

Article

Deriving Mobility Service Policy Issues Based on Text Mining: A Case Study of Gyeonggi Province in South Korea

Younghoon Seo ¹, Donghyun Lim ¹, Woongbee Son ², Yeongmin Kwon ³ , Junghwa Kim ⁴ and Hyungjoo Kim ^{1,*} 

¹ Intelligent Transportation System Lab, Advanced Institute of Convergence Technology, Suwon 16229, Korea; y8seo@snu.ac.kr (Y.S.); ruok54@snu.ac.kr (D.L.)

² Gyeonggi Autonomous Driving Center, Advanced Institute of Convergence Technology, Seongnam 13449, Korea; swb76@snu.ac.kr

³ Policy Research Team, Incheon International Airport Corporation, Incheon 22382, Korea; ymkwon@airport.kr

⁴ Department of Urban and Transportation Engineering, Kyonggi University, Suwon 16227, Korea; junghwa.kim@kyonggi.ac.kr

* Correspondence: hyungjoo@snu.ac.kr

Received: 3 November 2020; Accepted: 12 December 2020; Published: 15 December 2020



Abstract: Mobility services facilitate various tasks related to transportation and passenger movements. Because of the Fourth Industrial Revolution, the importance of mobility services has been recognized by many countries. Thus, research is ongoing to provide more convenience to passengers and to obtain more efficient transportation systems. In the Republic of Korea, the officials of Gyeonggi Province are interested in providing an advanced mobility service to its residents; however, they still do not have any specific or detailed policies. This study aimed at deriving the key issues facing mobility services, especially in the case of Gyeonggi Province, by using a text mining technique and a clustering algorithm. First, a survey was taken by traffic and urban experts to collect reasonable plans for Gyeonggi-Province-type mobility service, and a morpheme analysis was then used for text mining. Second, the results reveal that the term frequency–inverse document frequency (TF-IDF) algorithm has better performance than frequency analysis. Third, the K-means application results in six clusters and six mobility service policy issues were determined by combining the words in each cluster. Finally, the methodology confirmed the validity and effectiveness of the proposed method by showing that the results reflect the current situation in the province.

Keywords: mobility service; text mining; TF-IDF; clustering; Gyeonggi Province; expert survey

1. Introduction

For hundreds of years, transportation has been a major factor in establishing and maintaining social relationships. Additionally, it is a tool that facilitates the purchase of large amounts of daily necessities and lets people commute long distances to their workplaces [1]. Therefore, mobility service can be defined as a service that provides transportation to users and offers both higher convenience and shorter travel times, as it depends on advanced technology. Mobility services somehow depend on buses and subways (low-cost hub-oriented public transportation), but they mainly depend on taxis (high-cost personalized services). The two aforementioned types of mobility services revealed a critical weakness, as both price and convenience cannot simultaneously be achieved. This naturally led to the birth of mobility services in the sharing economy, such as Uber and Lyft, which have become the main keywords since 2010.

The rapid improvement in technology has contributed to the demand for personal mobility means, especially when covering short distances [2–5]. Such means have been used at relatively lower prices compared with taxis since late 2010. Additionally, several researchers have investigated Mobility as a Service (MaaS) [6–8] and other mobility services that depend on autonomous driving technology [9–11], which is one of the essential components in the artificial intelligence industry. By following the trend of the mobility-related technology development up to date, there were two possible expectations: the diversification of the types of mobility and the unification of all the mobility services using one platform. Thus, it is important to construct a mobility service system that can transport people rapidly to their destinations and to establish operational plans that satisfy the sustainability, convenience, and safety factors of mobility services. One of the related studies has shown that safety is the most considerable factor in providing future mobility services [12,13].

With the traffic situation in the Republic of Korea, both the traffic volume and travel time on public transportation tend to be proportional to the number of residents in a city or a province [14]. According to previous reports, the number of passengers in the Seoul metropolitan area in percentage was more than 71% in 2018, which led to the conclusion that the records are greater than twice that of all the other provinces and metropolitan cities [15]. Another aspect is the population movement within the Seoul metropolitan area, as there is an increase in the number of people moving from Seoul to Gyeonggi Province because of the continuous rise in the housing prices in Seoul since 2014 [16]. Therefore, the traffic volume and travel time between Gyeonggi Province and Seoul as well as in Gyeonggi Province is expected to grow based on the above statements. To deal with the potential issue concerning the increase in traffic congestion and the decrease in the convenience of the existing mobility services during rush hours, Gyeonggi Province needs to have a mobility service plan that reflects its current status to provide additional mobility services while maintaining the quality of the existing ones.

Despite the efforts of the officials of the Seoul metropolitan area and those of Gyeonggi Province, with proceeding research on providing future mobility services, traditional types of mobility services still dominate in both regions. Considering the current status of mobility services around the world, both governments need to establish policy issues to provide competitive mobility services as soon as possible. Especially in Gyeonggi Province, the goal of mobility service is to create a Gyeonggi-Province-type mobility ecosystem without traffic congestions and fine dust. It includes (1) the establishment of a systematic foundation for activating mobility, (2) the creation of an environment for the use of tangible mobility, and (3) the introduction of mobility at public buildings such as subway stations. However, detailed policies fitted to the province have still not been formulated, and it cannot achieve the goal without them.

This study aimed to derive a text-mining-based mobility service policy issues for evidence-based bottom-up policy making in a Gyeonggi-Province-type future mobility service to help the government formulate future and competitive mobility service policies. The methodology comprises text mining to extract meaningful information from experts through surveys and clustering algorithms to compare the similarities between words and create clusters with similar words. Since the words in the expert surveys are significantly relevant to the questions and have specific characteristics, detailed plans can be established to provide local customized mobility services if the methodology is applied.

The composition of this study is as follows. Section 2 introduces the existing studies in text mining, mobility service policies, and the status of Gyeonggi Province. Section 3 describes the methodology, and Section 4 verifies the application results and the validity of this research. Finally, Section 5 presents the conclusions of this study and future works.

2. Related Works

2.1. Mobility Service

The research topics on mobility services are divided into two types: mobility service platforms [7,8,17,18] and mobility [9,19–24]. As mentioned earlier, the studies conducted on mobility

have contributed to the invention of various means of transportation (e.g., electric kickboards and electric bicycles), which shorten the necessary duration to arrive at a certain destination. Furthermore, many passengers now no longer have to walk for short distances.

In terms of mobility service platforms, unification is the main topic, meaning that all the means of transportation should be available in one application. Whim in Finland shows how such platforms help users. In the case of the Republic of Korea, Seoul and Gyeonggi Province, which have the most traffic, are in charge of most of the mobility research.

According to several reports written by the Seoul Metropolitan Government, a strategy has been prepared for smart mobility services (e.g., smart parking spaces), where it introduced MaaS and demand–response services in addition to a mobility market environment that can maintain the main public characteristics and build a sustainable smart mobility foundation in connection with urban redevelopment projects [25].

In November 2019, the Gyeonggi provincial government announced a strategy for promoting smart mobility, and it entailed setting up parking spaces for only smart mobility users at public facilities, new towns, and plans to expand smart parking spaces throughout the Gyeonggi Province [26]. However, there is a limitation in the lack of mobility services in the urban or provincial units, and the announced strategy merely focuses on personal mobility.

On the other hand, objects and actions need to be defined in order to create sustainable mobility policy issues. For mobility plans and policies, it is important to set and analyze objectives and actions [27]. In addition, research related to mobility policy needs to define inputs, activities, outputs, outcomes, and goals before formulating policies [28].

Therefore, components for mobility service issues will be required and described in Section 3 to clarify objectives and definitions.

