

## NORTH KOREA'S ECONOMY: A BIG DATA ANALYSIS

**Namkil Kang**

Professor, College of Liberal Arts, Far East University, South Korea

### ABSTRACT

A specific goal of this paper is to analyze 29 articles obtained in Google in 2024 regarding North Korea's economy. A major point of this paper is that the keyword *Korea* was the most frequently occurred one in 29 articles of Google, followed by the keyword *north*, the keyword *system*, the keyword *plan*, the keyword *planning*, the keyword *year*, and the keyword *economy*, in descending order. A point to note is that in the so-called word cloud, the keyword *Korea* is deemed to be an indispensable one in 29 articles of Google. A further point to note is that the keywords *Kim*, *imbalances*, *committee*, and *executive* showed up in the middle of the map. From this, it seems evident that these four keywords count as the most centralish in 29 articles of Google. Particularly noteworthy is that clusters are different from topics in that the latter are made up of the subject of a document(s), but the former form a cohesive group since they are keywords with similar properties and attributes in a document(s). Finally, when it comes to the sentiment analysis of North Korea's economy, while there are 3 positive keywords with respect to North Korea's economy, there are 11 negative keywords. More specifically, while we obtained -59 with respect to negative keywords, we obtained +13 with respect to positive keywords. From all of this, it seems clear that 29 articles of Google think that North Korea's economy is in bad shape even though there are some positive keywords with respect to it.

**Keywords:** big data, North Korea, economy, word cloud, sentiment analysis, network,

### 1. Introduction

The main goal of this paper is to analyze 29 articles obtained in Google in 2024 regarding North Korea's economy. A specific goal of this paper is to provide the so-called keyword, word cloud, network, and sentiment analyses. First, we aim to contemplate the frequency of keywords that showed up in 29 articles of Google. The so-called keywords are made up of main topics, which in turn form a document. This analysis leads us to see what words consist of main topics, which in turn constitute a document(s). Second, we aim at providing the word cloud with respect to 29 articles of Google. As well-known, python produces the so-called word cloud on the basis of the keyword analysis. Note, however, that this word cloud leads us to see what keywords are the representative ones in 29 articles of Google. Third, we attempt to probe into the core networks of 29 articles of Google. As well-known, in the networks, core keywords turn up in the middle of the map, whereas less pivotal ones show up outside the center, which leads us to capture core words that consist of 29 articles of Google. Fourth, we aim at inquiring into a cluster, which indicates grouping the related words in 29 articles of Google. The relevant words co-occur in a document

or documents. Finally, we try to go over the sentiment analysis in which we classify keywords into positive and negative ones. This analysis leads us to evaluate North Korea's economy.

## 2. Results

### 2.1. Samples

In the following, we aim to look into the samples of 29 articles. As mentioned earlier, we obtained 29 articles of Google regarding North Korea's economy. The following samples are the first sentence of each article:

**Table 1 First Sentence of Each Article**

<b>Article 1</b>	This article is about the economy of Democrati...
<b>Article 2</b>	Though North Korean imports of goods like tire...
<b>Article 3</b>	In 2023, North Korea intensified its political...
<b>Article 4</b>	North Korea has a command (centralized) econom...
<b>Article 5</b>	North Korea's economy is a centrally planned a...
<b>Article 6</b>	In the mid- to late 1990s, the North Korean ec...
<b>Article 7</b>	The country of North Korea, officially known a...
<b>Article 8</b>	Since North Korean leader Kim Jong Un formally...
<b>Article 9</b>	NORTH KOREA has become increasingly antagonist...
<b>Article 10</b>	North Korea's leader Kim Jong Un made a rare a...
<b>Article 11</b>	Kim Jong Un ended 2023 sounding confident, bot...
<b>Article 12</b>	As Pyongyang suspended freight shipments from ...
<b>Article 13</b>	After years of disruption from United Nations ...
<b>Article 14</b>	North Korea's economy shrank for a third conse...
<b>Article 15</b>	To date, assistance for North Korea's economy ...
<b>Article 16</b>	North Korea has faced enormous challenges in p...
<b>Article 17</b>	Viewed from afar, North Korea may appear bizar...
<b>Article 18</b>	The North Korean economy stagnated following t...
<b>Article 19</b>	The North Korean economy shrank for the third ...
<b>Article 20</b>	In recent weeks, there has been significant di...
<b>Article 21</b>	U.S. policy toward North Korea is in need of a...
<b>Article 22</b>	For most of North Korea's history, internation...
<b>Article 23</b>	North Korea, poor and largely sealed off from ...
<b>Article 24</b>	In North Korea, many enjoy making money for th...
<b>Article 25</b>	Whether North Korea's economy is in a state of...
<b>Article 26</b>	When North Korean leader Kim Jong-un and US Pr...
<b>Article 27</b>	North Korea's economy contracted for the third...
<b>Article 28</b>	This year marks the 100th anniversary of Kim I...

