

Research Article

Analysis of Public Big Data Management under Text Analysis

Yue Zhu ¹ and Ho Yin Kan ²

¹School of Law, Tsinghua University, Beijing, China

²Centre for Continuing Education, Macao Polytechnic University, Macao, China

Correspondence should be addressed to Ho Yin Kan; hykan@ipm.edu.mo

Received 7 June 2022; Revised 4 July 2022; Accepted 11 July 2022; Published 30 July 2022

Academic Editor: Lianhui Li

Copyright © 2022 Yue Zhu and Ho Yin Kan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Based on text analysis, public big data management is studied. The public data management of Mount Wutai tourism network travel notes is discussed. The positive, neutral, and negative effects of the naive Bayesian classification model and decision tree classification model on the tourism sentiment attitude of Mount Wutai are compared. The relationship between tourism resources, tourism facilities, tourism services, tourism environment, and tourism sentiment and attitude of Wutai Mountain is analyzed. The results show that the true positive rate, true negative rate, and F-measure of the Bayesian decision tree classifier to classify positive text are 86.64%, 81.27%, and 84.62%, respectively. The true positive rate for neutral text is 82.05%, the true negative rate is 78.89%, and the F-measure is 77.11%. The true positive rate for negative text is 83.67%, the true negative rate is 98.29%, and the F-measure is 82.83%. The Bayesian decision tree classifier can evaluate positive and negative texts better than neutral texts. The true positive rate of the C4.5 decision tree classifier for positive text is 91.44%, the true negative rate is 86.57%, and the F-measure is 89.45%. The true positive rate for neutral text is 90.17%, the true negative rate is 83.28%, and the F-measure is 84.06%. The true positive rate for negative text is 91.84%, the true negative rate is 99.05%, and the F-measure is 90.91%. The decision tree classifier has a better evaluation effect on positive and negative texts than on neutral texts. The ROC curve of the evaluation effect of the two classifiers shows that the evaluation effect of the two classifiers has a better evaluation effect on positive text than that of the neutral and negative texts, and the evaluation effect of the C4.5 decision tree classifier is better than that of the Bayesian classifier. The promotion degree of tourism resources and facilities in forwarding online travel notes is obviously higher, and there is a high correlation between tourism resources and facilities and forward online travel notes. In negative online travel notes, the promotion degree of tourism service and tourism environment is high, and the correlation between tourism service and tourism environment and negative online travel notes is high. In summary, improving the quality of tourism services and the tourism environment of Mount Wutai scenic spots can better enhance the recognition and satisfaction of tourists with Mount Wutai tourism.

1. Introduction

With the development of Internet technology, network public text data play an important role in people's search for and mastery of information. The management and classification integration of public text data can effectively improve data utilization [1, 2]. Tourism plays an important role in the development of a city. Understanding the status quo and problems of regional tourism is of positive value to guiding the development of the local economy [3, 4]. Tourists' emotional attitudes and recommendation indexes have an important influence on the development of tourist

attractions [5, 6]. Exploring and analyzing tourists' recognition of scenic spots, which is of reference value to guide the improvement and perfection of tourist attractions, is a research hotspot [7, 8].

Tourists' evaluations and emotional attitudes toward scenic spots are expressed mainly online and offline, and online searches have shortcut lines and universality [9]. The online travel notes of travel apps are an important tool to understand tourists' emotional attitudes toward scenic spots. Currently, widely used travel apps mainly include Ctrip, Qunar, Hornet's Nest, Tuniu, and Tongcheng [10]. Mount Wutai, located in Xinzhou City, is a national 5A-level scenic

spot with beautiful scenery and numerous temples. It is a Buddhist holy land with a beautiful natural landscape and profound cultural landscape. It is cool and comfortable in summer and a good place for summer vacation [11]. Mount Wutai scenic spot has many tourists every year, which greatly promotes local economic development and has great development potential. Understanding Mount Wutai's tourism recognition and tourist reputation is of great value to the economic development of Mount Wutai. The emotional attitude assessment of scenic spots based on online travel notes requires text processing, classification, data organization, and management and correlation analysis. Currently, the commonly used machine learning classification methods mainly include naive Bayesian classification models, decision tree classification models, and support vector machine classification models [12–14]. Naive Bayesian classifiers have high classification accuracy and fast running speed [15, 16]. C4.5 decision tree classification can effectively process data with missing values [17, 18].

In this study, 575 online travel journal texts from the Ctrip, Qunar, Mafengwo, Tuniu, and Tongcheng websites are selected to compare the classification and evaluation effects of the naive Bayesian classification model and decision tree classification model on Mount Wutai tourism sentiment and attitude. The relationship between tourism resources, tourism facilities, tourism services, the tourism environment, and Mount Wutai's tourism sentiment and attitude is also analyzed to provide guidance for tourists and a reference for the development and improvement of Mount Wutai scenic spots.

2. Methods

2.1. Data Acquisition. The public data selected in this study are mainly from online travel notes published on various travel websites. The online travel notes published on the Ctrip, Qunar, Horner's Nest, Tuniu, and Tongcheng websites are screened. After the above websites are selected, the keyword "Mount Wutai" in the travel notes of each website is searched. Octopus acquisition software is used to climb Mount Wutai network text data of the above tourism websites one by one. After screening, a total of 575 Mount Wutai travel notes from January 2016 to December 2018 are selected, with a total of 412,528 words. The number of texts on each website is shown in Table 1.

2.2. Data Feature Extraction and Typing. By comparing other similar literature and consulting relevant analysts, the public data of Mount Wutai tourism network text are divided into three types, namely, positive text data, neutral text data, and negative text data. Based on the content description of online travel notes, the description information of travel notes will contain specific descriptive words and short sentences of the Mount Wutai tourism event. For example, the event description in positive text includes keywords expressing positive evaluation such as "satisfactory," "beautiful scenery," and "high-cost performance." Negative text event description includes words expressing

TABLE 1: Mount Wutai tourism network text public data collection.

Website	Number of articles
Ctrip	328
Qunar	72
Hornet's nest	27
Tuniu	45
Tongcheng tourist	103

negative evaluation such as "inconvenient transportation," "expensive tickets," and "poor accommodation." Neutral event descriptions mainly include "check in" and "handle." These specific words are set as keywords and extracted from the input data set for text data classification (Table 2).

