

지능형 사물인터넷 기술 교육을 위한 머신러닝 모델 활용 사례 개발

Development of Machine Learning Model Use Cases for Intelligent Internet of Things Technology Education

허 경*

경인교육대학교 컴퓨터교육과

Kyeong Hur*

Department of Computer Education, Gyeong-In National University of Education, Anyang 13910, Korea

[요약]

지능형 사물인터넷인 AIoT는 IoT 디바이스가 측정된 데이터를 수집하고 머신러닝 기술을 적용해 예측 모델을 만들어 활용하는 기술을 의미한다. AIoT 기술 교육을 위한 기존 연구에서는 교육용 AIoT 플랫폼 구축하고 사용법을 교육하는 데 초점을 맞추었다. 그러나, IoT 디바이스가 측정된 데이터로부터 머신러닝 모델이 자동 생성되고 활용되는 과정을 교육하는 사례 연구는 부족하였다. 본 논문에서는 AIoT 기술 교육을 위한 머신러닝 모델 활용 사례를 개발하였다. 본 논문에서 개발한 사례는 AIoT 디바이스의 데이터 수집, 데이터 전처리, 머신러닝 모델 자동 생성, 모델별 정확도 산출 및 유효 모델 결정, 유효 모델을 활용한 데이터 예측 단계들로 구성되었다. 본 논문에서는 AIoT 디바이스의 센서들이 서로 다른 범위의 값들을 측정하는 것을 고려하였고, 이에 따른 데이터 전처리 사례를 제시하였다. 또한 여러 머신러닝 모델들을 자동 생성하고 이 모델들 중 정확도가 높은 유효 모델을 결정하여, AIoT 디바이스가 어떤 정보를 예측할 수 있는가를 스스로 결정하는 사례를 개발하였다. 개발한 사례를 적용하면, AIoT를 활용한 예측기반 사물 제어와 같은 AIoT 활용 교육 콘텐츠를 다양하게 개발할 수 있다.

[Abstract]

AIoT, the intelligent Internet of Things, refers to a technology that collects data measured by IoT devices and applies machine learning technology to create and utilize predictive models. Existing research on AIoT technology education focused on building an educational AIoT platform and teaching how to use it. However, there was a lack of case studies that taught the process of automatically creating and utilizing machine learning models from data measured by IoT devices. In this paper, we developed a case study using a machine learning model for AIoT technology education. The case developed in this paper consists of the following steps: data collection from AIoT devices, data preprocessing, automatic creation of machine learning models, calculation of accuracy for each model, determination of valid models, and data prediction using the valid models. In this paper, we considered that sensors in AIoT devices measure different ranges of values, and presented an example of data preprocessing accordingly. In addition, we developed a case where AIoT devices automatically determine what information they can predict by automatically generating several machine learning models and determining effective models with high accuracy among these models. By applying the developed cases, a variety of educational contents using AIoT, such as prediction-based object control using AIoT, can be developed.

Key Words: AI, AIoT education, IoT, Machine learning, Numerical data prediction

<http://dx.doi.org/10.14702/JPEE.2024.449>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 4 April 2024; **Revised** 22 April 2024

Accepted 13 May 2024

***Corresponding Author**

E-mail: khur@ginue.ac.kr

I. 서론

지능형 사물인터넷(IoT)이라는 AIoT는 1단계 연결형 IoT로부터 상황 인지, 단순 측정 및 제어, 예측이 가능한 2단계 지능형 IoT를 거쳐 자율적인 판단과 자율 제어까지 가능한 3단계 자율형 IoT로 발전해 가고 있다[1]. 즉, AIoT는 IoT 디바이스가 측정된 데이터를 수집하고 머신러닝 기술을 적용해 예측 모델을 만들어 활용하는 기술을 의미한다.

AIoT 기술 교육을 위한 기존 연구에서는 교육용 AIoT 플랫폼 구축하고 사용법을 교육하는 데 초점을 맞추었다. 참고 문헌[2]에서는 머신러닝과 IoT를 피지컬 컴퓨팅으로 교육하는 플랫폼을 제시하였다. 파이썬 비주얼 프로그래밍을 기반으로 머신러닝을 적용할 수 있도록 하였으며, 피지컬 컴퓨팅을 위한 AIoT 보드로 라즈베리파이를 활용하였다. 참고문헌 [3]에서는 파이썬을 활용한 AIoT 교육 프로그램으로 24차시 내용을 제시하였다. 24차시 내용은 AIoT 기초, 디지털 AI의 이해와 디지털 AI의 활용으로 구성하여 AIoT 장치에 구현된 기능을 사용하는 방법을 교육하고자 하였다. 그러나, IoT 디바이스가 측정된 데이터로부터 머신러닝 모델이 자동 생성되고 데이터를 예측하는 데 활용되는 과정을 교육하는 사례 연구는 부족하였다.