<b>Article 29</b>	North Korea, the world's most isolated state, ...
-------------------	---

It is vital that after processing the original texts in Table 1, we obtained only nouns for five analyses, as mentioned earlier. Put differently, we eliminated verbs, adverbs, adjectives, relative pronouns, to-infinitives, etc.

## 2.2. The Keyword Analysis

In what follows, we aim at going over 40 keywords that mainly occurred in 29 articles of Google. These 40 keywords are the main ones that showed up frequently in 29 articles of Google:

**Table 2 Keywords**

Number	Keyword	Frequency	Number	Keyword	Frequency
1	Korea	59	21	gnp	11
2	north	43	22	work	11
3	system	34	23	committee	10
4	plan	32	24	end	10
5	planning	26	25	gdp	10
6	year	23	26	Kim	10
7	economy	21	27	movement	10
8	production	21	28	output	10
9	state	19	29	industry	9
10	factory	17	30	method	9
11	government	17	31	budget	8
12	committee	16	32	capita	8
13	management	16	33	goals	8
14	south	16	34	goods	8
15	party	15	35	units	8
16	workers	15	36	consumer	7
17	people	14	37	country	7
18	enterprise	13	38	debt	7
19	growth	13	39	enterprises	7
20	targets	13	40	farm	7

What is interesting is that the word *Korea* obtained the highest frequency among keywords. The reason why this took place is that we obtained keywords related to North Korea's economy. It is noticing that the word *Korea* is followed by the word *north*. The keyword *north* obtained 43 tokens (the second highest). It is interesting to point out that the keyword *system* obtained 34 tokens (the third highest). It seems to me that the reason why it happened is that North Korea's economy is a planned economy system. That may be why this keyword obtained 34 tokens. It must be stressed that the keyword *plan* is the fourth highest in its frequency. Similarly, the keyword *planning* was the fifth most occurred one (26 tokens) among major keywords. Particularly noteworthy is that the





occurred around the center of the map. From this, it seems clear that these keywords are the core ones, but not so important as the keywords *Kim*, *imbalances*, *committee*, and *executive*. It is interesting to observe that the keywords *collapse*, *administration*, *gdp*, *control*, *errors*, etc. showed up outside the center of the map. This seems to suggest that these keywords are less pivotal and important in 29 articles of Google. We thus conclude that the keywords *Kim*, *imbalances*, *committee*, and *executive* count as the most centralish in 29 articles of Google.

### 2.5. Cluster

This section is focused on looking into the so-called cluster that refers to a group. We obtained a large group in which the relevant keywords occur. Topics are larger than keywords and a document(s) is larger than topics. Maybe the so-called cluster is larger than topics. In what follows, we aim at providing a cluster in which many keywords occurred:

