

## Article

# Recommending K-Wave Items Tailored for Small-Sized Exporters by Incorporating Dense and Sparse Vectors

Jimin Lee <sup>1</sup>, Eunjeong Na <sup>2</sup>, Keejun Han <sup>2,\*</sup> and Donggil Na <sup>3,\*</sup>

<sup>1</sup> Department of AI and Big Data, Soonchunhyang University, Asan 31538, Republic of Korea; dlwls7359@sch.ac.kr

<sup>2</sup> School of Computer Engineering, Hansung University, Seoul 02876, Republic of Korea; eunjeong42na@hansung.ac.kr

<sup>3</sup> Intelligent Convergence Research Laboratory, Electronics and Telecommunications Research Institute, Daejeon 34129, Republic of Korea

\* Correspondence: keejun.han@hansung.ac.kr (K.H.); dgna@etri.re.kr (D.N.)

**Abstract:** As K-wave has been strengthened via recent K-contents, K-wave items such as cosmetics and electronic devices have also gained attention globally. For small-sized export sellers who purchased the items and exported them to different countries, it is significant to discover which K-wave items are trending in specific countries. To do so, we proposed an ensemble recommender system by producing the dense vector, which is generated by a variant of Bidirectional Encoder Representations from Transformers (BERT), and balancing the vector with a sparse vector in order to ensure the efficient execution speed and recommendation accuracy. Based on the data we have collected specifically for potential K-items, our experiment showed that the proposed model outperforms the various baselines, which are used for content-based filtering.

**Keywords:** K-wave; BERT; export sellers; trending; recommender systems; content-based filtering; dense vector; sparse vector



**Citation:** Lee, J.; Na, E.; Han, K.; Na, D. Recommending K-Wave Items Tailored for Small-Sized Exporters by Incorporating Dense and Sparse Vectors. *Sustainability* **2023**, *15*, 16098. <https://doi.org/10.3390/su152216098>

Academic Editors: Young-Chan Lee and Runhui Lin

Received: 10 August 2023

Revised: 30 October 2023

Accepted: 6 November 2023

Published: 20 November 2023



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

## 1. Introduction

As the Korean Wave (K-wave), a cultural phenomenon in which South Korean popular culture has gained global popularity, strengthens with recent K-contents, K-wave items such as cosmetics and electronic devices have also garnered increased global attention [1]. This rapid spread is mainly handled by small export sellers with limited resources, implying that deciding what items to sell and the target countries should be carefully considered. Since trends change quickly, it would be beneficial to provide insightful information and trending predictions based on analysis from import and export data to help exporters make better-informed decisions.

The most straightforward approach for discovering trends is to retrieve relevant articles about items and the target countries. Specifically, when exporters investigate some items and countries, they could gain more confidence in making their final purchasing and exporting decisions by reading more relevant articles about the items and countries. To fulfill this need, we approach the recommendation problem by suggesting more relevant articles to the current article they are interested in.

A conventional way of building a recommender system is to use user rating data for collaborative filtering [2]. However, articles regarding the items and target countries are often context-only information, indicating that content-based filtering would be an alternative by measuring the similarity among the articles. Although the recommendation results are built for the target country, the items would vary, degrading user satisfaction as they prefer to continue reading the articles that are simultaneously relevant to the items and target country. This implies that the item category should also be considered so that the uniformity of the recommendation results is strengthened. Every item to be

exported has a kan-code: a serialized unique code. By utilizing the code, recommendation results can be filtered to only include the identical category. However, there is often a latent relationship between two items with different kan-codes. For instance, Korean fried chicken and Korean beer, though unrelated in terms of kan-code, became popular in Mainland China because they co-occurred together in a scene in a K-drama [3]. The conventional content-based filtering approach is also limited to capturing the latent information as it is mainly based on matching two different sparse vectors generated by the term dictionary.

To overcome this limitation, we propose an ensemble recommender system by producing the dense vector, which is generated by a variant of Bidirectional Encoder Representations from Transformers (BERT) [4], and balancing the vector with a sparse vector to ensure efficient execution speed and recommendation accuracy. BERT is the most well-known transformer-based approach to first learn the latent relationships between terms and sentences from massive unlabeled data such as Wikipedia. The model is then fine-tuned to perform a specific task, indicating that the proposed approach exploits the latent information for the final prediction. Based on the data we have specifically collected for potential K-items from three different categories—clothing, cosmetics, and electronic devices—we conducted an experiment with baselines widely used in existing recommendation systems. Throughout the experiment, we observed that the proposed ensemble approach promises further improvement for recommendation results.

In summary, the contributions of this paper can be summarized as follows:

- We build a decision support system for small-sized exporters by providing articles on which K-items are trending in targeted countries by proposing an ensemble recommendation approach utilizing sparse and dense vectors simultaneously;
- We build a benchmark dataset across different categories that can be used for local and global logistics big data analysis;
- By adapting the variant BERT, which is fine-tuned for the logistic domain, we achieve higher effectiveness compared to the baselines.

The remainder of this paper is structured as follows: Related Work reviews related studies and the Methods section describes the procedures of the proposed ensemble model for the recommendation. The experiments section demonstrates the effectiveness of the proposed model through our experiments and presents the results. Lastly, the conclusion section discusses limitations, future work, and concludes the study.

## 2. Related Work

### 2.1. Personalized Services

With the development of artificial intelligence technology, interest in personalized recommendation systems using big data is increasing, and research is actively in progress. Personalized services improve user satisfaction and experience by providing customized services in consideration of user tastes and preferences. Accordingly, it is possible to promote the user's purchase decision and obtain business results.

