BALI TOURIST VISITS CLUSTERED VIA TRIPADVISOR REVIEWS USING K-MEANS ALGORITHM

Ufik Alngatiq Hidayat Wamulkan A.S[1*]; Nengah Widya Utami[2]; I Nyoman Yudi Anggara[3]

[1,2,3] Teknik Informatika
[1,2,3] Primakara Universiity
www.primakara.ac.id
ufikhidayat51@gmail.com[1*], Indonesiawidya@primakara.ac.id[2], Inyomanyudi@primakara.ac.id[3]
(*) Corresponding Author

Abstract—*Bali is one of the provinces with the most popular destinations for tourists. However, there are obstacles in developing tourist destinations in the province of Bali in terms of absorbing more tourist visits. Tripadvisor, the world's largest tourism information platform. In order to improve its service to users, Tripadvisor conducts online reviews to obtain ratings based on travel experience. The purpose of this study is to determine clustering and accuracy in tourist visits to tourist destinations in the province of Bali. Clustering is done using 3 clusters using the KDD method. The first process is data selection, then data processing which consists of several stages, first deleting rows of empty data, then cleaning duplicate data and the final result is 5261 clean data then data transformation, so that data can be read by python, The next process is data mining, this process uses the K-Means clustering algorithm which produces 3 clusters with cluster 1 being medium with 1495 data, high cluster 2 with 2315 data, and low cluster 3 with 1451 data. The Davies Boldin Index is used to evaluate the K-Algorithm means clustering, the result is 0.3 where the value is very good because it is not minus and the value is close to zero.*

Keywords: *data mining, dbi, kdd, k-means clustering, tripadvisor.*

Abstrak—*Bali merupakan salah satu provinsi dengan destinasi yang paling banyak diminati oleh wisatawan. Namun, terdapat kendala dalam mengembangkan destinasi wisata di provinsi Bali dalam hal menyerap lebih banyak kunjungan wisatawan. Tripadvisor, platform informasi pariwisata terbesar di dunia. Dalam rangka meningkatkan pelayanannya kepada pengguna, Tripadvisor melakukan tinjauan online untuk mendapatkan penilaian berdasarkan pengalaman perjalanan. Tujuan penelitian ini yaitu untuk mengetahui clustering dan akurasi dalam kunjungan wisatawan ke desinasi wisata di provinsi Bali. Clustering dilakukan dengan menggunakan 3 cluster menggunakan metode KDD. Proses pertama yaitu data selection, kemudian processing data yang terdiri dari beberapa tahap, pertama menghapus baris dari data yang kosong, kemudian membersihkan data yang duplikat dan hasil akhir data bersih sebanyak 5261 selanjutnya tranformasi data, agar data dapat dibaca oleh python, Proses selanjutnya adalah data mining, proses ini menggunakan algoritma K-Means clustering yang menghasilkan 3 cluster dengan cluster 1 sedang dengan jumlah data 1495, cluster 2 tinggi dengan jumlah data 2315, dan cluster 3 rendah dengan jumlah data 1451. Davies Boldin Index digunakan untuk mengevaluasi dari algoritma K-means clustering, hasilnya 0.3 dimana nilai tersebut sangat baik karena tidak minus dan nilai sudah mendekati nol.*

Kata Kunci*: data mining, dbi, kdd, clustering k-means, tripadvisor.*

INTRODUCTION

Tourism in Bali is the most important sector, as it serves as a significant source of income for both the local population and the provincial government of Bali. However, there are challenges in developing tourist destinations in Bali, particularly in achieving an even distribution of tourist visits among different attractions. This imbalance has led to a lack of economic improvement in specific sectors (Wayan Suparta, 2021).