2.2. Text Mining

When data analysis is in progress with unstructured data, text mining algorithms are required to extract meaningful numeric information from a corpus of words before applying machine learning algorithms [29]. These algorithms are essential for conducting natural language processing (NLP) tasks, such as machine translation, question answering, and speech recognition. In addition, text mining algorithms are applied to various fields such as medical field [30,31], education [32], transportation [33,34], politics [35,36], and smart city [34,37], etc. The following paragraphs have brief descriptions of three fundamental text mining algorithms.

One of the simplest ways for transforming text data into numerical data is to utilize the frequency of words. Bag-of-Words (BoW) model creates a list containing specific words based on a certain topic, and the frequency of each word in each document is recorded. These lists can be used to determine the importance of words. For example, words related to mathematics that are gathered into a list may be compared with words related to other subjects.

The document-term matrix (DTM) is a matrix based on the frequency of words in each document. Therefore, each column consists of the frequency of words and each row indicates the number of all words in a document. DTM is a generalized version of BoW and the importance of a word is determined by the frequency as well.

Unlike the above two methods, TF-IDF is able to solve the problem: not reflecting the true word importance. Two indicators in TF-IDF, which determine the weight of the importance of words, can be defined in various ways. DTM is also included in the process of calculating an indicator in TF-IDF, so this study selects TF-IDF for the text mining process to overcome the limitation of using frequency-based methods. Recently, several new algorithms have been proposed [38,39] to deal with the case of multiple documents in multiple objects. Since our data consist of one short document from each expert, TF-IDF, the proven and most frequently used text mining algorithm to extract numeric information, seems to be appropriate.

The further description and explanation of TF-IDF will be in the next section.

2.3. The Status and Major Issues of Mobility in Gyeonggi Province

2.3.1. Description of the Study Site

The Gyeonggi Province has an area of 10,184 km², approximately 10% of the entire country, and it is a part of the Seoul metropolitan area as it surrounds the capital area. Its population is more than 13 million as of 2020, which is the biggest number among Korean provinces and cities. The province comprises 31 local governments, and the population in ten cities has already surpassed 500,000 or even one million, which is the basic condition of becoming a metropolitan city. Figure 1 illustrates the location and the above-mentioned information about the province. The current status and main issues of mobility in Gyeonggi Province are identified through Strengths, Weaknesses, Opportunities, Threats (SWOT) analysis in the following section.

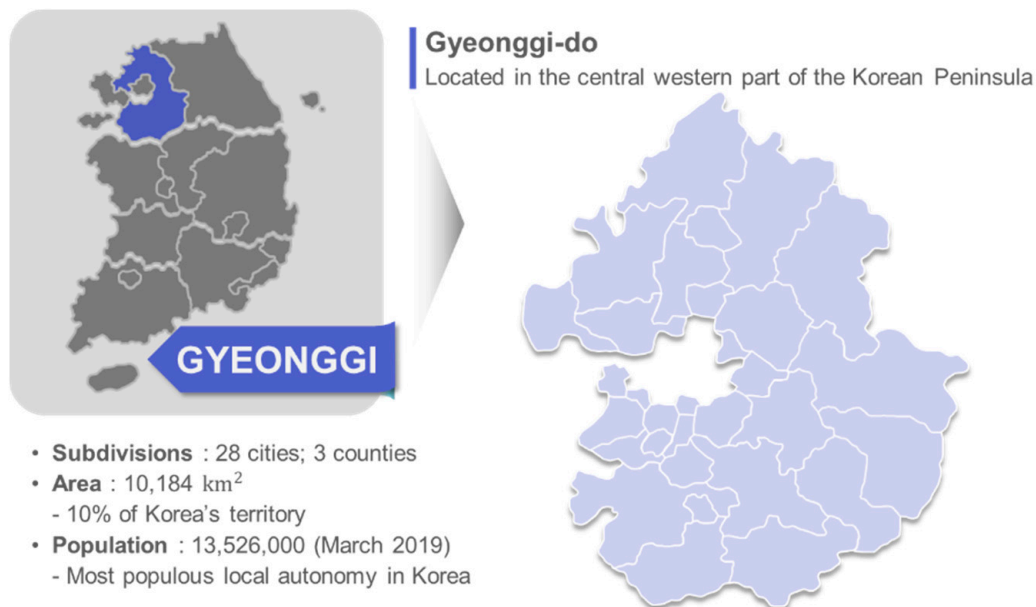


Figure 1. Gyeonggi Province description.

2.3.2. SWOT Analysis on Mobility in Gyeonggi Province

Before applying our proposed methodology, a report examining the status of the mobility services in Gyeonggi Province was conducted by the Gyeonggi Autonomous Driving Center [40]. Several methods were applied to understand the background in the province in terms of planning mobility service policies. In this study, the SWOT analysis appears to be appropriate for verifying the mobility service policy issues using the proposed process. Figure 2 visualizes the results of the SWOT analysis.



Figure 2. SWOT analysis result in Gyeonggi Province.

- Strengths

The Gyeonggi Province has the fifth largest area in the Republic of Korea and is one of its most populous regions. Furthermore, it has high accessibility to Seoul and builds a network environment to discover potential private resources. Additionally, many people prefer to live there because of its convenient conditions, such as housing, industry, and transportation.

- Weaknesses

In Gyeonggi Province, the accessibility limitations due to the geographical breadth, as mentioned in the strength section, were analyzed as weaknesses. More significantly, there is a relatively heavy regional bias in the transportation services among 31 administrative districts (28 cities and 3 counties). Other mobility-related weaknesses include the inability to predict the growing demand for transportation, lack of opportunity to use resources by excessive development restriction zones, etc.

- Opportunities

The population of Gyeonggi Province, which was only 8.98 million in 2000, has been rapidly growing every year, with 11.38 million in 2010 and 13.34 million in 2020. Because of this trend, the first and second new town developments have been successfully completed, and the third one is in progress. The major opportunity factor was determined by raising the necessity of introducing new means of transportation during the construction of the new town development. Other opportunity factors include recognizing the balanced developments in underprivileged communities and supporting the need for introducing eco-friendly transportation.

- Threats

One of the biggest threats, which is related to the external environment in Gyeonggi Province, is the high cost of research and development caused by the continuously soaring land prices. For next-generation mobility systems, such as autonomous driving and cooperative-intelligent transport systems, to be successfully applied, advanced road infrastructure must be established. However, increasing the local government burden and management costs can increase the development costs. Other threats include decreasing the rural population, imperfect the legal systems related to smart mobility, overlapping urban functions, and repeating the implementation of the same projects.

3. Methodology

This chapter explores a research methodology for deriving Gyeonggi-Province-type mobility service policy issues. In this study, topics related to the final mobility service policy were selected through data collection, data preprocessing, and data analysis phases, as shown in Figure 3. The selected topics in conjunction with the SWOT analysis of the current status of Gyeonggi Province were verified as policy measures for the Gyeonggi Province mobility services. In short, our goal is to define customized mobility service policy issues for Gyeonggi Province. Inputs are experts' answers, and activities include text preprocessing, text mining, and clustering algorithms. Outputs are the result from text mining, and outcomes will be the results from K-means. Each component will be described in the next section.

