

<https://doi.org/10.7236/JIIBC.2023.23.5.47>  
JIIBC 2023-5-8

## 데이터 저장소를 이용한 이상치 및 결측치 보정 시스템

# Adjustment System for Outlier and Missing Value using Data Storage

김광호\*, 김능회\*\*

Gwangho Kim\*, Neunghoe Kim\*\*

**요약** 현재 4차 산업혁명 시대에 오면서 다양하고 많은 데이터가 쌓여왔다. 농업사회도 스마트팜이나 노지에서 작물의 성장에 영향을 주는 환경 데이터를 센서로 수집해왔다. 환경 데이터는 측정하는 지역에 따라 특징을 가지고, 측정하는 시기에 따라라도 특징을 보이는 데이터이다. 수집된 농업 데이터를 활용해 통계, 인공지능을 사용하여 성장을 예측하거나 수확량을 예측하는 연구가 진행되어왔다. 이러한 연구는 기반이 되는 데이터에 따라 결과가 크게 차이난다. 이에 성능 향상을 위해서 데이터의 품질을 개선하기 위한 연구 또한 지속해서 진행되어왔다. 높은 성능을 위하여 많은 양의 데이터가 필요하고, 양이 충분하여도 데이터의 이상치나 결측치가 있을 경우 결과에 큰 영향을 미친다. 따라서 이상치와 결측치 값의 보정은 데이터 전처리 과정에서 필수이다. 이에 본 논문에서는 실제 농가에서 수집된 데이터를 통합하고 이를 기반으로 이상치와 결측치 보정 시스템을 제안한다.

**Abstract** With the advent of the 4th Industrial Revolution, diverse and a large amount of data has been accumulated now. The agricultural community has also collected environmental data that affects the growth of crops in smart farms or open fields with sensors. Environmental data has different features depending on where and when they are measured. Studies have been conducted using collected agricultural data to predict growth and yield with statistics and artificial intelligence. The results of these studies vary greatly depending on the data on which they are based. So, studies to enhance data quality have also been continuously conducted for performance improvement. A lot of data is required for high performance, but if there are outlier or missing values in the data, it can greatly affect the results even if the amount is sufficient. So, adjustment of outlier and missing values is essential in the data preprocessing. Therefore, this paper integrates data collected from actual farms and proposes a adjustment system for outlier and missing values based on it.

**Key Words** : Data Storage, Outlier Data, Missing Data

\*학생회원, 군산대학교 소프트웨어학과

\*\*정회원, 군산대학교 소프트웨어학과(교신저자)

접수일자 2023년 4월 7일, 수정완료 2023년 9월 3일

게재확정일자 2023년 10월 6일

Received: 7 April, 2023 / Revised: 3 September, 2023 /

Accepted: 6 October, 2023

Corresponding Author: nunghoi@kunsan.ac.kr

School of IT Convergence Engineering, Shinhan University,  
Korea, Department of Software Science & Engineering, Kunsan  
National University, Korea

## I. 서 론

현재 4차 산업혁명 시대에 오면서 기술적 혁신이 계속 진행되었다. 기술적 혁신은 다양한 분야에서 성과를 내보였고, 이로 인해 다양한 분야에서 많은 양의 데이터들이 쌓여왔다<sup>[1]</sup>. 농업사회에서도 기술의 발전으로 스마트팜이나 노지에서 작물의 성장에 영향을 주는 환경 데이터를 센서로 수집해왔다<sup>[2,3]</sup>. 환경 데이터는 측정하는 지역의 특성과, 측정하는 시기 등에 따라 다양한 데이터를 보인다<sup>[4]</sup>.