To address these issues, efforts are needed to distribute tourist visits more evenly. One approach is to group potentially developable tourist attractions using a clustering method. In response to the challenges faced by the tourism industry, various online platforms have been developed to offer services such as online booking for flights, hotels, and accommodations, and provide tourism information for various regions. One of the leading online tourism platforms adopted globally is TripAdvisor, the world's largest tourism information platform. To enhance its services, TripAdvisor conducts online reviews to collect ratings based on travel experiences in various

destinations. Data mining can serve as a tool to manage extensive online review data (Seimahuira, 2021).

The utilization of data mining techniques for the analysis and categorization of tourist destinations represents a significant approach in the governmental context. Data mining, as a process, is employed to extract essential information or patterns from extensive datasets, thereby generating new knowledge and insights (Utami et al., 2019). An integral aspect of data mining is clustering, an unsupervised method that involves grouping data into classes based on similarity. Clustering techniques, such as Fuzzy C-means, K-means clustering, and Hierarchical Clustering, are instrumental in this regard, with the K-Means Clustering algorithm being a preferred choice for data grouping (Herlinda & Darwis, 2021).

K-Means, introduced by MacQueen in 1967, is a non-hierarchical clustering method known for its simplicity and generality. The algorithm efficiently groups data with similar characteristics, demonstrating a capacity to handle large datasets with relative computational speed. However, it is essential to acknowledge a drawback associated with the determination of initial cluster centers (Ikotun et al., 2023; Hasibuan et al., 2021).

Previous research endeavors have exemplified the practical application of data mining and the K-Means algorithm in identifying popular tourist destinations. Syarah Seimahuira's study on TripAdvisor data highlighted the best destinations in East Asia, , while Nero Caesar Gosari's research in South Sulawesi identified that Malaysia had the most tourists and was followed by Singapore, Japan, South Korea, Taiwan, China, India, Philippines, Hongkong, Thailand, Australia, USA, UK, Netherlands, Germany, France, Russia, Saudi Arabia, Egypt, United Arab Emirates, Barhain and Switzerland (Seimahuira, 2021;Gosari & Rismayani, 2023). Bambang Setyo Purnomo and Putri Taqwa Prasetyaningrum's work in Yogyakarta emphasized Gembira Loka Zoo as the top tourist attraction. These studies collectively underscore the efficacy of the K-Means algorithm in diverse geographical contexts (Purnomo & Prasetyaningrum, 2021).

Observations and interviews with Mr. Agus Putu Yudiantara, responsible for tourism attraction development, unveil critical issues in the distribution of tourist visits across Bali's regencies. The uneven distribution has resulted in a lack of economic improvement, as evidenced by a significant decrease in tourist visits from 2020 to 2022. The data, sourced from entry points such as the I Gusti Ngurah Rai International Airport, the Gilimanuk Port, and the Padang Bai Port, underscores the need for strategic interventions (Yuwono & Kristini, 2023).

To address these challenges, clustering emerges as a crucial strategy. It is proposed that clustering can assist the Bali provincial tourism department in identifying regencies requiring more attention and promotion. This strategic approach aims to stimulate economic sectors and job opportunities, fostering sustainable tourism development in Bali. The proposed integration of clustering techniques aligns with the overarching goal of optimizing the economic potential of each regency, contributing to the holistic advancement of Bali's tourism landscape(Amrita et al., 2021).

Based on the described issues, the author conducted a study titled "Clustering Tourist Visits to Tourist Destinations in Bali Province Based on TripAdvisor Online Reviews Using the K-Means Algorithm." This research aims to assist the Bali provincial tourism department in grouping tourist destinations as a reference for promoting these tourist destinations.

MATERIALS AND METHODS

The methodology employed in this investigation is the Knowledge Discovery in Databases (KDD) approach, a method applied across various phases of the research continuum. TripAdvisor serves as the primary data source for this study, and the data collection technique encompasses web scraping facilitated by data mining application.

