correlation(R^2 _score) with the actual value and the lowest root mean square error(RMSE) compared to other algorithms. Additionally, hyperparameter optimization was performed to improve the predicted value of the model, allowing error value to reduce by 6.1%.

Finally, based on the measurement of Yeomsan AWS, Yeonggwanggun ASOS, and prediction value using kriging interpolation, the LGBM-based power generation prediction model was learned and tested, resulting in that the model using kriging had higher performance.

Based on the experience of this study, development and utilization of a wind power prediction model are intended, that can be used by collecting and utilizing local forecast model (LDAPS) information along with actual wind power generation, weather, and operation information.

I. 서론

1.1 연구의 배경

기후변화에 관한 정부 간 협의체(이하 IPCC)가 2021년에 공개한 제6차 평가 보고서(The Sixth Assessment Report, AR6)에 따르면 지구온난화가 가속화됨에 따라 온실가스를 지금처럼 배출하는 경우 산업화 시절 대비 지구 기온 상승 제한폭인 1.5°C에 도달하는 시점이 기존 예측치인 2030~2052년에서 2021~2040년으로 10여 년 앞당겨질 것으로 전망하였다. 현재 지구 표면 온도는 산업화 이전보다 1.1°C가량 높아졌으며 이에 따라 기록적인 폭염, 폭우, 폭풍, 가뭄, 산불 증가로 이어지고 이러한 추세가 지속되어 1.5°C 증가에 도달한다면 폭염 발생 빈도가 현재 대비 2배에 가까워지며 산호의 77~90%가 소멸하는 등 극단적 기후 변화가 이루어지며 2°C 증가에 도달하면 세계 육지의 20~30%가 사막화되고 생물 종의 54%가 멸종하는 초 극단적인 위기가 발생할 것을 경고하고 있다.

이러한 기후 위기에 대응하기 위하여 세계 주요국들은 파리협정(Paris Agreement)을 체결하여 산업화 이전 대비 지구 평균 온도 상승을 1.5°C 이하로 제한하기 위한 전 지구적 장기 목표하에 각 당사국이 국가결정기여(Nationally Determined Contribution, 이하 NDC) 온실가스 감축목표를 설정하고 이행하도록 규정하였으며 2021년 제26차 당사국총회(COP26)에서 국제 탄소시장 지침을 타결함으로써 파리협정 세부 이행규칙(Paris Rulebook)을 완성하였다.

우리나라에서도 파리협정 1.5°C 목표 실현을 위하여 2030 상향된 NDC와 2050 탄소중립 목표를 설정하여 2018년 대비 40% 이상 온실가스 감축과 2050년 모든 석탄발전 폐지를 약속하고 신재생에너지의무할당제(RPS) 강화 등 탄소중립 정책을 강화하고 기업에서도 RE100 도입하는 등 범국가적 노력을 기하고 있다. 2030 NDC 상향 안에서는 신재생에너지 비율을 기존 9차 전력 수급 계획에 제시된 신재생에너지 비율 20.8%에서 30.2%로 상향 목표를 제시하고 있으며, 이는 2034년까지 기존 신재생에너지 용량 77.8GW(풍력 24.9GW, 태양광 45.6GW)보다 더욱 강화된 비율로 신재생에너지 설비의 증가가 필수적인 상황이다. [1-3]

이렇게 신재생에너지 설비 용량이 증가하고 전력 계통 내 점유율이 증가함에 따라 통제할 수 없는 기상 환경에 의해 발전량이 결정되는 신재생에너지의 간헐적 특성으로 계통주파수 불안정과 전력 수요와 공급 불균형으로 인한 전력망 과부하 등의 문제를 야기하고 있다. 최근 제주에서 2015년 3회의 출력 제한을 시작으로 2021년까지 총 225회의 풍력발전 출력 제한 조치가 있었고 2022년 상반기에만 60회 이상의 출력 제한 조치가 진행되었다.

풍력발전은 구축된 지형 요소에 따른 출력의 간헐성과 변동성으로 초단기적으로는 전력 계통주파수 안전성에 영향을 주며 단기적(수십 분 ~ 수일)으로는 전력 계통 내 수요-공급 측면에서의 전력공급 계획 수립 및 정비계획에 영향을 준다. 풍력발전 단지 구축을 위한 사업성 평가 측면에서 장기간 예측 발전량에 대한 시나리오 도출이 필수적으로 사업 영역 전주기에 대한 발전량 예측의 중요성이 강조되고 있다. [4]

Table 1. Application of wind power generation prediction over Time Range

category	range	application
Ultra short term	seconds/minutes	<ul style="list-style-type: none"> • Electric market clearing • Real-time power grid operation • Regulation actions • Active wind turbine control • Grid frequency stabilization reference
short term	hourly	<ul style="list-style-type: none"> • Economic load supply plan • Rational load determination • Support for operation in the power market
middle term	weekly/monthly	<ul style="list-style-type: none"> • Determining the plan to shut down the generator • Determination of reserve power • Determination of additional input of generator
long term	annual	<ul style="list-style-type: none"> • Establishment of maintenance plan • Operation management • Determination of optimized operating price • Feasibility study for wind power complex design • Reviewing the effects related to climate change

1-2. 연구의 목적

풍력발전 예측에 가장 중요한 요소는 풍력발전기가 설치된 위치에서의 기상 정보가 필요하지만 정확한 풍속, 풍향 등 자료를 확보하기 위해서는 1억원 이상의 구축비와 지속적인 유지보수가 필요한 기상탑 혹은 2.5억원 이상의 윈드라이다같은 고가의 장비가 필요하다.

해당 장비가 구축되지 않은 지역에서의 수익성 판단 및 운영 중 발전량 예측을 위하여 본 연구에서는 접근 및 활용이 쉬운 “공공데이터포털”에서 제공하는 풍력발전기 발전량 정보와 자동 기상관측장비(AWS), 종간 기상관측장비(AOS)의 기상정보(풍속, 풍향, 온도, 습도)를 활용하여 목표 지역의 인근의 기상정보 확보한 뒤 목표 지점에서의 기상 정보를 추정하고 이를 바탕으로 기계학습 기반 풍력발전기 발전량 예측 모델을 개발하고 검증하고자 한다.

II. 이론적 배경

2.1 풍력발전기(Wind Turbine)

풍력발전기는 공기 유동 에너지를 통해 풍력 블레이드를 회전시켜 기계적 운동 에너지로 변환시키고 이를 활용해 전자기 유도를 통한 전기에너지로 변환시켜 전력 계통에 공급하는 기계장치이다.

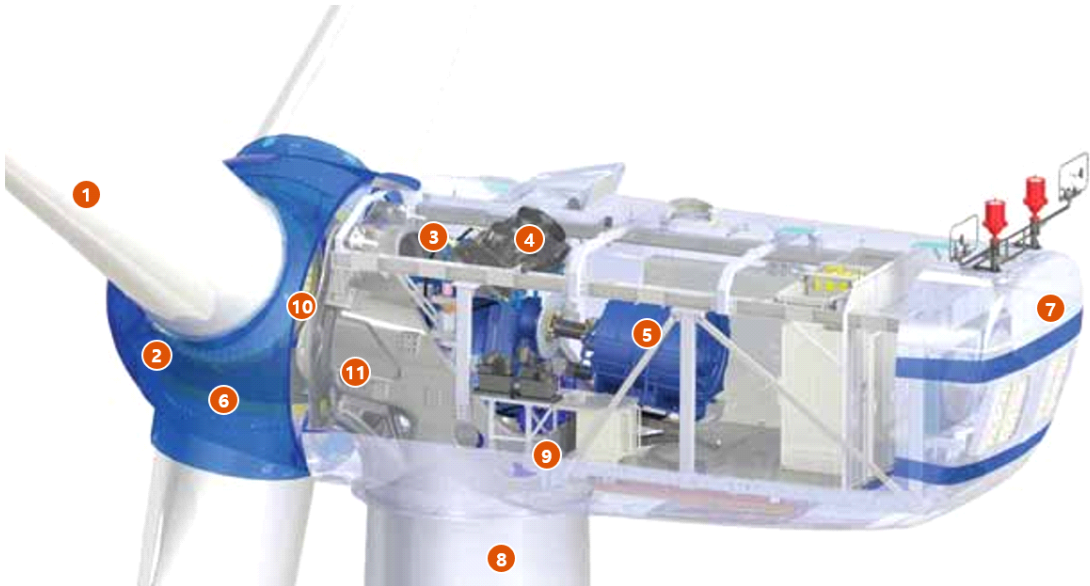


Fig. 1. Main components of wind turbines

(출처: UNISON 2.5MW-U113 Catalog)

풍력발전기는 8,000여 개의 부품(2MW 기준)으로 구성되어 있으며 주요 구조로 기계장치부와 제어장치부로 구분할 수 있다.

기계장치부에는 바람을 받아 회전하는 ①블레이드(blade), 블레이드와 연결되어 회전력을 회전축으로 전달하는 ②회전자(Rotor), 전달된 회전력을 증속기로 전달하는 ③회전축(shaft), 초기 저속 회전력을 기어를 이용해 고속 회전력으로 증폭시키는

④증속기(Gearbox), 전달된 회전력을 이용해 전자기 유도를 통해 전기를 생산하는 ⑤발전기(Generator), 블레이드를 회전자에 고정해주는 ⑥허브(hub), 주요 부품을 내장하고 발전기가 받는 공기 흐름을 조정하는 ⑦나셀(Nacelle) 그리고 풍력발전기를 지탱해주는 구조물인 ⑧타워(Tower)로 구성되어 있다.

제어장치부에는 바람의 방향이 바뀌게 되면 나셀을 바람 방향으로 회전시키는 ⑨요잉 시스템(yawing system), 풍속에 따라 블레이드 경사각(pitch)을 조절하여 발전량을 능동적으로 제어하는 ⑩피치 제어시스템(pitching control system), 중단 풍속 이상 바람이 부는 경우 양력이 블레이드에 작용하지 못하도록 조절하는 ⑪스톨 제어시스템(Stall control system)이 있다.

풍력발전은 설치 장소에 따라 육상풍력(Onshore)과 해상풍력(Offshore)으로 구분할 수 있다. 육상풍력은 입지 특성상 주로 산림지역 또는 산 능선부를 포함한 산지에 77%가 입지하고 있으며 나머지는 해안가나 농지에 입지하고 있다. 해상풍력은 기초구조물 설치에 따라 고정식/부유식으로 구분되며 육상풍력 대비 입지 제약에서 자유롭고 대형화로 높은 발전량을 확보할 수 있다.

풍력발전기의 출력은 식(1)에 따라 결정되며 시동풍속(Cut-in wind speed)에 전력 생산을 시작하여 정격 풍속(Rated wind speed)에서 최대 출력을 내고 중단(최대) 풍속(Cut-out wind speed) 이상의 과풍속 환경에서는 발전을 중단한다.

$$P = \frac{1}{2} \rho A V^3 \quad (1)$$

ρ : 공기밀도[$1.225\text{kg}/\text{m}^3$]

A : 블레이드의 회전 면적[m^2]

V : 풍속[m/s]

이러한 출력 특성으로 블레이드 크기가 클수록, 설치된 지형의 풍속이 빠를수록 더 많은 발전량을 생산할 수 있는데 ‘베츠의 법칙(Bet’s Law)’에 따라 바람을 에너지로 변환 시 효율의 한계치는 약 59.26%로 그 이상의 효율을 달성하기 어렵다. [5]

이러한 베츠의 법칙 한계 내에서 블레이드 크기와 형상, 발전기의 개량, 송배전망 손실 최소화 등 최대 효율을 내는 개발이 지속적으로 이루어지고 있다.

2-2. 지상기상관측장비

본 연구에서는 풍력발전기의 기상정보를 확보하기 위하여 공공데이터 포털을 통해 자동기상관측장치(Automatic Weather System, 이하 AWS)와 종간기상관측장치(Automatic Synoptic Observing System 이하 ASOS)에서 취득된 자료를 수집하여 분석에 활용하였다.

2-2-1. 자동기상관측장치(Automatic Weather System)

AWS는 과거 사람이 직접 관측하던 정보를 자동으로 관측할 수 있도록 설계된 장비로 실시간 측정, 연산, 저장, 표출 등 전 과정으로 자동으로 처리한다. 측기탑 상당 10m의 풍향 풍속계, 지면으로부터 1.5m 높이에 온도, 습도계, 강수유류계가 설치되어 있으며 처리기 내 강수량계가 설치되어 있어 관측지점에서의 다양한 대기 상태를 실시간으로 관측하고 있다.

