Article

Forecasts of the Amount Purchase Pork Meat by Using Structured and Unstructured Big Data

Ga-Ae Ryu [1], Aziz Nasridinov [1], HyungChul Rah [2],* and Kwan-Hee Yoo [1],*

[1] Department of Computer Science, Chungbuk National University, Cheongju 28644, Korea;
garyu@chungbuk.ac.kr (G.-A.R.); aziz@chungbuk.ac.kr (A.N.)
[2] Department of Management Information Systems, Chungbuk National University, Cheongju 28644, Korea
* Correspondence: hrah@cbnu.ac.kr or rah.remnant@gmail.com (H.R.); khyoo@cbnu.ac.kr (K.-H.Y.);
Tel.: +82-43-261-3636 (H.R.); +82-43-261-2788 (K.-H.Y.)

check for updates

Abstract: It is believed that the huge amount of information delivered to the consumers through mass media, including television and social networks, may affect consumers' behavior. The purpose of this study was to forecast the amount required to purchase pork belly meat by using unstructured data such as broadcast news, TV programs/shows and social network as well as structured data such as consumer panel data, retail and wholesale prices and production outputs in order to prove that mass media data release can occur ahead of actual economic activities and consumer behavior can be predicted by using these data. By using structured and unstructured data from 2010 to 2016 and five forecasting algorithms (autoregressive exogenous model and vector error correction model for time series, gradient boosting and random forest for machine learning, and long short-term memory for recurrent neural network), the amounts required to purchase pork belly meat in 2017 were forecasted and compared with the actual amounts to validate model accuracy. Our findings suggest that when unstructured data were combined with structured data, the forecast pattern is improved. To date, our study is the first report that forecasts the demand of pork meat by using structured and unstructured data.

Keywords: agri-food; purchase forecast; unstructured big data; social network service; pork meat

1. Introduction

Recently, there was an outbreak of African swine fever in South Korea, which severely affected pork meat consumption and price [1]. It has been believed that news on agri-food was delivered to the consumers through mass media including television and social networks, and the information may affect the behavior of consumers' behavior who were exposed to this huge amount of information [2–4]. The purpose of this study was to forecast pork consumption in terms of the amount required to purchase pork belly meat by using unstructured data, such as broadcast news, TV programs/shows and social network as well as structured data, such as consumer panel data, retail and wholesale prices and production outputs in order to prove that mass media data release, including various unstructured data can predate actual economic activities, and consumer behavior can be predicted by using these data.

Prediction of economic activities by using social network data or internet search data ahead actual activities has been reported in the stock market, marketing and tourism [5–7]. Recently, prediction of economic activities by using social network data or internet search data have been reported in agriculture [4,8–11]. However, the data of broadcast news and TV programs/shows, for which recipes have been provided during popular cooking shows and through social network, have never been applied to forecast either demands or prices of agri-foods. In this study, we aimed to demonstrate that broadcast news, TV programs/shows and social network as unstructured data combined with

structured one could be used to forecast demands of agri-food by using pork belly meat data. This study could help to understand the effects of broadcast news, TV programs/shows and social network on agri-food consumption eventually.

2. Materials and Methods

2.1. Data

We previously developed a prediction model of agri-food demand by unstructured and structured bigdata, in which structured data on agri-food production and sales and unstructured data from mass media including broadcasting programs and social network were collected and saved in Mongo database as seen in Figure 1 [2,12]. These bigdata were used to predict the demand of agricultural products in Korea by collecting and analyzing structured and unstructured data together.

In this paper, Agri-food consumers panel data provided by Rural Development Administration (RDA), wholesale market data of Outlook and Agricultural Statistics Information System (OASIS) of Korea Rural Economic Institute, retail price data of Korea Agricultural Marketing Information Service (KAMIS), pork production data of Korean Statistical Information System (KOSIS) were used as structured data whereas data from broadcasting programs and blogs were used as unstructured data [2,12–15].

Structured data of production and sales of pork belly meat from 2010 to 2017 were extracted from Mongo database and prepared for analysis in order to predict the demand pork belly meat such as amounts to purchase pork belly meat, daily retail prices of pork belly meat, daily wholesale prices of pork carcass, and others as seen in Table 1. Agri-food consumers panel data were from Korean consumer panels and have feature of amount required to purchase from 2010 to 2017. Data frequencies of structured data were daily, monthly, quarterly and yearly according to the data source as seen in Table 1 except Agri-food consumers panel data of which transaction data were summed for daily average per consumer panel.

Unstructured data that matched a keyword search were collected from broadcasting programs and blogs. Speech from broadcasting programs were converted into text or the transcripts were collected. Unstructured data is data that indicates broadcasting programs and social network where the term "pork belly meat" was mentioned, such as broadcast news, television program/shows, and Blogs in Korea as seen in Table 2. Unstructured data were transaction data which were summed for daily frequencies to match with structured data.
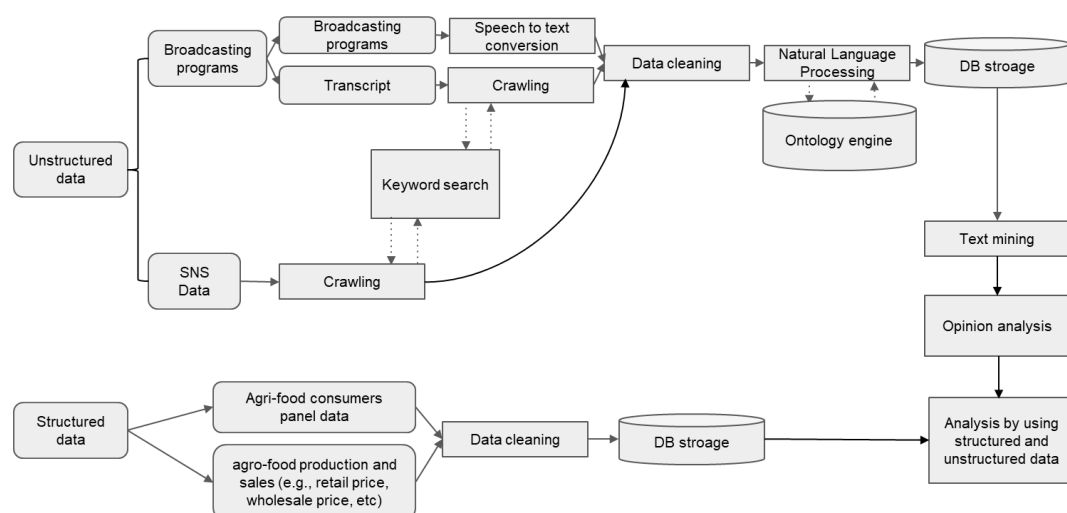

Figure 1. Agri-food related structured and unstructured bigdata modified from [2]. SNS for social network service, DB for database.