Method of Collecting Data

Data collection in this research involves scraping using data mining tools, and here are the steps involved:

1. Opening the Data Miner Extension in Google Chrome: Open Google Chrome and access the Data Miner extension.
2. Configuring the Target Website: Configure the Data Miner extension to work with the TripAdvisor website. This may involve specifying the URL, setting up authentication if required, and defining the structure of the data you want to extract.
3. Performing Searches with Defined Parameters: Use the Data Miner extension to initiate searches on the TripAdvisor website. You can specify search criteria or parameters to retrieve specific data.
4. Collecting Data from Each Page: Data Miner can be set to automatically navigate through multiple pages of search results, and for each page, it can extract the data you specified. This may include details such as reviews, ratings, locations, and more.
5. Saving Data in CSV Format: Once the data has been extracted, you can use the Data Miner

extension to save it in a structured format, such as CSV (Comma-Separated Values). CSV is a commonly used format for tabular data that can be easily imported into data analysis tools(Christian & Hakim, 2019).

In this process, web scraping is conducted to collect data from TripAdvisor. This data can then be used for various data mining and analysis purposes as part of the Knowledge Discovery in Databases (KDD) process. Web scraping is a valuable method for gathering data from websites like TripAdvisor to extract valuable information and patterns for further analysis and insights.

Datasets

The dataset employed in this study pertains to tourist destination visits sourced from the TripAdvisor website. To procure this dataset, the author systematically employed data mining tools, resulting in the extraction of pertinent information. The harvested data was subsequently organized and stored in CSV format under the nomenclature "data_kunjungan.csv." A representative example of the tourist destination visit dataset is delineated in Table 1, comprising a comprehensive total of 5,686 individual data points.

The data encapsulated within the "data_kunjungan.csv" file provides a nuanced insight into the myriad facets of tourist destination visits, as recorded on the TripAdvisor platform. Utilizing data mining techniques enabled the systematic curation of a dataset that encapsulates critical information germane to tourist preferences, behaviors, and trends. This dataset serves as a foundational resource for the subsequent analytical processes integral to the study's objectives.

Table 1 serves as a visual representation of the structured dataset, offering a glimpse into the voluminous nature of the compiled information. The 5,686 data points within this dataset are indicative of the robustness and comprehensiveness of the collected information. This meticulously curated dataset from TripAdvisor forms the cornerstone of the empirical foundation upon which the ensuing analyses and inferences will be predicated. The utilization of such a rich and extensive dataset ensures a rigorous and thorough examination of the dynamics underpinning tourist destination visits, thereby facilitating the attainment of nuanced insights and scholarly contributions to the field.

Table 1. Dataset

No	Name	Number of Reviews	Rating	Facilities	District	Category
1	Kanva Ubud	17	5	Free Parking	Gianyar	Hotel, Resort
2	RIMBA by AYANA Bali	3.004	4.5	Includes Breakfast	Badung	Hotel, Resort
3	Amnaya Resort Kuta	2.414	5	Free Parking	Badung	Hotel, Resort
4	AYANA Resort Bali	4.850	4,5	Swimming Pool	Badung	Hotel, Resort
5	The Sankara Resort Ubud by Pramana	2.567	5	Free Wi-Fi	Gianyar	Hotel, Resort
6	The Seminyak Beach Resort & Spa	3.791	4.5	Free Parking	Badung	Hotel, Resort
7	manis Villas	18	4,5	Free Wi-Fi	Badung	Hotel, Resort
8	Padma Resort Ubud	4.837	5	Free Parking	Gianyar	Hotel, Resort
9	Hard Rock Hotel Bali	8518	4	Free Wi-Fi	Badung	Hotel, Resort
10	The Kayon Resort	2.998	5	Free Parking	Gianyar	Hotel, Resort
....	Pantai legian	3.520	4	Free Wi-Fi	Badung	Beach
5.684	Pulau Menjangan	1,922	4	Free Parking	Buleleng	Museum, Park
5.685	Krisna Funtastic Land	53	4	Free Wi-Fi	Buleleng	Museum, Park
5.686	Rumah Taman	47	4	Free Parking	Gianyar	Museum, Park

