

# 머신러닝 적용 과일 수확시기 예측시스템 설계 및 구현

(Design and Implementation of Fruit harvest time Predicting System based on Machine Learning)

오정원\*, 김행곤\*\*, 김일태\*\*\*

(Jung Won Oh, Hangkon Kim, Il-Tae Kim)

## 요약

최근에 머신러닝 기술은 의료, 제조, 마케팅, 금융, 방송, 농업 등 사회 전반에 많은 영향을 미치고 있고 미래에도 인류의 생활에 많은 도움을 줄 것으로 예상된다. 본 논문에서는 인류의 생존에 가장 큰 영향을 주는 먹거리 즉, 농업 분야에 머신러닝 기술을 적용하는 방법을 연구한다. 농업 분야에 IoT(Internet of Things) 기술을 접목하는 스마트 팜 (Smart Farm) 분야는 생육환경을 실시간으로 모니터링 하여 농작물의 생육환경을 최적으로 유지 하는 방법을 중점적으로 연구한다. 최근 KT에서 출시된 기가 스마트 팜 솔루션 2.0 에서는 머신러닝 기술을 사용하여 온실내의 온습도를 최적으로 유지하는 기술에 머신러닝을 적용하였다. 기존의 스마트 팜 분야 연구가 생육환경 조절에 중점을 두어 생산성 증대에 집중되어 있지만 본 연구에서는 과일을 최상의 품질 상태에서 수확하여 좋은 가격으로 출하할 수 있도록 수확시기에 머신러닝을 적용하는 방법을 연구한다. 스마트 팜 분야에 머신러닝 기술을 적용하기 위해서는 풍부한 빅 데이터의 확보가 무엇보다 중요하므로 정확한 머신러닝 기술을 적용하기 위해서는 지속적으로 빅 데이터 수집이 가능해야 한다. 본 논문에서 수확시기 예측에 필요한 인자로는 온실 내에서 재배되는 과일의 색상 값과 무게 값, 내부 온습도 값을 색상센서 와 무게센서, 온습도센서를 사용하여 실시간으로 수집하여 확보한다. 본 논문에서 제안하는 FPSML은 유사 과일 재배에 반복적으로 사용할 수 있는 아키텍처를 제공하며 지속적으로 빅 데이터가 축적될수록 보다 정밀한 수확시기를 예측할 수 있다.

■ 중심어 : 머신러닝; 인류의 생존; 수확시기;스마트 팜;빅 데이터

## Abstract

Recently, machine learning technology has had a significant impact on society, particularly in the medical, manufacturing, marketing, finance, broadcasting, and agricultural aspects of human lives. In this paper, we study how to apply machine learning techniques to foods, which have the greatest influence on the human survival. In the field of Smart Farm, which integrates the Internet of Things (IoT) technology into agriculture, we focus on optimizing the crop growth environment by monitoring the growth environment in real time. KT Smart Farm Solution 2.0 has adopted machine learning to optimize temperature and humidity in the greenhouse. Most existing smart farm businesses mainly focus on controlling the growth environment and improving productivity. On the other hand, in this study, we are studying how to apply machine learning with respect to harvest time so that we will be able to harvest fruits of the highest quality and ship them at an excellent cost. In order to apply machine learning techniques to the field of smart farms, it is important to acquire abundant voluminous data. Therefore, to apply accurate machine learning technology, it is necessary to continuously collect large data. Therefore, the color, value, internal temperature, and moisture of greenhouse-grown fruits are collected and secured in real time using color, weight, and temperature/humidity sensors. The proposed FPSML provides an architecture that can be used repeatedly for a similar fruit crop. It allows for a more accurate harvest time as massive data is accumulated continuously.

■ keywords : Machine Learning; Smart Farm; Harvest Time ; Big Data ; Human Survival