Bolun et al. [5] developed a customized lightweight memory profiler called DroidPerf to provide personalized memory optimization to mobile developers. This tracks the efficiency of using memory with objects generated by the app and optimizes memory layout, access patterns, and allocation patterns based on information to improve the performance of the app. Yueshen et al. [6] developed a new personalized storage recommendation service using multimodal characteristic learning for developers. This helped us quickly find a suitable repository for sharing software, alleviating the challenge of finding the right one. Zhan et al. [7] proposed a hybrid approach for providing personalized item recommendations to users by analyzing user emotions and user preferences from product images and review text. The mixed model consists of two parallel modules performing multi-scale semantic and visual analysis (MSVA), demonstrating the performance improvement of MSVA for various products. Liu et al. [8] proposed an emotion-based personalized music recommendation system that meets the emotional needs of users and improves their mental state. Using the LSTM-based model, the most suitable music is selected based on

the user's previous and current emotions. Guo et al. [9] proposed personalized education to group students and recommended courses for the purpose of improving cooperative learning. The frequency of words and Word2Vec are used to extract student characteristics, and the K-means algorithm is used to form interest-based groups. Lee et al. [10] proposed a personalized exercise recommendation algorithm using personal propensity information to recommend accurate and meaningful exercises for users. The k-nearest neighbors algorithm (KNN) is used to classify users according to the criteria for obesity, and the singular value decomposition (SVD) algorithm provides users with personalized exercise recommendations. Kim et al. [11] introduced a customized cosmetics recommendation system using product search records, skin characteristics, and sensitivity data. Based on multiple user data, a collaborative filtering-based recommendation system was proposed that recommends items preferred by users with similar tendencies. Yun et al. [12] implemented a personalized recommendation system by utilizing demographic variables, psychological variables, and behavioral variables along with rating information for each food menu. Their system constructs the group most similar to the user and uses the group's ratings to predict the user's preference for the food menu item.

## 2.2. News Recommendation

A recommendation system is operated using purchase behavior, ratings, and search records, and recommends customized information to users. It requires a lot of time and effort to find information regarding the user's interests among the numerous pieces of information; however, it can assist users in understanding their interests and preferences and immediately recommend desired information, eliminating information overload and saving time. It can also increase user satisfaction and utilization. As a result, recommender systems are becoming more sophisticated and the use of deep learning is increasing.

Amir et al. [13] investigated the strengths, weaknesses, and trends of news recommendation models using deep learning methods. Their study deals with the classification of deep learning-based news recommendation models and conducts a performance comparison of the models. Zhang et al. [14] proposed a "Prompt4NR" framework that improves news recommendations by introducing a new "Prompt Learning" paradigm. This framework transforms the user's news click prediction task with close-style mask prediction tasks and validates effective results with various prompt templates and ensembles. Gharahighehi et al. [15] proposed a session-based recommendation system because traditional collaborative filtering was ineffective in recommending news. They proposed a scenario to improve the diversity of session-based recommendation systems and solve the filter bubble phenomena. This was proven to be effective based on four news datasets. Kim et al. [16] proposed a news recommendation system using text mining and LOD techniques to determine important news articles. Through K-means clustering, duplicate news articles are removed and important keywords for each news article are extracted using the TF-IDF model and TextRank algorithm. This method provides a function to summarize news articles that users want to read through the LOD interface. Jeong et al. [17] proposed a personalized news recommendation system by calculating users' preferences for tags in news articles. The user's preferred tag is recommended by utilizing text and tags together, extracted from news article images. The method introduced by Hong et al. [18] creates a profile using the user's SNS information and recent posts to identify the user's recent preferences. Using this for news services, a system was proposed to recommend news that suits one's preference based on similarity.

## 2.3. Content Filtering and Collaborative Filtering

Recommender systems include collaborative filtering utilizing user records, content filtering using news articles and keywords, and hybrid filtering that combines the two technologies. In this paper, a news recommendation system was established using a content filtering service. The following study employs content filtering and collaborative filtering, which are the main methods used in our recommendation system.

Kim et al. [19] developed recommendations of similar media content to an OTT platform using content-based filtering. Similar content was recommended based on text data, and genre classification and hashtag classification were predicted. Bhaskaran et al. [20] proposed an improved recommendation system that provides personalized recommendations to learners in educational fields such as e-learning. The user's learning style is classified through a server blog, and similarity is calculated through content-based filtering to create a recommendation list. Kim et al. [21] developed a craft recommendation system using content-based filtering. Crafts are dependent on expert recommendations, but search records and attribute information of crafts were compared for customized recommendations by users. Crafts with high cosine similarity were sorted and recommended, and content-based filtering was able to solve the cold start problem.

Behera et al. [22] proposed a collaborative filtering model for movie recommendations considering the temporal and dynamic effects of user-item interactions. The proposed model was evaluated on a standard dataset and showed improved performance over state-of-the-art models. Hwang et al. [23] proposed a recommendation system on a hobby item rental platform using collaborative filtering. Users may receive a recommendation only if they have a history of inquiring about an item. Both the number of clicks and the time spent on the page are also reflected so that effective recommendations can be received. Kim et al. [24] proposed a university liberal arts recommendation system using collaborative filtering. Student history data is used to assist students who are having difficulty finding their preferred classes due to the large number and variety of subjects. Using cosine similarity, the most similar students are found and subjects are recommended based on information on liberal arts and student lecture evaluations.

Vuong Nguyen et al. [25] proposed a hybrid recommendation method that utilizes the characteristics of movie content instead of analyzing user rating data to improve the collaborative filtering recommendation system. By combining word embedding and collaborative filtering, they measured the similarity between movies and improved the efficiency of recommendations. Walek et al. [26] proposed an e-shop recommendation system by combining a fuzzy expert system with collaborative filtering and content-based approaches. This system recommends suitable products in consideration of user preferences and e-store activities. Jomsri et al. [27] proposed a hybrid recommendation method that combines collaborative filtering and content-based filtering for digital libraries developed by multiple online publishers. Their model achieved better performance than traditional methods with a hybrid recommendation system model that considers book categories, user reading habits, and information source elements.

