

Article

Influence Relationship between Online News Articles and the Consumer Selling Price of Agricultural Products—Focusing on Onions

Jiyoung Ha ¹, Seunghyun Lee ^{1,*} and Sangtae Kim ²

¹ Global & Business Inc., Gwangju 61089, Republic of Korea; namu1223@hanmail.net

² Department of Food Industrial Management, Dongguk National University, Seoul 04620, Republic of Korea; kst0320@dongguk.edu

* Correspondence: shleemis@hanmail.net

Abstract: This study aimed to verify the influence relationship between the news articles on onions produced in Korea and the consumer selling price of onions. The analysis methods were the LDA topic modeling technique and the multiple regression analysis. As a result of the analysis, a total of eight topics were found in onion-related news articles. This study analyzed which articles out of the eight topics affected the consumer selling price of onions. As a result, Topic 1 (hypermarket onion sales-related articles), Topic 5 (onion supply and demand stabilization measures), and Topic 6 (inflation) had a statistically significant influence relationship. These results meant that as the number of hypermarket-related articles increased, the consumer selling price increased, and as the macroeconomic articles such as supply and demand stabilization measures and inflation increased, the selling price decreased. The significance of this study was that it revealed that news articles related to onions did not affect the selling price in the consumer market as a whole, and that only the articles directly related to the consumption market (distributors, macroeconomic indicators, etc.) had an effect.

Keywords: onions; online news article; LDA (Latent Dirichlet Allocation); topic modeling



Citation: Ha, J.; Lee, S.; Kim, S. Influence Relationship between Online News Articles and the Consumer Selling Price of Agricultural Products—Focusing on Onions. *Agriculture* **2023**, *13*, 1707. <https://doi.org/10.3390/agriculture13091707>

Academic Editors: Myung Hwan Na, Wanhyun Cho and In Seop Na

Received: 11 July 2023

Revised: 23 August 2023

Accepted: 25 August 2023

Published: 29 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tomatoes are the most consumed vegetable worldwide, followed by onions. Onions are used as an essential ingredient in various dishes. Recently, the prices of vegetables such as onions have soared in many countries around the world and inflation has worsened, threatening the global supply [1]. Onion price fluctuations are affected by a variety of factors, both internal and external. In a report [2] published by the Korea Rural Economic Institute, the factors for the price fluctuations of onions were presented in terms of supply and demand. On the supply side, there were increases in labor costs due to a shortage of agricultural workers, bad weather (rainy season, typhoon, cold wave in winter, etc.), and a temporary decrease in cultivation areas. On the demand side, there were changes in eating habits due to infectious diseases such as COVID-19 (increased consumption of HMR foods and home-cooked meals), an increase in disposable income due to government subsidies, and the price of imported onions. In fact, looking at the Korean news article [3] reported in the first half of the previous year, as demand decreased due to the sluggish restaurant industry caused by COVID-19, onion storage remained. As the onion price fell, it was reported that farmers plowed the onion fields right before harvest and the article predicted a political struggle to go to Seoul (Government). In addition, according to a press release by the National Statistical Office of Korea [4], crop yields were sluggish due to deteriorating weather conditions, resulting in a decrease in the fresh weight per onion. As the production (−24.2%) and cultivated area (−4.3%) decreased from 2021, onion price declines were mentioned. As such, various issues related to agricultural production and

consumption have been reported, and these articles have the potential to affect selling prices in the consumer market. This study aimed to verify the influence of media reports on price fluctuations of agricultural products, focusing on these aspects. The purpose of this study was to confirm the influencing relationship of media reports on specific agricultural products and consumer sales prices. As a specific research method, by extracting the main topics based on the keywords used in the news articles, we attempted to analyze the influencing relationship of the topics that had an influence on the selling price of agricultural products. As a research sample, the news articles written by major Korean media outlets about onions produced in the Republic of Korea were applied to the analysis. The world's onion production totaled 93,226,400 tons, with China's production (23,907,509 tons) being the largest, but the yield per 1 ha was much higher in Korea (65,277 kg) than in China (22,000 kg) [5]. In addition, among the vegetables produced in Korea, onions ranked second with the highest per capita food supply [6]. Onions are used as an important food ingredient in Korean cuisine. The government is closely examining the intensity of the fluctuations in the consumer consumption–purchase psychology depending on the issue at the time since onions are a price-sensitive item. Therefore, research is needed to verify which issues appear at a given time and whether they have an impact on the selling price.

2. Literature Review

The influence of mass media has been researched for a considerable time. Blood and Phillips [7] found that economic news could potentially affect consumer sentiment, and in the long run, it could have an impact on economic conditions. The 'presumed media influence model (PMIM)' by Gunther and Storey can be related to this [8]. In the absence of direct information to measure the effects of media, researchers revealed that people eventually tend to change their attitudes or behaviors by generally regarding the media as a significant influence on public opinion. Tal-Or et al. [9] used the PMIM in the food field to indirectly manipulate the perceptions of the influence of news stories on anticipated sugar shortages. Consequently, they measured the behavioral intentions by manipulating the perceived exposure to the news articles. Attavanich et al. [10] analyzed the market influence of media reports related to swine flu on future prices of pork meat, live cattle, corn, and soybeans. The result was that the media reports had a significant temporary negative influence on hog futures prices, but little influence on other futures prices. Han Mu Moungh Cho et al. [11] analyzed the correlation between the changes in news keywords for avian influenza and changes in egg prices. As a result of analyzing news keywords based on their frequency of appearance, the top keywords for news related to avian influenza were 'egg', 'price', and 'government'. It was found that the price of eggs soared when the keywords appeared. Additionally, as a result of analyzing the relationship between news keywords and egg prices, it was found that the keywords representing specific events such as 'Made in the United States', 'holidays', and 'prices' had an impact on price fluctuations rather than keywords related to avian influenza. In other words, the main issue rather than the frequency of appearance affected the market price.

