

# Agriculture Big Data Analysis System Based on Korean Market Information

Tserenpurev Chuluunsaikhan<sup>1</sup>, Jin-Hyun Song<sup>1</sup>, Kwan-Hee Yoo<sup>1</sup>, Hyung-Chul Rah<sup>2</sup>,  
Aziz Nasridinov<sup>1\*</sup>

## Abstract

As the world's population grows, how to maintain the food supply is becoming a bigger problem. Now and in the future, big data will play a major role in decision making in the agriculture industry. The challenge is how to obtain valuable information to help us make future decisions. Big data helps us to see history clearer, to obtain hidden values, and make the right decisions for the government and farmers. To contribute to solving this challenge, we developed the Agriculture Big Data Analysis System. The system consists of agricultural big data collection, big data analysis, and big data visualization. First, we collected structured data like price, climate, yield, etc., and unstructured data, such as news, blogs, TV programs, etc. Using the data that we collected, we implement prediction algorithms like ARIMA, Decision Tree, LDA, and LSTM to show the results in data visualizations.

**Key Words:** Agriculture, Big Data Analytics, Data Visualization, Text Mining.

## I. INTRODUCTION

Agriculture is one of the cornerstones of the world because food is a vital subject of humankind. People have been working in agriculture for thousands of years. However, we have not been able to fully take control of agriculture because this sector is dependent on many other factors, such as climate, weather, and people's income. Recently, people have been trying to cope with the problems of agriculture using computer technologies. The life of agriculture is becoming easier with the help of computer technologies such as smart farms, automation systems, etc. Also, the impact of big data is increasing in this sector. This sector produces big data related to agriculture like climate, weather, yield, price, and news. From big data, we obtain relevant information, such as to predict climate, price, and yield. If we can predict the future, we can use it for our purposes, such improving production efficiency, building a sustainable and competitive industry, and coping with climate hardship.

People who work in agriculture face problems, such as climate change, water availability, and drought. On the other hand, customers must worry about quality and price. As a government, investment, imports, and exports are problems that they must consider. Big data and the analysis of it are one way to solve these problems. For instance, we can use historical data to predict future results, image data to research products process, and text data to get trend topics in agriculture. But in most cases, we cannot get all of this data from one source. That's why we collect agricultural data from different sources, analyze it, and present it clearly for people. In this research, our target is the Korean agricultural market. There are so many real and synthetic-free datasets that we can implement for our research. For instance, in Korea, some governmental and non-governmental organizations provide open datasets such as climate, price, and yield. Also, we collect text data from Korean news, blog platforms, and social networks.

This paper introduces an agricultural big data analysis system. We intend to build a system that helps to make data-

---

**Manuscript received November 30, 2019; Revised December 23, 2019; Accepted December 25, 2019. (ID No. JMIS-19M-11-045)**  
Corresponding Author (\*): Aziz Nasridinov, (28644) Chungdae-ro 1, Seowon-Gu, Cheongju, Korea, +82-43-261-3597, aziz@chungbuk.ac.kr.

<sup>1</sup>Department of Computer Science, Chungbuk National University, Cheongju, Korea, {teo,thdwlsgus,khyoo}@chungbuk.ac.kr

<sup>2</sup>Big Data Cooperative Course, Chungbuk National University, Cheongju, Korea, hrah@chungbuk.ac.kr

---

driven decisions for the government, farmers, and customers. The system consists of two main contributions, such as the collection of agricultural big data, and analyzing agricultural big data.

- Agricultural big data collection: Agricultural data is collected into one database whether it is structured or unstructured. By doing so, we can analyze the big data in many ways and see the interdependence between structured and unstructured data.
- Agricultural big data analysis: The collected big data is analyzed to obtain relevant information such as prediction of future results, detection of hidden values, and understanding the market process. Our result of big data analysis helps to understand the Korean agricultural market and make decisions.

We organized the paper as follows: Section 2 provides related studies that include similar research and systems. We describe our proposed method in Section 3. The proposed method consists of data collection, data pre-processing, data analysis and visualization. Also, we introduce our experiment results in Section 4 with subsections such as experimental setup and experimental results. At the end, we describe our conclusion and future works.

## II. RELATED STUDY

In this section, we describe some research works that we studied to get ideas and motivation for our research work. We studied other works and try to find the advantages and disadvantages. We intend to include the advantages and to solve the disadvantages in our system. This section consists of agricultural big data research papers and agricultural big data systems.