이와 같이, 데이터를 수집하고 가공하여 머신러닝 모델을 통해, 다양한 현상을 예측하는 교육 시에는 워크플로우 및 데이터 시각화를 쉽게 구성할 수 있고 분석할 수 있는 도구가 필요하다[4-7]. Orange3는 직관적이며 복잡한 데이터의 시각화 및 머신러닝 분석을 수행할 수 있는 오픈소스 플랫폼으로서 최신 머신러닝 모델 및 공개된 학습 데이터를 제공한다[7]. 실제 AIoT 시스템 구현에는 Orange3를 사용하지 않으나, Orange3는 머신러닝 모델이 생성되고 데이터를 예측하는 전체 과정을 한눈에 볼 수 있어 머신러닝 교육에 효과적이다. 따라서 본 논문에서는 AIoT 시스템의 머신러닝 모델 활용 과정을 Orange3로 설명하였다.

본 논문에서는 AIoT 기술 교육을 위한 머신러닝 모델 활용 사례를 개발하였다. 개발한 사례는 AIoT 시스템의 데이터 수집, 측정 데이터 전처리, 머신러닝 모델 자동 생성, 모델별 정확도 산출 및 유효 모델 결정, 유효 모델을 활용한 데이터 예측 단계들로 구성되었다. AIoT 디바이스의 센서들이 서로 다른 범위와 특성을 갖는 값들을 측정하는 것을 고려하였고, 이에 따른 데이터 전처리 사례를 제시하였다. 또한 여러 머신러닝 모델들을 자동 생성하고 이 모델들 중 정확도가 높은 유효 모델을 결정하여, AIoT 디바이스가 어떤 정보를 예측할 수 있는가를 스스로 결정하는 사례를 개발하였다. 개

발한 사례를 적용하면, AIoT를 활용한 예측기반 사물 제어와 같은 AIoT 활용 교육 콘텐츠를 다양한 분야에 적용하여 개발할 수 있다.

본 논문에서는 AIoT 디바이스의 A~D 4종 센서가 측정된 수치 데이터 세트를 3개 설명변수(원인) 그룹으로 나누었다. 그리고 AIoT 디바이스가 측정된 4종 데이터들로부터 유효한 머신러닝 모델을 자동으로 도출하기 위해, 하나의 설명변수 그룹을 지정하고 나머지 4개 데이터 그룹들 중 하나의 목적변수(결과)를 지정하여 훈련데이터를 자동 생성하였다. 이렇게 자동 생성된 62종 훈련데이터로부터 학습된 62종 머신러닝 모델의 정확도를 모든 경우에 대해 도출하였다. 이중 정확도가 기준치를 넘는 유효한 모델을 자동 선택하였다. 이후 유효한 모델에 입력되는 설명변수 데이터를 랜덤하게 발생시켜, AIoT가 측정하는 특정 항목 값의 발생 범위를 예측하였다. 이러한 AIoT 시스템의 머신러닝 모델 활용 과정을 Orange3로 제안하였으며, 머신러닝 모델의 자동생성을 구현하는 방법과 AIoT 디바이스의 구현방법은 본 논문에서 제시하지 않았다.

2장에서는 AIoT 시스템과 데이터 수집 및 전처리 과정을 설명하였다. 3장에서는 머신러닝 모델 자동 생성 과정을 기술하였다. 4장에서는 모델별 정확도 산출 및 유효 모델 결정에 대해 설명하였다. 5장에서는 유효 모델을 활용한 데이터 예측에 대해 기술하였다. 6장에서는 결론을 통해, 제안한 AIoT 머신러닝 모델 활용 사례의 적용 효과와 추후 연구에 대해 설명하였다.

II. AIoT 시스템과 데이터 수집 및 전처리

그림 1은 클라우드 기반 AIoT 시스템 구조를 나타낸다[3]. 그림 1은 피지컬 컴퓨팅 IoT 디바이스에 클라우드 기반 AI가 결합된 형태이다. AIoT 디바이스 입력 장치로서 다양한 센서들이 데이터를 수집하여 클라우드 기반 AI에게 학습 데이터를 제공한다. 인터넷이 연결된 PC를 통해 클라우드 상 AI 라이브러리에서 머신러닝 모델을 불러오고 학습시킨 후, AIoT 디바이스의 MCU로 예측 결과를 전송하고 입출력 장치를 제어하는 구조이다[3]. 일반적인 AIoT 머신러닝은 공기 성분과 품질, 온도, 진동, 압력, 소리, 습도 등과 같은 측정된 수치 데이터의 패턴을 식별하고, 예측값과 비교하여 이상을 감지하는 기능을 제공한다[3].

표 1부터 표 3은 특정기간동안 AIoT 디바이스의 A센서, B센서, C센서 그리고 D센서가 측정된 수치 데이터의 항목

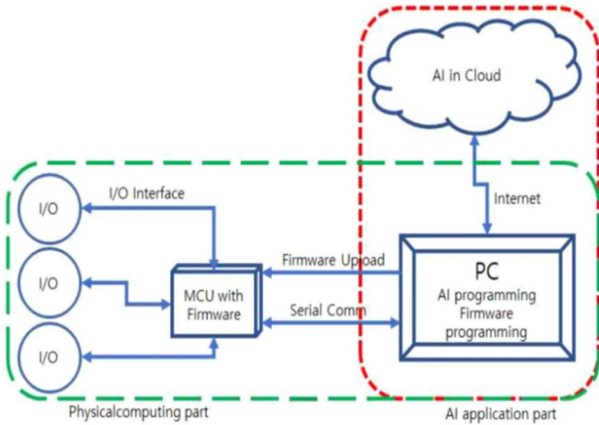


그림 1. 클라우드기반 AIoT 시스템 구조[3]

Fig. 1. Cloud-based AIoT system architecture [3].