Knowledge Discovery in Databases

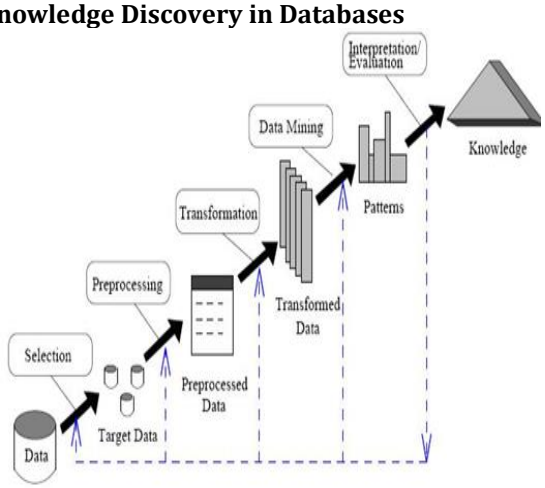

Figure 1. KDD Stage

KDD (Knowledge Discovery in Databases) is a procedure designed to unveil valuable patterns or insights within extensive datasets. The KDD stages comprise various steps, with "Data Mining" being a pivotal phase. Here is a concise overview of the stages in the KDD process. Data Selection, opting for a pertinent subset of data from expansive sources for subsequent analysis. Data Preprocessing, refining, altering, or merging data to eradicate flaws and ensure high data quality. Data Transformation, converting raw data into a more suitable format for analysis, potentially involving normalization, aggregation, or recoding. Data Mining, the extraction of patterns or knowledge from data through specific algorithms and techniques, encompassing methods like classification, clustering, regression, or association. Evaluation of

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 19, No. 2 September 2023 | DOI: 10.33480/pilar.v19i2.4571

Data Mining Results, gauging and scrutinizing the outcomes of the data mining process to ensure the significance and durability of the identified patterns or knowledge. Knowledge Utilization, assimilating newly uncovered knowledge into decision-making processes or pertinent systems(TÜRK, 2023).

Data Selection

Data selection, part of the Knowledge Discovery in Databases (KDD) process, involves extracting a subset of operational data before the actual information extraction phase in KDD (Savitri et al., 2021). The resulting data will be used for subsequent processes and stored separately from the operational database. The search parameters in this research include the name of the tourist destination, number of reviews, rating, facilities, and location of the tourist destination. The data obtained from scraping using data mining tools amounted to 5686 data. Data validation was carried out using a sampling technique by taking 1000 data randomly and then comparing them on the Tripadvisor website. The data scraping process does not take personal data such as user names to maintain the privacy of each user.

Preprocessing

During the preprocessing or cleaning phase, data refinement involves the removal of duplicate entries, rectification of disruptive data, verification of data consistency, correction of typographical errors, and augmentation of data with pertinent information for the Knowledge Discovery in Databases (KDD) process (Hasibuan et al., 2021). Following this preprocessing stage, the dataset, initially comprising 5,686 entries, undergoes refinement, resulting in a reduced dataset size of 5,261 entries.

Transformation

One pivotal step in the transformation process involves coding selected data to prepare it for subsequent Knowledge Discovery in Databases (KDD) processes. This transformative stage, guided by patterns identified in the database, is contingent upon the nature or pattern of the information present (Purnomo & Prasetyaningrum, 2021). During this phase, the data undergoes a conversion process, utilizing Python for classification to render it in a readable and processable format. The transformation is executed by converting the data into numeric values, such as assigning districts with numerical codes (e.g., 1= Jembrana, 2= Buleleng, 3= Tabanan, etc.) to facilitate computational processing with Python.

Data Mining

It is the process of applying algorithms or methods for knowledge discovery. At this stage, the mining process is carried out using the K-means Clustering algorithm.