## I. 서론

최근 인류의 먹거리 부분을 해결하기 위해서 스마트 팜

\*학생회원, 대구가톨릭대학교 컴퓨터공학과

\*\*교신저자, 대구가톨릭대학교 컴퓨터공학과

\*\*\*학생회원, 한국폴리텍대학대전캠퍼스 정보통신시스템과

접수일자 : 2018년 11월 02일

수정일자 : 2019년 01월 06일

게재확정일 : 2019년 01월 31일

교신저자 : 김행곤 e-mail : hangkon@cu.ac.kr

(Smart Farm) 분야가 활발하게 연구되고 있다. 대부분의 스마트 팜(Smart Farm) 분야는 IoT 기술을 사용하여 농작물 재배에 필요한 온도, 습도, 토양 영양상태 등을 각종 센서로부터 측정하여 분석 후 각종 제어 장치를 구동하여 농작물이 잘 자랄 수 있는 환경을 개선, 관리하는 데 중점을 두어 연구가 이루어지고 있다[1]. 미국 실리콘밸리의 농업 스타트업 기업인 ‘플렌티(Plenty)’는 중국 대도시에 식물공장 300개를 지었다. 수직농장은 날씨와 상관없이 온도 및 습도, 빛, 물 등을 인위적으로 통제할 수 있어 연중생산이 가능하다는 장점을 지닌다. 또한 해충이나 질병의 발생 통제가 용이해 농약을 사용할 필요가 없는 유기농이며, 도심에 위치함으로써 유통 거리가 짧아 소비자의 식탁에 더 신선한 농작물을 더 빨리 제공할 수 있다[2]. KT는 26일 서울 광화문에서 제 3회 KT 퓨처 포럼을 열고 ‘기가 스마트 팜 솔루션 2.0’ 솔루션에 대해 소개했다. 기가 스마트 팜 솔루션은 센서가 달린 사물인터넷(IoT) 기기를 통해 재배시설의 온도·습도·일사·CO2·토양 등을 분석해 자료를 모은다. 또 분석 결과에 따라 제어장치를 구동해 최적의 작물 재배환경의 조성에 기여한다는 설명이다[3]. 기존의 스마트 팜 분야 연구는 농작물의 생육환경을 최적으로 유지하여 농작물의 생산성을 증대시키는 데 중점을 두고 연구되었다. 그러나 농작물의 생산성을 증대시키는 것도 중요하지만 농작물을 최고의 품질일 때 수확하여 좋은 가격으로 판매하는 것 또한 농부들의 수익 증대를 위해서 무엇보다 중요하다. 인류의 먹거리를 해결하기 위해서는 농업 분야의 수익성을 증대시켜서 농부들의 생활을 안정시키는 것 또한 인류의 먹거리 해결을 위해서 매우 중요하다. 본 논문에서는 다양한 스마트 팜 영역 중 스마트 온실 분야를 연구한다. 스마트 온실에서 재배하는 과일의 수확시기 상관인자(과일의 색상, 과일의 무게, 온실의 온습도)를 색상센서 와 무게센서, 온습도센서를 이용하여 실시간으로 아두이노로 수집한 후 MQTT(MQ Telemetry Transport)와 Kafka를 사용하여 실시간으로 수집 한다. 실시간으로 수집된 데이터들은 머신러닝이 적용되어 수확시기를 정확히 예측한다. 본 논문의 모델에서는 성능을 고려하여 Machine Learning Model Server 와 Application Server는 분리 한다. 실시간으로 수집되는 빅 데이터들은 HDFS(Hadoop Distributed File System)에 저장되어 Machine Learning Model Server의 학습 데이터로 제공한다. 사용자들은 Application Server로부터 PC 와 Smart Phone을 사용하여 서비스를 제공받을 수 있다. 또한 사용자들은 실시간으로 수확시기의 변화 상황을 모니터링 할 수 있다. 본 논문에서는 수확시기 예측에 중점을 두어 연구하며 기존 스마트 팜 관련 시스템에서 중점을 두었던 생육환경 자동 조절 부분은 제외 한다. 본 논문에서 수확시기를 예측하는 알고리즘으로는 과일색상 값 과 과일무게 값, 온실의 온습도 값을 사용하여 과일의 수확시기 예측 값을 예측하므로 다중선형회귀를 사

용한다. 수집된 빅 데이터에 머신러닝을 적용하는 프레임워크로는 Apache Spark를 사용한다. 샘플 과일은 우리 국민들이 가장 많이 사먹는 사과를 선택한다. 최근 한국식품커뮤니케이션포럼에 따르면 농촌진흥청 조영빈 연구관팀이 수도권 732가구가 4년간(2009년 12월~2013년11월) 10대 과일을 산 자료를 분석한 결과 이같이 나타났다[4]. 사과, 배 등은 주로 노지에서 재배되는 과일 이지만 시설에서 재배하면 병충해 예방에도 유리하며 최적의 생육 환경을 유지할 수 있다. 본 논문에서 사과를 샘플 과일로 사용했지만 FPSML의 제안 아키텍처는 타 과수를 재배할 때 반복적으로 적용될 수 있다. 본 논문에서는 FPSML을 프로토타이핑 수준으로 구현하였지만 향후 상용화 시스템으로 개발될 때 스켈리톤(skeleton) 으로 제공될 수 있다.

## II. 관련연구

### 1. 기존 스마트 팜 연구 분야

스마트 팜 (Smart Farm)은 농업 과 ICT 기술을 융합하여 시공간의 제약 없이 농업 환경을 수집, 분석하며 생산, 유통, 소비의 각 분야에서 ICT를 기본으로 하여 시스템화하고 이를 농업에 적용하여 생산성 향상, 비용 절감뿐만 아니라 농업인 삶의 질을 향상시키는 고효율 지향의 농업형태를 의미한다[5].

표 1. 스마트 팜 분야별 적용 모델

분야	내용
스마트온실	PC 또는 모바일을 통해 온실의 온도, 습도, CO2등을 모니터링 하고, 창문 개폐, 영양분공급 등을 원격 자동으로 제어하여 작물의 최적 성장환경을 유지 관리
스마트과수원	PC 또는 모바일을 통해 온도, 습도, 기상상황 등을 모니터링 하고 원격, 자동으로 관수, 병해충 등을 관리
스마트축사	PC 또는 모바일을 통해 온도, 습도 등 축사환경을 모니터링하고, 사료 및 물 공급시기와 양 등을 원격자동으로 제어 유지 관리

표 1은 국내 스마트 팜 적용 모델을 보여준다[6]. 본 논문에서는 스마트 팜 적용 모델 중 스마트온실 모델을 적용한다. 기존 스마트 온실 모델에서는 온실 내의 생육환경을 최적으로 유지하는 것이 중점적으로 연구 되었지만 본 논문에서는 수확시기 예측에 중점을 둔다.

