

DataPlatform as a Service(DPaaS)를 위한 분산 클라우드 기반의 딸기 생산량 최적화 예측 모델 선정에 관한 연구

이 세 연*, 김 준 경*, 김 혜 린**, 유 정 민**, 손 아 영°

A Study on Optimized Prediction Model Selection of Strawberry Production for DPaaS in Distributed Cloud Computing

Se-Yun Lee*, Junkyeong Kim*, Hye-Rin Kim**, Jeong-Min Yoo**, A-Young Son°

요 약

IoT 기기들의 발전으로 다양한 데이터가 수집되고 서비스가 제공되면서 데이터를 효율적으로 관리할 수 있는 데이터 플랫폼이 요구된다. 이에 따라 데이터 플랫폼 연구들이 진행 중이지만, 실제 활용하기 위한 활성화 방안이 필요하다. 본 논문에서는 DPaaS를 위한 분산 클라우드 기반의 데이터 플랫폼을 설계하며, 스마트팜 서비스 시나리오를 통해 데이터 플랫폼의 활용을 보인다. 구상한 시나리오는 딸기 생산량 예측 모델로 LSTM, convLSTM, GRU 알고리즘을 사용하여 비교 분석하였고, 실험 결과 LSTM 알고리즘의 RMSE 값이 0.6828로 최적의 알고리즘으로 선정하여 데이터 플랫폼 모델에 적용하였다. 향후 사용자 접근이 편리한 형태로 개선하여 데이터 플랫폼을 활용하고자 하며, 이를 통해 데이터 플랫폼 뿐만 아니라 스마트팜 관련 연구에도 활용될 수 있기를 기대한다.

키워드 : 분산 클라우드, 데이터 플랫폼, 예측 분석 모델링, 데이터 학습, 최적화, 데이터 전처리

Key Words : Distributed Cloud Computing, Data Platform, Modeling for Prediction and analysis, Data Learning, Optimization, Data Preprocessing

ABSTRACT

With the development of IoT devices and diverse of service based on various of data, a data platform that can efficiently manage data is required. Accordingly, existing work for data platform have been researched. It still need to method to boost utilization for using of data platform. In this paper, we design DPaaS based distributed cloud and demonstrate the utilization of the data platform through a smart farm service scenario. According to scenario we evaluate the using the LSTM, convLSTM, and GRU algorithms for prediction of strawberry production. As result of the experiments the RMSE value of the LSTM algorithm was 0.6828, which was selected as the optimal algorithm and applied to the learning model in DPaaS. In the future, we plan to utilize the data platform by improving for convenient user access. It is expected that DPaaS will be used not only data platform research but also related research for smart farm.

※ 본 연구는 차세대융합기술연구원의 지원을 받아 수행되었습니다.(AICT-2021-0001)

• First Author : Advanced Institute of Convergence Technology, seyuncom@snu.ac.kr, 정희원

° Corresponding Author : Advanced Institute of Convergence Technology, ayths10@snu.ac.kr, 정희원

* Advanced Institute of Convergence Technology, junkyeong@snu.ac.kr

** 소속없음(차세대융합기술연구원 일경협수련생), yoonist08@gmail.com; hea990@ajou.ac.kr

논문번호 : 202111-310-0-SE, Received October 31 22, 2021; Revised December 28, 2021; Accepted January 11, 2022

I. 서론

최근 IoT(Internet of Things, 사물인터넷) 기술과 빅데이터, 인공지능 등의 기술 발달과 함께 분산된 클라우드 기반으로 다양한 형태의 서비스들이 제공되고 있으며 그 과정에서 많은 데이터들이 생성되고 있다. 데이터는 미래 경쟁력을 좌우하는 원유로 표현¹⁾되고 있어 그 관심과 중요성은 매우 높다고 할 수 있다.

다양한 분야에서 대용량의 데이터가 쏟아지고 있어 이를 수집 및 분석하여 관리할 수 있는 데이터 플랫폼 구축과 관련된 연구가 진행되고 있지만, 실제 활용하는 사례는 적은 편이다. 윤창희의 연구²⁾에서는 데이터 플랫폼 중에서도 공공분야의 인공지능 플랫폼에 관한 연구로 정부·민간·학계가 협력하여 인공지능 플랫폼 구축이 필요하다는 주제로 연구를 진행하였고, 김아영의 연구³⁾에서는 오픈 소스인 Kubernetes 및 Docker를 사용하여 플랫폼을 구축하고 TensorFlow를 통해 효율적인 딥러닝 분석을 할 수 있는 데이터 플랫폼에 대한 연구를 진행하였다. 또한 데이터 3법이 국회를 통과하면서 정부나 공공기관들이 내부 데이터를 개방 및 공유하기 위해 데이터 플랫폼에 투자를 하고 있다. 그러나 개방된 데이터들은 권한이나 기간에 따라 제한적인 경우가 많아 활성화 시키기 위해서는 데이터 및 플랫폼을 제공하는 기관과 사용자들의 노력이 요구된다.

데이터 플랫폼을 기반으로 한 서비스 중 스마트팜(Smart Farm) 서비스는 센서를 통해 수집한 환경 정보(온도, 습도, 일사량, 이산화탄소 등) 및 생육 정보(잎의 크기, 줄기 굵기, 꽃 개수, 열매 개수 등)를 관리를 통해 수확량, 품질 등을 향상시켜 수익성을 높일 수 있는 방법들이 제시되고 있다. 또한 이렇게 관리된 데이터는 소비자들에게 신뢰도를 높여줄 수 있어 중요성이 커지고 있다. 그러나 현재의 스마트팜은 단순히 센서들을 이용해 데이터를 수집하고, 이를 PC나 스마트폰 등을 통해 원격으로 모터를 동작시켜 자동으로 비닐하우스 창 등을 개방하거나 양액을 공급하는 등의 용도로만 사용되는데 그치고 있어 이에 대한 활용 방안이 필요하다.

스마트팜 활용을 위해 다양한 연구가 진행되었다. 노희선의 연구⁴⁾에서는 토마토 스마트팜의 생육 데이터를 통해 상관 관계를 고려한 능형회귀분석(Ridge Regression) 모델을 적용하여 수확량의 연관성을 분석하였고, 홍성은의 연구⁵⁾에서는 다양한 학습 모델을 사용하여 실험하고, 결과적으로 성능이 높았던 ConvLSTM(Convolution Long Short Term Memory)

를 사용해 토마토 생산량 및 성장량 예측 모델에 관한 연구를 진행하였다. 대부분의 연구들은 토마토 작물에 관한 것으로 다른 작물로 연구를 확장하여 활용성과 신뢰성을 높이는 연구가 필요하다.

본 논문에서는 분산 클라우드 기반의 데이터 플랫폼의 실제적인 활용을 위해 다음과 같은 목표를 가진다.

