ENTHUSIASTIC
INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE
Volume 4, Issue 1, April 2024, pp. 37-50

Analyzing Tourist Satisfaction Using Factor Analysis and Text Mining: An Ecotourism Study in Girpasang Village

Kariyam [a,1,*], Tasya Apriliana [a,2], Nur Aulia Maknunah [a,3], Hafis Muhammad Nizam [a,4], Rizky Mardhatillah [a,5], Rahma Fatwa Nova [a,6]

[a] Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, Jl. Kaliurang KM 14.5, Yogyakarta 55584, Indonesia
[1] kariyam@uii.ac.id *; [2] 20611088@students.uii.ac.id; [3] 20611093@students.uii.ac.id; [4] 20611097@students.uii.ac.id; [5] 20611100@students.uii.ac.id; [6] 20611122@students.uii.ac.id
* Corresponding author

ARTICLE INFO

Keywords
tourist satisfaction
factor analysis
text mining
rural tourism
Girpasang village

ABSTRACT

In the second half of 2022, the tourism industry started recovering from the vast impacts of the COVID-19 pandemic. Tourism is one of the most feasible sources of income for the small, rural village of Girpasang, situated at the heights of Mount Merapi. Tourist satisfaction has been attributed to the success of tourist destinations and is, therefore, a benchmark for their development. This study aimed to explain the factors that affected tourist satisfaction and other underlying aspects that call for improvement, using confirmatory factor analysis and text mining. The data used was collected from a total of 102 respondents at Girpasang Village within two days. The results showed that there were five common factors affecting tourist satisfaction: staff attitude, reliability of tourist facilities, comfort of tourist facilities, comprehensiveness of facilities provided, and tangible condition of the environment. Based on text mining results of tourist critics, it was found that access roads were the most profound complaint.

1. Introduction

As economic activities have returned to normal and restrictions on community activities have lifted after the COVID-19 pandemic, tourism activities have resumed. Among them are nature tours, or ecotourism, which offers an escape from the burdens of city life. This type of tourism is also called rural tourism. According to United Nations World Tourism Organization (UNWTO), rural tourism is a type of tourism activities that allow the tourists to experience various elements such as nature-based activities, agriculture, rural lifestyle and culture, angling, and sightseeing. It involves engaging with a diverse range of products and experiences typically found in rural settings [1]. Therefore, during this post-pandemic era, numerous tourist areas are experiencing a surge in the volume of tourists.

A study examining the phenomenon of selfie spot management has emphasized that managers of tourist attractions properly maintain selfie facilities to ensure tourists' safety. If accidents occur, the negative news will quickly spread on social media and negatively impact the tourist attractions.

p-ISSN 2798-253X
e-ISSN 2798-3153

Although the attractions, such as the cable car and suspension bridge available at the village of Girpasang, are a strong attraction for tourists, both carry a significant risk for passengers or tourists who ride them if they are not appropriately maintained [2]. Thus, this research was conducted with the aim to create a sense of urgency to ensure the satisfaction and comfort of Girpasang tourism, thereby enabling this tourist attraction to uphold its status as the top destination in Klaten.

Some works have done to guide the direction of this research. A study conducted in 2021 by researchers in Nigeria that aimed to assess the relationship between service quality, student satisfaction, and loyalty in the Nigerian higher education experience utilized a survey method with structured questionnaires. A total of 265 students from a private university in Nigeria participated in the study by providing data through these questionnaires. The SERVQUAL measurement model from a highly cited research paper [3], was utilized to determine service quality in the higher education institution of Nigeria [4]. The SERVQUAL model is a widely recognized tool for measuring service quality based on five dimensions: reliability, responsiveness, assurance, empathy, and tangibility. By applying this model, the researchers were able to assess the quality of services provided by the educational institution and its impact on student satisfaction and loyalty in the Nigerian context.

Research conducted in 2021 utilized the Confirmatory Factor Analysis (CFA) method along with Structural Equation Modelling (SEM) and descriptive analysis to study the influences of airport service quality, tourist satisfaction, and image on behavioral intention towards destination visit [5]. The data collection involved distributing online and paper-based questionnaires with 7-point Likert-type scale questions based on dimensions of service quality, satisfaction, airport image, and behavioral intention. The results of analysis on the CFA model, which evaluated indicators such as the Goodness of Fit Index (GFI), comparative fit index (CFI), Tucker–Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA), concluded that the model suited the data well. The interpretation on the CFA results suggested that passengers were more likely to return when improvements of airport service quality were made. The higher the satisfaction level of passengers the higher the likelihood of return visits, and passengers are more likely to have behavioral intent of visiting destinations within the country when they are satisfied with the airport services [5].

Research conducted in 2022 analyzed the successes and shortcomings of companies in dealing with public opinion networks, to provide the right decision support for companies. In the text mining method, the evolution of public opinion networks, semantic networks, and sentiment trends were compared and then analyzed. Through comparative analysis, it was known that comments on Pinduoduo's website focused on corporate institutions and labor laws revolved around several important nodes. Meanwhile comments on Huolala's website focused on the truth and improvements the company was making. This research has put forward appropriate strategies and suggestions for companies to effectively respond to online public opinion crises [6].

This study employed CFA on survey data to investigate the factors affecting tourist satisfaction in Girpasang Village. Other methods including text analysis and descriptive analysis were also utilized to provide further insight on the respondents' profile as well as examine open-ended opinions and feedback from the respondents.

2. Method

The sample selection method used was accidental sampling, a nonprobabilistic technique where a respondent meeting the criteria is spontaneously chosen when encountering the researcher [7]. Based on this method, the sample of the population was tourists who visited the Girpasang tourism site on Sunday and Saturday. These responders were above the age of 15.

2.1. Research Variables

The researcher determined the independent variables in this study based on the five dimensions of satisfaction, which were used to measure satisfaction towards service quality such as in [4]. The five dimensions are as follows.

p-ISSN 2798-253X
e-ISSN 2798-3153

a. Tangibles: physical facilities, equipment and appearance of personnel encountered.
b. Reliability: the ability to fulfill promised services accurately and reliably.
c. Responsiveness: the desire to help customers and immediately in carrying out services.
d. Assurance: knowledge and courtesy of employees and their ability to create a sense of comfort.
e. Empathy: a sense of caring, special attention given by a business to its customers.