들과 범위를 나타낸다. A 센서측정 데이터세트는 A-1~A-6으로, B 센서측정 데이터세트는 B-1~B-5로, C 센서측정 데이터세트는 C-1~C-3로, D 센서측정 데이터세트는 D-1~D-5로, E 영역 데이터세트는 E-1~E-9로, 구성되었다. 여기서 A-6는 A-1~A-5 값들을 합한 데이터, B-5는 B-1~B-4 값들을 합한 데이터, E-9는 E-1~E-8 값들을 합한 데이터이다. E 영역 데이터세트는 C 센서측정 데이터세트와 D 센서측정 데이터세트로 구성되었다. A-6, B-5 및 E-9와 같은 합계 값이 설명변수(원인)에 포함된 이유는, 특정 센서가 측정하는 세부 측정 항목값들과 별개로 전체적인 특성이 담긴 합계 값이, 다른 특성의 측정값과 상관성이 높을 수 있고 정확도가 높은 머신러닝 모델이 도출될 수 있기 때문이다. 본 논문의 머신러닝 모델 활용 사례에서는 AIoT 디바이스의 A, B, C, D 센서들이 기 설정된 특정 기간동안 측정한 결과로 155개 데이터세트가 각각 A, B, C, D, E 영역에 수집되었고 이 데이터세트를 머신러닝 모델에 필요한 학습데이터로 사용하였다.

표 1. 특정기간동안 AIoT 디바이스의 A센서가 측정한 수치 데이터의 항목들과 범위

Table 1. Items and range of numerical data measured by the A sensor of the AIoT device during a specific period

센서	측정항목	수치값범위
A	A-1	0~16
	A-2	0~28
	A-3	0~20
	A-4	0~20
	A-5	0~16
	A-6(합계)	0~100

표 2. 특정기간동안 AIoT 디바이스의 B센서가 측정한 수치 데이터의 항목들과 범위

Table 2. Items and range of numerical data measured by the B sensor of the AIoT device during a specific period

센서	측정항목	수치값범위
B	B-1	0~56
	B-2	0~16
	B-3	0~20
	B-4	0~8
	B-5(합계)	0~100

표 3. 특정기간동안 AIoT 디바이스의 C, D 센서가 측정한 수치 데이터의 항목들과 범위

Table 3. Items and range of numerical data measured by the C and D sensors of the AIoT device during a specific period

영역	센서	측정항목	수치값범위	
C		C-1(=E-1)	0~10	
		C-2(=E-2)	0~10	
		C-3(=E-3)	0~10	
		D	D-1(=E-4)	0~6
			D-2(=E-5)	0~6
D-3(=E-6)	0~6			
D		D-4(=E-7)	0~6	
		D-5(=E-8)	0~6	
		E-9(합계)	0~60	

표 4. 기존 A 센서측정 데이터세트의 전처리 과정

Table 4. Preprocessing process of existing A sensor measurement dataset

센서	측정항목	기준범위	조정상수	신규범위
A	A-1	0~16	x(100/16)	0~100
	A-2	0~28	x(100/28)	0~100
	A-3	0~20	x(100/20)	0~100
	A-4	0~20	x(100/20)	0~100
	A-5	0~16	x(100/16)	0~100
	A-6	0~100	x(100/100)	0~100

표 1부터 표 3과 같이 본 개발 사례에서는 AIoT 디바이스의 A, B, C, D센서들이 서로 다른 범위와 특성을 갖는 값들을 측정한다. 본 데이터들이 머신러닝 모델에 필요한 학습데이터로 사용된다. 머신러닝 모델의 학습 효율을 향상시키기 위해, 기존 측정값에 (100/최대측정값)에 해당하는 조정상수를 곱하여 모든 측정 수치값들을 0~100 범위로 변환하였다. 표 4부터 표 8은 기존 A~E 데이터세트가 전처리되어 변환되는 과정을 설명하고 있다.

표 5. 기존 B 센서측정 데이터세트의 전처리 과정

Table 5. Preprocessing process of existing B sensor measurement dataset

센서	측정항목	기존범위	조정상수	신규범위
B	B-1	0~56	x(100/56)	0~100
	B-2	0~16	x(100/16)	0~100
	B-3	0~20	x(100/20)	0~100
	B-4	0~8	x(100/8)	0~100
	B-5	0~100	x(100/100)	0~100

표 6. 기존 C 센서측정 데이터세트의 전처리 과정

Table 6. Preprocessing process of existing C sensor measurement dataset

영역	측정항목	기존범위	조정상수	신규범위
C	C-1	0~10	x(100/10)	0~100
	C-2	0~10	x(100/10)	0~100
	C-3	0~10	x(100/10)	0~100