2-2-2. 종간기상관측장치(Automatic Synoptic Observing System)

ASOS는 주로 기상 관서에 설치되어 있으며 국제적 정해진 시간의 대기 상태를 파악하기 위하여 전 관측소에서 동일 시간대 대기 정보를 수집하는 지상관측장비로 일조량, 초상온도, 지중온도 등 AWS보다 많은 기상정보를 자동 관측하며 일부 지점에서는 지면상태, 증발량, 구름, 일기 현상 등 육안으로 관측하여 기록한다. [6]

2-3. 공간 보간법(spatial interpolation)

본 연구에서는 풍력발전기가 위치한 지역의 정보를 추정하기 위하여 관측지역의 값을 활용하여 다른 미관측 지역의 값을 추정하는 공간 보간법을 활용하였다.

관측지점의 값을 기반으로 공간적 수치를 계산하는 공간보간법은 두 가지 유형으로 분류할 수 있는데 관측지점의 자료가 갖는 영향권을 대표할 수 있는 이항 함수를 설정하고 그 함수의 변수를 이용하여 추정하는 '적합 함수 접근법(Fitted function interpolation)'과 어떤 범위에 속하는 관측점들의 영향 치를 직접 계산하여 특정 지점의 자릿값을 얻는 '가중평균법(Weighted average interpolation)'이 있다. 가중평균법은 거리에 기초한 가중평균법과 면적에 기초한 가중평균법으로 다시 구분할 수 있으며 대표적 내삽법으로는 역거리 가중(Inverse Distance Weighting, IDW) 보간법, 크리깅(Kriging) 보간법 등이 있다.

크리깅(Kriging) 보간법은 광물 매장량, 광범위한 지역의 지하수위, 대수층, 오염물질의 확산 정도 등을 예측하기 위해 많이 사용되고 있는 지구통계학 기법으로 1951년 남아프리카 공화국의 D.G Krige라는 광산 기술자가 광맥의 관측 데이터를 기반으로 데이터 간 가중치를 부여하여 미지점의 새로운 광맥을 찾기 위해 사용한 경험적 방법을 1962년 Goldberger, 1973년 Matheron이 수학적으로 정립하여 오차 분산을 최소화하는 가중선형 조합의 보간법으로 활용되고 있으며 정규 크리깅(Ordinary Kriging), 일반 크리깅(Universal Kriging), 이동 크리깅(Moving Kriging) 등 다양한 크리깅 기법들이 연구되고 있다.

인접 데이터에서 가중 평균을 계산한 영역의 한 점에서 값을 예측한다. 위·경도 격자지도에서 특정 지점의 예측값은 식(2)과 같이 추정된다. [7-9]

$$z_0^*(x_0, y_0) = \sum_{i=1}^n \lambda_i z_i(x_i, y_i) \quad (2)$$

z_i : 지점 i 에서의 관측값

λ_i : 지점 i 에서의 크리깅 가중치

x_i : 지점 i 의 x좌표

y_i : 지점 i 의 y좌표

각 지점 간 공간적 연속성은 분산도(variogram) 분석으로 검증되며 각 지점 값 간의 분산도는 거리에 대한 함수로 표현되기에 가중치 결정에 중요한 인자는 지점 간의 분리거리(separation distance)이다. 분리거리가 h 인 두 지점 간 분산도 $2\gamma(h)$ 는 거리에 따른 값의 공분산(covariance)으로 식(3)과 같이 정의되며, 이러한 지점이 N 개 존재할 때 반분산도(semi-variogram) $\gamma(h)$ 는 식(4)과 같이 표현된다.

$$2\gamma(h) = Var[z_{i+h} - z_i] \quad (3)$$

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^N (z_{i+h} - z_i)^2 \quad (4)$$

공간적으로 연속된 두 지점의 거리가 가까울수록 분산도는 작게 나타나고 거리가 멀어짐에 따라 분산도는 커지며 의미 없는 값을 가지게 된다. 이렇게 무의미한 경향을 보이는 분리거리를 상관거리(range) a , 이에 대응하는 분산도 값을 문턱값(sill) C_0 라 정의한다. 이러한 값을 추종하는 수학적 모델 중 일반적으로 사용되는 모델은 선형(Linear), 구형(Spherical) 모델, 지수형(Exponential) 모델, 가우시안(Gaussian) 모델 등이 있다.

Table 2. Feature and formulas of semi-variogram

Model	Describe	$\gamma(h)$	h
Linear	A simple model Assuming that the correlation of data changes linearly, there is a limit to the explanation of the distribution.	$C_0 \frac{h}{a}$ c	$h \leq a$ $h > a$
Spherical	Commonly utilized models. It is expressed in the form of a third-order polynomial and the half variance value matches the threshold.	$C_0 \left[\frac{3}{2} \frac{h}{a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right]$ c	$h \leq a$ $h > a$
Exponential	In the Spherical model, the position where the tangent line drawn at the separation distance $d=0$ intersects the Sill value is located at $2/3$ of the correlation. As the separation distance increases, it converges to the Sill value.	$C_0 \left[1 - \exp\left(-\frac{3h}{a}\right) \right]$	$h \geq 0$
Gaussian	Used to indicate a strong correlation of small separation data or to use a model with strong continuity. Use the separation distance of 95% of the threshold as the actual correlation	$c \left[1 - \exp\left(-\frac{h^2}{a^2}\right) \right]$	$h \geq 0$

정규 크리깅(Ordinary Kriging)에서는 목표 지점의 예측값(z_0^*)을 얻기 위하여 근처 지점(z_i)에 대한 가중치(λ_i)를 결정하기 위한 예측값은 편향되지 않아야 하며 오차분산이 최소화되는 조건을 만족해야 한다. 이 편향되지 않는 조건을 만족하기 위한 제약조건은 식(5)과 같다.

$$\sum_{i=1}^n \lambda_i = 1 \tag{6}$$

정규 크리깅은 식(5)을 만족하는 조건에서 오차 분산이 최소화되는 가중치를 구하는 기법으로 이를 수식화하면 식(6)과 같다.

$$\sigma^2 - 2 \sum_{i=1}^n \lambda_i \sigma_{0i}^2 + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \sigma_{ij}^2, \sigma_{ij}^2 = Cov(z_i, z_j) \quad (6)$$

상기 식을 라그랑지 승수법(Lagrange multiplier method)을 사용하여 재정의하고 라그랑지 목적함수 L 을 최소화하기 위하여 가중치 λ_i 와 라그랑지 승수 μ 에 대해 편미분(Partial Derivative) 값이 0이 되게 정의하면 다음식으로 표현할 수 있다.

$$\frac{\partial L}{\partial \lambda_i} = -2\sigma_{0l}^2 + 2 \sum_{i=1}^n \lambda_i \sigma_{il}^2 - 2\mu = 0, l = 1, 2, \dots, n \quad (7)$$

$$\frac{\partial L}{\partial \lambda_i} = \left(2 \sum_{i=1}^n \lambda_i \right) = 0 \quad (8)$$

식(7)과 식(8)을 재정의하면 식(9)으로 나타낼 수 있고 이를 행렬화하면 식(10)으로 나타낼 수 있으며 이를 매트릭스 방정식으로 풀면 식(2)에 요구되는 지점별 크리깅 가중치를 계산할 수 있게 된다.

$$\sum_{i=1}^n \lambda_i \sigma_{ij}^2 - \mu = \sigma_{0l}^2, l = 1, 2, \dots, n \quad (9)$$

$$\begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1n}^2 & 1 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2n}^2 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \dots & \sigma_{nn}^2 & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \\ \omega \end{pmatrix} = \begin{pmatrix} \sigma_{01}^2 \\ \sigma_{02}^2 \\ \dots \\ \sigma_{0n}^2 \\ 1 \end{pmatrix} \quad (10)$$

본 연구에서 기상 자료 수집에 활용한 지상기상관측장비는 10m 높이에서 수집된 자료로 정확한 풍력 발전량 예측을 위해서는 연구 대상의 허브 높이인 100m 고도에서의 풍속이 필요하다. 관측된 고도의 풍속을 활용하여 다른 고도의 풍속을 예측하는 방법으로는 디컨(deacon) 방정식을 사용하였다. 디컨 방정식은 식(11)의 수식을 통해 보정된 고도의 풍속과 고도를 구할 수 있으며 이때 풍속 분포는 식(12)에서처럼 지면의 거칠기 길이를 구하여 계산할 수 있다. [10]

$$U(z) = U(z_a) \times \left(\frac{z}{z_a} \right)^\alpha \quad (11)$$

$z, U(z)$: 보정된 고도와 보정 고도에서의 풍속
 $z_a, U(z_a)$: 실측 고도와 실측 고도에서의 풍속
 α : 풍속할증계수

$$\alpha = a + b \ln(U_2) \quad (12)$$

$$a = \frac{1}{\ln\left(\frac{z_g}{z_0}\right)} + \frac{0.088}{1 - 0.088 \ln\left(\frac{z_a}{10}\right)}$$

$$b = \frac{-0.088}{1 - 0.088 \ln\left(\frac{z_a}{10}\right)}$$

z_g : 해발고도의 기하평균

z_0 : 지면의 거칠기 길이

z_a : 풍속 측정 고도

2-4. 기계학습(Machine Learning)

기계학습은 인공지능의 한 분야로 인간이 경험을 통해 유사한 상황에서 판단을 내리는 ‘학습’이라는 능력을 컴퓨터로 구현하는 방법으로 Arthur Lee Samuel은 기계 학습을 “기계가 일일이 명령하지 않은 동작을 데이터로부터 스스로 학습하여 실행할 수 있도록 방법을 개발하는 연구 분야”라고 정의하였다. [11]

즉 일정 규모 이상의 데이터를 제공하여 컴퓨터가 데이터 분석을 통해 일정한 규칙을 인식하고 해당 규칙을 기반으로 다른 데이터를 분류하거나 예측하는 기술이다. 이런 기계학습의 유형에는 보통 지도학습(Supervised Learning), 비지도 학습(Unsupervised Learning), 강화학습(Reinforcement Learning)으로 분류한다. 지도학습은 컴퓨터에게 문제(독립변수)와 정답(종속변수)을 알려주고 그사이에서의 연관성을 찾아내고 모델화하여 새로운 데이터에 대한 출력을 예측하는 학습 방법이다. 지도학습에는 회귀 방법(Regression)과 분류 방법(Classification)이 있다. 비지도 학습의 경우 지도학습과 달리 입력값에 대한 정답을 주지 않아 어떤 연관된 규칙을 찾기보다는 데이터의 숨겨진 패턴이나 구조를 파악하는 방식으로 군집(clustering)이 일반적인 비지도 학습법이다. 강화학습은 학습을 통해 능력을 향상시키는 방법으로 어떤 이벤트에 대한 보상을 주고 그 보상이 최대가 되는 방향으로 학습하는 방법이다. [12]

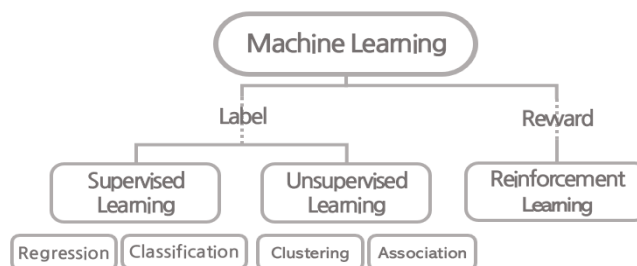


Fig. 2 Types of machine learning

이러한 단일 모델을 통한 학습보다 예측력이 강한 모델을 만들기 위하여 약한 단일 모델을 여러 개 연결하여 투표(voting) 절차를 통해 더 강력한 예측 모델을 만드는 기법을 앙상블(Ensemble) 기법이라 한다. 모델 비교를 통해 최종 모델의 예측값을 결정하는 투표는 크게 직접 투표(Hard voting)와 간접 투표(Soft voting)로 분류된다. 직접 투표는 각각의 약한 모델들의 예측 결과값을 다수결 투표를 통해 결정하는 방식이며 간접 투표는 예측 확률값의 평균 또는 가중치 합을 활용하여 결정하는 방식이다.