Researchers study many sections of agriculture using big data, such as market analysis, coping with climate change, smart farming, and IoT. Jianghui et al. [1] introduced an integration and analysis of agricultural market information based on web mining. The authors focused on how to effectively extract and use data found on the Internet. They proposed three kinds of methods such as focused, incremental and custom based on web crawling technology. Through these data collecting methods, they collected a large amount of data and analyzed them by web mining and visualization technologies. Amirhossein et al [2] used machine learning techniques to detect nearby and distant lightning hazards by using four kinds of parameters: air pressure, air temperature, relative humidity, and wind speed. They claim that their model has statistically considerable predictive skill for lead times up to 30 min. Yunong et al. [3] detected the growth of apples using factors such as size, color, cluster density, and others. They find that traditional detection methods cannot be adapted to different growth stages using a same model, and they

proposed an improved YOLO-V3. They use the DenseNet method for process feature layers with low resolution, combine it with YOLO-V3 and named it YOLOV3-dence. They proved the proposed method is superior to the YOLO-V3 and can be applied in the actual environment of orchards. Jirapon, et al. [4] proposed a smart sensor system for watering crops. The smart system consists of three kinds of components: the control box, web application, and smartphone application. The control box collects data on crops, the web application manipulates the details of crop data, and the smartphone application controls the crop watering. They applied data mining techniques to predict the suitable temperature, humidity, and soil moisture for crops in the future. The result showed that this work was suitable for use in agriculture.

To get an idea for our system we studied some big data analyzing systems. For example, Google Trends, SomeTrend, and BigFos. Google Trends [5] helps to analyze and understand a data trend based on the google search engine. Google Trends works by visualizing trends by a search keyword, comparing by search keywords and selecting options such as region, date, and category. There is also a system named SomeTrend [6] which was developed in Korea, similar to Google Trends. SomeTrend has features, such as Issue Analysis, Big Data Analysis, and Insight Report. We can learn about a recent trend search keyword and analyze big data by custom search keyword. When we insert a search keyword, it shows us visualizations of trend analysis, related words, and emotional analysis. BigFos [7] is a supply and demand prediction system of five major vegetables which are cabbage, radish, dried red pepper, onion, and garlic. BigFos consists of a real-time big data collection framework, price forecast information using AI, and prediction service through collecting agricultural climate information, information exchange and policy related to supply and demand through community communication.

Table 1. The comparing table of big data analysis systems.

	Advantages	Disadvantages
Google Trends	<ul style="list-style-type: none"> <li>• Using Google source</li> <li>• User-friendly</li> <li>• Comparing search keywords</li> <li>• Embedding visualization</li> </ul>	<ul style="list-style-type: none"> <li>• Do not have many types of visualizations</li> <li>• Cannot see in very detail</li> </ul>
SomeTrend	<ul style="list-style-type: none"> <li>• User-friendly</li> <li>• Has many types of visualizations</li> </ul>	<ul style="list-style-type: none"> <li>• Can see only last 1 month</li> <li>• Cannot select any options like categories</li> </ul>
BigFos	<ul style="list-style-type: none"> <li>• Many types of data sources</li> <li>• Many types of data analytics</li> <li>• Well explained</li> </ul>	<ul style="list-style-type: none"> <li>• Only five kinds of vegetables</li> <li>• A User has to wait unknown time until the account is enabled</li> </ul>

From these related studies, we can see that big data helps solve many problems in agriculture. Because big data

analysis gives us a chance to obtain hidden values from massive amounts of data. Therefore, we also try to contribute to this sector with our system. Table 1 shows a comparison of big data visualization systems. From the table, a big data system has to focus on being user-friendly, the definition of visualizations, and data sources. The advantages and disadvantages of similar systems are provided to use the advantages and avoid the disadvantages in our system.

### III. PROPOSED METHOD

In this section, we describe our system in detail. Section III.A describes the overall structure of our system. We explain our methods to collect data in Section III.B. In Section III.C we introduce some data analytic methods which are implemented in our system, and we explain the data visualization step of the system in Section III.C.