표 7. 기존 D 센서측정 데이터세트의 전처리 과정

Table 7. Preprocessing process of existing D sensor measurement dataset

영역	측정항목	기존범위	조정상수	신규범위
D	D-1	0~6	x(100/6)	0~100
	D-2	0~6	x(100/6)	0~100
	D-3	0~6	x(100/6)	0~100
	D-4	0~6	x(100/6)	0~100
	D-5	0~6	x(100/6)	0~100

표 8. 기존 E 영역 데이터세트의 전처리 과정

Table 8. Preprocessing process of existing E category dataset

영역	측정항목	기존범위	조정상수	신규범위
E	E-1	0~10	x(100/10)	0~100
	E-2	0~10	x(100/10)	0~100
	E-3	0~10	x(100/10)	0~100
	E-4	0~6	x(100/6)	0~100
	E-5	0~6	x(100/6)	0~100
	E-6	0~6	x(100/6)	0~100
	E-7	0~6	x(100/6)	0~100
	E-8	0~6	x(100/6)	0~100
	E-9	0~60	x(100/60)	0~100

III. 머신러닝 모델 자동 생성 과정

표 9와 같이, 수집된 데이터들로부터 유효한 머신러닝 모델 탐색을 위해, 생성 가능한 모든 모델을 그룹 단위로 생성

표 9. 머신러닝 모델 그룹별 자동 생성

Table 9. Automatic creation of machine learning model groups

그룹 번호	설명변수 데이터세트(원인)	목적변수 데이터세트(결과)
1	A	B, E
2	B	A, E
3	C	
4	D	A, B
5	E	

표 10. 머신러닝 모델그룹 1번의 모델번호 자동 할당 결과

Table 10. Automatic assignment of model number to machine learning model group 1

그룹 번호	모델 번호	설명변수 데이터세트	목적변수 데이터세트
1	1	A-1~A-6	B-1
	2	A-1~A-6	B-2
	3	A-1~A-6	B-3
	4	A-1~A-6	B-4
	5	A-1~A-6	B-5
	6	A-1~A-6	E-1
	7	A-1~A-6	E-2
	8	A-1~A-6	E-3
	9	A-1~A-6	E-4
	10	A-1~A-6	E-5
	11	A-1~A-6	E-6
	12	A-1~A-6	E-7
	13	A-1~A-6	E-8
	14	A-1~A-6	E-9

하고 모델 번호를 부여하였다. 표 10과 같이, 모델그룹 1번에서는 A-1~A-6 측정 결과 데이터들을 설명변수들(원인요소들)로 하여 B-1~B-5 및 E-1~E-9의 값들 중 하나의 값을 목적변수(결과)로 예측하는 총 14개의 데이터 예측 모델들이 생성되었다. 표 11과 같이, 모델그룹 2번에서는 B-1~B-5 측정 결과 데이터들을 설명변수들(원인요소들)로 하여, A-1~A-6 및 E-1~E-9의 값들 중 하나의 값을 목적변수(결과)로 예측하는 총 15개의 데이터 예측 모델들이 생성되었다.

표 12와 같이, 모델그룹 3번에서는 C-1~C-3 측정 결과 데이터들을 설명변수들(원인요소들)로 하여, A-1~A-6 및 B-1~B-5의 값들 중 하나의 값을 목적변수(결과)로 예측하는 총 11개의 데이터 예측 모델들이 생성되었다. 표 13과 같이, 모델그룹 4번에서는 D-1~D-5 측정 결과 데이터들을 설명변수들(원인요소들)로 하여, A-1~A-6 및 B-1~B-5의 값들 중

표 11. 머신러닝 모델그룹 2번의 모델번호 자동 할당 결과

Table 11. Automatic assignment of model number to machine learning model group 2

그룹 번호	모델 번호	설명변수 데이터세트	목적변수 데이터세트
2	15	B-1~B-5	A-1
	16	B-1~B-5	A-2
	17	B-1~B-5	A-3
	18	B-1~B-5	A-4
	19	B-1~B-5	A-5
	20	B-1~B-5	A-6
	21	B-1~B-5	E-1
	22	B-1~B-5	E-2
	23	B-1~B-5	E-3
	24	B-1~B-5	E-4
	25	B-1~B-5	E-5
	26	B-1~B-5	E-6
	27	B-1~B-5	E-7
	28	B-1~B-5	E-8
	29	B-1~B-5	E-9

표 12. 머신러닝 모델그룹 3번의 모델번호 자동 할당 결과

Table 12. Automatic assignment of model number to machine learning model group 3

그룹 번호	모델 번호	설명변수 데이터세트	목적변수 데이터세트
3	30	C-1~C-3	A-1
	31	C-1~C-3	A-2
	32	C-1~C-3	A-3
	33	C-1~C-3	A-4
	34	C-1~C-3	A-5
	35	C-1~C-3	A-6
	36	C-1~C-3	B-1
	37	C-1~C-3	B-2
	38	C-1~C-3	B-3
	39	C-1~C-3	B-4
	40	C-1~C-3	B-5

하나의 값을 목적변수(결과)로 예측하는 총 11개의 데이터 예측 모델들이 생성되었다. 표 14와 같이, 모델그룹 5번에서는 E-1~E-9 추정 결과 데이터들을 설명변수들(원인요소들)로 하여, A-1~A-6 및 B-1~B-5의 값들 중 하나의 값을 목적변수(결과)로 예측하는 총 11개의 데이터 예측 모델들이 생성되었다.