앙상블 알고리즘은 학습 방법에 따라 배깅(Bagging), 부스팅(Boosting)으로 분류될 수 있다. 배깅은 Bootstrap Aggregating의 약자로 기계학습의 알고리즘의 안정도와 정확도를 향상시키기 위하여 주어진 데이터셋을 임의로 샘플링하여 신규 데이터셋을 만들어내는 부트스트랩을 활용하는 방식으로 이를 통해 생성된 다수의 데이터셋을 바탕으로 약한 학습기를 훈련시킨 뒤 그 결과들을 투표하면서 모델을 강화하는 방식이다. 이러한 샘플링 특성 때문에 데이터 수가 적은 경우에도 분산과 과적합(overfitting)을 줄이며 성능을 향상시킬 수 있다. 부스팅 방식은 반복적으로 모델을 갱신하는데 이때 반복 결과에 따라 해당 데이터셋 샘플에 대한 가중치를 부여한다. 모델 갱신을 반복함으로써 샘플 가중치에 따라 각기 다른 모델 트리가 생성되며 최종적으로 모든 반복과정에서 생성된 모델의 결과를 투표하여 결정하는 방식이다. 이때 배깅 방식이 각 학습기가 의존적으로 동작하며 이전 학습의 오차가 개선된 트리를 생성한다는 점에서 배깅보다 높은 성능을 보여준다. [12]

본 연구에서는 사용된 기계학습 알고리즘으로는 트리 기반의 학습 알고리즘인 의사결정나무(Decision Tree), 랜덤포레스트(Random Forest), XGBoost(eXtreme Gradient Boosting), LGBM(Light Gradient Boosting Machine)이다.

트리 기반 학습 모델 중 대표적인 알고리즘인 의사결정나무는 데이터 변수를 분류하는 일련의 조건을 찾아 데이터를 분류, 회귀하는 지도학습 모형 중 하나로 결과 모델이 나무(Tree) 구조로 되어 있어 Decision Tree라는 이름을 가지고 있다. 분기 때마다 변수 영역을 두 개로 분류하고 지니(Gini) 혹은 엔트로피(Entropy) 지수로 불순도(Impurity)가 감소하는 방향으로 분기된다. 분기 과정에서 과적합

(over-fitting)을 막기 위하여 특정 노드 하부 트리를 제거하는 가지치기 과정을 통해 일반화 성능을 높인다. 의사결정트리는 모델의 시각화가 쉽고 변수의 전처리, 정규화 작업 없이 사용할 수 있는 장점이 있지만 통계적 유의성 없이 분기하며 과적합이 발생할 수 있다. [13]

이러한 의사결정 나무의 과적합 문제를 해결하고 성능을 향상시키는 앙상블 기법으로 배깅을 이용한 랜덤 포레스트(Random Forest) 모델이 있다. 부트스트랩을 이용하여 임의의 데이터 샘플을 만들고 생성된 샘플로 각각의 의사결정 트리를 만들어 가장 성능이 좋은 트리를 투표를 통해 찾는 기법이다. [14]

트리 기반 앙상블 기법 중 대표적인 부스팅 알고리즘으로 GBM(Gradient Boosting Machine)이 있다. 의사결정나무를 약한 학습기로 활용하는 모델로 잘못 분류된 오차에 대한 미분 값을 최소화하는 방식의 경사하강법(gradient descent)을 이용하여 반복 학습을 통해 실제값과 예측값 차이인 잔차(residual)를 줄여가는 방향으로 학습된다. 이러한 GBM의 단점은 가중치 계산의 반복 수행으로 학습 시간이 다른 모델에 비해 상대적으로 오래 걸리는 데 있다. [15]

이러한 GBM의 단점을 보완한 알고리즘이 XGBoost 알고리즘으로 eXtreme Gradient Boosting의 약자다. GBM과 가장 큰 차이점으로는 배깅 기법처럼 병렬 학습이 가능하여 학습 속도가 매우 개선되었다. 이외에도 조기 중단(Early Stopping) 및 가지치기 기능으로 과적합 규제(Regularization) 기능도 강화되었다. [16]

LGBM은 Lighten Gradient Boosting Machine의 약자로 XGBoost 알고리즘에 비해 훨씬 학습에 걸리는 시간이 적다는 것이 장점이다. 가장 큰 특징은 기존의 결정 트리의 깊이(max_depth)를 효율적으로 줄이기 위하여 균형 트리 분할(Level-wise tree growth) 방식에서 리프 중심 분할(Leaf-wise tree growth)을 수행하며 트리의 균형보다는 최대 손실 값을 가지는 리프 노드를 최대 깊이까지 빠르게 도달하여 다른 알고리즘 대비 빠른 속도와 성능을 확보할 수 있다. 하지만 적은 수의 데이터로 학습하는 경우 과적합 문제를 발생할 수 있어 과적합을 방지하는 초매개변수 최적화를 통해 관리해야 한다. [17]

2-5. 변수 및 모델 평가 지표

2-5-1. 피어슨 상관계수(Pearson Correlation Coefficient)

발전량과 다른 변수와의 연관성 강도를 측정하기 위하여 상관계수 중 하나인 피어슨 상관계수를 활용하였다. 피어슨 상관계수는 주어진 두 변수 간의 선형 상관관계를 계량화한 수치로 각 변수의 선형관계인 공분산(covariance)을 각각의 표준 편차의 곱으로 나눈 값이다.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (11)$$

n : 샘플 크기

x_i, y_i : i 에 인덱싱된 개별 샘플 포인트

\bar{x}, \bar{y} : 샘플 평균

상관계수의 값은 코시-슈바르츠 부등식(Cauchy-Schwarz)에 의해 항상 1과 -1의 범위 내에 있으며 절댓값 1인 경우 완벽한 선형 상관관계, 0은 상관관계 없음을 의미한다. 일반적으로 0.7 이상을 강한 상관관계로 판단한다.

2-5-2. 결정계수(Coefficient of Determination, R-Square, 이하 R2-Score)

선형회귀 알고리즘을 활용하여 학습하는 모델의 적합도를 판단할 때 활용되는 성능지표로 독립변수가 종속변수를 어느 정도 설명하지를 나타내며 1에 가까울수록 성능이 좋은 예측 모델이라 할 수 있다

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

y_i : 실제 값
 \hat{y}_i : 예측 값
 \bar{y} : 평균 값

2-5-3. 평균 제곱근 오차(Root Mean Square Error, 이하 RMSE)

기계학습 알고리즘으로 학습된 풍력 발전량 예측 모형의 성능 비교를 위하여 RMSE를 사용하여 모형별 예측값과 실제 발전량 값 간 오차를 검증하였다.

RMSE는 잔차의 제곱에 대한 평균값에 루트를 씌운 값으로 지표 자체가 객관적이며 예측변수와 단위가 같다. 잔차에 대한 제곱을 하기에 이상치에 민감하며 제공된 잔차를 다시 루트로 풀어주기에 잔차를 제공할 때 생기는 값의 왜곡이 적다.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (R_t - F_t)^2}{n}} \quad (13)$$

n : 예측에 활용된 데이터 수
 t : 관측 시점
 R : 실제 발전량
 F : 예측 발전량

RMSE는 오차에 대한 지표이기 때문에 값이 작을수록 예측 모형이 상대적으로 우수하다는 것을 나타낸다. [18]

III. 풍력 발전량 예측 모델 개발

본 연구는 풍력발전소 인근의 지상기상관측장비로 측정된 기상정보를 크리깅 보간법을 통해 발전소 위치의 풍속, 풍향 등의 기상정보를 수집하고 이를 이용하여 풍력 발전량을 예측하는 모델을 개발하는 연구다.

풍력발전소 발전량 데이터 및 기상관측장비의 기상정보 데이터는 ‘공공데이터포털’에서 취득할 수 있는 개방 데이터를 활용하여 발전량 예측 모델 개발 과정을 Fig. 3.과 같다.

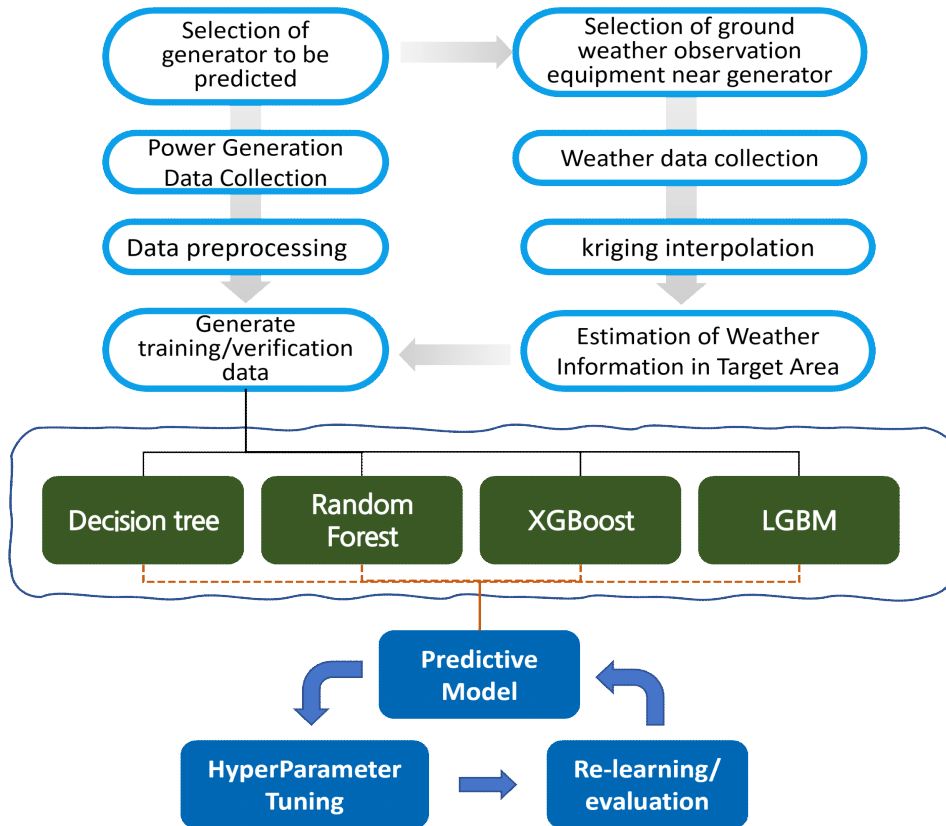


Fig. 3. Design and Validation Process for Wind Power prediction Model

3-1. 연구 대상 선정 및 분석

공공데이터포털에서 제공되는 자료 중 한국동서발전(주)에서 제공하는 영광백수풍력백수 발전단지 내 1번 풍력발전기의 10분 평균 출력 자료(51,586개)와 1시간 평균 출력 자료(8,604개)를 활용하였다. [19-20]

영광백수풍력발전단지는 전라남도 영광군 백수읍 하시리 간척지 일원에 100m 높이의 타워가 적용한 유니슨 2.3MW U113 풍력발전기 20기(①~⑳), 총 40MW 규모로 2015년 5월에 구축된 풍력발전단지로 연간 약 90GWh의 전력을 생산하고 있다. Fig. 4는 영광백수풍력발전단지의 배치도로 북쪽, 서쪽 방향으로 바다와 인접한 간척지에 설치되어 있다.

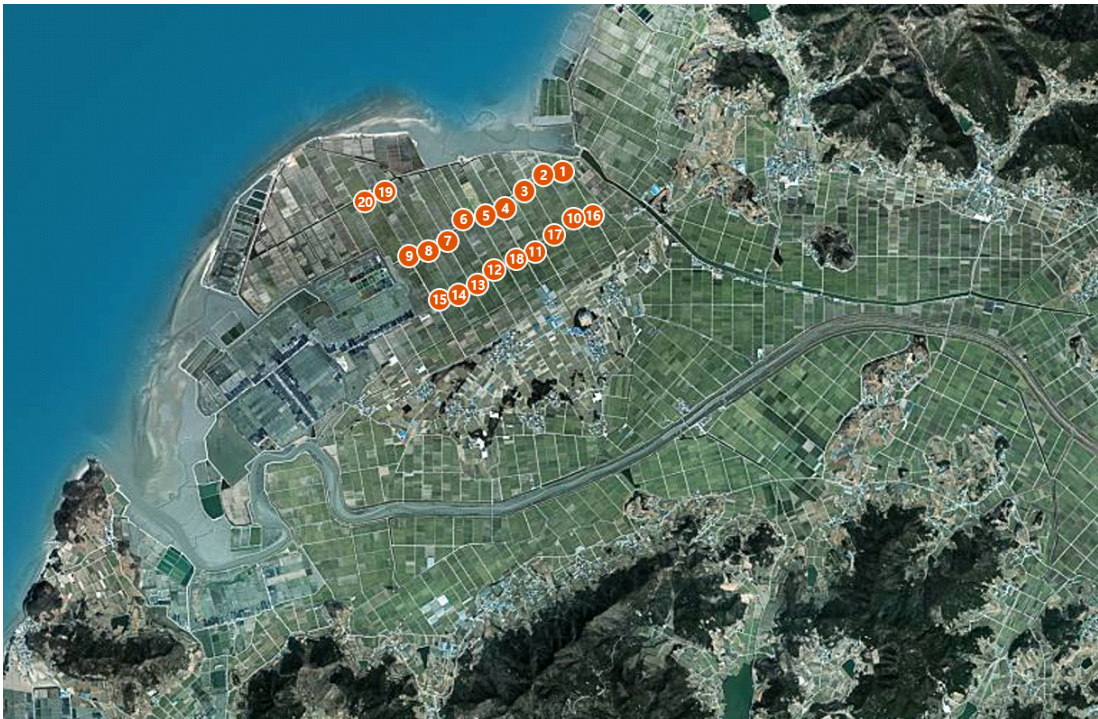


Fig. 4. Yeonggwang-Baeksu wind power plant layout

본 연구에서 연구 대상인 영광백수풍력단지 1번 풍력발전기는 U社의 U-113(2.3MW) 모델로 주요 제원은 Table 2.와 같다.

