

# 인공지능 알고리즘 기반 미국 California 산불 규모 예측 모델에 대한 연구

김정현<sup>1</sup>, 윤동규<sup>2</sup>, 유귀호<sup>3</sup>, 고보경<sup>1</sup>, 조석현\*

<sup>1</sup> 경기대학교, <sup>2</sup> 계명대학교, <sup>3</sup> 한남대학교, \*University of California, San Diego (UCSD)  
jmmisung@gmail.com, ydg4977@gmail.com, 625thousand@gmail.com, gobk0507@gmail.com,  
\*justinshcho@gmail.com

## Study on Artificial Intelligence-based Wildfire Size Prediction Model in California, USA

Junghyun Kim<sup>1</sup>, Donggyu Youn<sup>2</sup>, Gwiho Yoo<sup>3</sup>, Bogyong Ko<sup>1</sup>, Seokheon Cho\*  
Kyonggi University<sup>1</sup>, Keimyung University<sup>2</sup>, Hannam University<sup>3</sup>,  
University of California San Diego (UCSD)\*

### Abstract

Various research on wildfires have been conducted because they lead to significant losses in both economic and environmental aspects. We proposed the prediction models based on Artificial Intelligence algorithms to predict wildfire size in advance. We referred to two public datasets, such as California Fire Perimeters and Total Weather Data, and used the newly generated dataset consisting of features selected from these public datasets to train the wildfire size prediction models. K-Nearest Neighbor (KNN), Random Forest (RF), and Multi-Layer Perceptron (MLP) algorithms were considered to establish the binary classification models to distinguish between a large wildfire and a small wildfire according to the wildfire size. Two different experiments were carried out, which depends on how to split our new dataset into training and test sets; the first experiment where the entire dataset is divided into two subsets regardless of the calendar year in which each wildfire occurred; the second experiment where it is divided in accordance with the year in which each wildfire occurred. Since the major purpose of our study is to predict the actual large wildfires in advance, Recall is the most important evaluation metric for the prediction models, which is the proportion of actual large wildfires that were correctly classified as megafires. It was observed that the first experiment and the MLP-based wildfire size prediction model result in the best performance in terms of Recall and F1-score.

### I. 서론

캘리포니아는 미국에서 경제적 및 환경적으로 중요한 농업과 기술 산업 분야에서 핵심적인 역할을 하고 있는 대표적인 주 (state)이다. 이러한 캘리포니아에서는 매년 다양한 재해가 발생하지만, 그 중에서도 산불은 막대한 경제적 손실과 환경적 피해를 초래하는 대표적 재해이다. 또한, 산불의 빈도와 강도 또한 증가하고 있는 추세이다. 캘리포니아에서는 2000년부터 2023년까지 매년 평균 6,500~7,000건의 산불이 발생했으며, 이로 인한 연간 평균 손실은 약 5,000억 달러에 이를 정도로 막심하다 [1]. 날씨 변화, 건조한 대기 및 강풍 등이 산불의 주요 원인이며, 날씨 데이터를 이용한 분석은 산불 대응 전략 수립에 필수적이다.

K. Pham *et al.*는 California Fire Perimeters 데이터 세트와 날씨 데이터세트를 이용하여 캘리포니아에서 발생하는 산불 예측 모델 개발을 위해 Artificial Neural Network (ANN)과 Support Vector Machine (SVM) 등의 알고리즘을 이용하였다 [2]. 해당 연구는 전체 데이터세트와 카운티 (county)별로 데이터세트를 나누어 예측 모델 성능 비교를 진행하였으며, 카운티별로 분류한 데이터세트에 대해 학습 시 산불 예측 정확도가 향상되는 것을 확인하였다. 이는 지역적 특성을 반영한 산불 예측 분석의 중요성을 강조하는 것이다. 캘리포니아 북부 지역 산불 데이터세트를 기반으로 산불 발생 예측 모델을 개발한 결과를 제시한

