applied sciences

MDPI

Article

Methodology of Labeling According to 9 Criteria of DSM-5

Geonju Lee [1], Dabin Park [2] and Hayoung Oh [3,*]

[1] Department of Mathematics, Sungkyunkwan University, Seoul 03063, Republic of Korea; dlrjswn179@naver.com
[2] Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul 03063, Republic of Korea; dabin3178@g.skku.edu
[3] Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul 03063, Republic of Korea
* Correspondence: hyoh79@skku.edu; Tel.: +82-2-583-8585

Abstract: Depression disorder is a disease that causes a deterioration of daily function and can induce thoughts of suicide. The Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5), which is the official reference of the American Psychiatry Association and is also used in Korea to identify depressive disorders, sets nine criteria for diagnosing depressive disorders. The lack of counseling personnel, including psychiatrists, and negative social perceptions of depressive disorders prevent counselors from being treated for depressive disorders. Natural language processing-based artificial intelligence (AI) services such as chatbots can help fill this need, but labeled datasets are needed to train AI services. In this study we collected data from AI Hub wellness consultations and crawls of the Reddit website to augment and build word dictionaries and analyze morphemes using the Kind Korean Morpheme Analyzer and Word2Vec. The collected datasets were labeled based on word dictionaries built according to nine DSM-5 depressive disorder diagnostic criteria.

Keywords: DSM-5; natural language processing; morphological analysis; multi labeling; embedding

check for updates

Citation: Lee, G.; Park, D.; Oh, H. Methodology of Labeling According to 9 Criteria of DSM-5. *Appl. Sci.* **2023**, *13*, 10481. https://doi.org/10.3390/app131810481

Academic Editors: István Vassányi, István Kósa and László Balkányi

Received: 28 July 2023
Revised: 7 September 2023
Accepted: 18 September 2023
Published: 20 September 2023

1. Introduction

Depressive disorder is a disease that causes various cognitive and mental and physical symptoms, resulting in a decrease in daily function. Severe depressive disorder can lead to suicidal thoughts as well as repetitive thoughts of death 194 [1]. The prevalence of depressive disorders in Korea is 36.8%, and four of every 10 people in Korea suffer from depressive disorders or depression, which is the highest among OECD countries 194 [2].

The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5) is the official reference of the American Psychiatric Association (APA). The DSM-5 is also used around the world, and depressive disorders are included in its classification of mental disorders. Depressive disorders are determined by comprehensively considering the nine diagnostic criteria of the DSM-5, psychiatrists' clinical experiences, clinical interviews, psychological test results, and treatment progress.

Fear caused by negative social perceptions of mental disorders can prevent counselors, including psychiatrists, from being treated for depressive disorders. In addition, it can be difficult to intervene in a timely manner due to the lack of infrastructure to cope with depressive disorders when they are accompanied by a physical disease [3]. Accordingly, natural language processing-based artificial intelligence (AI) services that can be consulted in a timely manner are emerging [4,5]. Among them, conversational AI agents known as chatbots can help predict mental illness conditions by consulting a counselor's textual records.

The term chatbot refers to a robot that can converse with users. Natural-language processing and machine learning are essential for chatbots to carry on meaningful and complex conversations. In this case, supervised machine learning is the primary tool, and it is essential to label datasets to ensure correct answers are used during training sessions. In addition, because perceived empathy with users can lead to a positive evaluation of a chatbot's responses, it is effective to use empathy datasets [6].

This paper labels an empathetic dataset according to DSM-5 depressive disorder diagnostic criteria. A wellness consultation dataset provided by AI Hub [7] and a dataset obtained by crawling the depression subreddit channel of the Reddit internet forum were used. In addition, the corresponding datasets were classified and labeled through the use of a word dictionary that classifies words expressing the elements of depressive disorders according to nine criteria.

Artificial intelligence is becoming a part of society, and human–computer interaction technology is developing that automatically detects human emotions or gives appropriate responses to questions or utterances through chatbots. [8–12] However, there are currently no labeled datasets according to the DSM-5 depressive disorder diagnostic criteria. DSM-5, which is used in many countries, including Korea, has nine criteria for diagnosing true depressive disorder: 1. depressed mood, 2. loss of interest/pleasure, 3. weight loss or gain, 4. insomnia or hypersomnia, 5. psychomotor agitation or retardation, 6. fatigue, 7. feeling worthless or excessive/inappropriate guilt, 8. decreased concentration and 9. suicidal thoughts. If more than five of the nine symptoms, including one of No. 1 and No. 2, persist for more than two weeks, it is diagnosed as a depressive disorder. Accordingly, this paper labeled the dataset using the DSM-5 depression disorder diagnostic criteria, which are widely used by the medical community in various fields. In medicine, the accuracy of diagnosis can be improved through the use of corresponding datasets and machine-learning models, and in natural-language processing, the criteria can be used in chatbots programmed for depression counseling. For chatbots specifically, empathy datasets and actual counseling datasets can help elicit positive evaluations by revealing an empathetic attitude toward the user.

2. Previous Research

Kang Seung-sik et al. [13] analyzed the vocabulary usage characteristics of the short messaging service (SMS) character corpus, which is a spoken corpus, the Naver movie review corpus, and the Korean written primitive corpus, which is a written corpus. To measure the strength of the performance, a discriminatory performance vocabulary analysis methodology was used, and adjectives that appear mainly in the spoken and written forms were identified. In addition, a word-embedding technique was used to automatically build an emotional vocabulary dictionary based on adjectives with high performance intensity. A total of 343,603 emotional vocabulary dictionaries were automatically constructed using this technique. However, related research has not been able to manually select a large number of automatically constructed professional vocabularies.

Kim et al. [14] proposed a machine learning-based emotional-analysis system that detects user depression through SMS messages. Using an emotion dictionary of words related to depression and other emotions, search keywords were selected from text data on Twitter to build a learning dataset consisting of 1297 sentences that express depression-related emotions and 1032 sentences without them. Finally, a circulatory neural network, short- and long-term memory, and gate circulation units were compared and evaluated. Among them, a model based on gate circulation units achieved an accuracy of 92.2%. However, Kim et al. limited depression-related emotions to "sadness".