Table 3. Wind turbine(Unison U113 model) datasheet

Category	Description
Type	Horizontal axis, Upwind, Variable speed
Rated power	2.300kW
Rotor diameter	112.8m
Hub height	80m, 100,. Tubular steel tower/ 140m Hybrid tower
Power regulation	Pitch control
Rotational speed	6~15.4rpm
Cut-in/Cut-out wind speed	3m/s / 20m/s
Rate wind speed	10.5m/s
Extreme wind speed(V_{e50})	59.5 m/s
Design type class	IEC S(7.5m/s(avg), TI: 17%)
Design life time	20years
Temperature range	Operation: -10℃ to 40℃, Standstill: -20℃ to 50℃

(Source: UNISON 2.5MW-U113 Catalog)

제공된 제원을 식[1]에 대입하여 도식화하면 [Fig. 5]와 같은 출력 곡선(Power Curve)을 얻을 수 있다.

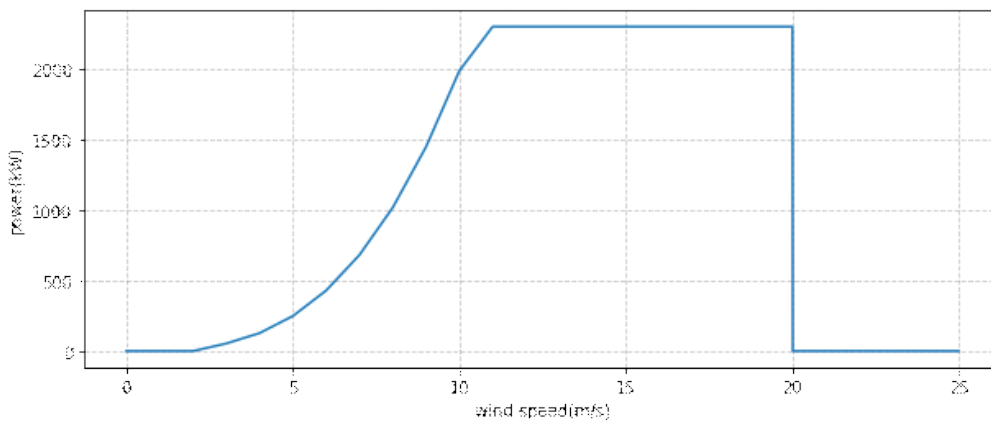


Fig. 5. Wind turbine(Unison U113 model) power curve

영광백수풍력단지 1번 풍력 발전기의 2020년 1월부터 12월까지의 발전량을 CSV 파일, 혹은 OpenAPI를 통해 XML, JSON 파일 형태로 받을 수 있으며 해당 데이터세트 중 본 연구에서는 ‘일시’, ‘평균 발전량(kWh)’, ‘평균 출력(kW)’, ‘평균 풍속(m/s)’ 데이터를 활용하여 1번 풍력발전기의 발전량을 도식화하여 분석하였다.

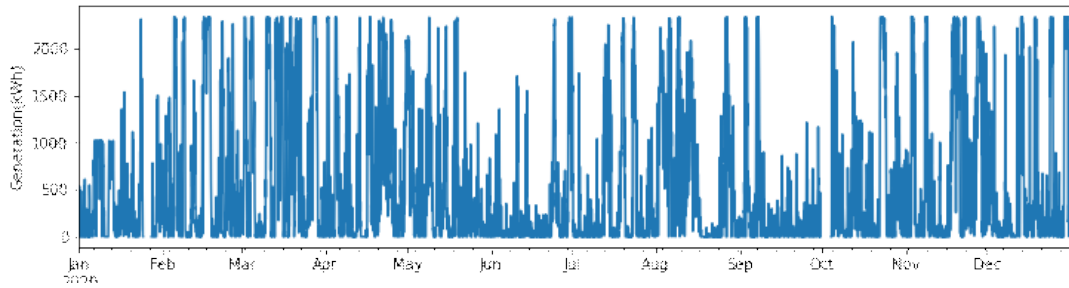


Fig. 6. Hourly power generation chart

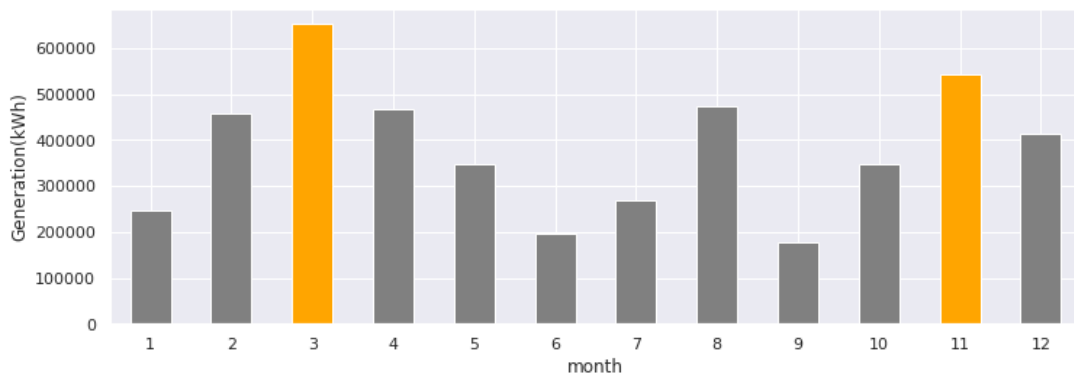


Fig. 7. Monthly power generation chart

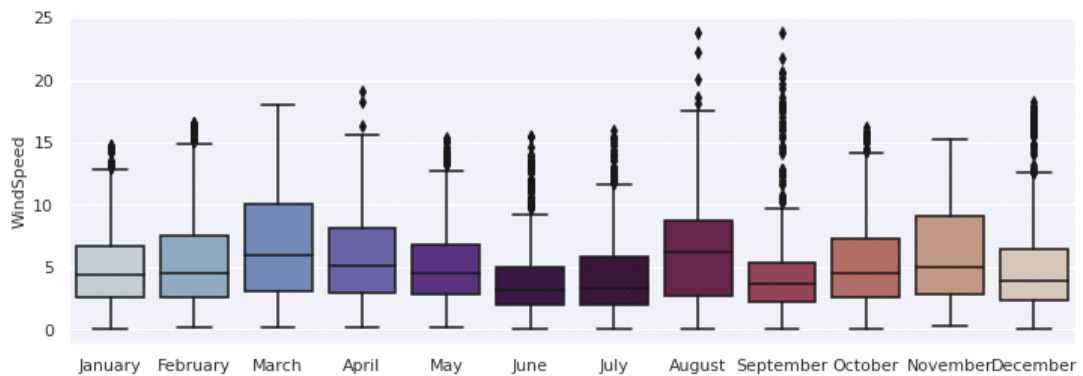


Fig. 8. Monthly box plot of wind speed values

월간 풍속의 상자 그림(Box plot)을 확인해보면 9월은 마이삭, 하이선 등의 태풍의 영향으로 최고 풍속은 높지만, 최댓값 범위를 넘어가는 이상치로 분류되며 가장 낮은 발전량을 보였고 사분위범수 Q1, Q3 풍속이 높고 범위가 넓었던 3월, 11월에 가장 높은 발전량을 보였다.

Fig. 9와 같이 풍속에 따른 출력 곡선을 도식화하면 이론상 계산된 출력 곡선과 유사한 경향을 보였으나 일정 구간에서 풍속이 기준보다 증가함에도 출력이 1MW 구간에 머물러 있음을 확인할 수 있다.

이에 풍속이 9m/s보다 크면서 평균 출력이 1~1.2MW인 구간을 확인 한 결과 Fig. 10.처럼 2020년 1월 6일부터 1월 9일까지 기준 풍속 이상에도 1MW 이상 출력이 증가하지 않는 문제를 확인할 수 있었고 정상 풍속 범위(3~10m/s)에서 출력이 0인 기간을 도식화하면 Fig. 11.에서 보는 것처럼 1월 10일~11일 유지보수를 통해 출력 문제가 정상화되었음을 예상해 볼 수 있다. 그 외 정상 풍속에서 출력이 없는 경우는 진단, 유지보수 등 작업이 있거나 태풍 등에 의한 가동 정지로 가정한다.

정지, 고장 및 유지보수 등의 가동 정지에 따른 이상치(outlier)를 제거한 정상 발전상태의 풍속에 따른 출력 곡선을 Fig. 12.와 같이 도식화하였다. 제공된 평균 풍속은 풍력발전기의 나셀 풍속계에서 계측된 풍속으로 블레이드의 회전, 나셀 모양, 발전기 주위의 순간적인 측면바람 등의 영향으로 실제 허브와 블레이드를 향한 풍속과 일부 다를 수 있음을 고려할 때 일부 이상치를 제외한다면 이론적으로 계산된 출력과 실제 출력 간 피어슨 상관계수는 0.99로 매우 유사한 경향성을 보이므로 발전량 예측에 적합한 데이터임을 확인할 수 있다.

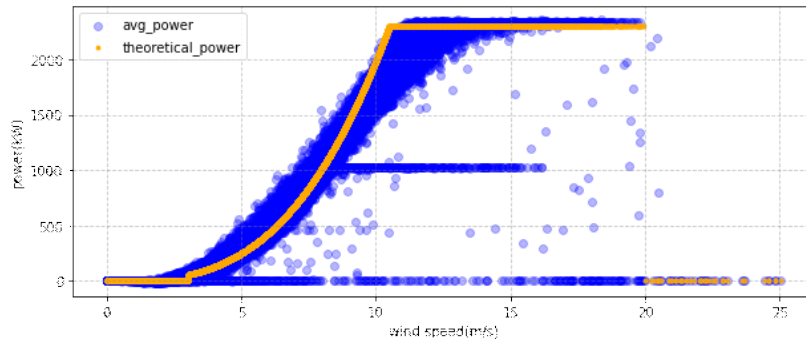


Fig. 9. Power curve before outlier removal

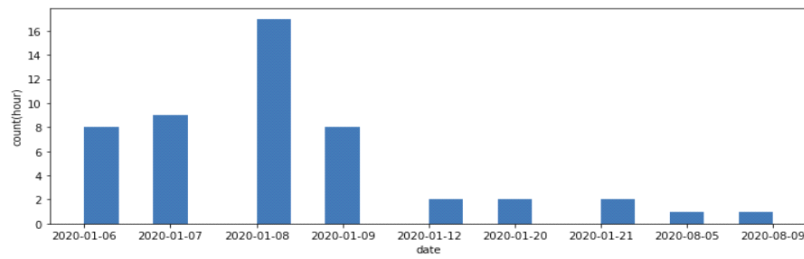


Fig. 10. Time frequency chart with denaturation range of 1 to 1.2 MW with wind speed higher than 9 m/s

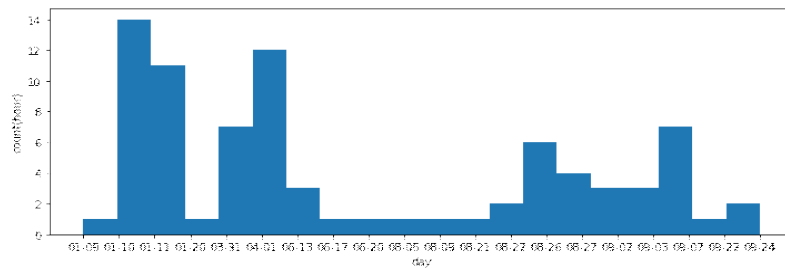


Fig. 11. Time frequency chart with zero power generation in the driving wind speed interval (3-20 m/s)

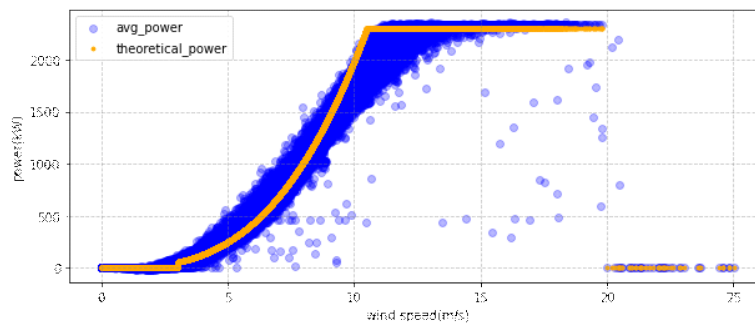


Fig. 12. Power curve after outlier removal

3-2. 풍력 발전량 예측을 위한 기상정보 추정

3-2-1. 연구 대상 인근 기상정보

풍력 발전량 예측을 위한 입력 변수로는 풍속, 풍향, 온도, 기압 등 외부 기상 정보와 기어박스 온도, 진동 데이터, 블레이드 피치 각도 등 내부 부품/장비 상태와 운전 조건을 확인할 수 있는 내부 정보로 별도 SCADA 데이터가 제공되지 않기에 풍력발전기 인근 지상기상관측장비의 기상정보를 수집하여 입력 변수로 활용하였다. [21-22]



Fig. 13. Weather observation equipment closest to the study target

연구 대상 인근의 지상관측장비는 영광군 ASOS(지점번호 252, 위도 35.28366, 경도:126.47784)와 염산AWS(지점번호 769, 위도 35.1993, 경도 126.3834)로 공공 데이터포털 API를 통해 해당 지상기상관측장비의 2020년 1월부터 12월까지의 기상정보(풍속, 풍향, 기온, 습도) 17,534행의 데이터를 수집하였다.