수집된 농업 데이터를 활용해 통계적 방법이나 인공지능 기술을 사용하여 성장을 예측하거나 수확량을 예측하는 등의 연구가 지속해서 진행되어왔다. 이러한 연구의 경우 기반이 되는 데이터에 따라 결과가 크게 차이가 나타난다. 이에 성능을 향상하기 위해서 기반이 되는 데이터의 품질을 개선하기 위한 연구 또한 함께 지속적으로 진행되어왔다. 높은 성능의 결과를 도출하기 위해서는 많은 양의 데이터가 필요하다. 양이 부족하다면 데이터가 충분한 정보를 제공하지 않기 때문에 정확한 값을 도출하는데 무리가 있고 결과의 신뢰성이 떨어진다. 하지만 데이터의 양이 충분하다고 하더라도 정상적인 값이 아닌 값을 가지는 이상치나 중간에 값이 비어있는 결측치 또한 결과에 큰 영향을 미친다. 따라서 이상치와 결측치 값의 보정은 데이터 전처리 과정에서 필수이며 중요하다<sup>[5,6,7,8]</sup>.

이에 본 논문에서는 실제 농가에서 수집되는 데이터들을 통합하고 이를 기반으로 하여 이상치와 결측치를 탐색하고 통합한 데이터를 기반으로 보정을 진행하여 데이터의 품질을 높이기 위한 데이터 저장소를 이용한 이상치 및 결측치 보정 시스템을 제안한다.

## II. 배경

### 1. 이상치 데이터

통계적 측면에서는 이상치 데이터를 관측된 데이터들이 모여 있는 지점이 아닌 멀리 떨어진 곳에 따로 있는 데이터를 이상치 데이터로 정의하였다. 이러한 이상치는 합리적인 이상치, 비합리적인 이상치 두 가지로 분류된다. 합리적인 이상치는 어떠한 영향을 받지 않고 정확하게 측정된 데이터로 확인되지만 다른 데이터와 비교에서 전혀 다른 특성이나 값을 나타내는 것을 말한다. 비합리적인 이상치는 입력 오류, 자료 오염처럼 데이터의 변형으로 인해 발생한 이상치를 말한다<sup>[9][10]</sup>.

### 2. 결측치 데이터

결측치 데이터는 크게 세 종류로 나누어지는데 완전 무작위 결측, 무작위 결측, 비 무작위 결측이다. 첫 번째 완전 무작위 결측은 전체 데이터에서 변수의 종류와 수치에 상관없이 비슷한 분포를 두고 빠진 데이터를 의미한다. 이 경우 빠진 데이터의 패턴을 확인할 수 있다. 하지만 실제 데이터에서 이러한 경우는 거의 없다. 두 번째 무작위 결측은 특정 변수의 데이터가 빠지는 경우이며 이 경우 데이터 안의 다른 변수와 관련성은 있지만 그 변수의 수치와는 관계가 없다. 마지막으로 비 무작위 결측은 대부분의 결측치 데이터가 여기에 해당한다. 빠진 부분이 무작위로 나타나며 결측치 값이 결측 사유와 관련성이 있는 경우이다. 예를 들면 데이터를 수집하는데 있어서 네트워크나 센서의 고장으로 데이터가 수집되지 않아서 결측치로 저장되는 경우가 여기에 해당한다<sup>[11]</sup>.

## III. 데이터 저장소를 이용한 이상치 및 결측치 보정 시스템

환경 데이터는 지역과 시기에 영향을 많이 받는다. 지역 환경에 따라 날씨도 다르고 토양도 달라서 지역에 따라 환경 데이터는 큰 차이를 보인다. 또한, 1년을 봄, 여름, 가을, 겨울로 구분하듯 시기 또한 영향력이 크다. 하우스 농업으로 하우스 내부 환경의 조정이 쉬워지고, 각종 설비를 사용해 유지하는 스마트팜이 생겼지만, 지역과 시기의 영향을 벗어날 수 없다. 이는 데이터 탐색과 보정을 진행할 때 같은 지역과 시기를 사용해야 하는 이유이다. 따라서, 탐색 및 보정은 같은 지역 및 근사한 시기의 데이터만으로 기준을 생성하고 진행한다<sup>[12]</sup>.