- DPaaS(Data Platform as a Service)를 위한 분산 클라우드 기반의 데이터 플랫폼 구조 설계
- DPaaS를 이용한 데이터 분석 서비스 제공에 따른 데이터 플랫폼의 활성화
- 스마트팜 서비스 제공을 위한 시나리오를 통해 딸기 작물에 관한 생산량 예측 및 최적화 모델 선정에 관한 연구

본 논문은 다음과 같이 구성되어 있다. 2장 관련 연구에서 기존 데이터 플랫폼에 대한 분석과 분산 클라우드에 대한 연구 방향을 살펴보고, 딸기 생산량 예측을 위한 방법론에 관해 기술한다. 3장에서 데이터 학습 플랫폼을 설계하고 주요 기능 및 시나리오를 통해 동작 과정을 보인다. 4장 실험 및 평가에서 딸기 생산량 예측을 위한 분석 모델을 구성하여 데이터 전처리 과정부터 모델 구성, 분석 및 성능 평가를 진행한다. 마지막으로 5장 결론에서 DPaaS를 위한 분산 클라우드 기반의 데이터 플랫폼의 향후 활용 방안과 연구 계획에 대해 제시하며 마친다.

II. 관련 연구

2.1 기존 데이터 플랫폼 분석

데이터 플랫폼이란 다양한 데이터들을 수집하고, 저장하며 수집한 데이터를 처리 및 분석하여 시각화 및 서비스를 제공해주는 IT 기술을 말한다.

데이터 플랫폼 현황 분석을 위해 키워드를 “데이터 플랫폼”으로 하여 최근 1년 동안의 국내 논문 653건을 비교·분석한 결과, 데이터 플랫폼 개발을 위해 진행되는 작업인 설계^{6,7)}, 연구^{2,8)}, 구축 방안^{9,10)}과 같은 주제가 559건(약 85.6%)으로 대부분을 차지하는 반면, 사례^{11,12)}, 분석^{13,14)}과 같이 실제 데이터 플랫폼을 활용하면서 작성된 주제의 논문은 94건(약 14.4%)으로 아직까지 데이터 플랫폼에 대한 연구는 설계하고 방안을 제시하는 연구적인 측면이 높았고 실제 수집 및 분석된 데이터를 활용하는 플랫폼 활용과 관련해서는 더디다고 볼 수 있었다.

정부에서도 2025년까지 금융, 교통, 지역 경제 등

15개 분야에서 대표적인 빅데이터 플랫폼을 구축^[15]하기 위해 발전 전략을 제시하는 것으로 미루어볼 때, 데이터 플랫폼이 가지고 있는 산업·경제적 효과와 국민·기업의 편익은 크다고 할 수 있어 관련 연구가 필수적이다.

다음 그림 1은 다양한 데이터를 수집하여 데이터베이스에 가공하여 저장하고, 데이터 분석을 수행한 후 시각화 및 서비스 과정을 비롯하여 데이터 보호, 모니터링 및 관리하는 과정 모두를 포함하는 일반적인 형태의 데이터 플랫폼을 나타낸다.

따라서 본 논문에서는 데이터 플랫폼에 대해 그림 1을 기본 구조로 하여 실제 서비스를 제공할 수 있는 『DPaaS를 위한 분산 클라우드 기반의 데이터 플랫폼』을 제안한다.

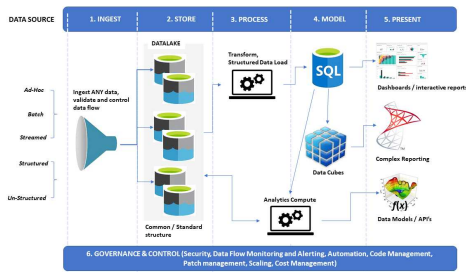


그림 1. 일반적인 형태의 데이터 플랫폼[16]
Fig. 1. The modern data platform[16]

2.2 분산 클라우드에 관한 기존 연구 방향

분산 클라우드란 기존의 클라우드 컴퓨팅의 장점을 최대한으로 활용하면서, 서비스를 받던 사용자가 다른 지역으로 이동하더라도 기존의 서비스를 연속적으로 받을 수 있도록 통신 오버 헤드, 비용 및 대기 시간을 줄일 수 있다는 특징이 있다. 또한 가트너가 뽑은 2021년 전략 기술 트렌드에도 분산 클라우드가 소개^[17]되고 있어 앞으로 그 중요성이 커지고 있다.

사물인터넷의 발달로 중앙 서버와 IoT 기기들의 중간 계층의 역할이 커져 완전 분산 모델인 에지-포그 클라우드에 대한 연구^[18]가 진행되고 있으며, IoT를 활용한 다양한 서비스를 통해 수집되는 데이터를 처리하기 위해 산업용 IoT의 엣지 컴퓨팅 플랫폼 기반의 데이터 분산 서비스를 제안^[19]하는 등 분산 클라우드 환경에서 데이터를 활용한 플랫폼과 관련된 연구도 진행되고 있다.

본 논문에서도 분산된 형태의 IoT Node들을 활용하여 효율적으로 데이터 수집, 분석, 서비스 순으로

제공하면서 분산 클라우드의 장점을 활용하기 위해 분산 클라우드 기반의 데이터 플랫폼인 DPaaS를 설계하여 서비스 제공을 하고자 한다.

2.3 딸기 생산량 예측을 위한 방법론

DPaaS의 활용을 위해, 스마트팜에 초점을 맞추어 딸기 생산량을 예측하여 데이터 플랫폼을 통해 수집·분석하여 서비스 형태로 제공하고자 한다.

기존 연구들에서는 환경 변수만을 가지고 분석^[20,21]하거나 생육 변수만을 가지고 분석^[4,22]하는 경향이 많았으나, 본 논문에서는 환경 변수와 생육 변수 모두를 고려하여 생산량을 예측할 수 있는 모델을 구성하였다.

딸기 생산량 예측 분석을 위해 사용한 데이터 분석 알고리즘으로는 다양한 변수(환경·생육 변수)와 딸기 성장 시간의 흐름에 따라 기록된 데이터를 활용하기 위해, 다변량 시계열 예측에 많이 활용되고 있는 RNN(Recurrent Neural Network) 기반으로 1997년에 Hochreiter&Schmidhuber가 제안한 LSTM(Long-Short Term Dependency)과 LSTM 알고리즘에 Convolution Layer를 추가한 ConvLSTM 알고리즘 및 LSTM보다는 단순한 형태로 2014년에 Cho, et al.에 의해 고안된 GRU(Gated Recurrent Unit) 알고리즘을 사용하였다.

III. 본 론

본 장에서는 DPaaS를 위한 분산 클라우드 기반의 데이터 플랫폼을 설계한 내용에 대해 기술한다. 1절에서는 구조와 데이터 분석 흐름도 및 특징을 나타내고 2절에서는 딸기 생산량 예측 모델을 시나리오로 하여 DPaaS의 동작 과정을 나타낸다.

3.1 DPaaS를 위한 분산 클라우드 기반 데이터 플랫폼 설계

3.1.1 DPaaS 전체 구조

DPaaS는 하나의 IoT 기기로 볼 수 있는 Edge들에서 실시간으로 데이터를 수집하고, 실시간 처리 및 단순한 분석이 가능해진 Edge Node^[23]를 활용한다. 대용량의 데이터 저장 및 관리와 복잡한 데이터 분석은 Main Server에서 처리하는 분산 클라우드 기반의 데이터 플랫폼 구조로, 다음 그림 2는 제안하는 데이터 플랫폼인 DPaaS의 구조이다.

