ISSN: 2633-4828 Vol. 5 No. 4, December, 2023

*International Journal of Applied Engineering & Technology*

# Predicting the Passage of Bill Using Machine Learning: Big Data Analysis of Factors Influencing the Probablility of Passage of Bill

**Sunchoon Park**

*PhD candidate, Department of Business Administration, Namseoul University Graduate School, Cheonan-si, Chungcheongnam-do, Republic of Korea,*

*psc@ihopper.co.kr*

***Date of Submission: 30th September 2023 Revised: 20th October 2023 Accepted: 05th November 2023***

**How to Cite*: Sunchoon Park, (2023), Predicting the Passage of Bill Using Machine Learning: Big Data Analysis of Factors Influencing the Probability of Passage of Bill, International Journal of Applied Engineering and Technology 5(4), pp. 104-109.***

Abstract: ***In this study, countries that are both presidential and parliamentary democracies tend to have a strong parliamentary authority. This can be seen in the vast number of bills submitted to parliament, particularly in countries such as the United States and South Korea. As the number of legal bills submitted to parliament increases, so too does the associated uncertainty, making it necessary to predict the probability of passage of legal bills in order to alleviate such uncertainty. This study focused on developing a machine learning-based artificial intelligence prediction algorithm for predicting the probability of bill passage and describing a series of procedures and methodologies for performing actual predictions and validating the results. This big data analysis aims to construct a model of the factors that influence the probability of passage of legal bills and collect the relevant data accordingly. Machine learning is then applied to the collected data to develop an algorithm. For this study, we developed a model for predicting the likelihood of a bill passing based on factors that may influence its passage. We constructed a dataset of these influencing factors and trained a machine learning algorithm to predict the probability of a bill's passage. The results were successful, with an accuracy rate of 95% in predicting the bill's passage, as well as its final form. This study demonstrates that using big data and machine learning techniques, it is possible to accurately predict the likelihood of a bill passing in the National Assembly of Korea. However, further research is needed to determine if it is possible to predict the probability of specific provisions within a bill passing using these techniques.***

***Keywords: Predict the probability of bill passage, Machine Learning, Bigdata Analysis, National Assembly of Korea, Apollo System***

## 1. INTRODUCTION

Korea is a parliamentary democracy country where the legislative power of the parliament is the strongest and the legislative activities are very active, along with the United States. When comparing the number of bills submitted to the parliament, the US submits 5,899 bills annually, while Korea submits 6,035 bills annually. Considering that Germany submits 151 bills annually, Switzerland submits 104 bills, Japan submits 202 bills, and France submits 461 bills, it explains that the legislative activities of both countries are noticeably more active compared to other OECD countries (National Assembly Library, 2019). The active legislative activities of the parliament mean that the enactment and revision of laws are frequent. The frequent enactment and revision of laws can also be interpreted as meaning that there is a high degree of legal uncertainty.

The purpose of this paper is to verify whether it is possible to accurately predict the likelihood of a legislative bill being passed using machine learning based on big data. Therefore, this study aims to use the following research questions (RQs):

**RQ1:** What factors influence the likelihood of a legislative bill being passed?

**RQ2:** How should modelling be performed using the factors that affect the likelihood of the passage of bill?

**RQ3:** What are the prediction results of the machine learning algorithm using predictive big data for the likelihood of the passage?

## 2. PRELIMINARY STUDY

A practical study on big data analysis and algorithm development for predicting the likelihood of passage of a legislative bill is being conducted by John Nay, a computer scientist and co-founder of the venture company Skopos Labs, based in Nashville, USA. The Nay research team downloaded data related to the legislative process of 68,863 bills submitted to the legislature between 2001 and 2015 from the public information sharing site 'Gov-Track' to simulate real-time predictions of the bills. Also, the team built the text and context models using bill text and metadata assuming that there are consistent semantic and syntactic rules in bills (John, 2016).

There is a company in the United States called Fiscal Note that has developed an algorithm for predicting the likelihood of a bill passing using big data and machine learning, separate from Nay's research team. Fiscal Note crawls government websites to extract data on factors affecting the passage of bills from over 1.5 million bills active in the US Congress, 50 state legislatures, and 9,000 municipal councils, and uses machine learning to develop its algorithm (Michael, 2018).