Based on the media influence theory, Walker [12] found that word of mouth was the final mechanism for opinion change, but that the media led the discourse. He studied the relationship between news media and the skyrocketing house prices in the UK. In the analysis of more than 30,000 news articles on the UK housing market between 1993 and 2008, it was found that the media caused real house price fluctuations, revealing that the media influenced opinions about the housing market. Park Jae-soo et al. [13] extracted positive and negative emotional indices related to apartment sales in Seoul of the Republic of Korea by using topic modeling techniques from online newspaper articles. Through a cross-correlation analysis and causal analysis, the influence relationship between the positive and negative emotional indexes and apartment sales price was analyzed. In the analysis results, it was found that the positive emotional index had a significant effect on the sales price of small apartments in Seoul, except for the downtown area. Seo Jeong-seok et al. [14] established a media report tone index for this by developing a positive/negative

emotional dictionary for article titles based on the combination of words that frequently appeared in newspaper articles related to the housing market to analyze the dynamic relationship between news on the housing market and the housing market. Based on this, the relationship between the real estate consumption sentiment, real transaction price index, and trading volume of apartments was analyzed. In the analysis results, the tone of the media reports on the housing market was found to have a positive effect on apartment prices. In addition, Atri et al. [15] analyzed the influence of COVID-19 news, panic, and media reports on oil and gold prices. The results showed that COVID-19 deaths and panic negatively affected crude oil prices.

By comprehensively considering the above previous studies, it can be seen that media reports can affect market prices. Recent studies analyzing the relationship between media reports and market prices have been conducted by combining unstructured and structured data to complement the traditional analysis method and expand the research area. Therefore, it was necessary to conduct research into various aspects using big data in the agricultural field as well. The differences of this study are as follows. In the previous studies, the frequency of the appearances in press reports, the emotional index of positive and negative emotions, and the influencing relationship between the price and article tone (positive–negative) were examined. However, in this study, while several issues about onion items were articulated, we focused on which issues affect the selling price in the consumer market.

3. Materials and Methods

The analysis process of this study proceeded in the order of collecting online news articles, conducting data preprocessing and a major issue analysis, collecting time series data, and performing an influence relationship analysis. First, in the stage for collecting online news articles, the data was collected using the NetMiner 4.3 News Data Collector from Cyram Co., Ltd. (Seongnam, Gyeonggi, the Republic of Korea). The collection period was a total of 12 months from January to December 2022, and the unit of analysis was one day. This was used to reflect the fact that online news articles report various contents by date. According to the 2022 Korea Press Yearbook [16] of the Korea Press Foundation, the rate of newspaper articles being used by Koreans was 25.1% for PC internet and 76.8% for mobile internet. The rate of using TV news was 76.8%, the daily average usage time of TV news was 41.5%, and the internet portal news usage rate was 75.1% for mobile and PC internet. Korea and Japan ranked first with 69% for the main path of digital news channels in search engines and news collection services.

The media that were collected included online Korean news articles from major domestic daily newspapers, i.e., Donga Ilbo, Chosun Ilbo, JoongAng Ilbo, Kyunghyang Shinmun, Hankyoreh, and Hankook Ilbo, and terrestrial broadcasters, i.e., KBS, MBC, and SBS. According to the results of a survey on the most trusted media outlets by experts (500 people) and the general public (500 people) in the Sisa Journal [17], a Korean current affairs weekly, it was found that KBS ranked first (27.2%) and MBC ranked third (21.6%) in the case of experts, and that KBS ranked first (36.2%), MBC ranked second (30.6%), and SBS ranked third (24.2%) in the case of the general public. In the case of newspapers, Hankyoreh was ranked first (14.6%) among experts and Chosun Ilbo was ranked first (18%) among the general public, indicating that it was the most trusted media.

The collection keyword was ‘onion’ and the total number of collections was 2162. The articles containing words such as ‘onion’, ‘onion crispies’, ‘onion hydrolysates’, etc., which were not related to the onion consumption market, were refined and the remaining 645 articles were applied to the analysis. The data preprocessing and major issue analysis were performed using the NetMiner 4.0 (Seongnam, Gyeonggi, the Republic of Korea). As a preprocessing task, words (farmhouse, farmers, etc.) with similar meanings were registered as synonyms, words that were frequently used in news articles (subscription, copyright, video editing, reporter, etc.) were registered as negative words, and native languages (Ministry of Agriculture, Food and Rural Affairs, etc.) were registered as designated words

and processed to prevent duplicate counting. Except for ‘onion’, which was a search term and a word used in most documents, the low TF-IDF (term frequency—inverse document frequency) were deleted in order to identify the important words mentioned in several documents. In the main issue analysis, a coherence index was analyzed to secure the predictive validity of the topic model, and based on this, the topic modeling analysis was conducted.

The materials and methods for analyzing the influencing relationship between the extracted topics and the consumer selling prices of onions are as follows. For the onion consumer selling price data, the onion consumer selling price of the SSM (super supermarket) and SM (supermarket) provided by the Agricultural Products Distribution Information Site [18] was collected. The collection period was the same as that of the news articles.

During the process of collecting the data by date, the price of the previous day was applied to public holidays where the selling price was not provided. The collected data was the average of retail prices surveyed from 10 locations in five major metropolitan cities. Individual vendors, origins, brands, specifications, etc. were not reflected. Prior to analyzing the influence relationship, the number of articles per day and selling price were indexed, and the unit root test was conducted to secure the stationarity of the selling price data. Lastly, the multiple linear regression analysis was performed to analyze the influence relationship between the onion-related major topics and the selling price. The analysis tool was SAS 9.4. The data preprocessing and analysis procedures is shown in Table 1.