표 13. 머신러닝 모델그룹 4번의 모델번호 자동 할당 결과

Table 13. Automatic assignment of model number to machine learning model group 4

그룹 번호	모델 번호	설명변수 데이터세트	목적변수 데이터세트
4	41	D-1~D-5	A-1
	42	D-1~D-5	A-2
	43	D-1~D-5	A-3
	44	D-1~D-5	A-4
	45	D-1~D-5	A-5
	46	D-1~D-5	A-6
	47	D-1~D-5	B-1
	48	D-1~D-5	B-2
	49	D-1~D-5	B-3
	50	D-1~D-5	B-4
	51	D-1~D-5	B-5

표 14. 머신러닝 모델그룹 5번의 모델번호 자동 할당 결과

Table 14. Automatic assignment of model number to machine learning model group 5

그룹 번호	모델 번호	설명변수 데이터세트	목적변수 데이터세트
5	52	E-1~E-9	A-1
	53	E-1~E-9	A-2
	54	E-1~E-9	A-3
	55	E-1~E-9	A-4
	56	E-1~E-9	A-5
	57	E-1~E-9	A-6
	58	E-1~E-9	B-1
	59	E-1~E-9	B-2
	60	E-1~E-9	B-3
	61	E-1~E-9	B-4
	62	E-1~E-9	B-5

IV. 모델별 정확도 산출 및 유효 모델 결정

1~5번 그룹 내 62개 모델에 필요한 지도학습 데이터를 그룹 2의 Orange3 머신러닝 모델 학습 워크플로우로 입력하여 머신러닝 모델을 학습시켰다. 그룹 2의 워크플로우를 다음과 같이 설명한다. 첫번째로, A-1~A-6, B-1~B-5, C-1~C-3, D-1~D-5, E-1~E-9 추정 항목값들로 구성된 것을 인스턴스로 정의한다. E-1~E-8 항목 값들이 C-1~C-3와 D-1~D-5 항목값들과 중복되므로 전체 20개 데이터 항목이 한세트가 되어 1개 인스턴스를 구성한다. 본 사례에서는 특정 시간구간

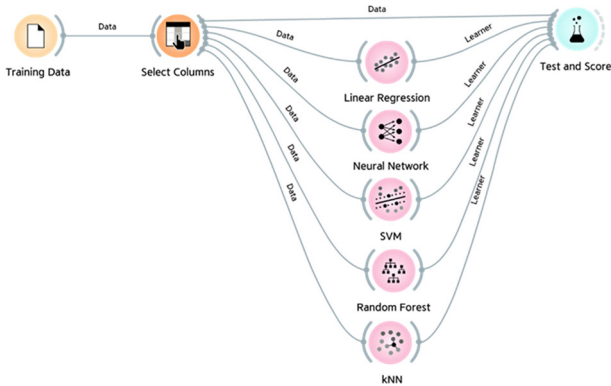


그림 2. Orange3 머신러닝 모델 학습을 위한 워크플로우
Fig. 2. Workflow for training Orange3 machine learning model.

동안 측정된 155개 인스턴스가 포함된 데이터 세트 파일을 그림 2의 Training Data 위젯에 입력하였다.

두번째로, Select Columns 위젯에서 62개 지도학습데이터 모델 중 하나의 모델을 지정하여, 해당설명변수들(Feature)을 지정하고, 하나의 목적변수(Target)를 지정하였다. 그외 나머지 측정값들은 Ignored에 넣어 무시하였다. 세번째로, Linear Regression, Neural Network, SVM, Random Forest, kNN 위젯을 연결하여, 특정 모델 번호에 해당하는 지도학습 데이터를 학습한 머신러닝 모델이 생성되었다. 네번째로, 각 머신러닝 모델의 정확도를 측정하기 위해 Test and Score 위젯을 연결하였다. Test and Score 위젯에서 신뢰성이 높은 결과를 얻기 위해 Cross validation 값을 10으로 설정하였다[8]. 이를 통해, 학습 데이터를 10개 파트으로 분할하고 1개 파트를 테스트 데이터로 사용하여 10회 반복한 평균 정확도를 산출하였다. Test and Score 위젯 내에서 각 모델의 예측 오차값으로 RMSE(Root Mean Square Error)값을 측정하였다. 본 논문에서는 100에서 RMSE 값을 뺀 값을 해당 모델의 정확도로 정

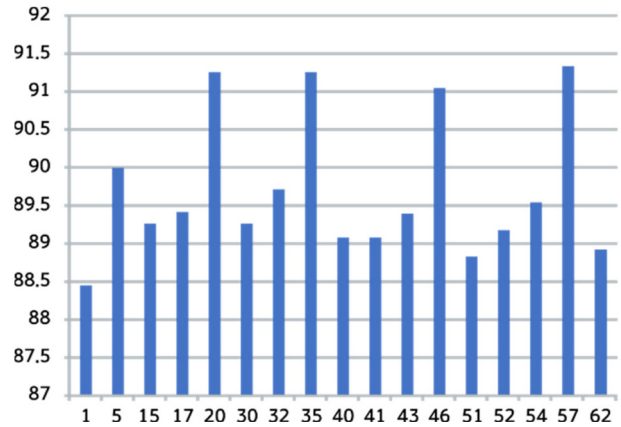


그림 4. 88% 이상의 평균 정확도를 나타낸 모델 번호
Fig. 4 Model numbers with average accuracy greater than 88%.

