International Journal of Advanced Culture Technology Vol.11 No.3 190-198 (2023)
DOI https://doi.org/10.17703/IJACT.2023.11.3.190

IJACT 23-9-26

Prediction Model of Inclination to Visit Jeju Tourist Attractions based on CNN Deep Learning

[1]YoungSang Kim

[1]Prof., Dept. of Computer Multimedia, Cheju Halla Univ., Korea
yskim@chu.ac.kr

Abstract

Sentiment analysis can be applied to all texts generated from websites, blogs, messengers, etc. The study fulfills an artificial intelligence sentiment analysis estimating visiting evaluation opinions (reviews) and visitor ratings, and suggests a deep learning model which foretells either an affirmative or a negative inclination for new reviews. This study operates review big data about Jeju tourist attractions which are extracted from Google from October 1st, 2021 to November 30th, 2021. The normalization data used in the propensity prediction modeling of this study were divided into training data and test data at a 7.5:2.5 ratio, and the CNN classification neural network was used for learning. The predictive model of the research indicates an accuracy of approximately 84.72%, which shows that it can upgrade performance in the future as evaluating its error rate and learning precision.

Keywords: *Visit Evaluation, CNN, Prediction, Sentimental Analysis, Opinion Mining, NLP*

1. INTRODUCTION

These days, as tour restrictions on tourism that has come to a crisis from the covid 19 pandemic epoch have been lifted, the number of tourists starts to show a clear increasing trend in Jeju. The current situation of tourist visits to Jeju indicates the increase by 2.4% from February of 2022 (1,029,503) to February of 2023 tally (1,054,299 people)[1]. Tourism is a high value-added activity which provides the local growth of the industry and economy and improves chances for income and job creation in Jeju. Generally, tourism improves the quality of life by increasing incomes and living standards due to positive results like revenue growth, job creation, and infrastructure improvement[2] but it may also have a detrimental influence on the lives of the residents[3]. Jeju has not only its attractiveness, but also the natural environment as a tourist destination, becoming a significant factor in visiting. The residents of the province have pride for this natural environment and infinite affection[4].

The government designates 2023-2024 as the Year of Visit to Korea and sets aside 1.229 trillion won as the 2023 tourism budget to make the Republic of Korea a tourist attraction country visited by people around the world[5], and Jeju provincial government conducts the policy of investing 19.6 billion won in the intelligent tourism service sector as a regional strategic industry to pursue the revival of the tourism industry[6]. Accordingly, in order to revitalize Jeju tourism, it is essential to analyze consumer opinions and development

Manuscript received: August 2, 2023 / revised: August 10, 2023 / accepted: August 29, 2023
Corresponding Author: yskim@chu.ac.kr
Tel:+82-64-741-7669
Professor, Dept. of Computer Multimedia, Cheju Halla University, Korea

of visit recommendation algorithms, and to research how to integrate 4th industrial revolution technologies such as artificial intelligence into tourism products. A good example is how to plan a personalized travel schedule using tourism information and recommended travel products in the Smart Tourism Information System[7].

There are important factors in improving predictive performance such as exploring a model that predicts sentiment and rating in TripAdvisor hotel review data to predict opinions in certain situations[8], analyzing sentiment polarity in tourists' reviews based on Bidirectional Encoder Presentation from Transformers (BERT) model[9], or solving high performance discrepancies between online review-based Support Vector Regression(SVR) models[10]. The consumer's unique sentiment in a tourist destination is a unique emotion felt in the tourist destination. Therefore, An unstructured data mining technique, Sentiment Analysis is a technique that determines if a text written by consumers is either positive or negative by identifying attitudes or emotions contained in it. Sentiment Analysis is able to be applied to all of the texts generated from blogs, websites, SNS and tweeter messages[11].

Consumer opinions have become an important data domain in such marketing areas as product sales and services, movie preferences, and travel destination recommendations. In particular, if a predictive model like repurchase, revisit, and recommendation is established by analyzing text opinions and quantitative ratings, it can be used as an active virtuous cycle function to improve the quality of product services. Of course, it is true that the criteria for writing consumer opinions and judging ratings are very diverse and subjective, and the degree of actual polarity is often inconsistent, which affects the accuracy of the prediction model. Nevertheless, it is possible to implement a predictive model with improved performance through preprocessing, learning, and training of big data through data sampling, combination, and elimination of missing values.