연구도 있다 [3]. 해당 연구에서는 날씨, 지형 그리고 식물의 건조 정도를 나타내는 정규 식생 지수 정보들을 활용하여 산불 예측 모델을 개발하였다. 특히, 날씨 데이터 포함 여부에 따른 Random Forest (RF) 기반의 예측 모델 성능을 비교하였으며 날씨 데이터가 포함된 경우 예측 모델의 더 높은 정확도를 보였다. 이는 날씨 데이터와 산불이 밀접한 연관성이 있음을 의미한다. 하지만, 이러한 선행 연구들은 산불 발생 여부에 초점을 두고 산불이 발생한 당일의 데이터만을 활용했다는 한계점이 존재한다. 따라서, 본 연구에서는 시계열 데이터 특징을 가지는 날씨 데이터 같은 경우는 과거 날씨 데이터를 추가함으로써 산불 규모 예측 모델의 성능을 향상시키고자 한다. 이러한 산불 규모 예측 모델은 산불 발생 후 대형 산불로의 발전 가능성을 예측하여 산불 발생 시 진압 우선순위를 설정할 수 있어 더욱 효과적으로 산불에 대응할 수 있도록 돕는 것에 목적을 두고 있다.

본 논문의 구성은 다음과 같다. 제 II 장에서는 본 연구에서 사용된 원본 데이터세트와 전처리 과정을 설명한다. 또한, 전처리 과정을 통해 새로이 생성한 데이터세트를 소개한다. 제 III 장에서는 산불 규모 예측 모델을 위해 사용한 인공지능 알고리즘들과 예측 모델들의 성능을 평가하기 위한 성능 지표를 소개한다. 제 IV 장에서는 인공지능 기반 산불 규모 예측 모델별 성능에 대한 분석을 수행하였다. 마지막으로 제 V 장에서는 본 연구의 결론과 향후 연구 과제를 제시하며 마무리한다.

## II. 원본 데이터셋 설명 및 전처리 과정

### 2.1 원본 데이터셋 설명

본 연구에서는 미국 캘리포니아 주에서 제공하는 California Fire Perimeters 데이터셋에서 2010년부터 2023년까지 최근 약 14년 간의 데이터를 사용한다 [4]. 또한, 산불 규모와 관련성이 있을 수 있는 날씨 정보를 고려하기 위해 Visual Crossing 에서 제공하는 Total Weather Data 라는 데이터셋을 활용한다 [5].

### 2.2 산불 데이터셋 전처리 과정

California Fire Perimeters 원본 데이터셋에서 Year (산불 발생 연도), Alarm\_date (산불 발생 일자), Unit\_ID (산불 관리 구역), Cause (산불 발생 원인), GIS\_calculated\_acres (산불 피해 면적) 등의 5 가지 특성들만을 선택하고 나머지는 제외하였다. 또한, Alarm\_date 에서 산불 규모를 예측하는데 있어 일시는 관련이 없지만 산불이 발생한 연도에 따라 데이터셋 분할 시 필요한 Alarm\_year (산불 발생 연도) 변수 그리고 계절 정보가 포함된 산불 발생 월과 산불 규모의 관련성 때문에 Alarm\_month (산불 발생 월) 변수를 새로이 생성하고 해당 정보를 추가하였다. 지역적 특성을 반영한 산불 규모 예측을 위해 Unit ID 를 County (산불 발생 카운티 지역) 단위로 변환하였으며, 이를 기반으로 산불이 발생한 지역을 캘리포니아 남부와 북부로 나눈 State\_region (남부/북부 캘리포니아 지역)이라는 새로운 변수를 생성하였다. 마지막으로, GIS\_calculated\_acres 변수 값에 따라 300 에이커 (acres) 이상이면 대형 산불 (1 = large wildfire)로 라벨링하고 300 미만은 일반 산불 (0 = small wildfire)로 라벨링하여 새로운 생성한 변수 Wildfire\_category 에 라벨링값을 갱신하였다.