#### 3.1. Overview

Fig. 1 shows the overview structure of our system. Our system consists of data collection, data preprocessing, data analysis and visualization. We collect agricultural big data from news, social networks, and other Korean agricultural open data platforms. the Korean open data platforms provide structured data such as price, climate, and yield. We also collect unstructured data using web crawling technology. We collect data from published news, broadcast news, blogs, and Twitter and store it in our database. Our dataset, which has been collected since 2010 has become large enough to analyze and visualize. We focus on three kinds of data analysis: statistical analysis, deep learning analysis, and machine learning analysis. In the end, our results are visualized using user interactive, functional charts.

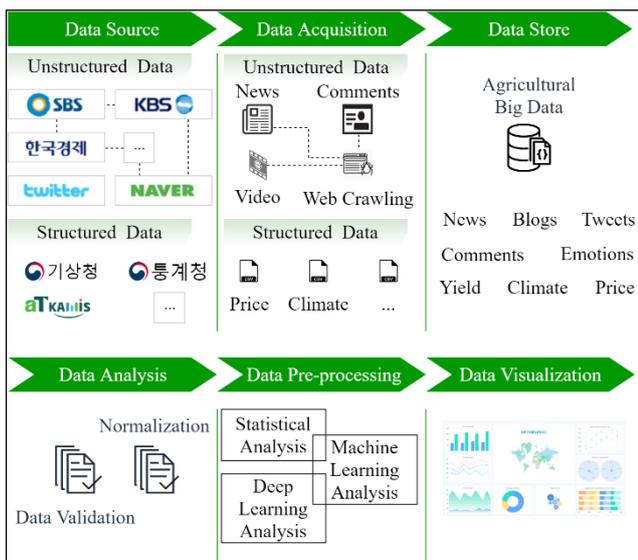


Fig. 1. Overall Structure of the System

#### 3.2. Data Collection

##### 3.2.1. Structured data collection

In this system, we collect structured and unstructured data related to the Korean agricultural market. The structured data is obtained from some organizations which provide open data. To collect text data from news, blogs, and social networks, web-crawling technologies are used

In Korea, there are some organizations which provide public or researchers with open data. For instance:

- Rural Development Administration. RDA is a governmental and agricultural organization in South Korea, and belongs to the Ministry of Agriculture, Food, and Rural Affairs. RDA helps the agricultural sector create more value through their efforts.
- Korea Meteorological Administration. KMA is the national meteorological service of South Korea. KMA is a governmental organization of the Republic of Korea under the Ministry of Environment. KMA collects climate information and is committed to analyzing and providing the information.
- Korea Agri-Fisheries & Food Trade Corporation (aT). This is a wholly government invested corporation which was established in 1967. The goal of the organization to increase the income of farmers and fishermen, and to contribute balanced development of the national economy by the law of Agricultural & Fishery Marketing Corporation. This organization provides Korean agricultural products' distribution information through Kamis which is a portal for this information.
- Korea Rural Economic Institute. KREI is a government-funded research organization. The purpose of the organization is, "To contribute to the nation's economic development and the enhancement of public welfare by conducting comprehensive surveys and research on the agricultural and forest economy and rural community development". This organization employs Outlook & Agricultural Statistics Information System. This system provides market panel price data.
- Korea Statistical Information Service. KOSIS is a gateway for Korea's official statistical information. As of 06.12.2018 (1:00), official statistics on 1089 subject matters (92 designated statistics and 997 general statistics) are officially approved under the Article 18 (or Article 20) of the Statistics Act [11].

Figure 2 shows the method of collection for structured data. Agricultural structured data is obtained from open data portals like RDA, OASIS, KAMIS, and KOSIS, etc. First, variables are selected to satisfy our research goals. We also collect data on daily, weekly, monthly, and yearly date ranges. After that, we clean the data and format it in from one into another. For data cleaning, we remove extra spaces, remove all blank or null cells, remove duplicates, and delete all formatting. At the end, we store the data in MongoDB.

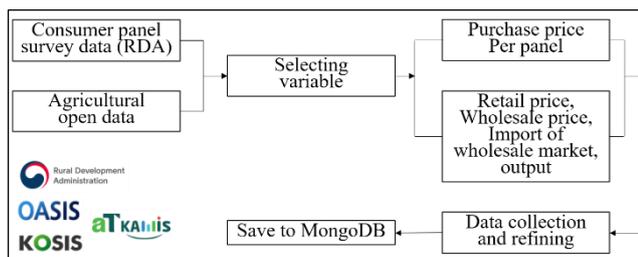


Fig. 2. Method of collecting structured data.