의하였다. 그림 2에서는 여러 머신러닝 모델을 개별로 활용하여 각 모델의 정확도를 산출하였다. 이 방법보다 스택킹(Stacking)을 이용한 앙상블 모델이 더 효과적일 수 있으며 향후 Orange3에서 적용하여 워크플로우를 개선할 수 있다 [9].

모델번호 1번~62번 모든 경우에서, Linear Regression, Neural Network, SVM, Random Forest 그리고 kNN [10] 5개 머신러닝 모델들을 적용하였고, 이중 정확도가 높지 않은 Neural Network 모델의 정확도 값은 정확도 평균값을 계산할 때 제외하였다. 그림 3과 같이, 전체 62개 모델들의 정확도 평균값은 82.6%으로 산출되었다. 그림 3은 62개 모델 분석 결과 중 88% 이상의 평균 정확도를 나타낸 모델 번호들을 나타낸다. 각 모델번호에 표시된 평균 정확도는 kNN, SVM, Random Forest, Linear Regression 4개 모델에 해당 설명변수들과 목적변수를 포함한 155개 인스턴스 지도학습 데이터를

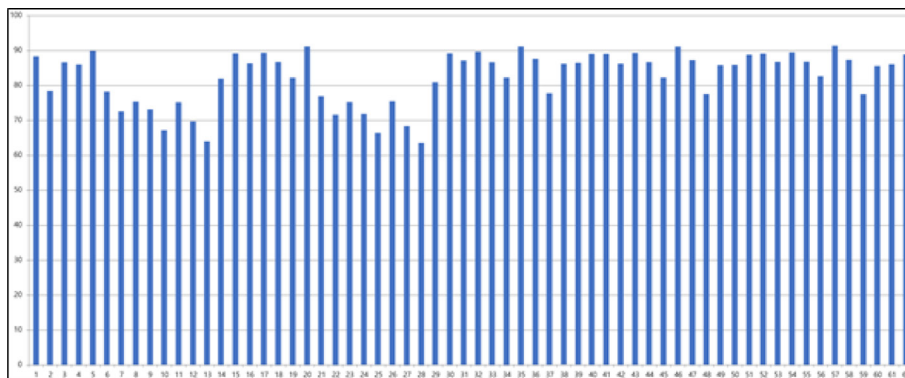


그림 3. 62개 모델들의 정확도
Fig. 3. Accuracy of 62 models.

표 15. 평균 정확도 90% 이상 유효한 모델번호

Table 15. Valid model number with average accuracy of 90% or higher

모델번호	설명변수데이터세트(원인)	목적변수데이터세트(결과)	평균정확도
5	A-1~A-6	B-5	90.0%
20	B-1~B-5	A-6	91.3%
35	C-1~C-3	A-6	91.2%
46	D-1~D-5	A-6	91.1%
57	E-1~E-9	A-6	91.3%

학습시킨 후 측정된 4개 모델 정확도의 평균 값이다. 그리고 본 논문에서는 평균 정확도가 90 이상일 경우, AIoT 시스템의 유효한 모델로 자동 결정되는 경우를 가정하여, 표 15와 같이 5개 유효한 모델이 결정되었다.

V. 유효 모델을 활용한 데이터 예측

그림 5는 유효 모델을 사용한 데이터 예측 워크플로우를 나타낸다. 첫번째로, 유효한 모델 5, 20, 35, 46, 57번에 필요한 설명변수 값들이 AIoT 디바이스 A, B, C, D 센서를 통해 특정시간구간동안 500개 측정되었다. 이 데이터 세트 파일을 Test Data 위젯에 연결하였다. 두번째로, Select Columns 위젯에 연결하여, 각 모델 5, 20, 35, 46, 57번에서 필요한 설명변수 값들만 Feature로 지정하고 Target 값은 미지정하였다. 세

표 16. 유효한 모델 5번에서 예측된 500개 목적변수값(결과)의 통계값

Table 16. Statistical values of 500 target variable values (results) predicted by valid model No. 5

모델번호	설명변수 데이터세트	목적변수 데이터세트	평균 정확도
5	A-1~A-6	B-5	90.0%
머신러닝모델	최소값	평균	최대값
kNN	63	78	93
Random Forest	70	96	100

번째로, Predictions 위젯에 연결하여, 측정된 설명변수 데이터로부터 예측된 목적변수 값이 Neural Networks를 제외한 4개 모델들로 500개 예측되었다. 4개의 모델에서 결과의 경향이 유사한 두 모델만 선정하여 예측결과의 신뢰도를 높였다. 그리고 Distributions 위젯을 연결하여 각 모델이 예측한 500개 목적변수 값의 분포를 확인하였다.