Table 1. List of features of structured data used in the study.

Data type	Data name	Feature name	Description
Structured data	Agri-food consumers panel data	Panel_purchase_amount	Daily average amount required to purchase pork belly meat per consumer panel
	Sales of pork meat	Retail_price_meat	Daily retail prices of pork belly meat
		Wholesale_price_carcass	Daily wholesale prices of pork carcass
		Wholesale_price_carcass_quarter_before	Daily wholesale prices of pork carcass in previous quarter
		Monthly_sales_trend_ton_meat	Monthly sales trend of pork meat (ton)
	Production of pork meat	Pig_bred_number_quarter_before	Number of pig bred in previous quarter
		Pig_slaughtered_number_quarter_before	Number of pig slaughtered in previous quarter
		Output_ton_year_before_carcass	Pork meat production in previous year (ton)
		Import_ton_year_before	Imported pork meat in previous year (ton)

Table 2. List of features of unstructured data used in the study.

Data Type	Data Name	Feature Name	Description	Data Counts
Unstructured data	Broadcast news	News_freq	Daily frequency that keyword term was mentioned in broadcast news	6655
		Emotions_Number_Angries	Daily frequency of comments with angry emoticon of broadcast news in which keyword term was mentioned	3979
		Emotions_Number_likes	Daily frequency of comments with like emoticon of broadcast news in which keyword term was mentioned	14,811
		Emotions_Number_sads	Daily frequency of comments with sad emoticon of broadcast news in which keyword term was mentioned	395
		Emotions_Number_wants	Daily frequency of comments with want more reports emoticon of broadcast news in which keyword term was mentioned	438
		Emotions_Number_warms	Daily frequency of comments with moved emoticon of broadcast news in which keyword term was mentioned	153
		News_comment_freq	Daily frequency of comment of broadcast news in which keyword term was mentioned	44,342
		News_positive_term_freq	Daily frequency of positive term of broadcast news in which keyword term was mentioned	35,319
		News_negative_term_freq	Daily frequency of negative term of broadcast news in which keyword term was mentioned	4429
	Television program/shows	Video_freq	Daily frequency that keyword term was mentioned in television program/shows other than broadcast news	1529
		Video_total_ranking_ave_p	Average television view rate of television program/shows in which keyword term was mentioned	1529
		Video_freq_times_viewrate	Video_freq times Video_total_ranking_ave_p	1529
		Video_positive_term_freq	Daily frequency of positive term of television program/shows in which keyword term was mentioned	119,396
		Video_negative_term_freq	Daily frequency of negative term of television program/shows in which keyword term was mentioned	4745

Table 2. *Cont.*

Data Type	Data Name	Feature Name	Description	Data Counts
		Blog_freq	Daily frequency that keyword term was mentioned in blog	75,035
		Blog_comments	Daily frequency of comment of blog in which keyword term was mentioned	109,950
	Blogs	Blog_likes	Daily frequency of comments with like emoticon of blog in which keyword term was mentioned	70,025
		Blog_positive_term_freq	Daily frequency of positive term of blog in which keyword term was mentioned	1,666,492
		Blog_negative_term_freq	Daily frequency of negative term of blog in which keyword term was mentioned	56,870

2.2. Forecasting Methodology

Forecasting models were developed in order to forecast the daily average amount required to purchase pork belly meat in 2017 by using data from 2010 to 2016 (in-sample period) as a training data set and data from 2017 (out-of-sample) as a test data set. Structured and unstructured data were used for training and testing, whereas structured data alone were also used to compare if unstructured data could improve models' forecasting. Different algorithms were used to develop forecasting models including the autoregressive exogenous model, vector error correction model as traditional time-series algorithms, gradient boosting and random forecast as machine learning algorithms, and long short-term memory as a neural network algorithm. This is because, in relation to price prediction, the time series analysis model is mainly used, and recently, analyses by using machine learning, artificial neural network, or deep learning model have been attempted. The machine learning model shows better predictive power than the regression analysis model [16]. Forecasted amounts required to purchase pork meat in 2017 were compared with actual amounts required to purchase it in 2017 in terms of mean absolute percentage error (MAPE) and mean absolute error (MAE) in order to compare the accuracy of the forecasting models. The Diebold-Mariano test was used for forecast comparison by using the DM.test function in multDM package in R [17,18].

2.2.1. Time Series: Autoregressive Exogenous Model and Vector Error Correction Model

In order to forecast the amounts required to purchase pork belly meat in 2017, time series analysis was used, including autoregressive exogenous modeling and vector error correction modeling. The autoregressive exogenous model (ARX) and vector error correction modeling (VECM) are models in the Multivariate Time Series. When the observed variable and the predicted variable is more than one, the multivariate time series approach is more appropriate [19]. The Autoregressive Exogenous (ARX) model is an autoregressive model with exogenous variables and is a representative and quantitative dynamics modeling approach that has often been used in time series analysis [20]. In order to forecast daily amounts to purchase pork belly meat in 2017 by using the ARX model, data from 2010 to 2016 and the arx function in the 'gets' package in R were used [17]. In order to forecast weekly amounts required to purchase pork belly meat in 2017 for comparison, daily amounts required to purchase pork belly meat were averaged on a weekly basis.