Seo et al. [15] collected conversation datasets from 2016 to February 2019 from users who tended to be depressed and those who were not and analyzed the subject and vocabulary characteristics of the dataset. The periods before and after the depression trend were compared, and the depression trend was analyzed by period. In addition, topic modeling, simultaneous word analysis, and emotional analysis were used to identify the characteristics and differences among topics of conversation of users with respect to depression and non-depression trends and to analyze the vocabulary used. However, the Seo et al. team's biased thoughts may have affected the outcome when classifying words associated with depression and non-depression.

Chin et al. [16] collected conversation datasets related to depressed emotions in chats between people and chatbots. They analyzed their data using text-mining techniques to identify conversation topics related to depression-related emotions. In addition, through qualitative analysis, the types of depression that users described when confiding in chatbots were categorized and classified, and differences were identified by comparing them with depression-related data on Twitter.

Płaza et al. [8] developed and validated a method to use machine learning algorithms to detect and classify customer emotions in contact center applications. This enabled companies to obtain useful information that can help improve their customer experience. The advantage of this study was that we develop and test emotion classification algorithms using data collected from real-world voice and text channels, which can help companies improve their customer experience. However, the limitations of this study are the problem of data imbalance and that the performance of some classifiers was low.

Rathnayaka et al. [9] proposed a meta-cognitive technology that leverages artificial intelligence chatbots to provide custom behavior activation and remote health monitoring, proposed a conceptual framework by synthesizing state-of-the-art research and technological advances, and dealt with the design, development, and participatory evaluation of chatbots. The advantage of this paper was that it can provide customized support through chatbots, and the limitations were that it had only been tested in small sample sizes and laboratory environments. In addition, it is necessary to verify its effectiveness in other cultural contexts.

Prottasha et al. [10] dealt with how to improve accuracy using transfer learning in BERT-based emotional analysis, and experiments were conducted using the Bangla language. Experiments showed that combining Bangla-BERT with LSTM yields 94.15% accuracy, and LSTM showed the most important results among the four word-embedding systems. This paper collected data from various sources, including Internet sites, social networking sites, online retailer product reviews, and film and book reviews, and provided useful insights into emotional analysis using transfer learning. However, this study only conducted experiments on the Bangla language, and there were limitations because it used only one language and one dataset. In addition, this study did not conduct a qualitative analysis of the results.

Graterol et al. [11] proposed a general emotion recognition framework that allowed social robots to detect and respond to human emotions using natural language processing (NLP) transformers and emotion ontology, which can improve social robot-to-human interaction and social behavior modeling.

Roca et al. [12] analyzed the performance impact of the slot tagging and training data length of joint natural language understanding models in Spanish drug management scenarios and presented design considerations for better NLU model development, but the study had limitations in that it did not consider the impact of other factors such as training data quality or pre-trained language models.

3. Related Techniques

3.1. Kind Korean Morpheme Analyzer

Morphological analysis refers to classifying sentences into morphemes and grasping the linguistic structure; it can distinguish roots, prefixes, suffixes, and parts of speech. The little morpheme analyzer [17] is a Korean morpheme analyzer that is less sensitive to spacing errors. For speed, an adjacent condition-inspection method using a mood dictionary was used, and for quality, a probability model based on the Heuristic and Hidden Markov Model was used. In the case of a heuristic model, morphemes were analyzed by considering the characteristics of the language, such as the part-verb of the previous word and the part-verb of the next word. Morpheme analysis results were then sorted to ensure the most appropriate information was obtained. A hidden Markov model is a probability model with both hidden and observed states. In the little morpheme analyzer, each morpheme is modeled as a hidden state, and the parts of the morpheme are modeled as observable. Morpheme analysis estimates the probability of the appearance of morphemes. Figure 1 depicts an example of the morpheme analysis process using a little morpheme analyzer.

[‘아버지’, ‘가‘, ‘방‘, ‘에’, ‘들어가’, ‘ㅂ니다‘, ‘.’]
[‘My father’, ‘is’, ‘enter’, ‘–ing’, ‘a room’, ‘.’]
[(‘아버지’, ‘my father’, ‘NNG’),
(‘가’, ‘is’, ‘JKS’),
(‘방’, ‘a room’, ‘NNG’),
(‘에’, ‘in’, ‘JKM’).,
(‘들어가’, ‘enter’, ‘VV’),
(‘ㅂ니다’, ‘–ing’, ‘EFN’),
(‘.’, ‘SF’)]

Figure 1. Example of the Kind Korean Morpheme Analyzer (made by IDS Lab of Seoul National University, Seoul, Republic of Korea). English after Korean is the translation.

3.2. Word2Vec

Word2Vec [18] is a word-embedding technology presented by Google in 2013. Word2Vec distributes the meaning of a word across multiple dimensions, expresses it as a vector, and calculates the similarity between word vectors. At this time, analysis of words that appear in similar contexts assumes that they are similar. In addition, after making the word blank, the researchers guessed the word to place in the blank using the word adjacent to the blank in the sentence and one of two methods. The continuous bag of words (CBOW) methods leaves the surrounding words and middle words blank, and Skip-Gram methods predict surrounding blank words as intermediate words. An example of the Word2Vec expression is shown in Figure 2.

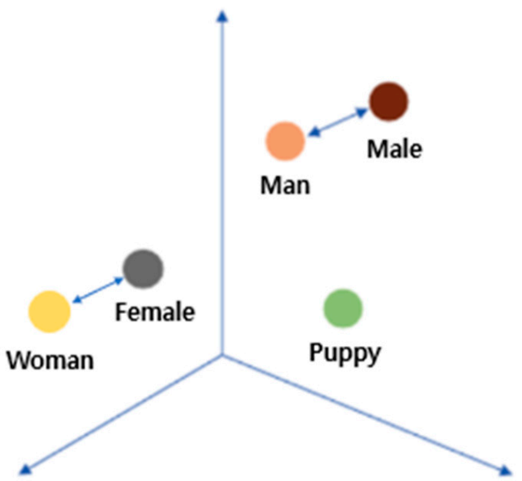

Figure 2. Example of Word2Vec Vector Space. Word2Vec preserves the contextual meaning of words when vectorizing them. For example, the distance between Man and Male is similar to the distance between Woman and Female.