Table 2 shows the sources of structured data. Consumer panel data is collected from the Rural Development Administration, and other open data from public agricultural data portals between 2010 and 2017. These portals provide many kinds of data with many variables. Needed variables from each open data source are selected, downloaded to a CSV file, and stored in MongoDB after refining the data.

Table 2. Structured data collection.

Data type	Variable	Collection site
Panel data	Consumer Panel Data (KRW) (Average purchases per week per person)	http://rda.go.kr
	Retail price (KRW)	https://kamis.or.kr
Wholesale and Retail price data	Wholesale price (KRW)	http://oasis.krei.re.kr/
	Wholesale import volume (kg)	
	Yield per unit (10a) (kg)	
	Cultivated area (ha)	http://kosis.kr/
	Production of year (kg)	
	Production of previous year	
Climate	Temperatures (C)	http://data.kma.go.kr

### 3.2.2. Unstructured data collection

We also collect text and video data from news, blogs, and twitter using web crawling. We select keywords related to agriculture such as vegetables, fruits, and meat. The keywords also include different names of products, similar words, translated words, etc. Using this keyword structure, we can collect all data without any loss. We use web crawling to obtain text data from the sources. Figure 3 shows the method of collecting unstructured data. Broadcast data consist of broadcasting such as video programs and their script that is transformed into text by jsoup library. News data is collected from news text of MBC, KBS, SBS, etc. Text data is stored in MongoDB after purifying it and processing the natural language.

We started our project in 2018. In the first phase of the project, we focused on collecting unstructured data by developing data collection software using jsoup in Java. In 2019, we started the second phase of collecting structured data and developing a big data visualization system. Before starting the second phase, we measured our data that was already collected. Table 3 shows the measurement of the

data that was collected in 2018 and from which we are still collecting date.

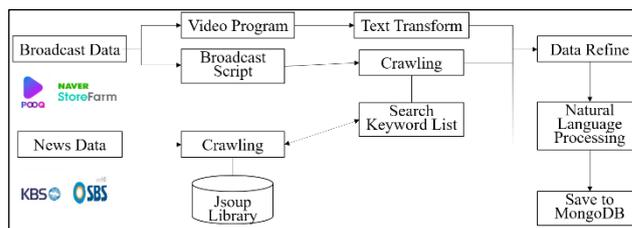


Fig. 3. Method of collection unstructured data.

Table 3. Volume of unstructured data collection.

Channel	Number of data	Amount of data /GB/
Twitter	31,874,996	17.06
Broadcast News	328,802	1.3
Published News	974,652	5.29
Blog	11,005,993	58.31
Total data	44,184,443	81.96

Fig. 4 shows the flowchart of web crawling. To collect data related to agriculture, we use the following steps:

- Search a document by an agricultural key word like onion, garlic, beef, etc.
- Extract a URL from the results of search.
- Receive an HTML document from the URL.
- Obtain needed information like title, date, content, etc., use CSS selector.
- Clean the parsed text such as removing blank spaces, removing formats, and others.
- Store the data in the database.
- Repeat the steps until the search result is finished.

Here, we explain the crawling steps in detail. First, we need to fetch and parse an HTML document from the web using a URL. After that, we can obtain the needed data from the document by using either DOM or selector which we chose. The selector is a CSS-like element selector, that finds elements by matching a query. For instance, "Element title = doc.select ("div.news-title").first();" returns the title of a document. The selector can find content by tag name (div, a, p, etc.), id ("#title"), class (".title"), and attribute ([href, src, etc.]). We collect title, content, published date, publisher, number of emotions, number of comments, and the comment itself. In other sources like blogs and Twitter, sections are a little bit different than news.

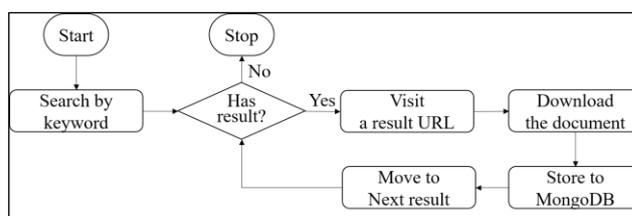


Fig. 4. Flowchart of web crawling.

### 3.3. Data Preprocessing

Preprocessing prepares raw data for analyzing and is a necessary step before data analysis. In our case, we implemented data pre-processing techniques in our data analysis to: check for missing values, handle categorical values, and split a dataset into training and test sets. The complicated thing about data preprocessing is that it is different in every situation. We collect various data, such as structured and unstructured; therefore, we implemented the different types of data pre-processing techniques based on our data analysis.