The vector error correction model (VECM) was developed by Engle and Granger and aimed to accommodate the insertion of short-term adjustments due to the presence of integration [19,21]. In VECM as well as VAR (Vector Autoregressive) models, more than one variable can be predicted because the interrelations between variables can be seen [19]. In order to forecast the daily amounts required to purchase pork belly meat in 2017 by using the VECM model, the lag selection of 1 was selected based on the information criteria function before the granger causality and cointegration degree were tested in Eviews 10.

2.2.2. Machine Learning: Gradient Boosting and Random Forest

Machine learning began with Samuel's paradox that if a computer could learn from experience, it would be a hassle [22]. Therefore, machine learning can be defined as the creation of a computer program that solves a problem with high performance by using data or experience that occurs in a specific area [23]. Machine learning methods include supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. It is appropriate to use supervised learning for consumption prediction. In this paper, we chose random forest and gradient boosting, which is one of the most used algorithms in map learning.

Random forest is a representative ensemble model, which is a combination of tree models having the same distribution extracted independently [24]. Random Forest is an ensemble technique algorithm based on decision trees. A decision tree is created, but a tree is created by randomly selecting some of the attributes used to create the tree instead of using the entire attribute. This process is repeated randomly to create slightly different trees. By using the forest of these trees, the predictions for the new data are voted for to choose the final predictions. In order to forecast daily amounts to purchase pork belly meat in 2017 by using Random Forest model, randomForest package in R was used. When structured and unstructured data were used, the parameters of ntree = 1500, mtry = 13 were used whereas structured data were used, the parameters of ntree = 1000, mtry = 5 were used.

Like Random Forest, Boosting is one of the most popular ensemble models. However, unlike random forest, which generates a tree randomly, boosting compensates for the error of the binary tree and creates the tree sequentially. The feature of the boosting model is that weak learners can come together to make excellent ensemble learners. Boosting gives each learner a weight based on past performance. Thus, a good model has a greater impact on the final prediction of the ensemble [25]. Gradient boosting is one of the boosting algorithms that improve the model by correcting residual errors in the previous model. In order to forecast the daily amounts required to purchase pork belly meat in 2017 by using the Gradient Boosting model, xgboost package in R was used. When structured and unstructured data were used, the parameters of max.depth = 15, eta = 0.03 were used whereas when structured data were used, the parameters of max.depth = 8, eta = 0.03 were used.

2.2.3. Long Short-Term Memory

The Long Short-term Memory (LSTM) is a modification of the existing RNN (Recurrent Neural Network) model, which is a typical improvement model that mitigates the vanishing gradient problem of RNN [26,27] as seen in Figure 2. LSTM is an artificial neural network commonly used for time-series data analysis. This model includes a memory cell in the hidden node which stores and outputs the value and adjusts the forgetting value. What is interesting in this model is that the LSTM consists of an input gate for the input value, an output gate for the output value, and a forgetting gate for the forgetting value. The learning algorithm of the LSTM uses Backpropagation in the same way as the RNN; the input data is the sequence data, the output data is the output data of LSTM. The LSTM model includes three gates, so that the number of weights and the number of biases is about four times that of typical RNN learning, which means that execution time and learning time of LSTM model are longer than the RNN models. Despite this, the vanishing gradient problem can be mitigated to obtain more accurate results.

In this study, the panel purchase amount was forecasted by using a basic LSTM model and two kinds of data sets, structured and unstructured data, and structured data alone for comparison. The data learning and prediction method is shown in Figure 3. First, the data is transformed by using the log, and then data is normalized by min max scaling. After that, the correlation coefficient is calculated for each column through the panel purchase amount which currently predicted through correlation. The correlation coefficients for each column were estimated through the correlation relation analysis, and the weighting was calculated based on correlation coefficient. To learn and predict time-series data, the LSTM model is created in many-to-one form with multiple inputs and one output. Then, values for the LSTM model for sequence length and hidden dimension were specified. Once the LSTM model

was determined, the data from 2010 to 2016 were trained for each sequence in order to forecast the daily average amount required to purchase pork belly meat in 2017. In this method, a total of 48 cases were generated to find optimized parameter values of the LSTM model with the best forecasting accuracy. Each case was made by changing the parameters that affect the model during learning in LSTM (e.g., sequence length, hidden dimension, and stack layers). Details of the selected experimental cases are listed in Table 3 and details of all 48 experimental cases are listed in Table S1.
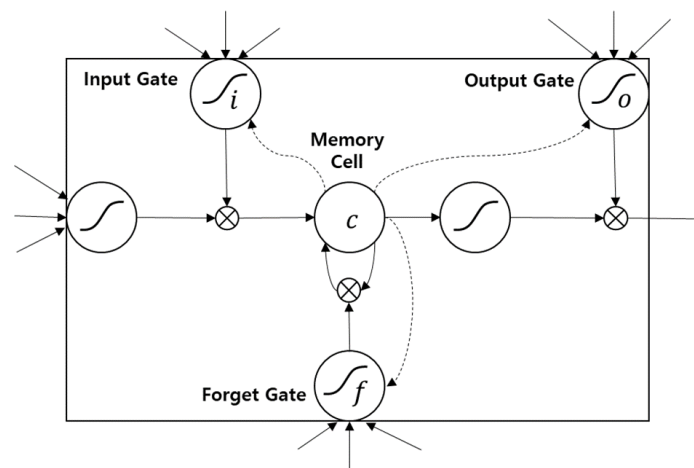
Figure 2. Long Short-Term Memory block modified from [26].
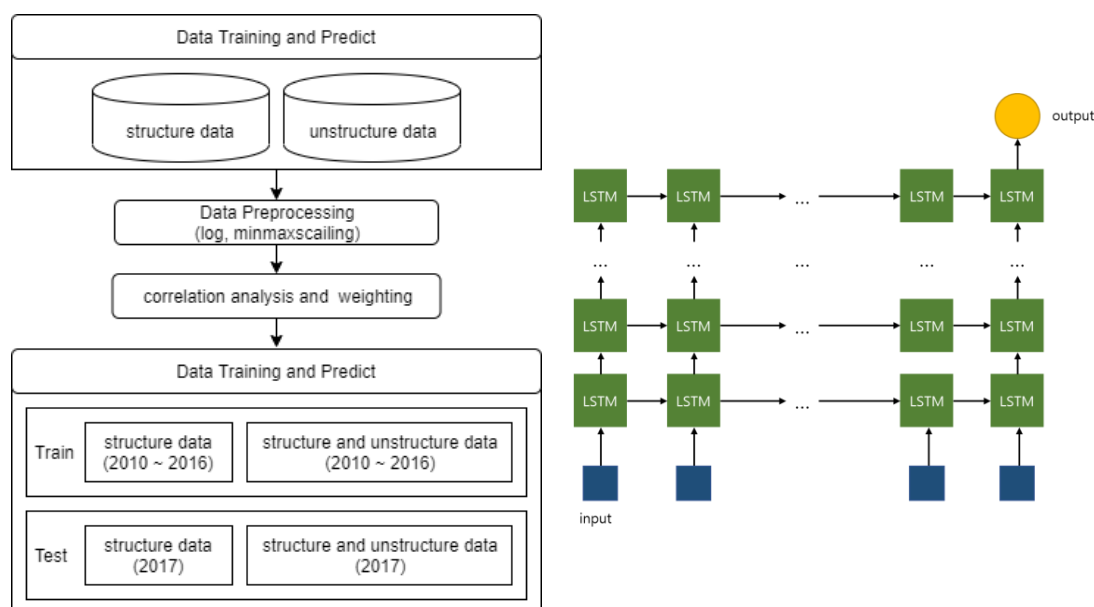
Figure 3. Data prediction method using Long short-term memory.