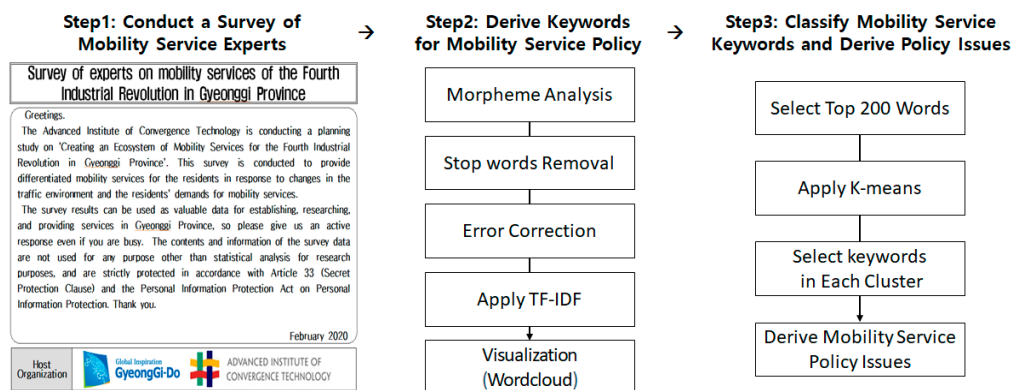


Figure 3. The framework of our proposed methodology.

3.1. Data Collection

The survey data were collected by experts working in public institutions and universities to come up with a plan to establish mobility services in Gyeonggi Province. The criteria for selecting experts were as follows: those who have obtained a doctorate degree in transportation or a related field or those who have conducted research in the mobility field for at least 5 years after obtaining a master's degree. Furthermore, their research area needed to be related to Gyeonggi Province.

The survey consisted of two parts: analytic hierarchy process (AHP) and short essay. Selected experts were required to evaluate factors in mobility service in AHP part (e.g., convenience vs. accessibility, safety vs. eco-friendliness) and write an essay of 500 characters or more about how to develop Gyeonggi-Province-type future mobility services. The survey was conducted and sent via e-mail twice in December 2019 and in February 2020 with a total of 36 experts.

The sociodemographic data of experts are summarized in Table 1. About 88.9% of the respondents were male. The age of the respondents was 41.1 years old on average, and the majority were between thirty-six and forty-five (61.6%). Most of the respondents had doctoral degrees (94.4%), and more than half of the respondents were professors.

Table 1. Descriptive statistics for survey sample.

Sample Attributes	Category	Number (%)
Sample size		36
Gender	Male	32 (88.9)
	Female	4 (11.1)
Age	<36	9 (25.0)
	36–45	22 (61.1)
	45>	5 (13.9)
Degree	Master	2 (5.6)
	Doctor	34 (94.4)
Major	Transportation	28 (77.8)
	Urban	8 (22.2)
Occupation	Professor	20 (55.6)
	Researcher	16 (44.4)

3.2. Text Preprocessing

Text preprocessing is a stage in which the collected data are revised before detecting the numerical patterns. To derive words related to mobility service policies in this study, the answers to the questionnaire must be first separated and transformed from sentences into words. Therefore, the morpheme analysis performed in the preprocessing stage was used to distinguish the words and then correct any grammatical errors, such as removing any special characters and typographical errors to obtain accurate analysis results. The morpheme analysis is a method in NLP, and it simply divides sentences into morphemes, which are the most basic forms of words. A custom dictionary was built for the morpheme analysis to recognize the specific words in the mobility service field. Then, each expert's opinion was divided into morphemes based on the dictionary.

Additional preprocessing work was conducted for the following classifications of morphemes: (I) Before removing the stop words, the unnecessary morphemes (e.g., conjunction and postpositional particle) were defined as stop words in each document. (II) The different morphemes that indicate the same words were unified for counting purposes. (III) The uppercase letters in the words were converted to lowercase letters for the same purpose.

3.3. Frequency Analysis

The frequency analysis is the most representative way to quantify the importance of words through their frequencies. In the field of smart cities, using text mining, a frequency analysis was performed in 116 documents to define the smart city concept, and 40 keywords were chosen to define smart cities. Various visualizations, such as the word cloud, were also conducted for further analysis [31]. In this study, the frequency analysis aimed to compare the top-ranking keywords related to mobility services using the frequency analysis and TF-IDF algorithm.

3.4. TF-IDF

TF-IDF is an algorithm that quantifies how important a word that exists in multiple documents is in each document by using several frequency types. In text mining, it is used to overcome the frequency analysis limitation: not finding out the essential keywords for representing a topic, as the importance of a word is only determined by the number of times it appears in all the documents. However, the TF-IDF penalized frequent words that existed in all the documents because it considered them not as important words in the answers of the survey but just as a part of the words related to the topic. The elements included in the TF-IDF are the term frequency (TF), document frequency (DF), inverse DF (IDF), and TF-IDF term weights. The definition and formula of each element are explained below.

3.4.1. TF

TF means the frequency of a word that exists in one document. To date, various methods of calculating the TF have been presented. Furthermore, a study has previously described other types of the TF, such as the Boolean frequency and augmented frequency [41], to prevent the divergence of weights when the length of a document is too large [42]. However, considering the 36 samples and the length of each document, in this study, the TF was defined as the number of each word in each document. If there are N documents and a function to obtain TF for a word w in document d is $f(w, d)$, the formula for calculating the TF in this study would be as follows:

$$tf(w, d) = f(w, d) \quad (1)$$

3.4.2. DF

DF refers to the number of documents in which a particular word exists, and it can also be represented by the total sum of Boolean frequencies. The DF is mainly utilized to calculate IDF in the TF-IDF. The related formulas of the DF are expressed as follows:

$$df(w, d) = \begin{cases} 1, & \text{if } w \text{ exists in } d \\ 0, & \text{Otherwise} \end{cases} \quad (2)$$

$$df(w) = \sum_{d=1}^N df(w, d) \quad (3)$$

3.4.3. IDF

The most important feature in TF-IDF is to lower the weight by imposing penalties on the words appearing in all the documents. This role is performed using IDF, which is simply the reciprocal of the DF. In some researches, the logarithm of the IDF is applied to its equation [43–45] in the case of divergence of weights based on the number of documents and the possibility of not penalizing words that exist in most of documents.

For the same purpose, the natural logarithm of the IDF has a representation in this study, and one was added to the denominator to prevent the condition from being undefined before evaluating the natural logarithm of the equation [46]. The formula for determining the IDF is as follows:

$$idf(w, d) = \log \frac{N}{1 + df(w)} \quad (4)$$

3.4.4. TF-IDF Term Weights

TF-IDF term weights represent both the importance of a word in total and in each document, and it is calculated by multiplying the TF and IDF. Unlike the BoW and DTM, the TF-IDF has the advantage of not having to consider the group of all the words from all the documents, which enables the quick computation and intuitive calculation of each frequency.

In this study, the 200 highest weights of the words in the TF-IDF were selected to derive the mobility service policy issues through the sum of the TF-IDF weights in each document.

$$tf - idf(w, d) = tf(w, d) * idf(w) \quad (5)$$

$$tf - idf(w) = \sum_{d=1}^N tf - idf(w, d) \quad (6)$$

3.5. Clustering

Clustering is a technique that is widely used in the unsupervised learning of machine learning and is utilized to classify unlabeled datasets. There are three main categories in clustering algorithms: partitioning, hierarchical, and model-based. In this study, K-means in partitioning clustering methods was chosen to determine the associations of the keywords selected using the TF-IDF.