Thus, the researcher arranged the independent variables into question items, as shown on Table 1. The response obtained on each question item were used to build the CFA model.

Table 1. Description of Research Variables

Dimension of Satisfaction	Variable	Description
Tangibles	X_1	Spacious parking area is available
	X_2	Restaurants or stalls providing food supply are available
	X_3	Toilets function properly
	X_4	Prayer room is available
	X_5	The entire area of Girpasang Village is clean
	X_6	The cable car is in good condition
	X_7	The suspension bridge is in good condition
	X_8	The road surface on the way to Girpasang Village is smooth
	X_9	Traffic on the way to Girpasang Village is not congested
	X_{10}	The journey to Girpasang Village is pleasant
Reliability	X_{11}	It is possible to enjoy the natural scenery from the cable car
	X_{12}	It is possible to enjoy the natural scenery from the suspension bridge
	X_{13}	High quality photos can be taken at the tourist attraction
	X_{14}	The price of food at the location matches its value
Responsiveness	X_{15}	Parking attendants at the location are very helpful
	X_{16}	Staff at the village gives prompt service
	X_{17}	Staff at the village are always willing to help
	X_{18}	Staff at the village offers help according to each tourist's needs
Assurance	X_{19}	Tourist feels safe when using the cable car
	X_{20}	Tourist feels safe when crossing the suspension bridge
	X_{21}	Tourist feels safe at Girpasang Village
	X_{22}	Staff at the village are knowledgable towards matters involving the tourist attraction
	X_{23}	Staff at the. Village are always courteous
Empathy	X_{24}	Staff are attentive towards tourists
	X_{25}	Staff are attentive towards the care and maintenance of the cable car
	X_{26}	Staff are attentive towards the care and maintenance of the suspension bridge
	X_{27}	Small, local businesses at the village are honest in their trade
	X_{28}	There are clear directions on the way to the village.

While the research focused on the analysis of factors of satisfaction, which is an outcome of the CFA method, such analysis is limited to scale-based quantitative data, which are known to cause information distortion due to the respondents' inability to express opinions accurately. Real-world scenarios often comprise indeterminacy that cannot be captured using rigid scale-based instruments [8]. Several variables were collected and analyzed using descriptive analysis to provide more comprehensive information; these variables were name, age, gender, city of origin, profession, source of information promoting the Girpasang Village, willingness to recommend Girpasang Village to others, and willingness to return as a tourist in the future. The distributed questionnaire allowed for open-ended responses, which resulted in respondents providing feedback and suggestions using natural language.

p-ISSN 2798-253X
e-ISSN 2798-3153

2.2. Data Collection Method

The data collection technique used in this research was a survey. A survey is a research method that involves asking standardized questions to a group of people, which can be analyzed statistically. An electronic and physical questionnaire was utilized. The team worked in pairs, with one person as the interviewer and one as the note-taker. This survey used a cross-sectional method, meaning that the data obtained was sampled at one time.

This study used the Likert scale, a psychometric scale commonly used in questionnaires, designed to measure respondents' attitudes, opinions, or perceptions toward a statement. In this study, five answer choices were used: strongly agree, agree, neutral, disagree, and strongly disagree [9].

2.3. Confirmatory Factor Analysis (CFA) Assumptions

CFA is a statistical analysis tool that reduces the number of factors that affect a variable to only a few indicators without dropping any meaningful information. The main goal is to simplify factors to provide a clear interpretation of them. Factor analysis was used for initial research, wherein the factors influencing a variable had not been identified properly. In factor analysis, all variables are checked for correlation patterns, resulting in the grouping of variables with similar roles [10].

The assumptions that are a requirement of factor analysis are:

a. Adequacy of variables for factor analysis

This assumption was measured by the Kaiser-Mayer-Olkin (KMO) index, also known as Measure of Sampling Adequacy (MSA), proposed by Kaiser in 1970. It checks for the eligibility of the sample to be analyzed by the factor analysis method. If the index is between 0.5 and 1, then the variables are sufficiently correlated. The formula of the KMO index for a dataset with p variables is as follows [11]:

$$MSA = \frac{\sum_{i \neq k}^{p} r_{ik}^2}{\sum_{i \neq k}^{p} r_{ik}^2 + \sum_{i \neq k}^{p} a_{ik}^2} \quad (1)$$

where $i = k = 1, 2, \ldots, p$, r_{ik} indicates correlation coefficient of variables *i* and *k*, and a_{ik} indicates partial correlation coefficient of variables *i* and *k*.

b. Multicollinearity

The Bartlett's Test of Sphericity was used to test for correlation between variables. The hypothesis is as follows:

H_0: $\boldsymbol{R} = \mathbf{I}$ (The correlation matrix is an identity matrix (the variables are not correlated with each other))

H_1: $\boldsymbol{R} \neq \mathbf{I}$ (The correlation matrix is not an identity matrix (there are significant correlations among some of the variables)).

The formula for Bartlett's Test statistic is as follows [11]:

$$\chi^2 = -\left[(n-1) - \frac{(2p+5)}{6}\right].\log(|\mathbf{R}|) \quad (2)$$

where n denotes number of observations and $|\mathbf{R}|$ = determinant of correlation matrix.

