

# 기상자료 예측을 위한 Facebook's Prophet 모델 적용

## Application of Facebook's Prophet Model for Forecasting Meteorological Data

김준석\* · 김태진\*\*

Kim, Junsuk\*, and Kim, Tae Jin\*\*

### Abstract

The wildfire risk index was calculated based on current meteorological information, for example, temperature, humidity, and wind speed. Thus, meteorological data forecasting could help estimate the probability of fire occurrence or spreading speed to prevent large wildfires. This study predicts meteorological data (e.g., temperature, humidity, and wind speed) using Facebook's Prophet library. We trained the Prophet model using meteorological data between 2016 and 2018 in Goseong, Gangwon-do (where the wildfire occurred in 2019) and predicted meteorological data for the first four months in 2019. We obtained that Facebook's Prophet model was effective in computing speed and predicting the overall trend. However, it could not predict sudden irregular changes satisfactorily. Considering its rapidity, these results could play an important role in future research, especially as a basic research for time-series forecasting.

**Key words :** Time Series Forecasting, Wildfire, Temperature, Wind, Humidity

### 요 지

지금까지 산불 발생 위험지수는 온도, 습도 및 풍속 등의 기상인자에 근거하여 산정되고 있다. 이에 따라 정확한 기상인자 예측은 향후 산불 발생 확률 감소 및 확산 가능성 예방하는데 필요한 상황이다. 본 연구에서는 현재 데이터 기반으로 신속성에서 향상된 시계열 예측 성능을 보여주고 있는 Facebook's Prophet 모델을 적용하여 온도, 습도, 풍속의 기상자료의 학습 및 예측에 대한 연구를 수행하였다. 2019년 산불이 발생한 강원도 고성의 2016-2018년 기상자료를 통한 모델 학습과 2019년 4개월간의 기상자료에 대한 예측에 대한 평가를 진행하였다. 적용 결과 Prophet 모델은 신속한 예측이 가능하며 기상자료의 전체적인 패턴 및 온도 인자에 대한 예측은 정확성을 보여주고 있지만 불규칙적인 급격한 변화가 있는 습도 및 풍속 인자에 대한 예측에 대한 정확성은 적은 것으로 판단되었다. 본 연구 결과는 기상인자 예측에서 정확성과 신속성의 두 가지 관점을 고려할 때 활용 가능한 모델에 대한 기초자료로 활용될 수 있을 것이다.

**핵심용어 :** 시계열 예측, 산불, 기온, 풍속, 습도

## 1. 서 론

산불은 자연발화, 인간의 활동에서 발생하는 실화 등 다양한 요인에 의하여 시작되지만, 산림에 피해를 줄 정도로 확산되기 위해서는 연소 환경을 구성하는 기상 인자의 영향에 의하여 결정된다(Cunningham and Martell, 1973; Won

et al., 2006; Kang et al., 2016). 일반적으로 산불의 확산과 기상인자는 상관관계가 있는 것으로 인식되고 있으며 특히 국내외적으로 산불위험지수는 습도, 풍속, 기온 등의 인자를 기반으로 산정되고 있는 상황이다(Won et al., 2012; Schunk et al., 2013; Kong et al., 2017). 또한, 현재 주로 사용되고 있는 산불위험지수는 관측된 기상 인자(예: 상대습도, 기온,

\*정회원, 동의대학교 산업ICT기술공학과 교수(E-mail: junsuk.kim@deu.ac.kr)

Member, Professor, Department of Industrial ICT Engineering, Dong-Eui University

\*\*교신저자, 정회원, 대구대학교 건설시스템공학과 교수(Tel: +82-53-850-5626, Fax: +82-53-850-6529, E-mail: kimtj@daegu.ac.kr)

Corresponding Author, Member, Professor, Department of Civil Engineering, Daegu University