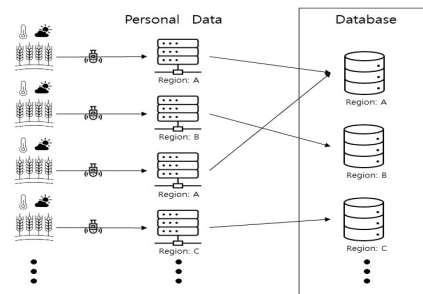


그림 1. 시스템 아키텍처  
Fig. 1. System architecture

그림 1은 본 논문에서 제안하는 시스템의 전체적인 아키텍처이다. 각 농가의 센서를 통해 수집한 환경 데이터를 저장소에 전송해 저장한다. 이렇게 전송된 데이터들을 통합한 하나의 기준 데이터베이스를 생성한다. 통합한 데이터베이스를 기준으로 수집된 데이터의 이상치와 결측치를 탐색하고 사용자에게 알람을 준다. 사용자는 알람이 울린 데이터를 확인하고 직접 수정하거나 시스템이 보정하도록 한다. 또한, 통합 데이터베이스는 실시간으로 데이터를 갱신한다.

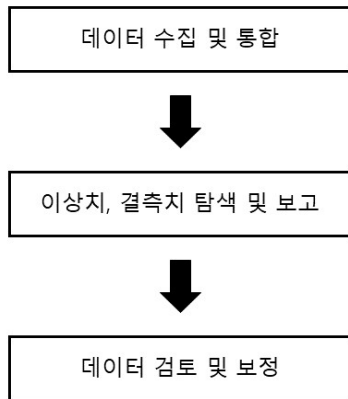


그림 2. 시스템 순서도  
 Fig. 2. System sequence

그림 2는 본 논문에서 제안하는 시스템의 실행 순서이다. 동일한 지역에서 수집된 농가의 환경 데이터들을 통합하고 통합된 데이터를 기반으로 근사한 시기를 통하여 이상치 및 결측치를 탐색하고 보정하는 시스템을 제안한다. 제안하는 시스템은 총 3단계로 데이터 수집 및 통합, 이상치, 결측치 탐색 및 보고, 데이터 검토 및 보정 단계로 이루어져 있다.

### 1. 데이터 수집 및 통합

작물의 성장을 위해서는 적당한 온도와 습도처럼 적합한 환경을 조성하는 것이 아주 중요하다. 이러한 환경은 생산하는 지역과 시기에 밀접한 관련성을 가지고 있다. 이에 본 연구에서는 지역별로 농가의 환경 데이터를 수집하고 수집된 환경 데이터에서 근사한 시기를 사용한다. 환경 데이터의 수집은 각 농가에 설치된 센서들을 통해 수집한다. 그리고 이렇게 수집된 데이터들은 데이터의 형태를 통일시키기 위한 전처리 과정을 거치고 지역별 하나의 통합된 데이터로써 가공한 후 데이터베이스에

저장한다. 그림 3은 데이터베이스에 저장된 통합 데이터의 예시이다.

Internal Humidity	Soil Humidity	...	Outside Temperature	Inside Temperature	Region
49.61	43.91	...	20.00	16.55	Gimje-si
60.43	42.90	...	20.77	14.39	Gimje-si

그림 3. 통합된 데이터의 예시  
 Fig. 3. Example of the integrated data

### 2. 이상치, 결측치 탐색 및 보고

농가들로부터 환경 데이터는 지속적으로 측정되고 수집된다. 수집된 환경 데이터를 활용하기 위해 추출한 환경 데이터의 이상치와 결측치를 확인한다.

동일 작물, 같은 지역의 데이터를 데이터베이스에서 추출해 데이터셋을 생성한다. 이후 추출된 데이터셋에서 근사한 시기의 범위를 정하고 측정 데이터마다 정해진 근사한 시기 내에 최댓값과 최솟값을 비교하여 최댓값과 최솟값을 벗어나는 이상치를 탐색한다. 그리고 측정 데이터에서 비어있는 결측치 또한 함께 탐색한다. 탐색은 한 농가씩 진행되며 선택된 한 농가와 나머지 다른 농가들의 평균값을 이용하여 통합 데이터 전부를 탐색한다. 전체 측정 데이터에 대해 탐색이 완료되면 이상치와 결측치의 탐색된 결과를 보고한다.

```

standard_data
target_data
error_data

for(i in range(standard_data)):
    MIN = i.min
    MAX = i.max
    for(j in range(target_data)):
        if(j[i] < MIN and j[i] > MAX):
            error_data.append(j[i])

return error_data
    
```

그림 4. 이상치, 결측치 탐색 및 보고  
 Fig. 4. Outlier, missing detection and report

그림 4는 이상치, 결측치 탐색 및 보고의 의사코드이다. standard\_data의 최댓, 최솟값을 도출한 후 이후 보정하려는 target\_data의 값과 비교를 통해 이상치와 결

측치 데이터를 탐색한다. 탐색된 데이터들은 error\_data에 저장해 두었다가 마지막에 사용자에게 보고하는 방식이다.