Table 1. Data preprocessing and analysis procedures.

Step 1: News data collection	<ul style="list-style-type: none"> • Collection tool: NetMiner 4.0 News Data Collector • Collection period: 1 January 2022–31 December 2022 • Media: six major domestic daily newspapers; three terrestrial broadcasters • Search keyword: onions • Number of cases collected: 2162 cases in total
▼	
Step 2: Data pre-processing and major issue analysis	<ul style="list-style-type: none"> • Analysis tool: NetMiner 4.0 • Excluding articles not related to the onion consumption market • Word filtering and dictionary registration (thesaurus, defined, exception) • Importing the final 645 news data into NetMiner • Remove common words with node filtering (TF-IDF) • Extracting meaningful words from the entire data (more than two letters) • Coherence index analysis (hyperparameter (α, β), appropriate number of topics (k)) • Extraction of the major topics using the topic modeling analysis • Changes in the topic weight by period
Step 3: Time series data collection	<ul style="list-style-type: none"> • Collection period: 1 January 2022–31 December 2022 • Data collection: onions distribution information (www.kamis.or.kr, accessed on 1 May 2023) • Collection path: SSM–SM selling price • Collection conditions: all varieties; grade high
Step 4: Influence relationship analysis	<ul style="list-style-type: none"> • Analysis tool: SAS 9.4 • Indexing and unit root test for the number of articles and selling price by topic • Multiple linear regression analysis

3.1. Topic Modeling

Topic modeling was first presented by Blei, Ng, and Jordan. This analysis method is one of the text mining analysis methods and it extracts potential topics by classifying the data based on a probabilistic model in a wide range of documents. Text mining refers to a method of identifying hidden patterns or meaningful knowledge by analyzing unstructured text data composed of natural language [19]. It is used in research such as drawing topics from document groups, analyzing trends, and creating knowledge maps. For an analysis to extract the topics from the text data, the topic modeling analysis based on the LDA (Latent

Dirichlet Allocation) is most often used since the LDA is most successfully proven when applied to the text data composed of long sentences. The LDA is a probabilistic model that describes which topics exist in a given document. The LDA classifies a document as a set of topics, and by analyzing the distribution of the number of words found in a given document, it can determine which topics the document covers. Additionally, since a document contains various words, it does not belong to just one topic. Instead, it can belong to many other topics. In contrast, the LDA can identify the main issue in an entire document by analyzing the words in the topic and the influence of each topic [20].

The document creation structure for LDA-based topic modeling is shown in Figure 1. M is the number of documents and N is the number of words in a particular document. K refers to the number of designated topics, and α and β are the hyperparameters. The colored nodes in the square are the observed parameters and the uncolored nodes are the latent variables. The nodes outside the squares represent the observed parameters. First, given the lexical distribution α of the topics following the Dirichlet distribution and the topic distribution β of the documents, a topic for each document is stochastically selected under the influence of α to extract the topic distribution (θ) of the document, and the word distribution (ϕ) of the topic is extracted by probabilistically selecting the words that exist in the extracted topic under the influence of β . The Bayesian inference is used to infer θ and ϕ to minimize the difference between the topic modeling result and the original document [21].

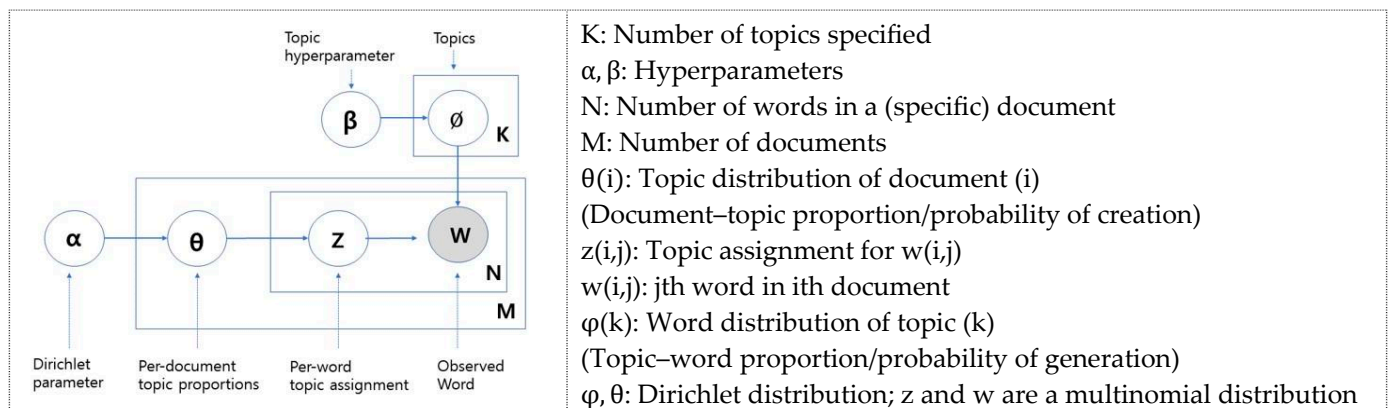


Figure 1. LDA model.

On the other hand, in the topic analysis, the researcher must set the K -value of the total number of topics and the hyperparameter values. The perplexity is used to evaluate how well a particular probability model predicts the actual observations. However, Chang et al. [22] identified that the perplexity did not always provide appropriate results, and Newman et al. [23] said that the number of topics could be determined by considering the consistency of the extracted topics and the possibility of interpretation. While there is ongoing discussion in the academic world about the appropriate number of topics, Cyram Co., Ltd. Developed a coherence index as an evaluation method for the topic models. The coherence index is an indicator that can determine whether the topic contents are consistent by classifying semantically similar words within the topic. As for the measurement method, it selects the top N words for each topic and calculates the association of those words (similarity, frequency of occurrence). The evaluation criteria are based on the c_v and u_{mass} values, and the closer the c_v value is to one and the closer u_{mass} is to zero, the more semantically consistent the topic [21]. Through the index, the appropriate number of topics and hyperparameter (α, β) values can be recommended, and by applying it, a topic model can be extracted to secure the validity of the prediction model.