In Korea, a company called CG INSIDE has also crawled legislative data from the National Assembly to build big data, and developed a machine learning algorithm to predict the likelihood of a bill passing (Jaeyou, 2023). Unlike Skopos Labs, which utilizes text analysis data of bills, CG INSIDE, similar to FiscalNote, has utilized approximately 660,000 data points on factors affecting the likelihood of a bill passing from 2016 to 2022, and the prediction accuracy is known to be at a level of 95%.

In Korea, research on legislative activities by the National Assembly has been qualitatively and quantitatively analyzed since 2003, thanks to the activation of legislative participation by citizens' interest in the information disclosure system and bills proposed by members(Park Yun Hee et. al., 2013). These studies can be broadly categorized into those that focus on the role of standing committees and those that focus on the role of the plenary session. They can also be divided into studies that focus on factors influencing the passage of legislative bills as the dependent variable (Park Yun Hee et al., 2013) and studies that empirically analyze the time required for the passage of legislative bills (Lee et al., 2017).

## 3. AFFECTING FACTORS MODEL OF BILL PASSAGE

To collect and analyze data that affects the passage of a bill, it is necessary to determine the factors that influence the passage of the bill, and this involves building a model of the factors that affect the passage of the bill. This is based on a deep understanding and research of the entire process, from the introduction to the consideration, review, and passage of the bill in the actual legislature, to create a model of the factors that influence the passage of the bill.

Figure 1 illustrates the factors in the dataset used in this study that affect the passage of a legal bill in the model.

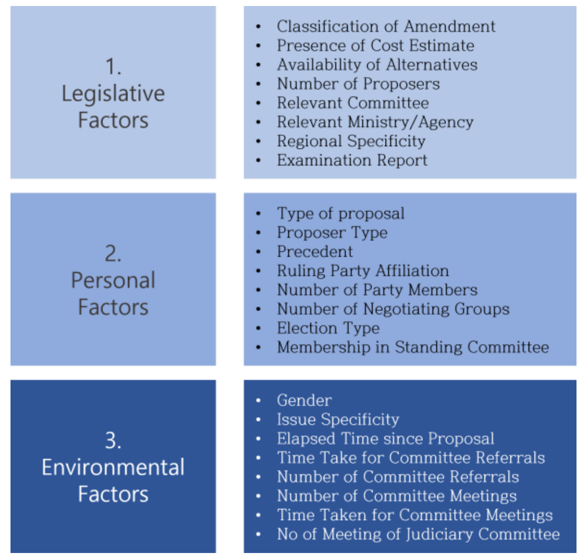



**Fig. 1:** Research Methodology

In this study, we utilized 24 factors influencing the passage of legislative bills in constructing our model. The total 24 influencing factors are classified into three categories: legislative factors, personal factors, and environmental factors.

### 3.1. Data Collection

We mainly collected data through the open information portal of the National Assembly of the Republic of Korea and the public data portal of the Korean government's Open API. Some data was also collected by developing web crawlers. From 2004 to 2020, I collected data on 24 factors included in all bills submitted to the National Assembly of the Republic of Korea that were passed. The collected data on impact factors amounted to about 1.6 million.

### 3.2. Data Coding

The collected data was coded using a binary classification method that distinguishes only between the final passage or failure of the bill in the National Assembly, or a multi-class classification method that categorizes the bills into seven categories, such as passage of the original bill, passage with modifications, rejection, and withdrawal. A separate coding program was developed and utilized for the binary or multi-class classification coding of the data. This coding was performed for the purpose of developing algorithms using the supervised learning approach.

### 3.3. Algorithm Development Environment

This study utilized Python and the artificial intelligence library Tensor Flow to develop a legal passage prediction algorithm. The main programs used were as follows:

- Python v3.7: Development language
- Jupyter Lab v1.0.6: Web-based Python development tool
- Keras v2.2.4: A library developed to allow consistent use of various deep learning libraries such as Tensor Flow.
- Tensor Flow v1.14.0: A deep learning library developed by Google.

### 3.4. Algorithm Development Environment

The legal passage prediction algorithm is developed based on neural network Algorithm. Neurons used in the neural network model the human nervous system. The left picture below shows a human neuron, and the right picture models it.

- Axon: Extending from the body like an arm, it is connected to the dendrites of other neurons.
- Dendrite: It is connected to the axon of another neuron and is attached to the body in a tree-like shape.
- Synapse: The point where the axon and dendrite are connected. Here, signals are transmitted from one neuron to another.