### 3. Methods

In this section, we propose a model that uses the news data together with the user's past news history. The KOBERT model allows you to classify news categories. Our approach is divided into three ways to recommend relevant news to users, and the following sections outline the proposed model in order of approach.

#### 3.1. Proposed Model Overview

Figure 1 illustrates the general structure of the content-based filtering approach for building recommender systems. This is a method of analyzing the characteristics of content that users have preferred in the past and recommending other content with similar characteristics. The data comprises daily news data spanning one year from Kotra Overseas Market News and Kita Korea International Trade Association's Overseas Market News site, including the news body and corresponding kan-code as variables. The kan-code is used in various industries, especially in sales, logistics, and manufacturing to efficiently manage the distribution of products. The kan-code used in this paper represents the standard code for product classification and is the classification code of the distribution product standard DB managed by the Korea Chamber of Commerce and Industry's Distribution and Logistics Agency. The kan-code represents information using a total of 6 digits: 2 digits

for each classification into major, middle, and small categories. In this study, labeling was conducted using only the numerical information representing the major classification. For the categories represented by the first digit, 9 indicates clothing, 3 indicates cosmetics, and 7 indicates electronics. Table 1 shows examples of the data. First, data is processed to remove unnecessary symbols from the body content. This also removes news that have multiple categories that correspond to one news item, as multiple categories are considered unclear news. Second, the preprocessed news data is labeled with a total of 6052 data items, which are labeled as 0 for clothing, 1 for cosmetics, and 2 for electronics based on the kan-code number. The labeled data is shown in Table 2. Then, a recommender system is implemented. This is modeled using the body of news that the user has read in the past. After tokenizing the content of the body, the KOBERT model determines the category of the news. After checking the category, news of the same category are recommended to the user.



**Figure 1.** Content-based filtering method.

**Table 1.** An example of news data.

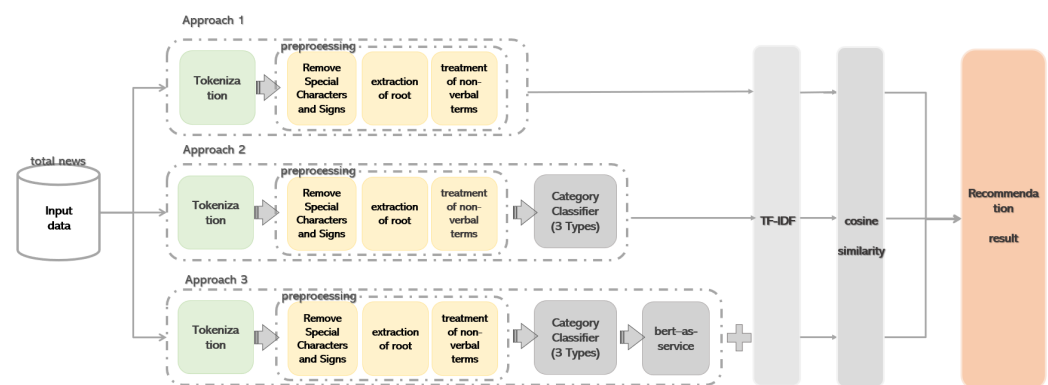
Id	Content	Corresponding Kan-Codes
59	K-AgriTech and ArgoTech start-ups entering France prioritize identifying local ecosystems	730030
128	Singapore, which is spurring unmanned systems, plans to utilize various military-private sectors	903030, 701010, 730030
180	The world's top three industrial exhibitions were held in a hybrid manner for the first time in the era of COVID-19	903030, 730010, 730030
209	2021 Sri Lanka Investment Forum Event, Video Due to COVID-19	110701, 730030
421	Attention is focused on the agricultural WEEK exhibition, 6th industry, and smart agriculture-related booths held in 2 years	903030, 110300, 730030

**Table 2.** An example of kan-code labeled data.

Content	Kan-Code
KAgriTech, AgroTech’s entry into France should prioritize understanding the local ecosystem	2
Invest Chile, Sep. 16 Webinar on how to efficiently proceed with investment projects in Chile	2
It oversees all of Thailand’s largest corporate group companies, which account for 0.16 of the market capitalization of listed companies in Thailand	2
Ecuador Airport Can Play a Geopolitically Latin American Hub to Sell Some Airport Operating Rights	2

### 3.2. Description of the Proposed Model

This section proposes a third approach. Figure 2 shows a detailed description of the proposed model.

**Figure 2.** Proposed model overview (build recommendation system).

- First, data entry is performed using uncategorized full news data. The news data consist of the news title and the content of the news body;
- Second, when there are two variables, the news title and the news body, the input variable is entered. We then proceed with tokenization. Tokenization refers to dividing natural language processing into units to use as a necessary step for preprocessing. We use Google’s Transformer for tokenization;
- Third, preprocessing is carried out. Text preprocessing is performed by removing special characters and signs, extracting roots, and processing non-verbal terms to remove information that is not relevant to the analysis. Out of a total of 20,756 data points, 20,686 data points were extracted through the data purification process. Special character and sign removal were able to eliminate the noise that hindered model learning. Because root extraction unifies words into basic forms, it can solve the problem of expressing the same words in various forms and reduce the size of the final text data. Terminology processing can highlight important information and improve the accuracy of the model; therefore, we contributed to improving the performance of the model by improving the quality of text data through the preprocessing process;
- Fourth, category discrimination is carried out. The KOBERT model was used to determine the categories. KOBERT is a Korean version of the BERT natural language processing model. This model is based on the Transformer architecture and is very effective at processing sequence data. Based on large-scale Korean text data, KOBERT can be fine-tuned for various natural language processing tasks, understanding words and contexts in sentences through two-way learning. This model improves Korean text processing and performs well in various NLP tasks such as emotion classification, text classification, text generation, question response, etc. First, for KOBERT model



learning, we used news data, which is classified into three categories, clothing, cosmetics, and electronics, using Kan-code. When the model is trained with Epoch 10, the accuracy of category classification is 0.829. We built a model to find the category classification and reliability of news by entering the title or body content of the news. For example, if you put the news title as input data, “Regulations you need to know to distribute cosmetics in the UK” from 2021, the news category was classified as clothing and the reliability was 0.9795;