3.2. Stochastic Time Series Analysis

This chapter was prepared for this study by referring to the book “Analysis of time series data using SAS and R” by Lee Seong-deok (2017) [24].

Time series data starts with the assumption of stationarity, which assumes that the stochastic nature of the time series model does not change with time. The first condition of normal time series data is that it should not show a pattern such as trend, seasonal, or cyclical irregular. Second, the range of change in the data must be constant. Third, there should be no intervals showing different autocorrelation patterns over time. Regarding the first condition, the trendiness is the variation in the case where the observed value has a continuously increasing or decreasing trend, and the seasonality is the variation caused by a periodic component, whether weekly, monthly, or seasonal. The cyclicity exhibits periodic changes. That is due to fluctuations when the cycle is long, such as the business cycle. The irregularity exhibits fluctuations caused by random causes rather than regular movements over time. If the time series is unstable like this, a unit root exists in the variable. In this way, the result of the regression analysis is that the data are not related to each other or are significantly related to each other, and it is called the problem of spurious regression. Therefore, it is important to determine whether the analyzed time series has a unit root. The unit root test is used to test whether a stochastic trend exists in the time series, and generally, the Dickey-Fuller (DF) test is applied. The DF test determines whether a given time series y_t is non-stationary and tests whether the estimate of the coefficient obtained by differentiating the given time series y_t and regressing it to y_{t-1} is zero or less than zero. In addition, the DF test can test the null hypothesis considering where the random walk process has drift and where it includes a non-probability trend. If the time series data follow the auto regressive (p) process, the ADF (Augmented Dickey-Fuller) test is performed. The ADF test is used to verify whether a model without an intercept (when $\Phi_0 = 0$, that is, $Z_t = \varphi_1 Z_{t-1} + \varepsilon_t$), a model with an intercept ($Z_t = \varphi_0 + \varphi_1 Z_{t-1} + \varepsilon_t$), and a model with an intercept and temporal trend ($Z_t = \varphi_0 + \beta_t + \varphi_1 Z_{t-1} + \varepsilon_t$) is $\varphi_1 = 1$. In general, when $\delta = 1 - \varphi_1$, $\psi_1 = -\varphi_{i+1}$. For the $AR(p)$ model without an intercept, $Z_t = \varphi_1 Z_{t-1} + \dots + \varphi_p Z_{t-p} + \varepsilon_t$ and the difference model is Equation (1). In the case of an $AR(p)$ model with an intercept, $Z_t = \varphi_0 + \varphi_1 Z_{t-1} + \dots + \varphi_p Z_{t-p} + \varepsilon_t$ and the difference model is Equation (2). An absolute first-order trend for the intercept and time can be expressed as Equation (3).

$$\nabla Z_t = \delta Z_{t-1} + \psi_1 \nabla Z_{t-1} + \dots + \psi_{p-1} \nabla Z_{t-p+1} + \varepsilon_t \quad (1)$$

$$\nabla Z_t = \varphi_0 + \delta Z_{t-1} + \psi_1 \nabla Z_{t-1} + \dots + \psi_{p-1} \nabla Z_{t-p+1} + \varepsilon_t \quad (2)$$

$$\nabla Z_t = \varphi_0 + \beta_t + \delta Z_{t-1} + \psi_1 \nabla Z_{t-1} + \dots + \psi_{p-1} \nabla Z_{t-p+1} + \varepsilon_t \quad (3)$$

4. Analysis and Results

4.1. Results of Topic Modeling

Prior to the topic extraction for the onion news articles, a coherence index analysis was performed to secure the validity of the prediction. When studying the standard values of c_v and u_mass , the appropriate number of topics was found to be eight, and the hyperparameter values were alpha 0.02 and beta 0.02. The results of the topic coherence is shown in Table 2 (Index: 16, Topic(k): 8, alpha: 0.02, beta: 0.02, u_mass : −1.18, c_v : 0.74). In the broken line shown in Figure 2, the dark orange value Topic in the middle maintained consistency without significant fluctuations.

Table 2. Results of the topic coherence.

Index	Topic (k)	Alpha	Beta	u _{mass}	c _v
1	5	0.01	0.01	−2.988	0.627
2	5	0.01	0.02	−1.87	0.641
3	5	0.02	0.01	−3.039	0.634
4	5	0.02	0.02	−2.758	0.583
5	6	0.01	0.01	−2.466	0.646
6	6	0.01	0.02	−2.08	0.659
7	6	0.02	0.01	−2.488	0.63
8	6	0.02	0.02	−2.76	0.611
9	7	0.01	0.01	−1.554	0.676
10	7	0.01	0.02	−1.956	0.639
11	7	0.02	0.01	−2.35	0.65
12	7	0.02	0.02	−2.626	0.645
13	8	0.01	0.01	−3.288	0.605
14	8	0.01	0.02	−1.496	0.712
15	8	0.02	0.01	−2.004	0.64
16	8	0.02	0.02	−1.18	0.74
17	9	0.01	0.01	−2.33	0.657
18	9	0.01	0.02	−3.268	0.613
19	9	0.02	0.01	−1.841	0.689
20	9	0.02	0.02	−2.776	0.613
21	10	0.01	0.01	−3.05	0.61
22	10	0.01	0.02	−1.833	0.645
23	10	0.02	0.01	−2.565	0.675
24	10	0.02	0.02	−2.941	0.594