풍속, 가뭄인자 등)에 따라 산불발생 확률 또는 위험정도를 계산하는데 초점이 맞춰져 있다(Lee et al., 2004). 하지만 이미 관측이 되어있는 기상 인자를 이용하여 산불 발생 위험도를 계산할 수 있다는 측면에서 매우 효과적이지만, 아직 관측되지 않은 앞으로 일어날 기상 변화에 따른 위험도 계산은 불가능하다는 단점이 있다. 만약, 산불 발생전의 기상인자의 예측이 가능하다면 산불의 발생 또는 확산 방지에 효과적으로 활용될 수 있을 것이다. 이에 따라 회귀 분석, 심층 신경망 등 다양한 기술을 활용한 시계열 예측 연구가 활발히 진행되고 있다(Yu et al., 2006; Kim et al., 2020). 하지만 다양한 인자를 바탕으로 시계열 예측이 진행되어 정확성은 향상되고 있지만 신속성에 대한 부분의 향상이 필요한 상황이다.

본 연구에서는 현재 시계열 예보시스템 등의 인위적인 전략을 사용하지 않고 데이터만을 활용하여 훌륭한 시계열 예측 성능을 보여주는 Facebook의 시계열 분석 라이브러리 Prophet (Taylor and Letham, 2018; AshaEmail et al., 2020; Thiyagarajan et al., 2020)을 활용하여 기상 변화를 예측하는 연구를 수행하였다. 본 연구에서는 2019년 산불이 발생한 고성 지역의 실제 관측 기상자료인 온도, 습도, 풍속 데이터를 기반으로 Facebook's Prophet 모델을 이용하여 예측 정확도를 평가하였다. 강원도 고성군 산불은 2019년 4월 4일 19시 17분에 특고압 전선 아크 불티를 통하여 발생하였으며 총 12.67 km<sup>2</sup>를 태우고 4월 5일 18시 00분에 진화가 완료되었음을 확인하였다(KFS, 2020).

## 2. 방법론

### 2.1 Prophet 모델

기존의 예측 모델은 시간적 순서가 존재하는 시계열 데이터를 분석하기 위해서 규칙적인 패턴을 파악하여 모델링하는 방법이 주로 사용되어 왔다(Cho et al., 2019). 특히, 시계열 분석을 통한 예측에서는 수치 모델링, 통계적 모델링 등이 사용되어 왔으며 현재 인공지능 기술을 포함한 다양한 종류의 시계열 예측 기법이 제안되고 있다. Kim et al. (2020)은 기존의 다중 회귀분석, 주성분 회귀분석뿐만 아니라 인공지능 기술인 심층신경망(DNN; Deep Neural Network)을 사용하여 습지의 수위 예측모형을 개발하였다. 또한 다양한 예측 모형간의 예측 정확도를 비교하여, 심층신경망을 활용한 수위 예측모형이 가장 뛰어난 예측 성능을 보이는 것을 보고하였다.

본 연구에서 활용되는 Facebook's Prophet 모델(Taylor and Letham, 2018; AshaEmail et al., 2020; Thiyagarajan et al., 2020)은 기본적으로 3개의 요소를 가지는 가산모델(additive model)로 구성되어 있다(Eq. (1)).

$$y(t) = g(t) + s(t) + h(t) + e \quad (1)$$

여기서,  $g(t)$ : 반복적인 요소가 없는 트렌드;  $s(t)$ : 반복적인 요소를 가지고 있는 변화;  $h(t)$ : 불규칙적으로 영향을 미치는 이벤트;  $e$ : 정규분포 에러 Prophet 모델은 먼저 시간 데이터의 각종 특징을 임베딩(embedding)하여  $s(t)$  추정을 하고, 나머지 데이터를 사용하여 구간별 선형회귀분석을 진행한 후 최종 모델을 완성한다. Prophet 모델의 특징은 기본적으로 가산모델을 따르고 있지만 이벤트(event) 등 새로운 외부 변화가 있을 때 모델을 쉽게 다시 훈련시킬 수 있다는 점에 있다. 그리고 기존의 확률론적이고 이론적인 모델 기반 접근법인 ARIMA (AutoRegressive Integrated Moving Average) 모델과 다르게 경험적 규칙(heuristic rule)을 사용하고 있다. 또한, 시계열 데이터의 정상성, 정규성 등과 같은 엄격한 통계적 요구가 필요하지 않으며, 결측치들을 모델에 포함시킬 필요가 없다. 뿐만 아니라, 인공지능 기반의 시계열 예측과 비교하여 모델링에 소요되는 시간이 적다는 장점이 있다.