발전량 예측에 활용한 기상정보를 선정하기 위하여 염산AWS와 영광군 ASOS에서 수집된 각각의 풍속(wind_speed), 풍향(wind_direction), 기온(temperature), 습도(humidity) 데이터를 일자별로 풍력발전기의 발전량(Generation), 풍속(Wind Speed_Avg) 데이터와 결합한 데이터셋을 생성하고 각 변수 간 피어슨 상관계수를 계산하고 히트맵 형태로 도식화하였다.

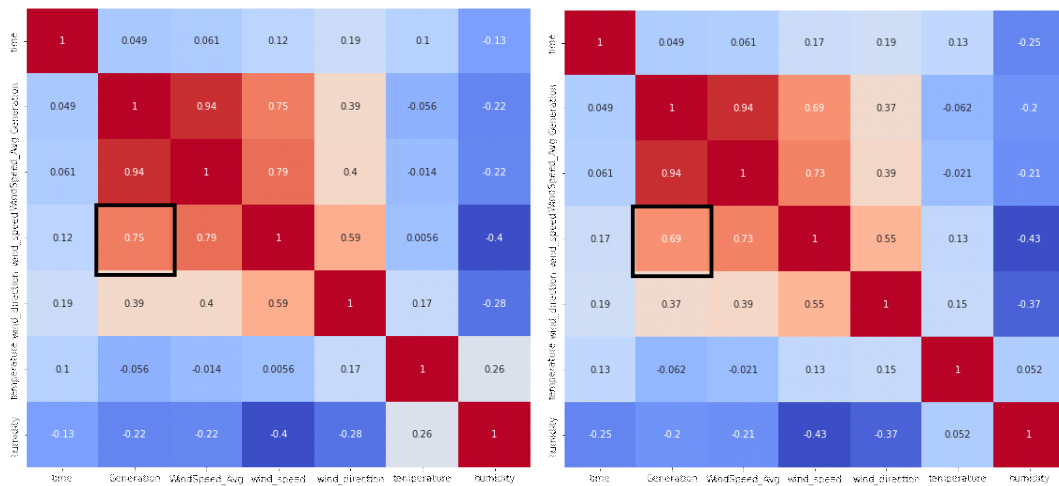


Fig. 14. Pearson's correlation heat map for YeomsanAWS(left)dataset and Yeongwanggun ASOS(Right) dataset

풍력 발전량 예측 모델 설계를 위하여 제거할 나셀 측정 평균 풍속을 제외하면 발전량과 피어슨 상관계수가 가장 높은 변수는 평균 풍속으로 영광군ASOS 풍속은 0.69, 염산AWS 풍속은 0.75다.

강한 상관관계를 나타내는 0.7 이상 혹은 근접한 상관계수를 보였지만 기록이 누락된 시간대의 기상정보 결측치가 존재하며 발전기 기준 시동풍속 3m/s 이상인 시간대에 풍속/풍향 데이터가 0인 이상치가 존재하여 발전량 예측 입력 변수로 활용하기 위해서는 이러한 이상치를 제거하고 유의미한 값으로 채워주어야 한다.

3-2-2. 정규 크리깅을 이용한 기상정보

이에 영광백수풍력발전단지 1번 풍력발전기 위치에서의 정확한 기상 데이터를 확보하고자 풍력 발전단지 인근의 기상관측장비(AWS, ASOS)의 기상정보를 수집하고 해당 관측장비의 위치정보와 기상정보를 활용하여 풍력발전기 위치의 기상정보를 추정하였다.

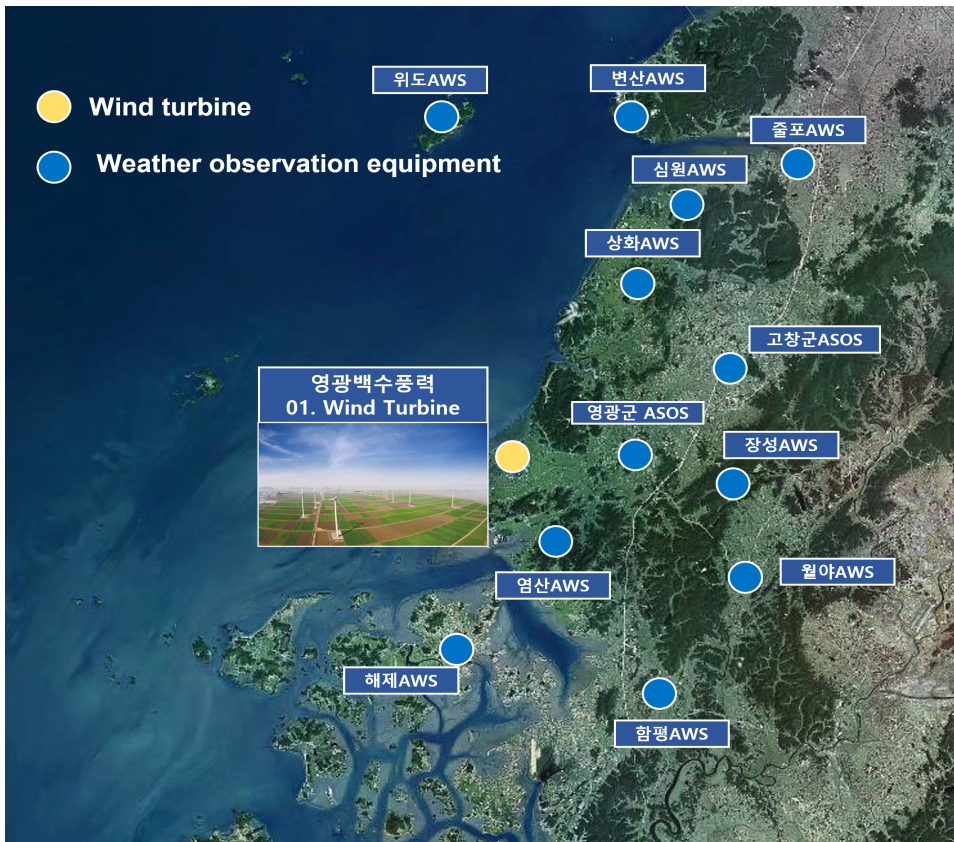


Fig. 15. Study target nearby weather observation equipment

영광백수풍력 1번 발전기 인근에 있는 지상 관측장비(AWS, ASOS)의 지점 중 풍향, 풍속 데이터가 누락된 낙월도 AWS 등의 지점을 제외하고 12개의 AWS, ASOS 지점을 선정하여 공공데이터포털 API를 통해 12개 지점의 위·경도 좌표, 관측일시, 기상 자료(풍속, 풍향, 습도, 온도) 130,109행의 데이터셋을 생성하였다.

생성된 데이터세트를 바탕으로 풍력발전기의 기상정보를 확보하기 위하여 일반적으로 사용되고 있는 크리깅 기법인 정규 크리깅(Ordinary Kriging)을 활용하여 풍력발전기 위치의 기상정보를 추정하였다. 크리깅을 이용하여 위·경도 격자 지도 내 풍력발전기의 기상정보를 추정하기 위해서는 모든 분리거리 h 에 대한 반분산도 $\gamma(h)$ 가 필요하며 확보된 지점별 데이터를 기반으로 예측 오차가 적은 반분산도 해석 모델을 선정해야 한다.

정규 크리깅을 통한 임의 지점의 기상정보 예측 모델을 학습하기 위하여 생성한 데이터세트에서 임의의 관측일시 기준 테스트 데이터를 샘플링하고 파이썬 라이브러리인 Pykrige 패키지를 활용하여 4개의 반분산도 해석 모형(Linear, Spherical, Exponential, Gaussian)을 적용한 정규 크리깅 보간을 통한 풍속 예측값과 임의 지점의 실제 풍속 값 간의 절대 오차를 평균하여 분석하였다.

각 반분산도 모델 적용 크리깅 예측 풍속 값과 실제 풍속 간 RMSE 결과는 Table 4.와 같다.

Table 4. Evaluation of Ordinary kringing prediction value by semi-veriogram models

	Linear	Spherical	Exponential	Gaussian
count	500	500	500	500
mean	1.065526	1.070682	1.058981	1.164915
std	0.404643	0.408866	0.394642	0.480338
min	0.337527	0.330709	0.33777	0.285731
25%	0.800472	0.809634	0.803176	0.849567
50%	0.988486	1.003386	0.983177	1.053762
75%	1.251541	1.263565	1.238478	1.392713
max	3.291128	3.378518	3.351625	4.020546

본 연구에서 정규 크리깅에 가장 적합한 반분산도 모형은 가장 작은 사분위 범위 값과 분산 값을 보인 지수(Exponential)모형로 해당 모델을 적용한 정규 크리깅을 통해 풍력발전소 좌표에 대한 풍속(ip_wind_speed), 풍향(ip_wind_direction), 습도(ip_wind_humidity), 기온(ip_temperature)을 추정하고 발전량 데이터와 결합하여 생성된 데이터와 발전량 간의 피어슨 상관관계를 분석한 결과 기존 염산 AWS 지점 상관계수 대비 풍속, 풍향 모두 0.05 만큼 향상되었음을 확인하였다.

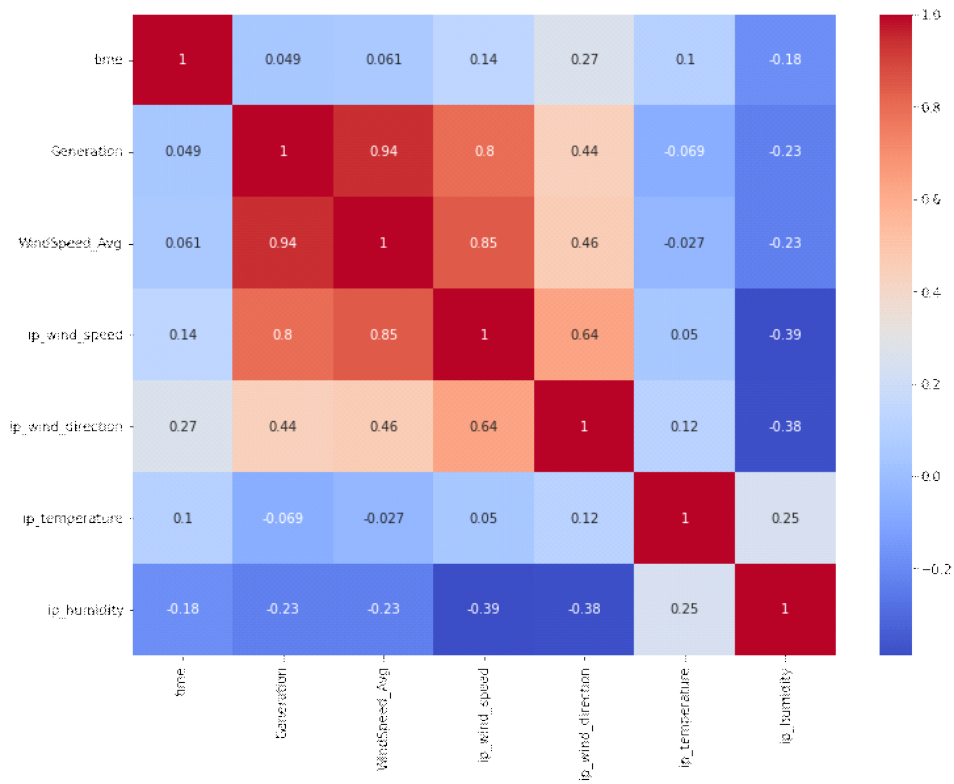


Fig. 16. Pearson's correlation heat map for kriging interpolation dataset

Table 5. Pearson correlation coefficient between major variables and power generation

	Yeonggwanggun ASOS data	Yeomsan AWS data	Ordinary Kriging interpolation data
Wind speed	0.67	0.72	0.80
Wind direction	0.37	0.39	0.44
Temperature	-0.062	-0.056	-0.069
Humidity	-0.2	-0.22	-0.23

3-3. 예측 모델링을 위한 변수 탐색 및 선정

3-3-1. 기상 파생 변수 탐색 및 추가

우선 선행 연구 대상 분석에서 확인한 이상치(고장, 진단, 유지보수 등)를 제외하고 나셀 측정 풍속 값을 제거하여 기본 데이터셋(Base Dataset)을 생성한다.

