Spektrum Industri
Vol. 23, No. 1, 2025, pp. 16-30
ISSN 1693-6590
http://journal3.uad.ac.id/index.php/spektrum

Analysis of Visitor Perceptions of Malang City Thematic Parks using a Text Mining Approach

Nur Izza Vania [a], Aisyah Larasati [a,*], Yuh-Wen Chen [b], Nikmatus Sholihah [a]

[a] Department of Mechanical and Industrial Engineering, Universitas Negeri Malang, Malang, 65145, Indonesia
[b] Department of Information Management, Da-Yeh University, Chang Hwa, 515006, Taiwan
* Corresponding Author: aisyah.larasati.ft@um.ac.id

ARTICLE INFO

Article history
Received August 23, 2024
Revised March 12, 2025
Accepted March 17, 2025

Keywords
Text mining;
Place branding;
Text summarization;
Thematic parks

ABSTRACT

Ruang Terbuka Hijau (RTH) play a crucial role in urban environments, not only supporting nature conservation but also fostering social interaction and contributing to economic growth. City parks, as representations of GOS, can be enhanced through thematic development and place branding to improve user engagement and functional value. This study aims to analyze visitor perceptions of thematic parks by identifying high-frequency keywords extracted from user-generated reviews. Text mining techniques were employed using Term Frequency-Inverse Document Frequency (TF-IDF) and Term Frequency-Relevance Frequency (TF-RF) methods, followed by text summarization using Cosine Similarity and Maximum Marginal Relevance (MMR). These methods effectively process large volumes of unstructured data to reveal meaningful insights. The analysis focused on three parks with the highest number of reviews: Alun-Alun Kota Malang (945 reviews), Taman Merjosari (552 reviews), and Alun-Alun Tugu (462 reviews). Keyword analysis showed prominent terms such as 'tugu', 'olahraga' (sports), and 'anak' (children) under both TF-IDF and TF-RF methods, with TF-RF emphasizing more context-specific vocabulary. Results indicate that Alun-Alun Tugu is perceived as a comfortable space near government offices featuring a lotus pond, Taman Merjosari is recognized for its sports facilities, and Alun-Alun Malang is identified as a child-friendly park with fountains. The study offers place branding recommendations by analyzing word associations in summarized user feedback. This study can contribute valuable insights for governments, architects, and urban designers in developing thematic parks that better reflect and accommodate user preferences.

1. Introduction

According to Law No. 26 of 2007, Ruang Terbuka Hijau (RTH) is a vital component of urban areas that serves multiple ecological and social functions. It plays a significant role in nature conservation, rainwater infiltration, pollution mitigation, ecosystem stability, provision of wildlife habitats, recreational opportunities, water purification, and the reduction of urban heat and noise (Ayu, 2019). RTH improves quality of life (Fikriyah et al., 2021), and includes urban parks, with thematic parks specifically designed for visitors based on location, size, and architecture (Levochkina, 2017). These parks also support SDG 15 "Life on Land" and enhance city image (Rista & Siregar, 2019). Place branding promotes competitive advantages through investment in tourism, regional economies, and city image, aligning with market expectations by providing relevant information and experiences (Agustian & Razali, 2023; Mihardja et al., 2020).

Branding through thematic parks has been successful in various Indonesian regions. For instance, Bandung's Superhero Park and Film Park have expanded RTH and boosted public interaction. Tangerang City's thematic parks increased daily visitors from 100 to 750 in 2023 (Irfan, 2023). Munawir et al., (2019) studied visitor perception and effectiveness of place branding strategies in thematic parks in Bandung City using text mining methods, including TF-IDF and text summarization, based on google maps user reviews. Their study demonstrates that these methods can be used to determine the extent to which a park can be a benchmark for place branding activities in Bandung City. Sari, Baskara, Prakoso, & Royani (2022) and Larasati et al., (2023) demonstrated that the TF-IDF algorithm effectively identifies the most important words used to classify novel genres and assess brand awareness in museums.

However, a problem arises when a term j_a appears in document 1 but not in document 2, yet both produce the same IDF value. To address this issue, TF-RF was introduced as an improved version of TF-IDF. The RF factor helps distinguish values across documents, with its development focusing on enhancing the discriminatory influence on (a) the number of documents containing a term and (d) the number of documents not containing a term. Malang City, recognized as Indonesia's most livable city with a score of 63.5 by the Most Livable City Index (MLCI) (Nugroho et al., 2022), features various RTH, including thematic parks like Trunojoyo Park and Merbabu Park. Evaluating Malang's thematic parks for their impact on city branding involves assessing visitor experiences and repeat visits (Juwanda & Widiastuti, 2023). Online reviews play a crucial role in this assessment, as sharing travel experiences online is enjoyable and helps map consumer behavior (Munawir et al., 2019; Oliveira et al., 2020) One of the indicators of MLCI's assessment is the availability of public and social facilities that facilitate community interaction, one of which is RTH in the form of city parks.

This research indicates that text mining methods have not yet been applied to evaluate thematic parks or assess the extent to which these parks are recognized and contribute to place branding in Malang City. In this context, text mining techniques using the TF-IDF word weighting method compared with TF-RF results, can be utilized as an analytical tool to measure the effectiveness of place branding strategies associated with thematic parks in Malang City.

One platform for writing reviews of a place online is Google Maps, a free mapping service from Google that allows users to provide comments freely. The research will utilize Google Maps to gather data on visitor review regarding thematic parks in Malang City. Data analysis will employ text mining methods to evaluate the effectiveness of thematic parks in building place branding and understanding public perceptions. Online reviews are considered effective as they avoid psychological biases.

Text mining methods are used for efficient analysis of large data sets and extraction of information from unstructured data. In the term weighting stage of text mining, word weighting methods are necessary to determine the importance level of words in the dataset. One commonly used method is Term Frequency - Inverse Document Frequency (TF-IDF), but it has a drawback as it doesn't consider the relevance level and documents that don't contain a term.

To address this, a word weighting method called Term Frequency - Relevance Frequency (TF-RF) will be combined, which considers the ratio of documents that contain a particular word and those that do not (Assidyk et al., 2020). The findings of this research can be used by the government, architects, and urban designers to design thematic parks that better align with user preferences. This research aims to understand visitors' perceptions of parks through the text mining by determining words with the highest weights based on reviews or feedback from previous park users.

2. Method

A successful branding can be assessed by how the location represents its identity and how people feel, understand, use, and connect with the location through visitor experiences. This computational text mining method is more effective when applied to large datasets (Thakur & Kumar, 2022).