2.4. CFA Model

$\boldsymbol{X} = [x_1 \quad x_2 \quad \cdots \quad x_p]$ is an observed random vector with p components, which has mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The factor model postulates that $\boldsymbol{XX}$ is linearly dependent on m unobserved random variables $F_1, F_2, \ldots, F_m$, which are called common factors; $\boldsymbol{X}$ is also dependent on p specific factors. The orthogonal model with m specific factors is [12]:

$$\begin{aligned} X_1 - \mu_1 &= +\ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= +\ell_{21}F_1 + \ell_{22}F_2 + \cdots + \ell_{2m}F_m + \varepsilon_2 \\ &\vdots \end{aligned} \quad (3)$$

$$X_p - \mu_p = + \ell_{p1}F_1 + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \varepsilon_p$$

where μ_i denotes mean of the *i*th variable, ℓ_{ij} denotes loading of *i*th variable on the *j*th factor, F_j denotes the *j*th common factor, ε_i denotes *i*th specific factor. $\boldsymbol{F}$ and $\boldsymbol{\varepsilon}$ are unobserved random vectors which satisfy the assumptions: (a) $\boldsymbol{F}$ and $\boldsymbol{\varepsilon}$ are independent, (b) $E(\boldsymbol{F}) = 0$, $cov(\boldsymbol{F}) = \boldsymbol{I}$, and (c) $E(\boldsymbol{\varepsilon}) = 0$, $cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$. Here, $\boldsymbol{\Psi}$ denotes a diagonal matrix. Covariance structure of orthogonal factors [10]:

$$cov(\mathbf{X}) = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} \quad (4)$$

$$cov(\mathbf{X}, \mathbf{F}) = \mathbf{L} \text{ or } cov(\mathbf{X}_i, \mathbf{F}_j) = \ell_{ij} \quad (5)$$

2.5. Estimation of Loadings and Communalities

The definition of the loadings of standardized variables is the correlation between the *i*th variabel and the *j*th common factor.

$$corr(x_i, f_j) = \ell_{ij} \quad (6)$$

Communality is the common variance of x_i due to the common factors. The formula to calculate communality is given by [12].

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2 \quad (7)$$

Sum of squared loadings on the *j*th column is the contribution of the *j*th factor towards the total sample variance. The proportion of variance due to the *j*th factor is given by the sum of squared loadings divided by the number of variables in the model.

$$\sum_{j=1}^{p} \ell_{ij}^2 = \ell_{1j}^2 + \ell_{2j}^2 + \cdots + \ell_{pj}^2 \quad (8)$$

$$\sum_{j=1}^{p} \ell_{ij}^2 / tr(\mathbf{R}) = \sum_{j=1}^{p} \ell_{ij}^2 / p \quad (9)$$

where $\mathbf{R}$ is the correlation matrix.

2.6. Factor Rotation

All loadings obtained from the initial loading through orthogonal transformations have the same ability to produce covariance or correlation matrices [12]. Sometimes the results of the estimated loading are difficult to interpret because the loading of all variables has a large value on only one particular factor, making it difficult to form groups [10]. Therefore, factors are rotated for interpretation purposes while maintaining all the qualities of the initial factors.

If $\hat{\mathbf{L}}$ is the matrix of estimated factor loadings with $p \times m$ dimensions, then the rotated loadings

$$\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T} \quad (10)$$

where $\mathbf{T}$ is an orthogonal transformation matrix. The most frequently used rotation is the varimax rotation proposed by Kaiser in 1958 [12]. The Varimax rotation procedure selects a transformation matrix $\mathbf{T}$ that maximizes the variance of loadings.

2.7. Factor Analysis Procedure

This stage initiated with determining the indicator variables that will be the input for analysis. After that, the partial correlation between variables was calculates and examined. The partial correlation test between variables was then conducted using Bartlett's Test of Sphericity. Following testing the partial correlation, the adequacy and eligibility of the data was tested to use the factor analysis method with the MSA test. Subsequently, the number of factors with the greatest contribution (eigenvalue greater than 1) was determined. Next, factors with varimax rotation were arranged to further understand the grouping of each variable. Last, he results of factor analysis were interpreted.

p-ISSN 2798-253X
e-ISSN 2798-3153

2.8. Text Analysis

Text analysis, also known as text mining, is a technique used to extract valuable and meaningful information from large amounts of unstructured text data. In this study, the feedback data on Girpasang Village was obtained through primary data collection, which required extensive preprocessing of the raw data. Thus, several preprocessing methods were employed, including tokenization, lemmatization, removal of stop words, and removal of nonalphabetic characters [13].

From the data, the term frequencies were extracted as they provided insight into the recurring topics within the tourist's feedback. The appearance of certain words simultaneously in one unit of sentence or paragraph can be extracted using correlational analysis [14]. The correlation coefficient was used to describe the degree to which a pair of terms appeared in a sentence unit. It ranges from 0 to 1, with 1 representing the highest degree of correlation. Based on the correlation of certain words with words with negative, positive, or neutral connotations, insights can be drawn from customer feedback [15].

3. Results and Discussion

The survey, which was conducted on Saturday, November 19 and Sunday, November 20, 2022, had 102 respondents who were 15 years of age or older. The measurement of tourist satisfaction was carried out based on 28 variables that had been formulated with reference to the SERVQUAL questionnaire. Data was analyzed using factor analysis method and had discrete ordinal properties from the use of a psychometric scale, namely the Likert scale with a range of 1 to 5. After inspection, there were no missing data in the 28 numeric columns.

3.1. Respondent Profile

By employing exploratory data analysis methods, this study attempted to describe and summarize the data by presenting it in graphical form, facilitating analysis, and drawing conclusions. Based on a survey conducted using 37 variables and involving 102 tourists who visited Girpasang Village, male and female tourists had an equal number of respondents, which was 51 respondents for each. Fig. 1. is a pie chart visualization for the number of respondents based on gender.

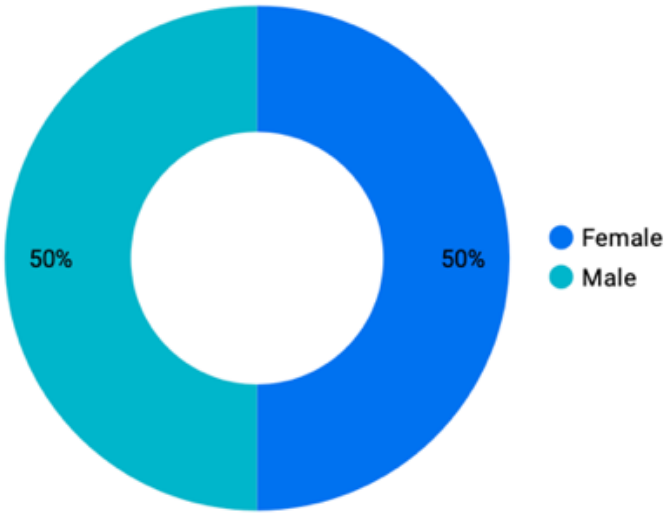

Fig. 1 Pie chart of the gender count of respondents.