Our data analysis consists of statistical analysis, machine learning analysis, and deep learning analysis. Also, the data pre-processing technique that we used is different in data analysis. For statistical analysis, we made data integration first because we collect data from different sources. Matching data from multiple sources is a problem for data integration. We solve this problem using metadata which is data about data. After that, we apply some data cleaning techniques based on the purpose. In machine learning and deep learning analysis, data integration and data cleaning are also required. Some data transformation techniques are also applied which transform data into suitable forms of mining.

### 3.4. Data Analysis and Visualization

To explore relevant information, we applied various types of data analysis such as statistical analysis, machine learning analysis, and deep learning analysis. In statistical analysis, we visualize structured data and unstructured data. We show the information on facility vegetable prices provided in public data and news frequency. We also show the relationship between price, climate, and production and predict purchase, price, and yield using machine learning and deep learning algorithm. For instance, we analyzed onion disease using LDA which is a machine learning algorithm. Thus, consumption price is predicted with structured and unstructured data based on LSTM which is a deep learning algorithm.

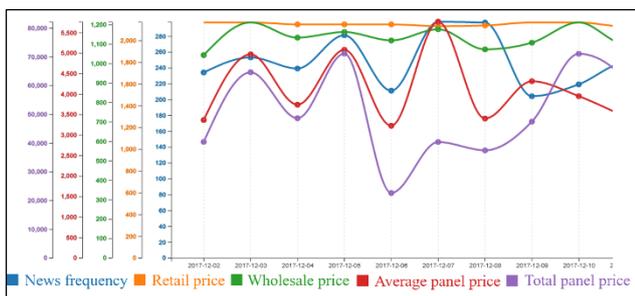


Fig. 5. Price total information.

#### 3.4.1. Statistical Analysis

Various agricultural data is collected, such as consumer price, climate, and yield, and the size of the unstructured data is one of our statistical data. To get a quick, clear understanding of the information, we analyze some data using statistical analysis. These analyses show relationships between different types of data and help to make future data-driven decisions. For instance, price is also an essential topic in agriculture. In most cases, price depends

on yield, product quality, and natural phenomena. In this section, we try to understand if there are relations between prices, news frequency, and news text.

Figure 5 shows a correlation between news frequency and prices. In this figure, The Y-axis represents news frequency, retail price, wholesale price, average consumer panel price, and total consumer panel price. The X-axis shows the date.

Figure 6 shows information about average temperature, yield, and crop area. The relation between weather and yield is shown in this figure. Thus, we can see a relationship between temperature and yield, temperature and price, yield and price, etc. In this figure, the X-axis shows the years and the Y-axis shows the size of average, yield, and area.

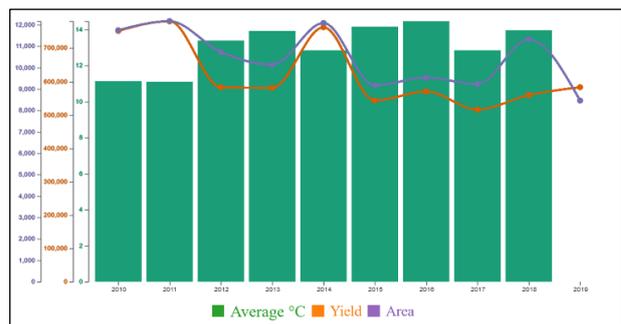


Fig. 6. A relation between yield and climate data.

#### 3.4.2. Machine Learning Analysis

One of the purposes of our system is to predict the future based on structured data. to achieve this, we use some machine learning algorithms which are ARIMA, Decision Tree Regression, and K-Means. We predict wholesale and retail price using ARIMA and Decision Tree Regression. ARIMA is a popular and widely used statistical method which is a model-based algorithm designed to predict time series data. ARIMA stands for Autoregressive Integrated Moving Average Model. Decision tree regression predicts data in the future with continuous output.

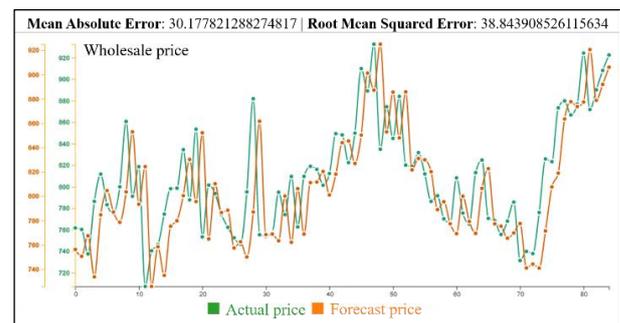


Fig. 7. Wholesale price prediction using ARIMA.