3.5.1. K-Means

K-means is the most representative clustering algorithm that creates groups based on the distance between numerical data. First, the number of groups is specified to be randomly grouped. Then, the center point of each group is randomly designated. Each group calculates the distance between the data coordinates and the centroids. Next, each data belongs to the nearest group. Subsequently, each centroid is recalculated through the average of the

coordinates of all the data in each group, and these steps are repeated until the result is optimized. The detailed description of the K-means is explained below in Algorithm 1:

Algorithm 1: K-means

Input: K : the number of clusters
 $D = \{x_1, \dots, x_N\}$: Dataset, N : the length of Dataset
Output: $C = \{C_1, \dots, C_K\}$: the set of clusters
 $c = \{c_1, \dots, c_K\}$: the set of the center of clusters

- 1 Randomly Initialize c_1, c_2, \dots, c_K ,
- 2 $\forall i, j$ Calculate $d(x_i, c_j)$, where $d(a, b)$ is the distance between a and b
- 3 $\forall i$ Find $j^* = \operatorname{argmin}_j d(x_i, c_j)$
- 4 $\forall i$ Assign x_i to C_{j^*}
- 5 $\forall k$ Relocate c_k ,

$$C_k^{t+1} = \frac{1}{n} \sum_{i=1}^N x_i, \text{ where } x_i \in C_k^t$$

- 6 Repeat Step 2 to 5, until convergence
-

3.5.2. Average Silhouette Method

The assessment metrics for determining the optimal number of groups, K , to be clustered can be mainly divided into internal and external assessments. The internal assessments measure the optimal K with the intra-cluster distances. The Dunn, Davis–Bouldin, and average silhouette method are typical internal clustering evaluation metrics. In this study, the average silhouette method was selected as a better metric than the aforementioned two metrics. The formulas related to the method are as follows:

$$a^{(i)} = \frac{1}{|C_i| - 1} \sum_{i \neq j, j \in C_i} d(i, j) \text{ where } i \in C_i, |C_i| \text{ is the number of data in Cluster } i \quad (7)$$

$$b^{(i)} = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j) \quad (8)$$

$$s^{(i)} = \frac{b^{(i)} - a^{(i)}}{\max(a^{(i)}, b^{(i)})} \quad (9)$$

$$S = \frac{1}{N} \sum_{i=1}^N s^{(i)} \quad (10)$$

Once calculating the silhouette coefficient $s^{(i)}$ for each group with the above equations, the average silhouette score S is determined for each K . Considering all S in each K , the method decides the optimal K where it corresponds to the highest S .

4. Results

4.1. Frequency Analysis Result

The results of the frequency analysis of the preprocessed survey data show that the words related to the survey topic, such as service, mobility, and Gyeonggi Province, are top ranked as described in Figure 4 and Table 2. The highest frequency word was “service” (243 instances) because of the nature of the survey. It seems that “service” appeared in most of the expert answers, and its frequency was about twice the frequency of the second most frequent word. The second most frequent word was “mobility,” which seems to have been written a lot for the same reason as for “service.”

The substantial words for providing mobility services start with the ninth-ranked word, “autonomous driving,” but it was clearly observed that there were insignificant words that cannot help to develop specific policies if the list of words up to the 30th most frequent word was considered.

This indicates that the frequency analysis results have limitations in carrying out this study, as seen in the example of the need for other measures to ensure that essential and specific words for planning and providing mobility services are ranked at the top. Additionally, the frequency of the words was one digit from the top 82nd place, indicating that only certain words monopolized the overall frequency. This can be confirmed through the plot, so other algorithms should complement these characteristics.

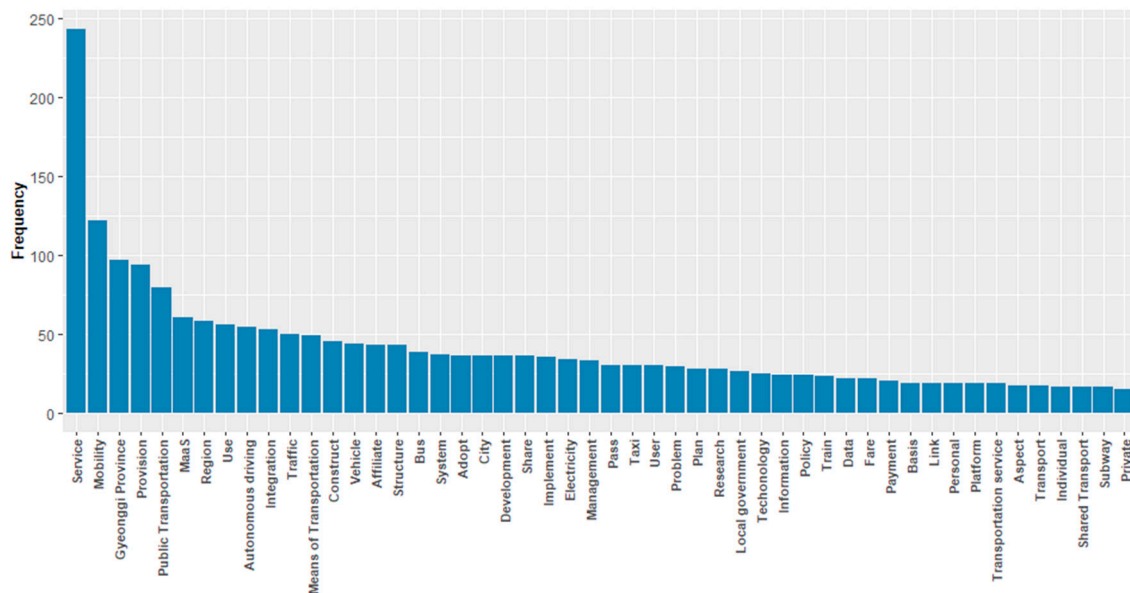


Figure 4. Bar chart for the frequency analysis result on the mobility service survey.

Table 2. Result of the frequency analysis of the mobility service survey.

Name	Frequency	% in Total	Document Frequency
Service	243	4.72%	32
Mobility	122	2.37%	28
Gyeonggi Province	97	1.88%	28
Provision	94	1.82%	26
Public Transportation	79	1.53%	24
MaaS	60	1.16%	19
Region	58	1.13%	23
Use	56	1.09%	23
Autonomous Driving	54	1.05%	19
Integration	53	1.03%	19
Traffic	50	0.97%	20
Means of Transportation	49	0.95%	18
Construct	45	0.87%	19
Vehicle	44	0.85%	14
Affiliate	43	0.83%	21
Structure	43	0.83%	20
Bus	38	0.74%	17

Table 2. Cont.

Name	Frequency	% in Total	Document Frequency
System	37	0.72%	20
Development	36	0.70%	17
Share	36	0.70%	17
City	36	0.70%	13
Adopt	36	0.70%	13
Implement	35	0.68%	18
Electricity	34	0.66%	11
Management	33	0.64%	19
User	30	0.58%	13
Taxi	30	0.58%	11
Pass	30	0.58%	13
Problem	29	0.56%	17
Plan	28	0.54%	12
Research	28	0.54%	13
Local	26	0.50%	12
Government	26	0.50%	12
Technology	25	0.49%	11
Information	24	0.47%	8
Policy	24	0.47%	15

MaaS, Mobility as a Service.

4.2. Result of the TF-IDF

Figure 5 and Table 3 show the results of applying the TF-IDF algorithm to overcome the limitations of the aforementioned frequency analysis. The words at the top of the list in the frequency analysis generally went down. It appears that the word with the highest weight was “vehicle,” as it is the word that could represent all the mobility services. “Electricity” has the second-largest weight, and it is another representative word for moving toward smart mobility services along with the popularization of electric vehicles. If the list of words up to the 10th place was considered, specific words such as “data,” “cities,” “taxis,” “payments,” and “autonomous driving” were distributed at the top to be able to form mobility service policy measures.

Therefore, “service,” the most frequent word in the frequency analysis, was typically ranked 45th because of the weight of the IDF, whereas “mobility” was ranked in 30th and 46th place in Gyeonggi Province. Similar to the frequency analysis, the average of all weights was 6.05, which is somewhat deviant compared with the weight of the top words. The difference between the words of a similar rank was not so significant that the graph in the TF-IDF is flatter than in the frequency analysis. The standard deviation of the TF-IDF algorithm was 5.51, which is half less than the value of the frequency analysis.