DPaaS는 크게 Main Server와 각각의 Edge Node들, 데이터 수집을 실제적으로 담당하는 Edge들로 구

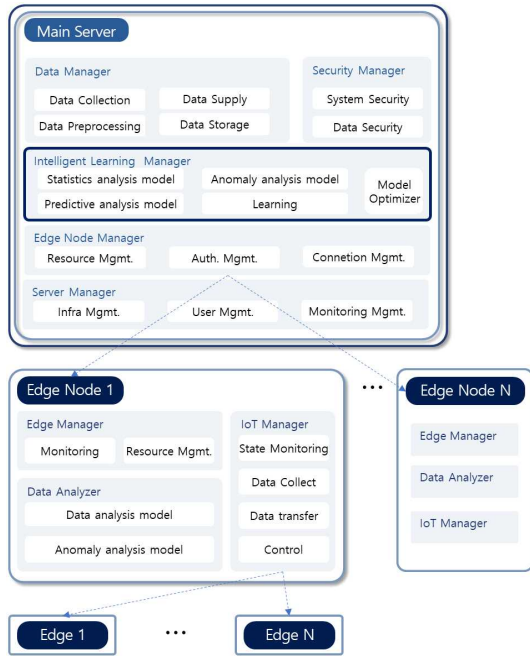


그림 2. 제안하는 DPaaS 구조
Fig. 2. Proposed DPaaS architecture

성되어 있다.

- **Main Server** : 데이터 플랫폼의 핵심 요소로 메인 서버 역할을 담당한다. 인프라, 사용자를 관리하는 Server Manager와 Edge Node들을 관리하는 Edge Node Manager, 데이터를 수집하고 전처리한 후 저장하여 시각화 및 서비스 형태로 제공을 해주는 Data Manager, 실제적인 데이터 분석 요소로 통계 및 예측 분석, 이상 모델 분석, 학습을 하는 Intelligent Learning Manager를 비롯하여 보안을 담당하는 Security Manager로 구성되어 있다.
 - **Server Manager** : 컴퓨팅 자원 관리를 하는 Infra Mgmt.(Management), 사용자의 인증 및 권한 관리를 하는 User Mgmt, Server System을 모니터링 하는 Monitoring Mgmt.로 구성되어 있다.
 - **Edge Node Manager** : Edge Node들의 자원을 관리하는 Resource Mgmt., 인증을 담당하는 Auth.(Authentication) Mgmt., 연결 상태를 유지하고 관리하는 Connection Mgmt.로 구성되어 있다.
 - **Data Manager** : 데이터를 수집하는 Data Collection, 데이터를 전처리하는 Data Preprocessing, 데이터를 저장하는 Data Storage, 데이터를 시각화하고 서비스화해 제공해주는

Data Supply로 구성되어 있다.

- **Intelligent Learning Manager** : 데이터 플랫폼의 핵심으로, 데이터 분석을 위해 통계 분석 모델, 예측 분석 모델, 이상 분석 모델, 학습 및 모델 최적화로 구성되어 있다. 사용자는 원하는 데이터 분석 결과에 맞추어 분석 방법을 정할 수 있고, 분석을 수행할 수 있다. 다음 3.3절에서 동작 과정을 보인다.
- **Security Manager** : Main Server가 실행 중일 때 항상 동작하고 있는 부분으로 시스템 보안을 담당하는 System Security, 데이터 보안을 담당하는 Data Security로 구성되어 있다.
- **Edge Node** : 다양한 지역에 분포되어 있는 작은 서버의 개념으로, 해당 Node의 자원 관리와 성능을 모니터링하는 Edge Manager와 데이터 분석을 수행하는 Data Analyzer, 데이터를 수집하고 전송 및 제어하는 IoT Manager로 구성된다.
 - **Edge Manager** : Edge Node의 상태를 관리하기 위해 시스템을 모니터링하는 Monitoring과 자원을 관리하는 Resource Mgmt.로 구성된다.
 - **Data Analyzer** : IoT Manager로부터 수집된 Data들을 분석할 수 있도록 저용량 데이터를 분석할 수 있는 데이터 분석 모델과 실시간 데이터 분석을 수행하는 이상 분석 모델로 구성된다.
 - **IoT Manager** : 각각의 Edge를 관리하기 위해 모니터링하는 State Monitoring, 제어하는 Control로 구성되어 있으며, Edge들의 데이터를 수집하는 Data Collect, 데이터 분석을 위해 Data Analyzer로 전송하는 Data transfer로 구성된다.
- **Edge** : 다양한 Edge(IoT 기기)들로 구성될 수 있다.

이러한 DPaaS의 구조는 2장 관련 연구에서 살펴본 옛지 클라우드와 구조적으로 비슷하나, 중앙 집중식 클라우드와 옛지 클라우드의 장점을 혼합한 형태이다.

Edge Node에서는 이상 분석 모델과 같이 실시간이 요구되는 데이터 분석을 다룬다. Edge Node에서 간단한 데이터 분석이 가능하지만, 학습된 데이터들을 함께 활용하고 재사용 등을 용이하게 위해 통계 분석, 예측 분석 등은 메인 서버에서 담당하며, 데이터의 저장도 메인 서버에서 이루어진다. Edge Node에서는 데이터를 수집하고, 전송하는 역할을 수행하며 별도로 저장하고 있지 않는다.

3.1.2 DPaaS 구조에서의 데이터 분석 흐름도

다음 그림 3은 데이터 분석을 위해, DPaaS에서 데이터를 수집하고, 전처리 과정을 거쳐 데이터 분석을 수행하고 성능 평가를 하는 과정 및 데이터 학습 모델을 선정하여 서비스 제공까지의 DPaaS의 전체 데이터 분석 흐름도를 나타낸다.

Data Collection을 통해 다양한 데이터를 수집한다. 다음은 Data Preprocessing 과정이다.

- Data Transformation : 데이터 형식을 맞춰준다.
- Missing Data 판별 : 데이터의 결측치 유무를 판별하여 보간법을 이용하여 처리한다.
- Outlier Data 판별 : 데이터의 이상치 유무를 판별하여 제거한다.

다음 과정인 Intelligent Learning 부분은 다음 절에서 설명한다.

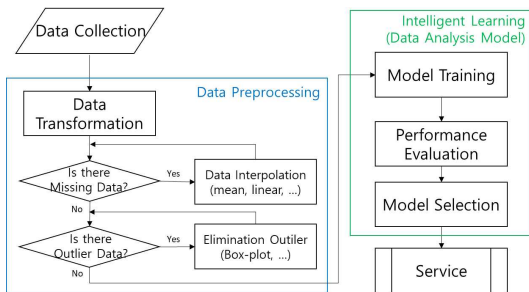


그림 3. 데이터 분석 흐름도
Fig. 3. Data analysis flowchart

3.1.3 DPaaS 구조의 특징

제안하는 DPaaS는 사용자 요청부터 서비스 제공까지 모든 과정을 포함하는 것으로, DPaaS는 PaaS(Platform as a Service)^[24]가 제공하던 플랫폼 서비스에서 데이터 분석에 필요한 요소들을 포함한다.