Therefore, in this study, park data will be selected from datasets categorized as large, specifically those containing more than 500 documents (Hickman et al., 2022). The method used in this research is illustrated in Fig. 1. The study collected visitor experiences at thematic parks in Malang City using web scraping from Google Maps, including user information, reviews, ratings, and dates. After collecting the data, it underwent pre-processing and word weighting with TF-IDF and TF-RF methods to identify important words, followed by text summarization. The aim was to evaluate the parks' effectiveness as a place branding tool for Malang City, which aims to be a "Green City" with 19.4% green open spaces (Wihardjo, 2020).

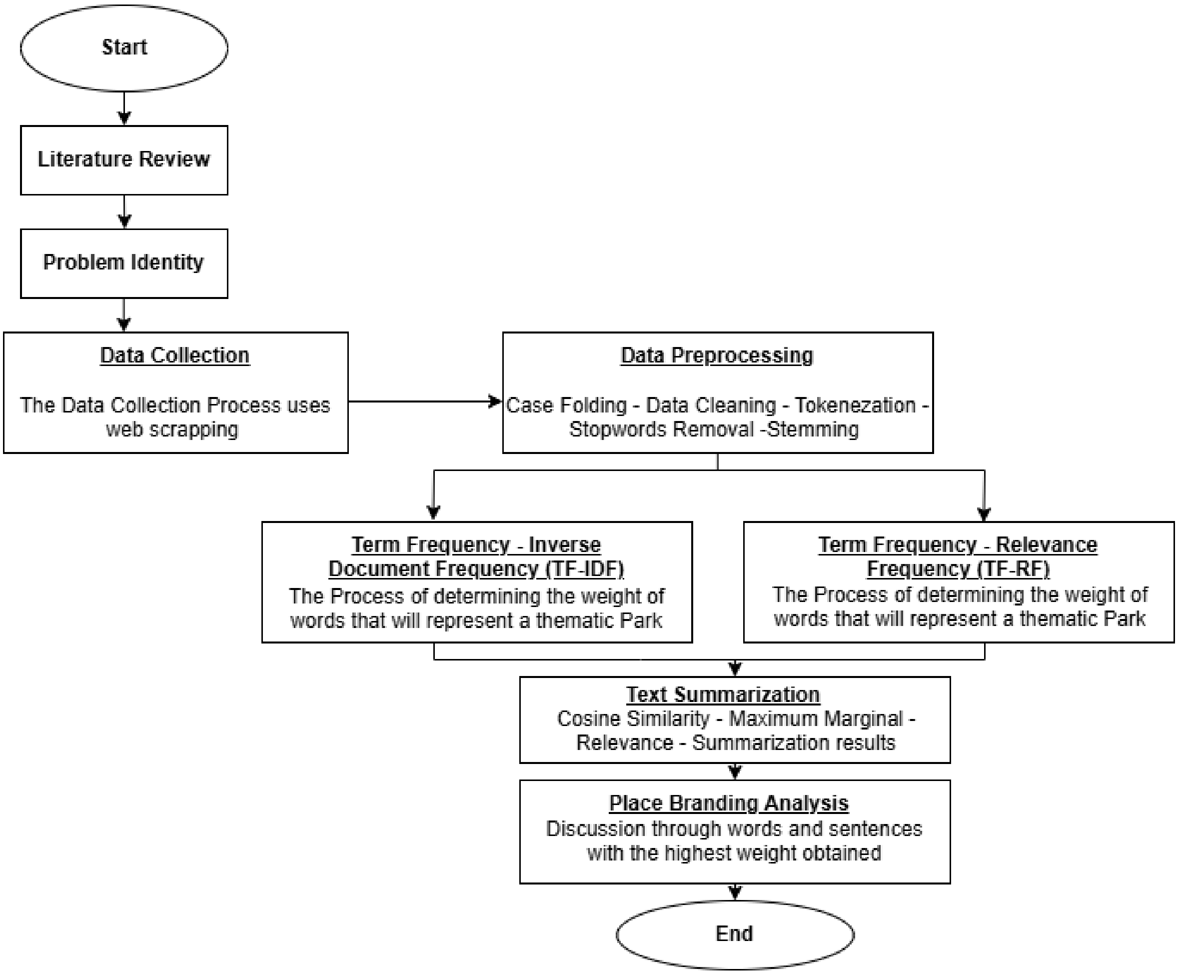


Fig. 1. Research Flow Chart

Actual data show in Table 1, Malang City has 16 thematic parks and 68 non-thematic parks, and the research focuses on using visitor reviews to enhance place branding and attract more visitors.

Table 1. Thematic Parks in Malang City

No.	Nama Taman Tematik	Alamat
1	Taman Alun-Alun Kota Malang	Jl. Merdeka Kelurahan Kidul Dalem
2	Taman Chairil Anwar	Jl Basuki Rahmat Kelurahan Kidul Dalem
3	Taman Alun-Alun Tugu	Jl. Tugu Kelurahan Klojen
4	Taman Kertanegara	Jl. Kertanegara Kelurahan Klojen
5	Taman Trunojoyo	Jl. Trunojoyo Kelurahan Klojen
6	Taman Dr. Sutomo	Jl. Dr. Sutomo Kelurahan Klojen
7	Taman Jl. Ijen	Jl. Ijen Kelurahan Oro-Oro Dowo
8	Taman Jl. Merbabu/Family Park	Jl. Merbabu Kelurahan Oro-Oro Dowo
9	Taman Slamet	Jl. Taman Slamet Kelurahan Gading Kasri
10	Taman Dempo	Jl. Dempo Kelurahan Oro-Oro Dowo

No.	Nama Taman Tematik	Alamat
11	Taman Cerme	Jl. Cerme Kelurahan Oro-Oro Dowo
12	Taman Merjosari	Jl. Mertojoyo Keluarahan Merjosari
13	Taman Mojolangu	Jl. Candi Mendut Kelurahan Mojolangu
14	Taman Kendedes	Jl. Jendral A. Yani Kelurahan Balearjosari
15	Taman Pandanwangi	Jl. LA Sucipto Kelurahan Pandanwangi
16	Taman Bundaran Halmahera	Jl. Halmahera Kelurahan Kasin

The research uses primary data from Google Maps collected over the 2021-2023 period, chosen due to decreased park visitors during the pandemic. Data collection began in 2021 after Malang City's PPKM regulations were eased to level 2, allowing limited access to public spaces. The data includes park locations, user information, reviews, and timestamps, extracted using Python libraries in Jupyter Notebook and stored in a dataset. Google Maps was selected for its easy API access and detailed, structured review data, which aids in comprehensive data extraction and analysis (Nurahaliza & Mulyadi, 2022). Reviews were gathered based on park names. Reviews from thematic and non-thematic parks are gathered to compare the popularity of each park among visitors and the effectiveness of branding strategies. The flowchart of text mining usage is depicted in Fig. 2, starting with the extraction of review data from Google Maps, then through the text pre-processing process: case folding, cleaning, tokenization, stop words removal, and stemming. Afterward, words with the highest weights are identified, and text summarization is conducted to recommend place branding strategies for thematic parks. This text mining computation method will work more effectively on large amounts of data (Thakur & Kumar, 2022). Therefore, in this study, the park data to be used are documents categorized as large, specifically datasets with more than 500 documents (Hickman et al., 2022).