One of the advantages of text mining is that various visualizations are possible other than simple graphs. The word cloud is a typical visualization method, and a very effective way to visually compare the size of words in proportion to their weight.

The result of the word cloud represents the top 200 words in terms of the weights of the conducted TF-IDF algorithm in this study, and the words are located from the center to the outside of the cloud in the order of the large weights. The library, “wordcloud2,” in R was used to create a word cloud in this paper, and Figure 6 reveals that the words around the center correspond to specific ways of building mobility services.

Table 3. Result of TF-IDF of mobility service survey.

Name	TF-IDF (w)	Avg.	Std.	Rank in Freq.
Vehicle	38.52	1.07	2.06	14
Electricity	37.35	1.04	1.95	24
Data	36.03	1.00	3.07	37
MaaS	35.27	0.98	1.78	6
City	34.00	0.94	1.79	21
Introduce	34.00	0.94	1.68	22
Information	33.27	0.92	2.19	34
Taxi	32.96	0.92	2.48	27
Payment	32.75	0.91	2.16	39
Autonomous	31.74	0.88	1.10	9
Driving	31.74	0.88	1.10	9
Transportation	31.31	0.87	1.31	12
Integration	31.15	0.87	1.18	10
Link	31.11	0.86	2.20	43
Fare	30.50	0.85	1.83	38
Train	29.46	0.82	2.07	36
Public	28.81	0.80	0.88	5
Transportation	28.81	0.80	0.88	5
Subway	28.67	0.80	2.58	49
Transportation	28.58	0.79	1.79	41
Service	28.58	0.79	1.79	41
Smart	28.56	0.79	3.40	61
Plan	28.52	0.79	1.54	30
User	28.33	0.79	1.47	26
Pass	28.33	0.79	1.36	28
Technology	27.47	0.76	1.44	33
Provision	27.04	0.75	0.79	4
Traffic	26.95	0.75	0.89	11
Ecosystem	26.88	0.75	3.00	51
Local	26.48	0.74	1.90	32
Government	26.48	0.74	1.90	32
Construct	26.45	0.73	1.03	13
Research	26.44	0.73	1.56	31
Mobility	26.38	0.73	0.76	2
Platform	26.34	0.73	2.05	44
Bus	26.34	0.73	1.02	17
Development	24.95	0.69	1.01	19
Share	24.95	0.69	1.03	20
Region	23.52	0.65	0.77	7

4.3. K-Means Application Result

To derive the policy issues, the top 200 words were clustered based on the TF-IDF weight ranking. To proceed with the K-means, the weights of the words in each document were vectorized. The range was set from 2 to 50 to determine the optimal K. After the clustering process, the average silhouette scores were calculated for all the Ks, and the highest score was recorded when K was 6, as shown in Figure 7.

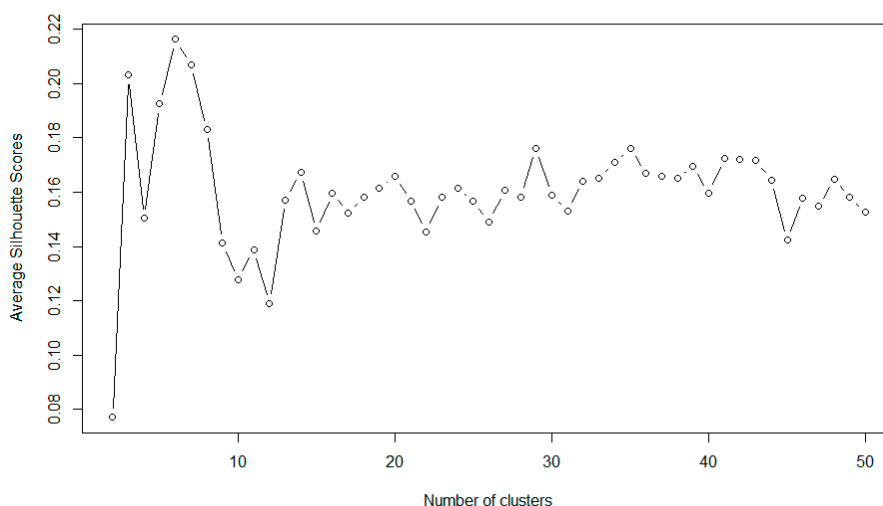


Figure 7. Line chart of the average silhouette score for the K-means.

Table 4 shows a list of the keywords in each cluster, and Table 5 shows the results of establishing the Gyeonggi-Province-type mobility services policy plans by combining a list of the keywords in each cluster. In Table 4, keywords in each cluster were selected based on similarity with the top-ranked word in each cluster and the Gyeonggi Province’s mobility service policy measures were mainly divided into a mobility service, a service provision, and a research area. The six mobility policy issues listed in Table 5 were a combination of the keywords in each cluster and are verified by experts so that those policy issues may reflect the mobility service status of the study site. In the areas of the mobility services, clusters 1, 2, and 3 were included. Clusters 4, 5, and 6 represent the service provision and research areas. The details of the Gyeonggi-Province-type mobility policy issues derived by combining the words contained in the clusters are explained as follows.

Table 4. List of the keywords in the K-means on the mobility service survey.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Vehicle	City	Transportation Seoul	Data	Introduce	Ecosystem
Taxi	Autonomous Driving	Metropolitan Area	Information	Local Government	Viable
Payment	Public Transportation	Smart Mobility	User	Research	Market
Link	Subway	Shared Transport	Platform	Consultative Group	Utilization Rate
Train	Pass	Great Train eXpress	Integrative	Control	
Fare System Integrated Transportation Service			Vitalization		

Table 5. List of the policy issue from K-means on mobility service survey.

Gyeonggi-Province-Type Mobility Policy Issues	
1.	Integrated Transportation Service with Taxi and Public Transportation
2.	Urban-based Autonomous Driving Public Transportation
3.	GTX-based Smart Shared Mobility in Seoul Metropolitan Area
4.	Data Platform for Providing Transportation Service
5.	Mobility Service Research through Institutional Collaboration
6.	Viable Mobility Market

GTX, Great Train eXpress.

The words in each cluster were arranged in the order of the TF-IDF weights. In the case of Cluster 1, there were words about the mobility itself, such as taxis, railways, vehicles, and payment systems. The policy in the cluster was to build an integrated transportation service that combines taxis and public transportation. As mentioned earlier, Finland succeeded in creating Whim, which uses both taxis and public transportation.

Cluster 2 included city, autonomous driving, and public transportation, so it is reasonable to think about combining autonomous driving and mobility, which has become the most controversial issue in the era of the Fourth Industrial Revolution. For Cluster 3, the Great Train eXpress (GTX), the Seoul metropolitan area, smart mobility, and shared transportation were included, so wide-area mobility services could be configured.

Since Cluster 4 contained words such as data, information, and platform, it is possible to think of a data platform that can store big data and provide users with raw data or information generated by mobility. Cluster 5 included words such as local government, research, consultative group, and control, which can be considered as a consultative body that researches mobility services.

Finally, a combination of words in Cluster 6 can refer to a competitive mobility service market not only in Korea but also globally. Given the current semi-public transportation system in Korea, building a mobility service that can perform well in the long term is important. Therefore, it is also necessary to create a market for self-sustainable mobility services that guarantee consistent availability and high utilization rates.

4.4. Verification of the SWOT Analysis Result in Gyeonggi Province

Figure 8 shows the result of the SWOT analysis of the current status of the Gyeonggi Province and the association with the mobility service policy measures derived from each cluster.