3.3. Dataset Labeling and Classification

Dataset labeling provides information about the data, and dataset classification assigns data to groups. In general, machine-learning models use labeled datasets to solve classification problems. For machine-learning models to learn efficiently and perform predictions, high-quality labeled datasets are required. Datasets can be labeled manually or automatically. Manual labeling assigns datasets to experts and other categories and is often highly accurate, but it can be costly, in terms of both money and time. Semi-graphic learning combines hand-labeled and non-labeled data, which can reduce labeling costs at the expense of mislabeled or missing data.

4. Dataset

4.1. Dataset Description

"Wellness counseling data" were created by processing records of mental health counseling at Severance Hospital in Sinchon. Only data related to mental symptoms were used. A total of 5231 questions and 1034 responses were included. The "intent" column has 19 labels: depression, sadness, loneliness, anger, lethargy, emotional control abnormalities, loss, loss of appetite, appetite, insomnia, anxiety, fatigue, guilt, concentration, self-confidence, despair, suicide impulse, and anxiety.

In the wellness counseling data, the average length of a counselor's utterance is 28 characters (minimum 2, maximum 117), and the average length of a counselor's utterance is 31 characters (minimum 5, maximum 67). Figure 3 shows the results of visualizing words that mainly appear in each counselor's speeches and responses as a word cloud.

Figure 3. Wordcloud of 'utterance (2(차) 'column and 'response (공감)' column. In the picture on the left, words such as "Thinking", "People", "Work", "Sleep", "Friend", "Worried", "Mind", "Feeling", "Home", "Day", "My husband", "Other", "First time", "Mom", "Drinking" are drawn in word cloud in the order of the letter size. In the picture on the right, words such as "You", "Heart", "Worried", "Thinking", "People", "Hospital", "Work", "Help", "Concern", "Feeling", "How", "Everything", "A moment", "Situation", and "Stress" are drawn in word cloud.

Data from the r/depression subreddit channel were extracted from Reddit, a mega-social community site in the United States. In this paper, 30,000 posts and comment data from 2010 to 2016 were randomly extracted. In the "depression" subreddit, any violation of the notice and detailed rules will be deleted regardless of the number of recommendations. The notice and detailed rules are shown in Figure 4.

Users of r/depression can only ask for help with depressive disorders for themselves or someone close to them, and all answers should be sympathetic. In addition, answers to posts asking for help are only allowed in public comments, and personal answers and posts are not allowed. Responses without meaningful content are not allowed, and responders should not pose as a role model or someone with a superior perspective. Meaningless encouragement, promises that cannot be kept, and medical knowledge, advocacy, and opposition to treatments are also not permitted. In addition, it is impossible to certify self-harm methods and self-destructive content, and content that deviates from the subject or goes against the rules should not be encouraged. Finally, it is impossible to request and provide discussions, unauthorized surveys, self-promotion, and money, goods, services, etc.

Community Basics	r/depression Rules
We offer a peer-support space for anyone dealing with a depressive disorder in themselves or someone close to them. Please stay on topic. Depression is both important and difficult to talk about so focus is essential. Posts here need to be support requests specifically related to depression, and comments need to be supportive of the OP. If you want to talk about thoughts or risk of suicide, please post at /r/SuicideWatch. If you've lost someone to suicide, /r/SuicideBereavement is the best community to get support. It might seem that we have a lot of rules, but we've found they're all necessary to maintain as much emotional and physical safety as possible. Most people are surprised by at least some of our policies so please read all of them carefully before jumping in. Please click "report" to let us know of any inappropriate content you see here - we'd like to know and handle it as soon as we can. If your post or comment is not appearing, it may have been removed for a rule violation or it may be stuck in the filter. Please message us and we'll look into it. We are not a crisis service. We can't guarantee an immediate response, and there are times when this subreddit is relatively quiet. This does not mean no one cares. If you need to talk to someone at once, you may want to take a look at the hotlines list from /r/SuicideWatch Note: the Android Reddit app currently has a bug that causes wikis not to work - if links do not open in the app, use your browser to open the links instead	1. Posts must request personal support for depressive disorders in yourself or someone close to you 2. All replies must be empathetic and responsive to OP. No tough love or debate. 3. Offer help ONLY via public comments. No "I'm here to help" posts or PM/chat requests 4. No low-content posts or responses 5. Don't role-model or put yourself above others. No achievement/advice posts or AMAs 6. No general uplifting content of any kind, including "it gets better" or other unkeepable promises 7. Do not give or request clinical advice or advocate for or against treatments & self-help strategies 8. NO methods for any type of self-harm, or validating self-destructive intent 9. Don't encourage rule-breaking or off-topic content 10. No activism, debate or unapproved surveys including casual or general questions and polls 11. Absolutely no self-promotion of any kind, including creative writing or other arts 12. Do not request or offer money, goods, or services

Figure 4. Community basics and rules of r/depression.

4.2. Dataset Preprocessing

In the case of wellness counseling data, the "utterance" column, which involves counselors' utterances, and the "response" column, which involves counselors' responses, were used. After that, the missing value was deleted. In r/depression data, there is data in which two adjectives are bound by "/". If "/" is contained, there is no clear analysis, so preprocessing is required. The way we came up with it was deletion. Thus, we deleted sentences containing adjectives on either side of a "/", such as "depressed/anxious". An example of a sentence is shown in Figure 5.

depressed/anxious for years and years
I always have been on the depressed/anxious side, and very hard on myself, especially about my looks
what are your experiences with being or having a depressed/anxious parent?
when i am with her i feel more relaxed and less depressed/anxious
maybe i should just focus on getting a simply job so that i can do it even if depressed/anxious
I am depressed/anxious and it is just wearing me down
additionally, depressed/anxious people tend to be sensitive to things that may seem insignificant to others
so for as long as I can remember i have been having depressed/anxious/suicidal feelings almost all the time
I get depressed/anxious on sundays/mondays really bad
sometimes i feel depressed/anxious and can't stop thinking about the future and what i have it do and what i can't do

Figure 5. r/depression data examples including "depressed/anxious".

Translations were made using the Google Translate API. An example of an original text and the translation result is shown in Figure 6.