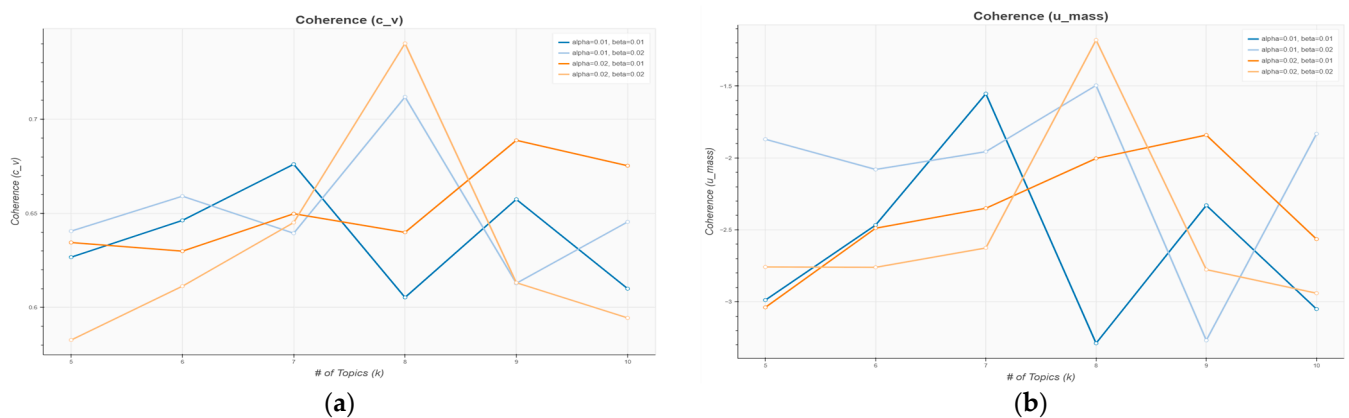
**Figure 2.** Topic coherence graph: (a) c_v ; (b) u_{mass} .

Table 3 shows the results from analyzing which topics were covered in news articles related to ‘onion’ in 2022 were divided. Among the total of eight topics, Topic 6 (inflation) was the top topic followed by Topic 5 (onion supply and demand stabilization measures) and Topic 4 (drought damage).

Table 3. Topic of an onion news article.

Topic	1st~10th Keywords	Topic Naming	Document
1	Sale, market, mart, merchandise, hypermarket, plant, packaging, mind, event, piece	Hypermarket onion selling individually	63
2	Ingredients, cooking, food, salmon, meat, recipe, vegetables, sauce, health, salt	Onion cooking recipe	67
3	Support, onions, agriculture, farmhouse, eco-friendly, business, rural, manpower, product, region	Support for eco-friendly growers	79
4	Drought, damage, area, precipitation, reservoir, farmhouse, garlic, average year, potato, Jeonnam	Drought damage	102
5	Price, farmhouse, cabbage, shipment, market, kimchi, cultivation, government, countermeasures, supply and demand	Onion supply and demand stabilization measures	108
6	Price, prices, Korean cabbage, Chuseok, rise, government, onions, vegetables, imports, kimchi	Inflation	140
7	Agriculture, smart, farmhouse, harvest, garlic, cultivar, potato, technology, machinery, crops	Open-field smart farm	35
8	Rise, inflation, climate change, price, consumer, index, oil, Ukraine, economy, crisis	Climate change food problem	51

4.2. Proportion of Topics by Period

The results of the Topic proportion by period is shown in Table 4. In January, the proportion of articles on Topic 1 (hypermarket selling onions individually, 21.7%) and Topic 2 (onion cooking recipe, 21.7%) was high. In February, articles on Topic 4 (drought damage, 33.9%) appeared for the first time, accounting for the largest portion of the month. In March, Topic 5 (onion supply and demand stabilization measures, 48.2%) had the highest share, and in April, there were many articles on Topic 2 (onion cooking recipe, 23.3%) and Topic 5 (onion supply and demand stabilization measures, 23.3%). Subsequently, in May, there were many articles on Topic 2 (onion cooking recipe, 17.4%). In June, there were many articles on Topic 4 (drought damage, 28.2%) and Topic 6 (inflation, 27.2%). Articles on inflation continued to appear in July (28.6%), August (51.1%), and September (66.1%). In October, articles on Topic 5 (onion supply and demand stabilization measures) and Topic 3 (support for eco-friendly farms) were equally high at 21.4% each. In November, there were many articles on Topic 4 (drought damage, 26.6%) and Topic 5 (onion supply and demand stabilization measures, 28.1%). In December, there were numerous articles on Topic 2 (onion cooking recipe, 25.8%).

The basic statistics of the analysis data are shown in Table 5. The eight topics for the independent variable (onion article) and the dependent variable (onion selling price) were the indexed values. Considering the mean (standard deviation), individual hypermarket onion selling was shown to be 0.430 (0.941), onion cooking recipe was 0.503 (1.091), eco-friendly farmhouse support was 0.352 (0.575), drought damage was 0.475 (1.177), onion supply and demand stabilization measures was 0.346 (0.731), inflation was 0.439 (0.808), open-field smart farm was 0.115 (0.629), climate change food problem was 0.223 (0.864), and the onion selling price of SSM–SM was 0.530 (0.093).

Table 4. Topic proportion by period.