- Fifth, the Bert-as-service model was used to concatenate the KOBERT model and TF-IDF. Bert-as-service is a sentence encoding service that maps variable-length sentences to fixed-length vectors. KOBERT was used in the Bert-as-service model and TF-IDF was concatenated together. There was a difference between the shape of the Bert-as-service and the shape of the TF-IDF; therefore, dimensional reduction was required to perform the concat. The TF-IDF was dimensionally reduced to (1768) to match the shape of the model;
- Sixth, TF-IDF was employed, which refers to the weights used in text mining. The statistical figures show which words are important within a particular document. It is a method of weighting the degree of importance within a document according to the frequency of a word. TF-IDF determines that the more frequently words appear in all documents, the less important the word is, and that words appearing frequently only in a specific document are more important. If the word  $i$  is in document  $j$ , the formula for TF-IDF is  $tfi_{i,j}$ .  $dfi$  indicates the number of documents containing the word  $i$ .  $N$  is the total number of documents. The following formula is used to obtain the TF-IDF value. For the third approach, the Bert-as-service and TF-IDF are concatenated as described earlier.

$$wi_{i,j} = tfi_{i,j} \times \log\left(\frac{N}{dfi}\right) \quad (1)$$

- Seventh, the cosine similarity of the concatenated data results is determined in the most relevant news. The cosine similarity refers to the similarity between two vectors obtained using the cosine angle. The more frequently a word appears within a document, the larger the size of the vector. The equation for cosine similarity is as follows:

$$similarity = \cos\theta = \frac{A \times B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n Ai \times Bi}{\sqrt{\sum_{i=1}^n (Ai)^2} \times \sqrt{\sum_{i=1}^n (Bi)^2}} \quad (2)$$

The most similar news data is extracted by comparing the vector values of the captured Bert-as-service and TF-IDF with the vector values of the data entering input data using the above formula. Among them, the top news is judged to be the most relevant;

- Finally, the recommendation results are obtained. In this step, news is recommended to users based on the top news obtained earlier through cosine similarity.

Here, we described the three approaches for our proposed model. For the first approach, the recommended news results are extracted from all data without category classification via cosine similarity and TF-IDF. Therefore, it selects recommended news within all categories without categorizing the data categories. For the second approach, news is categorized into KOBERT models using preprocessed data into three categories: clothing, cosmetics, and electronics, not the entire data. Within the same category of news as the category of news received by the input data, the cosine similarity and TF-IDF are used to extract the recommended news results.

#### 4. Experiments

The accuracy is 0.58 when the body of the news item is used for the KOBERT model in our proposed method. The ratio of training to test set was 9:1, and cross-validation was used. The TextRank algorithm was used to improve accuracy, which is a graph-based ranking model that uses sentences or words in a document to calculate the ranking of

sentences. The Textrank algorithm was used to extract important sentences from the news body content. If the extracted important statements are used, the accuracy is improved to 0.64. When the ratio of the train and test set is changed to 7:3, the highest accuracy was achieved, as shown by several experiments. In the same way, when the text is used separately from the TextRank algorithm, the accuracy is 0.759, and it improves by changing the ratio of the train and test set. When using the TextRank algorithm, the accuracy was 0.751, similar to that of using the text as is. By increasing the number of epochs, the results show that the epoch is 10 and the category prediction accuracy is the highest at 0.829 when using the text rank algorithm.

#### 4.1. Data Introduction

In this experiment, KOBERT was pre-trained and news data were used for category classification. This data was collected from 20 September to 21 October on the Kotra Overseas Market News, Kita Korea International Trade Association's Overseas Market News site. This is the daily news data, with a total of 20,756 data points. Data preprocessing and labeling eventually led to the use of 6052 data points. Data variables before categorization include each unique number, ID, publication date, title, content containing the body of the news, URL, and the address of the news article. Through the KOBERT model, the category classification process adds the kan-code variable, a category classification number contained in each news article, to the news data. After labeling, the kan-code ranges from 0 to 2, where 0 means clothing, 1 means cosmetics, and 2 means electronics.

#### 4.2. Determining Categories through KOBERT

The category can be determined by entering the body or title of the news using KOBERT. For model learning, we used the labeled data above, including news titles, news bodies, and categories. If the news title and news body data are entered into the learned KOBERT model, the category is determined. Table 3 shows an example of the data used for the KOBERT recommendation system. If "Continuation of Anti-Dumping Regulations on Large Washing Machines Made in the U.S. and China," is entered as the title, the category of the news was determined to be electronics, and the reliability was high at 0.9802. If the main content of the news is regarding an online women's clothing platform that reduces the hassle of visiting a clothing store, and the body of the news states, "The advantage is that you can select your favorite clothes with a few clicks on your phone...", then the category of the news was determined as clothing, and the reliability was also high at 0.9876.

**Table 3.** An example of news data used in the KOBERT recommendation system.

Title	Content
Russian notebook market drops 0.15 due to supply shortages and price increases	Last year, notebook sales surged as telecommuting and online classes were activated due to the influence of COVID-19
Changes in the U.S. Employment Market and Successful Job Strategies After COVID-19	The U.S. job market should prepare a job visa strategy along with the COVID-19 shock recovery employment strategy
2021 New York Comic Con Observation	The largest entertainment fair in the eastern United States will be held successfully
The automotive parts manufacturing industry has been on the decline for the past two years	One of India's key industries, the auto parts manufacturing sector, is sluggish

#### 4.3. Establishment of Recommendation System

Using our method, the news data entered in KOBERT is added to the 0th category data index after category determination. We embed and vectorize the news body content based on data. Through the TF-IDF and cosine similarity test, the news that is most relevant to the news entered is extracted as recommended news. Only the title or the text of the



news content can be used to extract relevant top news. It is also possible to extract relevant top news using both the title and the body of the news simultaneously. Top news can be organized at a glance with a data frame. Table 4 shows the top 5 news data.