I don't have an addictive personality, so I've been adding tramadol to my antidepressants for the past 2 years.
I look drunk sometimes.
But over the last 28 months I've been under high stress, I've been fast at it, and I've excelled at it.
I had problems driving. It probably has more to do with something that drains all my energy.
It changed my life.
Tramadol works after a few hours and lasts about 12 hours.
You can get it online.
Similar to opioids, but with far fewer side effects, no headaches or dry mouth.
The big downside is that it's addictive.
I think it's really a bitch that something so useful and works so well and so fast is addictive.

나는 중독성 성격이 없어서 지난 2년 동안 항우울제에 트라마돌을 보중하고 있습니다.
나는 가끔 술에 취해 보인다
하지만 지난 28개월 동안 나는 높은 스트레스를 유지했고 일을 빠르게 진행했으며 그 일에 탁월했습니다
나는 운전에 문제가 있었습니다. 그것은 아마도 내 모든 에너지를 소모하는 일과 더 관련이 있을 것입니다.
그것은 삶을 변화시켰습니다.
tramadol은 몇 시간 후에 작동하고 약 12 지속됩니다.
당신은 그것을 온라인으로 얻을 수 있습니다
아편류와 유사하지만 부작용이 훨씬 적고 머리가 흐릿하거나 입이 마르는 등의 현상이 없습니다.
중독성이 있다는게 큰 단점
너무 유용하고 잘 작동하고 빠르게 작동하는 것이 중독성이 있다는 것은 정말 개년이라고 생각합니다.

Figure 6. Translation examples of r/depression data.

5. Dictionary of Expression Words

A word dictionary was created based on the nine DSM-5 diagnostic criteria for depressive disorder. Medical staff working in the Department of Mental Health Medicine collected expressions for each of the nine depressive disorder criteria and generated a dictionary, as shown in Figure 7.

1. Depressed mood most of the day, nearly every day.
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day.
3. Significant weight loss when not dieting or weight gain, or decrease or increase in appetite nearly every day.
4. Insomnia or hypersomnia, nearly every day.
5. A slowing down of thought and a reduction of physical movement (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt nearly every day.
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day.
9. Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

Figure 7. DSM-5 diagnosis criteria for depressive disorder. Red letters are the key content in each criterion.

The number of words in the word dictionary for each criterion is 59, 45, 44, 33, 37, 27, 45, 26, and 47. Figure 8 provides an example of the first of the nine criteria: "most of the day, almost daily depressed mood".

1. Most of the day, nearly every day; may be subjective (e.g. feels sad, empty, hopeless) or observed by others (e.g. appears tearful); in children and adolescents, can be irritable mood
a feeling of depression (sadness, despair, a feeling of helplessness, a sense of worthlessness)
Be hopeless
Heart aches
Be worthless

Figure 8. Example of the word dictionary of the first DSM-5 diagnosis criteria for depressive disorder.

6. Methodology

6.1. Morphological Analysis

Each sentence of the preprocessed data was analyzed using a little morpheme analyzer. The preprocessed word dictionary was also morphemically analyzed and stored for each of the nine criteria. An example of each result is shown in Figure 9.

[('아이', 'a baby', 'NNG'),
('가지', 'have', 'VV'),
('고', '-ing', 'ECE'),
('나', 'after', 'JKM'),
('서', 'cannot be translated, just an end of a word', 'JKM'),
('우울하', 'feel depressed', 'VV'),
('어', 'cannot be translated, just an end of a word', 'ECS')]

[('덧', 'meaning', 'XPN'),
('없다', 'less', 'UN')]

Figure 9. Results of morphological analysis of datasets and word dictionaries. English next to Korean is the translation of Korean.

6.2. Morphological Screening

Korean is an agglutinative language. In an agglutinative language, one morpheme performs a function; natural language processing in Korean is usually carried out on a morpheme basis. Korean morphemes are divided into real morphemes and formal morphemes depending on the presence or absence of lexical meanings. Morphemes with lexical meanings are called real morphemes, and morphemes with grammatical meanings are called formal morphemes. To extract only words with specific meanings, we selected the noun, the proverb, the root, and the adverb corresponding to the real morpheme. Among the formal morphemes, prefixes, which are morphemes that give new meanings to vocabulary, were also added and selected. The morpheme analysis results of the DSM-5 word dictionary can be found in Figure 10. These were applied to the part-time tag table of the little morpheme analyzer (Figure 11).

6.3. Data Classification

Each sentence of the consultation dataset was searched to determine if the words were included in the dictionary, and automatic classification was performed accordingly in a cloud-based Jupiter Notebook (made by Jupyter, New York, NY, USA) environment using a Tesla T4 (made by NVDIA, Santa Clara, CA, USA) graphical processing unit accelerator. An example of the classification process is shown in Figure 12.

1. [['우울', '기분', '슬프', '절망감', '무력감', '무가', '치감']
1. [['depressed', 'mood', 'sad', 'despair, 'feeling of helplessness', 'feeling of', 'worthlessness']
2. [['가라앉'], ['감정', '안', '느껴지'], ['감흥', '없']
2. [['sink'], ['can, 'not', 'feel'], ['no', 'interest']
3. [['굶'], ['꾸역꾸역'], ['끼니', '거르'], ['다이어트'], ['더부룩']
3. [['starve'], ['pig out'], ['skip', 'a meal'], ['diet'], ['bloated']
4. [['수면', '초기', '중기', '말기'], ['깊', '못', '자']
4. [['early', 'mid-', 'late-', 'sleep'], ['not', 'sleep', 'soundly']
5. [['사람', '눈치', '채', '정도', '평소', '말', '행동', '느리', '지']
5. [['human', 'become', 'aware of', 'degree', 'usual', 'speak', 'act', 'slow', 'delay']
6. [['전반적', '이', '신체증'], ['갈리'], ['고되'], ['근육통']
6. [['overall', 'be', 'body sym'], ['wear'], ['tough'], ['muscle ache']
7. [['나', '때문', '이'], ['나', '자신', '실망'], ['나', '가망', '없']
7. [['be', 'because of', 'me'], ['I', 'disappoint', 'myself'], ['I', 'no', 'hope']
8. [['결정', '장애'], ['결정', '힘들'], ['계속', '힘들']
8. [['decision', 'paralysis'], ['difficult', 'decide'], ['still', 'tired']
9. [['자살', '또는', '자해'], ['고통', '벗어나', '싶']
9. [['suicide', 'or', 'self-injury'], ['want', 'get away', 'pain']

Figure 10. Results of morphological analysis of word dictionaries. The English below Korean is the translation of Korean.