수집되는 IoT 데이터들을 Edge Node에서 분석 가능한 부분은 분석하여, 실시간 처리가 가능하다는 점과 동시에 메인 서버의 부하를 줄일 수 있고 높은 수준의 보안을 유지할 수 있다는 분산 클라우드 컴퓨팅의 장점을 최대한 활용하고자 하였다. 또한 Main Server 혹은 Edge Node에 서비스 장애 발생시, 다른 Edge Node에는 영향을 주지 않아 끊임 없는 효율적인 서비스 이용을 할 수 있도록 하였다.

기존의 데이터 플랫폼은 데이터를 수집 및 전처리하여 저장하고, 데이터 분석 결과를 사용자에게 시각화하여 보여주는 것에 그쳤다면, DPaaS의 특징은 다

음과 같다.

첫째, DPaaS는 실제 데이터 분석을 할 수 있도록 다양한 데이터 분석 알고리즘을 수행해보고 최적화하여 제공하는데 특징이 있다. DPaaS를 통해 적용되는 최적화된 알고리즘은 오류 지표, 사용자가 원하는 데이터 분석 방향, 데이터 학습 시간 등에 가중치를 주어 선정하는 것으로 데이터 분석 결과에 신뢰도를 높일 수 있다. 또한 각각의 알고리즘을 사용자는 따로 수행해볼 필요 없이 최적의 알고리즘으로 선정하여 데이터 분석을 할 수 있도록 하여 데이터 플랫폼의 사용 편의성을 제공하여 접근성을 높이고자 한다.

둘째, 시각화하여 제공하는 것에서 그치지 않고 서비스 형태로 만들어 제공한다. 데이터 분석 결과를 눈으로 한번 보는 것이 아닌 지속적으로 사용할 수 있도록 서비스화하여 제공함으로써 DPaaS가 데이터 플랫폼으로서 꾸준히 활용되는 것을 기대할 수 있다.

그림 4는 DPaaS가 가지는 중요한 요소인 실제 데이터 분석을 수행하는 Intelligent Learning Manager에서 실제 데이터 분석이 이루어지고, 성능 평가를 하고 모델을 선정하는 모델 최적화 동작 과정을 보인다. Intelligent Learning Manager에서는 이미 전처리된 데이터(Prepared Data)를 가지고 데이터 분석을 수행한다. 이때 데이터 분석 종류(통계 분석, 예측 분석, 이상 감지 등)는 사용자에게 미리 입력받아 선택하여 분석을 수행한다. 각각의 데이터 분석 방법별로 학습 알고리즘(본 논문에서의 사용 예: LSTM, convLSTM, GRU 등)들을 사용하여 사용자의 목적에 따라 최적의 알고리즘을 선정하고, 데이터 분석을 수행한다. 선정된 알고리즘의 데이터 분석 결과로 시각화 및 서비스 화시켜 사용자에게 제공한다.

DPaaS에서 수집 및 전처리하는 데이터들은 데이터 분석을 한 뒤 삭제하는 것이 아니라 Data Storage에 저장한다. 저장된 데이터는 Intelligent Learning Manager의 Learning(데이터 학습)을 통해 재구성하여 학습에 활용한다. 지속적인 데이터 학습을 통해 데이터 분석시 정확도를 높일 수 있으며, 생성되는 학습

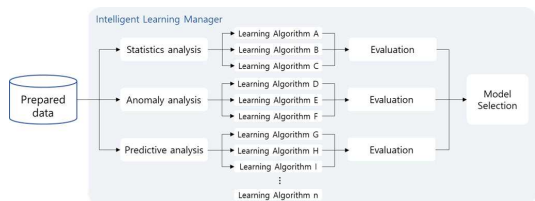


그림 4. 모델 최적화 동작 과정
Fig. 4. Procedure for model optimization

데이터는 또 다른 서비스로 제공될 수 있어 높은 확장성을 가진다.

3.2 데이터 플랫폼 동작과정: 딸기 생산량 예측 모델 시나리오

3.2절에서는 DPaaS를 위한 분산 클라우드 기반 데이터 플랫폼이 다양한 환경에서 폭넓게 사용 가능하다는 점과 실제 사용자가 접근해서 사용하기 편하게 제공됨을 보이기 위해 딸기 생산량 예측 모델 시나리오에 맞추어 DPaaS 동작 과정을 기술한다.

다음 그림 5는 딸기 생산량 예측 모델 시나리오에 맞춰 구성한 데이터 플랫폼의 동작 과정을 나타내며, Main Server는 제안하는 분산 클라우드 기반의 데이터 플랫폼인 DPaaS를 의미하며, Edge Node는 데이터를 수집하여 분석 및 전송 등을 수행할 수 있는 스마트팜에 있는 게이트웨이 혹은 컴퓨터 등이 될 수 있다. Edge는 스마트팜에서 사용하고 있는 각종 센서 Node들 혹은 IoT 기기 및 사용자의 기기가 될 수 있다. User는 스마트팜을 운영하는 농부, 혹은 딸기 데이터를 통해 연구를 하고자 하는 연구자 등이 될 수 있다.

세부 동작 과정은 다음과 같다.

- ① 사용자가 데이터 플랫폼에 접근하여 인증을 받고, 원하는 데이터 분석 종류(본 시나리오에서는

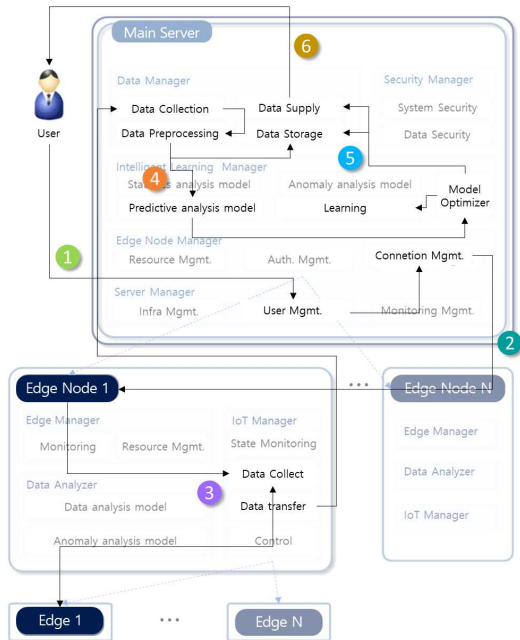


그림 5. 시나리오에 따른 DPaaS 동작 과정
Fig. 5. DPaaS procedure by scenario

Predictive analysis model, 예측 분석)를 선택하고, 사용자가 2주 동안의 해당 농장의 환경 정보 및 생육 정보를 입력함

- ② Edge Node와 연결함
- ③ 필요한 데이터를 Edge들을 통해 수집하고, 수집한 데이터를 Main Server로 전송함 (딸기 스마트팜에 연동되어 있는 환경 정보 센서를 이용하고, 생육 정보 데이터는 사용자에게 입력받음)
- ④ Main Server의 Data Manager는 데이터를 수집한 후, 전처리하여 저장함. 전처리된 데이터를 Intelligent Learning Manager에서 예측 분석 모델을 생성하고 데이터 분석을 수행함. 분석 결과에 따라 성능 평가를 진행하고 최적의 모델을 선정함
- ⑤ 데이터 예측 분석 결과를 Data Supply를 통해 시각화 및 서비스로 구성하고, Data Storage에 저장하며, 만들어진 모델을 데이터 플랫폼에서 학습할 수 있도록 함
- ⑥ 사용자에게 데이터 분석 결과를 서비스 형태로 제공(딸기 작물의 수확량을 예측하여 값을 알려줌)하며, 서비스를 지속적으로 사용 가능하도록 제공함