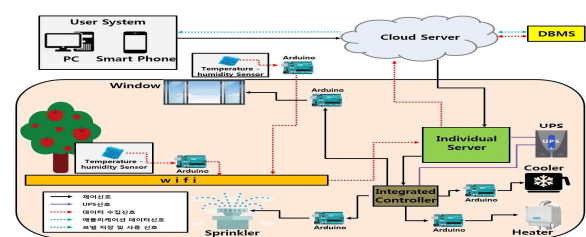


그림 1. 기존의 스마트 온실 아키텍처[7]

그림 1 에서는 기존에 연구되던 스마트 온실의 아키텍처의 예를 보여준다. 상단 그림 1 에서는 실내외에 있는 Temperature-humidity 센서로부터 온습도를 측정 한 후 실시간으로 온실 내에 있는 Individual Server 로 수집된 측정값들이 전달된다. 수집된 값들은 Cloud Server로 전송되어 가공된 후 User System에 현재 생육환경 상태 값들의 모니터링 기능으로 제공된다. 사용자들은 모니터링 화면에서 생육환경을 조절할 수 있으며 사용자들로부터 피드백 된 값들은 Integrated Controller에 전송되어 Heater, Cooler 등의 제어 장치를 구동한다.

### 2. 선형회귀 분석

선형 회귀 분석은 데이터 값 하나가 다른 데이터 값과 선형 관계를 가진다는 가설 하에 결과를 예측하는 분석의 방법이다.

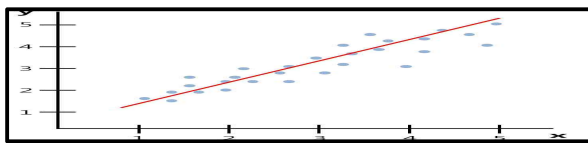


그림 2. 선형관계 그래프

단일 회귀 분석의 경우 x 값이 y 값과 선형관계가 있다면 선형 회귀 분석에서는 학습 데이터를 사용해서 “ $y = ax + b$ ” 라는 수식으로 y값이 정확히 도출될 수 있는 Machine Learning Model을 찾아낸다. “ $y = ax + b$ ” 수식을 반복적으로 적용하여 학습 데이터에 알맞은 최적의 수식을 찾아내는 방법이 선형 회귀 분석이다. 그림 2 에서 보여주는 내용이 x값과 y값이 선형관계에 있음을 나타내는 것이다. 독립 변수가 두 개 이상이면 다중 선형 회귀 분석에 속한다.

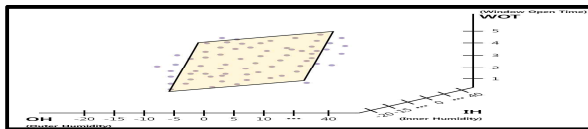


그림 3. 다중 선형관계 그래프

그림 3은 다중 선형 관계를 설명한다. 다중 회귀 분석에서는 “ $y = ax_1 + ax_2 + b$ ” 와 같은 형태로 도출되는 Machine Learning Model을 찾아내는 것이다.

### 3. 아파치 스파크

빅 데이터 처리를 위한 가장 많이 알려져 있는 솔루션이 하둡

일 것이다. 하둡의 시작은 두 개의 논문에서다. 첫 번째 논문은 2003년에 발표한 “The Google File System” 이고 HDFS의 시작이 되었다. 두 번째 논문은 2004년에 발표한 “MapReduce: Simplified Data Processing on Large Clusters” 이고 맵리듀스의 기초가 되었다. 작업을 병렬로 처리하는 프레임워크인 하둡은 HDFS와 MapReduce로 구성된다. MapReduce는 분산 저장되어있는 데이터를 처리하는 프레임워크이며 HDFS(Hadoop Distributed File System)는 분산 파일 시스템이다. 하둡은 하나의 네임노드에 여러 개의 데이터노드의 정보가 저장되어 네임노드를 이용하여 처리를 위한 데이터가 어느 데이터노드에 저장되어 있는지를 알 수 있으며 MapReduce 프레임워크로 병렬 처리작업을 실행한다. 맵 프로세스는 분산된 데이터들을 병렬로 처리하고, 리듀스 프로세스는 맵 프로세스의 결과 값들을 조합한다. 하둡 2.0 에 추가된 Yarn 덕분에 자원 사용의 효율성이 좋아졌다. 그러나 하둡은 처리결과를 저장하는 방식으로 파일시스템만 제공하여 반복적으로 데이터를 처리하고 실시간처리가 필요한 머신 러닝 작업에는 적합하지 않다. 이에 반해 아파치 스파크는 빅 데이터의 실시간 처리가 유리하고 머신 러닝 라이브러리를 별도로 제공하기 때문에 머신 러닝 작업을 수행하기에 적합한 프레임워크이다.