This study is meaningful in implementing a propensity prediction machine learning system that can conduct emotional analysis of major tourist attractions in Jeju and judge new evaluation opinions by applying CNN deep learning techniques used in various fields. This study consists of as following. Chapter 2 includes the opinion mining, the sentiment analysis, and natural language processing, in addition, ChatGPT and the research related to CNN deep learning. Chapter 3 performs sentiment analysis and conducts CNN machine learning on the big data of the reviews from each Jeju tourist destination which has been collected on Google from October 1st to November 30th, 2021 and ends up proposing an either positive or negative propensity prediction model about the opinions of the visit evaluation. Chapter 4 estimates the performance of the predictive model by conducting experiments, and Chapter 5 discusses the performance improvement of this model and concludes by suggesting the future direction[12].

2. RELATED WORKS

2.1 Opinion Mining and Sentiment Analysis

In general, text mining refers to a technique that extracts a pattern or relationship from text data and finds and interprets meaningful information or value. In other words, text mining can be used to analyze consumers' opinions or sentiments by analyzing the structure of text data included in websites, office files, XML, blogs, SNS, articles, reviews, etc. and making it statistically processable. Its major processing technologies include natural language processing techniques for extracting necessary information, classification techniques for learning, and keyword correlation analysis techniques. Opinion mining, a field of text mining, analyzes the opinions of consumers in marketing services, which is a very appropriate technique to understand public opinion or to examine responses to services.

Machine learning is critical to process huge amounts of text data. Furthermore, natural language processing should be used in order to find and analyze such sentiment information as reviews, preferences, and ratings for services or products, and end up making judgments on the tendencies for positive or negative position. H.Tacherdoost lately analyzed the search of the Scopus literature between 2012 and 2022 about how to differentiate between competitors' products and services, showing that it is used in the order like this: first, sentiment analysis, second, deep learning, third, machine learning, and lastly artificial intelligence[13].

Sentiment analysis refers to identifying ideas, emotions, and attitudes expressed in a source material[14].

Therefore, sentiment analysis can lead to interpretation errors if it is too simple or complicated, depending on whether the word used in the consumer's opinion of a product or service is positive or negative. However, in spite of some errors in individual sentences, there is no great difficulty in grasping the trend if the data is vast. Sentiment analysis includes spam and false data, domain dependencies, negations, natural language processing (NLP) burdens, bipolar terms, and extensive vocabulary use, especially data preprocessing that removes punctuation, emoticons, and spaces[15]. Susnjak proposed a process of verifying preliminary results using both a step of performing sentiment analysis using a pre-trained language model and an interpretable machine using 5643 abstract datasets collected from a scientific journal on the subject of chronic Lyme disease, a medical domain[16].

2.2 Natural language processing(NLP) and ChatGPT

ChatGPT is creating a revolutionary syndrome around the world, which makes the recent era of search end. The key to ChatGPT is to let computers learn the probability relationship between words and words in natural language processing. In other words, hundreds of billions of language data are learned to present words with the highest probability relationship among them, and reinforcement learning is added to compensate for the awkwardness of sentences. In December, 2022, OpenAI released ChatGPT, which is known to own 100 trillion parameters equivalent to synapses in the brain of human beings. Therefore, By applying artificial intelligence technology which on the spot gives answers to user questions, it will appear as a progressive technology in the area of natural language processing. In addition, as the answer is on the basis of an enormous amount of learning information, the response of ChatGPT might not be a fact in truth, where there are a lot of disadvantages such as generalization of either new data or continuity of conversation. Nonetheless, ChatGPT is a large-scale data model which is able to be sufficiently enough to be recognized for its performance from forming various documents of low importance, coding, to delivering information as a work aid. Figure 1. (a) hows ChatGPT's answer to the question of recommending 10 Jeju attractions. Figure 1. (b) shows the output of ChatGPT when asked to classify the abstract of sentiments from 1 to 5 for Lyme disease's ratings in reference [16].