Large categoryz	Sejong POS tag		Shim Gwangseop POS tag		KKMA single tag V 1.0					
	tag	explanation	class	explanation	group 1	group 2	tag	explanation	probability tag	storage dictionary
substantive	NNG	common noun	NN	noun	N	NN	NNG	common noun	NNA	noun.dic
	NNP	proper noun					NNP	proper noun		
	NNB	bound noun	NX	bound noun			NNB		NNB	simple.dic
			UM				NNM			
	NR	numeral	NU	numeral		NR	NR	numeral	NR	
	NP	pronoun	NP	pronoun		NP	NP	pronoun	NP	
predicate	VV	verb	VV	verb	V	VV	VV	verb	VV	verb.dic
	VA	adjective	AJ	adjective		VA	VA	adjective	VA	
	VX	auxiliary predicate	VX	auxiliary verb		VX	VXV	auxiliary verb	VX	
			AX	auxiliary adjective			VXA	auxiliary adjective		
	VCP	affirmative copula	CP	predicative particle 'be'		VC	VCP	affirmative copula, predicative particle 'be'	VCP	raw.dic
	VCN	negative copula					VCN	negative copula, adjective 'not'	VCN	
determiner	MM	determiner	DT	common determiner	M	MD	MDT	common determiner	MD	simple.dic
			DN	numeral determiner			MDN	numeral determiner		
adverb	MAG	common adverb	AD	adverb		MA	MAG	common adverb	MAG	
	MAJ	connective adverb					MAC	connective adverb	MAC	
interjection	IC	interjection	EX	interjection	I	IC	IC	interjection	IC	

Figure 11. Tag table of KKMA.

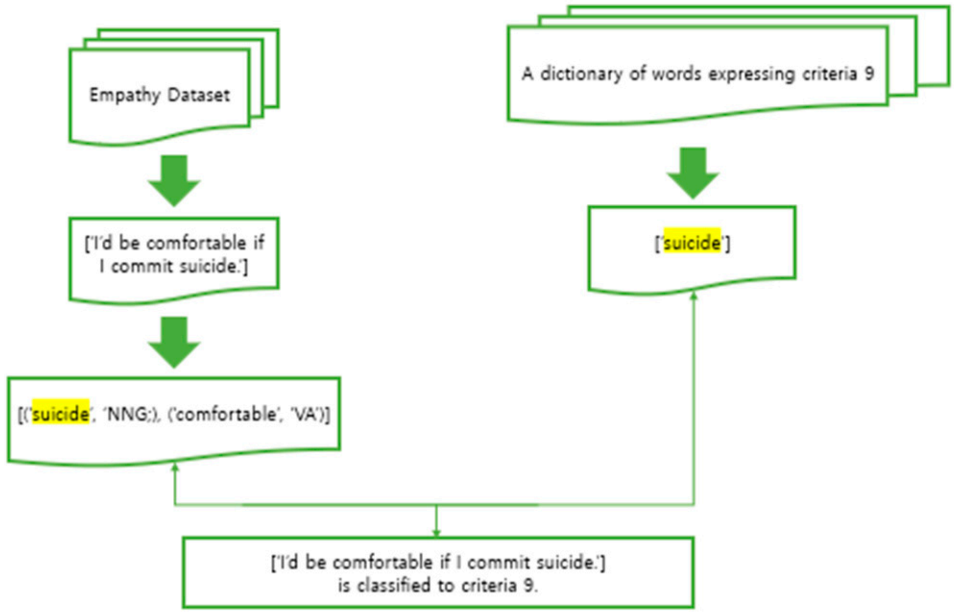

Figure 12. Example of a classification process. The input sentence is classified as criteria 9 because there is a morpheme called 'suicide' highlighted in the input sentence, which exists in the word dictionary corresponding to DSM-5 Depressive Disorder Criteria 9.

6.4. Data Evaluation

In Section 7.1, it can be confirmed that the number of datasets belonging to diagnostic criteria 3, 4, 6, 7, and 8 was small. Therefore, we augmented the words, and the whole process of word augmentation is shown in Figure 13.

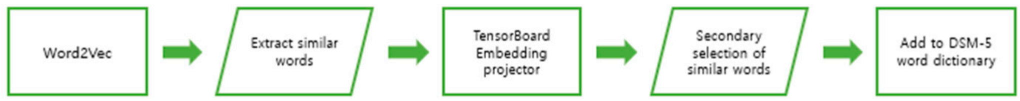

Figure 13. Process of word augmentation.

The similarity of words for the corresponding diagnostic criteria was therefore searched using Word2Vec. The hyperparameters of Word2Vec were set as follows. The larger the window size, the better it is to capture topic or domain information [19]. Therefore, we used five left and right words of the word as context by setting the window to 5. min_count is a parameter that specifies the frequency to be used for learning. We set min_count to 5, the default value of genism, and excluded words that appeared fewer than 5 times from learning. Finally, the CBOW algorithm with a relatively high learning speed was used by designating sg, a parameter that specifies the algorithm, as 0. Word2Vec was used to calculate word similarity and extract words filtered by the set parameters into similar words. Figure 14 shows the results of searching for words similar to "lonely", "terrible".

After that, similar words were visualized using TensorBoard Embedding Projector (www.projector.tensorflow.org, accessed on 20 March 2023), and secondary selection was performed. Figure 15 shows the results of visualizing words similar to 'tough'.

Secondary selected expressions were added to the DSM-5 expression word dictionary to increase the data.