In the case of this study, we divided the approaches into three categories. First, the recommended news is extracted from the entire news data without categorization. Second, the KOBERT model is used to categorize news data into clothing, cosmetics, and electronics categories and extract recommended news. Third, we extract recommended news by combining different weights of the BERT and TF-IDF models.

**Table 4.** Top five news data frames related to user-read news.

ID	Watched News	First Recommendation	Second Recommendation	Third Recommendation	Fourth Recommendation	Fifth Recommendation
16861	EU Summit Agrees 4th Russian Sanctions Focused on Trade and Energy	EU to consider additional Russian sanctions (draft)	EU sanctions on Russia's crude oil, agreement stalled on Hungary opposition	EU steel industry urges extension of safeguard and introduction of carbon tax	EU carbon border, focus on steel, cement and power	EU to self-sufficiency in EV fuel cells by 2025
14873	Regional Gross Domestic Product (GRDP) and the Impact of COVID-19 in the Third Quarter of Vietnam's Three Big Cities	The 4th Korean Wave in Japan through interviews	Japan's demand for home appliances plunges recently due to COVID-19	[Kim Seokwoon's Vietnam News] Forecast of economic recovery	4 Big Chinese Smartphone Companies Tensions Over India's Spread of COVID-19	China's production and sales of new energy vehicles increased 1.1 times in the January-April period
14531	Semiconductor stabilizes in the second half of next year... Emergency response, such as finding substitute products	Is the supply and demand of semiconductors getting longer... the sense of touch in the automobile industry	Automotive semiconductors will continue to lack next year	Japan Semiconductor Giant's Fire Forecasts 1.6 Million Automobile production	The semiconductor crisis is stopping the production of cars... have no countermeasures	Semiconductor investment... Why are they flocking to America?
10284	[Korea Trade Insurance Corporation] Mubo holds export consulting week... Help novice export companies advance overseas	[Korea Trade Insurance Corporation] Korea Trade Insurance Corporation will provide special support for overseas expansion of excellent companies in safety management activities	Korea Posts Trade Insurance Incentives to Outstanding ESG Pilot Assessment Companies	[Ministry of Trade, Industry and Energy] The Ministry of Industry strengthens safety management of outdoor exercise equipment and electric board products	MubO, 4.8 million buyers' credit information will be released free of charge by the end of the year	The government imports 20,000 tons of urea water from Australia... Measures to stabilize long-term supply and demand are also in place

Table 4. Cont.

ID	Watched News	First Recommendation	Second Recommendation	Third Recommendation	Fourth Recommendation	Fifth Recommendation
3405	Japan's economic weekly trend (2.22 to 2.28)	Japan's economic weekly trend (1.11 to 1.17)	U.S. demand for women's skincare products rises	Japan's economic weekly trend (11.2 to 11.8)	Japan's economic weekly trend (12.28 to 1.10)	[Vietnam Market News] Vietnam saw a 4.5-fold increase in the number of overseas entrants between January and May this year

#### 4.4. The First Approach

In the first approach, recommended news is extracted from the total news data that is not categorized, and the body and title of the news are received as input values. After tokenizing this, relevant news is extracted from the entire news data using TF-IDF and cosine similarity. Figure 3 illustrates the first approach.



Figure 3. Overview of the first approach.

#### 4.5. The Second Approach

In the second approach, using the KOBERT model, the total news is classified into three categories: clothing, cosmetics, and electronics. The body and title of the news are received as input values. The news is then received as the input value is also categorized using the KOBERT model. Table 5 shows the kan-code labeled data. If the news is determined to be clothing, TF-IDF and the cosine similarity are used within clothing news to extract related news. Figure 4 illustrates the second approach. For the second approach, unlike the first approach, KOBERT involves determining categories.

Table 5. An example of kan-code labeled data using the KOBERT.

Content	Kan-Code
KAgriTech, AgroTech's entry into France should prioritize understanding the local ecosystem	2
Invest Chile, Sept. 16 Webinar on how to efficiently proceed with investment projects in Chile	2
It oversees all of Thailand's largest corporate group companies, which account for 0.16 of the market capitalization of listed companies in Thailand	2
Ecuador Airport Can Play a Geopolitically Latin American Hub to Sell Some Airport Operating Rights	2

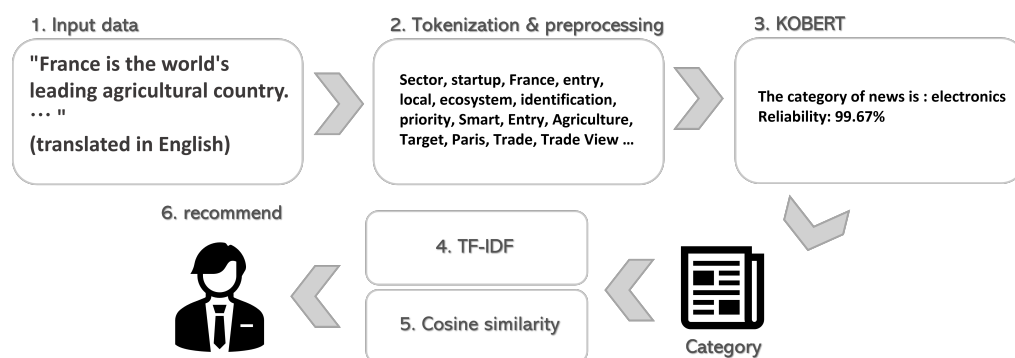


Figure 4. Overview of the second approach.