Topic	Topic Title	January	February	March	April	May	June	July	August	September	October	November	December
1	Hypermarket onion selling individually	21.7	16.1	7.1	3.3	10.9	11.7	6.6	6.4	10.7	16.7	7.8	-
2	Onion cooking recipe	21.7	7.1	3.6	23.3	17.4	7.8	4.4	19.1	1.8	14.3	7.8	25.8
3	Support for eco-friendly growers	13.0	10.7	10.7	20.0	15.2	5.8	15.4	12.8	5.4	21.4	12.5	16.1
4	Drought damage	-	33.9	19.6	10.0	15.2	28.2	5.5	2.1	3.6	11.9	26.6	9.7
5	Onion supply and demand stabilization measures	4.3	21.4	48.2	23.3	13.0	10.7	11.0	4.3	3.6	21.4	28.1	9.7
6	Inflation	8.7	1.8	-	3.3	10.9	27.2	28.6	51.1	66.1	11.9	10.9	12.9
7	Open-field smart farm	13.0	8.9	1.8	3.3	6.5	-	11.0	2.1	7.1	2.4	1.6	16.1
8	Climate change food problem	17.4	-	8.9	13.3	10.9	8.7	17.6	2.1	1.8	-	4.7	9.7

Table 5. Basic statistics of the analysis data.

	Variables	Mean	S.D.	Min.	Max.
Independent variable	Hypermarket onion selling individually	0.430	0.941	0.000	5.056
	Onion cooking recipe	0.503	1.091	0.000	5.634
	Eco-friendly farmhouse support	0.352	0.575	0.000	2.864
	Drought damage	0.475	1.177	0.000	9.600
	Onion supply and demand stabilization measures	0.346	0.731	0.000	4.839
	Inflation	0.439	0.808	0.000	5.094
	Open-field smart farm	0.115	0.629	0.000	5.926
	Climate change food problem	0.223	0.864	0.000	7.692
Dependent variable	SSM–SM onion selling price	0.530	0.093	0.391	0.680

4.3. Unit Root Test

In the results from examining the stationarity of the original data, as shown in Figure 3, the pattern of the autocorrelation function (ACF) of the original time series had the highest autocorrelation coefficient at 1 lag. It can be seen that the time difference decreased as time passed. Therefore, a trend clearly existed. However, it was unknown whether this trend simultaneously included a “stochastic trend” and an “absolute trend”. Therefore, the ADF unit root test was performed on the original data. Considering the Tau statistic and significance probability $Pr < \text{Tau}$ from the ADF test results shown in Table 6, the model without an intercept (zero mean), the model with an intercept (single mean), and the model with the intercept and the temporal trend were all larger than the significance level of 5%, so the null hypothesis could not be rejected. Therefore, it was observed that a stochastic trend due to the unit root existed. Accordingly, in the results from performing the first difference to secure the stability of the time series data, it was found that there was no trend and that the stationarity was satisfied. In the results from the repeated ADF test to check the stationarity of the data, for the zero mean, single mean, and trend, the p-value was less than the significance level of 5%, so the null hypothesis could be rejected. Therefore, it could not be said that a stochastic trend existed in the first-difference data.

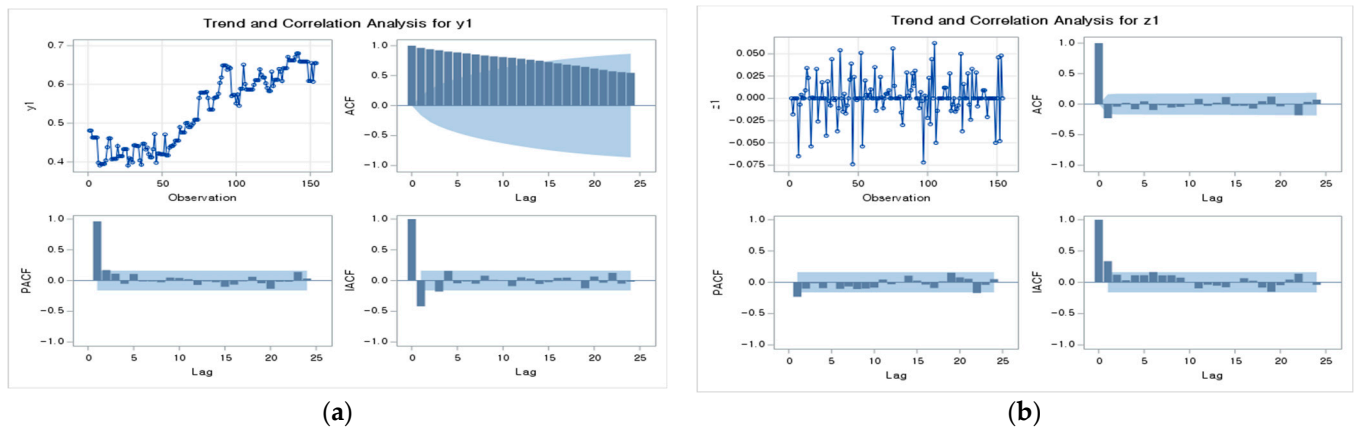


Figure 3. ADF unit root test: (a) original data; (b) first-order difference.

Table 6. Results of the ADF unit root test.

Type	Lags	Original Data		First-Order Difference	
		Tau	Pr < Tau	Tau	Pr < Tau
Zero Mean	0	0.39	0.796	−15.50	<0.000
	1	0.62	0.848	−10.48	<0.000
	2	0.76	0.877	−7.96	<0.000
	3	0.78	0.881	−7.23	<0.000
	4	0.88	0.897	−6.05	<0.000
Single Mean	0	−1.25	0.650	−15.50	<0.000
	1	−0.87	0.795	−10.52	<0.000
	2	−0.78	0.821	−8.01	<0.000
	3	−0.78	0.822	−7.30	<0.000
	4	−0.67	0.849	−6.13	<0.000
Trend	0	−4.45	0.003	−15.48	<0.000
	1	−3.91	0.014	−10.50	<0.000
	2	−3.59	0.034	−8.00	<0.000
	3	−3.65	0.029	−7.30	<0.000
	4	−3.54	0.039	−6.13	<0.000