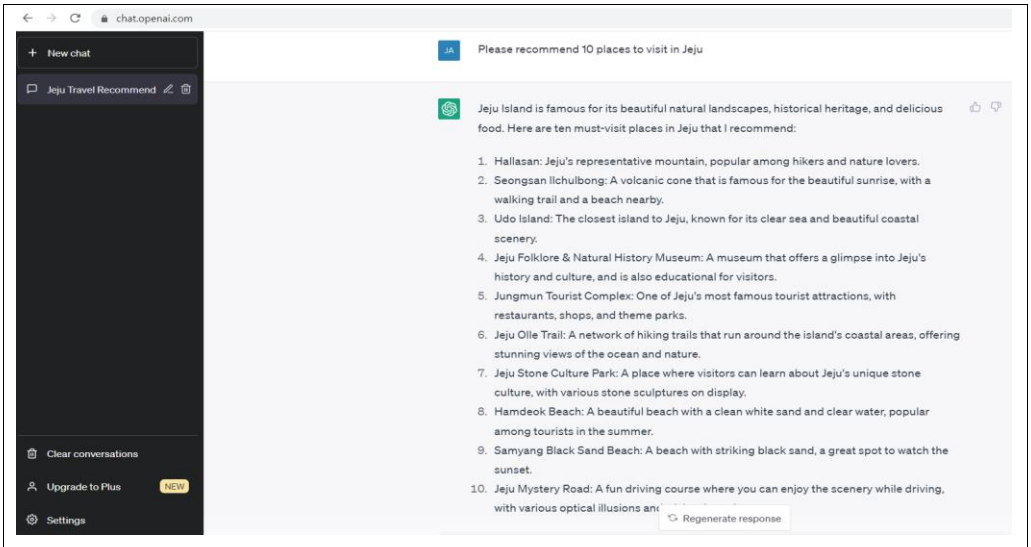

Figure 1. (a) ChatGPT's answer to the question of recommending 10 Jeju attractions

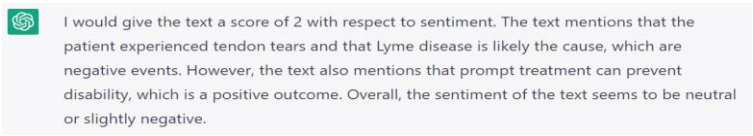

Figure 1. (b) ChatGPT Response to Lyme Disease Ratings

2.3 Deep Learning & CNN

Machine learning is divided into supervised learning and unsupervised learning depending on whether the learning data has labels or not. Therefore, machine learning should require a person to define a learning model in advance in a way that solves problems through methods such as classification, prediction, and clustering. Deep learning is a specific field of machine learning, which has the advantage of gradually learning meaningful expressions in continuous layers and is a method of learning hierarchical expressions from data.

CNN (Convolutional Neural Network) is a deep learning algorithm widely used for image recognition and voice recognition, but it can also be used for natural language processing projects. For example, 'It's so good' and 'I want to come back' can be analyzed as a positive grade, and 'very dissatisfied' and 'too bad' can be analyzed as a negative one. The CNN structure consists of convolution(Conv) that extracts features by creating a feature map, as shown in Figure 2, and pooling that reduces size and emphasizes specific features[17].

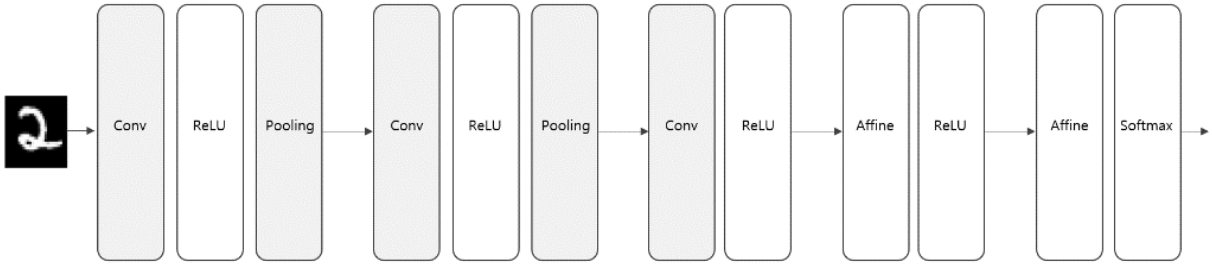

Figure 2. Example of each layer of a network made up of CNNs

To apply the CNN classification model for sentence analysis, As shown in Figure 3 of Y. Kim, sentences consisting of n words are embedded in k-dimensional row vectors for each word. CNN filters are 2 and 3, meaning Bigram and Trigram models, respectively[18].