### 2.2 모델 학습 및 예측

분석에 사용된 기상데이터는 기상청 기상자료 개방포털(KMA, 2020)에서 제공하는 온도, 습도, 풍속 데이터를 사용하였다. 특히, 실제로 발생한 대형 산불 사례에 적용해보기 위하여 강원도 고성군 지역의 2016년 1월 1일부터 2018년 12월 31일까지 3년간의 관측자료를 이용하여 Prophet 모델 학습에 활용하였으며 2019년 1월 1일부터 2019년 4월 30일 까지의 4개의 관측자료를 통하여 예측자료에 대한 검증을 수행하였다. 즉, 일별 온도, 습도, 풍속 관측 기상자료를 사용하였으며 3년간의 예측 모델 학습 및 4개월간의 예측을 가지는 두 개의 데이터 시퀀스(sequence)가 사용되었다.

본 연구에서는 온도, 습도, 풍속 데이터를 각각 활용하는 일변량 예측분석을 진행하였으며 세 종류의 데이터를 함께 활용하여 예측하는 다변량 예측분석은 진행하지 않았다. Table 1에 Prophet 모델에 사용된 인자를 정리하였다.

**Table 1.** Parameters for Prophet Modelling

Parameter	Value
Growth	Linear
Changepoints	Not specified
N_changepoints	Not specified
Changepoint_range	1
Changepoint_prior_scale	0.1
Seasonality_prior_scale	10
Seasonality_mode	Additive
Yearly_seasonality	Auto
Weekly_seasonality	Auto
Daily_seasonality	Auto
Interval_width	0.8

Prophet 모델에 사용되는 인자 ‘growth’는 트렌드 모델링의 방식, ‘changepoints’는 트렌드 변화시점을 명시한 리스트 값, ‘changepoint\_prior\_scale’는 트렌드의 유연성 조절을 위한 값, ‘n\_changepoints’는 changepoint의 개수, ‘changepoint\_range’는 changepoint 설정 가능 범위를 의미한다.

### 2.3 모델 결과 검증

관측값과 예측값은 인공지능 알고리즘을 기반으로 시계열 예측 평가에 활용(Cho et al., 2019; Kim et al., 2020)되고 있는 Index of Agreement (IOA, Eq. (2)), Mean Error (Eq. (3)), Normalized Root Mean Squared Error (NRMSE, Eq. (4))를 이용하여 검증하였다.

$$IOA = 1 - \frac{\sum_i (|O_i - M_i|)^2}{\sum_i (|M_i - \bar{M}| + |O_i - \bar{O}|)^2} \quad (2)$$

$$Mean\ Error = \frac{1}{n} \sum_i (O_i - M_i) \quad (3)$$

$$NRMSE = \frac{100}{\bar{O}} \sqrt{\frac{1}{n} \sum_i (O_i - M_i)^2} \quad (4)$$

여기서,  $O_i$ : 관측값;  $M_i$ : 예측값;  $\bar{O}$ : 평균 관측값;  $\bar{M}$ : 평균 예측값; 및  $n$ : 자료수

IOA는 관측값과 예측값의 시계열 유사성을 나타내는 척도로 0과 1사이의 값을 가지며, 1에 가까울수록 관측값과 예측값이 일치함을 의미한다. Mena Error는 관측값과 예측값 간의 평균 편향을 나타내는 지표로서 0에 가까울수록 편향이 적음을 나타낸다. NRMSE는 평균제곱근 오차를 표준화한 값이며 0에 가까울수록 예측이 실제값과 유사하다는 것을 의미한다. 또한 추가적으로 두 개의 시계열의 상관관계를 수치화한  $r$  값을 계산하였다.