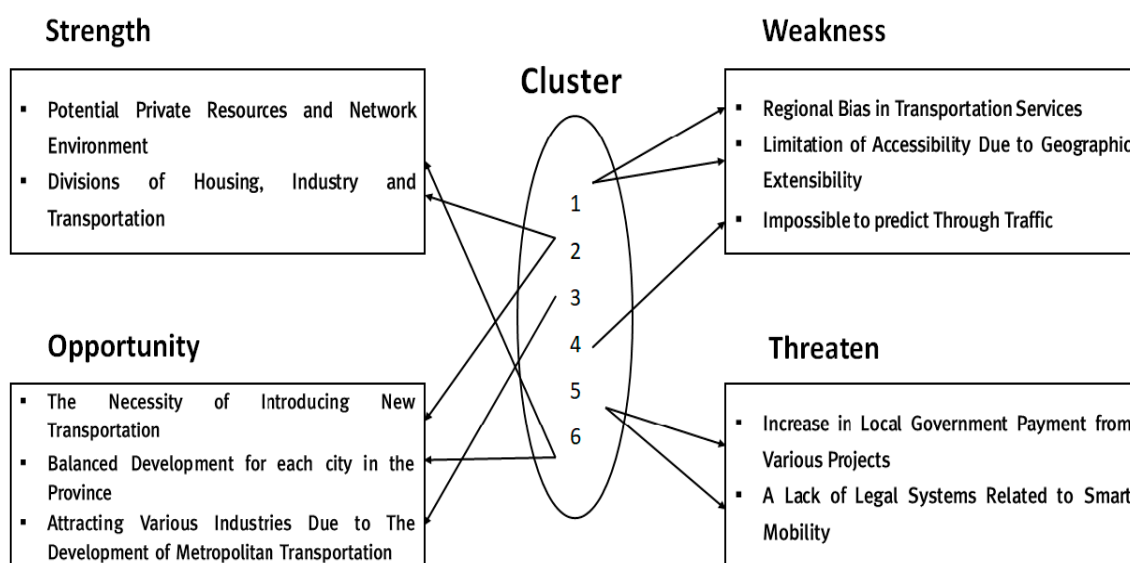


Figure 8. Results of the confirmation of the association between the policy plans and the Gyeonggi Province SWOT analysis.

First, the mobility policy issue to establish an integrated payment system for taxis and public transportation is suitable for situations in which it is difficult for public transportation alone to effectively cover all the areas due to the accessibility limitations. Therefore, the first policy issue may overcome the deepening regional bias and geographical breadth of transportation services, which are weaknesses held by the Gyeonggi Province. Since the infrastructure of public transportation is relatively insufficient for those who live in the underprivileged areas of the Gyeonggi Province, Gyeonggi Province will likely be able to distribute its share of residents to rural areas if it develops a mobility platform that can make passengers arrive at their destinations with taxis and/or public transportation together.

Given the current situation in which the population of Gyeonggi Province continues to rise, autonomous-driving public transportation in urban areas can satisfy both the needs of introducing new transportation with ongoing new town projects. Furthermore, an advanced public transportation system can establish various routes without expanding labor costs. If level 5 autonomous-driving public transportation becomes a reality, it can prevent the overcrowding of certain road sections in advance by providing demand–response mobility services.

The GTX-based shared transportation service in the Seoul metropolitan area can lead the era of the Fourth Industrial Revolution by attracting various industries due to the development of wide-area transportation. The Ministry of Land, Infrastructure, and Transport plans to expand the wide-area transportation networks in metropolitan areas by establishing various wide-area transportation policies, including GTX, which connects Seoul and the Gyeonggi Province [47]. Through this, it would be possible to encourage the establishment of various companies, such as manufacturing, finance, and IT companies, based on the low costs and high convenience that are associated with wide-area mobility services. Additionally, it would be possible to secure competitiveness in secondary and tertiary industries by utilizing the potential resources of the Gyeonggi Province.

Future mobility services can become a reality with the task of building data platforms for providing transportation services. Currently, the population of the Gyeonggi Province is increasing, which makes it difficult to predict the traffic volume or transit demand in certain sections. Therefore, a data collection platform for data analysis should be established to lay the foundation for conducting research on mobility services as well as forecasting the traffic demand.

The research on mobility services through institution collaborations is a measure to address the increase in the local government's burden, the management costs of implementing the current Gyeonggi Province project, and the lack of smart-mobility-related legal and institutional systems. To reduce service costs, cooperation between the Gyeonggi Province and commercial companies is essential, and a fair system should be established to minimize the unfair cases involving damages to users to secure the safety of mobility services.

The self-sustaining mobility market is a significant policy issue for the utilization of potential private resources and networks in Gyeonggi Province and for the balanced development of each local government. According to the current status of mobility services in Gyeonggi Province, the losses incurred in the public transportation sector are covered by the provincial government. Although public institutions can bear the burden by defining the mobility services in Gyeonggi Province as a public sector for a while, securing the economic feasibility of the mobility service market is most urgent to provide equal mobility services in all the cities in Gyeonggi Province. Therefore, measures should be devised to reflect the resources and networks possessed by Gyeonggi Province in the mobility market.

5. Conclusions

5.1. Summary and Implications

In this study, an exploratory research was conducted on the development of the forthcoming mobility service policy by utilizing a text mining technique. The content was collected and analyzed through interviews with 36 experts from various fields of transportation, cities, policies, and autonomous driving while focusing on the case of Gyeonggi Province, which has the nation's largest local government and is at the center of the regional and inter-regional traffic issues. Additionally, the search for future mobility service policies in the Seoul metropolitan area was highly useful, as various transportation policies are also centered in Seoul and Gyeonggi Province.

Here, the K-means, one of the most popular clustering methods, was borrowed to derive mobility policy measures in the Seoul metropolitan area, and key keywords for the Gyeonggi-Province-type mobility services were derived based on the TF-IDF algorithm. Ultimately, a clustering analysis was conducted to derive the Gyeonggi-do mobility service policy issue. In the paradigm of the

Fourth Industrial Revolution, exploratory approaches to shared transportation, autonomous driving, and integrated transportation services were implemented.

As mentioned above, the clustering results showed the Gyeonggi-Province-type mobility policy issues, which can be concluded in six plans as follows: (1) transportation services based on the integration of public transportation, (2) the introduction of a self-driving public transportation system in rural areas, (3) GTX-based inter-regional (capital area) and shared transportation services, (4) the establishment of a data platform to provide transportation services, (5) deriving mobility services through cooperation with various institutions, and (6) the creation of a self-sustaining mobility ecosystem.

Compared with the strengths, weaknesses, opportunities, and threats of Gyeonggi Province, these policy proposals, which were created using our proposed methodology, agree with the conditions of the abundant resources and traffic locations in Gyeonggi Province, the introduction of new means of transportation due to the spread of urban planning, and the convenience enhancement in various industries and transportation processes through the expansion of wide-area transportation networks.

This study is valuable as an exploratory study. Policy issues related to future mobility have been addressed around the keywords, so the possibility of classification analyses can be investigated by extracting the major words from the experts' answers in various fields in the future. By calculating the importance of specific words in a particular topic, the same methodology can be applied in similar fields in the future to verify their validity.

Implications of our proposed methodology can be applied to various areas. First of all, it is the first research to create customized mobility service policy issues for Gyeonggi Province using text mining and clustering algorithms. Academically, our research has revealed that not only can Seoul's mobility policy issues be derived through the same process, but also the methodology can formulate customized policy issues related to any area with expert surveys. Second of all, results from our proposed methodology will help each local government's officials to establish policies. It is important for them to summarize the opinions of experts. Our methodology will give them an opportunity to save time for summarization and understand key issues before making decisions on formulating policies.