1. The initial phase involves ascertaining the optimal quantity of clusters to be established.
2. Next, the centroid value (center point) needs to be determined.
3. When establishing the centroid value at the commencement of the iteration, the initial centroid is selected through a random process. Conversely, if the determination of the centroid value constitutes a phase within the iteration, the subsequent formula is employed.:

$$\overline{V_{ij}} = \frac{1}{N_i}\sum_{k=0}^{N_i} x_{kj} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

   a. V_{ij} is the centroid/average of the I to cluster for the j variable
   b. N_i is the amount of data that is a member of the cluster
   c. i, k are the indices of the cluster
   d. j is the index of the variable
   e. x_{kj} is the kth data value in the cluster for the j to variable
4. After that, the distance between the centroid point and each data object is calculated.

$$d(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \cdots + (x_{ki} - x_{kj})^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

   Information
   $d(i,j)$= distance of data i to the center of cluster j
   x_{ki} = i data in k data attribute
   x_{kj} = jth center point on kth attribute
5. Grouping objects to determine cluster members is done by taking into account the minimum distance of the object.
6. Iteratively revert to stage 2, persisting until the attained centroid value exhibits constancy, and the membership status of clusters ceases to undergo relocation to alternative clusters (Seimahuira, 2021).

Interpretation/Evaluation

The process of producing understandable output comes from the data mining process of information patterns. In this process, the Davies Bouldin Index (DBI) technique is used.

RESULTS AND DISCUSSION

In this stage, the Knowledge Discovery from Data (KDD) process will be employed. The implementation of KDD is carried out using Python Google Colab and Microsoft Excel.

DBI (Davies Bouldin index)

The Davies Bouldin Index (DBI) is a metric for evaluating or considering the results of a clustering algorithm. The first was introduced by David L. Davies and Donald W. Boudin in 1979 (Anggi Riyanto et al., 2022). By using DBI, a factory is considered to have an optimal clustering scheme, namely one that has a minimum DBI. The steps for calculating Davies Bouldin Index are: As follows :

a. Sum Of Square Within-Cluster (SSW)
To understand cohesion in a cluster, one way is to calculate the value of the Sum of Squares Within Clusters (SSW). With the following formula:

$$SSW_i = \frac{1}{mi} \sum_{j=i}^{mi} d(X_j, C_j) \text{ ..... (3)}$$

Where:
mi = the amount of data in the cluster
ci = centroid of the cluster
$d(X_j, C_j)$ = distance of each data to centroid I which is calculated using the Euclidean distance.

b. Sum Of Square Between-Cluster (SSB)
The Sum of Square Between-Cluster (SSB) calculation aims to find out the separation or distance between clusters. with the following calculation formula:

$$SSB_{ij} = d\ (X_i, X_j) \text{ ..... (4)}$$

Where :
$d\ (X_i, X_j)$= distance between i and j in another cluster.

c. Ratio (Ratio)
This ratio calculation (Ri, j) aims to find out the comparative value between the ith cluster and the jth cluster to calculate the ratio value held by each cluster. Indices l and j represent the number of clusters, where if there are 4 clusters then there are 3 indexes, namely i, j, and k. To determine the ratio value use the following formula:

$$R_{ij\ldots\ldots,n} = \frac{SSW_i + SSW_j + \cdots + SSW_n}{SSB_{i,j} + \cdots + SSB_{ni,nj}} \text{ ..... (5)}$$

Where :
SSW_i = Sum Of Square Within-Cluster at centroid i
$SSB_{i,j}$ = Sum of Square Between Cluster data i with j on different cluster
In the calculation formula n will continue the number of clusters selected with these conditions is not the same as nj.

d. Davies Bouldin Index (DBI)
The ratio value obtained from the ratio formula is used to find the DBI value using the following calculation:

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max\ _{i \neq j}(Ri, j, \ldots k) \text{ ..... (6)}$$

Where, Ri, is the basic ratio of SSW and SSB through calculation, then it can be known that there are several clusters. From the calculation of the Davies Bouldin Index (DBI), it can be concluded that the smaller the Davies Bouldin Index (DBI) value obtained (non-negative >= 0), the better the cluster (Firman Ashari et al., 2022).