2.3. Data Classification

2.3.1. Naive Bayesian Classification Algorithm. The Bayesian classifier supports incremental learning and has high classification accuracy. The classifier assumes that terms are independent of each other and that each sample X in the text set consists of a set of attribute values $(a_1, a_2, a_3, \dots, a_n)$, where a_k is the value of term A_k . W is the classification variable, and w is the value of W . It is supposed that there are two classes, namely, + (positive class) and - (negative class). According to the Bayesian rule, sample X is the probability p of class w , as shown in (1), where X is classified as $W = +$ if and only if and represented by the Bayesian classifier as $f_b(X)$, as shown in (2).

$$p(w|X) = \frac{p(X|w)p(w)}{p(X)}, \quad (1)$$

$$f_b(X) = \frac{p(W = +X)}{p(W = -X)} \geq 1. \quad (2)$$

If the given values of class variables and the terms are independent of each other, the probability p can be expressed as $p(X|w)$, as shown in (3). The naive Bayesian classifier, $f_{nb}(X)$, is obtained, as shown in (4).

$$p(X|w) = p(a_1, a_2, \dots, a_n|w) = \prod_{i=1}^n p(a_i|w), \quad (3)$$

$$f_{nb}(X) = \frac{p(W = +)}{p(W = -)} \prod_{i=1}^n \frac{p(a_i|W = +)}{p(a_i|W = -)}. \quad (4)$$

$P(a_1|w)$, $p(a_2|w)$, $p(a_3|w)$, \dots , and $p(a_n|w)$ can be estimated by the training sample. The posterior probability of each class can be calculated separately, and the class with the highest posterior probability is the prediction class.

2.3.2. Decision Tree Classification Algorithm. C4.5 decision tree classification is a common method in inductive reasoning, which has great advantages in dealing with continuous attributes and discrete attributes. The measure of the uncertainty of feature vector A is the entropy of feature vector A , expressed as $H(A)$, as shown in (5). The conditional entropy of class B of feature vector A under given conditions

TABLE 2: Results of feature extraction in the training set.

Describe	Feature extraction
Wutai Mountain has beautiful scenery and is very comfortable in summer	Beautiful scenery, comfortable
Tickets are expensive and accommodation consumption is high	Tickets are expensive, consumption is high
There are many volunteers in the scenic spot with comprehensive service	Comprehensive service
A good place for self-cultivation and picturesque scenery	Good place, picturesque scenery

is expressed as $H(A/B)$, as shown in (6), which represents the uncertainty of classification of feature vector A under given conditions of class B , where p_k is the probability that A is a_k , $P(A = a_k) = p(a_k)$ and $k = 1, 2, \dots, n$. $P(A_i|B_j)$ is the joint probability distribution. The difference between $H(A)$ and $H(A/B)$ is information gain, and the larger the information gain is, the stronger the classification ability is

$$H(A) = - \sum_{a_i} p(a_i) \log_2 p(a_i), \quad (5)$$

$$H(A|B) = - \sum_j p(b_j) \sum_i p(a_i|b_j) \log_2 p(a_i|b_j). \quad (6)$$

In C4.5, the information gain rate is used to determine the selection of feature test points by measuring the correlation between feature A and class B and the entropy value of feature A and class B , as shown in the following equation:

$$\text{Gain Ratio}(A|B) = \frac{H(A) - H(A|B)}{H(A)}. \quad (7)$$

The C4.5 decision tree algorithm adopts the information gain rate as the feature selection method sets a threshold α and takes it as the stop condition. The feature with the maximum information gain rate is placed on the root node. If the information gain rate of the feature is less than the threshold value, the node constitutes a single-node tree. The class is marked as the class with the most samples in the dataset that satisfies the path condition from the root node to the local node. If the information gain rate of the feature is greater than the threshold, a branch is generated for each feature value of the node, and the training sample is assigned to the corresponding branch. If there are no samples in a branch or all samples belong to the same category, the branch ends; otherwise, the branch node is repeated until all features have been traversed.

2.4. Data Organization and Management. Data organization and management are conducive to better use of data and to improving the efficiency of data search. The obtained network travel notes text data are organized and managed, including time information, emotion and attitude information, and programmatic information. Time information includes year, month, and day, and sentiment and attitude classification includes positive text, neutral text, and negative text. According to the factors affecting tourism evaluation, the compendium of tourism resources, tourism facilities, tourism services, and the tourism environment is divided into three categories. The data organization management framework is shown in Figure 1.

2.5. Data Correlation Analysis. Data correlation analysis is conducted on the factors that affect the text category of online travel notes, and correlation analysis is conducted by using the Apriori algorithm to calculate the support degree, confidence degree, promotion degree, and confidence degree of each factor with positive and negative online travel notes.

2.6. Observation Indexes. The included online travel notes are statistically classified. The number of Mount Wutai online travel notes on different websites in different years is counted, the number of Mount Wutai online travel notes with different emotional attitudes is counted, and the number of online travel notes with revisit intentions and recommendation intentions is counted.

The evaluation results of the Bayesian decision tree classifier and C4.5 decision tree classifier on the emotional attitude of online travel journal texts are calculated, including positive text, neutral text, and negative text. The evaluation indexes mainly include the true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), false negative rate (FNR), precision, recall, and comprehensive evaluation index (F-measure). The calculation method is shown in equations (8)–(14), where TP is true positive, FN is false negative, TN is true negative, and FP is false positive.

$$TPR = \frac{TP}{TP + FN}, \quad (8)$$

$$TNR = \frac{TN}{TN + FP}, \quad (9)$$

$$FPR = \frac{FP}{TN + FP}, \quad (10)$$

$$FNR = \frac{FN}{TP + FN}, \quad (11)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%, \quad (12)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%, \quad (13)$$

$$F\text{-measure} = \frac{2 \times TP}{2 \times TP + FP + FN}. \quad (14)$$

The C4.5 decision tree and Bayesian classifier are drawn to evaluate the receiver operating characteristic curve (ROC) of web travel notes with different emotional attitudes.

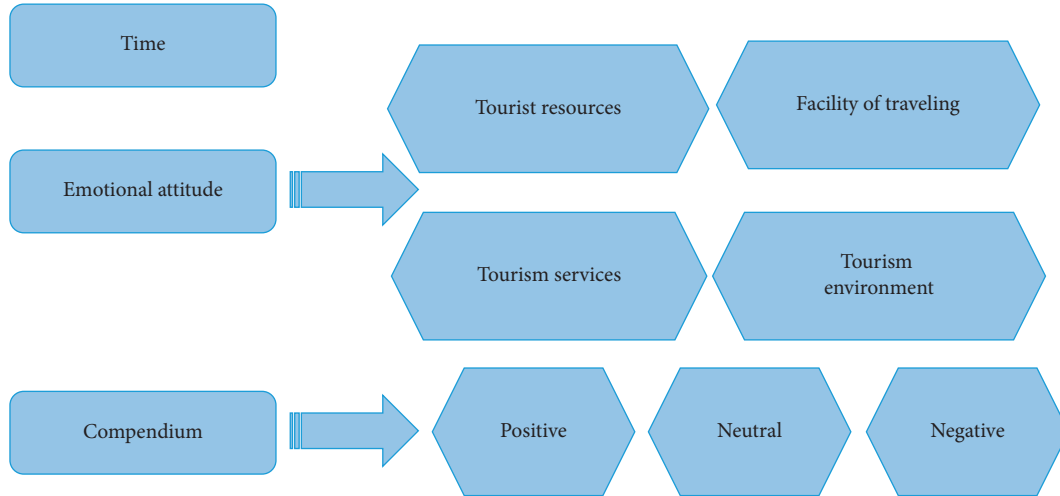


FIGURE 1: Text data organization framework of Mount Wutai tourism network travel notes.