본 논문에서 구상한 딸기 생산량 예측 모델 시나리오에서는 Main Server에서 데이터 분석하는 것이 중시이지만, Edge Node와 연결 시 데이터를 수집하고 전송 시키는 동안 연결이 잘 유지될 수 있도록 해야한다. 또한 Edge를 통해 얻은 데이터 중, 이상 분석 등 실시간 분석이 필요한 경우 Edge Node에서 데이터 분석을 수행하고 실시간으로 결과를 알려줄 수 있어야 한다.

딸기 생산량 예측 모델 시나리오에서는 1~4개의 농가를 분석한 기존 논문들^{[4],[5],[20]-[22]}과는 달리 다수의 사용자가 활용했을 때 신뢰도를 높이기 위해, 최대한 많은 농가의 데이터를 이용해 학습하고자 하였다.

다음 4장에서 관련된 실험 내용과 평가를 보인다. 이후 동작 과정에 따라 학습 알고리즘을 선정하여 딸기 생산량 예측 서비스를 제공하고자 한다.

다음 그림 6은 1, 2주 전의 데이터(환경 변수, 생육 변수)를 입력하여 분석을 수행하고, 예측을 원하는 주(이

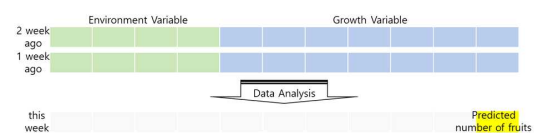


그림 6. 학습 모델 구성
Fig. 6. Configuration for training model

번주)의 착과수를 예측하는 학습 모델 구성을 보인다.

각각의 모델은 입력 데이터를 2주, 4주, 6주 등 다양하게 실험해본 결과 2주 동안의 데이터를 분석하여 그 다음 1주의 착과수를 예측하는 모델이 사용자가 입력하기에 편리하고 예측 정확도도 높아서, 2주간의 데이터를 사용해서 다음주의 착과수를 예측하는 학습 모델을 구성하였다.

IV. 실험 및 평가

4장에서는 시나리오로 제시된 딸기 생산량 예측 분석 모델에 대한 실험 및 평가를 진행한다. 그림 6에 따라 데이터 분석을 위한 학습 모델을 구성하고, 실험을 진행하였다.

1절에서는 실험을 진행한 환경 및 데이터 수집 출처, 학습 데이터 구성에 대해 설명하고, 2절에서 딸기 생산량 예측 모델로 실제 데이터를 전처리하고, 분석 모델을 만들어 성능 평가 수행 후 최적의 알고리즘을 선정하는 과정을 보인다.

4.1 실험 환경

본 논문에서 활용된 딸기 관련 스마트팜 데이터는 공공 데이터 포털에서 제공하는 “농촌진흥청_스마트팜 우수농가 공개용 데이터”^[25]에서 딸기 부분 데이터와 스마트팜 코리아에서 개방한 “시설원에 작기 데이터셋”^[26]에서 딸기 데이터를 사용하였다.

학습 데이터는 주간별로 환경 변수와 생육 변수 모두가 포함되어있는 데이터 형태로, 주간을 데이터 분석을 위한 인덱스(Index) 값으로 지정하였고, 환경 변수 4개(외부 일사량, 내부온도, 내부습도, 내부 CO2)와 생육 변수 7개(조장, 엽장, 엽폭, 엽병장, 엽수, 관부직경, 착과수)를 사용하였다.

표 1. 실험 환경
Table 1. Evaluation environment

Environment		Specifications
Physical Environment	CPU	AMD Ryzen 5 5600X 6 Core Processor 3.70 GHz
	RAM	16GB
	OS	Windows 10 (64 bit)
IDE (Integrated Development Environment)	Interface	Jupyter Notebook
	Notebook Server	Version 6.0.2
	Python	Version 3.7.10
	TensorFlow	Version 2.1.0

다음 표 1은 딸기 생산량 예측을 위해 분석 학습에 활용한 실험 환경이다.

4.2 딸기 생산량 예측 모델

4.2.1 데이터 전처리

데이터 분석에 앞서 다음과 같이 데이터를 전처리하는 과정을 거쳤다. 먼저 주간 단위로 데이터를 정리하여 환경 변수와 생육 변수를 컬럼을 합한 데이터셋을 구성하였다. 이때 환경 변수를 주간 단위로 변환하는 과정은 Python에서 제공하는 pandas. DataFrame에서 resample 함수^[27]를 이용하여 주간 단위로 데이터를 정리하였다.

수집한 학습 데이터는 각 농장들이 측정할 날짜에 따라 차이가 있었지만, 최대한 겹치는 기간인 2021년 1/10~6/13까지 측정하여 수집된 데이터를 사용하였고, 너무 많은 값을 결측치로 처리하게 되면 데이터 분석에 좋지 않은 영향을 미칠 수 있어 환경이나 생육 변수의 결측치가 70% 이상인 농가는 제외시켰다.

데이터 분석 흐름도(그림 3)에 따라 결측치와 이상치 값을 있고 없음을 판단하여 처리 및 제거해주었다.

환경 변수는 센서 데이터를 활용해서인지 대부분 데이터가 정확하게 있었지만, 결측치가 다소 있었다. 외부 일사량은 해당 농장이 위치한 지역과 날짜를 고려하여 평균치 값으로, 내부 온도와 습도, CO2의 경우에는 해당 농장과 날짜를 고려하여 평균치 값을 사용하여 결측치를 처리하였다. 생육 변수의 결측치는 중앙값, 평균값을 구하는 방식과 선형 보간법을 사용하여 결측치를 처리하였다.

결측치 처리를 거친 후, 이상치(outlier) 값 판별을 위해 Box-plot을 이용해 데이터 분포를 파악하였고, 이상치 값을 가진 농장들을 제거하였다. Box-plot은 통계학에서 활용되는 데이터 분포를 개괄적으로 파악할 수 있는 그래프로 IQR(InterQuartile Range)*1.5 범위 밖에 자리하는 값들을 이상치로 판단한다. 이러한 이상치 값들은 데이터 분석 결과가 이상치 값에 치우쳐질 수 있어 제거해주었다.

4.2.2 데이터 분석 모델

딸기 생산량 예측 모델을 만드는데 많이 고려하였던 것은, 다수가 활용할 수 있는 데이터 플랫폼 서비스를 위해 최대한 많은 스마트팜의 데이터를 활용하여 결과를 얻는 것이었다. 데이터 전처리 때부터 최대한 많은 농가의 데이터를 확보하기 위해 노력하였으나 데이터 결측치, 이상치 값이 존재하는 등의 문제가

발견되어, 수집한 73개의 농가 데이터 중 최종적으로 15개의 농가를 활용하여 데이터 분석을 수행하였다.

