JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 3, Juni 2024, hlm. 467 - 474 ISSN 2550- 020 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v10i3.2894
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

IMPLEMENTATION OF DATA ANALYSIS HOTEL RATING LEVELS IN BALI USING THE K-MEANS ALGORITHM AND DECISION TREE

Hamdani[1*], Dedy Hartama[1]
[1]Information Engineering, STIKOM Tunas Bangsa Pematang Siantar
email: *dhannymotovlog@gmail.com

Abstract: The service dramatically affects the number of guests staying at the hotel. Bali is the most visited tourist area by foreign tourists. Therefore, improved service is crucial for determining the rating level of a hotel. This research aims to combine two data mining algorithms: clustering and classification. This research is expected to contribute to hospitality in improving the best services for tourists, especially in the City of Bali. Clustering algorithms are used to group the best number of hotels based on the four clusters selected from the k-means clustering algorithm. The classification algorithm using C4.5 determines the factors most dominant in determining the hotel rating level based on the gain ratio. The data used in this study results from observations on the website agoda.com in Bali of 51 data. The results of this study explained that cluster_0 is the highest-rated cluster, with a total number of 19 hotels found in claster_0. Data cluster0 is used for classification analysis using a decision tree, and the most dominant factor is the service factor, with an accuracy of 80%.

Keywords: data mining; kmeans; decision tree; hotel; bali;

Abstrak: Pelayanan sangat mempengaruhi jumlah pengunjung yang menginap dihotel. Bali merupakan daerah wisata paling banyak dikunjungi oleh wisatawan mancanegara. Oleh karena itu, peningkatan pelayanan sangat penting untuk penentuan level rating dari hotel. Tujuan dari penelitian ini untuk menggabungkan dua algoritma data mining yaitu clustering dan klasifikasi. Dengan penelitian ini diharapkan dapat memberikan kontribusi bagi perhotelan dalam meningkatkan pelayanan yang terbaik bagi wisatawan khususnya di Kota Bali. Algoritma Clustering digunakan untuk mengelompokkan dari jumlah hotel yang terbaik berdasarkan empat cluster yang dipilih dari algoritma clustering berupa k-means. Algoritma klasifikasi menggunakan C4.5 digunakan untuk mengetahui faktor apa yang paling dominan dalam menentukan level rating hotel berdasarkan gain ratio. Data yang digunakan dalam penelitian ini hasil observasi di website agoda.com di bali sebanyak 51 data. Hasil dari penelitian ini menjelaskan dataset cluster_0 merupakan cluster rating tertinggi dengan jumlah 19 hotel yang terdapat di cluster_0. Data cluster_0 digunakan untuk analisis klasifikasi menggunakan decesion tree, didapat faktor yang paling dominan adalah faktor layanan dengan nilai akurasi sebesar 80%.

Kata kunci: data mining; kmeans; decision tree; hotel; bali;

INTRODUCTION

The influx of foreign tourists to Indonesia will increase the country's foreign exchange reserves and improve the economy of residents in tourist areas [1]. Denpasar, the capital city of Bali, is one of the destinations for foreign tourists visiting Indonesia. Indonesia has many beautiful natural attractions, tourist parks, and traditional parks that support the country's tourism development. Indo nesia's geographical location, which has many natural beauties, has attracted foreign tourists to come and see these natural wonders[2]. With excellent and well-planned management, tourism in Indonesia will be able to attract foreign tourists to visit. One closely related area is accommodation. By providing adequ ate and quality facilities, hotel managers must evaluate based on visitor feedback data from various hotels.

The hotel industry prioritizes service to customers[3]. Government Regulation of the Republic of Indonesia No. 65 of 2011, coinciding with September 31, 2001, Article 1, states, "A hotel is a building specifically provided for people to stay or rest, receive services and other facilities for a fee, including other buildings that are integrated and managed and owned by the same party except for shops and offices." Hotels in Bali require new management ideas to increase profits. One idea that can be tried is to set up a hotel management strategy[4]. Such a scheme requires a tourist or customer database and their feedback[5]. The comfort of hotel guests in using hotel services is an essential requirement, so hotel managers must provide the best possible comfort to maintain accommodation services. In the technology and data analysis era, hotel managers can improve hotel service ratings through a data mining approach. The problem in managing hotels[6].