One neuron is connected to several axons of other neurons, and the strength of the connected synapses determines the influence of the connected neurons. When the sum of these influences exceeds a certain value, a signal is generated and transmitted to other neurons through the axon. This corresponds to the modeling in the right picture as follows:

- x0, x1, x2: The amount of signal transmitted from the axon of the input neuron.
- w0, w1, w2: The strength of the synapses, indicating the influence of the input neuron.
- w0x0 + w1x1 + w2*x2: The sum of the amount of input signal and the strength of the corresponding synapse.
- f: The rule that determines the amount of signal transmitted to other neurons based on the final sum, which is called the activation function.

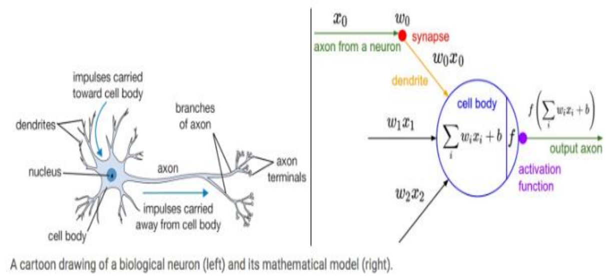



**Fig. 2:** Neural Network Model

### 3.5. Correlation between Data

The impact of 24 data factors on the passage of a bill through the legislature varies significantly, and furthermore, the influence on the likelihood of a bill's passage depends on the interrelationships between these data and the method of combining the data. To analyze the relevance of these data factors, the correlation index between variables was calculated and visualized using a heatmap. In the heatmap, the vertical and horizontal axes represent each variable, and the abbreviations are as follows.

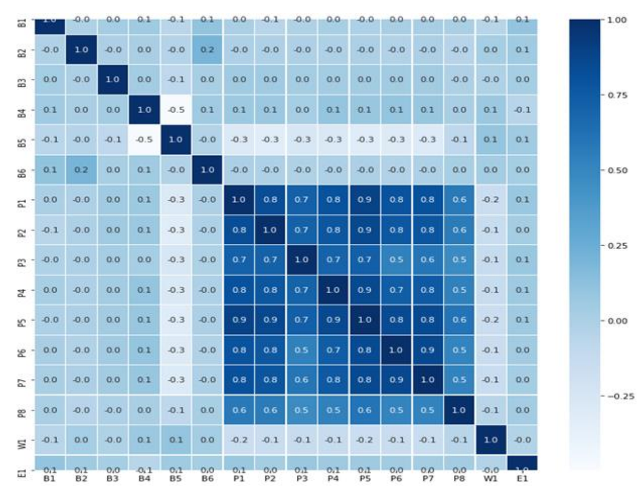


**Fig. 3:** Correlation Heatmap between Variables

## 4. EXPLORATORY DATA ANALYSIS

We extracted some of the data from the 24 factors that influence the passage of a bill, and conducted individual data analysis.

### 4.1. Correlation between Data

We analyzed the data on the number of bills submitted to the Korean parliament compared to the number of bills passed from 2004 to 2020. According to the analysis, there is a trend of increasing bill submissions each year, and as the number of bill submissions increases, there is a slight decrease in the probability of bills being passed. However, the actual probability of bills being passed remained largely unchanged at a level of 30% to 40%, suggesting that while the number of bill submissions may have had an impact on the passage rate, it was not a significant factor.

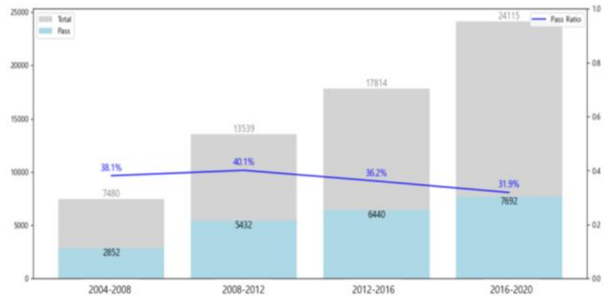
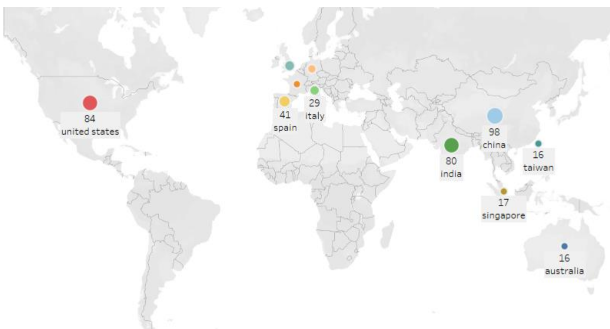



**Fig. 4:** Number of Bill Proposed vs Number of Bills Passed




### 4.2. Number of Co-Proposers

It is possible to hypothesize that a bill with many Co-proposers is more likely to pass through the legislature. However, analyzing all the data on bills from 2004 to 2020 showed that contrary to the general hypothesis, there was no correlation between the number of co-proposers and the likelihood of a bill passing. In fact, there was a tendency for the likelihood of passage to decrease as the number of Co-proposers exceeded 100, as shown in Figure 5.