**Table 3 Cluster**

<p>academy accordance account accounting accounts act action activities activity actual addition address adherence administration administration21 adoption affairs aftermath agricultural agriculture aid airports allocation amalgamate amounts analysis analyst anniversary application approach approaches approval areas armstrong army article asia aspects assessment assistants attempts attention august auspices autarky authorities authority autonomy average balance bank base basis battles beginning belief benefactors benefits benefits21 bias21 billion45 bloc bonuses bonuses48 branch brigades budget buffer buildings burden bureaucratism cadres calendar call campaign campaigns capacities capita capital care causes cells cement central century ch changes channel charge charles chido chief china china41 citations city claims co collapse colonial commission committee committee21 committees comparison competition concern confidence confirmation conflict consensus construction consumer consumers contain contract contraction contrast control cooperation cooperative cooperatives coordinate coordination correct costs council countries country county coverage cpc creation creditors culture21 currency czech data data27 day deal debt decade decade21 december decision decisions decline deductions default defense deficit degree demise departments departure deputies destruction detail detailed development deviation difficulties direction directive directives director disadvantages disasters discretion disputes21 disruptions diversion division dollars door dr draft drafts drought east economic economist economy economy21 education effect efficiency effort efforts electricity emphasis emulation enactment131415161718 encouragement end endorsement engineer engineers enterprise enterprises entirety21 entrepreneurship environment equipment errors establishment establishments estimate estimates estimating evaluation example exceeds exchange exchanges executive exhibition exhortations existence exodus expansion expedient expenditure expenditures expenses experimental experimentation expertise export expression extent facilities fact factor factories factors factory factory21 failure farm farmers farms february fees fellow festival field figures finance first five food force formalism formulation forward</p>
--

frequent functionaries funds funds21 gap garden gdp generation gnp goals goals57 goods goods33 government government21 grain grants great grievances group groups growth guidance guidance21 half handicrafts health high highways historian historical history hospitals hotels hungary ideas ideology imbalances impact implementation importance improvement incentives income increase increasing1112 indication indications individual individuals industries industries21 industry inefficient influenced information information21 infrastructure innovation input inspection inspections institute institutionalization institutions instrument interest interests introduced introduction investment iran46 irrigation issues items jangmadang japan japanese jong journals47 juche july june kim korea korean koreans koreas kwp kyōngsŏn lack lands lankov large largely law leadership leadership21 leap leaving legislation level level21 levels30 liberalisation liberalization liberalization21 light line lines living loans logistics long longevity loyalty machinery magnitude man management manager managers manner mao march margin market market49 marketing markets marxist mass masters material materials means measure measures members membership metals method method21 methods micro mid military mimura minimum ministry mitsuhito modernization month months move movement moves multi need needs neighbor new ni northeast number numbers objective objectives observers october officer offices officials oil onsŏng operating operation operations opportunities order organization organizations output overestimation pace parallel part participation partner partners parts party payroll penalties peninsula peninsula39 people20 per percentage performance period period4243 periods personnel place places planners planning plans plant plants plots21 pluralism policy politico population portion post practice practices precision preparation pressure price prices pricing principles priorities problem problems proceeds producer product production productivity products profit profits profits21 programs progress projects proportion province provinces purchase purchases pyongyang quality quarter quarterly quarters railroad range ranks rate readiness reduction refinement reform reforms regime regions relations reliance reliant reluctance remainder remnants reorganization report reported reports republic republics research resources responsibility result return revenue revision rewards rise roads role room rules russia saebunhwa sales sanctions21 sanitation say saŏpch scale schemes schools science sciences second secretary sector sector122526 sector55 sectors sectors57 seeks senior sense services sets seven seventeenth share ship shops short shortage shortages site size slogans socialist society society41 source sources south south47 southward soviet spending spin spot staff stage standard standards station statistical steel storage stores strategy4950 students subteam success sufficiency5152 summaries sung supply support supreme surface surplus system tactics taean taeanŭi target targets task taxes teams technicians techniques technology television ten tendency tenets term termination th theories third thirteenth three thrust thus time times titles today tongnip tons topic tour tourism trade trading tradition transportation trend trends turn turnover typically un uncertainties unification union unit units up us use user vacations venture ventures visit visits volume wangjaesan war war39 water57 welfare whole work workers workshops world year year31 years youth zedong zones

Note that this cluster was produced by python. These keywords are co-related to one another, which in turn constitute 29 articles of Google. Clusters are different from topics in that the latter are made up of the subject of a document(s). On the other hand, the former form a cohesive group since they are keywords with similar properties and attributes in a document(s). Put differently, they are co-related to one another in a document(s).