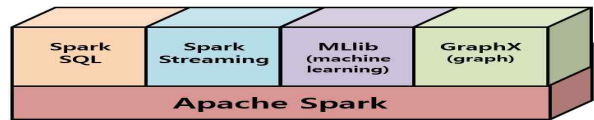


그림 4. 아파치 스파크의 라이브러리 구성

그림 4 는 아파치 스파크에서 제공되는 라이브러리 구성이다 [9]. 아파치 스파크는 디스크 상에서 데이터를 처리할 때 하둡의 맵리듀스에 비교해 보면 약 10배가 빠르며, 메모리 안에서 처리할 때는 약 100배가 빠르다. 아파치 스파크는 스칼라, 자바, 파이썬 프로그램 언어를 지원하며 병렬 처리를 위한 80개 이상의 연산자를 제공한다. 아파치 스파크는 StandAlone 방식과 Cloud 방식으로 모두 접근이 가능하다. 또한 HDFS, Cassandra, HBase, S3 등 다양한 데이터소스를 제공한다[11]. 클러스터는 여러 대의 서버가 하나의 서버처럼 연결되어 동작하는 것이다. 스파크는 클러스터메니저가 사용할 수 있는 독립적인 클러스터 모델을 제공한다.

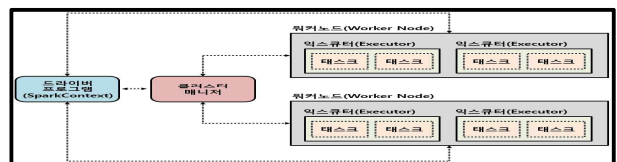


그림 5. 스파크의 클러스터 모드

그림 5 에서 알 수 있듯이 스파크의 클러스터의 구성은 클러스터매니저, 드라이버, 워커(노드), 익스큐터로 이루어진다. 드라이버 프로그램은 SparkContext 인스턴스를 생성하며 클러스터매니저와 연결하는 처리를 담당하고, 작업을 실질적으로 수행하는 서버는 워커노드이며 각 워커노드에 생성된 프로세스가 익스큐터(Executor)이다. 태스크라는 실질적인 작업 단위가 익스큐터(Executor)에 전달된다[10]. 아파치 스파크는 메모리 기반 실시간 데이터 분석이 가능하다.

#### 4. MQTT Broker 및 Kafka

##### 가. MQTT Broker

MQTT(MQ Telemetry Transport)는 아두이노나 라즈베리파이 같은 임베디드 장치 간 통신을 위한 경량 디바이스를 위한 메시징 프로토콜이다. 즉, 대역폭이 작은 통신에 적합하다. FPSML 시스템에서는 각종 센서로부터 수집되는 과일 정보 값이나 온실 생육환경 정보를 아두이노로부터 실시간으로 수집해야 하므로 메시지 푸시 기능처리로 MQTT를 사용하는 것이 적합하다. MQTT의 Server로 사용하는 장치가 MQTT Broker이다.

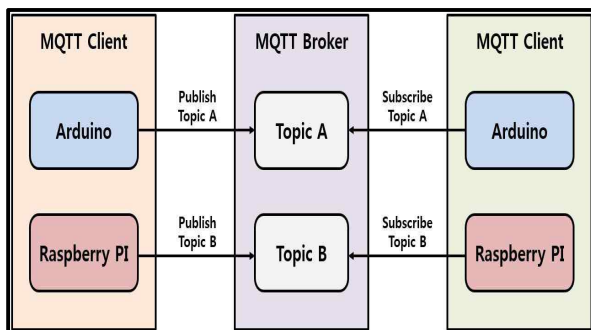


그림 6. MQTT 브로커의 메시지 전송 구조

그림 6은 MQTT Broker의 발행(Publish) 과 구독(Subscribe) 구조를 보여준다. MQTT Broker는 메시지를 푸시 하는 장치로부터 메시지를 수신하여 메시지를 수신하는 장치로 메시지를 송신하는 중계역할을 수행한다. FPSML 시스템에서는 MQTT Broker는 아두이노로부터 메시지를 수신하여 카프카로 릴레이 한다. MQTT Broker에는 Topic을 여러 개 생성할 수 있으며 Topic 별로 메시지를 발행(Publish) 하고 구독(Subscribe) 할 수 있다[12]. 본 논문에서는 색상 값을 수신하는 Topic, 무게 값을 수신하는 Topic, 온습도 값을 수신하는 Topic을 사용하였다.

##### 나. Kafka

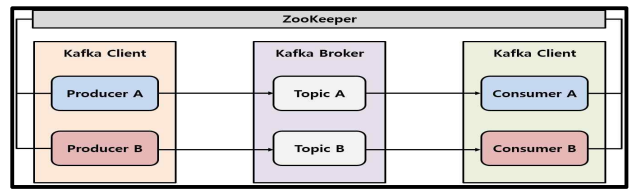


그림 7. 카프카 메시지 전송 구조

Kafka는 비동기 처리를 위한 메시징 큐로서, 생산자와 소비자를 사용하여 메시지를 송수신한다. 그림 7은 Kafka의 생산자와 소비자 구조를 보여준다. 본 논문의 FPSML 시스템에서는 빅 데이터가 실시간 방식으로 스파크로 수집되므로 실시간 메시지에 대한 안정적인 트랜잭션 처리가 필요하다고 판단되어 Kafka를 메시징 큐로 사용했다. Kafka는 노드관리를 위해서 Zookeeper를 사용하므로 Kafka Server를 운영하기 위해서는 반드시 Zookeeper 서버를 먼저 실행해야한다[13].