Results Using Google Colab

The first stage involves data selection, which is the initial step in data classification. The data used pertains to tourist visits via Tripadvisor. The author selects the data to be used using Microsoft Excel with the datakunjungan.csv file format (Winarta & Kurniawan, 2021).

The second stage involves data preprocessing, where the obtained data is cleaned to enhance the accuracy of the data classification process. Data with missing values/null, duplicate data, and similar issues are cleaned. The following shows the count of missing values/null.

Table 2. Quantity Data null

No	0
name	0
number of reviews	15
rating	48
facilities	415
district	1
category	0
dtype: int64	

Subsequently, data with missing values/null are removed, resulting in data as presented in the following table.

Table 3. Final Data

No	5261
name	5261
number of reviews	5261
rating	5261
facilities	5261
district	5261
category	5261
dtype: int64	

The next step is to eliminate duplicate data, although, in this research, no duplicate data was found, as shown in the table 4.

Table 4. Removal of Duplicate Data

Number of duplicate rows	0
Number of rows before duplicate removal	5261
Number of rows after duplicate removal	5261

Next is the removal of unused data such as numbers and names.

The third stage involves data transformation. The data will be transformed using Python for classification, converting it into a format that can be read and processed with Python. The initial step

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 19, No. 2 September 2023 | DOI: 10.33480/pilar.v19i2.4571

before data transformation involves dividing the data into X and Y for each data column. The variable X serves as the lister ent (feature), while the dependent variable Y (target) will be used as the target in model training. The lister ent variable will serve as the feature used to predict the target.

The fourth stage involves the data mining process. In this research, clustering is performed using the K-Means algorithm with 3 clusters and a random state value of 30 for consistent random initialization across code runs. The clustering results are explained in Figure 2.

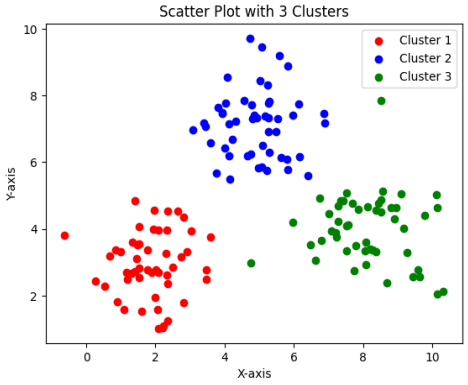

Figure 2. Clustering Results

Figure 2 displays the data distribution in each cluster. Cluster 1 is marked in red, cluster 2 in blue, and cluster 3 in green.

The fifth stage is the interpretation or evaluation process. In this phase, iterations are performed from 2 to 10 to test various numbers of lister in the K-means algorithm. For each iteration, clustering is conducted using K-means with different numbers of lister, the Davies-Bouldin index is calculated for evaluation, and the results are stored. By doing so, a results dictionary is obtained, containing Davies-Bouldin indices for each tested number of the lister. The results are shown in the Figure 3.

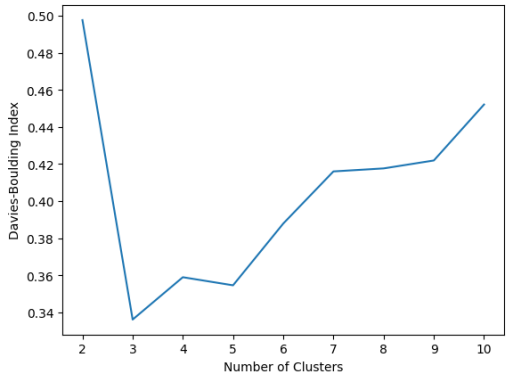

Figure 3. DBI Results

The graph indicates that the DBI results for clusters 2 through 10 suggest that cluster 3 has the best accuracy. Further details can be observed in the following table. The choice of several clusters will certainly affect the clustering results and accuracy of the K-Means itself, and that is a weakness of the K-Means algorithm. The accuracy result of clustering using the K-means algorithm is 0.33.