본 논문에서는 딸기 생산량 예측을 위해 3가지 알고리즘을 사용하여 비교·분석하였다.

첫 번째로 실험해본 것은 다변량 시계열 예측에 많이 활용되고 있는 LSTM 알고리즘이다. 두 번째는 Convolution Layer가 추가된 형태인 convLSTM으로 수행하였다. 그러나 스마트팜 딸기 데이터의 복잡도가 높지 않아서 Convolution Layer가 추가되면 오히려 성능이 떨어지는 양상을 보여 LSTM보다 단순한 구조로 이루어져 연산량을 줄일 수 있는 GRU 알고리즘으로 세 번째 실험을 진행하였다.

이 때 데이터 분석에 활용한 생육 변수는 농작물을 관리하는 농부의 목표량(크기가 큰 딸기, 개수가 많은 딸기, 당도가 높은 딸기 수확 등) 혹은 방식(잎이나 꽃을 따는 빈도 등)에 따라서도 달라질 수 있으나, 수집된 데이터에서의 착과수를 기준으로 하여 데이터 분석을 수행하였다.

4.2.3 성능 평가

성능 평가를 위해 오류 지표는 RMSE(Root Mean Squared Error, 평균 제곱근 오차) 값을 사용하였다. RMSE는 직관적인 지표였던 MSE(Mean Squared Error)에 제곱근 값을 구하여, 오류의 제곱을 구해 실제 오류 평균 값보다는 크게 나왔던 MSE보다 왜곡이 적은 것이 특징으로 데이터 분석 결과에 대한 제시, 평가 통계 등에 더 적합하여[28] 활용하였다. 공식은 식 (1)에 나타내었다. n 은 데이터 수, y 는 실제 착과수 값, \hat{y} 는 예측한 값을 나타낸다.

또한 학습 시간(T, Training Time)을 측정하여 성능 평가 지표로 활용하였으며 공식은 식 (2)에 나타내었다. i 는 epoch의 횟수로, 학습 사이클(Training cycle) 수를 의미한다. end_time 은 학습이 끝난 시간, $start_time$ 은 학습을 시작한 시간을 의미한다.

$$RMSE = \sqrt{\frac{\sum (y - \hat{y})^2}{n}} \quad (1)$$

$$T = \sum_{i=1}^{epoch} ((end_time)_i - (start_time)_i) \quad (2)$$

학습 모델에 따라 높은 정확도의 값을 얻기 위한, 최적의 하이퍼파라미터(Hyperparameter) 값이 달라질 수 있어 실험을 통해 하이퍼파라미터 값을 설정하는

표 2. 하이퍼파라미터 튜닝
Table 2. Hyperparameter tuning

Hyper Parameter	LSTM	conv-LSTM	GRU
Activation Function	tanh	softplus	tanh
Optimizer Function	Adam	Stochastic Gradient Descent	Adam
Dropout	0.2	0.2	0.2
Loss Function	Huber	MeanSquare-d Error	Huber

작업이 필요하다. 다음 표 2는 실험에 사용한 학습 모델 알고리즘별 하이퍼파라미터 튜닝을 나타낸다.

활성화(Activation) 함수는 입력값에 대한 출력값을 비선형 방식으로 도출하기 위해 설정하였고, 최적화(Optimizer) 함수는 학습 모델에 맞게 오차를 줄일 수 있는 함수로 설정해주었다. 드랍아웃(Dropout)은 과적합 방지를 위해 설정하였고, 성능 평가를 위한 오차 계산을 위해 손실(Loss) 함수를 설정하였다.

표 3은 실험 결과로, 생산량 예측 모델의 성능 평가 분석 결과이다. 학습은 Epoch 값을 100, 300, 500번으로 설정하여 수행하였으며, RMSE의 평균 값과 총 학습 시간을 구하였다.

평균 정확도가 가장 높았던 모델은 RMSE 값이 가장 작았던 LSTM 알고리즘을 Epochs 100으로 했을 때이고, 학습 시간이 가장 적게 걸린 모델은 GRU 알고리즘을 Epochs 100으로 했을 때로 나타났다.

실험 결과 Epochs을 100으로 했을 때 RMSE 값이 낮고, Training Time이 적게 걸려, 각 알고리즘별로 Epochs을 100으로 했을 때의 15개 농장의 RMSE 값

표 3. 생산량 예측 모델 실험 결과
Table 3. Result of production prediction model

Epochs	Model	RMSE	Training Time (Second)
100	LSTM	0.6828	0.6708
	convLSTM	0.9650	1.0619
	GRU	0.7074	0.6022
300	LSTM	0.7646	1.8575
	convLSTM	1.0259	3.1475
	GRU	0.7721	1.6879
500	LSTM	0.6986	3.1395
	convLSTM	0.9761	5.2718
	GRU	0.7467	2.8090

을 비교해보았고, 다음 그림 7에 그래프로 나타내었다.

그림 7의 그래프 상에서 빨간색은 LSTM, 초록색은 convLSTM, 파랑색은 GRU 알고리즘의 RMSE 값을 나타낸다. RMSE 최대값은 K 농장 ConvLSTM 알고리즘의 1.9183, 최소값은 D 농장 LSTM 알고리즘의 0.0910이다. 농장별 RMSE 값을 비교했을 때, 세 가지 알고리즘 중 LSTM 알고리즘의 RMSE 값이 낮은 농장은 6개, convLSTM이 4개, GRU 알고리즘이 5개 농장으로 나타났다.

최적의 알고리즘을 선정하기 위해 가중치를 주어 구하는데, 딸기 생산량 예측 모델에서는 정확도에 가중치를 주어 RMSE 값이 낮은 알고리즘을 선정하고자 하였다. 최종적으로 LSTM이 적합하다고 판단하여 이를 구성하는 데이터 플랫폼의 딸기 생산량 예측 모델에 사용하는 알고리즘으로 적용하였다.

다음 그림 8은 A, B, C 세 농가의 알고리즘별 딸기 생산량 예측 그래프를 나타낸다.

성능 평가에서는 RMSE 값에 가중치를 두어 알고리즘을 선정하였고, 예측 결과 값을 그래프로 나타내었다. 검정색은 실제 농장의 데이터, 빨간색은 LSTM, 초록색은 convLSTM, 주황색은 GRU 알고리즘의 예측 값을 나타낸다.

데이터가 시작하고 첫 2주동안의 데이터를 분석하여 그 다음 1주를 예측하는 형식으로, 첫 2주간은 실제 농장의 데이터 값만 표기되었다. 이후 예측되는 마지막 1주 값은 예측 분석을 수행한 결과 값을 나타낸다.

최종적으로 DPaaS는 최적으로 선정된 LSTM 알고리즘으로 딸기 생산량 예측 분석 결과 서비스를 제공한다. 서비스를 제공하며 다양한 농가들의 데이터를 학습할수록 예측 결과 값은 더 정확해지고, 신뢰할 수 있을 것이다.