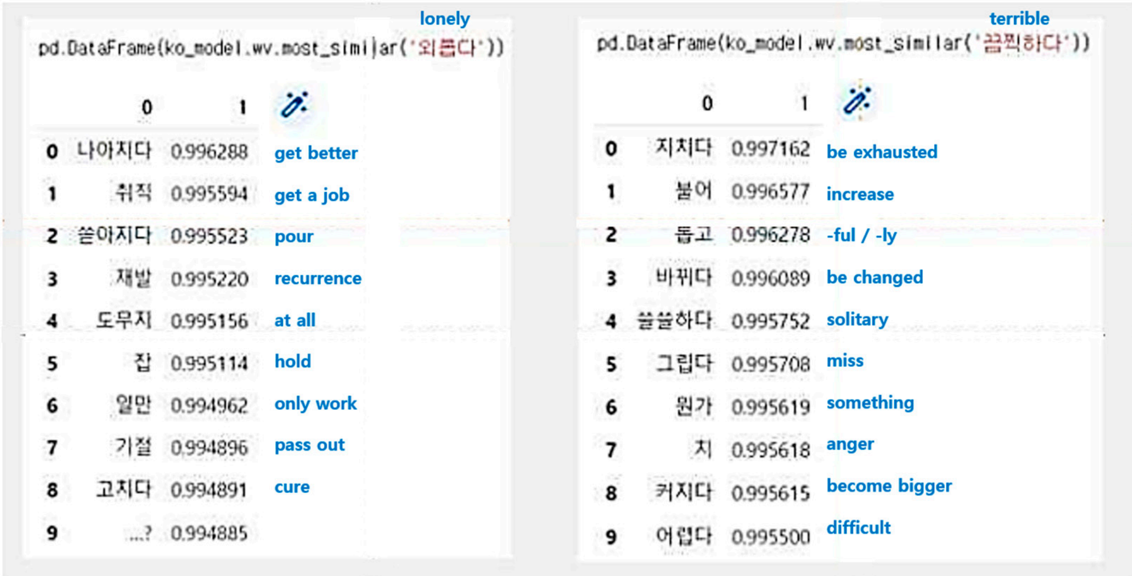

Figure 14. Words similar to "lonely" and "terrible". Translation of Korean is blue-colored.

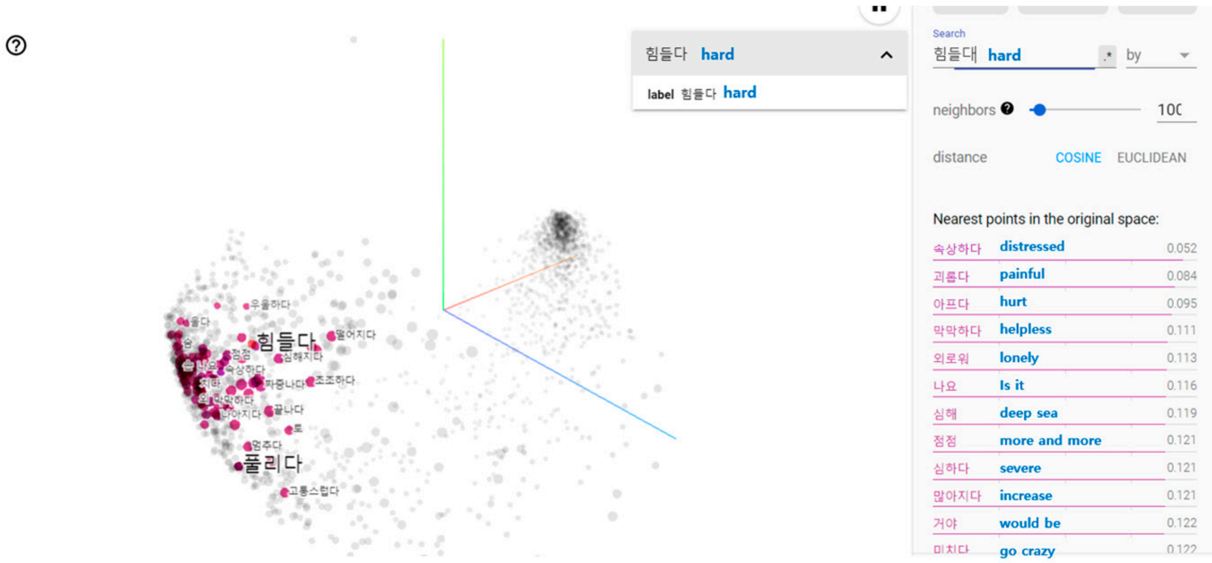

Figure 15. Visualization and words similar to "tough". Similar words appear up close, so the contents seem to be overlapped. Similar words can be found in detail on the right side of the figure. Translation of Korean is blue-colored.

7. Result

7.1. Data Labeling

Classified data sets were labeled according to the DSM-5 depression disorder diagnostic criteria. For example, if the word dictionary for the third diagnostic criterion, "loss or gain weight without diet, loss or increase of appetite almost every day", contained a "pig" in the counseling dataset sentence, the sentence was given a label of number 3, which means the third diagnostic criterion. Figure 16 shows the results of labeling wellness consultation data according to the nine diagnostic criteria and visualizing the frequency.

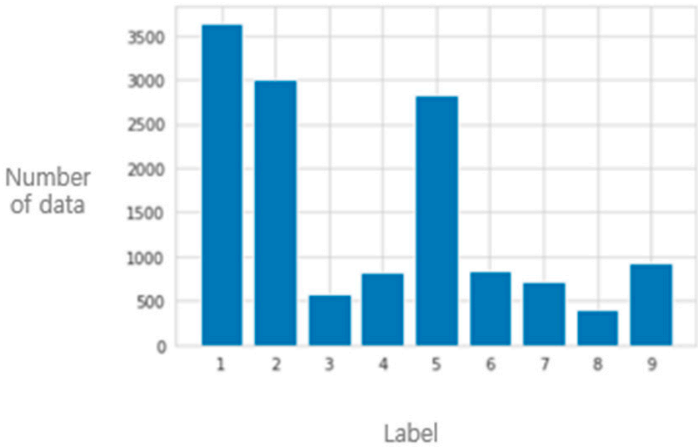

Figure 16. Visualization of dataset classification.

7.2. Comparing the Frequency of Each of the 9 Criteria

Figure 17 compares the labeling result of wellness counseling data before and after word pre-augmentation, and Figure 18 compares the labeling result of r/depression data from Reddit before and after augmentation. As can be seen from the figures, there was no significant difference before and after the word dictionary was augmented.