### 3. 데이터 검토 및 보정

앞서 탐색된 이상치와 결측치 데이터들을 측정자에게 검토받고 보정하는 단계이다. 측정자는 시스템이 탐색해 보여주는 결과를 검토한다. 검토 후 이상치와 결측치 데이터를 직접 입력해 수정하거나 시스템이 보정하도록 할 수 있다. 시스템이 보정하는 경우에는 이상치와 결측치 모두 동일한 방식으로 보정한다. 이상치와 결측치에서 정해진 근사한 시기 내에 최댓값과 최솟값의 평균값을 사용한 통계적 기준을 사용하여 보정을 진행한다. 완료 후 이상치와 결측치가 보정된 환경 데이터를 활용한다.

## IV. 실험

본 논문에서 제안하는 시스템의 이해를 돕기 위해 실제 환경 데이터를 활용하여 시스템의 각 단계별로 설명한다.

### 1. 데이터 수집 및 통합

본 논문에서 제안하는 방법에 따르면 환경 데이터의 수집은 각 농가에 설치된 센서들을 통해 수집한다. 하지만 수집 기간으로 인해 스마트팜 데이터 마트에서 제공하는 환경 데이터를 이용하였다<sup>[13]</sup>. 이는 실제 농가들에 설치된 센서들을 통해 수집된 데이터이다.

사용한 데이터는 2021년 전라북도 김제시에서 딸기를 생산하는 농가들의 환경 데이터이다. 하우스 형태에 온실 규모는 3000평 이상의 대규모 농가들로 9월부터 12월까지 측정한 환경 데이터이며 이를 데이터베이스에 저장하여 활용하였다. 그림 5는 사용한 통합 데이터이다.

Measurement Date	Internal Humidity	Insolation	Outside Temperature	Inside Temperature	Region
09-10	93.69	144.62	22.64	23.66	Gimje-si
09-11	96.55	280.18	24.11	25.08	Gimje-si
09-12	98.40	177.04	24.64	25.44	Gimje-si
...	...	...	...	...	...
12-31	61.78	99.17	-3.65	7.20	Gimje-si

그림 5. 사용한 통합 데이터  
Fig. 5. Integrated data used

### 2. 이상치, 결측치 탐색 및 보고

통합 데이터에서 한 농가의 환경데이터를 탐색 기준으로 선정하여 이상치, 결측치를 생성하였다. 그리고 본 논문의 제안에 따라 이상치, 결측치를 탐색하고 보고를 진행하였다. 근사한 시기는 이상치, 결측치를 기준으로 전과 후 한 달의 범위를 정하였고 현재의 한 농가와 나머지 다른 농가들의 평균값을 이용하여 탐색하였다. 그림 6은 탐색이 완료된 데이터 중 일부이다. 그림에 표시된 데이터처럼 벗어나는 이상치나 비어있는 결측치는 데이터의 날짜와 속성값을 저장해 사용자에게 반환한다.