The organization and management results of public data of travel notes on the Mount Wutai tourism network are collected, mainly including the number of tourism resources, tourism facilities, tourism services, and the tourism environment.

The support degree, confidence degree, promotion degree, and confidence degree of each factor and positive and negative online travel notes text are calculated. Support represents the proportion of events containing both X and Y to all events, as shown in (15). Confidence represents the proportion of events containing X that also contain Y events, as shown in (16). The promotion degree represents the proportion of events containing X that also contain Y events, as shown in (17). The calculation of confidence is shown in (18).

$$\text{Support} = P(X\&Y), \tag{15}$$

$$\text{Confidence} = \frac{P(X\&Y)}{P(X)}, \tag{16}$$

$$\begin{aligned} \text{Lift} &= (P(X\&Y)/P(X))/P(Y) \\ &= P(X\&Y)/P(X)P(Y), \end{aligned} \tag{17}$$

$$\text{Conviction} = \frac{1 - \text{Support}}{1 - \text{Confidence}}. \tag{18}$$

2.7. Statistical Methods. SPSS 20.0 is used for statistical analysis of the data, and a T test is used. Rate (%) represents counting data. ROC curves of the C4.5 decision tree and Bayesian classifier are plotted to evaluate online travel notes with different emotional attitudes. $P < 0.05$ is considered statistically significant.

3. Results

3.1. Public Data Result Statistics Based on Text Analysis. Figure 2 shows the statistics of Mount Wutai tourism articles of different categories. In Ctrip, there are 169 positive travel

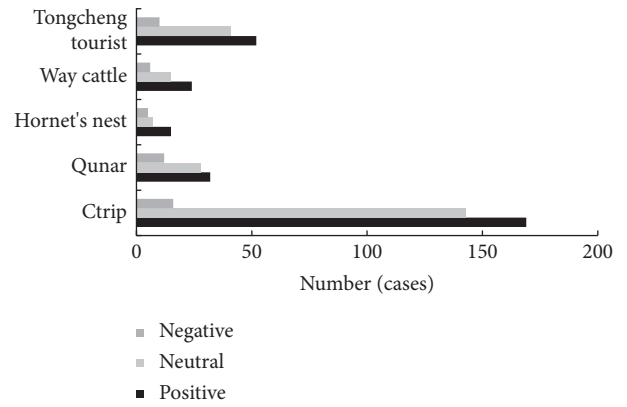


FIGURE 2: Statistics of Mount Wutai tourism articles of different categories.

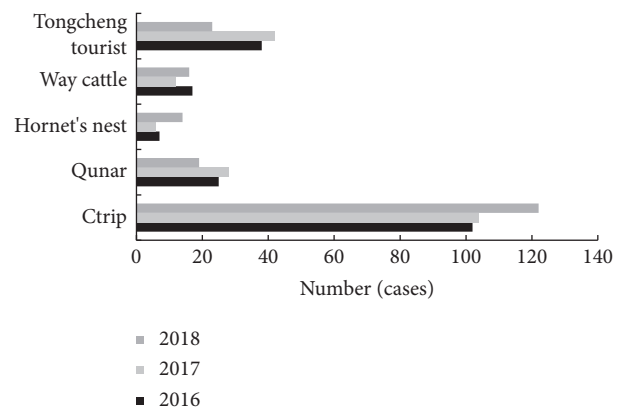


FIGURE 3: Statistics of Mount Wutai tourism articles at different times.

TABLE 3: Statistical table of emotional attitude of online travel notes text.

	Yes	No	Neutral
Revisit	59	22	494
Recommendation	158	65	352

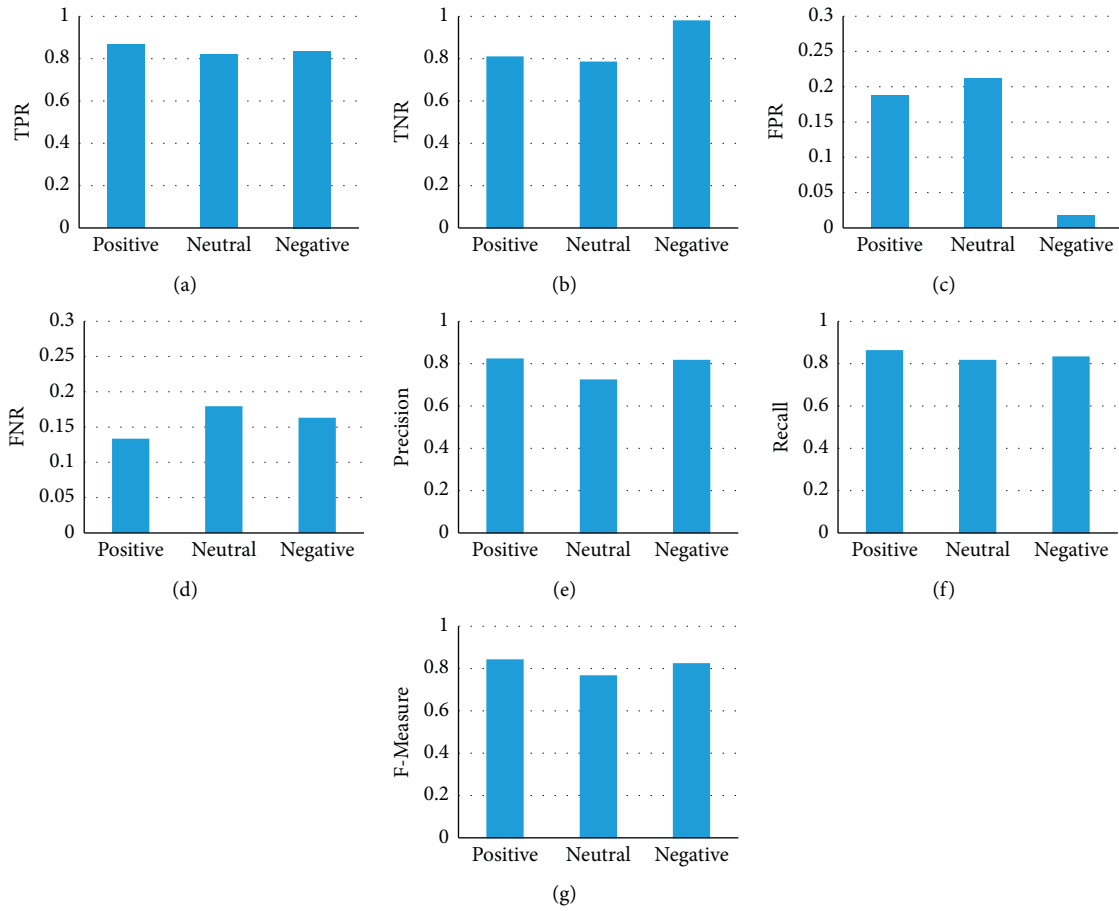


FIGURE 4: Text analysis-based Bayesian decision tree classification evaluation results of public data: (a) TPR, (b) TNR, (c) FPR, (d) FNR, (e) precision, (f) recall, and (g) F-measure.