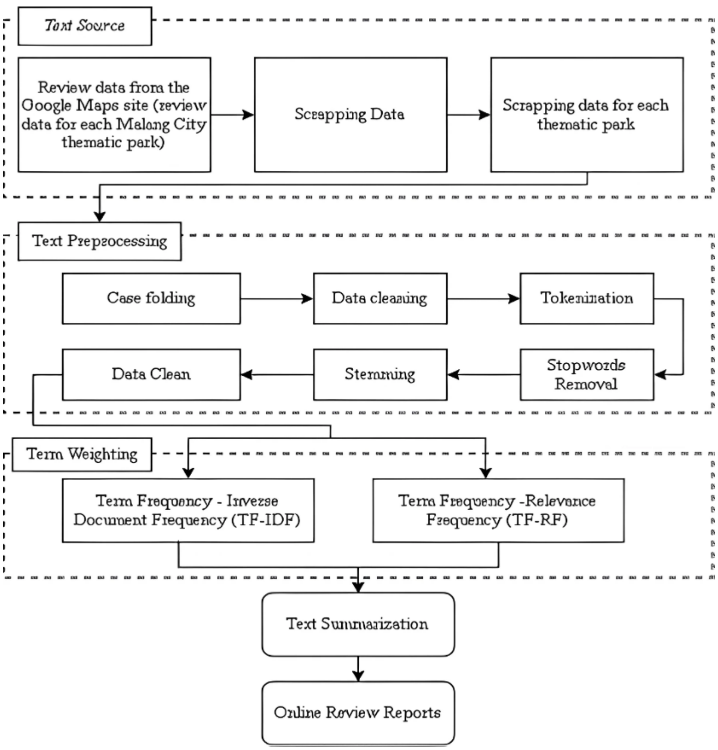


Fig. 2. Text Mining Flow Chart. Source: Adaptation from Ahmad Gozali & Alfan Rosid, 2020

2.1. Scrapping Data

The first process of text mining is scrapping user reviews from Google Maps using the Python Selenium library. The stages of this process are: Library preparation; Loading the thematic park URL; Data extraction from HTML code; and Data will be collected in .xlsx format for the next step.

2.2. Text Pre-processing

The The first stage is case folding, converting uppercase letters to lowercase for consistency. Next is data cleaning, which removes symbols, punctuation, and numbers, leaving only alphabetic characters (Kasanah et al., 2019). The following stage is tokenization, where words are separated by spaces and called tokens. Then, stopwords removal filters out common words like articles, conjunctions, and pronouns. The final step is stemming, which reduces words to their basic forms, simplifying unstructured text into structured data by retaining only the root terms.

2.3. Term Frequency - Inverse Document Frequency (TF-IDF)

After data cleaning and word preparation, the TF-IDF process was carried out to determine word weights. In Eq. (1), TF shows the word occurrence value, while IDF gives the relative weight for each word. The TF-IDF calculation formula is:

$$\textbf{TF-IDF} = \frac{n_k}{n} \text{ x } log \left(\frac{N}{N_k}\right) \quad (1)$$

Source : Kotu & Deshpande, 2019

Information:

TF : Term Frequency
IDF: Inverse Document Frequency
nk : number of keywords that appear in the document (k is the keyword)
n : total terms in document
N : total document
Nk : total dokumen with keyword k, k for word

2.4. Term Frequency - Inverse Document Frequency (TF-IDF)

Word weighting is also carried out using the TF-RF method, a development of TF-IDF, which takes into account the importance of terms based on their frequency in document categories. The RF factor increases the discrimination effect on the values of a (the number of documents containing a term) and d (the number of documents that do not contain a term). The TF-RF value is the product of TF and RF Eq. (2).

$$\textbf{RF} = \log\left(2 + \frac{a}{d}\right) \quad (2)$$

Source : Luthfi & Lhaksamana, 2020

In Eq. (3), the value of TF-RF is obtained from multiplying the Term Frequency (TF) and RF (Relevance Frequency) values, which can be seen in the formula below:

$$\textbf{TF - RF} = \text{TF x RF} \quad (3)$$

Source : Luthfi & Lhaksamana, 2020

Information:

TF : Term Frequency
RF : Relevance Frequency
a : total document with terms k, k for word
d : total document without terms k, k for word

2.5. Equations

Text summarization is used to create concise summaries of large texts (Manjari et al., 2020). This research employs an extractive text summarization approach using the Maximum Marginal Relevance (MMR) method, which measures text similarity to a query and reduces redundancy. The cosine similarity method is used for its stability in comparing documents of different lengths. The lambda value of 0.7 is applied in the MMR method, as it has shown high precision and good performance in previous studies. The text summarization process begins with calculating cosine similarity. The results are then used to calculate the MMR score, and the text summarization is obtained based on the highest score of each sentence.

2.5.1. Count the Cosine Similarity

The cosine similarity value is calculated to determine the similarity between documents and queries obtained from term weighting, as well as to assess the similarity between documents. In mathematics Eq. (4), cosine similarity is calculated using the following formula:

$$CosSim(dj, q) = \sum_{i=1}^{t}\left(W_{i,j}norm \times W_{i,q}norm\right) \quad (4)$$

Source : Saraswati et al., 2018

$CosSim(dj,q)$	= cosine similarity between doc dj and query q
$\sum t,i=1$	= sum of all words (t) and dimensions (i) of the vector
Wi,jnorm	= TF-IDF normalized weight for word i in document dj
Wi,qnorm	= TF-IDF normalized weight for word i in query q

2.5.2. Count the MMR Score

MMR performs document summarization by measuring the similarity between text segments, aiming to obtain sentence scores based on similarity to the query while reducing redundancy. In this research, a lambda value of 0.7 is used. In Eq. (5), The lambda value (λ) ranges $0 < \lambda < 1$. If $\lambda = 1$, the MMR score will be more relevant to the original document, while if $\lambda = 0$, the score will be more relevant to previously extracted sentences.