### 2.3 날씨 데이터셋 전처리 과정

Visual Crossing 에서 제공하는 Total Weather Data 데이터셋에서 선택한 특성들은 다음과 같다: Temp\_min (산불 발생 당일 최저 기온), Temp\_max (산불 발생 당일 최고 기온), Temp\_avg (산불 발생 당일 평균 기온), DewPoint\_avg (산불 발생 당일 평균 이슬점 온도), Humidity\_avg (산불 발생 당일 평균 습도), Precip\_avg (산불 발생 당일 평균 강수량), WindSpeed\_avg (산불 발생 당일 평균 풍속) 등 7 개이다. 이 7 개의 변수들 중에서 결측값 (missing value)이 발견되면 결측값이 발생한 일자를 중심으로 3일 전부터 3일 후까지 해당하는 값의 가중 평균을 이용하여 대체하였다. 가중 평균값은 다음과 같이 정의할 수 있다.

$$\bar{v} = \alpha * (v_{-1} + v_1) + \beta * (v_{-3} + v_{-2} + v_2 + v_3),$$

where

$$2\alpha + 4\beta = 1$$

여기에서,  $\bar{v}$ 는 결측값이 발생한 해당 변수의 가중 평균값을 의미한다.  $v_i$ 는  $i$ 가 양의 정수이면 산불 발생 당일로부터  $i$ 일 후의 해당 변수의 값을 의미하고  $i$ 가 음의 정수이면 산불 발생 당일로부터  $i$ 일 전의 해당 변수의 값을 의미한다. 또한, 산불 발생 당일로부터 전후 3일 간의 해당 변수 값에 두 개의 다른 가중치  $\alpha$ 와  $\beta$ 를 할당한다.  $\alpha$ 와  $\beta$ 는 양의 정수이며  $2\alpha + 4\beta = 1$ 를 만족한다. 산불 발생 당일로부터 전후 1일 간에 측정된 날씨 데이터가 전후 2일 또는 3일 간에 측정된 날씨 데이터보다 산불 발생 당일의 해당 데이터와 더욱 높은 관련성이 있기 때문에, 가중치  $\alpha = 0.3$ 와  $\beta = 0.1$ 을 할당하였다. 또한, 날씨 데이터가 시계열 데이터임을 고려하여 날씨 데이터 관련 다음의 두 그룹의 특성들을 추가하였다:

산불 발생 당일을 기준으로 전날의 날씨 데이터들과 전날부터 일주일 전까지의 평균 날씨 데이터들이다. 예를 들면, 산불 발생 전일 최저 기온과 전날부터 일주일 전까지 최저 기온들의 평균 기온은 각각 Temp\_min\_1day 와 Temp\_min\_1week 라 정의한다. 또한, 산불 발생 전일 평균 이슬점 온도와 전날부터 일주일 전까지의 평균 이슬점 온도는 각각 DewPoint\_avg\_1day, DewPoint\_avg\_1week 라 명명한다.

### 2.4 산불 규모 예측을 위한 새로운 데이터셋 생성

본 연구에서 원본 데이터셋으로 고려하는 California Fire Perimeters 데이터셋과 Total Weather Data 데이터셋을 전처리 과정 수행 후 Alarm\_date (산불 발생 일자)와 County (산불 발생 카운티 지역)를 기준으로 결합하였다. 상기 두 개의 데이터셋을 결합한 후, Alarm\_date 변수와 County 변수는 각각 Alarm\_month (산불 발생 월)와 State\_region (남부/북부 캘리포니아 지역) 변수들로 변환하고 모델 학습 시 제외시켰다. 이와 같은 과정을 통해 생성한 데이터 셋을 Constructed Feature Dataset (CFD)라고 정의하고 미국 캘리포니아 산불 규모 예측 모델 개발을 위해 사용하고자 한다. 표 1은 CFD에 포함된 21개의 독립변수 (Independent Variable)들과 1개의 종속변수 (Dependent Variable)를 보여주고 있다. 종속변수인 Wildfire\_category는 산불이 300 에이커 이상으로 발생한 대형 산불 (1 = large wildfire)과 300 에이커 미만으로 발생한 일반 산불 (0 = small wildfire)의 두 개의 클래스로 분류되었다.