#### 4.6. The Third Approach

For the third approach, the recommended news is extracted using a method of combining the two models by different weights ( $a \times \text{TF-IDF} + (1 - a) \times \text{BERT}$ ), where  $a$  is from 0 to 0.9. TF-IDF is the same as used in the two approaches described above. BERT uses KOBERT as the second approach to the Bert-as-service model. BERT's shape was (1768), and TF-IDF's shape was dimensionally reduced to (1,651,517) to proceed with the concat. A cosine similarity test is performed on the combined vector and the relevant news is extracted. Figure 5 illustrates the third approach.

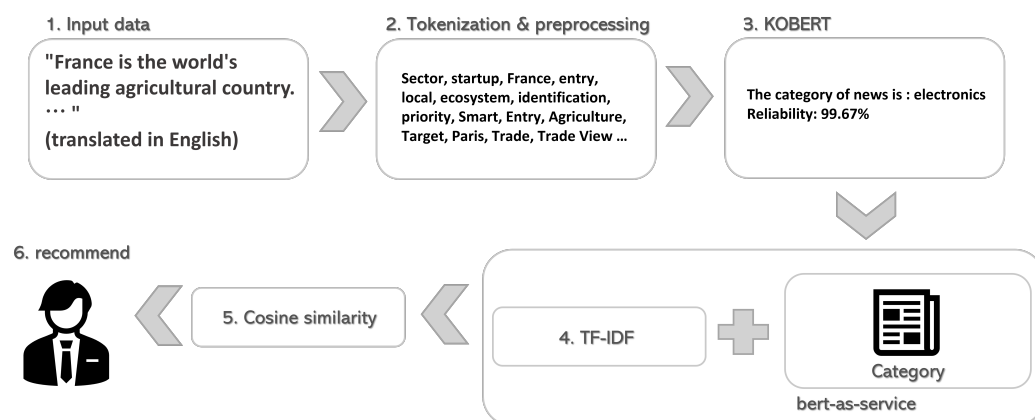


Figure 5. Overview of the third approach.

## 5. Results

The purpose of the study is to evaluate the performance of recommendations by balancing dense and sparse vectors. The following results are based on approaches 1, 2, and 3. In the case of the news, we used “Korea Cosmetics Buyer Interview in Nigeria” as an example.

### 5.1. Recommended News

We used the title and body of the “Korea Cosmetics Buyer Interview in Nigeria” news as input values. Table 6 shows the recommendation results based on the first approach, second approach, and third approach. For example, only the title of the news is extracted, and the body of the news can be extracted in the same way.

**Table 6.** Recommended news corresponding to Approach 1, Approach 2, and Approach 3.

	Watched News	First Recommendation	Second Recommendation	Third Recommendation	Fourth Recommendation	Fifth Recommendation
Approach 1	Nigerian cosmetics buyer interview	Argentine Lubricant Distributor Interview	In order to succeed in the Chinese cosmetics market, it needs to be both marketing and quality	Japanese resource developers ponder whether to continue Russian business	China's Online Market Is Now a Customized Consumption Era	Air fares will rise following maritime fares in the event of a prolonged Suez Canal accident
Approach 2	Nigerian cosmetics buyer interview	In order to succeed in the Chinese cosmetics market, it needs to be both marketing and quality	Growing Ayurveda Market in India	K-beauty in Montenegro? Maxima Pharmacy in-depth interview	Indonesian palm oil export ban until when..a local farmer's tear	Why did the online fashion giant enter the cosmetics market?
Approach 3	Nigerian cosmetics buyer interview	[Patent Office] Fostering future industrial talents armed with ideas and intellectual property	EU talks suspension of bilateral trade deal over UK-Northern Ireland trade row	Impact of Russia's Ukraine Crisis on Uzbekistan's Economy	Interview with founder of Beeducation Adventures, an online education company in Malaysia	[Malaysia] Announcement of a total export ban on chicken and an extension of the price cap system

### 5.2. Evaluation

For our evaluation, we collected the top 20 recommended news, one by one, in order of approach. Relevance annotation was performed on all collected news and the relevance results were secured for news obtained through the n models. This is because there is no set of correct answers to compare, so if more than two out of three people are determined to be liberated, it is determined to be relevant. Table 7 shows the relevant results for the news data. The third approach was divided into four models: bert 1.0 + tf-idf 0.0, bert 0.7 + tf-idf 0.3, bert 0.5 + tf-idf 0.5, and bert 0.3 + tf-idf 0.7. For the evaluation, *Precision@N*, Mean Reciprocal Rank (*MRR*) was used. *Precision@N* refers to the percentage of the *N*-recommended items that users are interested in, and the formula is written as follows:

$$precision @N = \frac{\text{items of interest to the user}}{\text{total recommended items}(N)} \quad (3)$$

Mean Reciprocal Rank (*MRR*) focuses on the location of where preferred items are on the list. The higher the value of the Mean Reciprocal Rank, the higher the user's preferred item is in the list. The formula is written as follows:

$$MRR = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{1}{k_u} \quad (4)$$

**Table 7.** News data relevance results.

Category	Title	Recommendation-Title	Recommendation-Content	Evaluation
Clothing	Retail market trends and online shopping patterns after COVID-19 seen by Taiwanese think tanks	How to prepare for the holiday season amid huge job shortages in the U.S. retail industry?	Consumer spending expected to rise this year's holiday shopping season following the previous year	Relevant
Clothing	Retail market trends and online shopping patterns after COVID-19 seen by Taiwanese think tanks	Use 200% of Shopee, an online shopping platform in Southeast Asia	Shopee, an online shopping mall in Southeast Asia	Relevant
Clothing	Retail market trends and online shopping patterns after COVID-19 seen by Taiwanese think tanks	Let's take the first step in exporting to Japan on a crowdfunding platform!	Japan's leading crowdfunding company MAKUAKE enters Korea	Irrelevant

Table 8 shows the Precision @N, Mean Reciprocal Rank evaluation for each model with the clothing category. Furthermore, Tables 9 and 10 show the Precision @N, Mean Reciprocal Rank evaluation for each model with cosmetics and electronics, respectively.