Based on the histogram of respondents' ages, it was found that the highest frequency was in the age range of 20–25 years (28 tourists), which was the age of most tourists visiting the Girpasang Village. Typically, respondents of this age group are university students or young working adults. Additionally, there were 24 tourists ranging from 15–20 years old, 11 from 25–30 years old, and 10 from 50–55 years old. Meanwhile, other age ranges were relatively low.

p-ISSN 2798-253X
e-ISSN 2798-3153

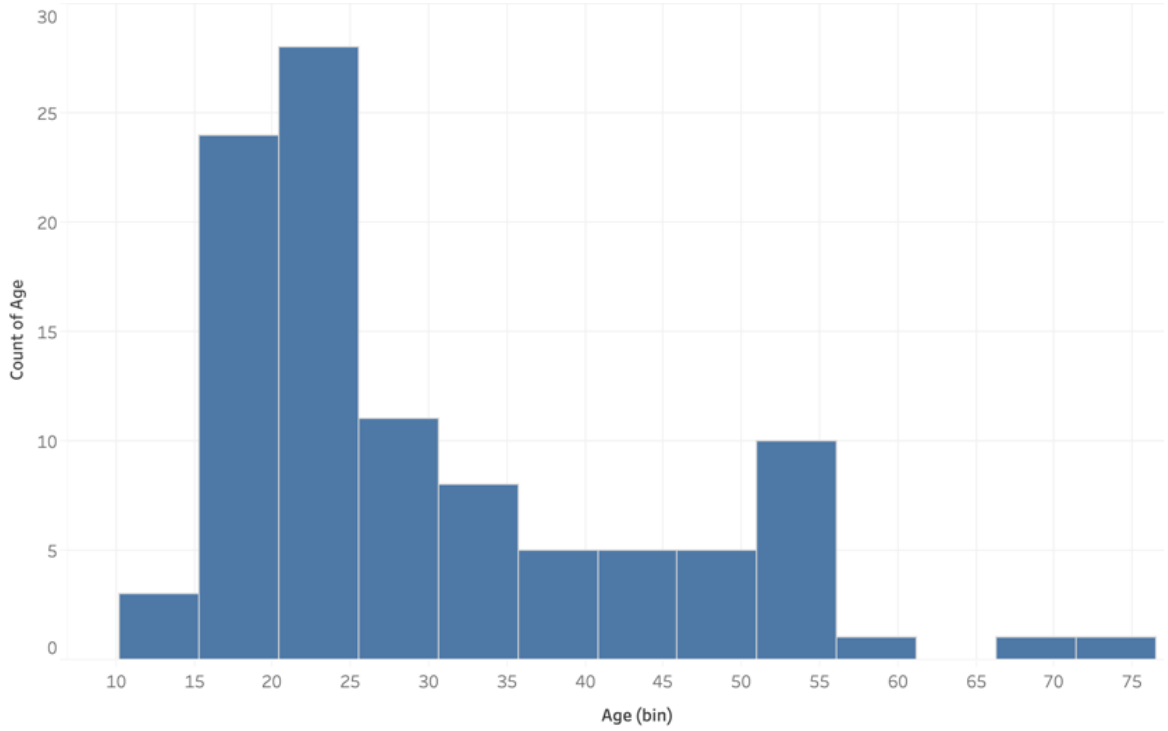

Fig. 2 Histogram of the age distribution of respondents.

Survey data of respondents' province of origin areas was visualized using the tree map chart (Fig. 1), where the dark purple color indicates the province where most respondents came from, namely tourists from Central Java province with a total of 75 people. The second place was the Special Region of Yogyakarta, which had 16 tourists. East Java and regions outside Java were in third place, with 4 tourists each. The last place was Jakarta, with several tourists of as many as 3 people.

Fig. 3 Tree map of respondents' place of residence.

The number of respondents based on their profession is presented as a bar chart. Most tourists came from other professions (not within the determined options). The number of tourists in this category was 30. They were then followed by the student and private sector employees, both at 29

p-ISSN 2798-253X
e-ISSN 2798-3153

persons. A total of seven respondents worked as civil servants, while the rest belonged to other categories.

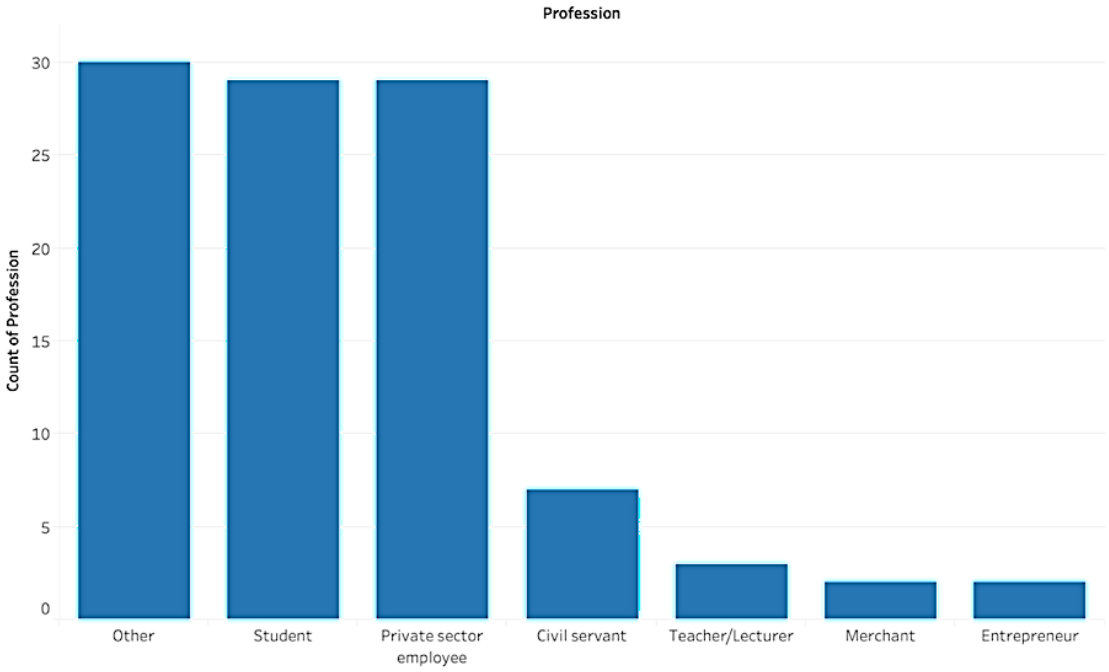

Fig. 4 Bar chart of respondents' profession.