표 1. CFD 데이터셋 구성

Name	Type	[Unit] /{Value}
<b>Independent Variables</b>		
Alarm_year	Integer	{2010, ..., 2023}
Alarm_month	String	{1, ..., 12}
State_region	String	{Southern, Northern}
Cause	String	{1, ..., 19}
Temp_avg, Temp_avg_1day, Temp_avg_1week	Double	[°F]
Temp_max, Temp_max_1day, Temp_max_1week	Double	[°F]
Temp_min, Temp_min_1day, Temp_min_1week	Double	[°F]
DewPoint_avg, DewPoint_avg_1day, DewPoint_avg_1week	Double	[°F]
Humidity_avg, Humidity_avg_1day, Humidity_avg_1week	Double	[mph]
Precip_avg	Double	[%]
WindSpeed_avg	Double	[mph]
<b>Dependent Variable</b>		
Wildfire_category	Double	{0 = small wildfire, 1 = large wildfire}

그림 1 은 본 연구에서 고려하는 CFD 의 종속변수인 Wildfire\_category 의 클래스별 구성을 보여주고 있다. 전체 5,061 개의 데이터 샘플들 중에서 다수 클래스인 small wildfire (= 0)와 소수 클래스인 large wildfire (= 1)에 해당하는 데이터 샘플 수는 각각 4,067 개와 994 개이다. 이는 다수 클래스와 소수 클래스의 비율이 약 80.4% : 19.6%로 극심한 불균형 데이터임을 확인 할 수 있다.

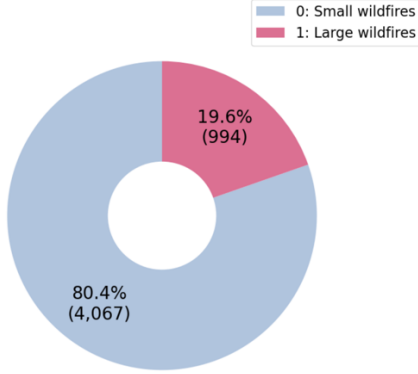


그림 1. CFD 의 Wildfire\_Category 의 클래스 분포

### III. 알고리즘 및 성능 평가 지표

본 연구에서는 K-Nearest Neighbors (KNN), Random Forest (RF)과 Multi-Layer Perceptron (MLP) 등 3 개의 인공지능 알고리즘을 이용하여 산불 피해 규모를 분류하고 예측 모델들에 대한 성능 평가를 수행하였다.

#### 3.1 K-Nearest Neighbors (KNN)

KNN 은 새로운 데이터 포인트의 클래스나 값을 예측할 때 K 개의 가장 가까운 이웃 데이터를 참조하여 결정을 내리는 지도 학습 기법 중 하나이다 [6]. KNN 은 간단하고 직관적인 모델이지만, 데이터의 스케일에 민감하며 모델 성능에 영향을 미치는 중요한 요소 중 하나는 적절한 K 값 (이웃의 수)을 선택하는 것이다.

#### 3.2 Random Forest (RF)

RF 는 앙상블 학습 기법의 일종으로 다수의 Decision Tree 의 결과를 조합하여 데이터 분석을 수행한다 [7]. 분류 문제에서는 각 트리의 예측 결과에서 가장 많이 선택된 클래스를 최종 예측 값으로 결정한다. 개별 결정 트리가 과적합에 취약한 반면에, RF 는 여러 트리의 예측을 결합함으로써 더 뛰어난 일반화 성능을 제공한다.

#### 3.3 Multi-Layer Perceptron (MLP)

MLP 는 Artificial Neural Network (ANN)의 한 유형으로 가장 기본적인 형태의 Deep Learning 모델이다. MLP 는 여러 개의 뉴런으로 이루어진 층들이 쌓여서 구성되며 입력 데이터를 받아 분류 문제를 해결한다. 특히, 본 연구에서 사용한 RProp 은 'Resilient Propagation' 의 약자로 경사 하강법의 단점을 극복하기 위해 만들어진 학습 알고리즘이다 [8]. 이 알고리즘은 활성화함수를 이용하여 복잡한 비선형 관계를 학습할 수 있다는 장점을 가지고 있다.