Figure 7 shows the result of the ARIMA model based on prediction wholesale price. The Y-axes show actual and forecast prices. The X-axis represents time. As can be seen in Fig. 7, there is not much difference in actual and forecast prices, which means the model is qualified for our data.

Figure 8 shows the result of the Decision Tree Regression model based on prediction wholesale price. The Y-axes shows actual and forecast prices. The X-axis represents

time. Unlike the ARIMA model, the Decision Tree Regression model does not show enough results.

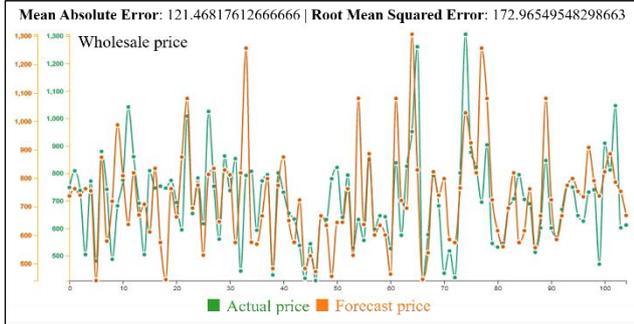


Fig. 8. Wholesale price predict using Decision Tree Regression.

We also use K-Means clustering algorithm to understand our data. K-Means is a simple and popular unsupervised machine learning algorithm. It can make inferences without any labeled data. This algorithm finds K groups of data. Figure 9 shows a correlation between yield size and the average temperature during a year.

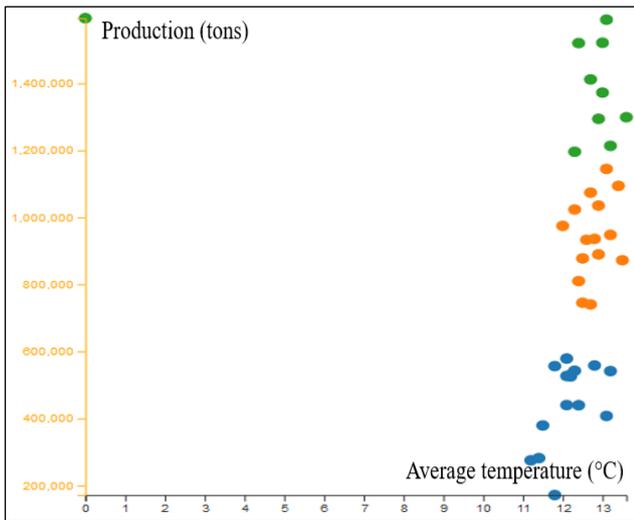


Fig. 9. Segments of yield and average temperature.

This study identified the statistically significant correlations between paprika purchases with data from news and blogs [7]. News, blogs, and social media in which paprika was mentioned was used between 2010 and 2017 and the results of this study could be applied to promote bell-pepper consumption. Figure 10 shows cross-correlation coefficients between the broadcasting entertainment programs mentioning bell-peppers and diet and the consumer panel data between 2010 and 2017. Based on the results, what could be suggested may include the followings:

- For promoting consumption of bell-peppers, publicity by using broadcast entertainment programs and blogs appears more effective than news.
- For promoting consumption of bell-peppers through broadcasting programs and blogs, it is necessary to emphasize it as a diet food.

Another research work of ours is related to onion disease. Previous some studies tend to improve onion's taste and other ingredients. But this study proposes a method for extraction of impact factors related to onion disease [5]. Because the quality of onion is directly related to the demand of consumers. If the quality is high, it can have a positive effect on demand.