The next stage involves finding the final centroid values for each cluster, as shown in the table 6.

Table 5. Final Centroid Results

Cluster	Number of reviews	rating	facilities	district	category
1	355.178834	4.476415	2.421598	4.952916	3.092441
2	214.326671	4.729841	10.556168	6.070296	2.043418
3	456.114381	4.226087	10.880936	3.886957	2.004682

The table 5 displays the average values for each cluster. Cluster 1 has an average of 355 reviews, a rating of 4.4, 2 facilities (toilet), European food type, district 5 (Denpasar), and category 3 (restaurant). Cluster 2 has an average of 214 reviews, a rating of 4.7, 10.5 facilities (parking and Wi-Fi), European and African food types, district 6 (Gianyar), and category 2 (hotel and resort). Cluster 3 has an average of 456 reviews, a rating of 4.2, 11 facilities (parking and cafe), European and American food types, district 4 (Badung), and category 2 (hotel and resort).

From the data in the clustering results in the table, we can get policies for the Bali provincial tourism office and tourist destinations, such as carrying out promotions in Badung district which has the lowest cluster, which is a result of ratings and facilities. Then for hotel and restaurant owners from the clustering results using the K-Means algorithm, it is recommended to improve facilities.

In the subsequent step, the number of data in each cluster is calculated for a total of 3 clusters. The results are presented in the table 6.

Table 6. Number of Data in Each Cluster

No	Data
1	1495
2	2315
3	1451

Based on the number of data in each cluster, Cluster 1 falls into the medium category with 1495 data points, Cluster 2 is in the high category with 2315 data points, and Cluster 3 is in the low category with 1451 data points.

CONCLUSION

The utilization of data mining techniques in clustering tourist visits to destinations within the Bali province, employing the data_kunjungan.csv dataset and the K-Means algorithm, yields insightful conclusions. Firstly, the clustering process categorizes tourist visits into three clusters. Cluster 1, categorized as medium, comprises 29% of the data with a total of 1495 data points. Cluster 2, classified as high, constitutes 44% of the data with 2315 data points. Cluster 3, characterized as low, encompasses 27% of the data, totaling 1451 data points. Secondly, the evaluation of the K-Means algorithm, assessed through the Davies-Bouldin index, yields a small validity value of 0.3, indicating its efficacy compared to alternative cluster numbers. Further refinement of tourist visit clustering can be achieved through the exploration of alternative algorithms, such as Hierarchical clustering, and conducting comparative analyses across multiple algorithms.

Analyzing the clustering results from the dataset offers actionable insights for the Bali provincial tourism office and tourist destinations. Specifically, strategic interventions can be devised, such as targeted promotions in Badung district, which is identified as having the lowest cluster, potentially attributable to factors like ratings and facilities. Furthermore, recommendations for improvement are extended to hotel and restaurant owners based on K-Means clustering results, emphasizing the enhancement of facilities to align with tourist preferences and expectations. This data-driven approach provides a foundation for informed decision-making and policy formulation in the realm of tourism management in Bali.

REFERENCE

Amrita, N. D. A., Handayani, M. M., & Erynayati, L. (2021). Pengaruh Pandemi Covid-19 Terhadap Pariwisata Bali. *Jurnal Manajemen Dan Bisnis Equilibrium*, *7*(2), 246–257. https://doi.org/10.47329/jurnal_mbe.v7i2.824

Anggi Riyanto, A., Daryanto, D., & Abdurrahman, G. (2022). Text Mining Untuk Clustering Buku Di Perpustakaan Menggunakan Metode K-Means. *National Multidisciplinary Sciences*, *1*(6), 835–845. https://doi.org/10.32528/nms.v1i6.239