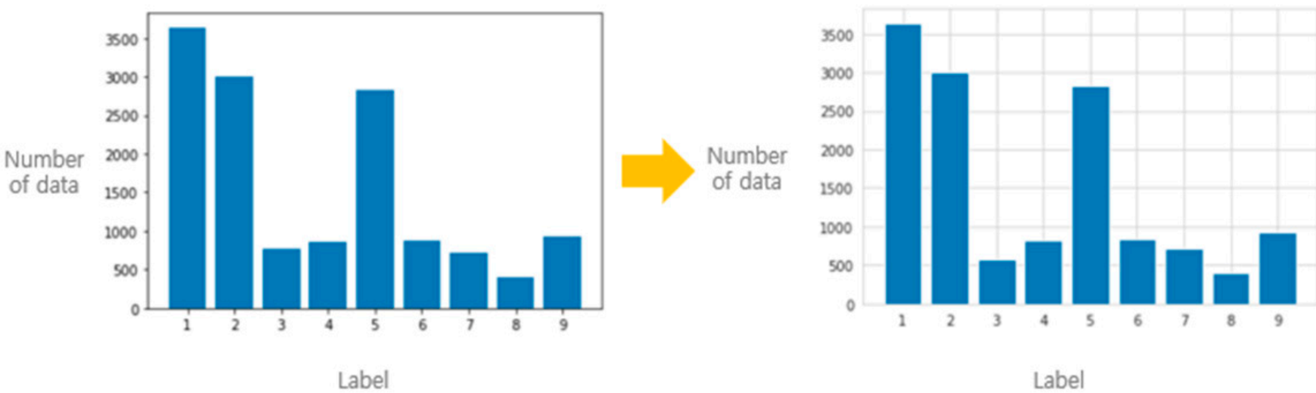

Figure 17. Comparison of the wellness dataset classification.

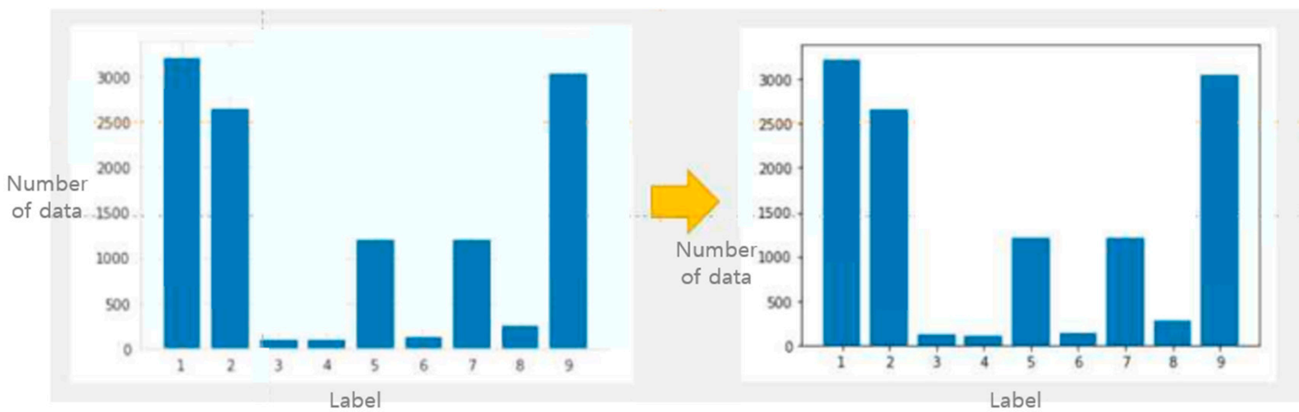

Figure 18. Comparison of the Reddit dataset classification.

Combining the wellness consultation data and Reddit data produced the following frequency visualization (Figure 19).

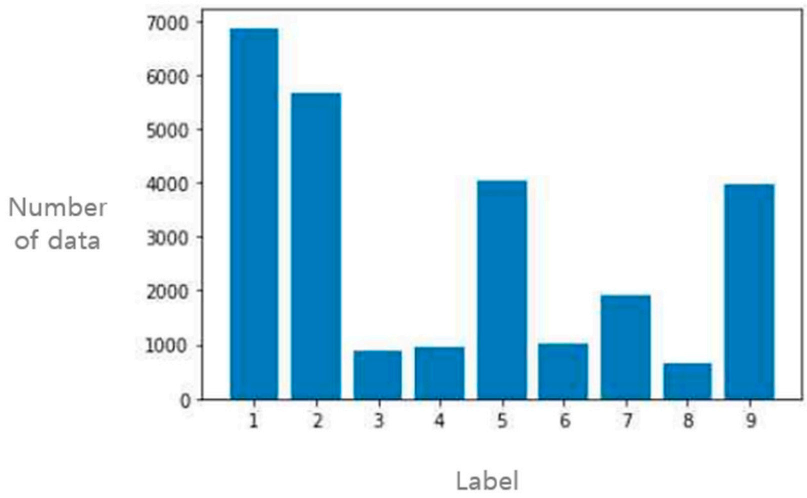

Figure 19. Comparison of entire dataset classification.

7.3. Frequency Result Inference

As a result of establishing a null hypothesis that there is a difference in data before and after augmentation and conducting a t-test, the p-value was over 0.8, which was greater than 0.05. Therefore, we confirmed that there was no significant difference before and after word dictionary augmentation. No significant difference was evident in the results before and after augmentation of the word dictionary because many words were not added during dictionary enhancement. In addition, we also identified that there was no significant difference in the counseling dataset collected in this study due to the lack of data corresponding to the label.

However, criteria 1 and 2 are "most of the day, almost every day of depressed mood", and "most of the day, almost every day of interest or pleasure in almost all daily activities is significantly reduced". This indicates that the expression words are more ambiguous than other criteria and are often used. Compared with other criteria, the number of expressions such as "loss or gain weight without diet control, loss or increase of appetite almost every day" is small, but clear. However, diagnostic criteria 1 include "tears", "sad", and "disappointed", which are ambiguous and frequently used in other contexts. As a result, we concluded that the number of data is inevitably high when the expression word is an ambiguous and widely used label.

8. Conclusions

This paper proposes a methodology for classifying and labeling datasets based on DSM-5 diagnostic criteria for depressive disorder. Wellness counseling data and Reddit crawling data consisting of empathy-related content were used to augment AI services for depressive counseling. A word dictionary built on medical knowledge was used, and we analyzed the dataset and word dictionary using a little morpheme analyzer, with the resulting dataset labeled 1 to 9. In addition, if the number of words was insufficient, the word dictionary was augmented using Word2Vec.