Clustering methods such as K-Means and classification methods like C4.5 can be utilized in the approach [7]. Clustering aims to sort hotel data based on the variables used. In this case, the clustering result yields data clusters with optimal values [4]. From the clustering results, the next step is to perform a classification process using the C4.5 decision support system algorithm. The next step from the optimal cluster data is determining what factors greatly influe nce the hotel service quality rating. The decision support system is designed for the decision-making process. Almost all companies, from small to large scales, use decision support system methods to support their business activities in increasing profits. With the K-Means algorithm, we can cluster based on the highest ranking and apply the C4.5 algorithm to determine the factors causing hotel ratings to increase or decrease[8]. A previous study conducted by Intan Utnasari, titled "Clustering Analysis with K-Means for Grouping Product Sales at the Newton Hotel," concluded that the clustering performed by data mining in this study resulted in three separate cluster lists: top-rated, moderately famous, and less popular product categories.

However, this study still has some weaknesses, such as not explaining whi ch factors dominate Newton Hotel's prod uct sales. Therefore, this study combines two algorithms, clustering, and classifi cation decision tree, to select hotels with the best ratings in Bali using the K-Means algorithm. Then, the C4.5 decis ion tree algorithm determines the most dominant factors in the rating level: Cle anliness, Location, Service, Facilities, Room Comfort, Price, Quality, and Value Rating[9]. It is expected that this research will be beneficial for related institutions and hotel managers in improving hotel service ratings. The novelty of this research lies in applying two algorithms to analyze hotel rating levels in Bali, using the K-Means algorithm and the decision tree algorithm, with a case study in Bali[10].

METHOD

The research used an approach that involved data mining science, algorithms, and applications.

Data Mining

Data mining has several similarities, such as knowledge discovery or pattern recognition. These two terms have their accuracy. The term knowledge discovery or insight discovery is appropriate because the vital goal of data mining is to obtain insights that are still hidden in lumps of data[11]. The term pattern recognition or pattern identification is also appropriate because the understanding that will be explored is in the form of patterns that may also need to be extracted from within the chunk of data being experienced. If the term data mining is used in this article, this is based more on the popularity of this term in data mining activities [12].

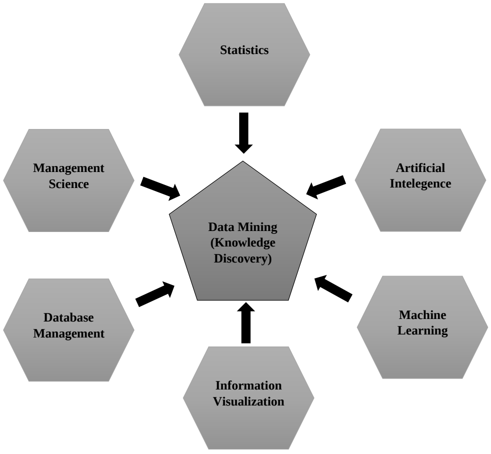


Image 1. Data Mining Architecture

K-Means

The k-means algorithm The grouping equation divides n observation subjects into k clusters, assigning each subject to the cluster with the closest mean. This procedure aims to create a representative point for each group by iterating through various algorithms to correct the data. K-Means is a non-hierarchical (dividing) data grouping method that attempts to partition the data found into 2 or more groups[13]. This method system partitions data into groups. Finally, data with similar characteristics is put into one group, and data with different characteristics is grouped into another group.

The k-means algorithm is an algorithm that requires an input standard of k and sorts a group of n subjects into k clusters so that the level of similarity between units in one cluster is significant. In contrast, the level of similarity with units in other clusters is minimal. The similarity of units to a cluster is measured by the subject's familiarity with the mean number in the cluster or can be said as the cluster centroid or center of mass[14]. This section is the sequence of research methods carried out, including starting with system design, data collection, data processing, then Decision Tree, grouping using K-Means, and evaluating and verifying the results; each sub-chapter will be explained in the next section [15].

Decision Tree

A decision support system is specifically designed to collect decisions[16]. Almost all industries, starting from small ratios or large ratios in carrying out employee income activities, are beginning to make decisions using decision support systems to support the capabilities of their company's activities[17]. Data mining using decision tree procedures is widely used to deal with cases with large amounts of information. This decision tree procedure is a grouping procedure that is widely used because it is built relatively quickly, the results of the form formed are easy to understand, and the estimation results are solid, so they can help gather decisions [18].