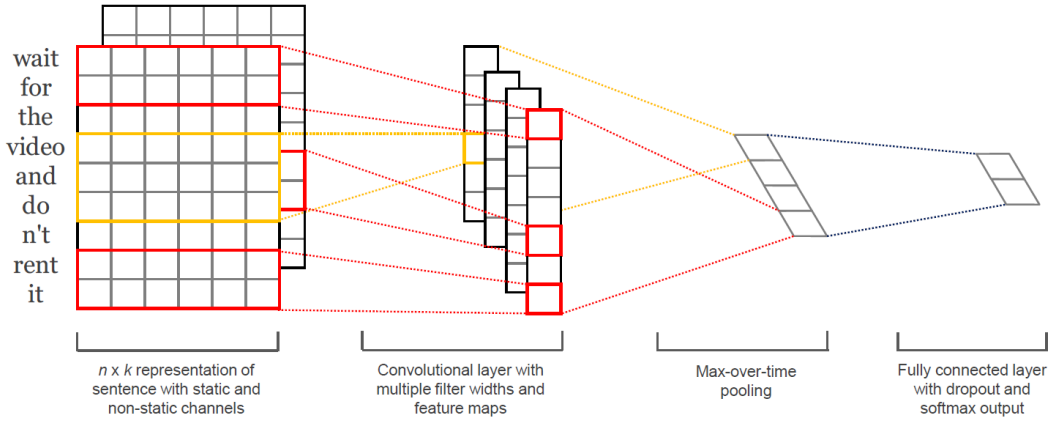

Figure 3. Model Architecture with two Channels for an Example Sentence

Sentiment analysis is conducted through a total of four stages from left to right in Figure 3. In the last step, a fully connected neural network enables final classification results (positive/negative) based on features extracted from input data, when the softmax function is used.

3. PROPOSED METHOLOGY AND PREDICTION MODEL

3.1 Deep Learning Overview

A language model is a model that calculates the probability of each word in a sequence of words such as a sentence. Large language models such as ChatGPT are pre-learned based on sentence and language data. The pre-trained model performs the transfer learning that optimizes the performance by pre-learning large amounts of data and learning additional necessary data. In the process of fine tuning a pre-learned model, labeled data such as sentiment analysis and natural language reasoning are used for learning. This study installed TensorFlow and Keras 2.12 for CNN deep learning.

The sentiment analysis modeling shown in this paper proceeds in the sequence of big data collection, normalization and analysis of learning data, separation of verification data and learning data, CNN model composition of visit evaluation opinions and ratings, learning and testing, and then prediction. The usage data summary information in this study is the same as Table 1.

Table 1. Usage Data Summary Information

Table classification	Description
Original	Data collected by NEXTEZ Co., Ltd. from visit evaluation of Jeju tourist attractions extracted by Google from October 1st, 2021 to November 30th, 2021
Data structure	tourist destination name, tourist destination type name, author type value, total number of evaluation, total number of images, visit evaluation opinion, image information, writing out time, collection date, rating
Total number of collected data	183,507 cases(49MB)
Number of actually used data (excluding data without feedback)	95,890 cases(15MB)
The range of ratings	1 ~5points
The Shortest assessment	1 letter
The longest assessment	2000 letters
The average length of assessment comments	46. 11 letters
Number of tourist attractions for visit evaluation	527 places

3.2 Data Normalization and Preprocessing

The 95,890 real learning data used in this paper are at random divided into training data set versus test data set at the ratio of 7.5:2.5. Training data is separated into validation data and training data. As it is difficult to estimate the point at which the error of the actual data rises with just the learning data, implementing data normalization doesn't bring about overfitting of overly learning the learning data. Figure 4 is the frequency analysis WordCloud for the training data.