### III. 머신러닝 적용 과일 수확시기 예측시스템의 분석 및 설계

#### 1. 시스템 아키텍처

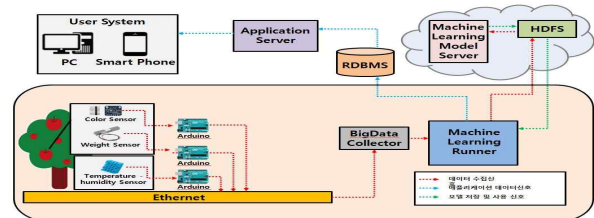


그림 8. 시스템 아키텍처

그림 8은 머신러닝 적용 과일 수확시기 예측시스템의 시스템 아키텍처이다. 각종 센서는 과일의 수확시기 관련 정보 값(색상 값, 무게 값, 실내온습도 값)을 실시간으로 측정한다. 각종 센서에서 수집된 수확시기 관련 정보 값은 아두이노로 수집된다. 아두이노에 수집된 색상 정보 와 무게 정보, 실내온습도 정보는 MQTT Broker(Mosquitto : 이하생략)를 사용하여 Kafka에 전송된다. Kafka에 수집된 데이터들은 트랜잭션 처리를 하여 안정적으로 BigData Collector에 메시징 처리를 한다. Big Data Collector는 Machine Learning Runner를 경유하여 인터넷 통신으로 HDFS(Hadoop Distributed File System)에 저장되어 Machine Learning Model을 생성하고 지속적으로 개선하는데 사용된다. Machine Learning Runner에서 생성된 수확시기 관련 예측 값들은 DBMS(Database Management System)에 저장된 후 Application Server를 거쳐 사용자들에게 서비스 된다. 사용자는 PC 나 Smart Phone을 사용하여 서비스를 제공 받는다. Machine Learning Model Server에서는

지속적으로 HDFS(Hadoop Distributed File System)로부터 빅 데이터를 읽어 들여 Machine Learning Model을 개선한다. 시스템을 계속해서 사용하면 더 많은 빅 데이터가 쌓이므로 보다 정밀한 수확시기를 예측하는 모델을 생성할 수 있다. Machine Learning Model Server에서 생성된 모델은 Machine Learning Runner에서 읽어 들여 실시간으로 수집되는 데이터에 적용한다.

### 2. Software Stack

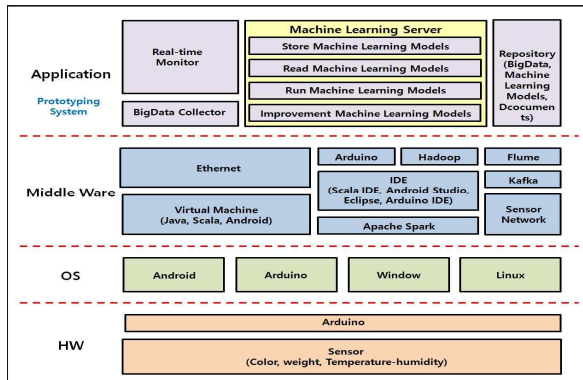


그림 9. Software Stack

그림 9는 머신러닝 적용 과일 수확시기 예측시스템의 Software Stack이다. 프로토타이핑 시스템은 Application Layer 중심으로 구현하였다. OS에 추가되어 있는 아두이노는 아두이노 소프트웨어를 의미하며 HW 부분의 아두이노는 아두이노 디바이스를 의미한다.

### 3. 머신러닝에 사용되는 특성 값 도출

머신러닝 모델을 학습시키거나 테스트하기 위해서는 특성 값들을 도출해야 한다. 각종 센서로부터 실시간으로 수집되는 데이터들은 머신러닝을 적용하기에 최적화되어 있지 않다. 수확시기 상관 값들을 머신러닝에 적용하기 위해서는 머신러닝을 적용하기에 적합한 특성 값들로 변경하여야 한다.

표 2. FPSML의 머신 러닝 데이터 셋의 특성들

측정값	특성명	측정단위	측정여부
내부온도	fpInnerTemp	섭씨(°c)	측정
내부습도	fpInnerHumi	백분율(%)	측정
과일빨강색상	fpRedColor	0~255	측정
과일녹색상	fpGreenColor	0~255	측정
과일파랑색상	fpBlueColor	0~255	측정
과일무게	fpWeight	그램(g)	측정

표 2는 과일 수확시기 예측에 사용되는 특성 값들을 보여 준

다. 과일의 수확시기를 예측하는 머신러닝 모델 단위인 FPHarvestMLModel은 내부온도, 내부습도, 과일빨강색상, 과일녹색상, 과일파랑색상, 과일무게 특성 값을 머신러닝에 적용하여 과일의 예상 수확시기를 시간단위로 예측 한다. 과일의 수확시기에 가장 영향을 미치는 기준은 과일의 무게와 색상이다. 또한, 현재의 습도와 온도가 과일의 무게와 색상을 변화시키는 속도에 영향을 주기 때문에 수확시기 상관인자에 포함시켰다. 각 특성 값들을 다중 선형회귀 알고리즘을 이용해서 머신러닝을 학습하고 테스트 한다. 본 논문에서는 아파치 스파크에서 제공되는 LinearRegression 클래스와 LinearRegressionModel 클래스를 사용하여 머신러닝을 적용하였다.

### 4. 특성 값들을 사용한 수확시기 예측단계

그림 10은 학습데이터를 학습하여 머신러닝모델을 생성하고 실시간 수집된 실 데이터에 머신러닝을 적용하여 수확시기 예측 값을 도출하는 결과 화면이다.