4.4. Multiple Regression Analysis

The results from analyzing the influence relationship between the number of articles of the major topics by period and the onion selling prices are as follows (Table 7). Topic 1 (hypermarket onion selling individually) had a positive (+) influence relationship on the selling price at the significance level of 0.01 ($t = 2.99$). Topic 5 (onion supply and demand stabilization measures) had a negative (−) influence relationship on the selling price at the significance level of 0.01 ($t = -2.53$), and Topic 6 (inflation) had a negative (−) influence relationship on the selling price at the significance level of 0.05 ($t = -1.97$). The remaining five topics (Topics 2, 3, 4, 7, and 8) were not statistically significant. This meant that as the number of articles related to the hypermarket increased, the selling price of consumers increased, and that the selling price decreased as the supply–demand stability and inflation-related articles increased. In particular, it was observed that the extent of the price decline was more sensitive to the onion supply and demand stabilization measures article than inflation. The Durbin–Watson test, which confirmed the independence of the residuals after the regression analysis, showed that there was no autocorrelation when it was closer to two, which was 2.477 in this study. Finally, studying the intercept revealed that it was not significant. However, the model’s goodness of fit (R-square) and F-value seemed to be rather low since several topics in the article contained various changes, which made it difficult to explain with a single model. In general, it can be seen as a feature that

can appear when a model is fitted using the time series data that includes observations at multiple points in time.

Table 7. Results of the multiple regression analysis.

Model Fit Measures				Overall Model Test		
R ²	Adj-R ²	Durbin–Watson D	Root MSE	F-Value	DF	Pr > F
0.1003	0.0503	2.477	0.02219	2.01	8	0.0497
Variable		Parameter Estimate	Standard Error	t-Value	Pr > t	
Topic 1	Hypermarket selling onions individually	0.00681	0.00228	2.99	0.003 **	
Topic 2	Onion cooking recipe	−0.00036	0.00169	−0.22	0.830	
Topic 3	Support for eco–friendly onion growers	0.00715	0.00369	1.94	0.055	
Topic 4	Drought damage	−0.00171	0.00167	−1.03	0.307	
Topic 5	Onion supply and demand stabilization measures	−0.00660	0.00261	−2.53	0.013 **	
Topic 6	Inflation	−0.00517	0.00263	−1.97	0.051 *	
Topic 7	Open-field smart farm	−0.00203	0.00309	−0.66	0.512	
Topic 8	Climate change food problem	−0.00174	0.00211	−0.83	0.411	

As shown in Table 7, Pr denotes statistical significance. An * is *p*-value < 0.05, ** is *p*-value < 0.01.

5. Conclusions

This study extracted the main topics of news articles about onions produced in Korea and verified the influence relationship between several topics in the articles and consumer selling prices of onion. As a result of the topic analysis of the news articles, a total of eight categories were classified. Topic 1 was for hypermarket selling onions individually, Topic 2 was for onion cooking recipes, Topic 3 was for support for eco-friendly farming families, Topic 4 was for drought damage, Topic 5 was for onion supply and demand stabilization measures, Topic 6 was for inflation, Topic 7 was for open-field smart farms, and Topic 8 was for climate change food problems. In the results from analyzing the influence relationship between the number of articles and the selling prices of the eight topics, Topic 1 (hypermarket onion selling individually) had a positive (+) influence relationship on the selling price at the significance level of 0.01 ($t = 2.99$), Topic 5 (onion supply and demand stabilization measures) had a negative (−) influence relationship on the selling price at the significance level of 0.01 ($t = -2.53$), and Topic 6 (inflation) had a negative (−) influence relationship on the selling price at the significance level of 0.05 ($t = -1.97$). The remaining five topics (Topics 2, 3, 4, 7, and 8) were not found to be statistically significant. As observed from the analysis results, as the number of articles related to the hypermarket in contact with the consumer market increased, the selling price of consumers increased. Additionally, it was observed that the selling price decreased as the number of macroeconomic articles such as the supply and demand stability and inflation increased.

The practical significance of the study results is as follows. First, by extracting the main topics of the onion-related news articles and examining the changes in the articles by period, this study contributed to related institutions and businesses by understanding the flow of media reports related to onions. Second, through the analysis of the influence relationship between the news articles and the consumer selling prices of onions, it was verified that the hypermarket onions selling individually and macroeconomic indicators (onion supply and demand stabilization measures, inflation), which were directly related to the consumption market, were the variables that had a significant impact on the price. These results are meaningful in that they can serve as the basic data that can be utilized by policy institutions for price stabilization in the consumer market.

The academic significance of this study is as follows. First, it was confirmed that the all the news articles related to onions did not affect the selling price in the consumer market. Only certain articles did. In this regard, these result are meaningful in that no

study has been conducted to verify the influence relationship between an article topic and the selling price in the existing studies. Second, the influence relationship was verified by not applying the variables that affected the consumer selling price limited to a specific field, but by applying realistic variables that occurred randomly in the market.

In terms of the limitations of the study, first, the analysis sample was limited to the conditions of Korea. Second, the analysis period was limited to the year 2022. We plan to analyze the changes in article topics by year in the future. Third, the sales price changes in the onion consumption market was examined in relation to news articles only. Therefore, in a future study, we intend to analyze the influence relationship by considering various variables in a complex manner.

Author Contributions: Conceptualization, J.H.; formal analysis, J.H.; funding acquisition, S.L. and S.K.; methodology, J.H.; resources, S.K.; supervision, J.H.; validation, S.L.; visualization, J.H.; writing—original draft, J.H.; writing—review and editing, J.H. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the research fund from the Rural Development Administration (RDA) (RS-2023-00215013).

Institutional Review Board Statement: The study protocol was approved by the RDA (RS-2023-00215013).