Table 3. Selected experimental cases to find forecasting model parameters.

Case18	Case27	Case33	Case36
seq_length = 14	seq_length = 21	seq_length = 21	seq_length = 21
hidden_dim = 15	hidden_dim = 20	hidden_dim = 10	hidden_dim = 5
forget_bias = 0.5	forget_bias = 0.5	forget_bias = 0.5	forget_bias = 0.5
stacked_layers = 12	stacked_layers = 12	stacked_layers = 12	stacked_layers = 12
keep_prob = 0.5	keep_prob = 0.5	keep_prob = 0.5	keep_prob = 0.5
epoch = 1000	epoch = 1000	epoch_num = 1000	epoch = 1000
learning_rate = 0.01	learning_rate = 0.01	learning_rate = 0.01	learning rate = 0.01

3. Results

3.1. Forecasted Daily Amounts Required to Purchase Pork Belly Meat

Five algorithms were developed to forecast daily amounts required to purchase pork belly meat in 2017, and model accuracy was compared by using MAPE and MAE. Two different data sets were used to develop the forecast model, which include structured data alone and structured and unstructured data, as seen in Table 4. Among the ten forecast results, LSTM with structured data showed the lowest MAPE whereas ARX with structured and unstructured data showed the lowest MAE, which showed no statistically significant difference in forecasts comparison by using Diebold-Mariano test (DM statistic = 0.7041, *p*-value = 0.4814).

Table 4. Comparison of forecasted daily amounts to purchase pork belly meat in 2017.

Forecasting interval	Data	Algorithm	Mean Absolute Percentage Error (MAPE)	Mean Absolute Error (MAE)	Remark
Daily	Structured	Autoregressive exogenous	15.18	2547.5	lag = 1
		Vector error correction model	14.62	2601.4	lag = 1
		Random Forest	15.95	2690.3	ntree = 1000, mtry = 5
		Gradient boosting	18.31	3141.6	max.depth=8, eta=0.03
		Long short-term memory	14.51	2570.5	seq_length = 21 hidden_dim = 5 forget_bias = 0.5 stacked_layers = 12 keep_prob = 0.5 epoch = 1000 learning rate = 0.01
	Structured and Unstructured	Autoregressive exogenous	15.05	2522.7	lag = 1
		Vector error correction model	17.7	2823.2	lag = 1
		Random Forest	17.74	2859.6	ntree = 1500, mtry = 13
		Gradient boosting	16.9	2863.5	max.depth = 15, eta = 0.03
		Long short-term memory	14.59	2629.0	seq_length = 21 hidden_dim = 20 forget_bias = 0.5 stacked_layers = 12 keep_prob = 0.5 epoch = 1000 learning rate = 0.01

Four forecasted models were compared with actual amounts in the graph including ARX with structured and unstructured data, LSTM with structured and unstructured data, ARX with structured data, and LSTM with structured data as seen in Figure 4. In Figure 4, the patterns of ARX with structured and unstructured data and ARX with structured data alone mimics the pattern of actual daily amounts whereas the patterns of LSTM with structured and unstructured data and LSTM with structured data stay close the mean. The patterns of ARX with structured and unstructured data display a more similar pattern to the actual pattern than that of ARX with structured data alone in terms of height and depth.

3.2. Forecasted Weekly Forecased Amounts Required to Purchase Pork Belly Meat

Daily forecasted amounts required to purchase pork belly meat were averaged on a weekly basis in order to see if forecasting errors could be reduced, as seen in Table 5. Among the ten forecast results, ARX with structured and unstructured data showed the lowest MAPE and MAE; however, it did not show a statistically significant difference compared to ARX with structured data in forecasts

comparison by using Diebold-Mariano test (DM statistic = 1.3432, *p*-value = 0.1792). The same four models were compared with actual amounts, as seen in Figure 5. The patterns of ARX with structured and unstructured data and ARX with structured data alone, which stay close to each other, mimics the pattern of actual weekly amounts better than the patterns of LSTM with structured and unstructured data and LSTM with structured data.

Table 5. Comparison of forecasted weekly amounts to purchase pork belly meat in 2017.