Figure 4. Frequency analysis of train data WordCloud

Korean data preprocessing makes only Korean words left with spaces in the sentence as delimiters through normalization, tokenization that divides sentences and part-verb tagging which determines the part-verb corresponding to each word. After that, it extracts the word stems with a morpheme analyzer, and removes postpositions, punctuations, and non-terms such as 'ㅎㅎ', 'ㅋㅋ', and 'ㅠ'. It generates a word dictionary as shown in Figure 5 and converts the visiting evaluation opinion sentence into each index vector.

Index	Type	Size	Value
63	list	3	['블로그', '많이', '다르다']
64	list	12	['소음', '심하다', '별로', '볼', '게', '없다', '특히', '흐리다', '날', '엔', ...]
65	list	23	['마라도', '짱콩', '맛있다', '여기', '서', '오토바이', '다리', '개', '짜다', '빌리다', ...]
66	list	11	['주차장', '크게', '자다', '갖추다', '있다', '마농', '마늘', '치킨', '맞다', '집도', ...]
67	list	14	['요즘', '제주도', '가장', '핫', '하다', '곳', '이네', '요', '카페', '많다', ...]
68	list	1	['좋다']
69	list	1	['좋다']
70	list	6	['어리다', '시절', '되다', '돌다', '가다', '같다']
71	list	2	['매우', '좋다']
72	list	10	['해변', '검은색', '돌', '알갱이', '관리', '부족', '쓰레기', '보이다', '점', '아쉬움']
73	list	6	['고', '즈', '넉', '하고', '분위기', '좋다']
74	list	2	['역사', '학습']
75	list	3	['산책', '하다', '좋다']
76	list	16	['투어', '내용', '생각', '보다', '깊다', '않다', '금요일', '오후', '방문', '하다', ...]
77	list	2	['풍경', '멋지다']

Figure 5. Examples of Segmentation of Visiting Evaluation Opinion Sentences and Word Dictionaries

It shows a word dictionary which is a simple list of words such as a bag of words and a list of words which is a sort of word appearance frequency. When distinguishing the words into positive and negative words, the rating information of the visiting evaluation opinions is used to classify them as positive/ negative. For training data and test data, ratings from 1 to 5 are classified into two groups, where 4 and 5 ratings are labeled as positive and 1~ 3 ratings are labeled as negative.

3.2 Creating a Sentiment Analysis Model

Similar to the method that is used to make image classification, for sentiment analysis, CNN's input data is made up of a one-dimensional data array. In this process, this study uses a method of increasing one dimension for each word while embedding words in a sentence consisting of n words, which creates a feature map as many as the number of CNN filters and outputs a positive or negative tendency through Max-pooling.

This paper sets the CNN model and designate Epochs values and Batch size for full-fledged learning. In

the course of learning, if the verification is less accurate than before, this system will stop learning as following in Figure 6 so as to prevent overfitting as well as the loss rate.

```
750/750 [==============================] - 2s 2ms/step - loss: 0.3742 - accuracy: 0.8472
750/750 [==============================] - 1s 2ms/step - loss: 0.3742 - accuracy: 0.8472
[0.3741779029369354, 0.8471550345420837]
Epoch 1/10
750/750 [==============================] - 43s 56ms/step - loss: 0.3735 - accuracy:
0.8496
Epoch 2/10
750/750 [==============================] - 42s 56ms/step - loss: 0.3153 - accuracy:
0.8739
Epoch 3/10
 94/750 [==>...........................] - ETA: 41s - loss: 0.2311 - accuracy: 0.9139
```

Figure 6. Stop learning if verification accuracy is reduced

4. RESULT OF EXPERIMENT AND DISCUSSION

As a result, the loss rate and the rate of change in accuracy are presented in Figure 7 to suggest that the performance of this model is excellent.