풍력발전 예측에 가장 중요한 변수는 풍속으로 지상기상관측장비는 10m 높이에서 수집된 자료이기 때문에 정확한 풍력 발전량 예측을 위해서는 연구 대상의 허브 높이인 100m 고도에서의 풍속이 필요하다. 관측된 고도의 풍속을 활용하여 다른 고도의 풍속을 예측하는 방법으로는 디컨방정식을 사용하였다. 연구 대상 주변 지형에 따른 지면 거칠기 값을 계산하고 10m 풍속에 대입 및 보정하여 100m 고도의 풍속 변수(100m_wind_speed)를 생성하였다.

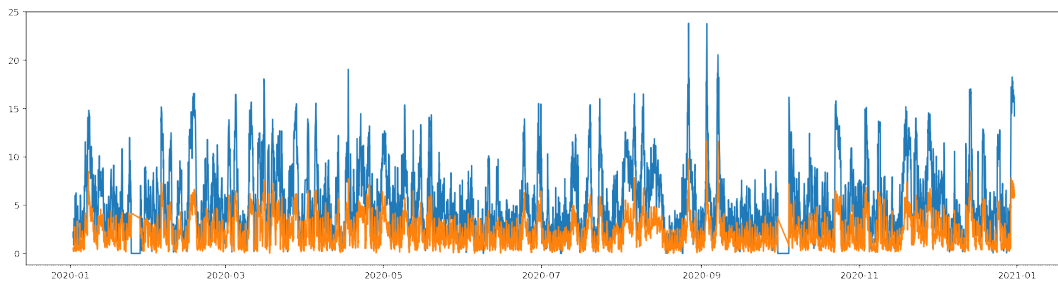


Fig. 17. Comparison of ordinary kriging interpolation wind speed (10 m) and nacelle wind speed

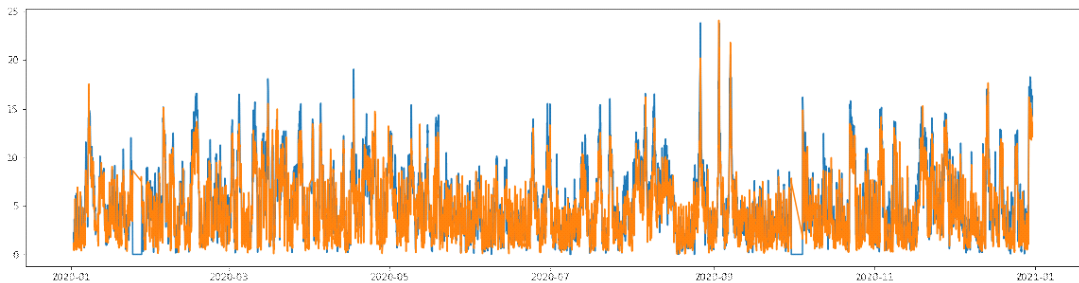


Fig. 18. Comparison of ordinary kriging interpolation wind speed (100 m) and nacelle wind speed

우리나라는 대륙과 해양의 중간지역의 반도 지형으로 계절적으로 풍향을 달리 하는 몬순(Monsoon)기후로 계절풍의 영향을 분석하기 위하여 계절 변수(season)를 추가하고 추정된 풍향, 풍속, 발전량을 바탕으로 계절별 풍배도를 도식화하였다.

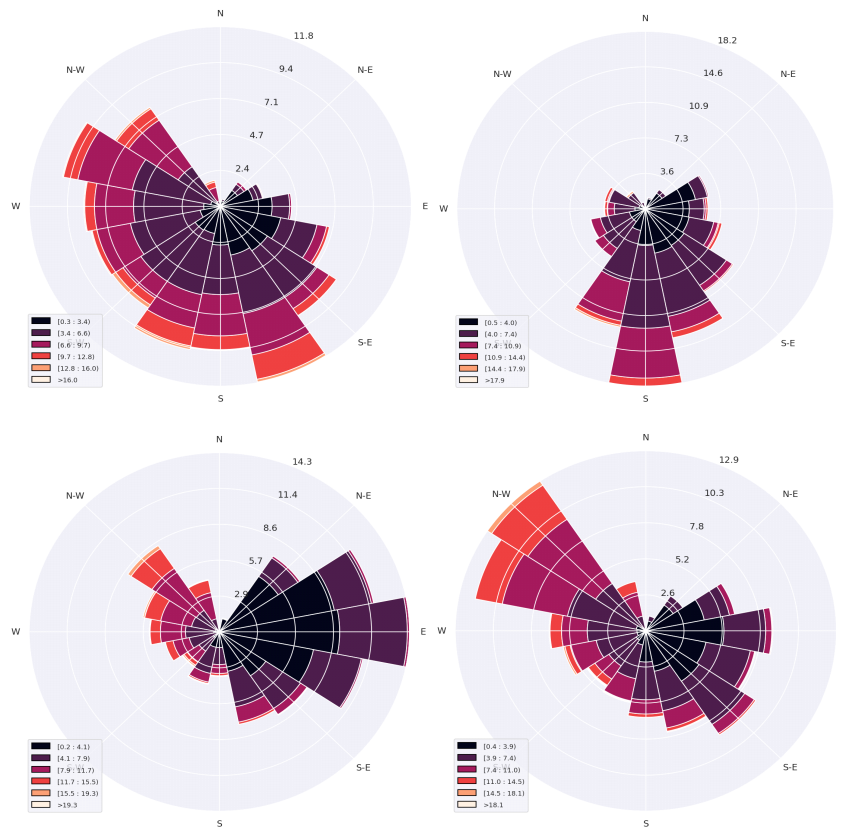


Fig. 19. Wind roses by Season

그 외 바람은 방향과 크기를 가지는 벡터이기에 풍향, 풍속 데이터를 활용한 바람 벡터(wind_cos_vector, wind_sin_vector)값을 생성하고 풍력 발전량은 출력 공식에 따라 풍속의 3제곱에 비례하므로 풍속의 세제곱 변수(windspeed³)를 생성하였다.

3-3-2. 시계열 파생 변수 탐색 및 추가

시간적 특성을 확인하기 위하여 월(month), 일(day), 시간(time) 변수를 생성하고 삼각함수를 통해 해당 변수의 연속형 변수(con_hour, con_time)를 생성하였다.

예측 목표인 풍력 발전량 데이터는 시간에 순차적으로 관측한 값들의 집합인 시계열 데이터로 과거 데이터의 패턴을 기반으로 미래의 발전량을 예측하는 시계열 기반 예측 알고리즘을 적용하기 위해서는 과거의 특정한 주기성을 가진 패턴이 미래에도 지속된다는 데이터의 안전성이 확보되어야 한다. 발전량에 대한 시계열 패턴을 파악하기 위하여 python의 statsmodels 라이브러리를 사용하여 가법 모형(additive model) 가정하에 시계열 분해(Time Series Decomposition)를 Fig. 18.과 같이 도식화하고 분석하였다.

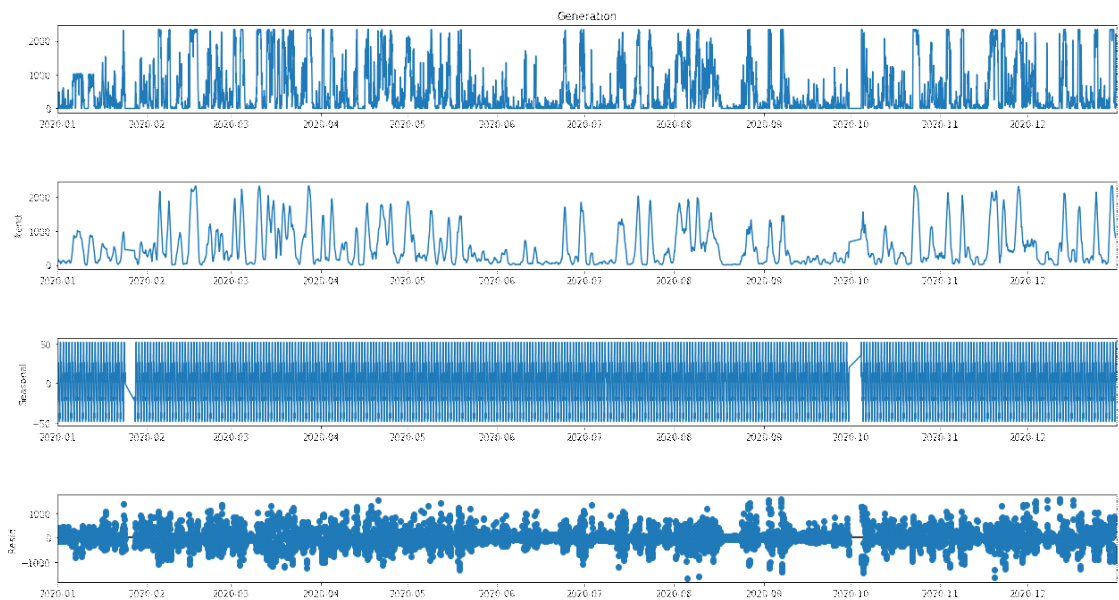


Fig. 18. Time series decomposition chart based on additive model of power generation data

우선 추세(Trend)와 계절성(Seasonality)을 제외한 잔차(Residual)가 특정 패턴 없이 불규칙한 분포를 띄고 강한 계절성과 불규칙한 추세를 보이기에 시간적 정상성(stationarity)은 낮다고 할 수 있다.

3-3-3. 최종 입력 변수 결정

정규 크리깅을 통해 추정한 기상정보를 바탕으로 탐색적 데이터 분석(Exploratory Data Analysis)을 통해 발전량 예측 모델링에 적용한 최종 변수를 확정하고 피어슨 상관계수를 Fig. 19와 같이 히트맵 형태로 도식화하였다.