Forecasting interval	Data	Algorithm	Mean Absolute Percentage Error (MAPE)	Mean Absolute Error (MAE)	Remark
Weekly	Structured	Autoregressive exogenous	6.33	1123. 5	lag = 1
		Vector error correction model	7.98	1460.8	lag = 1
		Random Forest	6.84	1212.0	ntree = 1000, mtry = 5
		Gradient boosting	9.78	1731.1	max.depth = 8, eta = 0.03
		Long short-term memory	7.25	1335.8	seq_length = 14 hidden_dim = 15 forget_bias = 0.5 stacked_layers = 12 keep_prob = 0.5 epoch = 1000 learning rate = 0.01
	Structured and Unstructured	Autoregressive exogenous	6.15	1091.7	lag = 1
		Vector error correction model	9.18	1538.6	lag = 1
		Random Forest	9.04	1535.2	ntree = 1500, mtry = 13
		Gradient boosting	7.75	1359.6	max.depth = 15, eta = 0.03
		Long short-term memory	6.5	1163.8	seq_length = 21 hidden_dim = 10 forget_bias = 0.5 stacked_layers = 12 keep_prob = 0.5 epoch = 1000 learning rate = 0.01

3.3. Forecasted Errors in LSTM When Structure Data and Unstructured Data Were Used Over Structure Data Alone

The forecasting results of the various LSTM cases are listed in Figure 6 with MAPE. Most of the daily and weekly results show fewer MAPE results when both structure data and unstructured data were used than when structure data alone was used. In the case of daily forecasts, the lowest MAPE was 14.59 of case 27 when structure and unstructured data were used. When structure alone data were used, the MAPE was 14.51 of case 36. In the case of weekly forecasts, the