Data Availability Statement: When requested, the authors will make available all data used in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yonhap News Agency. Onion, the World's Second-Largest Vegetable Consumption, Is Also a 'Disruption'. *Yonhap News*. 24 February 2023. Available online: <https://www.yna.co.kr/view/AKR20230224125300009?input=1195m> (accessed on 20 August 2023).
2. Kook, S.Y.; Seo, H.S.; Seo, D.J.; Kwon, S.W.; Kim, K.J. The latest agricultural price fluctuations and implications, Korea Rural Economic Research Institute. *Agric. Policy Focus Issue Paper*. **2021**, *10*, 1–58.
3. Hankyoreh Daily. Farmers Who Plowed Their Fields Due to the Plunge in Onion Prices. . . March 10th Announcement of the Struggle to Come to Seoul. *Hankyoreh Daily*. 4 March 2022. Available online: <https://www.hani.co.kr/arti/area/honam/1033565.html> (accessed on 20 August 2023).
4. Statistics Korea. Results of a Survey on Wheat, Garlic, and Onion Production in 2022. *Korea Briefing*. 20 July 2022. Available online: <https://www.korea.kr/briefing/pressReleaseView.do?newsId=156517342#goList> (accessed on 27 July 2023).
5. Onion-Producing Countries Around the World. Available online: www.atlasbig.com (accessed on 20 August 2023).
6. Park, G.S.; Domestic and Foreign Production and Consumption of Onions. Rural Development Administration. 2022. Available online: Nongsaro.go.kr (accessed on 20 August 2023).
7. Blood, D.J.; Phillips, P.C. Economic Headline News on the Agenda: New Approaches to Understanding Causes and Effects. In *Communication and Democracy: Exploring the Intellectual Frontiers in Agenda-Setting Theory*; Maxwell, E.M., Donald, L.S., David, H.W., Mahwah, Eds.; Routledge Taylor & Francis Group: Hoboken, NJ, USA, 1997; pp. 97–113.
8. Gunther, A.C.; Storey, J.D. The influence of presumed influence. *J. Commun.* **2003**, *53*, 199–215. [[CrossRef](#)]
9. Tal-Or, N.; Cohen, J.; Tsifti, Y.; Gunther, A.C. Testing Causal Direction in the Influence of Presumed Media Influence. *Commun. Res.* **2010**, *37*, 1–24. [[CrossRef](#)]
10. Attavanich, W.; McCarl, B.A.; Bessler, D. The Effect of H1N1 (Swine Flu) Media Coverage on Agricultural Commodity Markets. *Appl. Econ. Perspect. Policy* **2011**, *33*, 241–259. [[CrossRef](#)]
11. Han, M.M.C.; Kim, Y.S.; Lee, C.K. Analysis of the Relations between Social Issues and Prices Using Text Mining: Avian Influenza and Egg Prices. *Smart Media J.* **2018**, *7*, 45–51. [[CrossRef](#)]
12. Walker, C.B. Housing booms and media coverage. *Appl. Econ.* **2014**, *46*, 3954–3967. [[CrossRef](#)]
13. Park, J.S.; Lee, J.S. An Investigation into the Causal Relationship and the Cross Correlation between Apartment House Sales Prices and Real Estate Online News: An Approach to the Sentiment Analysis Using Unstructured Big Data of Online News Articles. *JKPA* **2019**, *54*, 131–147. [[CrossRef](#)]
14. Seo, J.S.; Oh, J.H.; Kim, J.S. Exploring the Dynamic Relationship between Real Estate News and Housing Market: Focusing on the Development of Media Tone Index. *KRDA* **2021**, *33*, 89–112.
15. Atri, H.; Kouki, S.; Gallali, M.I. The impact of COVID-19 news, panic and media coverage on the oil and gold prices: An ARDL approach. *Resour. Policy* **2021**, *72*, 1–11. [[CrossRef](#)] [[PubMed](#)]

16. Korea Press Foundation. The Korea Press Yearbook. 2022. Available online: https://www.kpf.or.kr/front/research/yearbookDetail.do?miv_pageNo=&miv_pageSize=&total_cnt=&LISTOP=&mode=W&seq=595594&link_g_topmenu_id=676f2f0f377b4b19840685a46f69a233&link_g_submenu_id=b9afda7cce754998bab61a14ecc139de&link_g_homepage=F®_stadt=®_end (accessed on 20 August 2023).
17. Choi, S.J. KBS Ranked First in Influence and Reliability...Hankyoreh Ranks First in Expert Trust Newspapers. 16 August 2022. Available online: <https://www.hani.co.kr/arti/society/media/1054903.html> (accessed on 20 August 2023).
18. Korea Agro-Fisheries & Food Trade Corporation. At Agricultural Product Distribution Information. Available online: www.Kamis.or.kr (accessed on 30 April 2023).
19. Feldman, R.; Dagan, I. Knowledge discovery in textual databases. In Proceedings of the International Conference in Knowledge Discovery and Data Mining, Montreal, QC, Canada, 20–21 August 1995; pp. 112–117.
20. Cyram Inc. *Text Network Analysis Textbook*; Cyram Inc.: Seoul, Republic of Korea, 2018.
21. Cyram Inc. *Social Media Data Collection and Analysis Textbook*; Cyram Inc.: Seoul, Republic of Korea, 2021.
22. Chang, J.; Gerrish, S.; Wang, C.; Boyd-Graber, J.; Blei, D. Reading tea leaves: How humans interpret topic models. *Adv. NIPS* **2009**, *22*, 288–296.
23. Newman, D.; Lau, J.H.; Grieser, K.; Baldwin, T. Automatic evaluation of topic coherence: In human language technologies. In Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT '10), Los Angeles, CA, USA, 2–4 June 2010; pp. 100–108.
24. Lee, S.D. *Analysis of Time Series Data Using SAS and R*; FreeAcademy: Uiwang, Gyeonggi, Republic of Korea, 2017.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.