Research Stages

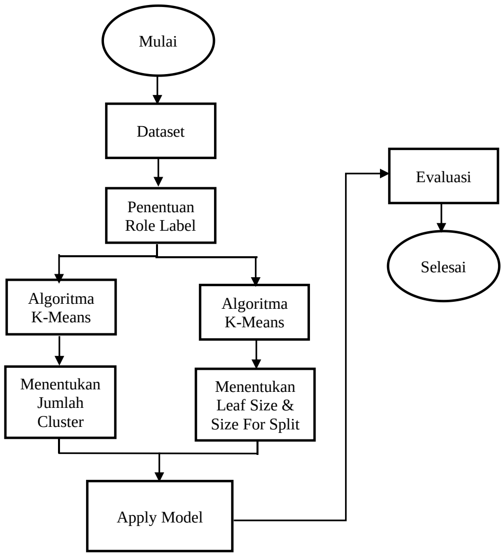


Image 2. Stages of the Research Process

They start by processing the dataset, which is the provision determined using the standard K-Means algorithm, and the decision tree, which is processed based on the existing dataset. Then, they choose the set of roles that become labels. The first stage, applying the K-Means algorithm, starts with entering the dataset and determining the number of clusters. After the results are available, the decision tree algorithm will be applied based on the results from the K-Means algorithm. Then, the leaf size and size for the split will be determined, and the algorithm will be validated.

Data Analysis

Information analysis is a way of processing data to create valuable data that can be used to make decisions for solving a case. This analysis method includes grouping information based on its characteristics, eliminating information, transforming knowledge, and creating forms of information to create meaningful data from that information. There are several types of information analysis when conducting research, namely qualitative and quantitative. Qualitative analysis uses an analytical method that does not use mathematics or statistics. In other words, this analysis is carried out by reading charts, diagrams, or other existing data obtained from various sources using unique information-gathering methods[19].

The purpose of qualitative analysis is to create meaning from the data. Quantitative analysis uses mathematical or statistical forms to work with the information. The results of the analysis are generally in the form of numbers that will be presented and explained by the researcher[20]. There are also methods used in quantitative analysis, namely descriptive analysis methods and inferential analysis methods, which have their respective uses.

Following this is the data taken from the Agoda site, which contains data on hotel accommodation in Bali. The table has categories: Cleanliness, location, service, facilities, room comfort, price, quality, price, value/rating, and 51 hotel data in Bali.

Table 1. Bali Hotel Dataset

Hotel Data	Cleanliness	Location	Service	…..	Rating
EDEN Hotel Kuta Bali	8.1	8.6	8.4	…..	Amazing
J Hotel Kuta	7.6	7.8	7.6	…..	Very Good
PrimeBiz Hotel Kuta	8.7	8.7	8.7	…..	Amazing
Amnaya Resort Kuta	9.4	8.9	9.5	…..	Spectacular
…..	…..	…..	…..	…..	…..
Mahogany Hotel	8.6	7.6	8.8	…..	Amazing

Determining Data Types

To implement this data, you need the help of an additional application, Rapid Miner V10, to determine the data type used in each category. Below are Categorical for each data used.

Table 2. Bali Hotel Data Type (K-means)

Hotel Name	Nominal
Hygiene	Real
Location	Real
Service	Real
Facilities	Real
Room Comfort	Real
Price & Quality	Real

Table 3. Bali Hotel Data Type (Decision Tree)

Hotel Name	Nominal
Hygiene	Real
Location	Real
Service	Real
Facilities	Real
Room Comfort	Real
Price & Quality	Real
Price	Integer
Value/Rating	Nominal

Design Rapid Miner

Data entered into Excel starts in column A1 and requires the help of software, namely Rapid Miner V10, to process the data. When the data is entered into the Repository in Rapid Mine, specify the Role Set or "Value/Rating" as the data label or target. The architectural design to apply this data to the Rapid Miner, as in the image 3:

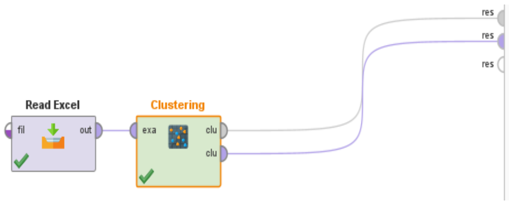


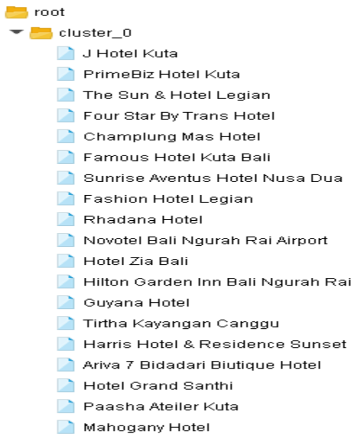
Image 3. K-Means Algorithm Design

After processing in the Rapidminer Application with the K-Means Algorithm with the following parameters:

Table 4. K-Means Algorithm Parameter

Algorithm	Parameters	Value
K-Means	Number of Clusters	4

Hotel ranking data is analyzed using the K-Means algorithm, with parameters set to give specific results based on total data. Get the following results:



root
cluster_0
- J Hotel Kuta
- PrimeBiz Hotel Kuta
- The Sun & Hotel Legian
- Four Star By Trans Hotel
- Champlung Mas Hotel
- Famous Hotel Kuta Bali
- Sunrise Aventus Hotel Nusa Dua
- Fashion Hotel Legian
- Rhadana Hotel
- Novotel Bali Ngurah Rai Airport
- Hotel Zia Bali
- Hilton Garden Inn Bali Ngurah Rai
- Guyana Hotel
- Tirtha Kayangan Canggu
- Harris Hotel & Residence Sunset
- Ariva 7 Bidadari Biutique Hotel
- Hotel Grand Santhi
- Paasha Ateiler Kuta
- Mahogany Hotel

Image 4. Cluster Results (Cluster 0)

It can be concluded that this is the result of the Hotel Cluster in Bali, and the sample was taken from one of the clusters_0. There are 19 hotels in the claster_0. We will process this data again using the Decision Tree Algorithm. The decision tree algorithm parameters.

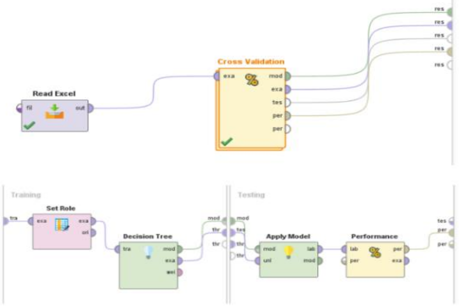


Image 5. Decision Tree Algorithm Parameter

After processing in the Rapid Miner Application with the K-Means Algorithm with the following parameters:

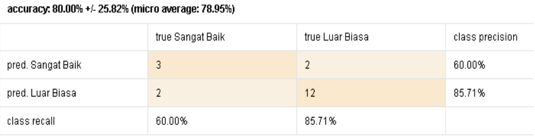
Table 5. Decision Tree Algorithm Parameter

Algorithm	Parameters	Value
Decision Tree	Set Role	Rating
	Minimal Life Size	2
	Minimal Size for Split	4

After the data is input and run by Rapid Miner V10, with the parameters in Table 5, the results of applying Data Mining in Rapid Miner are displayed in the Decision Tree below.

Decision Tree

The level of accuracy of the Decision Tree data is as follows :



accuracy: 80.00% +/- 25.82% (micro average: 78.95%)

	true Sangat Baik	true Luar Biasa	class precision
pred. Sangat Baik	3	2	60.00%
pred. Luar Biasa	2	12	85.71%
class recall	60.00%	85.71%	

Image 6. Decision Tree Data Accuracy

With data accuracy of up to 80.00%, it can be used as a reference for research where the data is more than 75%. Then, the results of the decision tree and the most influential factors.

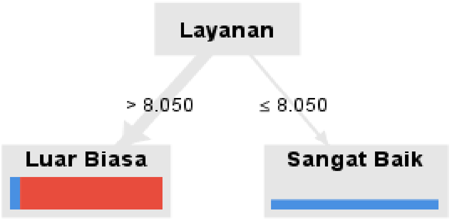


Image 7. The result of the Decision Tree Algorithm

On the results in Figure 7, the service is the most influential factor in determining the ho-tel ratings in Bali. The description of the Decision Tree reading is as follows:

Layanan > 8.050: Luar biasa
{Sangat Baik=1, Luar Biasa=14}
Layanan ≤ 8.050: Sangat Baik
{Sangat Baik=4, Luar Biasa=0}

CONCLUSION

Based on results and considerations, it can be concluded that for hotels in Bali, based on data from the Agoda website, the decision tree or factor that has the most significant influence is in terms of service, which means excellent service. The advantage of this research is the implementation of two different algorithms to identify the correct problem and serve as a reference for companies or hotel managers to improve hotel ratings so that they can evaluate each factor that can improve hotel rankings.