## 3. 모델 적용

### 3.1 예측 기간 자료 평가

Fig. 1은 Prophet 모델을 사용하여 온도, 습도, 풍속 시계열 데이터를 예측한 결과이다. 실제로 관측된 값은 검정색(2016.01.01~2018.12.31.)과 녹색(2019.01.01~2019.04.30.)이며, Prophet을 통해 예측한 값의 변화는 빨간색 실선으로 표현하였다. 또한 파란색 점선은 Prophet 모델이 예측한 트렌드와 함께 분포를 계산하여 얻어낸 불확실성 범위(Uncertainty Interval)를 나타내며, Prophet 라이브러리에서 기본으로 설정되어 있는 80%의 불확실성 범위를 보여주고 있다. Prophet을 통해 예측값과 관측한 값의 유사성을 정량

**Table 2.** Forecasting Performance Results

Metric	Data		
	Temp.	Hum.	Wind
IOA	0.78	0.38	0.36
ME	0.50	-5.02	0.12
NRMSE (%)	14.35	35.95	35.41

적으로 평가하기 위하여 IOA, Mena Error, NRMSE을 활용하였다. Table 2는 각각의 기상자료에 따른 예측 결과를 보여준다.

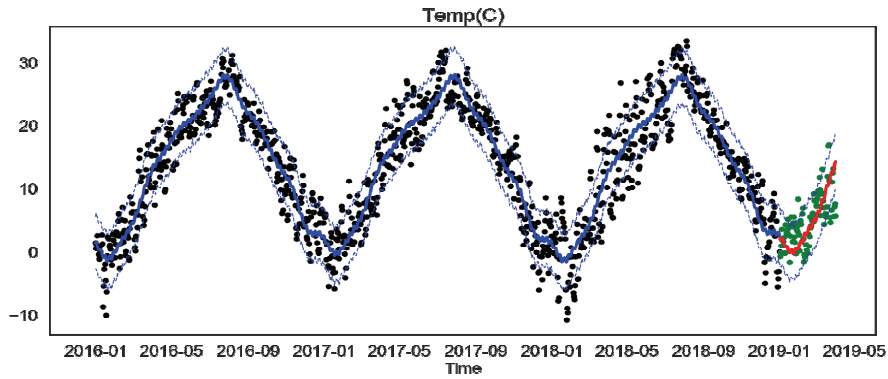
입력데이터 중에서 Prophet 모델은 기상 자료 중에 온도에 대한 예측은 가능하지만, 습도 및 풍속에 대한 예측은 기대치보다 좋지 않은 결과를 보여주었다. 상관관계 검증에서는 온도의 경우  $r = 0.64$ ,  $p < 0.01$ , 습도의 경우  $r = 0.29$ ,  $p < 0.01$ 로서 통계적으로 유의미한 유사성을 확인하였으며, 풍속의 경우  $r = 0.12$ ,  $p = 0.19$ 이었다.

Fig. 2는 각각의 기상데이터의 상관관계 분석 결과를 보여주며, x축은 실제 관측 값 y축은 Prophet 모델이 예측한 값을 나타낸다. 또한 파란색 실선은 상관계수, 즉  $r$  값이 1인  $y = x$  함수를 나타낸다.