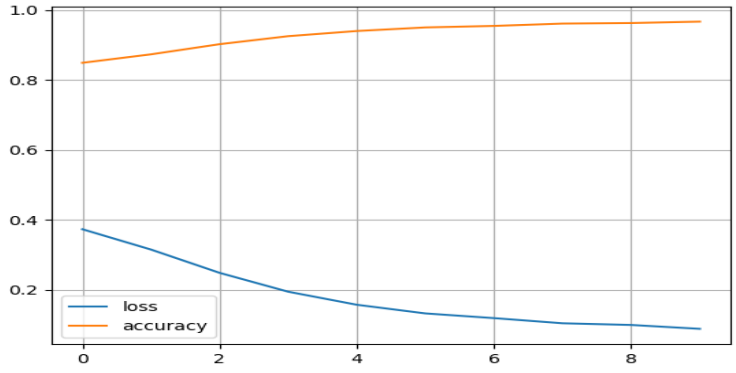

Figure 7. Visualization of loss rates and accuracy

There are 10 emotional words which play an important part in predicting the answer in this model. They show the order of either (positive) counts: (1.best, 2. enjoyable, 3. amused, 4. healed, 5. good, 6. beautiful, 7. valuable, 8. strongly recommended, 9. cool, 10. pretty) or (negative) ones: (1. nothing, 2. not bad, 3. just, 4. disappointed, 5. unkind, 6. so-so, 7. wasteful, 8. worthless, 9. normal, 10. worst). As a result, the accuracy of guessing the answers to the test data for the learning model was 84.72%.

Figure 8 shows the output results of the propensity prediction model of this study when a new evaluation opinion for a tourist destination is entered based on the previously learned content.

```
Enter a sentence of two or more words for the visit evaluation: 친절하고 좋다
1/1 [==============================] - 0s 75ms/step
82.12% Prob. POsitive Prediction

Enter a sentence of two or more words for the visit evaluation: 제주도 안좋음
1/1 [==============================] - 0s 73ms/step
59.15% Prob. NEgative Prediction

Enter a sentence of two or more words for the visit evaluation: 너무너무 좋다
1/1 [==============================] - 0s 62ms/step
56.78% Prob. POsitive Prediction

Enter a sentence of two or more words for the visit evaluation: 다시 안온다
1/1 [==============================] - 0s 70ms/step
52.20% Prob. NEgative Prediction

Enter a sentence of two or more words for the visit evaluation: 안좋고 불친절
1/1 [==============================] - 0s 76ms/step
69.53% Prob. NEgative Prediction
```

Figure 8. Output of propensity prediction model

In this model, it is confirmed that both positive and negative appraisals are appropriately output with probability values by model learning as for the sentences that contain emotional words which are already learned. Despite its good performance, positive-negative evaluations have occasionally changed in the case of combinations of words which have never been learned, because of either a lack of data which have been collected during a short period time or incompletely refined Korean sentences.

5. CONCLUSION

Consumers' experiences with travel or tourism are very suitable as the subject of sentiment analysis as they can tell stories in certain places and share or deliver information to others. As the interest in health and rest has increased since COVID-19, the travel industry is transforming into a future high value-added industry worldwide. Jeju supports the demand for tourism products and tourism-related job creation policies in major tourist destinations in order to respond to the digital transformation of the tourism industry and revive sustainable tourism.

This study provides a CNN learning model which can predict either positive or negative tendency if it is given a new evaluation opinion by carrying out an sentiment classifier that has learned consumer evaluation opinions and ratings on tourist attractions in Jeju. When establishing a policy to foster tourist attractions, the model can also be used in order to track consumers' movement routes and tourists' consumption patterns, and to attract new consumers on the basis of the results of the predictive model.

Future research plans will need to further develop existing learning models by ensuring the reliability of further learning, for instance, collecting more tourist big data and reconstructing data to address inaccuracies coming from sentence constraints. There is also a need for technology advancement strategies that can add functions that can correct wrong results on their own or overcome the limitations of sentence combinations. In the long run, it needs to break down the limitations of deep learning and then to implement an optimal propensity prediction program through continuous neural network learning training so that Jeju tourism can be branded with general and universal consumer sentiment.

REFERENCES

[1] Jeju Tourism Association, Current status of tourist arrivals in February, *http://www.visitjeju.or.kr/main.do*, Feb. 2023.