표 16과 표 17은 유효한 모델 5번과 35번에서 예측된 각 목적변수값(결과)의 최소값, 평균, 최대값을 나타낸다. 여기서 500개 목적변수값들(결과)은 모델 5번과 35번의 원인에 해당하는 설명변수들의 500개 측정값 세트들로부터 예측된 것이다. 이 결과는 AIoT 디바이스가 어떤 정보를 정확하게 예측하고, 예측된 정보의 통계값을 산출할 수 있음을 나타낸다. 이러한 예측 통계값들을 이용하여, 기준값에 따라 특정 목적을 달성하기위한 예측 기반 사물 제어가 가능하다. 한편, 그림 6과 그림 7은 유효한 모델 5번과 35번에서 예측된 500

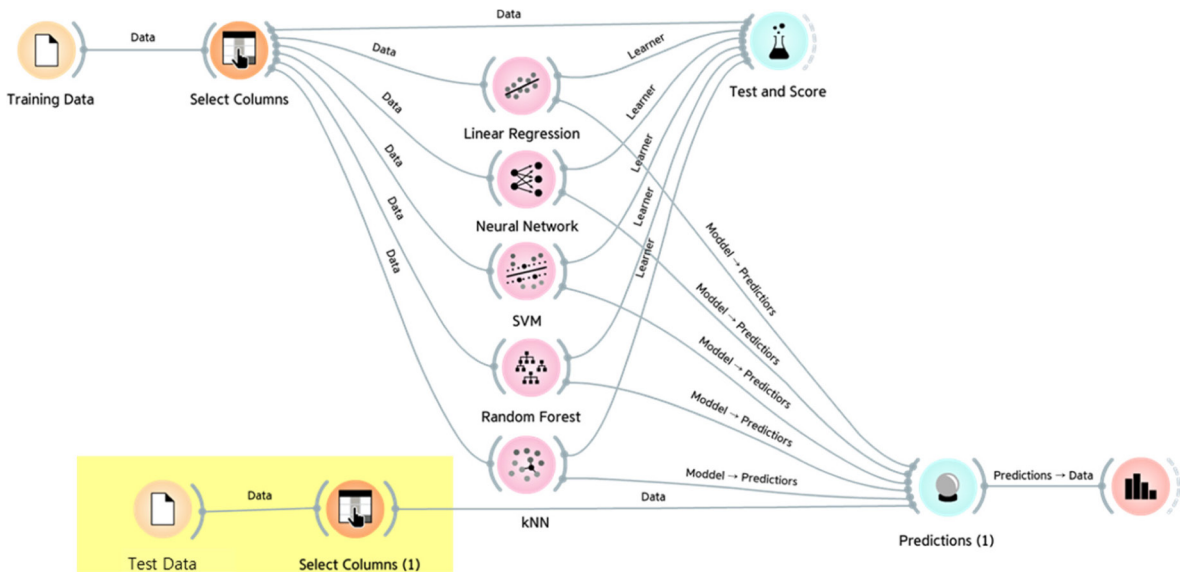


그림 5. 유효 모델을 사용한 데이터 예측 워크플로우

Fig. 5. Data prediction workflow using valid models.

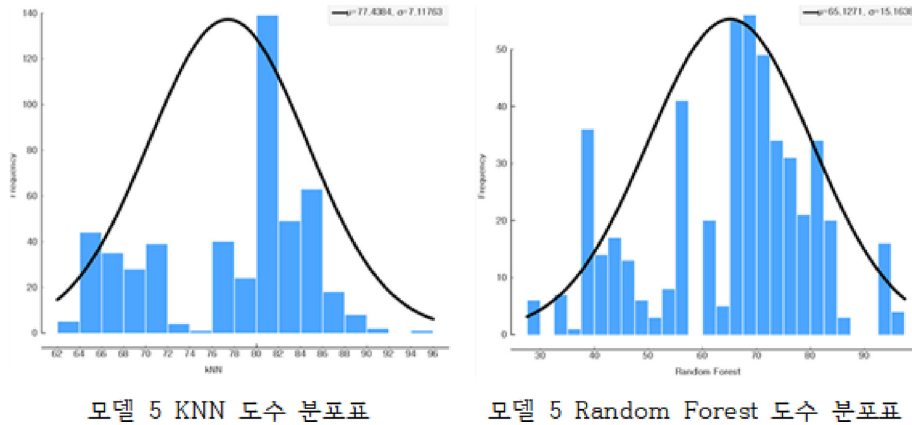


그림 6. 유효한 모델 5번에서 예측된 500개 목적변수값(결과)의 도수분포표
Fig. 6. Frequency distribution graph of 500 target variable values (results) predicted from valid model 5.

표 17. 유효한 모델 35번에서 예측된 500개 목적변수값(결과)의 통계값

Table 17. Statistical values of 500 target variable values (results) predicted by valid model No. 35

모델번호	설명변수 데이터세트	목적변수 데이터세트	평균 정확도
35	C-1~C-3	A-6	91.2%
머신러닝모델	최소값	평균	최대값
KNN	66	93	100
Random Forest	91	100	100

개 목적변수값(결과)의 도수분포표를 나타낸다. 모델 번호 20, 46, 57번에서도 이와 같은 500개 목적변수값들(결과)의 예측값과 통계값을 산출하였다.