Limitations also exist because of the keyword extraction analysis of the survey. Policy issues by our proposed methodology could be normative and declarative because they are not clustered based on quantitative data but on the importance of keywords. However, our proposed methodology showed that the importance of keywords, which is qualitative data, can be transformed into quantitative value. Furthermore, policy issues in the near future can be derived from a group of experts who have knowledge of the past progress in mobility services, the current awareness of problems, and future mobility predictions. Since Coronavirus disease of 2019 (COVID-19) has led to changes in all areas, including politics, economy, society, culture, and science, the mobility sector should also make an active breakthrough with the exploration of new policy issues and preemptive challenges.

5.2. Limitations and Future Research

This study is one of the leading works to suggest the establishment of future mobility service policies using text mining techniques. However, the lack of sufficient experts in the field of future mobility has the following limitations. First, there were only 36 experts who majored in urban and transportation sectors because of the lack of sufficient experts in the field of future mobility who understand the current situation in Gyeonggi Province. Although research with expert groups does not consider the number of samples significantly, obtaining at least 15 samples is encouraged [48,49] and there is some research that calculated the minimum required experts [50,51]. Based on the equation in the researches, the paper needs at least 33 experts with a statistical significance level of 0.05. However, the reliability of the study can significantly increase with the increase in the number of respondents to the survey [52]. Thus, further research needs to be conducted after finding more experts in more diverse areas. Second, since the research area has been limited to Gyeonggi Province, the results of the study may include some locality issues. Although the research applications can be used in other regional mobility policies, their effectiveness may vary depending on the regional characteristics (e.g., traffic level, infrastructure, and population). Therefore,

it is necessary to conduct future studies similar to this study as examples in other regions to compare and analyze the direction of inter-regional mobility policies.

Author Contributions: Conceptualization, Y.S., D.L., and H.K.; methodology, Y.S., D.L., W.S., Y.K., J.K., and H.K.; software, Y.S.; validation, Y.S., D.L., and W.S.; formal analysis, Y.S. and Y.K.; investigation, Y.S. and W.S.; resources, Y.S., J.K., and H.K.; data curation, Y.S.; writing—original draft preparation, Y.S., D.L., W.S., Y.K., and J.K.; writing—review and editing, Y.S., D.L., W.S., Y.K., J.K., and H.K.; visualization, Y.S. and D.L.; supervision, J.K. and H.K.; project administration, Y.S. and H.K.; funding acquisition, H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (no. 2020R1C1C1003296).

Acknowledgments: We are grateful to anonymous referees for evaluating the suitability of our proposed methodology on the topic.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Adey, P. *Mobility*, 2nd ed.; Routledge: Abingdon-on-Thames, UK, 2017.
2. Campisi, T.; Akgün, N.; Ticali, D.; Tesoriere, G. Exploring Public Opinion on Personal Mobility Vehicle Use: A Case Study in Palermo, Italy. *Sustainability* **2020**, *12*, 5460. [CrossRef]
3. Zagorskas, J.; Burinskienė, M. Challenges Caused by Increased Use of E-Powered Personal Mobility Vehicles in European Cities. *Sustainability* **2020**, *12*, 273. [CrossRef]
4. Wang, H.; Grindle, G.G.; Candiotti, J.; Chung, C.; Shino, M.; Houston, E.; Cooper, R.A. The Personal Mobility and Manipulation Appliance (PerMMA): A Robotic Wheelchair with Advanced Mobility and Manipulation. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August–1 September 2012; pp. 3324–3327.
5. Barr, S. Personal mobility and climate change. *Wires Clim. Chang.* **2018**, *9*, e542. [CrossRef]
6. Whim. All Transport in One App. Available online: <https://whimapp.com/> (accessed on 9 September 2020).
7. Jittrapirom, P.; Caiati, V.; Feneri, A.; Ebrahimigharehbaghi, S.; González, M.; Narayan, J. Mobility as a Service: A Critical Review of Definitions, Assessments of Schemes, and Key Challenges. *Urban Plan.* **2017**, *2*, 13–25. [CrossRef]
8. Motta, G.; Sacco, D.; Ma, D.; You, D.; Liu, K. Personal Mobility Service System in Urban Areas: The IRMA Project. In Proceedings of the 2015 IEEE Symposium on Service-Oriented System Engineering, San Francisco Bay, CA, USA, 30 March–3 April 2015; pp. 88–97.
9. Bösch, P.M.; Becker, F.; Becker, H.; Axhausen, K.W. Cost-based analysis of autonomous mobility services. *Transp. Policy* **2018**, *64*, 76–91. [CrossRef]
10. Dandl, F.; Hyland, M.; Bogenberger, K.; Mahmassani, H.S. Evaluating the impact of spatio-temporal demand forecast aggregation on the operational performance of shared autonomous mobility fleets. *Transportation* **2019**, *46*, 1975–1996. [CrossRef]
11. Szigeti, S.; Csiszár, C.; Földes, D. Information Management of Demand-responsive Mobility Service Based on Autonomous Vehicles. *Procedia Eng.* **2017**, *187*, 483–491. [CrossRef]
12. Gwon, Y.M.; Kim, H.J.; Lim, G.I.; Kim, J.H.; Son, W.B. A Research on Forecasting Change and Service Direction for the Future Mobility System. *Korea Inst. Intell. Transp. Syst.* **2020**, *19*, 100–115.
13. Kwon, Y.M.; Kim, S.J.; Kim, H.J.; Byun, J.H. What Attributes Do Passengers Value in Electrified Buses? *Energies* **2020**, *13*, 2646. [CrossRef]
14. Jeong, S.; Yoon, H. QAP Structural Characterization of Public Transportation Networks based on QAP Correlation. *J. Inf. Technol. Appl. Manag.* **2019**, *26*, 95–102.
15. Korea Transportation Safety Authority. 2018 Public Transportation Investigation. 2018. Available online: <http://stat.molit.go.kr/portal/cate/statMetaView.do?hRsId=483&hFormId=5499&hSelectId=5499&sStyleNum=1&sStart=2011&sEnd=2011&hPoint=00&hAppr=1&oFileName=&rFileName=&midpath=> (accessed on 3 September 2020).