#### 3.4 성능 평가 지표

산불 규모 예측 모델의 성능을 평가하기 위해 Precision, Recall, F1-score 및 Accuracy 를 지표로 선정하였다. 특히, 분류 모델이 실제 대형 산불을 얼마나 정확하게 예측하는지 평가하는 지표인 Recall 을 가장 중요하게 고려하였다. 또한, Precision 은 모델이 예측한 대형 산불 중 실제 대형 산불의 비율, F1-score 는 Precision 과 Recall 의 조화

평균 그리고 Accuracy 는 전체 예측 중 정확하게 예측된 비율을 의미한다. 상기 4 개의 성능 지표들은 모두 0 과 1 사이의 값을 가지며 1 에 가까울수록 성능이 우수함을 나타낸다. 이러한 성능 지표들은 모델의 예측 결과를 실제 결과와 비교하는 혼동 행렬 (confusion matrix)을 통해 계산된다. 혼동 행렬을 구성하는 요소들로는 True Positive (TP), True Negative (TN), False Positive (FP) 및 False Negative (FN) 등이 있다.

### IV. 산불 피해 규모 예측 결과 및 분석

KNN, RF 및 MLP 등 3 개의 인공지능 알고리즘들과 표 1 에 나열한 CFD 를 활용하여 산불 피해 규모를 예측하고 학습 모델들의 성능을 비교하고자 한다.

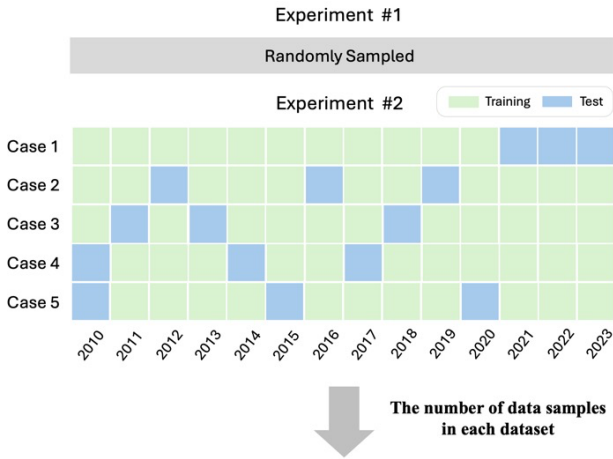
#### 4.1 학습 및 테스트 데이터세트 분류에 따른 실험 구성

산불 규모 예측 모델을 학습하고 평가할 때 사용하는 학습 (training) 및 테스트 (test) 데이터세트들을 분류하는 방식에 따라 2 개의 실험 (Experiment)을 구성한다. 실험 구성 방식에 따른 산불 규모 예측 모델의 성능을 비교 및 평가하고자 한다.

그림 2 와 실험마다 사용하는 학습 및 테스트 데이터 세트들을 분류하는 방법과 그에 따른 데이터세트별 데이터 샘플 개수 그리고 비율을 도시하고 있다. Experiment #1 에서는 2010 년부터 2023 년 사이에 발생한 산불들에 대해 발생 연도와 관계없이 무작위로 훈련 및 테스트 데이터 세트를 80% : 20% 비율로 분할하여 구성 하였다. 또한, Experiment #1 에서는 K-Fold Cross Validation 을 이용하여 일반화된 모델 성능을 확보하고자 하였다. 이 때 사용하는 K 값은 5 이다. Experiment #2 에서는 2010 년부터 2023 년 사이에 발생한 산불들을 각각 다른 5 가지 경우 (Case)로 나누어 학습 및 테스트를 진행 하였다. 모든 경우들에 있어서 훈련 및 테스트 데이터세트들의 분리는 산불 발생 연도를 이용하였고, 이는 훈련 및 테스트 데이터 세트 간의 중복 방지를 보장하는 것이다. 예를 들어, Case 1 의 학습 데이터세트는 산불이 발생한 연도가 2010 - 2020 인 데이터들이며 테스트 데이터세트는 산불이 발생한 연도가 2021 - 2023 인 데이터 들이다. 또한, 이 경우에 있어서 훈련 데이터세트와 테스트 데이터세트에 포함된 데이터 샘플의 개수는 각각 4,050 와 1,011 개이고 비율이 약 80%와 약 20%로 구성되어 있다. Experiment #2 의 5 개의 모든 경우들에 있어서도 테스트 데이터세트를 구성 하는 데이터들도 2010 년을 제외하고 중복되지 않음을 확인할 수 있다. 이는 5 개의 모든 경우에서 학습 및 테스트 데이터세트의 분포를 약 80%와 약 20%로 유사하게 설정 하여 실험 간의 오차를 줄이고자 함이다. Experiment #2 에 대한 예측 모델의 최종 성능 지표값들은 5 개의 경우들로부터 얻은 해당 성능 지표값들의 평균이다.