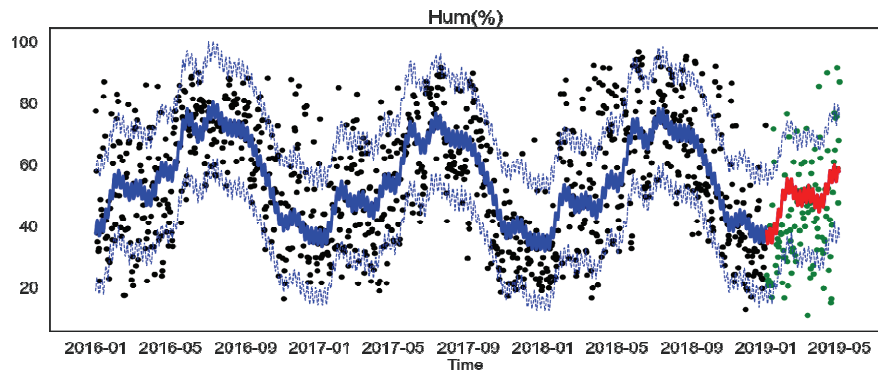
### 3.2 산불 발생 기간의 예측자료 평가

Kang et al. (2016) 연구에서 보고된 바와 같이 대형산불의 예보와 관련해서 풍속과 습도 변화를 예측하는 것은 매우 중요하다. 이 연구에서 제안한 대형산불 예보 기준에 따르면, 대형산불 주의보는 실효습도 30~45% 범위에서 2일 이상 지속되고 풍속이 7 m/s 이상일 때 발효하고, 대형산불 경보는 실효습도 30% 미만인 2일 이상 지속되고 풍속이 11 m/s 이상이 예상될 때 발효하는 것이 타당하다고 제안하였다. 이를 바탕으로 우리는 실제 대형 산불이 발생하였던 2019년 4월 4일 19시 17분부터 다음날인 5일 18시 00분까지의 실제 풍속과 습도를 획득하고, 또한 Prophet 모델을 통하여 예측한 풍속과 습도와의 차이를 살펴보았다.

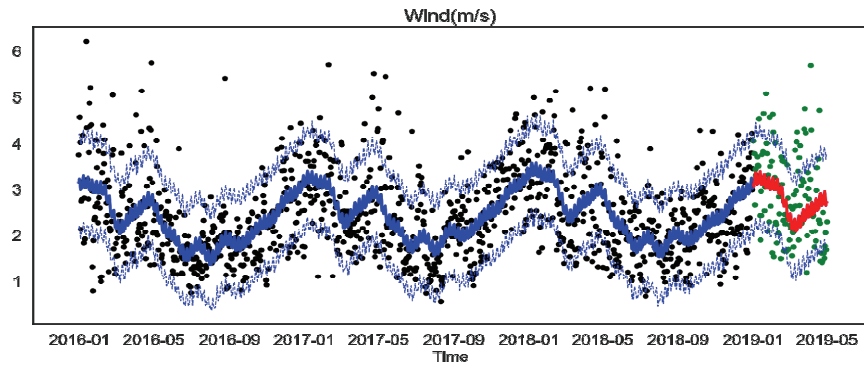
2019년 4월의 평균 습도는 49.56%, 평균 풍속은 2.51 m/s이었지만, 산불이 발생한 시간동안의 습도 값의 평균은 35.16%, 평균 풍속은 4.57 m/s이었다. 즉, 습도는 평소보다 매우 낮아 건조하고 풍속은 평소에 비하여 2배 가까이 높아 강한 바람이 발생하는 상황이었다. Fig. 3은 2019년 4월 4일 19시부터 다음날인 5일 18시까지의 실제 관측된 기상데이터 값(검정색 점)과 Prophet 모델이 예측한 값(파란색 실선)이다. 예측한 산불이 발생한 시간동안의 평균 습도는 46.19%로 관측값보다 31.4% 높게 예측되었으며, 평균 풍속은 2.64 m/s로 42% 낮게 예측되었다.