We deleted the entire sentence if it contained "/" in the data preprocessing process. In future work, only one of the front and back adjectives of "/" can be used or preprocessed in other ways. Reddit data are originally English data. Since we used this English data translated into Korean, the accuracy of the analysis may have been slightly reduced. In the future, it is expected that more accurate research will be conducted if Korean data that show depression are used.

A word dictionary built by medical staff and based on the DSM-5 depression disorder diagnosis criteria used by the APA was employed in this study. The proposed methodology can therefore be used in both medical care and natural language processing. Guidance learning techniques are widely used to create chatbots. Guidance learning or semi-learning can also be performed using datasets classified according to the DSM-5 depressive disorder diagnostic criteria, which are proven diagnostic criteria. In addition, through our proposed methodology, we can easily label new data to enable better learning, so we can expect excellent answers from natural language processing-based AI services such as chatbots. In addition, this approach can lead to a combination of AI and medical services. Training the model using these labeled datasets will enable early diagnosis in the case of not visiting the hospital due to negative gaze or cost problems, and it will be easy to monitor after diagnosis. If a sentence contains multiple emotions, it would be more reasonable to perform multi-label classification in future studies because even one sentence can have multiple labels.

Author Contributions: Conceptualization, G.L.; Formal analysis, G.L.; Methodology, G.L.; Supervision, H.O.; Visualization, G.L.; Writing—original draft, D.P.; Writing—review & editing, D.P. 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. NRF-2022R1F1A1074696), the Technology Innovation Program (or Industrial Strategic Technology Development Program-Source Technology Development and Commercialization of Digital Therapeutics) (20014967, Development of Digital Therapeutics for Depression from COVID19) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea), BK21 FOUR Project(Bigdata Research and Education Group for Enhancing Social Connectedness Thorough Advanced Data Technology and Interaction Science Research, 5199990913845) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea(NRF), AI Convergence Research Fund, Sungkyunkwan University, 2023 and SMC-SKKU Future Convergence Research Program Grant.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Na-Ju National Hospital. Depression. Available online: http://www.najumh.go.kr/html/content.do?depth=fc&menu_cd=02_02_00_01 (accessed on 20 March 2023).
2. Organization for Economic Co-Operation and Development. *A New Benchmark for Mental Health Systems: Tackling the Social and Economic Costs of Mental Ill-Health*; OECD Publishing: Paris, France, 2021.
3. Kim, S.Y.; Kim, J.E.; Lee, H.W. A Report of the Survey on the Current State of the Psychiatric Emergency Response in Seoul. *Ment. Health* **2021**, *11*, 38–54.
4. Song, G.H.; Kim, M.K.; Park, A.S. Promising Services Based on AI for Mental Health. *Electron. Telecommun. Trends* **2020**, *35*, 12–23.
5. Adam, M.; Wessel, M.; Benlian, A. AI-based chatbots in customer service and their effects on user compliance. *Electron. Mark.* **2020**, *31*, 427–445. [CrossRef]
6. Rashkin, H.; Smith, E.M.; Li, M.; Boureau, Y.L. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 5370–5381.
7. AI Hub. Wellness Conversation Dataset. Available online: https://aihub.or.kr/aihubdata/data/view.do?currMenu=120&topMenu=100&aihubDataSe=extrldata&dataSetSn=267 (accessed on 20 March 2023).
8. Płaza, M.; Trusz, S.; Kęczkowska, J.; Boksa, E.; Sadowski, S.; Koruba, Z. Machine Learning Algorithms for Detection and Classifications of Emotions in Contact Center Applications. *Sensors* **2022**, *22*, 5311. [CrossRef] [PubMed]
9. Rathnayaka, P.; Mills, N.; Burnett, D.; De Silva, D.; Alahakoon, D.; Gray, R. A Mental Health Chatbot with Cognitive Skills for Personalised Behavioural Activation and Remote Health Monitoring. *Sensors* **2022**, *22*, 3653. [CrossRef] [PubMed]
10. Prottasha, N.J.; Sami, A.A.; Kowsher, M.; Murad, S.A.; Bairagi, A.K.; Masud, M.; Baz, M. Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning. *Sensors* **2022**, *22*, 4157. [CrossRef] [PubMed]
11. Graterol, W.; Diaz-Amado, J.; Cardinale, Y.; Dongo, I.; Lopes-Silva, E.; Santos-Libarino, C. Emotion Detection for Social Robots Based on NLP Transformers and an Emotion Ontology. *Sensors* **2021**, *21*, 1322. [CrossRef] [PubMed]
12. Roca, S.; Rosset, S.; García, J.; Alesanco, Á. A Study on the Impacts of Slot Types and Training Data on Joint Natural Language Understanding in a Spanish Medication Management Assistant Scenario. *Sensors* **2022**, *22*, 2364. [CrossRef] [PubMed]

13. Kang, S.S.; Won, H.J.; Lee, M.H. Analyzing Vocabulary Characteristics of Colloquial Style Corpus and Automatic Construction of Sentiment Lexicon. *Smart Media J.* **2020**, *9*, 144–151. [CrossRef]
14. Kim, K.R.; Moon, J.H.; Oh, U.R. Analysis and Recognition of Depressive Emotion through NLP and Machine Learning. *J. Converg. Cult. Technol.* **2020**, *6*, 449–454.
15. Seo, H.R.; Song, M. An Analysis of the Discourse Topics of Users who Exhibit Symptoms of Depression on Social Media. *J. Korean Soc. Inf. Manag.* **2019**, *36*, 207–226.
16. Chin, H.J.; Baek, G.H.; Cha, C.Y.; Choi, J.H.; Im, H.S.; Cha, M.Y. A study on the categories and characteristics of depressive moods in chatbot data. In Proceedings of the Korea Information Processing Society Conference, Yeosu, Republic of Korea, 14–15 May 2021; pp. 993–996.
17. Kind Korean Morpheme Analyzer. Available online: http://kkma.snu.ac.kr/ (accessed on 20 March 2023).
18. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the Workshop at International Conference on Learning Representations, Scottsdale, AZ, USA, 2–4 May 2013.
19. Levy, O.; Goldberg, Y. Dependency-Based Word Embeddings. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Baltimore, MD, USA, 22–27 June 2014; Association for Computational Linguistics: Baltimore, MD, USA, 2014; pp. 302–308. [CrossRef]