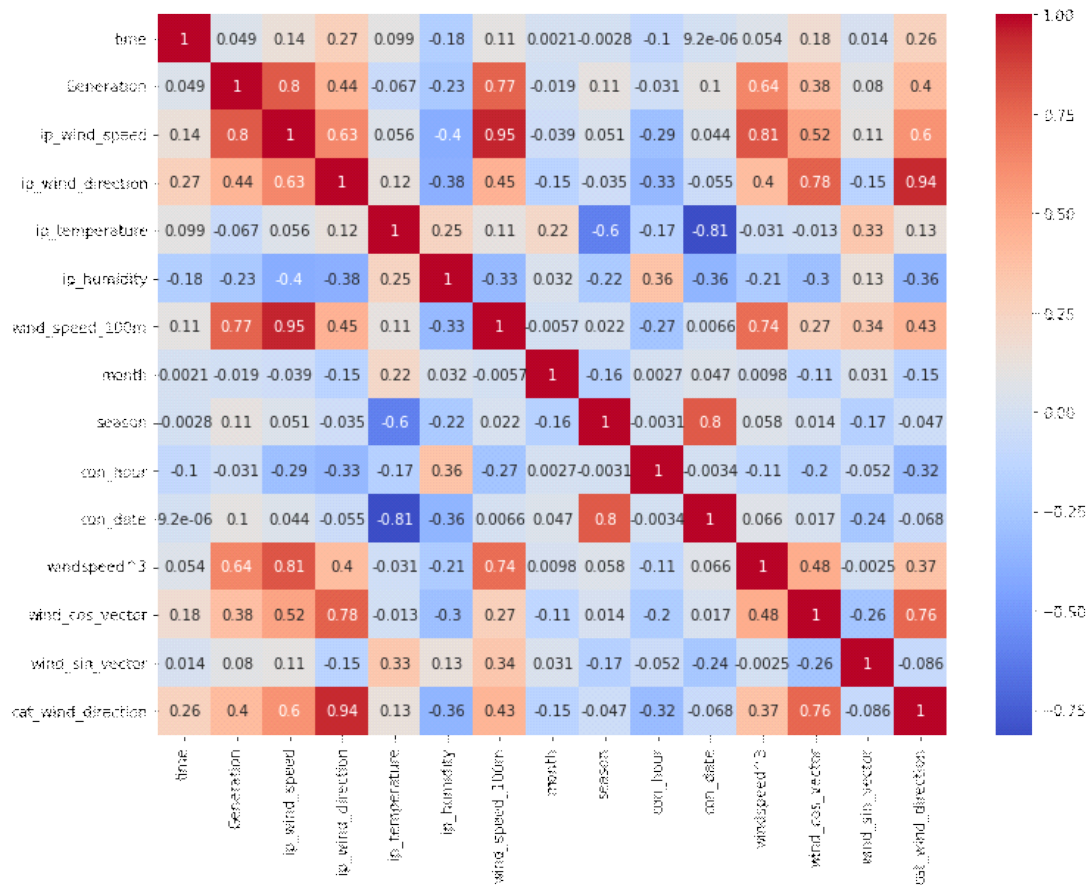


Fig. 21. Pearson's correlation heat map for the input variables

3-4. 발전량 예측 모델 학습 알고리즘 선정

3-4-1. 모델 예측 평가 선정 및 평가

시계열 분해 분석 결과 약한 정상성을 보이고 있으므로 시계열 연관성 기반의 CNN, LSTM 등의 모델보다 주어진 기상 조건을 바탕으로 변수 간 연관성을 찾아가는 트리 모델 기반의 예측 모델을 활용하여 발전량 예측 모델을 설계하였다.

학습에 사용된 트리 기반의 학습 알고리즘인 의사결정나무(Decision Tree), 랜덤포레스트(Random Forest), XGBoost(eXtreme Gradient Boosting), LGBM(Lighten Gradient Boosting Machine)이다.

모델링을 위해 생성한 데이터셋을 8:2로 학습, 검증데이터로 분리하고 타겟 값(발전량)과 입력 값을 분리하여 학습 및 평가 데이터셋을 생성하였다.

트리 기반의 4개 알고리즘을 바탕으로 예측 모델을 학습하였고 이때 초매개변수는 기본값을 사용하였으며 평가 방법은 K겹 교차검증(K-fold cross validation)을 통해 상호비교하였다. K겹 교차검증이란 데이터셋 전체를 K개의 부분 집합으로 분류하고 집합을 체계적으로 순환하면서 모든 데이터에 대한 모형의 성과를 검증하는 방식이다. 본 연구에서 K값은 10으로 설정하여 교차검증을 수행하였으며 각 학습 알고리즘에 대한 예측 모델 평가 결과는 Table 6.와 같다.

MODEL	R2_Score	RMSE
Decision Tree	0.657401	432.878391
Random Forest	0.823432	306.422256
XGBoost	0.812758	316.949053
LGBM	0.835142	295.627574

Table 6. Performance evaluation of learning models

K겹 교차검증을 통한 예측 결과, LGBM 알고리즘으로 학습한 예측 모델이 실제 측정값과의 상관관계를 나타내는 R2_Score도 다른 알고리즘의 예측 모형 대비 가장 높은 상관관계를 보였으며 실제값과 예측값과의 차이를 볼 수 있는 RMSE도 가장 작은 평균 제곱근 오차값을 보였다.

3-5. 매개변수 최적화를 통한 예측 모델 고도화

LGBM같은 트리 기반의 앙상블 알고리즘은 과적합 방지 및 모델의 정확도 향상을 위하여 트리 분기와 관련된 초매개변수의 조율로 모델의 정확도를 개선할 수 있다.

category	parameter	default	description
Boost Parameter	learning_rate	0.1	learning rate range: 0 to 1
	n_estimators	100	Number of weak learners to create
	max_depth	3	Tree Maximum Depth Apply values from 3 to 15
	min_child_samples	20	Minimum number of records to be the final leaf node
	num_leaves	31	Maximum number of leaves a tree has
	subsample	1	Data Sampling Ratio to Avoid Overfitting If the ratio value is 0.5, half of the data is used to create the tree Range: 0 to 1
	colsample_bytree	1	Random sampling of variables required for tree creation
	reg_alpha	0	L1 Regulazation Applied Value
	reg_lambda	1	L2 Regulazation Applied Value
Learning Task Parameter	objective	reg:linear	Definition of loss function
	eval_metric	rmse, error	Defining a Learning Test rmse, mae, logloss, merror, mlogloss, auc
	eval_set	None	Defining a Learning Test

Table 7. Types and definitions of hyperparameters of LGBM

최적의 예측 모델을 찾기 위해선 모델 검증 결과를 확인하고 초매개변수를 변경하여 재학습하고 검증하는 과정을 반복해야 한다. 선형회귀 알고리즘처럼 파라미터의 개수가 적거나 이진값으로 입력받는 경우 직접 파라미터를 수정해가며 모델을 검증할 수 있지만 LGBM은 Table 6.에서 보듯이 초매개변수의 개수가 많고 정수형 또는 실수형으로 인자를 받게 되면 직접 파라미터를 변경하며 검증하기가 어렵기 때문에 초매개변수 변경, 학습, 검증 과정의 최적화 과정을 용이하게 해주는 라이브러리를 활용할 필요가 있다.

본 연구에서는 파이썬 라이브러리인 Optuna를 활용하여 초매개변수 조정, 학습, 검증을 통해 최적의 초매개변수 값을 선정하였다.

초매개변수 최적화를 위하여 한 사이클당 10,000회 기준 과적합이 되지 않는 범위 내(early_stopping_rounds 값: 25)에 RMSE 값이 가장 적은 모델을 선정하도록 100사이클을 반복하였다.

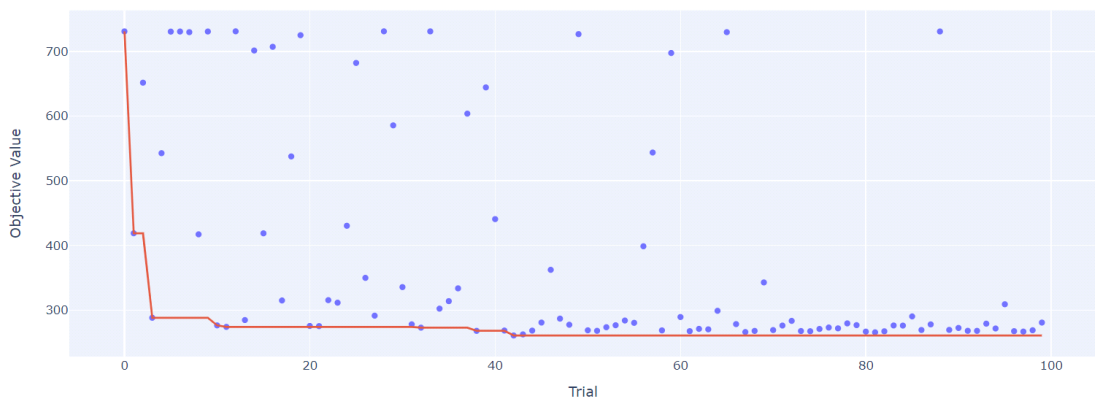


Fig. 22. Hyperparameter optimization history plot

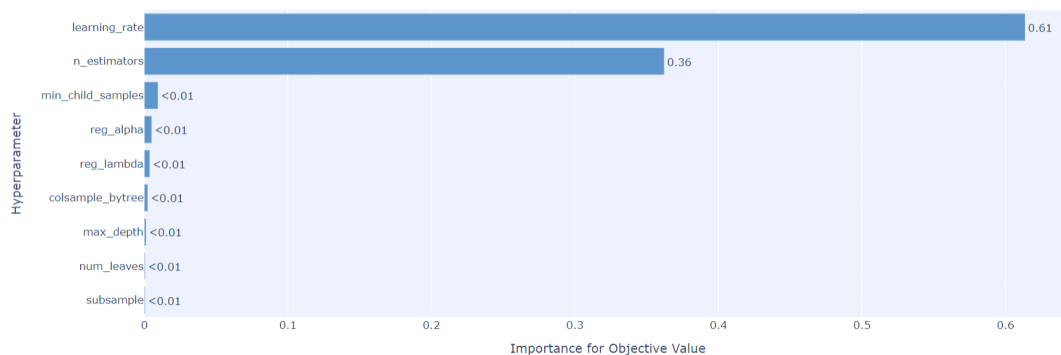


Fig. 23 Hyperparameter importance

최적화된 초매개변수 값은 Table 8.과 같다.

hyperparameter	range	result
num_leaves	2 ~ 1024	82
colsample_bytree	0.7 ~ 1.0	0.9536815308567923
reg_alpha	0 ~ 1	0.3216202770873735
reg_lambda	0 ~ 10	2.064957108666877
max_depth	3 ~ 15	13
learning_rate	$1 \times 10^{-8} \sim 1 \times 10^{-2}$	0.006273748227133646
n_estimators	100~3000	1,719
min_child_samples	5~100	15
subsample	0.4~1	0.5065056702148196

Table 8. Hyperparameter optimization range and result values

최적화된 초매개변수 값을 기준으로 LGBM모형을 재학습하고 동일 예측 기간의 기상정보를 입력 후 결과값을 비교한 결과 최적화 모델의 RMSE 값은 277.6160513685799로 기본 LGBM 알고리즘 적용 예측 모델보다 6.1%(18.01)가량 향상된 성능을 보였다.

IV. 발전량 예측 결과 분석

본 연구를 통해 풍력발전 단지의 기상정보가 제공되지 않는 경우 인근 지상기상관측장비에서 측정된 기상정보를 정규 크리깅을 통해 공간보간하여 자료를 수집하고 해당 데이터를 이용한 시간당 발전량 예측 모델을 개발하였다.

근접한 기상정보보다 정규 크리깅을 활용한 공간 보간을 통한 기상정보를 활용한 예측 모델의 성능을 검증하기 위하여 풍력발전기와 가장 인접한 염산AWS, 영광군 ASOS의 기상정보를 활용하여 동일한 예측 모델링 개발 과정을 거쳐 3개의 모델을 개발하였다. 이때 염산AWS 및 영광군ASOS 기상 데이터의 결측치는 시계열 기반 선형 보간법으로 해당 시간대 기상 데이터를 생성하였다.

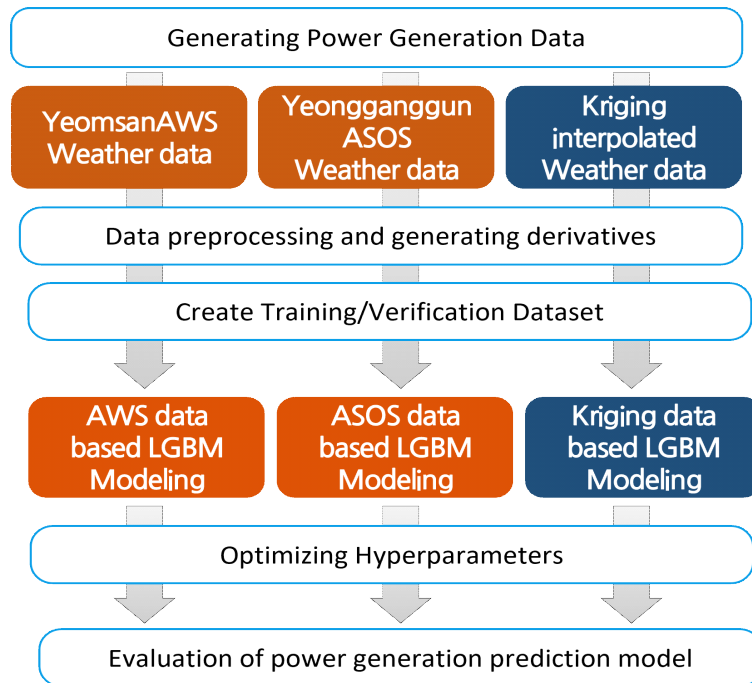


Fig. 24. Evaluation process of the prediction model based on weather data

학습된 3개의 모델의 R2_Score와 RMSE를 분석한 결과는 Table 9.에서 보듯이 정규 크리깅을 활용한 기상정보를 활용한 모델이 가장 좋은 성능을 보여주었다.