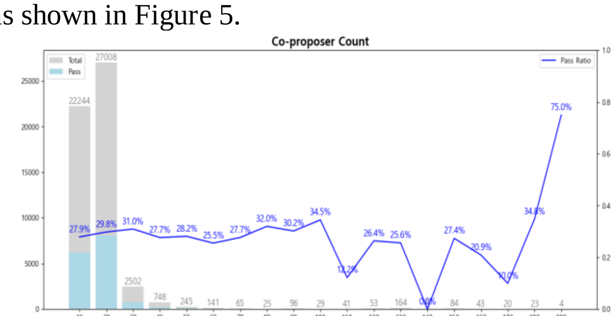



**Fig. 5:** The Probability of bill passing according to the Number of Co-proposers

### 4.3. Proposer Division

There are three proposer divisions for bills: general members of the National Assembly, the Speaker, and the government. Similarly, we analyzed the likelihood of a bill passing according to proposer division for all bills from 2004 to 2020. As shown in Figure 6 below, bills proposed by the Speaker showed a very high passage rate, approaching 100%. Bills submitted by the government also showed a relatively high passage rate of 50.8%, compared to the 29.0% passage rate for bills proposed by general members of the National Assembly.

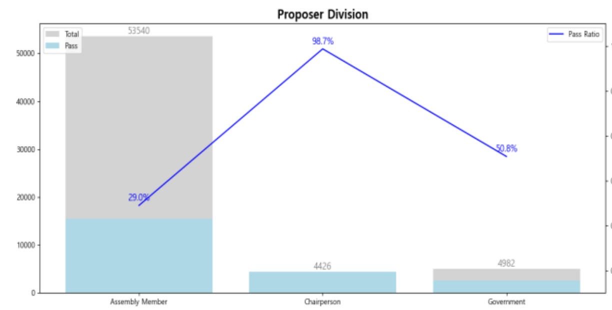



**Fig. 6:** The Probability of bill passing according to the Proposer Division

### 4.4. Number of Term

It is possible to hypothesize that the number of terms of a member of the National Assembly who proposes a bill may also affect the likelihood of the bill's passage. For example, if the number of terms is higher, the member may have greater influence in the legislature, which could have a positive effect on the bill's passage. In fact, we analyzed the probability of a bill's passage according to the number of terms for all bills from 2004 to 2020. According to the results of the data analysis, as shown in Figure 7 below, no correlation was found between the number of terms and the likelihood of a bill's passage.

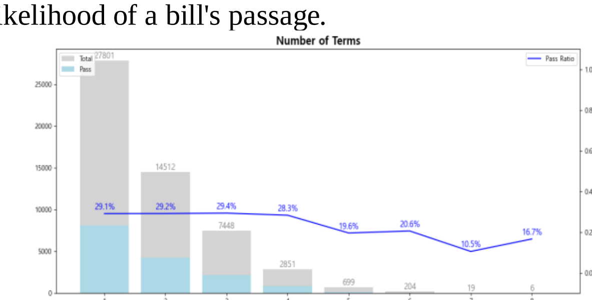



**Fig. 7:** The Probability of bill passing according to the Number of Term

On the contrary, contrary to the general hypothesis, a higher number of terms showed a tendency for a lower probability of a bill's passage. It is possible to infer that this is because a higher number of terms may indicate a relatively weaker enthusiasm for legislative activities, which could have a negative impact on the bill's passage.

## 5. RESULT AND DISCUSSION

To predict the possibility of a bill's passage, categorical data was transformed using one-hot encoding and numerical data was used without scaling. A total of 1,310,688 impact factor data were extracted from 54,612 bills (80% of the total 68,266 bills) and used as training data. The remaining 325,536 impact factor data were extracted from 13,654 bills and used as test data. To prevent over fitting of the prediction model, the Pass/Fail ratio was kept the same in both the training and test datasets.