$$\text{MMR} = \text{argmax}\ [\lambda * \ \text{Sim1}(\text{Si}, \text{Q}) - (1 - \lambda) * \max Sim2(Si, S')] \quad (5)$$

Source : Saraswati et al., 2018

MMR	= Maximum Marginal Relevance
λ	= parameters that influence the level of relevance
Si	= weight vector of candidate words
S’	= weight vector of words other than candidate
Q	= word weight vector of the query
Sim1 (Si,Q)	= similarity value between the ith sentence and the query
Sim2 (Si,S’)	= similarity value between the ith sentence and the extracted sentence

2.5.3. Summarization result

Summary results are derived from documents with the highest weights and no redundancy, based on the MMR score. These documents were analyzed to form branding recommendations from park visitor reviews using a two-word weighting method. Word relationships are visualized with a force-directed network graph (collocation graph), and the most important words are shown with a Word Cloud.

3. Results and Discussion

This research is a study of visitor perceptions and place branding strategies using reviews from Google Maps, involving 16 thematic parks with a total of 3204 reviews. Data was collected through scrapping using a web scrapper that applies machine learning algorithms in Jupyter Notebook. The distribution of reviews for each park in the last 2 years is shown in the graph in Fig. 3.

The latest review data (2021-2023) shows three parks with the most reviews: Taman Alun-Alun Kota Malang (945 reviews), Taman Merjosari (552 reviews), and Taman Alun-Alun Tugu (462 reviews). Each review can consist of more than one sentence per review column. The text mining method is more effective on large data sets (Thakur & Kumar, 2022), so this study uses datasets with more than 500 documents (Hickman et al., 2022).

From Fig. 3, Fig. 4, it can be concluded that the concept of developing a thematic park influences the number of reviews received by the park. The more reviews received, the higher the number of visits to the thematic park. This is supported by research conducted by (Rista & Siregar, 2019), which states that thematic parks do provide more attraction to visitors than non-thematic parks.

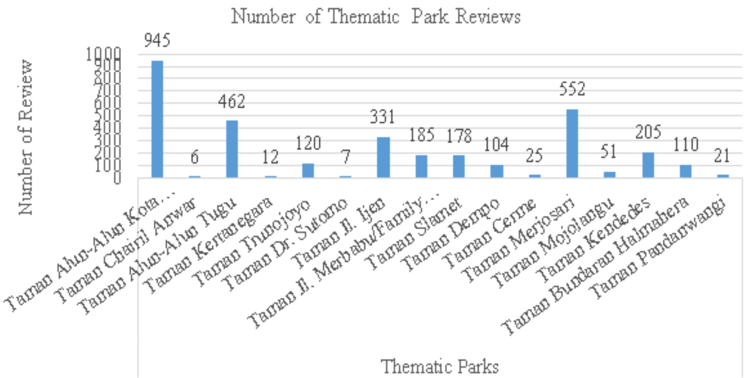


Fig. 3. Number of Reviews for Each Thematic Park in Malang City

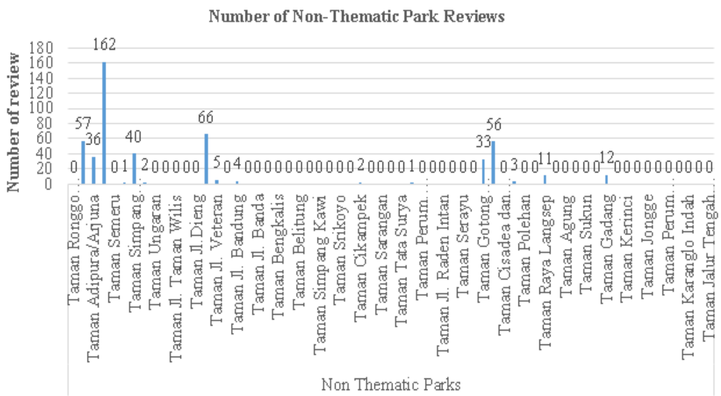


Fig. 4. Number of Reviews of Non-Thematic Parks in Malang City

3.1. Identify the Headings

The pre-processing stage was conducted using Jupyter Notebook software and the Python 3.0 programming language. The results of this pre-processing stage are shown in the following Table 2:

Table 2. Result from Text Pre-Processing

Pre-processing	Input	Output
Cleaning Data	I love this city.. ❤️ ❤️ 🤗 Selalu suka pemandangan tugu malang kapanpun..	I love this city.. Selalu suka pemandangan tugu malang kapanpun..
Case Folding	I love this city.. Selalu suka pemandangan tugu malang kapanpun..	i love this city.. selalu suka pemandangan tugu malang kapanpun..
Tokenization	i love this city.. selalu suka pemandangan tugu malang kapanpun..	'i', 'love', 'this', 'city' , 'selalu', 'suka', 'pemandangan', 'tugu', 'malang', 'kapanpun'
Stopwords Removal	'i', 'love', 'this', 'city' , 'selalu', 'suka', 'pemandangan', 'tugu', 'malang', 'kapanpun'	'love', 'city', 'suka', 'pemandangan', 'tugu', 'malang',
Stemming	'love', 'city', 'suka', 'pemandangan', 'tugu', 'malang',	'love city', 'suka pandang tugu malang'

3.2. Figures and Tables

The first word weighting process used is TF-IDF. TF refers to a value that indicates that the more often a word is found in a document, the higher its weight, and the word is considered more significant in the context of the document. DF describes how often the word appears in a single document. IDF gives higher weight to words that are uncommon and less frequently occurring in the entire dataset document. The TF-IDF weighting process in this research utilizes the python, pandas and math libraries. Below in Table 3, 5 words with the highest weight for each park studied from the TF-IDF weighting.

Table 3. Results of TF-IDF Thematic Parks in Malang City

Taman Tugu		**Merjosari Park**		**Alun Malang**	
Word	***TF-IDF***	***Word***	***TF-IDF***	***Word***	***TF-IDF***
bagus	1	taman	1	anak	1
malang	0.759931263	bersih	0.751359096	alun	0.965458419
nyaman	0.704057528	olahraga	0.742085513	main	0.877133348
kota	0.689434106	bagus	0.730488442	bersih	0.855362151
indah	0.597631291	nyaman	0.644672834	malang	0.774782638

Fig. 5. shows the Word Cloud visualization of the highest-weighted words in the thematic park corpus. In the Word Cloud, larger font sizes indicate higher weights and greater importance of the words, as detailed in Table 3.