#### 4.2 산불 규모 예측 모델 분석 결과

각 실험별 최적 하이퍼파라미터 (hyperparameter)는 다음과 같다: Experiment #1 에서 KNN 기반 예측 모델의 경우 K 가 3 이고, RF 기반 예측 모델은 트리 깊이 (tree depth)와 잎 노드 (leaf node)가 최소로 포함해야 할 데이터 샘플의 개수 (minimum split node size)가 각각 13 과 27 이다. 또한, MLP 기반 예측 모델은 은닉층의 개수 (hidden layer)와 은닉층의 포함 뉴런 개수 (hidden neuron)가 각각 1 과 18 이다. Experiment #2 에서는 KNN 기반 예측 모델의 K 가 3 이고, RF 기반 예측 모델의 tree depth 와 minimum split node size 가 각각 21 과 6 이며, MLP 기반 예측 모델의 hidden layer 와 hidden neuron 는 각각 1 과 44 로 최적 하이퍼파라미터를 설정하였다.



Datasets		Training	Test
Experiment #1		4,049	1,012
Experiment #2	Case 1	4,050	1,011
	Case 2	4,024	1,037
	Case 3	3,979	1,082
	Case 4	3,968	1,093
	Case 5	4,004	1,057
Ratio		~ 80%	~ 20%

그림 2. 실험별 학습 및 테스트 데이터셋 구성

표 2는 상기에서 정의한 2개의 실험에 따른 산불 규모 예측 모델별 성능 평가 지표값들을 보여준다. Recall, F1-score 및 Accuracy 등 모든 예측 모델 성능 지표 결과를 고려했을 때에는 전체 데이터셋(CFD)에서 무작위로 훈련 및 테스트 데이터셋으로 나누어 학습한 Experiment #1의 성능이 그림 2에서 보여준 것처럼 CFD에서 연도별로 훈련 및 테스트 데이터셋을 약 80% 대 약 20%로 분할하여 5개의 경우로 구성한 뒤 학습한 Experiment #2의 성능보다 전반적으로 좋음을 확인 할 수 있다. 단, RF 알고리즘 기반 예측 모델의 Accuracy에 있어서 Experiment #1이 Experiment #2보다 0.004가 작음을 확인할 수 있으나 무시할 정도이다. 또한, 산불 규모 예측 모델들의 성능이 우수한 Experiment #1의 결과를 확인해 볼 때, Recall과 F1-score는 MLP 알고리즘 기반 예측 모델이 우수하지만 Accuracy는 RF 알고리즘 기반 예측 모델이 우수하다.

표 2. 실험 방식에 따른 산불 규모 예측 모델별 성능 결과

	Experiment #1			Experiment #2		
	KNN	RF	MLP	KNN	RF	MLP
Recall	0.170	0.157	0.206	0.112	0.035	0.193
F1-score	0.221	0.235	0.275	0.153	0.065	0.241
Accuracy	0.764	0.799	0.783	0.759	0.803	0.765

하지만, 본 연구에서는 실제 대형 산불들을 올바르게 예측하는 비율인 Recall 값을 중요 지표로 선정하였기 때문에, 가장 높은 Recall과 그에 따른 F1-score를 가지는 MLP 기반 예측 모델이 가장 뛰어난 성능을 가진다고 판단하였다. 표 2에서는 2개의 실험 방식 전반에 걸쳐서 Accuracy를 제외하고 Recall와 F1-score 값이 좋지 못함을 확인할 수

있는데 이는 그림 1에서 보여준 것처럼 본 연구에서 고려하는 CFD의 극심한 데이터 불균형으로 기인한다. 따라서, 본 연구의 후속 과제로 다양한 샘플링 기법들을 시도하여 데이터 불균형 문제를 해결하여 Recall 및 F1-score를 향상시키고자 한다.