16. Korea Appraisal Board. Korea Housing Price Index. 2018. Available online: https://www.housing.or.kr/user/boardList.do?command=view&page=2&boardId=21825&boardSeq=24411&id=home_030300000000 (accessed on 3 September 2020).
17. Beutel, M.C.; Gökyay, S.; Kluth, W.; Krempels, K.; Samsel, C.; Terwelp, C. Product Oriented Integration of Heterogeneous Mobility Services. In Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014; pp. 1529–1534.
18. Ibrahim, A.; Zhao, L. Supporting the OSGi Service Platform with Mobility and Service Distribution in Ubiquitous Home Environments. *Comput. J.* **2009**, *52*, 210–239. [[CrossRef](#)]
19. Frändberg, L.; Vilhelmson, B. More or less travel: Personal mobility trends in the Swedish population focusing gender and cohort. *J. Transp. Geogr.* **2011**, *19*, 1235–1244. [[CrossRef](#)]
20. Alanne, K.; Cao, S. Zero-energy hydrogen economy (ZEH2E) for buildings and communities including personal mobility. *Renew. Sustain. Energy Rev.* **2017**, *71*, 697–711. [[CrossRef](#)]
21. Podobnik, J.; Rejc, J.; Slajpah, S.; Munih, M.; Mihelj, M. All-Terrain Wheelchair: Increasing Personal Mobility with a Powered Wheel-Track Hybrid Wheelchair. *IEEE Robot. Autom. Mag.* **2017**, *24*, 26–36. [[CrossRef](#)]
22. Schwanen, T. Innovations to transform personal mobility. In *Low Carbon Mobility Transitions*; Hopkins, D., Higham, J., Eds.; Goodfellow Publishers: London, UK, 2016; pp. 154–196.
23. Distler, V.; Lallemand, C.; Bellet, T. Acceptability and Acceptance of Autonomous Mobility on Demand: The Impact of an Immersive Experience. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–10.
24. Kwon, Y.M.; Son, S.H.; Jang, K.T. User satisfaction with battery electric vehicles in South Korea. *Transp. Res. Part D Transp. Environ.* **2020**, *82*, 102306. [[CrossRef](#)]
25. Hong, S.Y. *A Study on the Introduction of Smart Mobility Service in Seoul*; The Seoul Institute: Seoul, Korea, 2019; Available online: <https://www.si.re.kr/node/62534> (accessed on 4 September 2020).
26. Gyeonggi Province Government. *A Strategy for Smart Mobility in Gyeonggi Province-Type*; Gyeonggi Province Government: Seongnam City, Korea, 2019.
27. Cirianni, F.; Monterosso, C.; Panuccio, P.; Rindone, C. A review methodology of sustainable urban mobility plans: Objectives and actions to promote cycling and pedestrian mobility. In *Smart and Sustainable Planning for Cities and Regions*; Bisello, A., Vettorato, D., Laconte, P., Costa, S., Eds.; Springer: Cham, Switzerland, 2018; pp. 685–697.
28. Russo, F.; Rindone, C. The Planning Process and Logical Framework Approach in Road Evacuation: A Coherent Vision. In Proceedings of the 4th International Conference on Safety and Security Engineering, Antwerp, Belgium, 4–6 July 2011; pp. 415–425.
29. Amrita, C.; Paaauwb, T.; Alyc, R.; Lavrica, M. Identifying child abuse through text mining and machine learning. *Expert Syst. Appl.* **2017**, *88*, 402–418. [[CrossRef](#)]
30. Tang, V.; Siu, P.K.Y.; Choy, K.L.; Ho, G.T.S.; Lam, H.Y.; Tsang, Y.P. A web mining-based case adaptation model for quality assurance of pharmaceutical warehouses. *Int. J. Logist. Res. Appl.* **2019**, *22*, 325–348. [[CrossRef](#)]
31. Tang, V.; Siu, P.K.Y.; Choy, K.L.; Lam, H.Y.; Ho, G.T.S.; Lee, C.K.M.; Tsang, Y.P. An adaptive clinical decision support system for serving the elderly with chronic diseases in healthcare industry. *Expert Syst.* **2019**, *36*, e12369. [[CrossRef](#)]
32. Orellana, G.; Orellana, M.; Saquicela, V.; Baculima, F.; Piedra, N. A Text Mining Methodology to Discover Syllabi Similarities among Higher Education Institutions. In Proceedings of the 2018 International Conference on Information Systems and Computer Science (INCISCOS), Quito, Ecuador, 14–16 November 2018; pp. 261–268.
33. Jeong, H.R.; Park, S.M.; Jun, Y.J.; Choi, J.W.; Choe, B.H.; Park, K.H.; Yun, I.S. Reclassification of Traffic Crashes Using Traffic Crash Report Data and Keyword Analysis. In Proceedings of the 13th International Conference on Probabilistic Safety Assessment and Management (PSAM 13), Seoul, Korea, 2–7 October 2016.
34. So, J.H.; Kim, T.H.; Kim, M.J.; Kang, J.W.; Lee, H.; Choi, J.M. A Study on the Concept of Smart City and Smart City Transport. *J. Korean Soc. Transp.* **2019**, *37*, 79–91. [[CrossRef](#)]
35. Song, M.; Kim, M.C.; Jeong, Y.K. Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter. *IEEE Intell. Syst.* **2014**, *29*, 18–26. [[CrossRef](#)]
36. Charalampakis, B.; Spathis, D.; Kouslis, E.; Kermanidis, K. A comparison between semi-supervised and supervised text mining techniques on detecting irony in greek political tweets. *Eng. Appl. Artif. Intell.* **2016**, *51*, 50–57. [[CrossRef](#)]
37. Chae, Y.S.; Lee, S.H. Identification of Strategic Fields for Developing Smart City in Busan Using Text Mining. *J. Digit. Converg.* **2018**, *16*, 1–15.

38. Chen, K.; Zhang, Z.; Long, J.; Zhang, H. Turning from TF-IDF to TF-IGM for term weighting in text classification. *Expert. Syst. Appl.* **2016**, *66*, 245–260.
39. Yahav, I.; Shehory, O.; Schwartz, D. Comments mining with TF-IDF: The inherent bias and its removal. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 437–450. [[CrossRef](#)]
40. Gyeonggi Autonomous Driving Center. *A Study on the Development of Mobility Service Ecosystem*; Gyeonggi Autonomous Driving Center: Gyeonggi, Korea, 2020.
41. Lilleberg, J.; Zhu, Y.; Zhang, Y. Support Vector Machines and Word2vec for Text Classification with Semantic Features. In Proceedings of the 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing, Beijing, China, 6–8 July 2015; pp. 136–140.
42. Han, S.S.; Yang, D.W. Analysis of Research Trends Related to Start-Up Using Text Mining. *Korean Soc. Bus. Ventur.* **2017**, *12*, 1–12.
43. Robertson, S. Understanding inverse document frequency: On theoretical arguments for IDF. *J. Doc.* **2004**, *60*, 503–520. [[CrossRef](#)]
44. Havrlant, L.; Kreinovich, V. A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *Int. J. Gen. Syst.* **2017**, *46*, 27–36. [[CrossRef](#)]
45. Jing, L.P.; Huang, H.K.; Shi, H.B. Improved feature selection approach TFIDF in text mining. In Proceedings of the International Conference on Machine Learning and Cybernetics, Beijing, China, 4–5 November 2002; pp. 944–946.
46. Kim, D.; Seo, D.; Cho, S.; Kang, P. Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec. *Inf. Sci.* **2019**, *477*, 15–29. [[CrossRef](#)]
47. Metropolitan Transport Commission. Metropolitan Transportation 2030. 2020. Available online: <https://www.molit.go.kr/metro/main.jsp> (accessed on 4 September 2020).
48. Armacost, R.L.; Compton, P.K.; Mullens, M.A.; Swart, W.W. An AHP framework for prioritizing customer requirements in QFD: An Industrialized Housing Application. *IIE Trans.* **1994**, *26*, 72–79. [[CrossRef](#)]
49. Al-Harbi, K.M.A.S. Application of the AHP in project management. *Int. J. Proj. Manag.* **2001**, *19*, 19–27. [[CrossRef](#)]
50. Turoń, P.; Kubic, A.; Chen, F.; Wang, H.; Lazarz, B. A Holistic Approach to Electric Shared Mobility Systems Development—Modelling and Optimization Aspects. *Energies* **2020**, *13*, 5810. [[CrossRef](#)]
51. Turoń, P.; Kubic, A. Economic Aspects of Driving Various Types of Vehicles in Intelligent Urban Transport Systems, Including Car-Sharing Services and Autonomous Vehicles. *Appl. Sci.* **2020**, *10*, 5580. [[CrossRef](#)]
52. Dalkey, N.; Brown, B.; Cochran, S. Use of Self-ratings to Improve Group Estimates: Experimental Evaluation of Delphi Procedures. *Technol. Forecast.* **1970**, *1*, 283–291. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).