Measurement Date	Internal Humidity	Insolation	Outside Temperature	Inside Temperature	Region
10-01	56.41	204.53	20.49	14.91	Gimje-si
10-02		213.29	21.68		Gimje-si
10-03	54.75	221.35	-22.56	15.72	Gimje-si
...	...	...	...	...	...
10-30	59.56	20.65	11.9	9.57	Gimje-si
10-31	55.62	160.82	12.19	10.39	Gimje-si

그림 6. 탐색된 데이터  
Fig. 6. Detected data

그림 7은 사용자가 시스템으로 데이터 탐색 구간을 설정하는 화면이다. 본 논문에서는 탐색 구간을 10월로 선정해서 진행하였다. 본 시스템은 사용자가 선택한 구간에 따라 이상치를 탐색하는 기준치가 변경된다.

start date

2021 . 10 . 01 . 📅

end date

2021 . 10 . 31 . 📅

search

all date

그림 7. 탐색 구간 설정  
Fig. 7. Settings for search section

그림 8은 구현한 시스템을 사용해 실험한 데이터의 일부이다. 결측치의 경우 실제 데이터에는 빈 칸으로 되어 있지만 실험의 편의성을 위해서 nan으로 표시 하도록 초기화를 진행하였다.

2021-10-16	58.2	88.69	13.74	11.0	Gimje-si
2021-10-17	53.44	220.48	6.29	9.73	Gimje-si
2021-10-18	55.54	208.77	8.63	10.98	Gimje-si
2021-10-19	58.83	139.44	-10.98	9.93	Gimje-si
2021-10-20	57.07	185.94	8.53	10.09	Gimje-si
2021-10-21	56.81	175.89	8.94	10.0	Gimje-si
2021-10-22	55.54	179.65	8.98	9.64	Gimje-si
2021-10-23	54.26	199.88	9.64	9.96	Gimje-si
2021-10-24	56.18	nan	10.26	10.1	Gimje-si
2021-10-25	56.14	177.13	9.7	9.75	Gimje-si
2021-10-26	55.67	166.4	10.53	10.27	Gimje-si
2021-10-27	56.27	151.66	12.09	10.21	Gimje-si

그림 8. 사용한 데이터  
 Fig. 8. Used data

그림 9는 구현한 시스템을 사용하여 탐색된 데이터를 사용자에게 반환한 화면이다. 사용자가 선택한 기간의 데이터를 대상으로 기존 데이터의 최댓, 최솟값을 통해 이상치와 결측치를 탐색해 저장한다. 최종적으로 저장한 데이터만을 사용자에게 보여준다. 이후 데이터를 선택해 보정을 선택하면 보정 화면으로 진행하게 된다.

measurement date	Internal Humidity	insolation	outside temperature	inside temperature	region	outlier/missing	
2021-10-24	56.18	NAN	10.26	10.1	Gimje-si	insolation	edit
2021-10-02	55.09	213.29	21.68	15.57	Gimje-si	outside temperature	edit
2021-10-03	54.75	221.35	22.56	15.72	Gimje-si	outside temperature	edit
2021-10-04	57.0	163.98	23.79	16.37	Gimje-si	outside temperature	edit
2021-10-05	53.24	220.55	22.5	15.92	Gimje-si	outside temperature	edit
2021-10-06	52.97	228.83	21.48	15.35	Gimje-si	outside temperature	edit
2021-10-09	55.72	187.33	23.31	16.19	Gimje-si	outside temperature	edit
2021-10-10	55.38	180.94	23.7	16.48	Gimje-si	outside temperature	edit
2021-10-19	58.83	139.44	-10.98	9.93	Gimje-si	outside temperature	edit

home

그림 9. 시스템에서 탐색된 데이터  
 Fig. 9. Detected data in the system

save

measurement date	Internal Humidity	insolation	outside temperature	inside temperature	region	outlier/missing
2021-10-24	56.18	NAN	10.26	10.1	Gimje-si	insolation
2021-10-24	56.18	autoedit 120.88	autoedit 10.26	autoedit 10.1	autoedit Gimje-si	insolation

그림 11. 보정 화면  
 Fig. 11. Adjustment screen

### 3. 데이터 검토 및 보정

측정자는 반환된 데이터들을 검토하면서 데이터를 직접 입력해 수정하거나 시스템이 보정하도록 선택할 수 있다. 직접 수정하는 경우 측정자가 입력한 값으로 기존의 데이터를 수정한다. 그리고 시스템이 보정할 때는 탐색된 이상치나 결측치 데이터를 탐색 및 보고 단계에서 이상치를 분류하기 위해 추출된 최댓, 최솟값의 평균을 계산해 그 결과 값으로 데이터를 보정한다. 그림 10은 보정이 완료된 데이터 중 일부이다.