## 2.6. The Sentiment Analysis

In the following, we aim to evaluate North Korea's economy on the basis of keywords. We classified each keyword into a positive one and a negative one. Additionally, while we assigned +1 to a positive keyword, we assigned -1 to a negative keyword. Most importantly, we multiplied the frequency of the keyword by +1 or -1:

**Table 4 The Sentiment Analysis**

Number	Keyword	Frequency	Positive or Negative	Score
1	gnp	11	negative	-11
2	gdp	10	negative	-10
3	soviet	7	positive	+7
4	trade	7	negative	-7
5	central	5	negative	-5
6	control	5	negative	-5
7	collapse	4	negative	-4
8	income	4	negative	-4
9	lack	4	negative	-4
10	problems	4	negative	-4
11	China	3	positive	+4
12	modernization	3	negative	-3
13	imbalances	2	negative	-2
14	Russia	2	positive	+2

As exemplified in Table 4, while there are 3 positive keywords, there are 11 negative keywords. To be more specific, while we obtained -59 with respect to negative keywords, we obtained +13 with respect to positive keywords. Most importantly, as shown in Table 4, the total figure is -46 (negative keywords - positive keywords=-46). From all of this, it seems clear that 29 articles of Google think that North Korea's economy is in bad shape even though there are some positive keywords with respect to it. For the analysis of big data and AI, see Kang (Kang 2024a, 2024b, 2024c, 2024d, 2024e, 2024f, 2024g).



### 3. Conclusion

To sum up, we have analyzed 29 articles obtained in Google in 2024 regarding North Korea's economy. In section 2.2, we have maintained that the keyword *Korea* was the most frequently occurred one in 29 articles of Google, followed by the keyword *north*, the keyword *system*, the keyword *plan*, the keyword *planning*, the keyword *year*, and the keyword *economy*, in descending order. In section 2.3, we have contended that in the so-called word cloud, the keyword *Korea* is deemed to be an indispensable one in 29 articles of Google. In section 2.4, we have argued that the keywords *Kim*, *imbalances*, *committee*, and *executive* showed up in the middle of the map. From this, it seems evident that these four keywords count as the most centralish in 29 articles of Google. In section 2.5, we have further argued that clusters are different from topics in that the latter are made up of the subject of a document(s), but the former form a cohesive group since they are keywords with similar properties and attributes in a document(s). Finally, we have shown that while there are 3 positive keywords with respect to North Korea's economy, there are 11 negative keywords. More specifically, while we obtained -59 with respect to negative keywords, we obtained +13 with respect to positive keywords. From all of this, it seems clear that 29 articles of Google think that North Korea's economy is in bad shape even though there are some positive points with respect to it.

### References

- [1] Kang, N (2024a). A Big Data Analysis of a Hot Political Issue. *Studies in Linguistics* 70, 149-165
- [2] Kang, N (2024b). The Ukraine War in BBC News: A Big Data Analysis. *International Journal of Applied Engineering and Technology* 6(1), 1548-1555
- [3] Kang, N (2024c). A Study of the Prediction of the Model Logistic Regression in Machine Learning: Focusing on a Survey. *Semiconductor Optoelectronics* 43(1), 86-94
- [4] Kang, N (2024d). The Spring of Seoul: A Big Data Analysis. *Semiconductor Optoelectronics* 43(1), 264-272
- [5] Kang, N (2024e). A Study of the Prediction of the Three Models Logistic Regression, Gaussian Naive Bayes, and Multi-Layer Perceptron Classifier in Machine Learning. *Semiconductor Optoelectronics* 43(1), 253-263
- [6] Kang, N (2024f). An Analysis of the Prediction Power of Four Models in Machine Learning and Deep Learning: Focusing on the Reviews of the Movie Noryang. *A Jungang Journal of English Language and Literature* 66(1), 101-122

[7] Kang, N(2024g). An Association Analysis of Shopping Items in Machine Learning. Stochastic Modelling and Computational Sciences 4(1), 97-109