notes, 143 neutral travel notes, and 16 negative travel notes. In Qunar, there are 32 positive travel notes, 28 neutral travel notes, and 12 negative travel notes. There are 15 positive travel notes, 7 neutral travel notes, and 5 negative travel notes in Hornet’s nest. In Tuniu, there are 24 positive travel notes, 15 neutral travel notes, and 6 negative travel notes. There are 52 positive travel notes, 41 neutral travel notes, and 10 negative travel notes on the Tongcheng website. Notably, all websites have the most positive online travel notes on Mount Wutai tourism, followed by neutral and negative ones. Figure 3 shows the statistics of Mount Wutai tourism articles at different times. Ctrip had 102 travel notes in 2016, 104 in 2017, and 122 in 2018. In Qunar, there were 25 travel notes in 2016, 28 in 2017, and 19 in 2018. There were 7 travel notes in 2016, 6 in 2017, and 14 in 2018 on Hornet’s nest. There are 17 travel notes on Tuniu in 2016, 12 in 2017, and 16 in 2018. There were 38 travel notes on Tongcheng in 2016, 42 in 2017, and 23 in 2018. Table 3 is the statistical table of the emotional attitude of online travel notes. The number of tourists who have the intention to revisit and recommend is significantly higher than that of those who do not.

3.2. Public Data Bayesian Decision Tree Classification and Evaluation Based on Text Analysis. Figure 4 shows the classification and evaluation results of the public data

Bayesian decision tree based on text analysis. The true positive rate and true negative rate of the Bayesian decision tree classifier for positive text are 86.64% and 81.27%, respectively. The false positive rate was 18.73%, and the false negative rate was 13.36%. The precision is 82.68%, recall is 86.64%, and F-measure is 84.62%. The true positive rate of the Bayesian decision tree classifier for neutral text is 82.05%, and the true negative rate is 78.89%. The false positive rate, false negative rate, precision, recall, and F-measure are 21.11%, 17.95%, 72.73%, 82.05%, and 77.11%, respectively. The true positive rate of the Bayesian decision tree classifier for negative text is 83.67%, the true negative rate is 98.29%, the false positive rate is 1.71%, and the false negative rate is 16.33%. The precision is 82.00%, recall is 83.67%, and F-measure is 82.83%. The F-measure of the Bayesian decision tree classifier for positive and negative texts is higher than that for neutral texts. Hence, the Bayesian decision tree classifier is better than neutral text in evaluating positive and negative texts.

3.3. Public Data C4.5 Decision Tree Classification and Evaluation Results Based on Text Analysis. Figure 5 shows the C4.5 decision tree classification evaluation results of public data based on text analysis. The true positive rate of the C4.5 decision tree classifier for positive text is 91.44%, the true negative rate is 86.57%, the false positive rate is 13.43%, and

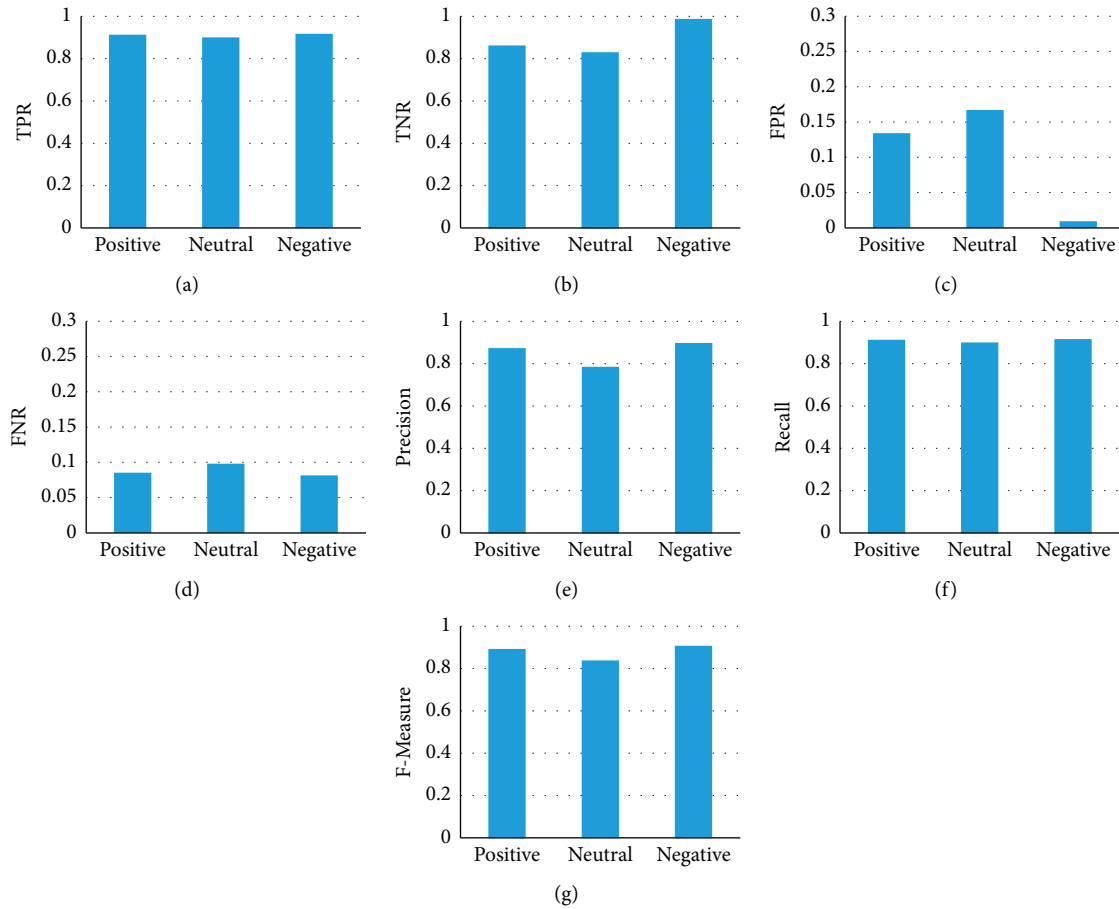


FIGURE 5: Public data C4.5 decision tree classification and evaluation results based on text analysis: (a) TPR, (b) TNR, (c) FPR, (d) FNR, (e) precision, (f) recall, and (g) F-measure.