The count of respondents' sources of information regarding Girpasang Village was visualized in the form of a bubble graph in Fig. 5, where 43.14% of respondents received information from friends/family/relatives. In second place, with a percentage of 22.55%, respondents received information from Instagram; the third position was other undetermined sources, with a percentage of 16.67%; the fourth position, with a percentage of 12.75%, was TikTok; the fifth position was online articles, where the percentage figure obtained was 4.90%. The last one is Facebook, with a percentage of 0.98%.

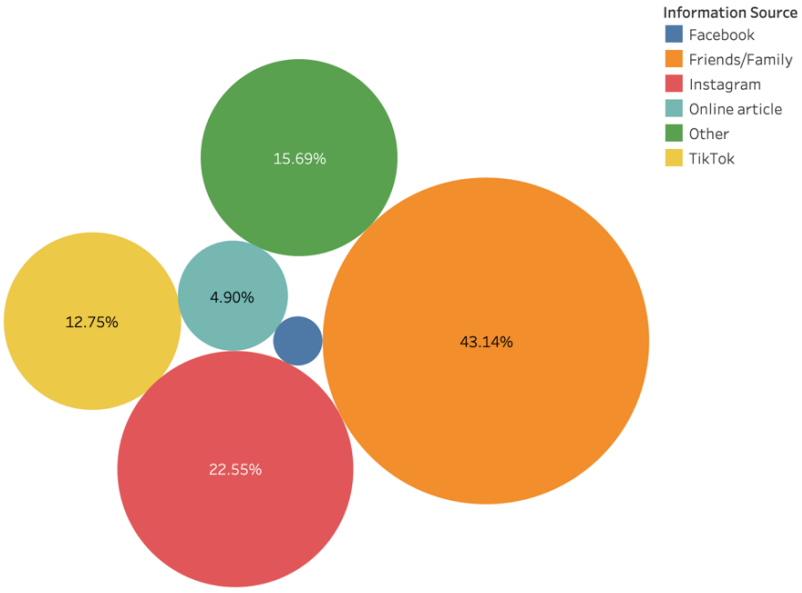

Fig. 5 Bubble chart of source of information on Girpasang Village.

Fig. 6 visualizes the survey results of tourists' desire to return to visit Girpasang Tourism. The majority of tourists who desired to revisit Girpasang Village were 97.06%. Meanwhile, very few tourists did not want to revisit Girpasang Village, 2.94 %. The high number of tourists who want to revisit the Girpasang Village is a positive sign as it indicates their satisfactory with this tourist destination.

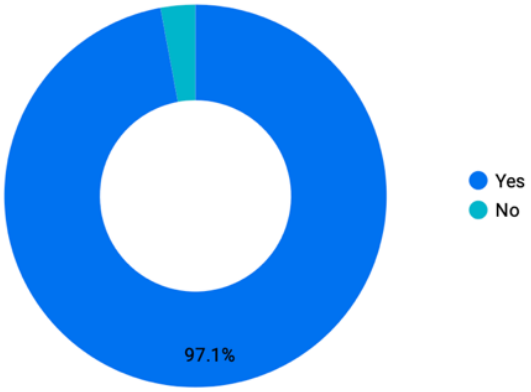

Fig. 6 Pie chart of willingness to return to Girpasang village.

Finally, survey data based on the willingness to recommend Girpasang Tourism to others is visualized in a pie chart on Fig. 7. The tourists were willing to recommend Girpasang Tourism Village to their friends, family, coworkers, or followers on social media. The percentage of those who recommended was 96.08%, and the rest who did not recommend was 3.92%.

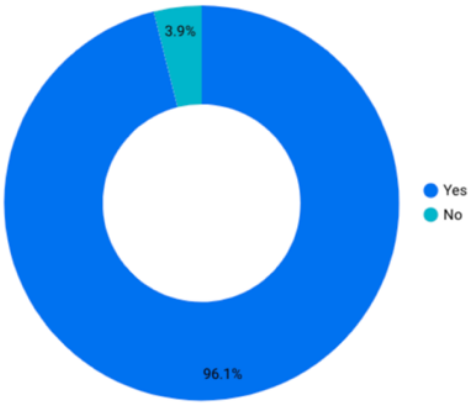

Fig. 7 Pie chart of willingness to recommend the tourist attraction.

3.2. Testing for CFA Assumptions

The results of calculating the MSA value on the survey data, as depicted in Table 2, showed that the variables were sufficient for each formed factor. The overall MSA was equal to 0.92. The partial MSA test also indicated that each variable was feasible to be analyzed using factor analysis.

Table 2. MSA Scores for Each Variable

Variable	MSA	Minimum
X_1	0.93	0.50
X_2	0.93	0.50
X_3	0.9	0.50
X_4	0.92	0.50
X_5	0.95	0.50
X_6	0.94	0.50
X_7	0.93	0.50

Variable	MSA	Minimum
X_8	0.81	0.50
X_9	0.92	0.50
X_{10}	0.85	0.50
X_{11}	0.82	0.50
X_{12}	0.89	0.50
X_{13}	0.96	0.50
X_{14}	0.94	0.50
X_{15}	0.92	0.50
X_{16}	0.89	0.50
X_{17}	0.91	0.50
X_{18}	0.93	0.50
X_{19}	0.90	0.50
X_{20}	0.91	0.50
X_{21}	0.92	0.50
X_{22}	0.96	0.50
X_{23}	0.94	0.50
X_{24}	0.9	0.50
X_{25}	0.92	0.50
X_{26}	0.92	0.50
X_{27}	0.91	0.50
X_{28}	0.94	0.50

Correlation tests on the variables obtained a p-value of less than 2.22×10^{-16}. This suggested that there was some correlation between the 28 variables, and the data was suitable for factor analysis.