[2] R. Harrill, "Residents' Attitudes toward Tourism Development: A Literature Review with Implications for Tourism Planning," *Journal of Planning Literature*, Vol. 18, No. 3, pp. 251-266, Feb. 2004. DOI: 088 5412203260306

[3] K.L. Andereck, G.P. Nyaupane, " Exploring the Nature of Tourism and Quality of Life Perceptions among Residents," *Journal of Travel Research*, Vol. 50, No. 3, pp. 248-260, May 2010. DOI:10.1177/0047287510362918

[4] M. Chang, The Effects of Community Attachment and the Perceptions of Tourism Impacts on the Intentions of Pro-Environmental Behavior and Support for Tourism Development on Jeju Island, Ph.D. Thesis. Jeju National Univ., KOREA, Feb. 2021.

[5] Ministry of Culture, Sports and Tourism, *https://www.korea.kr/news/visualNewsView.do?newsId=148910670&sectIds=&cateId=&pageIndex=1&repCode_C=*, January 7, 2023.

[6] S.S. Won, *https://www.headlinejeju.co.kr/news/articleView.html?idxno=508865*, Headlinejeju, March 05, 2023.

[7] T.S. Ko, B.J. Kim, J.W. Jwa, "Smart Tourism Information System and IoT Data Collection Devices for Location-based Tourism and Tourist Safety Services," *International Journal of Advanced Culture Technology*, Vol. 10, No. 2, pp. 310-316, 2022. DOI https://doi.org/10.17703/IJACT.2022.10.1.310

[8] K. Puh, M.B. Babac, "Predicting Sentiment and rating of Tourist Reviews using machine Learning," *Journal of Hospitality and Tourism Insights(JHTI)*, July 2022. DOI 10.1108/JHTI-02-2022-0078

[9] M. Chu et al.,"Language Interpretation in travel Guidance Platform: Text Mining and Sentiment Analysis of TripAdvisor Reviews," *Frontiers in Psychology*, pp. 1-8, Oct. 2022. DOI 10.3389/fpsyg.2022.1029945

[10] T.H. Hong, "Sentiment Analysis and Star Rating Prediction Based on Big Data Analysis of Online Reviews of Foreign Tourists Visting Korea," *Knowledge Management Research*, Vol. 23, No.1, Mar. 2022. https://doi.org/10.15813/kmr.2022.23.1.010

[11] Y.J. Park, "A Study on the Sentiment Analysis of Contemporary Pop Musicians and Classical Music Composers," *International Journal of Advanced Culture Technology(IJACT)*, Vol. 10, No. 3, pp. 352-359, Sep. 2022. https://doi.org/10.17703/IJACT.2022.10.3.352

[12] Y.S. Kim, "A Study on Sentiment Analysis of Jeju Tourist Destination Using Review Big Data with the Machine Learning," in *Smart Convergence of Culture Technology Letters(SCCTL 09)*, pp. 59-61, Jul. 2023.

[13] H. Taherdoost, M. Madanchian, "Artificial Intelligence and Sentiment Analysis:A Review in Competitive Research," *Computers*, Vol. 12, No. 37, Feb. 2023. https://doi.org/10.3390/computers12020037

[14] A. Ahad et al., " An Intelligent Data Mining System for Tweet Opinion Analysis using Combined Cluster based Classification Approach," *International Journal of Intelligent Systems and Applications in Engineering(IJISAE)*, Vol. 10, No. 4, pp. 604-609, Dec. 2022. *https://ijisae.org/index.php/IJISAE/article/view/2330*

[15] M.A. Palomino, F.Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Applied Sciencies* , Vol. 12. No.17, Aug. 2022. DOI 10.3390/app12178765

[16] T. Susnjak, "Applying BERT and ChatGPT for Sentiment Analysis of Lyme Diease in Scientific Literature", *Computation and Language(cs. CL)*, Cornell University, Feb. 2023. https://doi.org/10.48550/arXiv.2302.06474

[17] K. Saitoh, Deep Learning from Scratch, *Hanbit Media*, 2017.

[18] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing(EMNLP)*, pp. 1746-1751, Oct. 2014.