(a) Time Series Forecasting for Temperature Data Sequence



(b) Time Series Forecasting for Humidity Data Sequence



(c) Time Series Forecasting for Wind Speed Data Sequence

Fig. 1. Time Series Forecasting with Facebook's Prophet

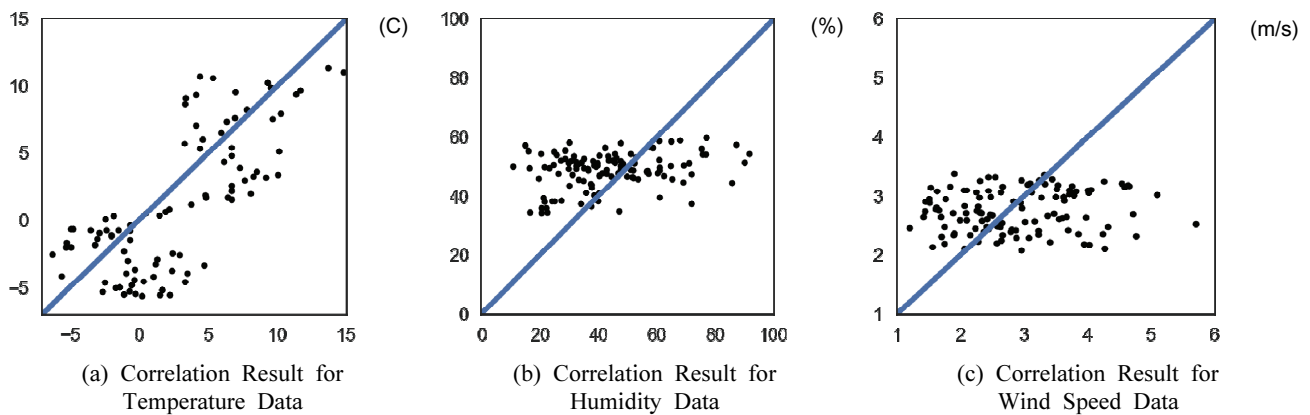
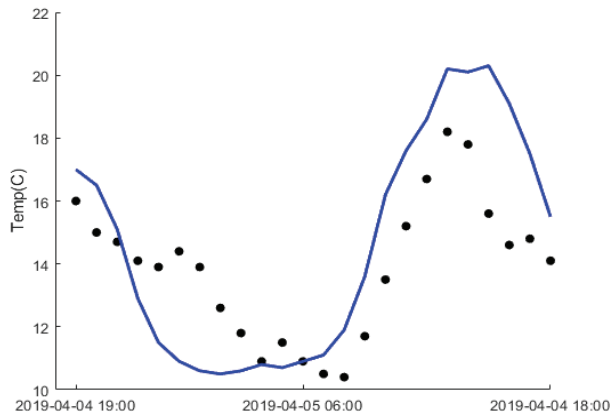
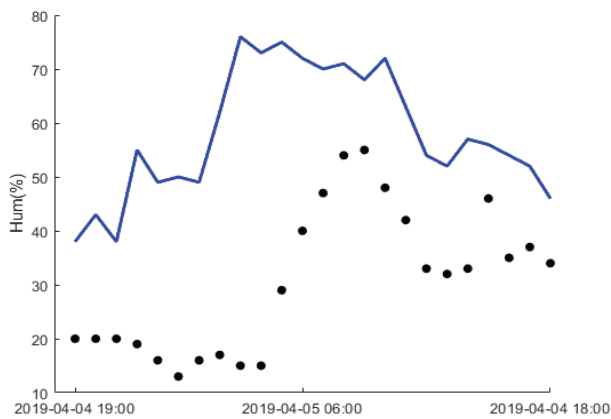


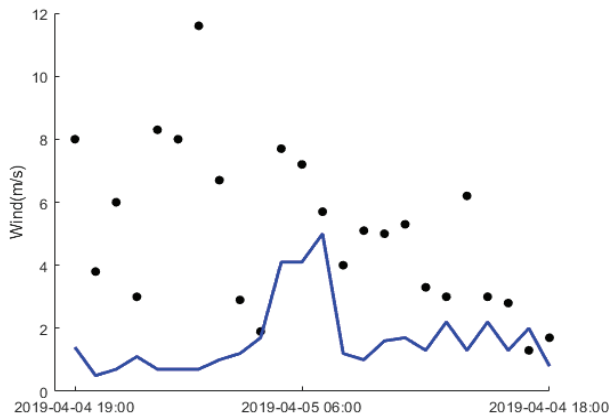
Fig. 2. Correlation Results



(a) Temperature Forecasting by Prophet  
(2019-04-04 19:00~2019-04-05 18:00)



(b) Humidity Forecasting by Prophet  
(2019-04-04 19:00~2019-04-05 18:00)



(c) Wind Speed Forecasting by Prophet  
(2019-04-04 19:00~2019-04-05 18:00)

Fig. 3. Prophet Forecasting Results

#### 4. 요약 및 결론

본 연구에서 Prophet 모델을 사용하여 강원도 고성군의 기상데이터(온도, 습도, 풍속)에 대하여 학습 및 예측을 수행하였다. 4개월간의 예측값은 관측값에 대하여 온도는 통계

학적으로 유사성이 보였지만 습도 및 풍속은 정확성이 온도 예측값에 비해 낮은 경향을 보여주었다. 특히, 산불발생기간인 2일 동안은 습도는 과다예측 및 평균 풍속은 과소예측을 하였다. 이에 대한 결과는 다음과 같이 정리할 수 있다.