the false negative rate is 8.56%. The precision, recall, and F-measure are 87.54%, 91.44%, and 89.45%, respectively. The true positive rate of the C4.5 decision tree classifier for neutral text is 90.17%, the true negative rate is 83.28%, the false positive rate is 16.72%, and the false negative rate is 9.83%. The precision, recall, and F-measure are 78.73%, 90.17%, and 84.06%, respectively. The true positive rate of the C4.5 decision tree classifier for negative text is 91.84%, the true negative rate is 99.05%, the false positive rate is 0.95%, and the false negative rate is 8.16%. The precision is 90.00%, recall is 91.84%, and F-measure is 90.91%. The F-measure of the C4.5 decision tree classifier for positive and negative texts is higher than that for neutral texts. Hence, the evaluation effect of the C4.5 decision tree classifier for positive and negative texts is better than that for neutral texts.

3.4. Evaluation Comparison between the C4.5 Decision Tree and Bayesian Classifier. Figure 6 shows the comparison of evaluation between the C4.5 decision tree and Bayesian classifier, where A is the positive text, B is the neutral text, and C is the negative text. The evaluation effect of the two classifiers on the positive text is significantly better than that of the neutral and negative text, and the evaluation effect of the C4.5 decision tree classifier is better than that of the Bayesian classifier.

3.5. Public Data Organization Manages Results. Figure 7 shows the organization and management results of public data of online travel notes of Mount Wutai tourism. There are 513 online travel notes related to tourism resources, 279 online travel notes related to tourism facilities, and 193 online travel notes related to tourism services. There are a total of 163 online travel notes related to the tourism environment, and a large number of them are related to tourism resources and tourism facilities.

3.6. Forward Common Data Association Analysis. Figure 8 shows the association analysis of forward public data, where A is the support degree, B is the confidence degree, C is the promotion degree, and D is the conviction degree. The promotion degree of tourism resources and facilities in forward online travel notes is significantly higher, indicating that tourism resources and facilities are highly correlated with forward online travel notes.

3.7. Negative Public Data Association Analysis. Figure 9 shows the association analysis of negative public data, where A is the support degree, B is the confidence degree, C is the improvement degree, and D is the confidence degree. The promotion degree of tourism service and the tourism

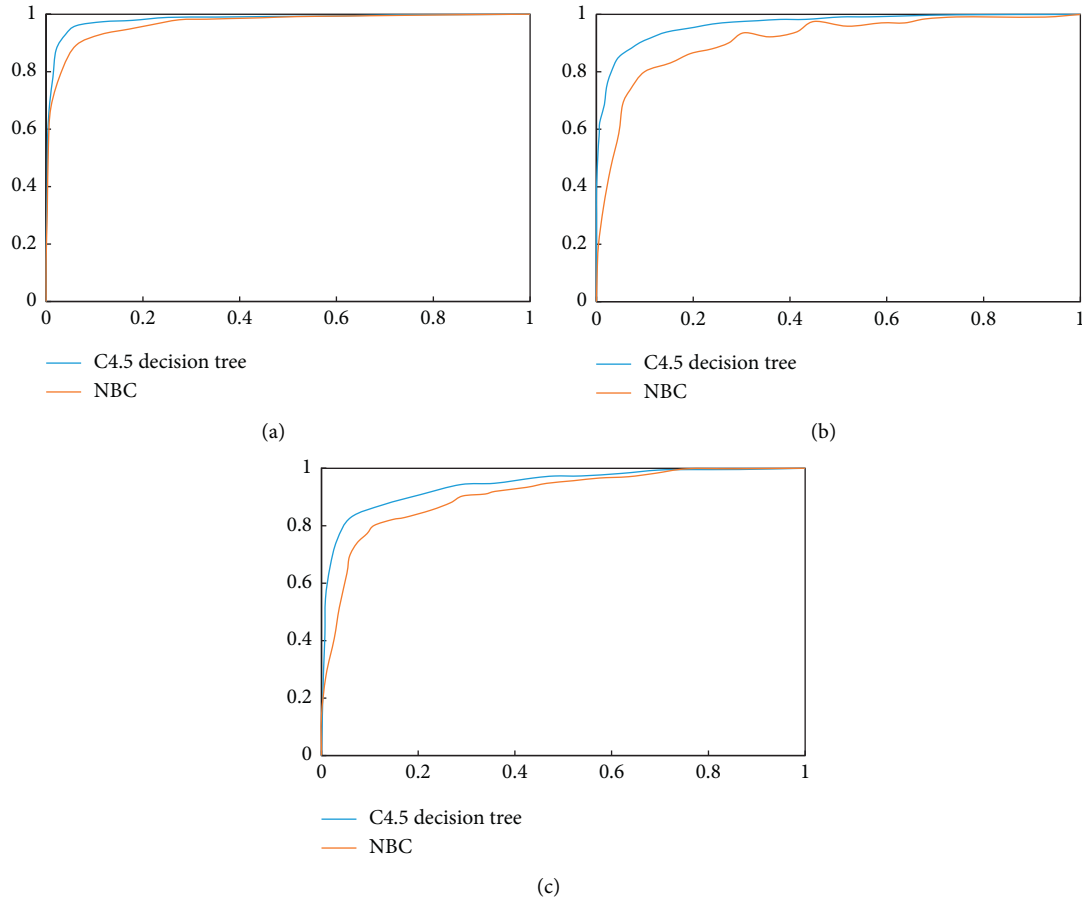


FIGURE 6: Evaluation comparison between the C4.5 decision tree and Bayesian classifier: (a) positive text, (b) neutral text, and (c) negative text.

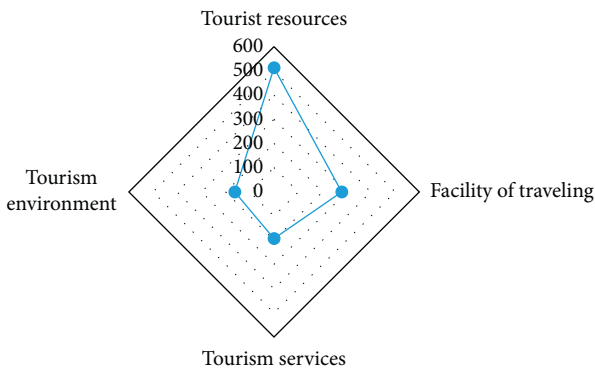


FIGURE 7: Mount Wutai travel notes text public data organization and management results.

environment in negative online travel notes is higher, indicating that tourism service and the tourism environment are highly correlated with negative online travel notes.