lowest MAPE was 6.5 of case 33 when structure and unstructured data were used. When structure data were used alone, the MAPE was 7.25 of case 18. In both forecasts we compared these results with actual values on a daily and weekly basis. The forecasting errors were lower when both structure data and unstructured data were used.
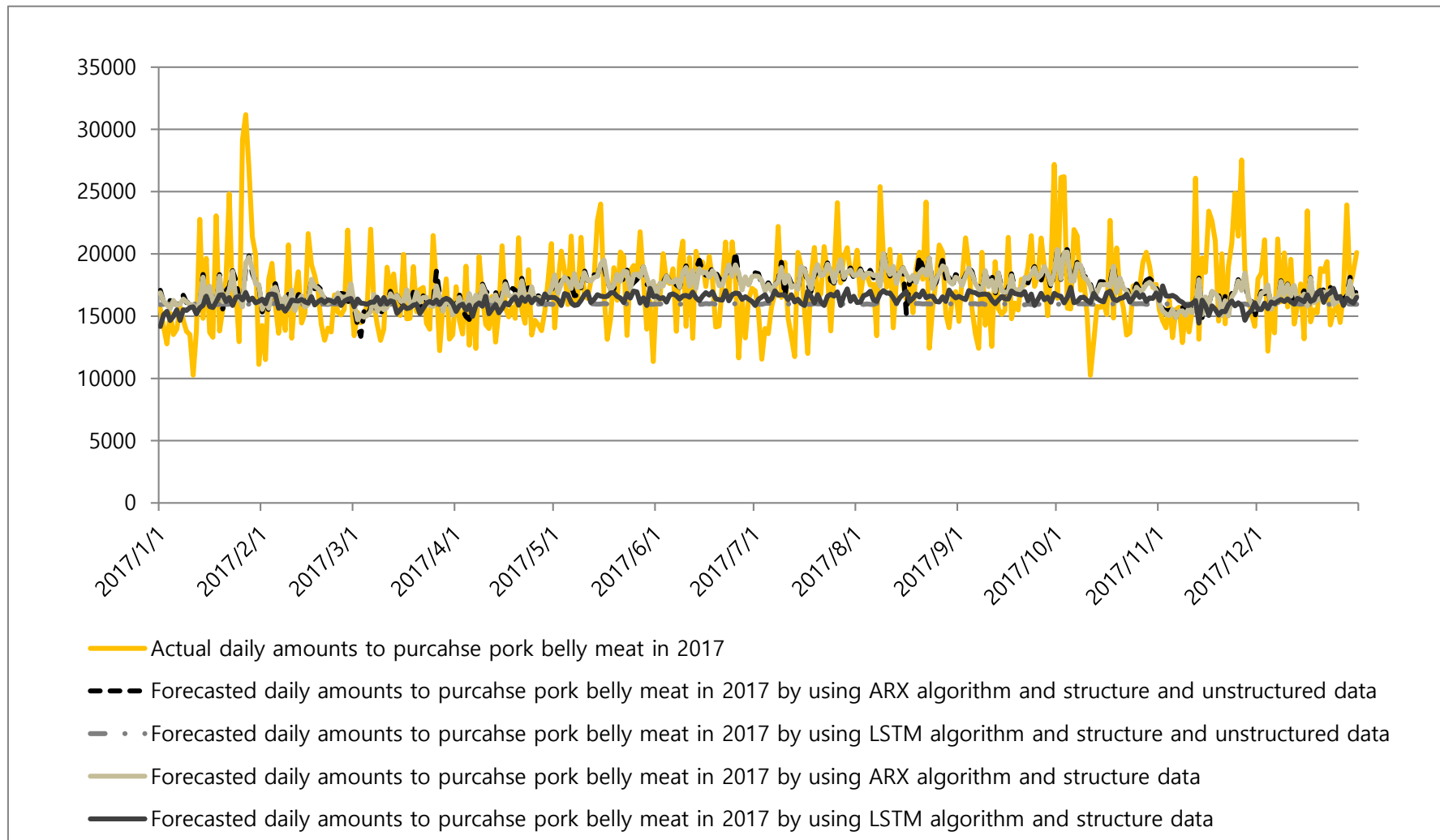
Figure 4. Comparison of forecasted daily amounts to purchase pork belly meat in 2017. ARX = Autoregressive exogenous, LSTM = Long short-term memory.
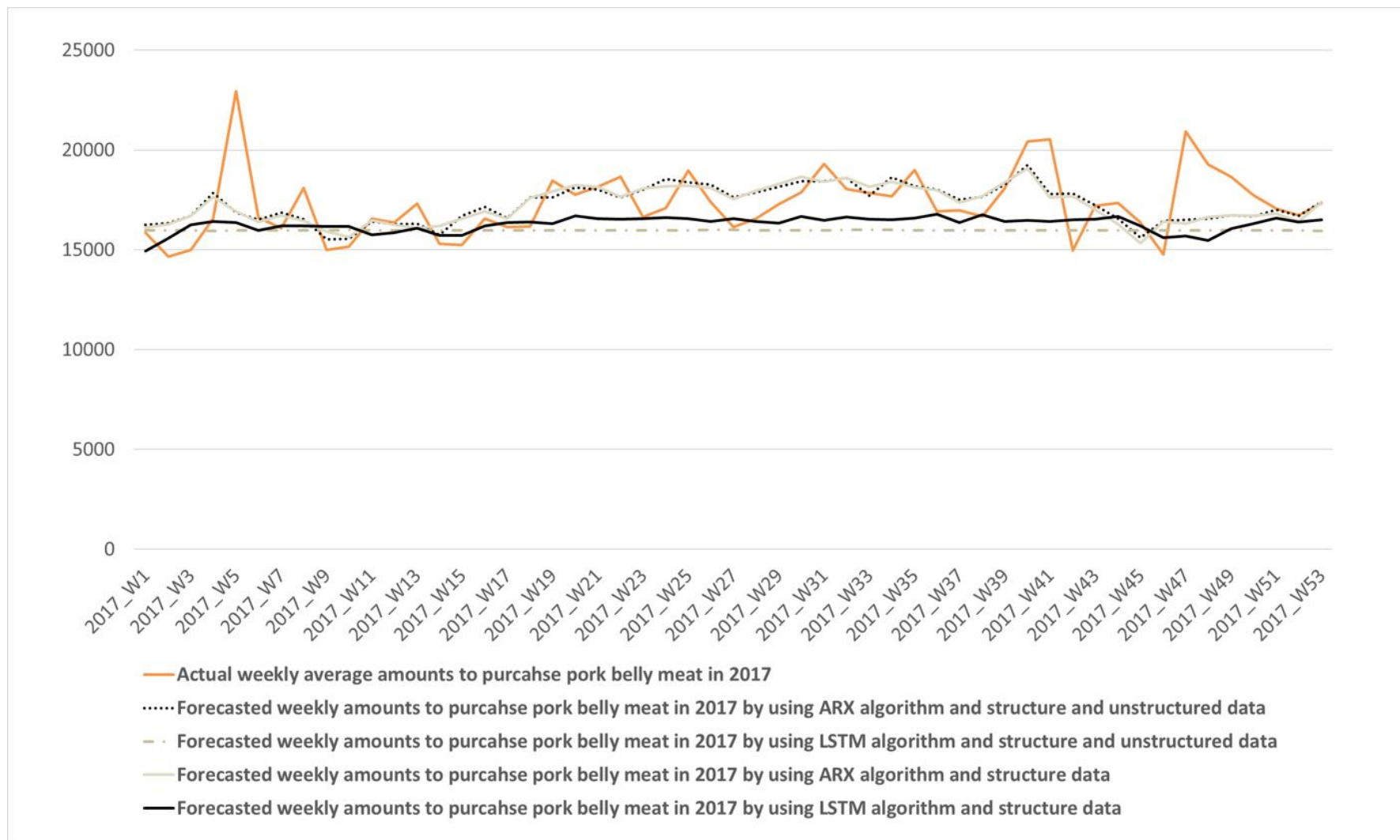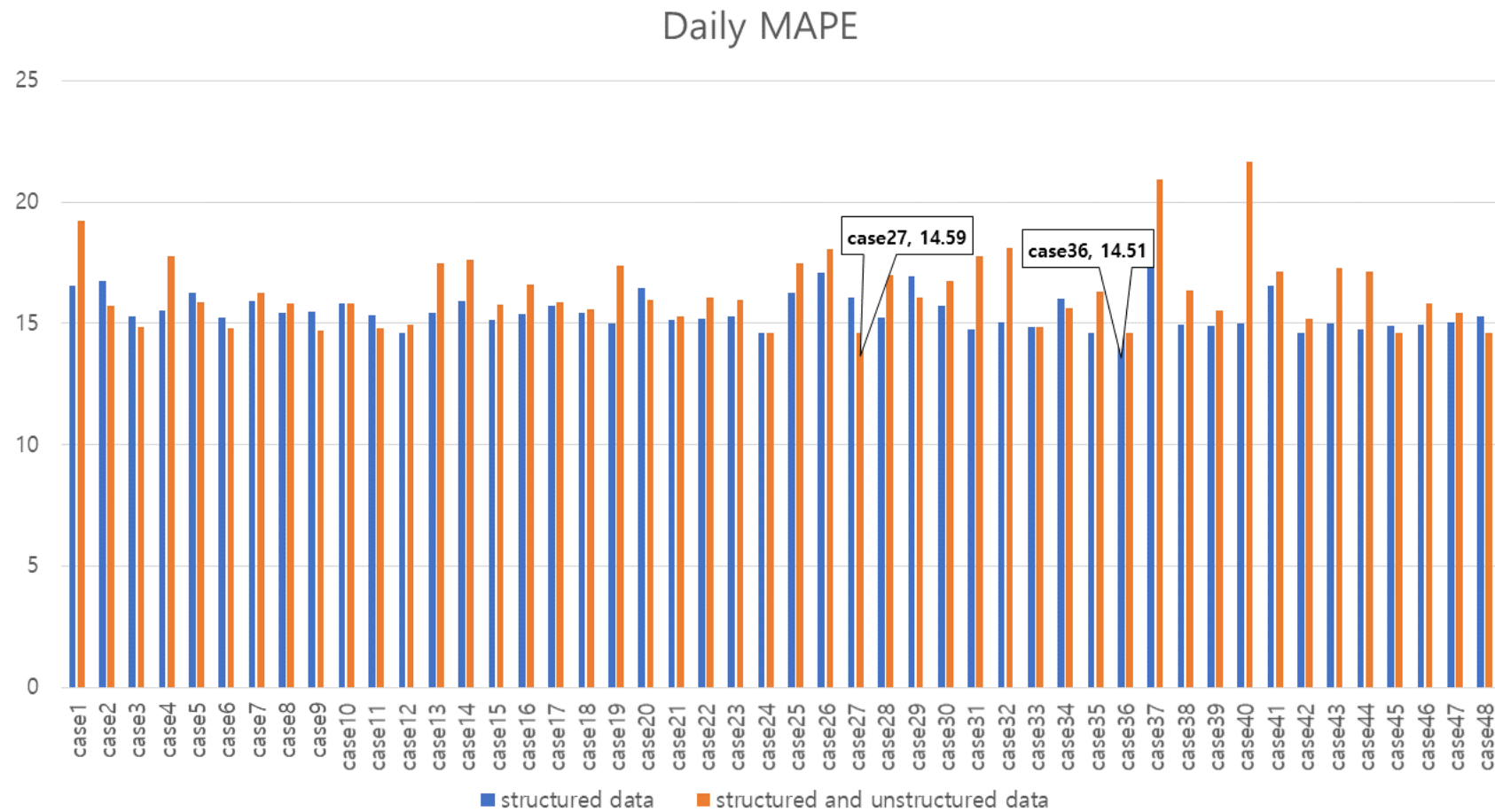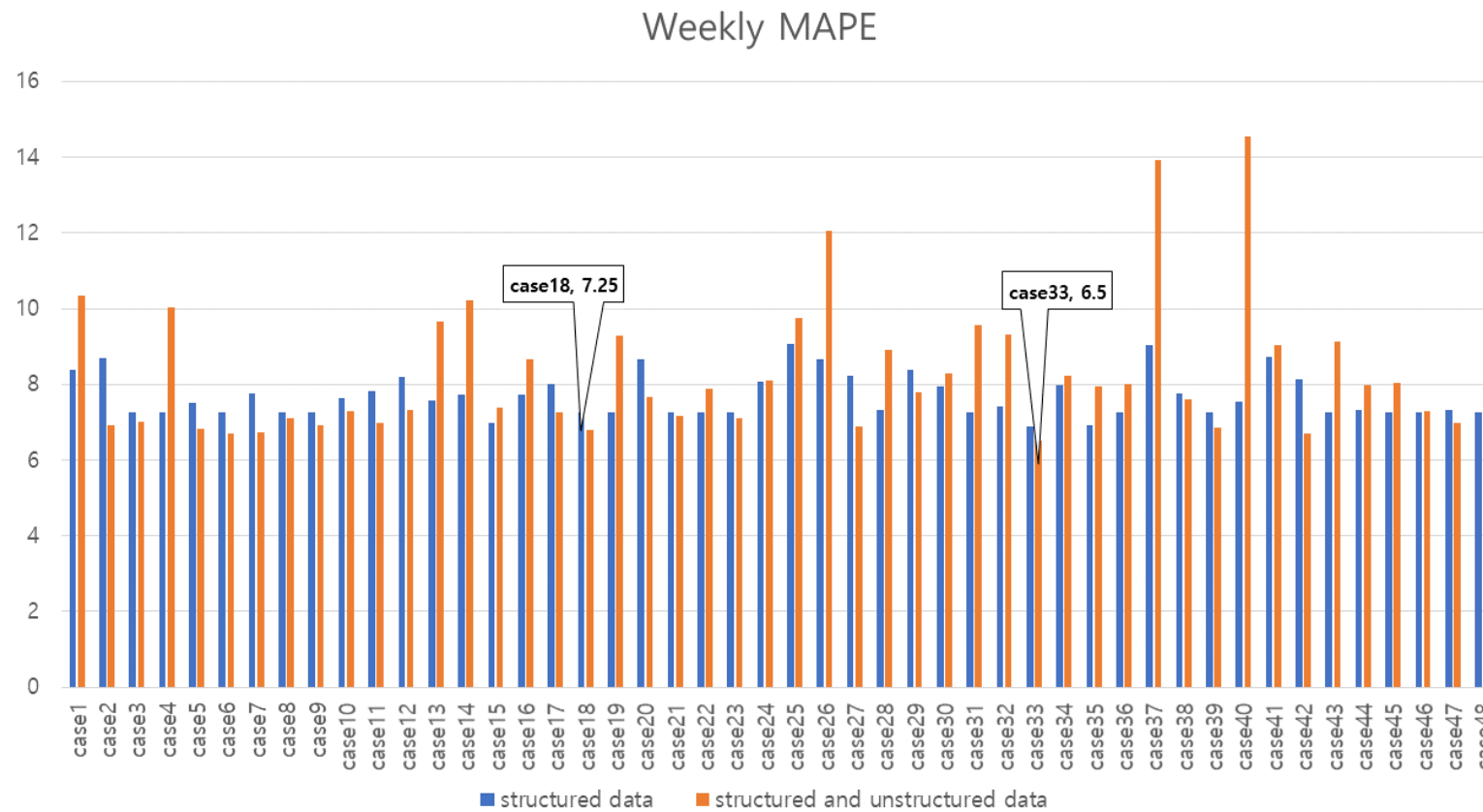
Figure 5. Comparison of forecasted weekly amounts to purchase pork belly meat in 2017.