VI. 결론

본 논문에서는 AIoT 기술 교육에 필요한 머신러닝 모델 활용 사례를 개발하였다. 본 개발 사례는 IoT 디바이스에 AI를 결합하여, AIoT 디바이스가 어떤 정보를 정확하게 예측할 수 있는가를 스스로 결정하는 사례로서, 미래 자율형 사물인터넷의 개발에 있어 활용도가 높다고 할 수 있다. 또한, AIoT 디바이스의 센서들이 측정하는 다양한 범위의 수치데이터에 대해 확장하여 적용할 수 있는 데이터 전처리 과정도 포함하였다. 실제 AIoT 시스템 구현에는 Orange3를 사용하지 않으나, 머신러닝 모델이 생성되고 데이터가 예측되는 과정을 교육 시에는 Orange3를 사용하는 것이 효과적이다. 추후 연구에서는 본 논문에서 개발한 사례를 적용하여, AIoT

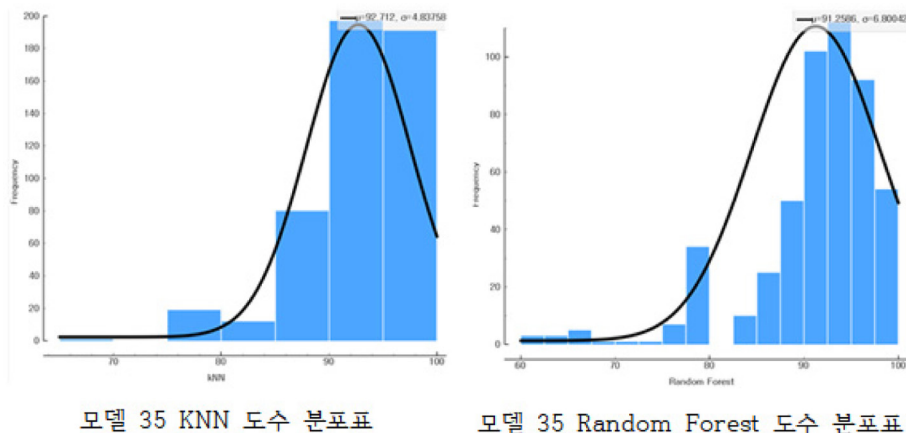


그림 7. 유효한 모델 35번에서 예측된 500개 목적변수값(결과)의 도수분포표
Fig. 7. Frequency distribution graph of 500 target variable values (results) predicted from valid model 35.

디바이스의 예측 기반 사물 제어가 가능한 실증 분야를 선정하고, 예측 기반 사물 제어의 효과성을 분석하고자 한다.

참고문헌

- [1] J. Y. Lee, "Technology trends of smart energy field using AIoT," *The Korean Institute of Electrical Engineers*, vol. 72, no. 12, pp. 10-16, December 2023.
- [2] S. H. Lee, S. M. Kim, and Y. H. Kim, "Design of python block coding platform for AIoT physical computing education," *Proceeding of Korea Society of Computer Information*, vol. 30, no. 1, pp. 1-2, 2022.
- [3] J. Y. Hong, "The impact of AIoT education programs using python on teachers' digital and AI literacy," *Journal of the Korean Association of Information Education*, vol. 28, no. 1, pp. 59-72, February 2024.
- [4] Y. J. Jang, "Searching for the direction of data science education in the era of the 4th industrial revolution," *Integrated Humanities Research*, vol. 9, no. 10, pp. 155-180, 2017.
- [5] Y. S. Park and S. J. Lee, "Study on the direction of universal big data and big data education-based on the survey of big data experts," *Journal of the Korean Association of Information Education*, vol. 24, no. 2, pp. 201-214, 2020.
- [6] J. S. Lee, "A study on visualization methods and expressions of information design for big data," *Journal of Basic Design & Art*, vol. 14, no. 3, pp. 259-269, 2013.
- [7] Orange3, University of Ljubljana, 2023, [Online]. Available: <https://orangedatamining.com/>.
- [8] Machine learning mastery, "A Gentle Introduction to k-fold Cross-Validation," [Online]. Available: <https://machinelearningmastery.com/k-fold-cross-validation/>.
- [9] K. S. Yang, T. S. Whang, D. S. Oh, C. J. Park, and H. S. Lim, "Cross-validated ensemble methods in natural language inference," *Journal of the Korean Institute of Information Scientists and Engineers (KIISE)*, vol. 48, no. 2, pp. 154-159, 2021.
- [10] IBM, "What is the k-nearest neighbors (kNN) algorithm?," [Online]. Available: <https://www.ibm.com/topics/knn>



허 경 (Kyeong Hur)_종신회원

1998년 : 고려대 전자공학과 학사
 2000년 : 고려대 전자공학과 석사
 2004년 8월 : 고려대 전자공학과 통신공학박사
 2004년 8월 ~ 2005년 8월 : 삼성종합기술원(SAIT) 전문연구원
 2005년 9월 ~ 현재 : 경인교대 컴퓨터교육과 교수
 <관심분야> 네트워크 MAC QoS, IoT, SW교육, AI교육, 데이터과학교육