4. Discussion

The development of big data provides convenient conditions for people to collect and use information. With the development of informatization, the Internet has become the

main way for people to release and search for information [19, 20]. The new media form of tourism has been widely recognized. An increasing number of tourists express their feelings and attitudes during travel through the Internet, showing their travel process and expressing their travel feelings in the form of online travel notes and comments [21, 22]. These network texts are authentic and extensive and play an important role in shaping the image of tourist destinations and providing a reference for tourists [23]. Based on the online travel notes of tourism websites, the evaluation and correlation factors of tourists' emotional attitudes on Mount Wutai are analyzed. Text classification plays an important role in text analysis and management. Mccallum and Nigam [24] compared the naive Bayesian text classification model with the unigram language model with integer words and found that the naive Bayesian text classification model performed well and had certain advantages. Tong and Koller [25] applied text classification with support vector machine active learning and achieved good results. Fesseha et al. [26] carried out text classification of Tigrinian based on a convolutional neural network and found that a convolutional neural network has higher accuracy in classification compared with other traditional machine learning models. Hutama et al. [27] used naive Bayesian and decision tree to create a classification model and classified the text

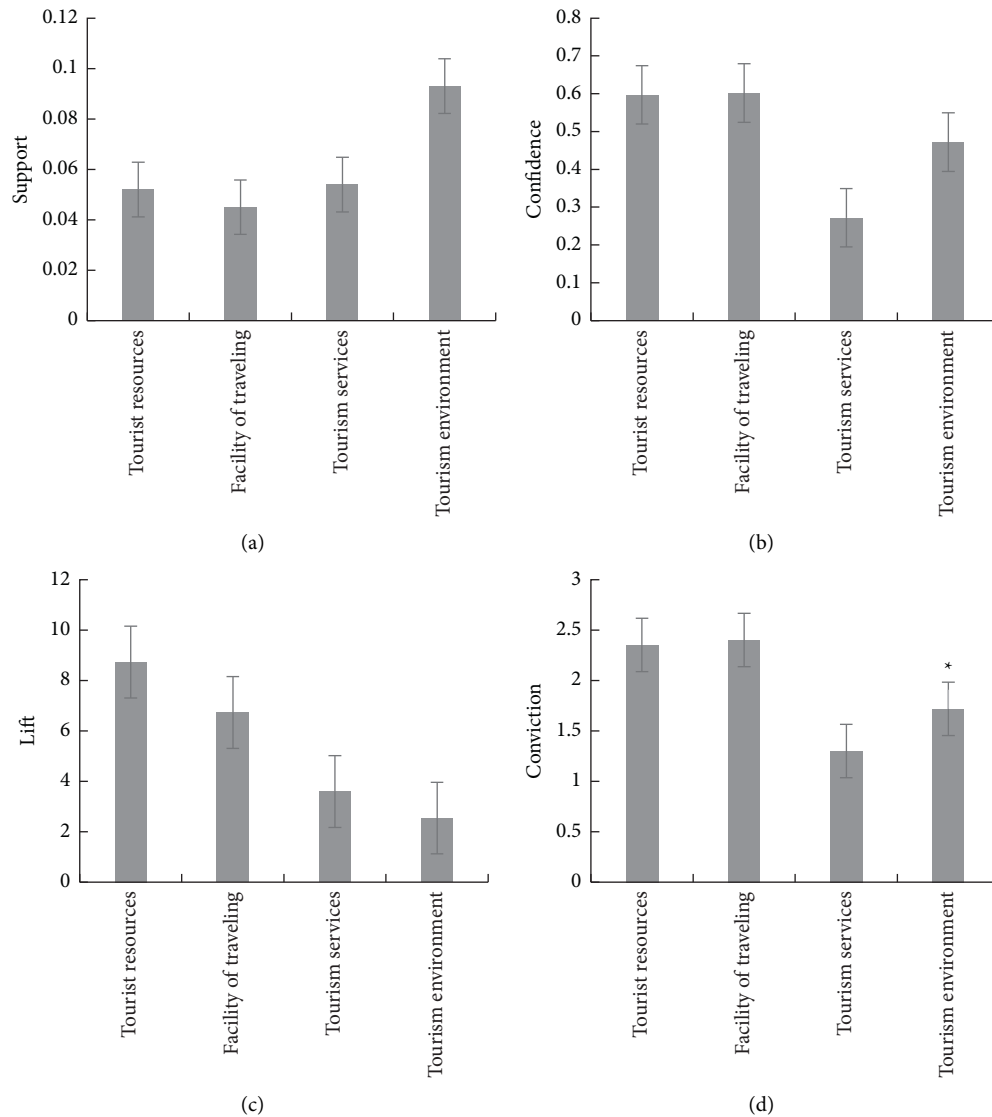


FIGURE 8: Positive association analysis of common data: (a) support, (b) confidence, (c) improvement, and (d) confidence.

data of work culture. They found that the accuracy of the three constructed work cultures was 33%, 66%, and 80%, respectively, and the accuracy of Bayesian was 83%, 50%, and 60%, respectively. Both methods had good performance.

The text classification method in this study adopts a naive Bayesian classification model and decision tree classification model [28–37]. The results show that the true positive rate of the Bayesian decision tree classifier for positive text is 86.64%, the true negative rate is 81.27%, and the F-measure is 84.62%. The true positive rate for neutral text is 82.05%, the true negative rate is 78.89%, and the F-measure is 77.11%. The true positive rate for negative text is 83.67%, the true negative rate is 98.29%, and the F-measure is 82.83%. Notably, the evaluation effect of the Bayesian decision tree classifier on positive and negative text is better than that of neutral text. The true positive rate of the C4.5 decision tree classifier for positive text is 91.44%, the true negative rate is 86.57%, and the F-measure is 89.45%. The true positive rate for neutral text is 90.17%, the true

negative rate is 83.28%, and the F-measure is 84.06%. The true positive rate for negative text is 91.84%, the true negative rate is 99.05%, and the F-measure is 90.91%. The C4.5 decision tree classifier has a better evaluation effect on positive and negative texts than on neutral texts. The ROC curve of the C4.5 decision tree and Bayesian classifier shows that the evaluation effect of the two classifiers on positive text is significantly better than that of neutral and negative text, and the evaluation effect of the C4.5 decision tree classifier is better than that of the Bayesian classifier.