3.3. CFA Results

The number of factors in the CFA model was determined by the number of eigenvalues of the partial correlation matrix with values above 1. There are five eigenvalues from the correlation matrix with values above 1, so a total of five factors were formed.

In the process of factor analysis, factor rotation was carried out to simplify the factor structure so that it was easier to interpret. In this analysis, orthogonal rotations were used, specifically the varimax rotation. As provided in Table 3, factor analysis without rotation produced large loadings on the first factor for all variables, but small on the other factors. This result complicated the interpretation because the variables were grouped into the factors that produced the largest loadings. Therefore, the results of analysis with varimax rotation were used.

Table 3. Values of Rotated Factor Loadings

	Initial Factor Loadings					Rotated Factor Loading					
Variable	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**Communality**
X_1	0,74	0.22	0.28	0.01	0.19	0.14	0.50	0.20	0.31	0.55	0.705
X_2	0,77	-0.25	0.25	0.12	-0.19	0.22	0.37	0.19	0.72	0.12	0.762
X_3	0,69	0.29	0.31	0.18	-0.04	0.12	0.69	0.18	0.29	0.31	0.693
X_4	0,71	0.01	0.16	0.37	0.07	0.03	0.48	0.45	0.43	0.21	0.664
X_5	0,72	-0.04	0.15	-0.35	0.03	0.44	0.20	0.01	0.45	0.49	0.667
X_6	0,68	0.14	-0.12	0.00	0.17	0.33	0.28	0.39	0.18	0.40	0.531
X_7	0,67	-0.13	0.10	-0.10	0.09	0.27	0.18	0.19	0.47	0.37	0.499
X_8	0,51	0.23	0.05	-0.20	0.30	0.22	0.20	0.14	0.06	0.57	0.443
X_9	0,64	-0.03	0.06	-0.20	-0.06	0.40	0.22	0.08	0.38	0.30	0.457
X_{10}	0,44	0.01	-0.16	0.19	0.26	0.09	0.11	0.48	0.12	0.25	0.326
X_{11}	0,55	-0.42	-0.14	0.23	-0.06	0.20	0.02	0.46	0.55	-0.05	0.558
X_{12}	0,69	-0.55	0.15	-0.06	0.06	0.20	-0.04	0.22	0.81	0.26	0.807
X_{13}	0,75	-0.26	0.16	0.05	0.00	0.21	0.23	0.26	0.65	0.26	0.656
X_{14}	0,73	-0.02	-0.13	0.17	-0.11	0.37	0.34	0.44	0.37	0.10	0.592

p-ISSN 2798-253X
e-ISSN 2798-3153

	Initial Factor Loadings					Rotated Factor Loading					
Variable	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**Communality**
X_{15}	0,74	-0.01	-0.09	-0.07	0.00	0.44	0.24	0.31	0.35	0.30	0.562
X_{16}	0,57	0.49	0.11	0.14	-0.23	0.29	0.73	0.13	0.03	0.13	0.651
X_{17}	0,73	0.26	0.21	0.01	-0.26	0.37	0.65	0.06	0.32	0.20	0.705
X_{18}	0,76	0.18	-0.17	-0.12	-0.08	0.58	0.35	0.28	0.20	0.29	0.662
X_{19}	0,71	-0.15	-0.34	0.21	0.10	0.35	0.12	0.66	0.34	0.15	0.701
X_{20}	0,72	-0.04	0.06	0.03	0.35	0.15	0.20	0.41	0.36	0.54	0.650
X_{21}	0,73	-0.14	0.13	-0.10	0.13	0.26	0.19	0.22	0.51	0.43	0.598
X_{22}	0,70	0.08	-0.30	0.05	-0.17	0.56	0.30	0.41	0.22	0.07	0.623
X_{23}	0,70	-0.07	-0.09	-0.29	-0.17	0.60	0.17	0.10	0.39	0.24	0.616
X_{24}	0,81	-0.02	-0.28	-0.20	-0.15	0.72	0.20	0.30	0.34	0.22	0.807
X_{25}	0,80	0.10	-0.27	0.11	0.01	0.48	0.34	0.53	0.24	0.23	0.733
X_{26}	0,84	0.05	-0.12	-0.14	0.01	0.54	0.29	0.32	0.35	0.38	0.748
X_{27}	0,61	-0.25	0.15	0.03	-0.22	0.26	0.25	0.12	0.60	0.05	0.509
X_{28}	0,68	0.36	-0.06	0.03	0.08	0.35	0.48	0.33	0.06	0.39	0.609

The standardized loading ℓ_{ij} represents the correlation of the *i*th variable with the *j*th factor. The closer the loading is to 1, the stronger the correlation between the variable and the factor. For example, the variable X12 has a loading of 0.81 on the 4th factor. The value indicates that X12 has a strong correlation with the 4th because the loading is close to 1.

The communality of each variable is presented on the last column in Table 3. This figure illustrates the proportion of variance in the variables that can be explained by the five factors that have been formed. For example, variable X2 has a communality value of 0.762. That is, 76.2% of the variance in variable X2 can be explained by factors 1, 2, 3, 4, and 5.

As shown on Table 4, the 4th factor had the largest sum of squared loadings and proportion of variance accounted for, which was 0.166. These factors could explain 16.6% of the total sample variance. The 5th factor could only explain 10% of the total sample variance. From the sum of the proportions of variance, it was found that the five factors could explain 62.6% of the total sample variance.