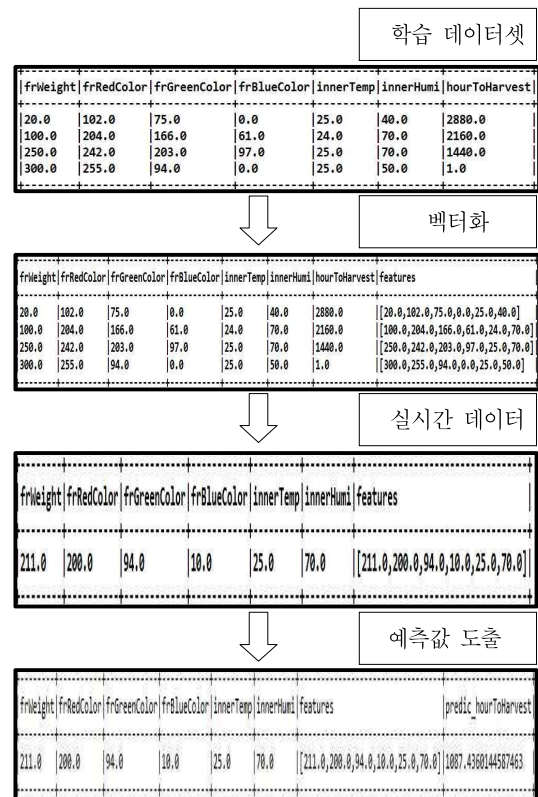


그림 10. 수확시기 예측값 도출

### 5. 머신러닝 모듈의 클래스 다이어그램

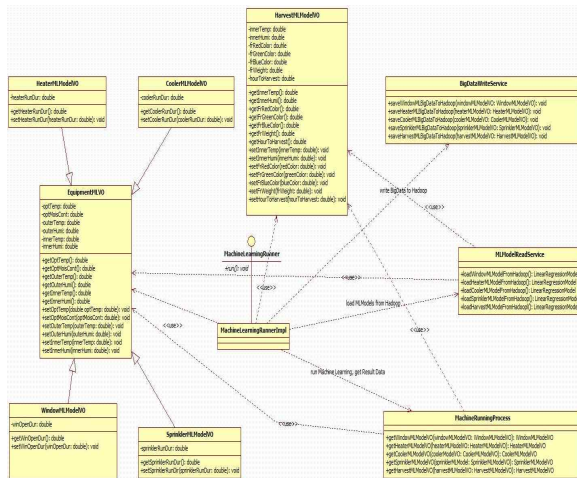


그림 11. 머신러닝 모듈의 클래스 다이어그램

그림 11은 머신러닝 모듈의 클래스 다이어그램을 보여준다. MachineLearningRunner는 머신러닝을 적용하는 시작점인 run() 메서드를 정의하는 인터페이스이다. MachineLearningRunnerImp 클래스는 MachineLearningRunner 인터페이스를 구현한다. MLModelReadService 클래스는 하둡으로부터 수확시기 예측에 사용되는 머신러닝 모델을 읽어 들이는 역할을 한다. MachineRunningProcess 클래스는 각종 센서로부터 실시간으로 수집된 수확시기 예측 관련 데이터들에 실질적으로 머신러닝을 적용하는 역할을 한다. BigDataWriteService 클래스는 하둡에 빅 데이터를 저장하는 역할을 하는 클래스이다.

## VI. FPSML의 프로토타이핑 및 실험

### 1. 센서로부터 아두이노로 데이터 수집

표 3. 수집된 데이터 형태

수집된 값	값 형태
무게 값	20g
온습도 값	Temperature : 22.1° c / Humidity : 22.6%
색상 값	Red : 220 Green: 66 Blue : 67

표 3은 센서에서 실시간으로 수집되는 측정값들이 아두이노로 수집되는 형태를 설명한다. 무게 값의 단위는 g 단위로 수집되고 온도 값의 단위는 섭씨(°c), 습도 값은 백분율(%)로 수집되며 색상 값은 RGB 단위로 수집된다.

### 2. 아두이노에서 Kafka로의 메시지 전송

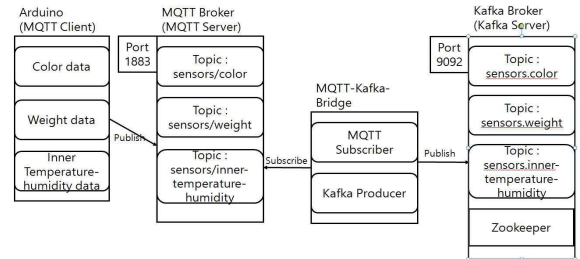


그림 12. 아두이노에서 Kafka로의 통신 구조

그림 12는 아두이노에서 Kafka로의 메시지가 전송되는 통신 구조를 보여준다. 아두이노가 MQTT Client로서의 역할을 하며 MQTT 프로토콜을 사용하여 MQTT Broker의 해당 토픽으로 실시간 측정된 수확시기 예측 관련 값들을 전송한다. MQTT-Kafka Bridge에서는 MQTT Broker의 관련 토픽으로부터 메시지를 구독한다. MQTT-Kafka Bridge는 Kafka Producer의 역할을 하며 수신한 메시지를 Kafka Broker의 관련된 토픽으로 발행한다.