BIBLIOGRAPHY

[1] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[2] M. P. Utami and H. Hardianti, "Rancangan Destination Branding Kabupaten Wajo: Strategi Membangun Citra Destinasi Wisata Untuk Peningkatan Daya Tarik Pariwisata," *Jurnal Pariwisata Tawangmangu*, vol. 2, no. 1, pp. 17–27, 2024.

[3] T. Hidayat and B. E. Putro, "Analisis Karakteristik Konsumen Hotel 'X' dengan Menggunakan Metode K-Means Clustering," *Jurnal Media Teknik dan Sistem Industri*, vol. 4, no. 2, p. 53, Sep. 2020, doi: 10.35194/jmtsi.v4i2.995.

[4] Y. Lee and D. Y. Kim, "The decision tree for longer-stay hotel guest: the relationship between hotel booking determinants and geographical distance," *International Journal of Contemporary Hospitality Management*, vol. 33, no. 6, pp. 2264–2282, 2020, doi: 10.1108/IJCHM-06-2020-0594.

[5] A. W. H. Fasa, D. Andriani, I. Haribudiman, and M. Berliandaldo, "Analisis Strategi Pengembangan Smart Destinations: Perspektif Service-Dominant Logic," *Altasia Jurnal Pariwisata Indonesia*, vol. 5, no. 2, pp. 76–91, 2023.

[6] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI

[7] S. S. Ningsih, A. Fauzi, and M. A. Syari, "Pendeteksi Mata Kantuk Pada Pengendara Mobil dan Sepeda Motor Menggunakan Metode Backpropagation," *Pelita Informatika: Informasi dan Informatika*, vol. 11, no. 1, 2022.

[8] F. A. N. Limantara, I. A. Kalpikawati, N. K. J. Rastitiati, and I. G. N. A. Suprastayasa, "Strategi Meningkatkan Kinerja Karyawan Reservasi di Hotel XYZ Nusa Dua, Bali," *Jurnal Ilmiah Pariwisata*, vol. 27, no. 3, pp. 264–276, 2022.

[9] N. Indriyani, H. S. Tambunan, and Z. A. Siregar, "Analisis Faktor Kepuasan Konsumen Terhadap Produk Roti Pinkan Bakery & Cake dengan Algoritma C4. 5," *JURAL RISET RUMPUN ILMU TEKNIK*, vol. 1, no. 2, pp. 76–90, 2022.

[10] F. Salsabila and S. M. Intani, "Implementasi Algoritma K-Means Dan C4. 5 Dalam Menentukan Tingkat Penyebaran Covid-19 Di Indonesia," *Jurnal Siliwangi Seri Sains dan Teknologi*, vol. 7, no. 1, 2021.

[11] S. Sinaga and A. Mahmud Husein, "Penerapan Algoritma Apriori dalam Data Mining untuk Memprediksi Pola Pengunjung pada Objek Wisata Kabupaten Karo," 2019.

[12] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan," vol. 15, no. 2, 2021.

[13] I. Pii, N. Suarna, and N. Rahaningsih, "Penerapan Data Mining Pada Penjualan Produk Pakaian Dameyra Fashion Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 423–430, 2023.

[14] Q. I. Mawarni and E. S. Budi, "Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 3, no. 4, pp. 522–528, 2022.

[15] E. Sutinah *et al.*, "Data Mining Untuk Klasifikasi Tamu Hotel Dengan Algoritma Apriori," 2019.

[16] B. A. Utami and A. Kafabih, "Sektor Pariwisata Indonesia Di Tengah Pandemi Covid 19," *Jurnal Dinamika Ekonomi Pembangunan*, vol. 4, no. 1, pp. 383–389, Jan. 2021, doi: 10.33005/jdep.v4i1.198.

[17] I. Utnasari, "Analisis Clustering Dengan K-Means Untuk Pengelompokkan Penjualan Produk Pada Hotel Newton," 2021.

[18] A. Wahyuni and S. Anggraini, "Implementasi Algoritma J48 Data Mining Untuk Inovasi Bisinis Perhotelan Di Masa Pandemi Covid-19," *Jurnal Ilmu Komputer dan Bisnis*, vol. 13, no. 1, pp. 182–192, May 2022, doi: 10.47927/jikb.v13i1.302.

[19] S. Wang, J. Cao, and P. S. Yu, "Deep Learning for Spatio-Temporal Data Mining: A Survey," Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.04928

[20] Y. Widayanti, "Meningkatkan hasil belajar peserta didik dengan modul pembelajaran berbasis problem based learning (PBL)," *Jurnal Pendidikan Ekonomi Undiksha*, vol. 12, no. 1, pp. 166–174, 2020.