Table 4. Variance Explained by Factors

Initial Eigenvalues	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Sum of Squared Loadings	3.800	3.396	2.873	4.657	2.810
Proportion of Variance	0.136	0.121	0.103	0.166	0.100
Cumulative Proportion of Variance	0.302	0.423	0.526	0.166	0.626

Using the varimax rotation, factors that represent a small number of variables were acquired. Each variable was grouped into the factor producing the largest loadings (Table 5).

Table 5. Grouping of Variables under Common Factors

Factor	**Variable**	**Description**
1	X_9	Traffic on the way to Girpasang Village is not congested
	X_{15}	Parking attendants at the location are very helpful
	X_{18}	Staff at the village offers help according to each tourist's needs
	X_{22}	Staff at the village are knowledgable towards matters involving the tourist attraction
	X_{23}	Staff at the. Village are always courteous
	X_{24}	Staff are attentive towards tourists
	X_{26}	Staff are attentive towards the care and maintenance of the suspension bridge
2	X_3	Toilets function properly
	X_4	Prayer room is available
	X_{16}	Staff at the village gives prompt service
	X_{17}	Staff at the village are always willing to help
	X_{28}	There are clear directions on the way to the village.

p-ISSN 2798-253X
e-ISSN 2798-3153

Factor	Variable	Description
3	X_{10}	The journey to Girpasang Village is pleasant
	X_{14}	The price of food at the location matches its value
	X_{19}	Tourist feels safe when using the cable car
	X_{25}	Staff are attentive towards the care and maintenance of the cable car
4	X_{2}	Restaurants or stalls providing food supply are available
	X_{7}	The suspension bridge is in good condition
	X_{11}	It is possible to enjoy the natural scenery from the cable car
	X_{12}	It is possible to enjoy the natural scenery from the suspension bridge
	X_{13}	High quality photos can be taken at the tourist attraction
	X_{21}	Tourist feels safe at Girpasang Village
	X_{27}	Small, local businesses at the village are honest in their trade
5	X_{1}	Spacious parking area is available
	X_{5}	The entire area of Girpasang Village is clean
	X_{6}	The cable car is in good condition
	X_{8}	The road surface on the way to Girpasang Village is smooth
	X_{20}	Tourist feels safe when crossing the suspension bridge

Each factor is given the appropriate name based on the results of the grouping of variables. Factor 1 is the attitude of the staff, factor 2 is the reliability of the tourist spot, factor 3 is the convenience of the tourist spot, factor 4 is the available facilities, and factor 5 is the tangible condition of facilities.

The factor analysis model obtained a Root Mean Square of the Residuals (RMSR) value of 0.04, where the maximum acceptable RMSR limit was 0.08. The RMSEA value was 0.066, indicating that the covariance matrix estimated by the model was quite close to the population covariance matrix. The TLI of factoring reliability was 0.899, indicating that the assumed model did not match the baseline model.

3.4. Text Analysis

The questionnaire allowed the respondents to express criticism and suggestions about their satisfaction with Girpasang Village. Statements of criticism and suggestions were more precisely analyzed using text analysis. The text analysis method used in this study aimed to understand how respondents described and expressed their perceptions of the conditions of Girpasang Village either positively, negatively, or neutrally. This analysis was essential to see how far this survey was as one of the instruments for disseminating information on the satisfaction of tourists who come to Girpasang Village. This research aimed to identify which words were most mentioned in the criticisms and suggestions of tourists. Fig. 8 shows the results of text analysis, obtaining 75 common words.

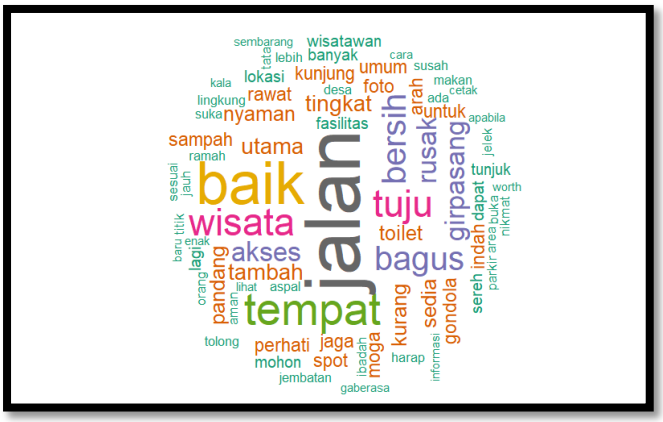

Fig. 8 Word cloud of common words from respondents' criticism and suggestions.

The words most frequently mentioned from the results of criticism and suggestions include "jalan" (road), "baik" (good), "bersih" (clean), "sampah" (garbage), "fasilitas" (facility), and other

words in Fig. 8. Next, an analysis was carried out for the words that repeatedly appeared based on the words “jalan” (street) and “tempat” (place). Table 6 displays the results of the words related to the two words and the correlation value of the association.

Based on the correlation coefficients provided in Table 6 and Table 7, the word “jalan” (road) had a close association with the word “tuju” (heading) with a correlation value of 0.5, while the association with the word “rusak” (broken) was 0.42. Then, the association with the word “baik” (good) was 0.39, “aspal” (asphalt) was 0.3, and “licin” (slippery) was 0.31, as well as the association with other words. Meanwhile, the word “tempat” (place) has an association relationship with the words “ibadah” (worship), “restoran” (restaurant), “sampah” (garbage), “bersih” (clean), “nyaman” (comfortable), and “parkir” (parking), which relationship value was 0.45, 0.44, 0.39, 0.37, 0.32, and 0.19, respectively. These results demonstrated that the word “tempt” (place) had a close relationship with these words; therefore, it could be concluded that the availability of prayer rooms was insufficient for many tourists to Girpasang Village. Despite its availability, a considerable number of tourists lacked precise location information, thus advocating for an improved navigation system within the village.