Measurement Date	Internal Humidity	Insolation	Outside Temperature	Inside Temperature	Region
10-01	56.41	204.53	20.49	14.91	Gimje-si
10-02	68.72	213.29	21.68	14.55	Gimje-si
10-03	54.75	221.35	13.32	15.72	Gimje-si
...	...	...	...	...	...
10-30	59.56	105.56	11.9	9.57	Gimje-si
10-31	55.62	160.82	12.19	10.39	Gimje-si

그림 10. 최종 보정 데이터  
 Fig. 10. Final adjustment data

그림 10의 강조된 부분이 보정된 내용이다. 표시된 데이터는 모두 시스템을 통해서 보정한 값으로 수정된 내용이다.

그림 11은 데이터의 보정을 선택한 후의 화면이다. 각 데이터의 현재 값과 변경 값을 입력하는 텍스트 박스가 출력된다. 사용자는 변경하려는 값을 직접 입력해 변경할 수 있고, 자동보정 버튼을 통해 보정을 진행할 수 있다. 그림 11의 insolation 속성의 입력된 값은 시스템의 자동보정을 통해 진행한 결과이다. 보정을 마치고 save를 통해 보정한 데이터를 저장한다.

그림 12는 보정된 데이터의 일부이다. 2021년 10월 24일 데이터의 3번째 속성이 그림 11에서 보정한 insolation으로 저장된 것을 볼 수 있다.

2021-10-16	58.2	88.69	13.74	11.0	Gimje-si
2021-10-17	53.44	220.48	6.29	9.73	Gimje-si
2021-10-18	55.54	208.77	8.63	10.98	Gimje-si
2021-10-19	58.83	139.44	-10.98	9.93	Gimje-si
2021-10-20	57.07	185.94	8.53	10.09	Gimje-si
2021-10-21	56.81	175.89	8.94	10.0	Gimje-si
2021-10-22	55.54	179.65	8.98	9.64	Gimje-si
2021-10-23	54.26	199.88	9.64	9.96	Gimje-si
2021-10-24	56.18	120.88	10.26	10.1	Gimje-si
2021-10-25	56.14	177.13	9.7	9.75	Gimje-si

그림 12. 보정된 데이터  
Fig. 12. Adjusted data

## V. 결론 및 향후 과제

본 논문은 지역별 농가로부터 수집되는 데이터를 통해 이를 기반으로 이상치와 결측치를 보정하는 시스템을 제안하였다. 실시간으로 농가의 데이터를 수집해 통합하고 이를 바탕으로 이상치와 결측치를 탐색해 사용자에게 알림을 주며 데이터를 자동 또는 수동으로 보정한다.

제안하는 시스템은 현 농가들의 측정 데이터를 기반으로 데이터를 보정하기 때문에 지역의 환경에 대한 특성과 시기에 따라 생기는 영향을 반영하여 데이터를 보정할 수 있다. 이러한 결과는 데이터의 신뢰성을 높이고 이상치와 결측치를 보정하여 품질도 개선한다. 또한, 추후 농가 데이터를 활용한 연구에서 더욱 정확한 결과를 도출할 수 있도록 긍정적인 영향을 줄 것이다.

추후 과제로는 실제 농가들과 연결해 실시간으로 데이터 보정을 진행하고 데이터를 축적하면서 이상치 데이터를 판단하기 위한 기간을 선정하는데 효율적인 기간을 비교해 구해내고자 한다.