딸기 수확량 예측을 할 수 있다면, 딸기 생산자들은 출하 계획을 세울 수 있고 수확량 증가를 위한 환경 변수들의 조절 등을 할 수 있어 스마트팜 운영의 비용 절감 및 수익 증대의 효과가 예상된다.

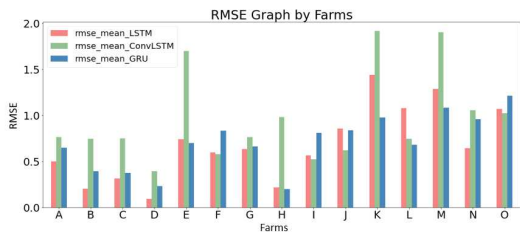


그림 7. 농장별 RMSE 그래프
Fig. 7. RMSE graph by farms

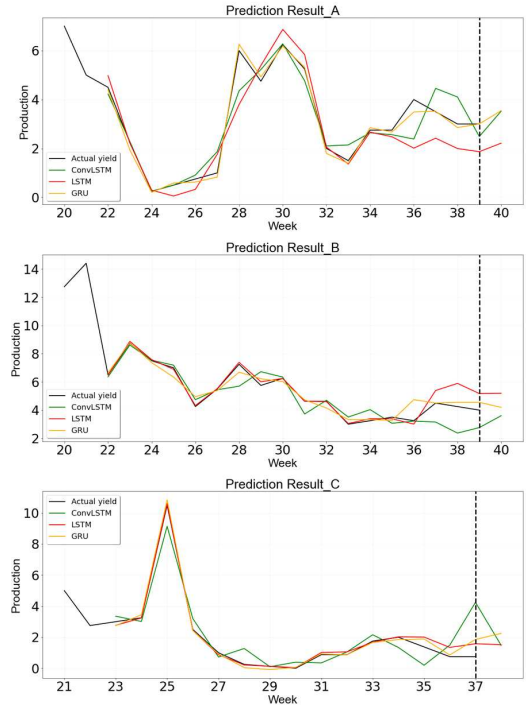


그림 8. 알고리즘별 예측 분석 그래프
Fig. 8. Prediction analysis graph by algorithm

V. 결론

다양한 분야의 대용량 데이터가 쏟아지면서 데이터에 대한 관심과 중요성은 커지고 있다. 이에 따라 데이터 플랫폼 관련 연구가 진행 중이지만, 데이터 플랫폼 설계 뿐만 아니라 이미 구축된 데이터 플랫폼을 제대로 활용하기 위한 활성화 방안이 필요하다.

본 논문은 데이터 플랫폼 설계 및 데이터 플랫폼 활성화를 위해 서론에서 살펴본 바와 같이, 다음과 같은 목표를 가지고 수행하였다.

첫 번째, DPaaS(Data Platform as a Service)를 위한 분산 클라우드 기반의 데이터 플랫폼 구조 설계 : 기존의 중앙 집중식 모델에서 분산 클라우드 기반의 데이터 플랫폼을 사용하면, 서버 과부하를 줄일 수 있고 실시간 처리를 빠르게 할 수 있다는 장점과 장애 발생시 피해를 최소화할 수 있다. 또한, 중간에 Edge Node, Edge로 사용되는 장치들도 IoT 기술의 발전으로 데이터 분석, 전송 등의 컴퓨팅 수행 능력이 향상되어 이를 활용하는 것이 효율적이므로 본 논문에서 제안하는 분산 클라우드 기반의 데이터 플랫폼 구조가 앞으로 사용되는데 적합하다고 할 수 있으며, 이를 기반으로 DPaaS 구조를 설계하였다.

두 번째, DPaaS를 이용한 데이터 분석 서비스 제공에 따른 데이터 플랫폼의 활성화 : 다양한 데이터 플랫폼의 연구와 기술 개발의 목적은 데이터 플랫폼의 활용일 것이다. 데이터 플랫폼의 활성화를 위해서는 데이터 플랫폼 구축에서 한 단계 더 나아가 많은 사람들이 쉽고 편하게 사용 가능해야 한다. 이를 위해 DPaaS는 사용자들이 복잡한 데이터 분석 과정을 알기 보다는 요구에 맞게 최적화된 알고리즘을 서비스 형태로 제공 받으므로써 높은 접근성을 가질 것이다. 또한, 서비스 이용에 대한 지속적인 활용을 기대할 수 있다.

세 번째, 스마트팜 서비스 제공을 위한 시나리오를 통해 딸기 작물에 관한 생산량 예측 및 최적화 모델 선정에 관한 연구 : 본 논문에서는 시나리오와 실험을 통해, 실제 스마트팜 시나리오를 구성하여 제공할 수 있도록 딸기 생산량 예측에 대한 데이터 분석을 통해 데이터 플랫폼의 활용 방안을 보였다. 실험은 전처리 과정을 거쳐 15개 농장의 데이터로 진행하였다. DPaaS에서 예측 분석을 하기 위해 세 가지 학습 알고리즘을 통해 실험을 진행하였고, 분석 결과 RMSE가 0.6828으로 낮았던 LSTM 알고리즘을 선정하였다. 결과적으로 사용자에게 DPaaS를 통해 오류 지표 값이 낮은 최적화된 알고리즘을 적용하여 정확도 높은 딸기 생산량 예측 분석 결과를 지속적으로 이용할 수 있도록 서비스화하여 제공하였다.

이를 통해 데이터 플랫폼 관련 연구 및 스마트팜 관련 연구에도 활용될 수 있으며 데이터 플랫폼 활성화에 기여할 것으로 기대된다.

향후 지속적인 학습을 통해 딸기 생산량 예측 모델의 오차를 줄여 신뢰성을 높일 수 있도록 할 것이며, 스마트팜에 DPaaS를 적용한 실증 연구를 진행할 예정이다. 또한 보다 편리한 형태로 사용할 수 있도록 데이터 플랫폼 접근 및 로그인 등에 필요한 웹페이지, 데이터 입력과 분석 결과를 받을 API 등을 개선하고자 한다.

References

[1] "Data is the New Oil," Retrieved Oct. 5, 2021, from <https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294>

[2] C. H. Yun and J. I. Kim, "A systematic review on ai platform for public sector," *J. KICS*, vol. 45, no. 11, pp. 1994-2003, Nov.

2020.