그림 13. 내부 온습도 값이 Kafka Topic으로 수신된 화면

그림 13은 이클립스에서 출력한 화면으로 실시간으로 수집되는 내부 온습도 값이 Kafka Topic으로 수집되는 화면이다. Kafka Topic으로 수집될 때는 온습도의 단위를 제거하고 머신러닝을 적용하기 적합한 형태의 실수타입의 데이터로 수집한다.

### 3. 과일 수확시기 예측 머신러닝 적용

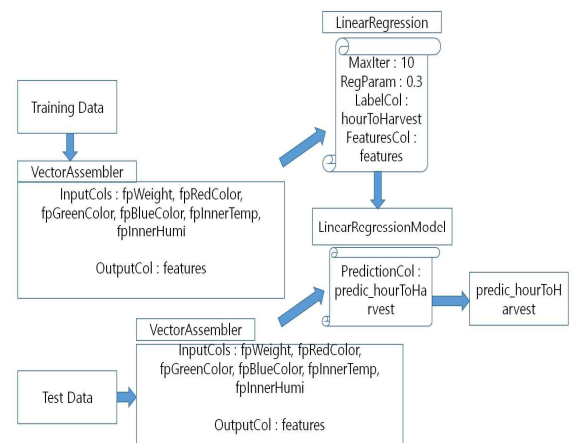


그림 14. API를 이용한 과일 수확시기 예측 머신러닝 적용 순서

그림 14는 아파치 스파크에서 제공하는 LinearRegression 클래스와 LinearRegressionModel 클래스를 사용해서 머신러닝을 적용하는 순서를 보여주는 그림이다. Training Data를 VectorAssembler 클래스를 사용하여 벡터화한 후 LinearRegression 알고리즘에 적용하여 과일 수확시기 예측에 사용되는 LinearRegressionModel을 생성한다. Test Data를 VectorAssembler 클래스를 사용하여 벡터화한 후 LinearRegressionModel에 적용하여 과일 수확시기 예측 값을 시간 단위로 얻어낸다. 본 문문의 머신러닝 학습에서는 과일의 색상 값, 과일의 무게 값, 온실 내부의 온도, 온실 내부의 습도 특성이 중요하게 사용되었다. FPSML의 경우 센서로부터 실시간으로 전송되는 값들을 빠른 속도로 분석해야 하므로 분석 속도를 향상시키기 위해서 하이퍼파라미터 값으로 사용된 RegParam 값의 경우 0.3 을 지정하여 Overfitting을 피하게 지정하였고, MaxIter 값의 경우는 10으로 지정하여 머신러닝 속도를 증가 시켰다. 수확시기 예측에 대한 학습데이터를 획득 하려면 5년 ~ 10년 혹은 그 이상의 기간이 필요하고, 농촌진흥청 등 기관에 문의 하였으나 빅 데이터를 획득할 방법이 없어서 본 논문에서 사용한 수확시기 예측에 관한 학습데이터는 임의의 데이터셋을 설계하여 사용하였다. 향후 학습에 필요한 빅 데이터가 축적될수록 더욱 정밀한 모델이 기대된다. 본 논문에서는 실제 데이터에 적용하여 검증하는 단계는 진행하지 않았고 아키텍처와 모델 제안 및 구현에 중점을 두어 연구되었다.

4. 서비스 실행 화면

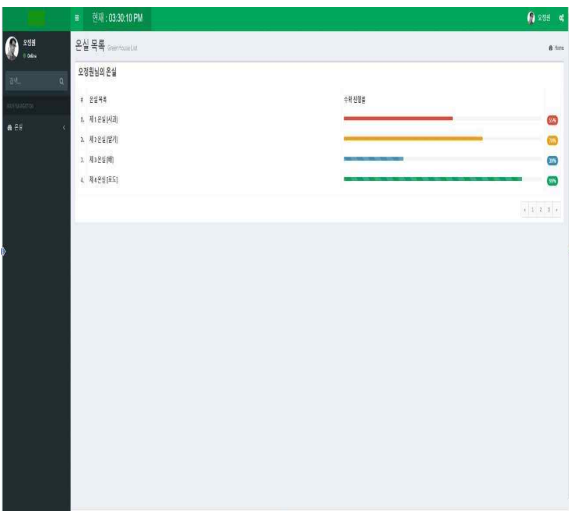


그림 15. PC에서 실행한 온실 목록 화면

그림 15는 온실 목록을 PC로 서비스 하는 화면이다. 사용자는 온실을 여러 개 등록해서 관리할 수 있으며 모니터링 화면에는 수확시기 진행률이 표시된다.



그림 16. Smart Phone 에서 실행한 온실 목록 화면

그림 16은 Smart Phone에 서비스된 온실 목록 화면을 보여 준다. FPSML 에서는 사용자에게 PC 와 Smart Phone 디바이스로 서비스를 제공한다.