**Table 8.** Overall average evaluation results combined with detailed evaluation results for each model whose category is clothing.

	Model A	Model B	Model C (1.0 + 0.0)	Model C (0.7 + 0.3)	Model C (0.5 + 0.5)	Model C (0.3 + 0.7)
P@1	0.1	0.1	0	0	0.1	0.2
P@5	0.08	0.12	0	0	0.08	0.24
P@10	0.06	0.07	0	0.02	0.1	0.17
P@20	0.055	0.075	0.01	0.04	0.1	0.16
MRR	0.22755	0.26012	0.00667	0.02926	0.15671	0.29778

If the category is clothing, the approach shows the highest accuracy in the third approach. Among the methods of the third approach, the method that weighted 0.3 for BERT and 0.7 for TF-IDF was the most accurate. Mean Reciprocal Rank, an evaluation technique that assesses the position of the user's preferred news in the list, ranked as the third approach among all approaches at 0.29778.

**Table 9.** Overall average evaluation result combined with detailed evaluation results for each model whose category is cosmetics.

	Model A	Model B	Model C (1.0 + 0.0)	Model C (0.7 + 0.3)	Model C (0.5 + 0.5)	Model C (0.3 + 0.7)
P@1	0.2	0.5	0	0.1	0.3	0.5
P@5	0.12	0.42	0	0.22	0.3	0.28
P@10	0.08	0.42	0.02	0.26	0.4	0.54
P@20	0.06	0.435	0.03	0.185	0.41	0.51
MRR	0.22054	0.69167	0.02226	0.26699	0.42889	0.57222

If the category is cosmetics, the approach shows the highest accuracy in the third approach. Based on the data, it was found that cosmetic keywords appeared frequently in both the title and the text content in the cosmetics category news. Among the methods of the third approach, the one that weighted 0.3 for BERT and 0.7 for TF-IDF was the

most accurate. In Mean Reciprocal Rank, an evaluation technique that focuses on where your preferred news is located in the list, the second of all approaches was the highest at 0.69167. Among the third approaches, BERT was weighted at 0.3 and TF-IDF at 0.7, followed by 0.57222.

**Table 10.** Overall average evaluation results combined with detailed evaluation results for each model whose category is electronics.

	Model A	Model B	Model C (1.0 + 0.0)	Model C (0.7 + 0.3)	Model C (0.5 + 0.5)	Model C (0.3 + 0.7)
P@1	0.4	0.3	0.3	0.1	0.2	0.5
P@5	0.26	0.38	0.3	0.08	0.2	0.3
P@10	0.21	0.38	0.22	0.16	0.28	0.36
P@20	0.18	0.395	0.19	0.14	0.23	0.33
MRR	0.5	0.55333	0.35033	0.1781	0.36087	0.56825

If the category is electronics, the approach shows similar accuracy in the second and third approaches. Among the third approach, the method that weighted 0.3 for BERT and 0.7 for TF-IDF had the highest accuracy. Mean Reciprocal Rank was the highest in the third approach of all approaches at 0.56825.

When comparing both accuracy and Mean Reciprocal Rank according to the approach in all categories, the highest evaluation results were obtained when weighted at 0.3 for BERT and 0.7 for TF-IDF among the third approaches.

## 6. Conclusions

In this paper, we proposed an ensemble-based K-wave item recommendation system that understands the latent information between items by exploiting dense vectors using a variant of BERT. We demonstrated that the proposed ensemble approach outperforms the baselines, including individual sparse and dense vector-based models. We attribute this result to the positive effect of the category classifier as the users would like to see more uniform results to support their final decision-making. Instead of using the kan-code explicitly, we determine the items' category with the BERT-based classifier, promising the latent information considered for category classification.

One limitation of the proposed model is that the limited categories: clothing, cosmetics, and electronic devices. Although those items represent the K-wave, further investigation with varied categories could provide more insight to exporters for decision-making.

Another limitation is that our model focuses primarily on news categories, and does not consider various ways of delivering information in other media formats. Therefore, further studies will consider a wider information ecosystem, including news as well as other data sources such as social media data. This will help users better understand the diversity of information and the interaction of personalized recommendations.

To overcome these limitations, we will leverage more data and various sources of information, and further studies will carry out more comprehensive experiments and analyses. We expect this to provide practical insights to enhance the understanding of the effectiveness and diversity of personalized recommendation systems and to improve the user experience. Nonetheless, the proposed model would potentially contribute to making a risk-alleviated business decision for small-sized exporters who have limited access to big data analytics by recommending the potentially trending K-items and their target countries throughout various text analytic approaches.

This study has shown that recommendation performance can be improved by exploiting the ensemble approach including dense and sparse vectors altogether. It also suggests that the dense-vector-based classifier can be used to recommend two different items, as the latent information between two items is learned. We acknowledge that more advanced transformer-based approaches that fully reflect the latent information between



different items can further improve the recommendation performance. We consider this as a potential area for future research.

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

**Funding:** This work was supported by the Technology Innovation Program (No. 20022899, Development of AI-based smart manufacturing process and equipment technology to strengthen the competitiveness of semiconductor materials, parts, and equipment) funded by the Ministry of Trade, Industry and Energy (MOTIE, Republic of Korea).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the fact that commercialization of the research is in progress.