## V. 결론

본 연구는 미국 캘리포니아 주에서 빈번히 발생하는 산불의 피해 규모를 예측하기 위해 산불 데이터셋(California Fire Perimeters)와 날씨 데이터셋(Total Weather Data)을 활용하여 인공지능 기반 산불 규모 예측 모델을 제시하였다. 이 두 개의 데이터셋들로부터 산불 규모 예측을 위해 필요한 특성들을 선택하여 새로운 데이터셋(Constructed Feature Dataset)을 생성하였다. 또한, 대상 산불들은 2010년부터 2023년까지 발생한 산불들이다. K-Nearest Neighbor(KNN), Random Forest(RF) 및 Multi-Layer Perceptron(MLP) 등 인공지능 알고리즘들을 고려하였다. Constructed Feature Dataset에서 산불이 발생한 연도와 관계없이 학습 및 테스트 데이터셋으로 분할하여 학습한 실험과 산불 발생 연도에 따라 학습 및 테스트 데이터셋으로 분할하여 학습한 실험을 대상으로 산불 규모 예측 모델들의 성능을 비교 분석하였다. 산불이 발생한 연도와 상관없이 학습 및 테스트 데이터셋으로 분할하여 학습한 실험이 전반적인 성능 지표들 값에 있어서 우수함을 확인할 수 있었다. 또한, 본 연구에서 목표로 하는 실제 300에이커 이상의 산불이 발생한 대형 산불을 올바르게 예측하는 비율인 Recall을 고려하면 MLP 알고리즘 기반 산불 규모 예측 모델이 가장 우수하였다. 하지만, MLP 알고리즘 기반 예측 모델의 Recall 값이 여전히 낮음을 확인할 수 있었는데 이는 본 연구에서 고려하는 데이터의 극심한 불균형에 기인한다.

이에 향후 연구에서는 다양한 샘플링 기법들을 통해 데이터셋의 데이터 불균형 문제를 해결하여 모델의 성능을 더욱 향상시키고자 한다. 또한, 산불 피해 면적에 따른 경제적 비용 데이터를 추가하여 예측된 산불 규모의 사회적 및 경제적 비용을 정량화하여 보다 효과적으로 산불 대응 전략을 수립하고자 한다.

## ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation) (2021-0-01393).

Following are results of a study on the "Leaders in Industry university Cooperation 3.0" Project, supported by the Ministry of Education and National Research Foundation of Korea.

## 참고 문헌

- [1] M. D. Hurteau, S. Liang, A. L. Westerling, and C. Wiedinmyer, "Economic Footprint of California Wildfires in 2018," Nature Sustainability, vol. 4, no. 3, pp. 252-260, 2021.
- [2] K. Pham, D. Ward, S. Rubio, D. Shin, L. Zlotikman, S. Ramirez, T. Poplawski, and X. Jiang, "California Wildfire Prediction Using Machine Learning," in 2022 21st IEEE International Conference on

Machine Learning and Applications (ICMLA), vol. 1, pp. 525–530, 2022.

- [3] A. Malik, M. R. Rao, N. Puppala, P. Koouri, V. A. K. Thota, Q. Liu, S. Chiao, and J. Gao, "Data-Driven Wildfire Risk Prediction in Northern California," *Atmosphere*, vol. 12, no. 1, p. 109, 2021.
- [4] California Open Data Portal, "California Fire Perimeters (1950+)," Department of Forestry and Fire Protection (CAL FIRE), Available: <https://data.ca.gov/dataset/california-fire-perimeters-1950>, [Accessed: Oct. 03, 2024].
- [5] Visual Crossing, "Total Weather Data – Historical. Weather Data and Weather Forecast Data," Visual Crossing Corporation, Available: <https://www.visualcrossing.com/weather-data>, [Accessed: Oct. 03, 2024].
- [6] O. Kramer, "Dimensionality Reduction with Un-supervised Nearest Neighbors: K-Nearest Neighbors," vol. 51, pp. 13–23, 2013.
- [7] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [8] M. Riedmiller and H. Braun, "A Direct Adaptive. Method for Faster Backpropagation Learning: The RPROP Algorithm," *IEEE International Conference on Neural Networks*, vol. 16, pp. 586–591, 1993.