### 5.1. XGBoost Classifier

Boosting is one of the Ensemble techniques that combines multiple weak Decision Trees. In other words, it creates a strong predictive model by sequentially reflecting the next learning model, weighting the learning errors of weak prediction models. The reason for using XGBoost in this study is that algorithms that usually show high predictability tend to have relatively slow implementation speeds. On the other hand, XGBoost was considered for its advantages of performing cross-validation and having additional features to find important variables. XGBoost (eXtreme Gradient Boosting) used a Machine Learning model based on Gradient Boosting to train the data, and the Hyper Parameters applied to the Machine Learning model are shown in Table 1. The 1,636,224 data on factors that influence the possibility of a bill passing were converted into numeric vectors using the One Hot Encoding method.

| Parameter | Value |
|---|---|
| booster | Gradient Boosting Tree |
| learning_rate | 0.3 |
| n_estimators | 100 |
| max_depth | 6 |
| max_leaves | 0 |
| min_child_weight | 1 |
| num_parallel_tree | 1 |

**Table. 1:** Hyper Parameter

### 5.2. Classification Result

We trained machine learning using 1,310,688 pieces of data on the factors that affect the passage of a bill, and validated the results of predicting the probability of passage with 325,536 pieces of data. The results showed a 95% prediction accuracy, which was similar for both Pass and Fail cases of the bill. However, it is noteworthy that in the Korean legislature, Pass cases are divided into three types: original passage, modified passage, and alternative reflection, and Fail cases are also divided into four types: rejection, disposal, withdrawal, and expiration. Therefore, this study measured the prediction accuracy for only Pass or Fail outcomes, not for the detailed processing types of Pass and Fail mentioned above. The result of the accuracy of predicting the passage of a bill are shown in Table 2 below.

| Label | Precision | Recall | F1-score |
|---|---|---|---|
| Fail | 0.96 | 0.95 | 0.95 |
| Pass | 0.94 | 0.95 | 0.94 |
| Macro avg | 0.95 | 0.95 | 0.95 |
| Weighted avg | 0.95 | 0.95 | 0.95 |
| **Accuracy** | **0.95** | | |

**Table. 2:** Classification Result

The following Table 3 shows the error rate of predictions. The error rate is 5.45% for predicting a bill as Fail when it actually passed, and 4.90% for predicting a bill as Pass when it actually failed.

| | Predict Pass | Predict Fail | Confusion Rate |
|---|---|---|---|
| Real Pass | 6,989 | 403 | 5.45% |
| Real Fail | 307 | 5,955 | 4,90% |

**Table. 3:** Confusion Matrix

We built a model of the factors that affect the likelihood of a bill being passed in the Korean National Assembly, collected data according to this model, and used machine learning with the constructed algorithm to derive a prediction accuracy of 95%. This figure is similar to the prediction accuracy of FiscalNote.

## 6. CONCLUSION AND FUTURE WORK

This study utilized various research techniques, such as formulating research hypotheses, collecting big data, developing algorithms, and utilizing machine learning. According to the results of this study, it was proven that it is possible to predict with 95% accuracy whether a bill submitted to the Korean National Assembly will pass or fail. The high level of accuracy in predicting the probability of a bill's passage was mainly due to the collection of massive data, totaling 1.62 million, based on the factors that influence the passage of a bill. The significance of this study is not limited to simply predicting the likelihood of a bill's passage. It opens the door to more in-depth follow-up research based on the various factors that affect the probability of a bill's passage.

In the future, two follow-up studies will be needed. The first is a study that accurately predicts not only whether a bill will pass or fail, but also which of the three types of pass or four types of fail it will fall under. The second is a study that can predict the probability of passage for each clause within a bill, as bills are composed of many clauses.

## 7. REFERENCES


[1] Anderson, J. E. (1984). Public policy and politics in america. New York: Harcourt Brace.

[2] Arnold, R. D. (2013). Congress, the press, and political accountability. New York: Princeton University Press.

[3] Clinton, J. D. & Enamorado, T. (2014). “The national news media’s effect on congress: How fox news affected elites in congress.” The Journal of Politics Vol.76, No.4. pp.928-943.

[4] Cox, G. W. & Terry, W. C. (2008). “Legislative productivity in the 93d–105th congresses.” Legislative Studies Quarterly Vol.33, No.4. pp.603-618.