첫째, 초단기 예측을 제외하면 풍속의 급격한 변화는 예측하기가 매우 어렵다. 인공지능 알고리즘인 순환신경망(Recurrent Neural Networks, RNN)을 사용하여 영종도 주위의 풍속을 예측한 연구(Kim and Woo, 2017)에서는 10~20분 후까지의 풍속의 초단기 예측 성능이 Mean Squared Error (MSE)는 0.5의 값을 보여 주었지만 20분 이후에는 MSE가 급격히 하락하는 것으로 보고하였다. 본 연구에서는 4개월 예측풍속으로 전체풍속의 패턴을 예측하였으나 2일간의 풍속예측은 기존의 연구와 유사한 결과를 보여주었다.

둘째, 산불발생에 중요 요소로는 습도와 온도보다는 상대적인 값의 변화가 더욱 중요하다. 기존의 Kang et al. (2016)의 연구에 따르면 산불이 대형화되는데 가장 큰 영향을 미치는 요소는 풍속과 함께 상대적인 습도의 변화량인 실효습도이다. 본 연구에서 Prophet을 이용하여 절대적인 습도 자체를 예측하였을 뿐, 상대적인 습도의 변화량을 예측하지는 않았다.

셋째, 정확도가 높은 예측을 위해서는 일변량 분석이 아닌 다변량 분석을 진행해볼 필요가 있다. Kim et al. (2019)의 연구에서 개발한 산불 위험지수를 살펴보면 온도, 습도, 풍속 등의 기상데이터 뿐만 아니라 강수량, 지역적인 특성, 식생건조 지수(VDI; Vegetation Dryness Index) 등을 종합적으로 고려하여 분석하여 연구를 진행하였다. 대형 산불 발생을 대비하기 위하여 개발된 산불위험지수는 현재의 기상데이터를 사용하여 계산하는 방식이므로, 아직 관측되지 않은 앞으로의 기상 변화에 따른 위험지수를 계산할 수 없다는 한계점을 보여주었다.

결론으로 본 연구에서 Facebook's Prophet 모델은 기상데이터 예측에서 전체적인 패턴 등을 모의 및 예측에는 효과적인 결과를 보여주었으나 풍속 등과 같은 단기간에 급격한 변화가 발생하는 기상 데이터의 예측에는 한계를 보여주었다. 간단하며 빠른 예측이라는 Prophet 모델의 장점이 분명히 존재하지만, 예를 들어 산불이 발생하고 확산하는 과정에서 진화를 위해 실시간 예측을 수행해야 하는 경우에 사용하기에는 정확도가 떨어진다는 단점 역시 존재한다. 따라서 향후 Facebook's Prophet 모델의 한계점을 극복할 수 있는 ARIMA, 신경망 모델 등과 같은 시스템 적용과 함께 적용하는 것이 모델의 정확성을 향상시킬 수 있을 것이다. 추후 연구로는 Prophet 모델을 구성하는 세 가지 요소 중에서 'h(t)', 즉 불규칙적으로 영향을 미치는 이벤트에 중점을 맞추어 학습을 진행하고 24시간 이내의 단기 예측, 실효습도의 예측과 함께 산불 발생 사례와 연관성 분석 및 풍속 및 습도 등의 개별 변수를 사용하여 예측하는 것이 아니라 개별 변수의 상관관계를 학습에 활용하는 다변량 분석을 수행하는 것을 진행할 것이다.