Table 9. Performance evaluation of predictive model based on weather data(AWS, ASOS, kriging interpolated data)

MODEL	R2_Score	RMSE
Yeomsan AWS data based LGBM Model	0.803854371647911	316.217502788815
Younggwanggun ASOS data based LGBM Model	0.745002967415126	360.54862077901055
Ordinary Kriging interpolated data based LGBM Model	0.8488192153689325	277.6160513685799

영광백수풍력발전단지 1번 풍력발전기 2020년 연간 발전량 예측 결과를 수집된 기상 데이터 종류별로 도식화하면 다음과 같다.

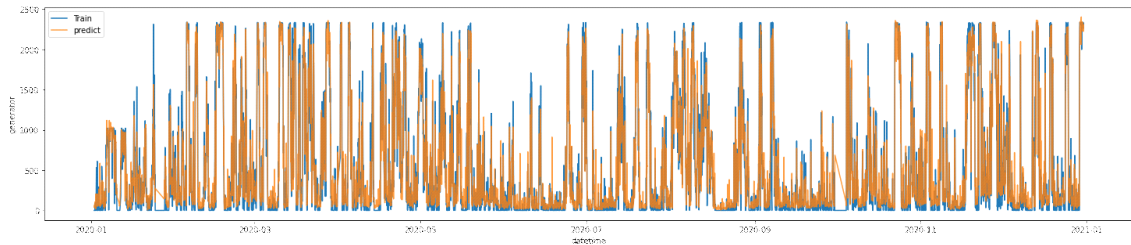


Fig. 25. Chart of hourly power generation prediction result based on weather data of Yeomsan AWS

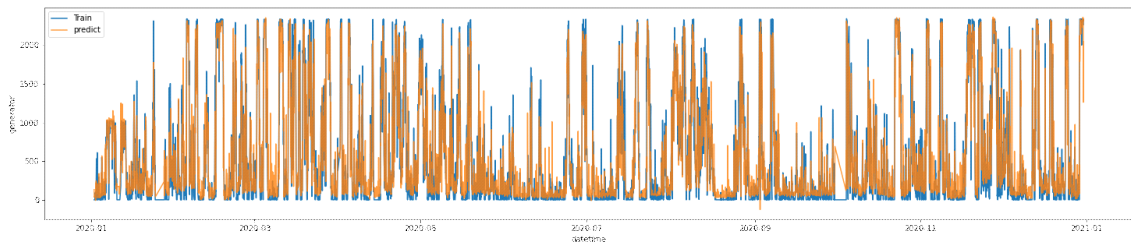


Fig. 26. Chart of hourly power generation prediction result based on weather data of Yenggwanggun ASOS

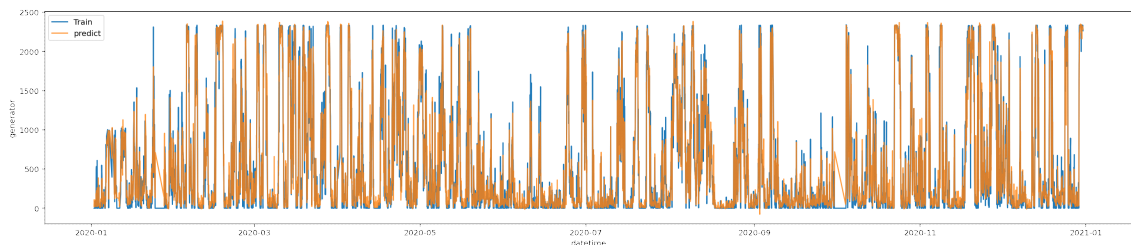


Fig. 27. Chart of hourly power generation prediction result based on ordinary kriging interpolated weather data

V. 결론

본 연구에서는 공공데이터포털에서 확보할 수 있는 풍력발전기의 발전량 정보와 발전소 인근 기상관측장비(AWS, ASOS)의 기상정보를 활용하여 정규 크리깅(Ordinary Kriging)기법을 통해 발전소 인근 기상정보를 추정하고 이를 바탕으로 트리 기반 기계학습 알고리즘으로 풍력발전기의 시간당 발전량을 예측하는 모델을 개발하는 연구이다.

(1) 연구 대상인 영광백수풍력단지의 가장 가까운 지상기상관측장비인 염산 AWS와 영광군ASOS의 풍속과 발전량 간의 피어슨 상관계수는 각각 0.75, 0.69이며 풍력발전기 인근 12개의 기상관측장비의 풍속을 정규 크리깅 기법으로 추정한 발전소 위치 기준 내삽 풍속과의 상관계수는 0.8로 기존 대비 0.05 만큼 향상됨을 확인하였다.

(2) 데이터 분석을 통해 트리 기반 기계학습 알고리즘(의사결정나무, 랜덤포레스트, XGBoost, LGBM)을 선정하여 예측 모델을 개발하고 K겹 교차검증을 통해 RMSE가 295.63으로 가장 낮았던 LGBM 알고리즘을 사용하여 발전량 예측 모델을 개발하고 LGBM 모델의 정확도를 향상시키기 위하여 초매개변수를 조정하면서 학습 모델을 최적화하였고 그 결과 기존 LGBM 기본 설정 예측 결과 대비 6.1% 향상됨을 확인하였다.

(3) 최종적으로 연구 대상에서 가까운 염산AWS, 영광군ASOS의 기상정보와 정규 크리깅으로 보간한 기상정보를 바탕으로 LGBM 기반 발전량 예측 모델 학습하고 테스트한 결과 정규 크리깅을 활용한 모델이 성능이 더 높은 것을 확인하였다.

본 예측 모델에 사용된 풍력발전 데이터 및 기상정보 데이터는 공공데이터 포털에서 확보된 제한적 데이터로 실제 풍력발전기 SCADA에서 확보할 수 있는 주요 부품의 상태 정보 등을 확보할 수 없었고 예측에 활용된 기상정보 역시 풍력발전소 인근의 10m 높이 복합 지형의 지상기상관측장치에서 계측된 기상정보를 바탕으로 추정된 정보로 정확한 기상정보 및 발전량 예측에는 한계가 있었다.

본 연구의 경험을 바탕으로 실제 풍력 풍력발전단지의 배치도를 기반으로 한 후류 손실 반영, 실제 운영 SCADA 데이터를 반영한 부품, 발전기 상태 정보 반영, 국지예보모델(LDAPS) 정보를 수집하여 공간, 시간 내삽기법을 활용한 고해상도 기상 예보 정보 반영 등 기상정보 및 발전량 예측 모델을 고도화하여 풍력발전량 예측 시스템을 개발하고 단기적으로는 발전량 예측 제도 참여를 위한 발전량 예측, 예측 발전량과 실제 발전량 차이를 통한 가동률 진단, 장기적으로는 단지 설계 이전 발전 수익 예측 검토 리포트 작성 등 다양한 형태로 활용하고자 한다.

참고문헌

- [1] Arias et al. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Technical Summary.
- [2] 외교부, 환경부. (2021). The Republic of Korea's Enhanced Update of its First Nationally Determined Contribution.
- [3] 산업통상자원부. (2020). 9차 전력수급기본계획(2020-2034).
- [4] Kim, D., Hur, J. (2016). 풍력발전 출력예측 기법 및 모형. *전기기술학회논문지*=*The Korean Institute of Electrical Engineers*, 65(12),12-18.
- [5] Ragheb, Magdi & Ragheb, Adam. (2011). Wind Turbines Theory - The Betz Equation and Optimal Rotor Tip Speed Ratio. 10.5772/21398.
- [6] Hyun, J., Park, H. & Chung, G. (2019). Effects of the Difference between ASOS and AWS Data on Runoff Characteristics. *Journal of The Korean Society of Hazard Mitigation*, 19(7), 443-449.
- [7] Choi, j. (2002). 공간정보 모델링-크리깅과 최적화 기법. 구미서관.
- [8] Woo, K., Park, J, & Lee, H. (2008). Sensitivity Analysis of Ordinary Kriging Interpolation According to Different Variogram Models. *Journal of the Computational Structural Engineering Institute of Korea*, 21(3), 295-304.

- [9] Hur, J., Park, B., & Jung, S. (2016). The Development of the Wind Power Prediction System using Ordinary Kriging. *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers. The Korean Institute of Illuminating and Electrical Installation Engineers*. 10.5207/jieie.2016.30.7.060.
- [10] Cheang, E., Moon, C., Jeong, M., Jo, K. & Park, G. (2010). The study for calculating the geometric average height of Deacon equation suitable to the domestic wind correction methodology. *Journal of the Korean Solar Energy Society*, 30(4), 9-14.
- [11] Geron, A. (2020). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow 2nd Edition. 한빛미디어.
- [12] Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3), 21 - 45.
- [143] Rokach, L., Maimon, O. (2005). Top-down induction of decision trees classifiers—a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 35(4), 476 - 487.
- [14] Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32.
- [15] Natekin, A., Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurobotics, Frontiers Media SA*,10.3389/fnbot.2013.00021.
- [16] Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System., *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

- [17] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems(NIPS)*, 3149 - 3157.
- [18] Bae, S, Kwon, Y, Moon, J, Nam, Y. Development of Performance Prediction Model for Water Source Heat Pump System based on Artificial Neural Network. *Korea Institute of Ecological Architecture and Environment*, 21(4), 99-104.
- [19] 공공데이터포털[한국동서발전(주)_영광백수풍력발전단지 1호기 10분 평균 발전량]
<https://www.data.go.kr/data/15091978/fileData.do>
- [20] 공공데이터포털[한국동서발전(주)_영광백수풍력 일일 시간대별 발전량]
<https://www.data.go.kr/data/15095657/fileData.do>
- [21] 기상자료개방포털[데이터:기상관측:지상:방재기상관측(AWS):자료]
<https://data.kma.go.kr/data/grnd/selectAwsRltnList.do>
- [22] 기상자료개방포털[데이터:기상관측:지상:중관기상관측(ASOS):자료]
<https://data.kma.go.kr/data/grnd/selectAsosRltnList.do>

감사의 글

지난 2년의 대학원 과정과 학위 논문을 마무리하며 지난 시간을 되돌아보니 직장 업무와 학업을 병행하는 힘들고 어려운 여정을 완주할 수 있도록 힘이 되어준 인연이 정말 많다는 것을 다시금 깨닫게 됩니다.

학문적 기초부터 연구에 대한 자세 하나하나 지도하여 주시고 산업 동향과 향후 연구와 사업화를 위한 인적/물적 네트워크까지 소개해주셨던 지도교수이신 김남진 교수님께 정말 감사드립니다. 항상 응원합니다.

기업 연구원으로서의 진로와 방향성을 고민하고 있을 때 대학원 진학을 추천해 주신 강영준 대표님, 많은 조언과 응원을 해주신 김경록 이사님께 이 이 자리를 빌려 감사의 말씀을 전합니다. 그리고 대학원 과정 전반을 함께하고 대학원 과정 전반 하나하나 가이드해준 연구실 강명보 선배님과 대학원 동기 민혁 씨에게도 감사의 인사를 전합니다. 그리고 마지막까지 논문 심사와 많은 조언을 해주셨던 강창남 교수님, 박윤철 교수님, 고진환 교수님께 감사드립니다.

직장 업무와 학업을 병행할 수 있도록 배려해주신 송기택, 부용혁 대표님과 논문 작성에 필요한 풍력 관련 정보를 제공해주었던 손진혁 책임연구원을 비롯한 연구소 동료들에게도 진심으로 감사한 마음을 전합니다.

대학원 진학을 응원해주시고 자기 일처럼 기뻐하셨던 조부모님과 어머니 그리고 무엇보다 에너지 넘치는 두 아들을 돌보면서도 항상 응원해주었던 세상에서 가장 사랑하는 아내 은주와 항상 아빠와 함께 놀고 싶다고 투정 부리던 두 아들 연우, 건우에게도 미안함, 고마움과 사랑을 전합니다.

지면에 다하지 못할 정도로 응원해주고 도움을 주셨던 사람들에게 감사를 표하며 이렇게 좋은 인연들이 내 자산이자 앞으로 나아갈 수 있는 원동력임을 다시 한번 느끼게 됩니다. 더 정진하는 모습으로 보내주신 응원과 사랑에 보답하겠습니다. 감사합니다.