(a) (b)

(c)

Fig. 5. TF-IDF Word Cloud Visualization for Taman Tugu (a); Merjosari park (b); and Alun-Alun Malang (c)

3.3. Term Frequency-Relevance Frequency (TF-RF)

The RF factor enhances the discriminative power by differentiating between the number of documents containing a term (a) and those not containing it (d). The results, showing the top 5 words with the highest weights for each park from the TF-RF weighting process, are displayed in Table 4.

Table 4. Results of TF-RF Thematic Parks in Malang City

Taman Tugu		**Merjosari Park**		**Alun-Alun Malang**	
Word	***TF-RF***	***Word***	***TF-RF***	***Word***	***Bobot TF-RF***
Malang	1	taman	1	anak	1
Kota	0.800659331	olahraga	0.406358461	main	0.861255617
Bagus	0.704503039	bagus	0.340239989	alun	0.853847273
Tugu	0.527279805	bersih	0.322559902	malang	0.782648786
Bersih	0.383543397	main	0.251952273	bersih	0.716572556

Word Cloud visualization for the thematic park corpus based on its weight in the TF-RF calculation, is shown in Fig. 6, as follows:

Spektrum Industri
Vol. 23, No. 1, 2025, pp. 16-30
ISSN 1693-6590

(a) (b)
(c)

Fig. 6. Word Cloud TF-RF visualization for Taman Tugu (a); Merjosari Park (b); and Alun-Alun Malang (c)

3.4. Text Summarization

The text summarization process uses the MMR method to identify high-weight sentences, ensuring minimal redundancy by considering their similarity to the query and other documents. These sentences, along with park scores calculated using a Python library, are then converted into concise, meaningful sentences for place branding by analyzing word relationships. Table 5, Table 6, shows the documents with the highest MMR values.

Table 5. Top 5 Documents with the Highest MMR Values TF-IDF

Taman Tugu		Taman Merjo		Alun-Alun Malang	
Document 219	0.17437	Document 73	0.133215	Document 866	0.153098
Document 759	0.166961	Document 1161	0.127084	Document 1745	0.138351
Document 582	0.153357	Document 228	0.121684	Document 2031	0.130783
Document 629	0.152727	Document 1152	0.109634	Document 1596	0.124988
Document 714	0.145153	Document 1109	0.108465	Document 179	0.119575

Table 6. Documents with the Highest MMR Values TF-RF

Taman Tugu		Taman Merjo		Alun-Alun Malang	
Document 786	0.190015	Document 1012	0.253904	Document 348	0.221191
Document 859	0.184061	Document 484	0.153069	Document 475	0.167991
Document 143	0.177343	Document 450	0.083943	Document 1000	0.133556
Document 184	0.165526	Document 315	0.07667	Document 2012	0.125621
Document 919	0.140149	Document 141	0.076291	Document 45	0.119958

3.5. Overview of Reviews and Distribution of Thematic Park Ratings

The study gathered recent reviews from 16 theme parks in Malang City on Google Maps from 2021 to 2023 to reflect current conditions. The top three parks with the most reviews are Alun-Alun Kota Malang Park (945 reviews), Merjosari Park (552 reviews), and Alun-Alun Tugu Park (462 reviews). These reviews were analyzed to provide place branding recommendations based on visitor feedback. Place branding involved creating high-weight sentences using two weighting methods: TF-IDF and TF-RF. Python in Jupyter Notebook was used to automate this time-consuming process.

The Maximum Marginal Relevance (MMR) method evaluated sentence relevance by measuring text similarity to reduce redundancy (Kurniawan & Humaidy, 2022). For Tugu Park, the highest-weighted words with TF-IDF were 'bagus, 'Malang', and 'nyaman' while with TF-RF, they were 'Malang' 'kota', and 'bagus'. The generated weights served as queries for document similarity using cosine similarity, which ranked documents for MMR summarization. The summarized sentences with the highest weights and relevance were used for place branding development. The research found 76 documents from TF-IDF and 96 from TF-RF for Tugu Park, 118 documents each from both methods for Merjosari Park, and 235 documents each from both methods for Alun-Alun Kota Malang Park.

3.6. Analysis of Term Weighting Result

The TF-IDF method has a drawback in that it doesn't account for the proportion of documents containing specific keywords versus those that don't, potentially reducing accuracy. For example, a word appearing in one document but not another might receive the same IDF value, skewing its importance. To address this, the TF-RF method is used, which considers the ratio of documents with specific words to those without, providing a more nuanced weighting (Assidyk et al., 2020).

This difference affects word weights; for instance, the word 'nyaman' in Alun-Alun Tugu has a weight of 0.704057528 with TF-IDF and 0.414231012 with TF-RF. Similar cases are found with other terms in Merjosari Park, such as 'anak' and 'olahraga', as seen in Table 3, respectively, and in Alun-alun Malang for terms like 'main' and 'bersih', as shown in Table 4.

Similar results are shown in the study by Sari et al., (2022), where the term 'buku' had a weight of 0.4 using TF-IDF and 0.6 using TF-RF. These differences stem from the weighting methods, impacting book classification and recommendations based on user keywords. Assidyk et al., (2020) also compared TF-IDF and TF-RF for classifying trending topics on Twitter, finding TF-IDF performed better for random topic searches. In this research, both methods are used to weight words for place branding from theme park reviews.

The study by Sari et al., (2022) found that TF-RF term weights were higher than TF-IDF, leading to fewer but more precise book recommendations. Similarly, Assidyk et al. (2020) found TF-IDF effective for predicting Twitter trends. This research uses TF-RF to address TF-IDF's limitation of only considering word occurrence, also factoring in documents without the terms (Harmandini & L, 2024; Sari et al., 2022). The recommended sentences for summarization in the MMR method are thus more relevant and non-redundant. The average MMR scores are shown in Table 7.