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

## References

1. Johnson, J. *The K-Wave and Its Impact on the South Korean Economy*; Ouachita Baptist University: Arkadelphia, AR, USA, 2023.
2. Ullah, R.; Zeb, A.; Kim, W. The impact of emotions on the helpfulness of movie reviews. *J. Appl. Res. Technol.* **2015**, *13*, 359–363. [[CrossRef](#)]
3. Ju, H. *The Korean Wave and Korean Dramas*; Oxford Research Encyclopedia of Communication: Oxford, UK, 2018.
4. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
5. Li, B.; Zhao, Q.; Jiao, S.; Liu, X. DroidPerf: Profiling Memory Objects on Android Devices. In Proceedings of the 29th Annual International Conference on Mobile Computing and Networking, Madrid, Spain, 2–6 October 2023; pp. 1–15.
6. Xu, Y.; Jiang, Y.; Zhao, X.; Li, Y.; Li, R. Personalized Repository Recommendation Service for Developers with Multi-modal Features Learning. In Proceedings of the 2023 IEEE International Conference on Web Services, Chicago, IL, USA, 2–8 July 2023; pp. 455–464.
7. Zhan, Z.; Xu, B. Analyzing review sentiments and product images by parallel deep nets for personalized recommendation. *Inf. Process. Manag.* **2023**, *60*, 103166. [[CrossRef](#)]
8. Liu, Z.; Xu, W.; Zhang, W.; Jiang, Q. An emotion-based personalized music recommendation framework for emotion improvement. *Inf. Process. Manag.* **2023**, *60*, 103256. [[CrossRef](#)]
9. Guo, Y.; Chen, Y.; Xie, Y.; Ban, X. An effective student grouping and course recommendation strategy based on big data in education. *Information* **2022**, *13*, 197. [[CrossRef](#)]
10. Lee, H.; Jeong, O. A personalized exercise recommendation system using dimension reduction algorithms. *J. Korean Soc. Comput. Inf.* **2021**, *26*, 19–28.
11. Kim, H.; Shin, W.; Shin, D.; Kim, H.; Kim, H. Beauty Product Recommendation System using Customer Attributes Information. *Inf. Syst. Rev.* **2021**, *23*, 69–84.
12. Yun, H.; Choi, K. A research of food menu recommendation system based on personal preference. *Hotel Manag. Res.* **2020**, *29*, 83–100.
13. Amir, N.; Jabeen, F.; Ali, Z.; Ullah, I.; Jan, A.; Kefalas, P. On the current state of deep learning for news recommendation. *Artif. Intell. Rev. Int. Sci. Eng. J.* **2023**, *56*, 1101–1144. [[CrossRef](#)]
14. Zhang, Z.; Wang, B. Prompt learning for news recommendation. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, Taipei, Taiwan, 23–27 July 2023.
15. Gharahighehi, A.; Vens, C. Diversification in session-based news recommender systems. *Pers. Ubiquitous Comput.* **2023**, *27*, 5–15. [[CrossRef](#)]
16. Kim, D.; Yoon, S.; Chae, S.; Yoo, S. A News Recommendation System using Text Mining and LOD Techniques. *J. Korean Inst. Inf.* **2022**, *20*, 1–6.
17. Jung, I.; Kim, B.; Kim, S.; Yoo, G. News Recommendation System Based on Text and Image Tag Data. *J. Digit. Content Soc.* **2020**, *21*, 479–486. [[CrossRef](#)]
18. Hong, M.; Oh, K.; Ka, M.; Cho, G. Social network-based content recommendation method for personalized news service. *Intell. Inf. Res.* **2013**, *19*, 57–71.
19. Kim, D. Similar content recommendation model based on content metadata using language model. *Intell. Inf. Res.* **2023**, *29*, 27–40.

20. Bhaskaran, S.; Marappan, R. Enhanced personalized recommendation system for machine learning public datasets: Generalized modeling, simulation, significant results and analysis. *Int. J. Inf. Technol.* **2023**, *15*, 1583–1595. [[CrossRef](#)]
21. Kim, J.; Kim, H.; Kang, W.; Hong, J. Development of a Craft Recommendation System Using Content-Based Filtering. In *Korean Society of Information Science and Technology Academic Papers*; Academic Press: Seoul, Republic of Korea, 2022; pp. 1489–1491.
22. Behera, G.; Nain, N. Collaborative Filtering with Temporal Features for Movie Recommendation System. *Procedia Comput. Sci.* **2023**, *218*, 1366–1373. [[CrossRef](#)]
23. Hwang, S.; Kim, D.; Kim, E.; Lee, Y.; Eom, Y.; Lee, J. Recommender System for the Hobby-Activity-Item Rental Platform using User-based Collaborative Filtering. *J. Digit. Contents Soc.* **2022**, *23*, 985–993. [[CrossRef](#)]
24. Kim, D.; Shin, W.; Han, G.; Lee, J.; Moon, K.; Lee, S.; Han, S.; Kwon, H.; Han, S. A system for recommending university liberal arts courses using collaborative filtering. In *Proceedings of the Fall Conference of the Industrial Engineering Society of Korea*, Seoul, Republic of Korea, 3–7 November 2020; pp. 2551–2556.
25. Nguyen, L.V.; Nguyen, T.; Jung, J.; Camacho, D. Extending collaborative filtering recommendation using word embedding: A hybrid approach. *Concurr. Comput. Pract. Exp.* **2023**, *35*, e6232. [[CrossRef](#)]
26. Walek, B.; Fajmon, P. A hybrid recommender system for an online store using a fuzzy expert system. *Expert Syst. Appl.* **2023**, *212*, 118565. [[CrossRef](#)]
27. Jomsri, P.; Prangchumpol, D.; Poonsilp, K.; Panityakul, T. Hybrid recommender system model for digital library from multiple online publishers. *F1000Research* **2023**, *12*, 1140. [[CrossRef](#)]

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