[5] Fenno, R. F. (1977). “US house members in their constituencies: An exploration.” American Political Science Review Vol.71, No.3. pp.883-917.

[6] Ferejohn, J. A. (1974). Pork barrel politics: Rivers and harbors legislation, 1947-1968. CA: Stanford University Press.

[7] Galtung, J. & Ruge, M. H. (1965). “The structure of foreign news: The presentation of the congo, cuba and cyprus crises in four norwegian newspapers.” Journal of Peace Research Vol.2, No.1. pp.64-90.

[8] Goodwin, G. (1970). The little legislatures: Committees of congress. Boston: Univ of Massachusetts Press.

[9] Hibbing, J. R. (1991). Congressional careers: Contours of life in the u.s. house of representatives. NC: UNC Press Books.

[10] Iyengar, S. & Kinder, D. R. (1987). News that matter: Television and american opinion. Chicago: The University of Chicago Press.

[11] John J. Nay. (2016). “Predicting and understanding law-making with word vectors and an ensemble mode.” PLoS One. 2017; 12(5): e0176999.

[12] Lowi, T. J. (1964). “American business, public policy, case-studies, and political theory.” World politics Vol.16, No.4. pp.677-715.

[13] _____. (1972). “Four systems of policy, politics, and choice.” Public

[14] administration review Vol.32, No.4. pp.298-310.

[15] Kyle Gulshen., Noah Makow & Pablo Hernandez. (2015). “Predicting BIll Paasage.” Stanford University-CS 229. FINAL PROJECT

[16] Lee, Ji Hyung & Park, Hyung Jun. (2019). “An analysis of factors that affecting the legislature’s bill passing: Focusing on Standing Committees in the 19th National Assembly.” Legislation and Policy, pp. 33 - 63, 2019, National Assembly Legislative Research Service

[17] Mayhew, D. R. (1974). “Congressional elections: The case of the vanishing marginals.” Polity Vol.6, No.3. pp.295-317.

[18] Melenhorst, L. (2015). “The media’s role in lawmaking: A case study analysis.” The International Journal of Press/Politics Vol.20, No.3. pp.297-316.

[19] Wu, J., Smith, A. L., & Zheng, T. (2022). Diagnostics and Visualization of Point Process Models for Event Times on a Social Network. Data Analysis and Applied Stochastic Modelling, 1(1), 11–20.

[20] Mamayusupov, K. (2023). Description of Newton maps of complex exponentials. Journal of Applied Data Analysis and Modern Stochastic Modelling, 1(1), 41–50.

[21] Miquel, G. P. I. & Snyder, J. M. (2006). “Legislative effectiveness and legislative careers.” Legislative Studies Quarterly Vol.31, No.3. pp.347-381.

[22] National Assembly Library, (2019), Statistical Analysis of Bill Submission and Passage Rates in the Legislatures of OECD Member Countries

[23] MICHAEL J. GAYNOR. (2018). “CAN BIG DATA PREDICT WHICH BILLS WILL PASS.” THE WASHINGTON POST. FEBRUARY 1, 2018

[24] Oleszek, W. J. (2007). Congressional procedure and policy. Washington, D.C: Congressional Quaterly Press.

[25] Page, B. I. (1996). “The mass media as political actors.” Political Science and Politics Vol.29, No.1. pp.20-24.

[26] Park Kyung-don. (2017). “The Possibilities and duration of Individual Bill Adoption.” Journal of Legislative Research Vol. 12, No.2. pp. 217-242

[27] Park Yun Hee & MyoungHo Park. (2013). “A Study on the Determinants of Members of Parliament Initiated Bills Passage at the Standing Committee: Focusing on the 18th National Assembly of Korea.” Journal of Legislative Research Vol. 8, No.2. pp. 29-56

[28] Ripley, R. B. & Franklin, G. A. (1982). Bureaucracy and policy implementation. IL: Dorsey Press.

[29] Jaeyou, Shin. (2023). “AI-based government risk management (GRM) platform gaining attention.” Sports Seoul. 2023.2.13.

[30] Tresch, A. (2009). “Politicians in the media: Determinants of legislators' presence and prominence in swiss newspapers.” The International Journal of Press/Politics Vol.14, No.11. pp.67-90.

[31] Wilson, J. Q. (1980). The politics of regulation. New York: Basic Books.

[32] Wilson, R. K. & Young, C. D. (1997). “Cosponsorship in the u.s. congress.” Legislative Studies Quartely Vol.22, No.1. pp.25-43.