- [3] A. Kim and J. Lee, "Data analyst platform using open source based deep learning engine," *2018 Int. Conf. PlatCon*, pp. 1-6, Jeju Island, Korea, Jan. 2018.
- [4] H. S. Noh and Y. S. Lee, "Determinants of growth variables on smart farm tomato production," *The Soc. Convergence Knowledge Trans.*, vol. 8, no. 3, pp. 17-25, Sep. 2020.
- [5] S. E. Hong, T. J. Park, J. I. Bang, and H. J. Kim, "A study on the prediction model for tomato production and growth using ConvLSTM," *J. KIIT*, vol. 18, no. 1, pp. 1-10, Jan. 2020.
- [6] S. Y. Lee, Y. H. Kim, and A. Y. Son, "Multi-channel platform architecture design for efficient data collection in distributed cloud computing environment," *Korea Comput. Congr. 2021*, pp. 1310-1312, Jeju Island, Korea, Jun. 2021.
- [7] D. H. Kim, B. S. Park, and J. S. Yang, "Architectural design for SPLE-based inertial navigation system software platform," *J. IEIE*, vol. 58, no. 3, pp. 79-86, Mar. 2021.
- [8] S. W. Kim, D. W. Kim, and J. H. Choi, "A cloud openlab platform for ai traffic data sharing environment," *Korean Soc. Civil Eng. 2020 Convention*, pp. 2013-2018, Jeju Island, Korea, Oct. 2020.
- [9] H. G. Han, C. Y. Lee, and Y. G. Park, "A plan to build an integrated data platform for real-time data collection and analysis of extreme ocean spaces," *J. Digital Contents Soc.*, vol. 22, no. 6, pp. 989-998, Jun. 2021.
- [10] H. J. Kim, J. H. Ra, W. R. Jeon, and P. K. Kim, "A study on construction of platform using spectrum big data," *Smart Media J.*, vol. 9, no. 2, pp. 99-109, Jun. 2020.
- [11] K. J. Lee and E. Y. Kim, "The role and effect of artificial intelligence (ai) on the platform service innovation: The case study of Kakao in Korea," *J. Knowledge Manag. Res.*, vol. 21, no. 1, pp. 175-195, 2020.
- [12] M. K. Nam and E. T. Jung, "A study on mobile terminal application of manufacturing

- data analysis platform-focused on the case of Changwon smart mountain cluster-,” *J. Kor. Soc. Design Culture*, vol. 26, no. 1, pp. 145-159, Mar. 2020.
- [13] J. W. Lee and S. J. Oh, “Analysis of success cases of insurtech and digital insurance platform based on artificial intelligence technologies: Focused on ping an insurance group ltd. in china,” *J. Intell. and Inf. Syst.*, vol. 26, no. 3, pp. 71-90, Sep. 2020.
- [14] H. T. Oh, J. M. Yu, C. J. Jung, and J. K. Choi, “Data analysis and consideration of radar-based contactless biometrics monitoring testbed for single elderly households,” *J. KICS*, vol. 46, no. 6, pp. 1056-1064, Jun. 2021.
- [15] “*Digital NewDeal - BigData Platform*,” Retrieved Oct. 5, 2021, from <https://www.bigdata-map.kr>
- [16] “*Modern Data Platform*,” Retrieved Oct. 5, 2021, from <https://medium.com/@vpowell4/the-modern-data-platform-8f82318ec70>
- [17] A. Y. Son and E. N. Huh, “Migration architecture based on container virtualization for efficiency improvement in distributed cloud computing,” *Korea Comput. Congr. 2015*, pp. 1250-1252, Jeju Island, Korea, Jun. 2021.
- [18] “*Top Strategic Technology Trends for 2021*,” Retrieved Oct. 5, 2021, from <https://www.gartner.com/smarterwithgartner/gartner-top-strategic-technology-trends-for-2021>
- [19] I. H. Bae, “Design and evaluation of an edge-fog cloud-based hierarchical data delivery scheme for iot applications,” *J. Internet Comput. and Serv.*, vol. 19, no. 1, pp. 37-47, Feb. 2018.
- [20] L. J. Kyung and S. B. Moon, “Intelligent smart farm a study on productivity : Focused on tomato farm households,” *Asia-Pacific J. Busin. Venturing and Entrepreneurship*, vol. 14, no. 3, pp. 185-199, Jun. 2019.
- [21] J. U. Park, H. H. Ahn, and B. K. Lee, “The agriculture decisionmaking system(ADS) based on DeepLearning for improving crop productivity,” *J. Kor. Inst. Inf., Electron. and Commun. Technol.*, vol. 11, no. 5, pp. 521-530, Oct. 2018.
- [22] S. K. Kim, J. H. Lee, H. J. Lee, S. G. Lee, B. H. Mun, S. W. An, and H. S. Lee, “Development of prediction growth and yield models by growing degree days in hot pepper,” *The Kor. Soc. Bio-Environ. Contr.*, vol. 27, no. 4, pp. 424-430, Nov. 2018.
- [23] M. S. Aslanpour, A. N. Toosi, C. Cicconetti, B. Javadi, P. Sbarski, D. Taibi, M. Assuncao, S. S. Gill, R. Gaire, and S. Dustdar, “Serverless edge computing: Vision and challenges,” *Australasian Comput. Sci. Week Multiconf.*, no. 10, pp. 1-10, Feb. 2021.
- [24] D. Beimborn, T. Miletzki, and S. Wenzel, “Platform as a service (PaaS),” *Busin. & Inf. Syst. Eng.*, no. 3, pp. 381-384, Oct. 2011.
- [25] “*Rural Development Administration - public use of smart farm excellent farms*,” Retrieved Oct. 5, 2021, from <https://www.data.go.kr/data/15042594/openapi.do>
- [26] “*SmartFarmKorea - Facility Gardening Dataset*,” Retrieved Oct. 5, 2021, from <https://www.smartfarmkorea.net/front/open/dsOpenFclgdnList.do?menuId=M01060301>
- [27] “*pandas API reference - DataFrame - resample*,” Retrieved Oct. 5, 2021, from <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.resample.html>
- [28] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)?,” *Arguments against avoiding RMSE in the literature, Geosci. Model Dev.*, vol. 7, pp. 1247-1250, Jun. 2014.

이 세 연 (Se-Yun Lee)



2013년 2월 : 경희대학교 컴퓨
터공학과 졸업
2015년 2월 : 경희대학교 컴퓨
터공학과 석사
2020년 3월~현재 : 차세대융합
기술연구원 연구원

<관심분야> 클라우드 컴퓨팅, 빅데이터, 가상화
[ORCID:0000-0003-1157-167X]

유 정 민 (Jeong-Min Yoo)



2018년 3월~현재 : 아주대학교
산업공학과 학사과정
<관심분야> 빅데이터, 기계학
습, 데이터마이닝, 컴퓨터비
전

김 준 경 (Junkyeong Kim)



2010년 2월 : 성균관대학교 고
분자시스템공학과 졸업
2012년 2월 : 성균관대학교 u-City
공학과 석사
2018년 8월 : 성균관대학교 건
설환경시스템공학과 박사
2019년 5월~현재 : 차세대융합기
술연구원 선임연구원

<관심분야> IoT 센서, 건전성 모니터링(SHM)
[ORCID:0000-0002-6503-7950]

손 아 영 (A-Young Son)



2013년 2월 : 경희대학교 컴퓨
터공학과 졸업
2015년 2월 : 경희대학교 컴퓨
터공학과 석사
2020년 2월 : 경희대학교 컴퓨
터공학과 박사
2020년 4월~2020년 12월 : 한
국건설기술연구원 연구원

2021년1월~현재 : 차세대융합 기술연구원 선임연구원
<관심분야> 클라우드 컴퓨팅, AI, 빅데이터
[ORCID:0000-0002-8291-6033]

김 혜 린 (Hye-Rin Kim)



2018년 3월~현재 : 아주대학교
산업공학과 학사과정
<관심분야> 데이터 엔지니어
링, 기계학습, 데이터마이닝