Table 7. MMR Values

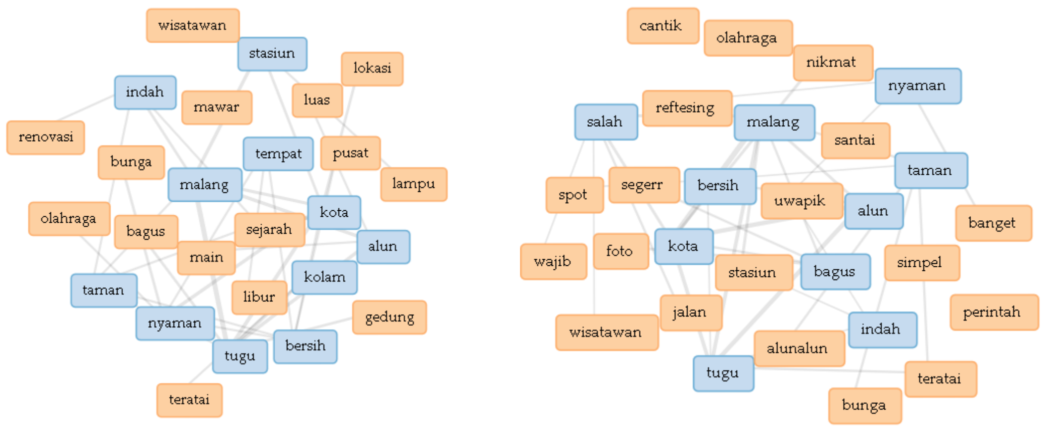
	TF-IDF	TF-RF
Taman Tugu	0.0748166	0.0350164
Taman Merjo	0.0647237	0.0392781
Alun-Alun Malang	0.0522542	0.0308379

In this study, TF-IDF MMR scores are consistently higher than TF-RF MMR scores for each theme park dataset. This is because TF-IDF assigns higher weights to frequently appearing words due to its focus on corpus variety. In contrast, TF-RF considers word relevance and occurrence within specific document categories, making it more suitable for datasets with less variation. Previous research by Siregar (2023) also found higher TF-IDF scores compared to TF-RF in text summarization using MMR, due to the high variability and range of topics in news articles, which led to lower TF-RF scores.

3.7. Recommended Place Branding for Thematic Parks in Malang City

Thematic parks play a crucial role in place branding, which involves promoting a region to enhance its recognition and attract visitors, investment, and sustainable economic growth (Rista & Siregar, 2019). Visitor reviews of theme parks help form place branding by reflecting current conditions and providing a unique identity (Satria & Fadillah, 2021). These online reviews are key indicators of branding success and recognition (Munawir et al., 2019). Electronic word of mouth (E-

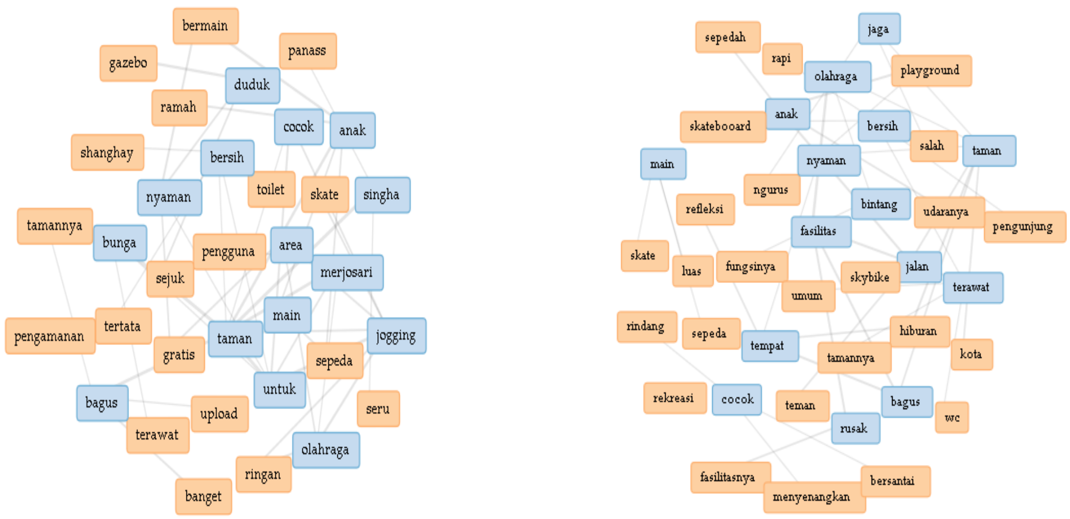
Wom) through social media platforms like Google Maps influences perceptions and decisions of new visitors. E-Wom helps build the image of destinations and encourages visitor retention. Given the vast number of reviews, text mining methods are applied to efficiently analyze and extract the essence of these reviews. This analysis supports recommendations for place branding to ensure theme parks meet their biological and social functions. Word weights are calculated using TF-IDF and TF-RF methods, resulting in two summaries for each park. Fig. 7, Fig. 8, Fig. 9, depict closely related keywords in force-directed network graphs.



(a) (b)

Fig. 7. TF-IDF (a) and TF-RF (b) Taman Tugu summary network graph

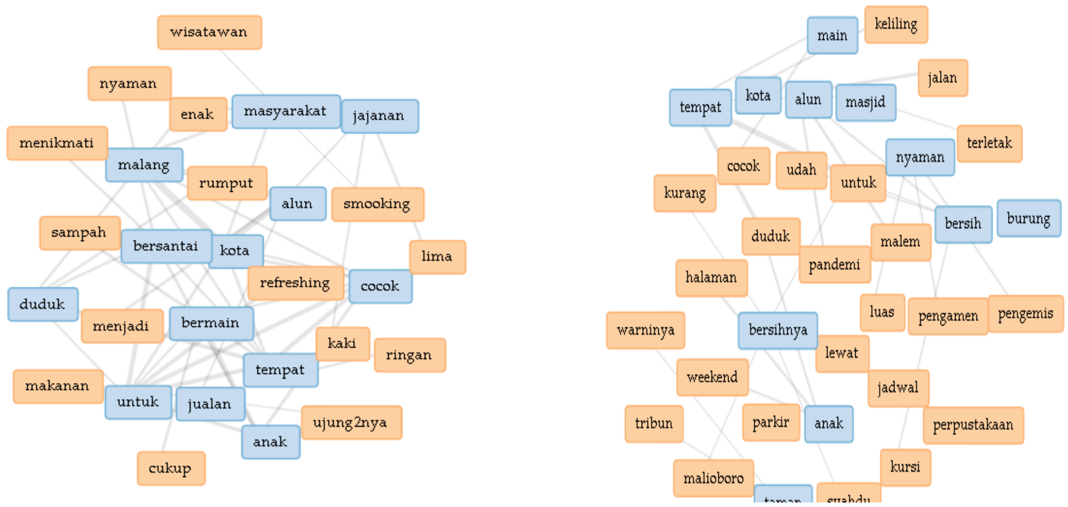
Based on Fig. 7, for the summary results of TF-IDF and TF-RF Taman Tugu, it can be seen that keywords in blue are collocated with terms in orange. This network graph is in accordance with the weighting results in table 4.6 for TF-IDF Taman Tugu and 4.9 for TF-RF Taman Tugu. Such as "bagus", "alun", and "kota" for TF-IDF, and "indah", "nyaman", and "taman" for TF-RF.