(a)

Figure 6. *Cont.*

(**b**)

Figure 6. Comparison of the forecasting results by structured data with the forecasting result by structured and unstructured data in daily basis (**a**) and weekly basis (**b**).

4. Discussion

In this study, we aimed to demonstrate that broadcast news, TV programs/shows and social network could be used to forecast demands of agri-food by using pork belly meat data, one of the popular meat products for Korean consumers. Our findings may suggest that when broadcast news, TV programs/shows and social network, which were grouped as unstructured data were combined with structured data, that include consumer panel data, retail and wholesale prices and production outputs, it improves the forecast pattern.

There have been a few reports that social network or internet search data were used to predict the price of agri-food; however, no paper was reported that data of broadcast news and TV programs/shows were used to predict either prices or demands of agri-food until a very recent one on paprika consumption prediction [4,8–11,28]. To date, our study is the first report that predicts the demand of livestock products by using unstructured data of broadcast news, TV programs/shows and social network as well as conventional structured data. Production of agri-food with better forecasts of prices and demand of agri-food by using structured and unstructured data could contribute to a stable supply of agri-food.

Limitations may include that the amounts required to purchase pork meat may have trends with other features (data not shown), which requires further study so that the effects of broadcast news, TV programs/shows and social network on consumption of agri-food could be clearly revealed. Recently, there was an outbreak of African swine fever in the Korean peninsula, which severely affected pork meat consumption and price. For future research, it is important to analyze how an outbreak of infectious diseases among livestock can affect meat consumption and in order to predict demand of agri-food by using unstructured data. Furthermore, prioritizing the impacts of various unstructured data on consumer demand is to be carried out so that these findings can be provided to policy makers to facilitate consumption of agri-food when over production occurs. These topics are lacking in our current study, and they constitute areas of future research.

References

1. Agriculture and Consumer Protection Department. ASF Situation in Asia Update. Available online: http://www.fao.org/ag/againfo/programmes/en/empres/ASF/situation_update.html (accessed on 25 November 2019).

2. Rah, H.; Park, K.; An, B.; Choi, S.; Chae, D.; Yoo, K.H. Development of Prediction Model of Agro-Food Demand by Unstructured and Structured Bigdata. In *The 5th International Conference on Big Data Applications and Services*; Korea Big Data Service Society: Jeju, Korea, 2017; Volume 5, pp. 122–127.

3. Shin, M.-H.; Oh, S.-H.; Hwang, D.-Y.; Seo, S.-S.; Kim, Y.-C. Effect of SNS Characteristics on Consumer Satisfaction and Purchase Intention of Agri-food Contents. *J. Korea Contents Assoc.* **2012**, *12*, 358–367. [CrossRef]

4. Kim, S.H. The Impact of Foot-and-Mouth Disease on Pork Consumption: Analysis of Consumer Response to Media. Master's Thesis, Seoul National University, Seoul, Korea, 2016.

5. Artola, C.; Pinto, F.; de Pedraza García, P. Can internet searches forecast tourism inflows? *Int. J. Manpow.* **2015**, *36*, 103–116. [CrossRef]

6. Choi, H.; Varian, H. Predicting the present with Google Trends. *Econ. Record* **2012**, *88*, 2–9. [CrossRef]

7. Bollen, J.; Mao, H.; Zeng, X. Twitter Mood Predicts the Stock Market. *J. Comput. Sci.* **2010**, *2*, 1–8. [CrossRef]

8. Kurumatani, K. Time series prediction of agricultural products price based on time alignment of recurrent neural networks. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 81–88.

9. Kim, J.; Cha, M.; Lee, J.G. A Model for Nowcasting Commodity Price based on Social Media Data. *J. KIISE* **2017**, *44*, 1258–1268. [CrossRef]

10. Meza, X.V.; Park, H.W. Organic Products in Mexico and South Korea on Twitter. *J. Bus. Ethics* **2016**, *135*, 587–603. [CrossRef]

11. Yoo, D.-I. Vegetable Price Prediction Using Atypical Web-Search Data. In Proceedings of the 2016 Annual Meeting, Boston, MA, USA, 31 July–2 August 2016; Agricultural and Applied Economics Association: Milwaukee, WI, USA, 2016.

12. Rah, H.; Oh, E.; Yoo, D.-I.; Cho, W.-S.; Nasridinov, A.; Park, S.; Cho, Y.; Yoo, K.-H. Prediction of Onion Purchase Using Structured and Unstructured Big Data. *J. Korea Contents Assoc.* **2018**, *18*, 30–37.

13. Statistics Korea. Korean Statistical Information System (KOSIS). Available online: http://kosis.kr/index/index.do (accessed on 2 March 2018).

14. Korea Agro-Fisheries & Food Trade Corporation. Korea Agricultural Marketing Information Service (KAMIS). Available online: https://www.kamis.or.kr/customer/main/main.do (accessed on 2 March 2018).

15. Korea Rural Economic Institute. Outlook and Agricultural Statistics Information System (OASIS). Available online: https://oasis.krei.re.kr/index.do (accessed on 2 March 2018).

16. Bae, S.; Yu, J. Predicting the real estate price index using machine learning methods and time series analysis model. *Hous. Stud. Rev.* **2018**, *26*, 107–133. [CrossRef]

17. *R Development Core Team R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.

18. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **2002**, *20*, 134–144. [CrossRef]

19. Suharsono, A.; Aziza, A.; Pramesti, W. Comparison of vector autoregressive (VAR) and vector error correction models (VECM) for index of ASEAN stock price. In *AIP Conference Proceedings*; AIP Publishing: College Park, MD, USA, 2017; p. 020032.

20. Fukata, K.; Washio, T.; Yada, K.; Motoda, H. A method to search ARX model orders and its application to sales dynamics analysis. In *Data Mining for Design and Marketing*; Chapman and Hall/CRC: New York, NY, USA, 2009; pp. 90–103.

21. Zhang, J.; Hu, W.; Zhang, X. The relative performance of VAR and VECM model. In Proceedings of the 2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering, Kunming, China, 26–28 November 2010; pp. 132–135.

22. Samuel, A.L. Some Studies in Machine Learning Using the Game of Checkers. *IBM J. Res. Dev.* **1959**, *3*, 210–229. [CrossRef]

23. Oh, I.-S. *Machine Learning*; Hanbit Academy: Seoul, Korea, 2017.

24. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

25. Lantz, B. *Machine Learning with R*; Packt Publishing: Birmingham, UK, 2013; p. 396.

26. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2222–2232. [CrossRef] [PubMed]

27. Lipton, Z.C.; Berkowitz, J.; Elkan, C. A critical review of recurrent neural networks for sequence learning. *arXiv* **2015**, arXiv:1506.00019.

28. Cho, Y.; Oh, E.; Cho, W.-S.; Nasridinov, A.; Yoo, K.-H.; Rah, H. Relations between Paprika Consumption and Unstructured Big Data, and Paprika Consumption Prediction. *Int. J. Contents* **2019**, *15*, 113–119.