## References

- AshaEmail, J., Rishidas, S., SanthoshKumar, S., and Reena, P. (2020). Analysis of temperature prediction using random forest and facebook prophet algorithms. *International Conference on Innovative Data Communication Technologies and Application*, pp. 432-439.
- Cho, K., Lee, B.Y., Kwon, M., and Kim, S. (2019). Air quality prediction using a deep neural network model. *Journal of Korean Society for Atmospheric Environment*, Vol. 35, No. 2, pp. 214-225.
- Cunningham, A.A., and Martell, D.L. (1973). A stochastic model for the occurrence of man-caused forest fires. *Canadian Journal of Forest Research*, Vol. 3, No. 2, pp. 282-287.
- Kang, S.C., Won, M.S., and Yoon, S.H. (2016). Large fire forecasting depending on the changing wind speed and effective humidity in Korean red pine forests through a case study. *Journal of the Korean Association of Geographic Information Studies*, Vol. 19, No. 4, pp. 146-156.
- Kim, D.H., and Woo, S. (2017). A study on the prediction of wind speed in Yeongjong Island using deep learning. *Proceedings of the Autumn Meeting of Korean Meteorological Society*, pp. 342-343.
- Kim, D., Kim, J., Kwak, J., Necesito, I.V., Kim, J., and Kim, H.S. (2020). Development of water level prediction models using deep neural network in mountain wetlands. *Journal of Wetlands Research*, Vol. 22, No. 2, pp. 106-112.
- Kim, Y.H., Kong, I., Chung, C.Y., Shin, I., Cheong, S., Jung, W.C., Mo, H.S., Kim, S.I., and Lee, Y.W. (2019). Wildfire risk index using NWP and satellite data: Its development and application to 2019 Kangwon wildfires. *Korean Journal of Remote Sensing*, Vol. 35, No. 2, pp. 337-342.
- Kong, I., Kim, K., and Lee, Y. (2017). Sensitivity analysis of meteorology-based wildfire risk indices and satellite-based surface dryness indices against wildfire cases in South Korea. *Journal of Cadastre & Land Informatix*, Vol. 47, No. 2, pp. 107-120.
- Korea Forest Service (KFS). (2020). <http://www.forest.go.kr/> [accessible 02 September 2020]
- Korea Meteorological Administration (KMA). (2020). <https://data.kma.go.kr/cmmn/main.do#> [accessible 02 September 2020]
- Lee, S.Y., Han, S.Y., Won, M.S., An, S.H., and Lee, M.B. (2004). Developing of forest fire occurrence probability model by using the meteorological characteristics in Korea. *Korean Journal of Agricultural and Forest Meteorology*, Vol. 6, No. 4, pp. 242-249.
- Schunk, C., Wastl, C., Leuchner, M., Schuster, C., and Menzel, A. (2013). Forest fire danger rating in complex topography results from a case study in the Bavarian Alps in autumn 2011, *Natural Hazards and Earth System Sciences*, Vol. 13, No. 9, pp. 2157.
- Taylor, S.J., and Letham, B. (2018). Forecasting at scale. *The American Statistician*, vol. 72, no. 1, pp. 37-45.
- Thiyagarajan, K., Kodagoda, S., Ulapane, N., and Prasad, M. (2020). A temporal forecasting driven approach using facebook's prophet method for anomaly detection in sewer air temperature sensor system, *techrxiv*, doi:10.13140/RG.2.2.31367.14245
- Won, M.S., Koo, K.S., and Lee, M.B. (2006). An analysis of forest fire occurrence hazards by changing temperature and humidity of ten-day intervals for 30 years in spring. *Korean Journal of Agricultural and Forest Meteorology*, Vol. 8, No. 4, pp. 250-259.
- Won, M.S., Lee, M.B., Lee, W., and Yoon, S. (2012). Prediction of forest fire danger rating over the Korean peninsula with the digital forecast data and daily weather index (DWI) model. *Korean Journal of Agricultural and Forest Meteorology*, Vol. 14, No. 1, pp. 1-10.
- Yu, P.S., Chen, S.T., and Chang, I.F. (2006). Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*, Vol. 328, No. 3-4, pp. 704-716.

---

Received	August 12, 2020
Revised	August 14, 2020
Accepted	September 29, 2020