(c) (d)

Fig. 8. TF-IDF (c) and TF-RF (d) Merjosari Park summary network graph

Similar to the previous results Fig. 8, displays keywords that are collocated with terms for Taman Merjosari. This is like the word 'taman' which collocates with the terms 'untuk', 'main' and 'olahraga' in the TF-IDF weighting, as well as the keyword 'fasilitas' which collocates with 'bagus' and 'terawat' in the TF-RF.



(e) (f)

Fig. 9. TF-IDF (e) and TF-RF (f) Alun-Alun kota Malang summary network graph

Meanwhile, in Fig. 9, the keyword ‘alun’ is collocated with ‘cocok’, ‘main’, and ‘burung’ in the TF-IDF summarization document, while the keyword ‘tempat’ is collocated with ‘untuk’, ‘duduk’, and ‘anak’. By utilizing the relationship between keywords and terms, a place branding recommendation is made from the relationship pattern. A summary of the terms used as recommendations for place branding is then displayed in Table 8.

Table 8. Summary of Keyword Relationships from Online Reviews of Thematic Parks

	TF-IDF	TF-RF
Taman Alun-Alun Tugu	Taman kota alun-alun tugu Malang adalah taman yang nyaman, rapi, dan indah. Terdapat kolam teratai dan cocok untuk berfoto. Dekat dengan stasiun, serta tugunya jadi simbol sejarah.	Alun-alun tugu kota Malang, dekat dengan stasiun & pusat pemerintahan. Tempat menikmati bunga yang cantik. Alun-alunnya bagus untuk berfoto & bersama keluarga.
Taman Merjosari	Taman Singha Merjosari adalah taman bunga yang cocok jadi tempat olahraga. Tempat yang sejuk dan rindang untuk duduk di gazebo. Taman anak untuk bermain skateboard, sepeda, dan refleksi.	Taman dengan fasilitas olahraga, playground, dan arena skateboard. Tempat yang bagus dan rindang. Cocok untuk bersantai dan bermain.
Taman Alun-Alun Malang	Alun-alun malang cocok untuk tempat bersantai, duduk, dan bermain bersama keluarga. Tempat untuk refreshing anak serta banyak jajanan. Tempatnya nyaman, luas, dan banyak burung, serta dekat dengan masjid.	Tempat untuk duduk santai dan refreshing. Taman yang nyaman dan cocok untuk bermain bersama anak.

4. Conclusion

The analysis of the visitors’ perception indicates that Alun-Alun Tugu is known for being a comfortable and beautiful park with a lotus pond, highlighted by the iconic Malang monument in the center. Merjosari Park is recognized as a sports park, supported by user mentions of various sports facilities like skateboarding, cycling, and reflexology. In addition, the Alun-Alun Malang Park is known as a child-friendly park with play and entertainment facilities specifically for children. These perceptions are supported by the highest word weights obtained for each park. In this study, TF-IDF MMR scores are higher than TF-RF MMR scores for each thematic park dataset. This is because TF-IDF assigns weights based on the variety of the corpus. The greater the variety in the dataset, the more unique weights it has. Thus, TF-IDF gives higher weights to words that appear most frequently in each document. Meanwhile, TF-RF considers word weights based on their occurrence and relevance in each document within the dataset. This method takes into account the occurrence of words in specific categories(Ghofany et al., 2022). Therefore, TF-RF is more suitable for datasets that emphasize relevance and specificity in a particular context, where the corpus is less varied. Previous research by Siregar (2023), which also combined TF-IDF and TF-RF for text summarization using MMR, showed higher TF-IDF scores compared to TF-RF MMR scores. This is due to the dataset being news articles with various topics, resulting in high corpus variability and term types, thus causing lower TF-RF scores compared to TF-IDF MMR. This study is limited to analyzing perceptions based on the highest weighted words using two word-weighting methods: TF-IDF and TF-RF. To further classify visitor perceptions, a follow-up study involving sentiment analysis would be highly beneficial for those interested in developing thematic parks, as it would help identify both positive and negative sentiments about the park.

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding: This This research received no external funding.

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

References

Agustian, E., & Razali, G. (2023). *Pengembangan Kota Berbasis City Branding: Pendukung Investasi, Komunikasi Pariwisata, dan Pemasaran* (1st ed.). CV Seribu Bintang. https://www.seribubintang.web.id/index.php?p=show_detail&id=105

Ahmad Gozali, H., & Alfan Rosid, M. (2020). Klasifikasi Keluhan Mahasiswa dengan Metode Naive Bayes dan Sastrawi. *Network, and Computer Science)* |, *3*(1), 22–26. https://joincs.umsida.ac.id/index.php/joincs/article/download/711/1709/

Assidyk, A. N., Setiawan, E. B., & Kurniawan, I. (2020). Analisis Perbandingan Pembobotan TF-IDF dan TF-RF pada Trending Topic di Twitter dengan Menggunakan Klasifikasi K-Nearest Neighbor. *E-Proceeding of Engineering*, *7*(2), 7773–7781. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/12794

Ayu, A. P. (2019). Peran Ruang Terbuka Hijau dalam Citra Kota Studi Kasus: Taman Suropati, Jakarta. *Jurnal Ilmiah Desain \& Konstruksi*, *18*(1), 53–66. http://dx.doi.org/10.35760/dk.2019.v18i1.1958

Fikriyah, Purnamasari, W. D., & Hasyim, A. W. (2021). Persepsi Pengguna Taman terhadap Kualitas Penggunaan Taman Kota di Kecamatan Klojen. *Planning for Urban Region and Environment*, *10*(1), 157–168. https://purejournal.ub.ac.id/index.php/pure/article/view/120

Ghofany, M. S. Al, Dwiyansaputra, R., Bimantoro, F., & Khairunnas. (2022). Indonesian SMS Spam Detection Using TF-RF Feature Weighting Method and Support Vector Machine Classifier. In *Proceedings of the First Mandalika International Multi-Conference on Science and Engineering 2022, MIMSE 2022 (Informatics and Computer Science)* (Vol. 1, Issue 1). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-084-8

Harmandini, K. P., & L, K. M. (2024). *Analysis of TF-IDF and TF-RF Feature Extraction on Product Review Sentiment*. *8*(2), 929–937. Doi: 10.33395/sinkron.v8i2.13376

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods*, *25*(1), 114–146. https://doi.org/10.1177/1094428120971683

Irfan, A. (2023). *Disbudpar: Jumlah pengunjung di Taman Tematik Tangerang meningkat*.