Christian, B., & Hakim, L. (2019). Penerapan Algoritma Fuzzy C-Means Pada Penentuan Lokasi Gudang Pendukung PT. XYZ. *AITI: Jurnal Teknologi Informasi*, *16*(1), 31–48. https://doi.org/10.24246/aiti.v16i1.31-48

Firman Ashari, I., Banjarnahor, R., Farida, D. R., Aisyah, S. P., Dewi, A. P., & Humaya, N. (2022). Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 6, Issue 1). http://jurnal.polibatam.ac.id/index.php/JAIC

Hasibuan, F. P. A., Sumarno, S., & Parlina, I. (2021). Penerapan K-Means pada Pengelompokan Penjualan Produk Smartphone. *SATESI: Jurnal Sains Teknologi Dan Sistem Informasi*, *1*(1), 15–20. https://doi.org/10.54259/satesi.v1i1.3

Herlinda, V., & Darwis, D. (2021). Analisis Clustering Untuk Recredesineling Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means. *Darwis, Dartono*, *2*(2), 94–99. http://jim.teknokrat.ac.id/index.php/JTSI

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178–210. https://doi.org/10.1016/J.INS.2022.11.139

Gosari, N. C., & Rismayani. (2023). Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Mancanegara Di Prov. Sulawesi Selatan Dengan K-Means Dan SVM. Jurnal Administrasi Dan Manajemen, 8(3), 308–317. https://doi.org/10.30591/jpit.v8i3.4554

Purnomo, B. S., & Prasetyaningrum, P. T. (2021). Penerapan Data Mining Dalam Mengelompokan Kunjungan Wisatawan di Kota Yogyakarta Menggunakan Metode K-Means. In *Journal of Computer Science and Technology JCS-TECH* (Vol. 1, Issue 1). https://doi.org/10.54840/jcstech.v1i1.9

Savitri, N., Pranata, R., Nadzario, A., Clara, M., Rahajeng, S., Politeknik, J. S., & Stis, S. (2021). *Pengelompokan Kunjungan Wisata Kabupaten Kulon Progo Tahun 2019 Menggunakan K-Means Clustering* (Vol. 12, Issue 1). https://jurnal.umj.ac.id/index.php/just-it/index

Seimahuira, S. (2021). Implementasi Datamining Dalam Menentukan Destinasi Unggulan Berdasarkan Online Reviews Tripadvisor Menggunakan Algoritma K-Means. Technologia: Jurnal Ilmiah, 12(1), 53-58.

Yuwono, W., & Kristini, V. (2023). Analisis Faktor-Faktor Yang Mempengaruhi Kepuasan Wisatawan Untuk Berkunjung Ke Kepulauan Riau. Jurnal Pengembangan IT, 8, 174–180.

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 19, No. 2 September 2023 | DOI: 10.33480/pilar.v19i2.4571

http://ejournal.urindo.ac.id/index.php/administrasimanajemen/index

TÜRK, F. (2023). Analysis of Intrusion Detection Systems in UNSW-NB15 and NSL-KDD Datasets with Machine Learning Algorithms. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, *12*(2), 465–477. https://doi.org/10.17798/bitlisfen.1240469

Utami, N. W., Sukajaya, N., Grana, E., & Dewi, A. (2019). *The Implementation of Data Mining to Show UKT (Students' Tuition) Using Fuzzy C-Means Algorithm (Case Study: Universitas Pendidikan Ganesha)*. https://doi.org/10.1109/ICACSIS47736.2019.8979933

Wayan Suparta. (2021). Recovery Pariwisata Bali Pasca Pandemi Covid 19 (Studi kasus: Penegakan Protokol Kesehatan. In *Regulation Number* (Vol. 12). https://doi.org/10.55115/sista.v1i1.1447

Winarta, A., & Kurniawan, W. J. (2021). Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemograman Python. *Jurnal Teknik Informatika Kaputama (JTIK)*, *5*(1). https://jurnal-backup.kaputama.ac.id/index.php/JTIK/article/view/466