This study also analyzes the correlation between the factors affecting the categories of online travel notes and the emotional attitudes of online travel notes. Using the Apriori algorithm for correlation analysis, the support degree, confidence degree, promotion degree, and confidence degree between tourism resources, tourism facilities, tourism services, and the tourism environment and online travel notes with different emotional attitudes are calculated. The results show that the promotion degree of tourism resources and

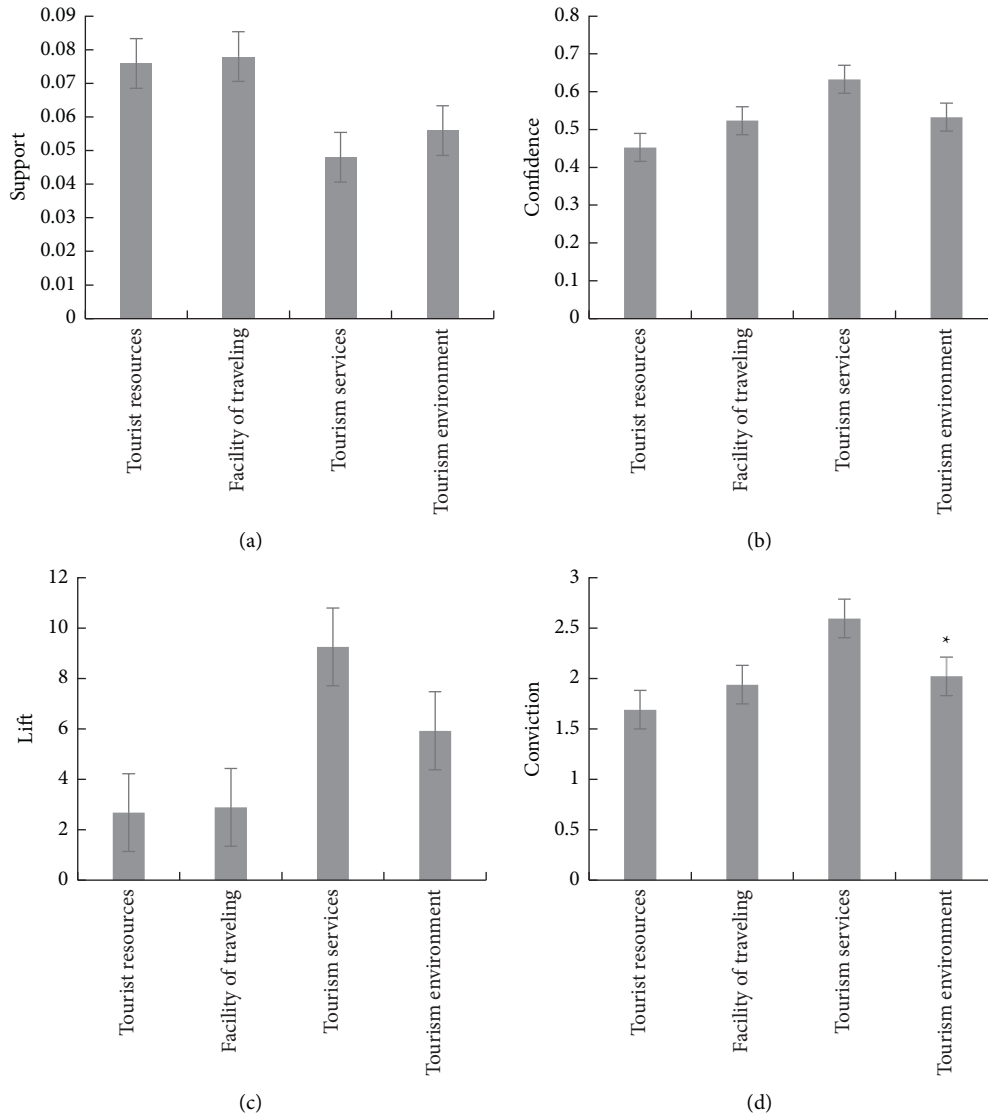


FIGURE 9: Negative association analysis of common data: (a) support, (b) confidence, (c) improvement, and (d) confidence.

facilities in the forward online travel notes is significantly higher, indicating that tourism resources and facilities are highly correlated with the forward online travel notes. The promotion degree of tourism service and the tourism environment in negative online travel notes is higher, indicating that tourism service and the tourism environment are highly correlated with negative online travel notes. The tourism service quality and environment of Mount Wutai scenic area should be greatly improved to make up the shortcomings and promote the development of Mount Wutai tourism.

5. Conclusion

In this study, a naive Bayesian classification model and decision tree classification model are used to classify the emotional attitude of Mount Wutai travel notes online, and it is found that the decision tree classification model has a better classification effect. The relationship between tourism resources, tourism facilities, tourism services, the tourism

environment, and the emotional attitude of online travel notes is discussed, and it is revealed that tourism resources and tourism facilities are strongly correlated with the positive text, while tourism services and the tourism environment are strongly correlated with the negative text.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] L. M. Spineli, C. Kalyvas, and K. Papadimitropoulou, "Continuous(ly) missing outcome data in network meta-analysis: a one-stage pattern-mixture model approach,"