Table 6. Correlation of Words Associated with “Tempat” (Place)

“Tempat” (Place)					
Word	**Correlation**	**Word**	**Correlation**	**Word**	**Correlation**
worship	0.45	access	0.2	sometimes	0.19
far	0.45	tour	0.2	land	0.19
sit	0.44	if	0.19	slippery	0.19
chair	0.44	new	0.19	see	0.19
table	0.44	waste	0.19	easy	0.19
restaurant	0.44	price	0.19	non	0.19
rubbish	0.39	information	0.19	parking	0.19
clean	0.37	want	0.19	I	0.19
visit	0.32	reach	0.19	any	0.19
comfortable	0.32	clear	0.19	pull	0.19
friendly	0.27	sell	0.19	of course	0.19
in accordance	0.27	connection	0.19	participate	0.19
system	0.27	return	0.19	inhabitant	0.19
toilet	0.21	hit	0.19	not enough	0.17

Table 7. Correlation of Words Associated with “Jalan” (Road)

“Jalan” (Road)			
Word	**Correlation**	**Word**	**Correlation**
tuju	0.5	tunjuk	0.31
rusak	0.42	aspal	0.3
baik	0.39	titik	0.3
akses	0.33	tolong	0.3
girpasang	0.32	tingkat	0.28
jalur	0.31	utama	0.28
kampung	0.31	wisata	0.23
kadang	0.31	arah	0.22
licin	0.31	wisatawan	0.19
muka	0.31	desa	0.16
semua	0.31	jauh	0.16
soal	0.31	suka	0.16
tahu	0.31	tata	0.16

p-ISSN 2798-253X
e-ISSN 2798-3153

4. Conclusion

Based on the factor analysis results, it can be concluded that five general factors influence the satisfaction of tourists visiting Girpasang Village. These factors are the attitude of the staff, the reliability of tourist attractions, the convenience of tourist attractions, the facilities available and the real conditions of the tourist environment. Meanwhile, based on text analysis, information was obtained from all the criticisms and suggestions collected from the survey results. It can be concluded that on the main road, damaged roads, slippery roads, places of worship, restaurants, trash found in the environment around the Girpasang tour, and parking lots could explain the contents of the criticisms and suggestions tourists gave. These results can be a consideration for Girpasang tourism to improve the satisfaction of tourists.

References

[1] UNWTO, Tourism and Rural Development: Understanding Challenges on the Ground – Lessons learned from the Best Tourism Villages by UNWTO Initiative. World Tourism Organization (UNWTO), 2023, doi: 10.18111/9789284424368.

[2] I.G.N.W. Arsana, I.N. Sunarta, and I.N.S. Arida, "Fenomena Pengelolaan Atraksi Wisata Swafoto dan Implikasinya bagi Masyarakat di Desa Wanagiri Kecamatan Sukasada Kabupaten Buleleng," *Jurnal Master Pariwisata (JUMPA)*, Vol. 5, No. 2, pp. 330–347, Jan. 2019, doi: 10.24843/JUMPA.2018.v05.i02.p07.

[3] A. Parasuraman, L.L. Berry, and V.A. Zeithaml, "Refinement and Reassessment of the SERVQUAL Scale," *Journal of Retailing*, Vol. 67, No. 4, pp. 420–450, 1991. [Online]. Available: https://www.proquest.com/scholarly-journals/refinement-reassessment-servqual-scale/docview/228675435/se-2?accountid=62100

[4] T.T. Borishade, O.O. Ogunnaike, O. Salau, B.D. Motilewa, and J.I. Dirisu, "Assessing the Relationship among Service Quality, Student Satisfaction and Loyalty: the Nigerian Higher Education Experience," *Heliyon*, Vol. 7, No. 7, pp. 1–8, Jul. 2021, doi: 10.1016/j.heliyon.2021.e07590.

[5] M. Saut and V. Song, "Influences of Airport Service Quality, Satisfaction, and Image on Behavioral Intention towards Destination Visit," *Urban, Planning and Transport Research*, Vol. 10, No. 1, pp. 82–109, Dec. 2022, doi: 10.1080/21650020.2022.2054857.

[6] G. Teng, J. Liu, J. Xiao, X. Ma, and Y. He, "Comparative Analysis and Strategy Research of Enterprises Dealing with Network Public Opinion Based on Text Mining," *Procedia Computer Science*, Vol. 214, pp. 801–808, 2022, doi: 10.1016/j.procs.2022.11.244.

[7] A. Fauzy, *Metode Sampling*, 2nd ed. Tangerang, Indonesia: Penerbit Universitas Terbuka, 2019.

[8] I. Kandasamy, W.B.V. Kandasamy, J.M. Obbineni, and F. Smarandache, "Indeterminate Likert Scale: Feedback based on Neutrosophy, Its Distance Measures and Clustering Algorithm," *Soft Computing*, Vol. 24, No. 10, pp. 7459–7468, May 2020, doi: 10.1007/s00500-019-04372-x.

[9] P. Leavy, *Research Design*, 2nd ed. New York, USA: The Guilford Press, 2017.

[10] M. Muhajir*, Modul Praktikum Statistika Multivariat Terapan*. Yogyakarta, Indonesia: Universitas Islam Indonesia, 2021.

[11] N. Shrestha, "Factor Analysis as a Tool for Survey Analysis," *American Journal of Applied Mathematics and Statistics*, Vol. 9, No. 1, pp. 4–11, Jan. 2021, doi: 10.12691/ajams-9-1-2.

[12] D.W. Wichern and R.A. Johnson, *Applied Multivariate Statistical Analysis*, 6th ed. New Jersey, USA: Pearson Prentice Hall, 2007.

[13] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations," *Organizational Research Methods*, Vol. 25, No. 1, pp. 114–146, Jan. 2022, doi: 10.1177/1094428120971683.

[14] H. Jung and B.G. Lee, "Research Trends in Text Mining: Semantic Network and Main Path Analysis of Selected Journals," *Expert Systems with Applications*, Vol. 162, pp. 1–12, Dec. 2020, doi: 10.1016/j.eswa.2020.113851.

[15] C.-Y. Chu, K. Park, and G.E. Kremer, "A Global Supply Chain Risk Management Framework: An Application of Text-Mining to Identify Region-Specific Supply Chain Risks," *Advanced Engineering Informatics*, Vol. 45, pp. 1–17, Aug. 2020, doi: 10.1016/j.aei.2020.101053.

p-ISSN 2798-253X
e-ISSN 2798-3153