V. 결론

본 논문에서는 최근 모든 산업 분야에서 가장 많은 변화를 야기하고 있고 미래의 인류에게 막대한 영향을 미칠 것으로 예상되는 머신러닝 기술을 스마트 팜 영역에 적용했다. 스마트 팜 영역 중 에서도 과수 분야에 집중하여 연구를 진행했다. 최근 스마트 팜 분야의 연구는 시설내의 생육 환경을 최적으로 유지하여 농작물의 생산성을 극대화 하는 데 연구가 중점 되어 있었다. 단순한 생산성 증대는 농작물 가격의 하락을 초래하는 부작용이 발생할 수 있으므로 농부들의 수익을 증대시키기 위해서는 생산성 증대도 중요하지만 농작물을 최고의 품질로 생산하여 수익성을 높이는 것이 중요하다는 점을 착안하여 본 논문에서는 과일의 수확시기를 정확하게 예측할 수 있는 방법을 연구 하였다. 과일의 수확시기를 정확히 예측하기 위해서 과일의 무게, 과일의 색상, 내부온도, 내부습도 등을 실시간으로 측정하여 머신 러닝에 적용하였다. 수확시기 관련 또한 실시간으로 발생하는 빅 데이터들을 하둡에 저장하여 지속적으로 정밀도 높은 머신러닝 모델을 개선해 가도록 설계하였다. 본 논문에서 개발한 시스템은 상용화 정도에 이르지는 못했으며 향후 유사 상용화 시스템 개발 시 기반 시스템으로 사용이 가능하다.

## REFERENCES

- [1] 오정원, 김행곤, "비파괴 당도센서와 색상센서를 사용한 빅 데이터 분석 적용 과일 수확시기 예측 시스템 아키텍처 개발", *인문사회과학기술융합학회*, 7권, 9호, 279-287쪽, 2017년 9월
- [2] 농부로 변신한 AI·머신러닝(2018년), <https://www.sciencetimes.co.kr>(accessed Dec., 3, 2018).
- [3] 알아서잘해요...KT, '스마트팜'에 머신 러닝 적용(2016년), <http://news.mk.co.kr/newsRead.php?no=378629&year=2016>(accessed Dec., 3, 2018).
- [4] 우리 국민이 가장 사랑하는 과일은?(2016년), <http://www.martjournal.com/news/articleView.html?idxno=177>(accessed Dec., 4, 2018).
- [5] 허정욱, 김현환, 이공인, "우리나라 스마트 팜 연구 동향", *원예과학기술지* 제34권, 별호, 33-33쪽, 2016년 12월
- [6] 김철영, "스마트 팜(Smart Farm)산업 - 농업과 ICT의 융합을 통한 고부가가치 6차 산업으로 육성 필요", *현대able Daily Market Issue*, 4-5쪽, 2016년 8월
- [7] 오정원, "빅 데이터 응용 머신러닝 기반 스마트 팜 시스템의 설계 및 구현", 박사학위논문, 2018. 8
- [8] Apache Spark, "Generality"(2018년), <http://spark.apache.org>(accessed Dec., 12, 2018).
- [9] Apache Spark, "Generality"(2018년), <http://spark.apache.org>(accessed Dec., 12, 2018).
- [10] 백성민, "빅데이터 분석을 위한 스파크2 프로그래밍", 위키북스, 175-180쪽, 2017년
- [11] Machine Learning 스테디 (1) Machine Learning이란?(2014년), <http://sanghyukchun.github.io/57>(accessed Dec., 13, 2018).
- [12] MQTT(2018년), <https://www.joinc.co.kr/w/man/12/MQTT/Tutorial>(accessed Dec., 14, 2018).
- [13] 카프카 운영자가 말하는 처음 접하는 카프카(2016년), <http://www.popit.kr>(accessed Dec., 16, 2018).
- [14] 이영훈, 김용일, "Hadoop 클러스터에서 네임노드와 데이터 노드가 빅 데이터처리 성능에 미치는 영향에 관한 연구", *스마트미디어저널*, 제6권, 제3호, 68-74쪽, 2017년 9월
- [15] 서희경, "IoT 및 네트워크 관리 지원을 위한 컴포넌트 아키텍처 개발", *스마트미디어저널*, 제6권, 제2호, 42-49쪽, 2017년 6월
- [16] 오정원, 김행곤, "IoT 기반 Apache Spark 분석기법을 이용한 과수 수확 불량 영역 모니터링 아키텍처 모델", *스마트미디어저널*, 제6권, 제4호, 58-64쪽, 2017년 12월

## 저자 소개



오정원(학생회원)

1994년 경희대학교  
입학과 학사 졸업.  
2004년 동국대학교  
컴퓨터공학 석사 졸업.  
2018년 대구가톨릭대학교  
컴퓨터공학 박사 졸업.

<주 관심분야 : MDD 개발 방법론, IOT, 빅 데이터, 딥러닝>



김행곤(정회원)

1985년 중앙대학교  
전자계산학과 학사 졸업.  
1987년 중앙대학교  
소프트웨어공학과 석사 졸업.  
1991년 중앙대학교  
소프트웨어공학과 박사 졸업.  
2014 영국 브리스톨대학  
소프트웨어공학 Ph.D

<주 관심분야 : 소프트웨어공학, 소프트웨어모델링, 융합 모델링, IOT개발 방법론, 빅 데이터응용 모델, 클라우드 컴퓨팅 개발 모델>



김일태

1990년 금오공과대학교  
전자통신공학과 학사 졸업.  
1993년 금오공과대학교  
전자통신공학과 석사 졸업.  
1995년 배재대학교  
정보통신공학과 박사 졸업.

<주 관심분야 : 5G시스템, 머신러닝, 빅 데이터 분석>