## References

- [1] Sang-Cheol Kim, "4th industrial revolution and development of smart farm technology", Magazine of the Korean Society of Agricultural Engineers, Vol. 59, No. 2, pp.10-18, May 2017.
- [2] Hyeong-Seok Lee, Hyo-Jung Kim, Seong-jun Kim, Su-Ho Jeong, Ga-Yun Yoon, Yeon-Jin Cho, Joo-Hye Kim, Jong-Mo Jung, Hee-Gon Kim, "Implementation of onion image-based growth and environmental data collection system", Proceedings of Symposium of the Korean Institute of communications and Information Sciences, Vol. 78, No. 1, pp.731-732, Jun 2022.
- [3] Won Suk Lee, Kyungja Son, Daeho Jun, Yongtae Shin, "Big Data Activation Plan for Digital Transformation of Agriculture and Rural", KIPS Transactions on Software and Data Engineering, Vol. 9, No. 8, pp.235-242, Jul 2020.  
DOI: <https://doi.org/10.3745/KTSDE.2020.9.8.235>
- [4] Mahn-Jo Kim, Uk Lee, Yun Mi Park, Suk-In Hwang, "Effects of Region and Age on Nut Quality of Major Chestnut Cultivars (Castanea spp.) cultivated in the Central Region of Korea", Korean Society of Forest Science, Vol. 100, No. 4, pp.616-622, Dec 2011  
DOI: <https://doi.org/10.14578/jkfs.2011.100.4.11>
- [5] Cheol Won Lee, Su-Yong An, Jae-Young Kim, Hyeongtae Ahn, "Sensor anomaly detection system in greenhouse-type smart farm using environmental data", Journal of the Korean Data And Information Science Society, Vol. 32, No. 6, pp.1237-1248 Nov 2021.  
DOI: <https://doi.org/10.7465/jkdi.2021.32.6.1237>
- [6] Yong hee Shin, Eun-jeong Lee, "Development of Long-term Crop Yield Prediction Using Seasonal Forecast Data", Proceedings of the Korean Society of Agricultural Engineers Conference, Vol. 2015, No. 0, pp.280-280, Oct 2015.
- [7] Sungho SHin, Mikyoung Lee, Sa-kwang Song, "A Prediction Model for Agricultural Products Price with LSTM Network", The Journal of the Korea Contents Association, Vol. 18, No. 11, pp.416-429, May 2018  
DOI: <https://doi.org/10.5392/JKCA.2018.18.11.416>
- [8] Dain Park, Sanghoo Yoon, "A Missing Value Replacement Method for Agricultural Meteorological Data Using Bayesian Spatio-Temporal Mode", Journal of environmental science international, Vol. 27, No. 7, pp.499-507, Apr 2018.
- [9] Jin-Hwi Kim, "Statistical methods for outlier detection", HIRA Policy Brief, Vol. 14, No. 1, pp. 49-56, Jan. 2020.
- [10] Jeong Yeon Seon, Ki Young Kim, Jin-Hwi Kim, "A Study on Statistical Methods and Application for Outlier Detection", Health Insurance Review & Assessment Service, Vol. 2019, No. 0, pp.1-84, Aug. 2019.
- [11] Eden Kim, Seok-gap Seok, Seung-cheol Son, Byeong-tak Lee, "Technical Trends of Time-Series Data Imputation," Electronics and Telecommunications Trends, Vol. 36, No. 4, pp. 145-153, Aug. 2021.  
DOI: <https://doi.org/10.22648/ETRI.2021.J.360414>
- [12] Jung-Tae Hwang, Young-Gon Kim, "Development of Multi-Crop Smart Farm Management System for User Convenience based on Lab-Vie", The Journal of the Institute of Internet, Broadcasting and Communication, Vol. 22, No. 1 pp.15-20, Feb 2022  
DOI: <https://doi.org/10.7236/JIIBC.2022.22.1.15>
- [13] <https://data.smartfarmkorea.net/structuredData/selectContHortiCultureDataViewLists.do>

## 저 자 소 개

### 김 광 호(학생회원)



- 2017년 3월 ~ 현재 : 군산대학교 소프트웨어학과 학사과정
- 주관심분야 : 빅데이터, 자연어처리

### 김 능 회(정회원)



- 2008년 8월 : 고려대학교 컴퓨터학과 (이학석사)
- 2017년 8월 : 고려대학교 컴퓨터 · 전파통신공학과 (공학박사)
- 2021년 3월 ~ 현재 : 군산대학교 소프트웨어학부 소프트웨어학 전공 조 교수
- 주관심분야 : 요구공학, 소프트웨어공학

※ 이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-현장연계 미래선도인재양성 지원사업의 지원을 받아 수행된 연구임(No. NRF-2022H1D8A3038669).