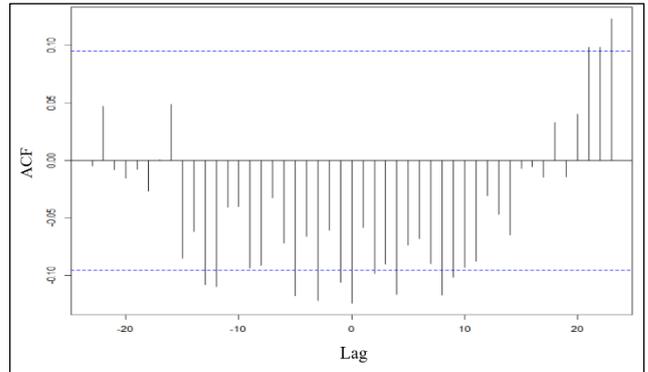


Fig. 10. Cross-correlation coefficients of video\_positive\_term\_freq\_paprika\_diet.ts and panel\_purchase\_amount.ts.

In this study, we analyze the collected articles using the LDA algorithm and extract impact factors related to onion disease. Figure 11 shows three topics that have been created. In the essential topic, words such as drought, rainfall, disaster, damage, and pests indicate that this topic is the climate is one of the onion disease impact factors.

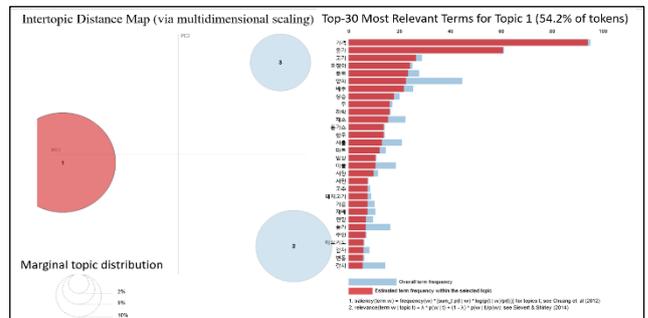


Fig. 11. Topic extraction using Pyldatryvis.

### 3.4.3. Deep Learning Analysis

Price of products depends on many factors. For example, demand, supply, and disease of products. We can calculate demand and supply easily because it has almost the same frequency. But it is difficult to predict disease. In this study, we propose a method to predict the purchase price of agricultural and livestock consumer panels using deep learning by utilizing the collected structured and unstructured data. The reason for estimating the purchase price of a consumer panel is that it is possible to understand the purchase pattern and consumption price of a consumer depending on the price of the panel. To predict the purchase price of the consumer panel, we apply the LSTM model. Figure 12 shows predicted result.

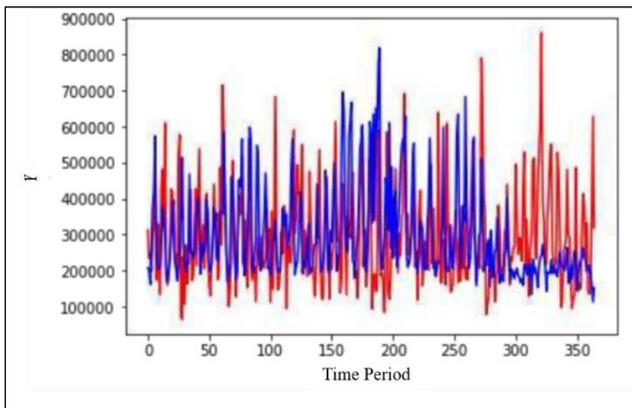


Fig. 12. Predicted result.

#### IV. Performance Analysis

In this section, we discuss the performance analysis. We first show the experimental setup in Section 4.1 and Section 4.2, we discuss experiment results.

##### 4.1. Experimental Setup

In this section, we discuss the evaluation of the agricultural big data analysis system. Table 3 shows the properties of the experimental environment. The system runs on an Ubuntu server with an Intel i7 processor and 16 GB of ram.

Table 4. Evaluation setup.

Property	Value
CPU	Intel (R) Core (TM) i7-4770 CPU @ 3.40GHz
RAM	Samsung PC3-12800 16GB
OS	Ubuntu 16.04.5 LTS
HDD	Seagate ATA STD500D 1TB (7200 RPM)

The development of the system consists of two main programming languages, such as Java and Python. We use Java for web crawling technology. There is a Java library named jsoup, which works with real-world HTML. It has a very suitable API for extracting data from HTML pages. We use the Python language for other tasks like data analysis, data visualization, etc. When visualizing data, a graph library is important. We choose D3.js for data visualizing, which is a Javascript library for manipulating documents based on data. D3 builds graphs using HTML, SVG, and CSS. Using the D3 we can build powerful visualization, but it is little bit complicated because it is the raw library.