- Statistical Methods in Medical Research*, vol. 30, no. 4, pp. 958–975, 2021 Apr.
- [2] T. C. Guetterman, T. Chang, M. DeJonckheere, T. Basu, E. Scruggs, and V. G. V. Vydiswaran, “Augmenting qualitative text analysis with natural language processing: methodological study,” *Journal of Medical Internet Research*, vol. 20, no. 6, p. e231, 2018 Jun 29.
 - [3] V. C. Kyara, M. M. Rahman, and R. Khanam, “Tourism expansion and economic growth in Tanzania: a causality analysis,” *Heliyon*, vol. 7, no. 5, p. e06966, 2021 May 6.
 - [4] J. Romão, “Tourism, smart specialisation, growth, and resilience,” *Annals of Tourism Research*, vol. 84, p. 102995, 2020 Sep.
 - [5] T. Dogru, U. Bulut, E. Kocak, C. Isik, C. Suess, and E. Sirakaya-Turk, “The nexus between tourism, economic growth, renewable energy consumption, and carbon dioxide emissions: contemporary evidence from OECD countries,” *Environmental Science and Pollution Research*, vol. 27, no. 32, pp. 40930–40948, 2020 Nov.
 - [6] Z. T. Shasha, Y. Geng, H. P. Sun, W. Musakwa, and L. Sun, “Correction to: past, current, and future perspectives on ecotourism: a bibliometric review between 2001 and 2018,” *Environmental Science and Pollution Research*, vol. 27, no. 19, p. 23514, 2020 Jul.
 - [7] A. Sinha, O. Driha, and D. Balsalobre-Lorente, “Tourism and inequality in per capita water availability: is the linkage sustainable?” *Environmental Science and Pollution Research*, vol. 27, no. 9, pp. 10129–10134, 2020 Mar.
 - [8] I. Bulatovic and K. Iankova, “Barriers to medical tourism development in the United Arab Emirates (UAE),” *International Journal of Environmental Research and Public Health*, vol. 18, no. 3, p. 1365, 2021 Feb 2.
 - [9] E. A. Grigorieva, “Adventurous tourism: acclimatization problems and decisions in trans-boundary travels,” *International Journal of Biometeorology*, vol. 65, no. 5, pp. 717–728, 2021 May.
 - [10] Z. Tian, Z. Shi, and Q. Cheng, “Examining the antecedents and consequences of mobile travel app engagement,” *PLoS One*, vol. 16, no. 3, p. e0248460, 2021 Mar 12.
 - [11] L. Niu and Z. Cheng, “Impact of tourism disturbance on forest vegetation in Wutai Mountain, China,” *Environmental Monitoring and Assessment*, vol. 191, no. 2, p. 81, 2019 Jan 17.
 - [12] X. Wang and Y. Tong, “Application of an emotional classification model in e-commerce text based on an improved transformer model,” *PLoS One*, vol. 16, no. 3, p. e0247984, 2021 Mar 5.
 - [13] C. Ling, X. Wei, Y. Shen, and H. Zhang, “Development and validation of multiple machine learning algorithms for the classification of G-protein-coupled receptors using molecular evolution model-based feature extraction strategy,” *Amino Acids*, vol. 53, no. 11, pp. 1705–1714, 2021 Nov.
 - [14] L. Chen, Q. Li, H. Song et al., “Classification of schizophrenia using general linear model and support vector machine via fNIRS,” *Phys Eng Sci Med*, vol. 43, no. 4, pp. 1151–1160, 2020 Dec.
 - [15] S. Bakheet and A. Al-Hamadi, “A framework for instantaneous driver drowsiness detection based on improved HOG features and naïve bayesian classification,” *Brain Sciences*, vol. 11, no. 2, p. 240, 2021 Feb 14.
 - [16] F. Li, Y. Shen, D. Lv et al., “A Bayesian classification model for discriminating common infectious diseases in Zhejiang province, China,” *Medicine (Baltimore)*, vol. 99, no. 8, p. e19218, 2020 Feb.
 - [17] X. Luo, X. Wen, M. Zhou, A. Abusorrah, and L. Huang, “Decision-tree-initialized dendritic neuron model for fast and accurate data classification,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 17, pp. 1–11, 2021 Mar.
 - [18] M. Amgad, L. A. Atteya, H. Hussein et al., “Explainable nucleus classification using decision tree approximation of learned embeddings,” *Bioinformatics*, vol. 38, no. 2, pp. 513–519, 2021 Sep.
 - [19] X. Zhou, W. Liang, K. I. K. Wang, R. Huang, and Q. Jin, “Academic influence aware and multidimensional network analysis for research collaboration navigation based on scholarly big data,” *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 1, pp. 246–257, 2021.
 - [20] J. Qiu, Y. Chai, Z. Tian, X. Du, and M. Guizani, “Automatic concept extraction based on semantic graphs from big data in smart city,” *IEEE Transactions on Computational Social Systems*, vol. 7, no. 1, pp. 225–233, 2020.
 - [21] Z. Yuan, “Big data recommendation research based on travel consumer sentiment analysis,” *Frontiers in Psychology*, vol. 13, p. 857292, 2022 Feb 28.
 - [22] W. Jaung and L. R. Carrasco, “Travel cost analysis of an urban protected area and parks in Singapore: a mobile phone data application,” *Journal of Environmental Management*, vol. 261, p. 110238, 2020 May 1.
 - [23] M. Fuchs, W. Höpken, and M. Lexhagen, “Big data analytics for knowledge generation in tourism destinations – a case from Sweden,” *Journal of Destination Marketing & Management*, vol. 3, no. 4, pp. 198–209, 2014.
 - [24] A. McCallum and K. Nigam, “A comparison of event models for Naive Bayes text classification,” *AAAI WORKSHOP ON LEARNING FOR TEXT CATEGORIZATION*, vol. 752, no. 1, pp. 41–48, 1998.
 - [25] S. Tong and D. Koller, “Support vector machine active learning with applications to text classification,” *Journal of Machine Learning Research*, vol. 2, no. 1, pp. 999–1006, 2002.
 - [26] A. Fesseha, S. Xiong, E. D. Emiru, M. Diallo, and A. Dahou, “Text classification based on convolutional neural networks and word embedding for low-resource languages: Tigrinya,” *Information*, vol. 12, no. 2, p. 52, 2021.
 - [27] N. Y. Hutama, K. M. Lhaksmana, and I. Kurniawan, “Text analysis of applicants for personality classification using multinomial naïve bayes and decision tree,” *JURNAL INFOTEL*, vol. 12, no. 3, pp. 72–81, 2020.
 - [28] A. Babbain, A. Abdulhakim, I. Elbatal, and E. M. Almetwally, “Bayesian and non-bayesian reliability estimation of stress-strength model for power-modified lindley distribution,” *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 1154705, 59 pages, 2022.
 - [29] L. Li, C. Mao, H. Sun, Y. Yuan, and B. Lei, “Digital twin driven green performance evaluation methodology of intelligent manufacturing: hybrid model based on fuzzy rough-sets AHP, multistage weight synthesis, and PROMETHEE II,” *Complexity*, vol. 2020, no. 6, 24 pages, Article ID 3853925, 2020.
 - [30] C. Li, S. Xiong, X. Sun, and Y. Qin, “Bayesian analysis for metro passenger flows using automated data,” *Mathematical Problems in Engineering*, vol. 2022, pp. 1–12, 2022.
 - [31] L. Li and C. Mao, “Big data supported PSS evaluation decision in service-oriented manufacturing,” *IEEE Access*, vol. 8, no. 99, pp. 154663–154670, 2020.
 - [32] F. Noor, S. Masood, M. Zaman et al., “Bayesian analysis of inverted kumaraswamy mixture model with application to burning velocity of chemicals,” *Mathematical Problems in Engineering*, vol. 2021, Article ID 5569652, 18 pages, 2021.

- [33] L. Li, T. Qu, Y. Liu et al., "Sustainability assessment of intelligent manufacturing supported by digital twin," *IEEE Access*, vol. 8, pp. 174988–175008, 2020.
- [34] S. Liao and Z. Liu, "Enterprise financial influencing factors and early warning based on decision tree model," *Scientific Programming*, vol. 2022, Article ID 6260809, 8 pages, 2022.
- [35] H. Liu and J. Liu, "Female employment data analysis based on decision tree algorithm and association rule analysis method," *Scientific Programming*, vol. 2022, Article ID 8994349, 11 pages, 2022.
- [36] L. Li, B. Lei, and C. Mao, "Digital twin in smart manufacturing," *Journal of Industrial Information Integration*, vol. 26, no. 9, Article ID 100289, 2022.
- [37] H. Zhang, L. Gao, H.-G. Luo, and Y. Zhai, "Research on the RFID anticollision strategy based on decision tree," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2913157, 7 pages, 2022.