Juwanda, V., & Widiastuti, D. A. (2023). Dampak Kepuasan Pengunjung sebagai Mediasi antara Hubungan Experiental Marketing terhadap Revisit Intention (studi kasus pada pariwisata di Pantai Carocok dan kawasan Mandeh). *Journal of Entrepreneurship, Management and Industry*. https://journal.bakrie.ac.id/index.php/JEMI/article/view/2430

Kasanah, A. N., Muladi, M., & Pujianto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *3*(2), 196–201. https://doi.org/10.29207/resti.v3i2.945

Kotu, V., & Deshpande, B. (2019). Data Science Concepts and Practice: Second Edition. In G. Jones (Ed.), *Data Science* (2nd ed.). Jonathan Simpson. https://doi.org/10.1016/c2017-0-02113-4

Kurniawan, A., & Humaidy, M. I. (2022). Penerapan Algoritma Maximum Marginal Relevance Dalam Peringkasan Teks Secara Otomatis. *Bulletin of Data Science*, *1*(2), 49–56. https://ejurnal.seminar-id.com/index.php/bulletinds/article/view/1358

Larasati, A., Editya, R. P., Chen, Y.-W., & Darmawan, V. E. B. (2023). The Effect of Data Splitting Ratio and Vectorizer Method on the Accuracy of the Support Vector Machine and Naïve Bayes Model to Perform Sentiment Analysis. *2023 8th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, 1–6. DOI: 10.1109/ICEEIE59078.2023.10334755

Levochkina, N. A. (2017). Theme Parks and Their Contribution to the Development of Territories. *IOP Conference Series: Earth and Environmental Science*, *13*(8), 41–48. https://doi.org/10.1088/1755-1315/204/1/012025

Luthfi, M. F., & Lhaksamana, K. M. (2020). Implementation of TF-IDF Method and Support Vector Machine Algorithm for Job Applicants Text Classification. *Jurnal Media Informatika Budidarma* , *4*(4), 1181–1186. https://doi.org/10.30865/mib.v4i4.2276

Manjari, K. U., Rousha, S., Sumanth, D., & Sirisha Devi, J. (2020). Extractive Text Summarization from Web pages using Selenium and TF-IDF algorithm. *Proceedings of the 4th International Conference on Trends in Electronics and Informatics, ICOEI 2020*, *Icoei*, 648–652. https://doi.org/10.1109/ICOEI48184.2020.9142938

Mihardja, E. J., Mulyasari, P., Widiastuti, T., & Bintoro, B. K. (2020). *Strategi City Branding* (Issue 1). Penerbit Universitas Bakrie. https://slims.bakrie.ac.id/index.php?p=show_detail&id=3791

Munawir, Koerniawan, M. D., & Dewancker, B. J. (2019). Visitor perceptions and effectiveness of place branding strategies in thematic parks in Bandung City using text mining based on google maps user reviews. *Sustainability (Switzerland)*, *11*(7), 1–20. https://doi.org/10.3390/SU11072123

Nugroho, F. I., Setyono, D. A., & Kurniawan, E. B. (2022). Identifikasi Kriteria Layak Huni Permukiman Di Kota Malang. *Tata Kota Dan Daerah*, *14*(1), 1–8. https://doi.org/10.21776/ub.takoda.2022.014.01.1

Nurahaliza, F. N. U., & Mulyadi, N. F. U. (2022). Aplikasi Sistem Informasi Dan Pemetaan Daerah Pariwisata Berbasis Web. *Jurnal Tika*, *7*(1), 48–54. https://doi.org/10.51179/tika.v7i1.1085

Oliveira, T., Araujo, B., & Tam, C. (2020). Why do People Share Their Travel Experiences on Social Media? In *Tourism Management* (Vol. 78, pp. 1–14). https://doi.org/10.1016/j.tourman.2019.104041

Rista, M., & Siregar, A. (2019). Komunikasi Kota Ruang Publik Taman Sebagai Pembentuk Citra Kota Hijau. *Jurnal Komunikasi Pembangunan*, *17*(1), 102–113. https://doi.org/10.46937/17201926595

Saraswati, N. F., Indriati, & Perdana, R. S. (2018). Peringkasan Teks Otomatis Menggunakan Metode Maximum Marginal Relevance Pada Hasil Pencarian Sistem Temu Kembali Informasi Untuk Artikel Berbahasa Indonesia. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *2*(11), 5494–5502. https://doi.org/10.1016/s1010-6030(01)00380-x

Sari, Y., Baskara, A. R., Prakoso, P. B., Royani, N., Mangkurat, U. L., & Korespondensi, P. (2022). *Perbandingan Metode Pembobotan TF-RF Dan TF-IDF dengan dikombinasikan Dengan Weighted Tree Similarity Comparison of Tf-Rf and . Weighting Methods Tf-Idf With Combination With Weighted Tree Similarity for a Book*. *9*(3). https://doi.org/10.25126/jtiik.202295709

Sari, Y., Baskara, A. Ri., Prakoso, P. B., & Royani, N. (2022). Perbandingan Metode Pembobotan Tf-Rf Dan Tf-Idf Dikombinasikan Dengan Weighted Tree Similarity Untuk Sistem Rekomendasi Buku. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *9*(6), 1323–1332. https://doi.org/10.25126/jtiik.2022935709

Satria, F., & Fadillah, F. (2021). Konsep City Branding dan Identifikasi Nilai Lokal pada Kota-Kota Indonesia dalam Mendukung Nation Branding Indonesia. *Jurnal Desain*, *8*(2), 147. https://doi.org/10.30998/jd.v8i2.8118

Siregar, J. S. (2023). *Perbandingan Metode Pembobotan Kata TF-IDF dengan TF-RF pada Peringkasan Teks Berita Otomatis dengan Metode Maximum Marginal Relevance*. Institut Teknologi Sumatera. https://repository.itera.ac.id/depan/submission/SB2305310009

Thakur, K., & Kumar, V. (2022). Application of Text Mining Techniques on Scholarly Research Articles: Methods and Tools. *New Review of Academic Librarianship*, *28*(3), 279–302. https://doi.org/10.1080/13614533.2021.1918190

Wihardjo, P. H. (2020). *Upaya Pemerintah Kota Malang Dalam Menyediakan Ruang Terbuka Hijau (Studi di Dinas Lingkungan Hidup Kota Malang)*. Universitas Katolik Widya Karya. http://repository.ukwk.ac.id/handle/123456789/503