##### 4.2. Experiment results

We use two common metrics such as the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to measure the accuracy of the machine learning models. The metrics are in the range of 0 to  $\infty$  with lower values being better. We apply ARIMA and Decision Tree Regression algorithms to predict prices in the future. The ARIMA model predictions are off by approximately 30.2 in MAE,

and 38.8 in RMSE. But in the Decision Tree Regression, performances are 121.5 and 172.97. From these performances, the ARIMA model is more suitable in our dataset.

In deep learning, to predict the purchase price consumer panel, we apply the LSTM model to the structured and unstructured data. The model predicts price with 82% accuracy over a 30-second time period.

#### V. CONCLUSION

Big data can be the future of agriculture. Data analytics of big data contains relevant knowledge to cope with the problems that the government and farmers face. This study presents the Agricultural Big Data Analysis System. This is a web application that helps to understand agricultural big data and make data-driven decisions. In the system, our purpose is to get useful information based on the data analytics of big data.

To achieve our purpose, we collected data related to agriculture using open data platforms and web crawling technologies. Based on the data, we implemented data mining techniques to obtain relevant information.

In the future, we are interested in extending the scope of the study. For instance, we will try to understand consumer opinions based on user comments and emotions which they write on news and blogs. We studied some research about consumer opinions of specific products or services. But news and blogs are different than products and service, which will be our challenge. Another idea is to predict drought using ground and surface water, and climate data.

##### Acknowledgement

This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET), funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (319003-01).

##### REFERENCES

- [1] J. Zhou, C. Cheng, Li Kang, and R. Sun, "Integration and Analysis of Agricultural Market Information Based on Web Mining," *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 778-783, 2018.
- [2] A. Mostajabi, D. L. Finney, M. Rubinstein, F. Rachidi, "Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques," *NPJ Clim Atmos Sci* 2, vol. 41, 2019.
- [3] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Computers and Electronics in Agriculture*, vol. 157, pp. 417 – 426, 2019.
- [4] J. Muangprathub, N. Boonnam, S. Kajornkasirat, N. Lekbangpong, A. Wanichsombat, P. Nillaor, "IoT and agriculture data analysis for smart farm" in *Computers*

and Electronics in Agriculture, vol. 156, pp. 467 – 474, 2019.

- [5] Google; <https://trends.google.com/> (accessed on Nov 2019)
- [6] SocialMetrics; <http://www.some.co.kr/> (accessed on Nov 2019)
- [7] <https://bigfos.kr/> (accessed on Nov 2019)

Convergence (PSM program) at Chungbuk National University as a faculty member.



**Aziz Nasridinov** is currently an Assistant professor of the Department of Software at Chungbuk National University. He received the M.S. and Ph.D. degrees in Computer Engineering from Dongguk University in 2009 and 2012. From 2012 to 2014, he was a

Postdoctoral researcher in the Department of Multimedia at Sookmyung Women's University. He was an Assistant professor of Computer Engineering at Dongguk University in 2014 and 2015. His research interest is in database, big data, data mining distributed processing system.

### Authors



**Tserenpurev Chuluunsaikhan** is currently a master student in the Department of Computer Science at Chungbuk National University, Korea. He received his bachelor's degree from the Computer Science and Management School, Mongolia.

His research interest is in the areas of big data.



**Jin-Hyun Song** is currently a master student in the Department of Computer Science at Chungbuk National University, Korea. He received his bachelor's degree from the Department of Software at Chungbuk National University in 2018. His research interests are text mining of big data.



**Kwan-Hee Yoo** is currently a professor of the Department of Software at Chungbuk National University since 1997. He received the M.S. and Ph.D. degrees from Korea Advanced Institute of Science and Technology in 1988 and 1995.

From 1988 to 1997, he was a Senior researcher in the Dacom. He was an Exchange Professor at Carnegie Mellon University between 2003 and 2005. His research interest is in Computer Graphics, Artificial Teeth Bodling, 3D Games, 3D Content.



**Hyung-Chul Rah** obtained his DVM degree at KonKuk University in Korea, and Master of Preventive Veterinary Medicine (MPVM) and Ph.D. degree in Comparative Pathology at School of Veterinary Medicine, University of California, Davis, USA. He was trained as

postdoctoral researcher and research associate at School of Medicine, University of California, Davis, USA and CHA University, Korea. His research interests are development of genetic markers for diseases including cancers and inherited disorders and its commercialization. Since 2014, he joined Graduate School of Health Science Business