SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Application of Data Mining using the K-Means Method for Visitor Grouping

Rahmayuni Syah[1)*], Marnis Nasution[2)], Irmayanti[3)]
[1,2,3)]Universitas Labuhanbatu, Indonesia
[1)] Rahmayunisyah20@gmail.com, [2)] marnisnst@gmail.com, [3)] irmayantiritonga2@gmail.com,

Submitted :April 4, 2024 | **Accepted** : April 15, 2024 | **Published** : April 22, 2024

Abstract: Grouping amusement ride visitor data is an important process that aims to identify certain patterns of visitors, enabling management to adjust marketing strategies and improve their services more effectively. This process begins with a data selection stage where relevant visitor data is collected and prepared for analysis. The next stage is data pre-processing, which involves cleaning the data from noise or irrelevant data, as well as ensuring the data is in a format suitable for analysis. After that, the data mining model design is carried out by selecting the most appropriate method for grouping visitor data. The next stage is testing and evaluating the model to verify its accuracy and effectiveness. The results of model testing show that visitor data can be categorized into three groups: C1 with 50 data, C2 with 20 data, and C3 with 48 data. The results of the model evaluation confirm that the designed model succeeded in classifying data with perfect accuracy, namely 100%. This success shows that the model is highly effective in identifying and segmenting visitor patterns, providing valuable insights for strategic decision making in service improvement and marketing. This success also opens up opportunities for the application of similar methods to other datasets in an effort to improve visitor experience and operational efficiency.

Keywords: Clustering; Confusion Matrix; Data Mining; K-Means; Scatter Plot

INTRODUCTION

Playgrounds are places where children and adults can enjoy various types of games and activities that are exciting and full of challenges. Usually located in amusement parks or recreation centers, rides offer a variety of attractions, from classic games such as swings and slides, to modern and sophisticated rides such as roller coasters, haunted houses and simulators. Each ride is designed to provide different sensations and experiences, with the main aim to entertain and stimulate visitors' adrenaline. Apart from that, playgrounds also often provide supporting facilities such as eating areas, souvenir shops, and rest zones, making them the perfect destination for families and friends who want to spend time together while enjoying the fun and excitement. On every playground, the presence of visitors is an important element that gives life and dynamics to the ride. Visitors come from various backgrounds and ages, bringing their enthusiasm and excitement to try the various attractions on offer. Interaction between visitors, whether it's excitement while waiting for their turn, screams full of adrenaline while enjoying the rides, or laughter after completing a game, adds to the lively and lively atmosphere in an amusement park. Their presence not only fills the rides with diverse stories and experiences, but also encourages managers to continue to innovate and maintain the rides so that they remain attractive and safe for everyone. In this way, visitors and playgrounds provide each other with value and unforgettable memories, creating a positive circle of entertainment and satisfaction.

In the world of amusement parks, visitor safety and security should be a top priority. However, sometimes there are still errors that visitors and ride owners may not immediately realize. Each playground is designed to specific specifications targeting visitor demographics based on age, gender and height to ensure their safety and comfort. For example, rides designed for toddlers typically have lower speeds, shorter routes, and fewer obstacles than rides for older children or adults. Unfortunately, sometimes errors occur when using this vehicle. For example, rides that should be specifically for toddlers are sometimes also used by children over 5 years old. This can be problematic because bigger, heavier children can accidentally cause damage to the ride or even harm themselves or other smaller children. In addition, overloading or use that does not comply with regulations can reduce the effectiveness of the ride's safety system, thereby increasing the risk of accidents.

Every ride in an amusement park is designed taking into account various safety aspects, including the age categories of visitors who are allowed to use them. This is done to ensure that every visitor, especially children, can enjoy the rides safely without risk of injury. These age categories are not just guidelines, but are the result of

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

rigorous research and testing to assess the safety of rides for various age groups. Age restrictions on rides are important because younger children may not have enough physical strength, height, or mental maturity to handle the challenges presented by some attractions. For example, rides that move quickly or that involve significant heights may not be safe for young children. In addition, the safety systems on these rides are often designed to protect bodies that have reached a certain size and weight, so children who are too small may not be properly protected in the event of an accident. Therefore, it is very important for theme park managers to strictly enforce and monitor these age restrictions. Parents and caregivers must also pay attention to and comply with these rules for the safety of their children. Educating visitors about the importance of these restrictions is also an important part of safety efforts, to avoid unwanted incidents. Proper management of these rules not only ensures visitor safety but also improves the overall experience at a theme park, making it a fun and safe place for all visitors, regardless of their age.

Therefore, it is very important to group or cluster visitors based on age, gender and weight for each playground. This clustering ensures that every child using the rides is in the appropriate category, thereby increasing safety and reducing the risk of injury. By using data mining methods, amusement park managers can effectively identify and group visitors into homogeneous groups. This process not only makes it easier to enforce rules relating to age restrictions and physical specifications, but also helps in customizing a safer and more enjoyable gaming experience for each visitor, based on their physical characteristics and needs. This is a strategic move in theme park management aimed at maximizing visitor satisfaction while maintaining high safety standards. Implementing this clustering not only improves security but also allows theme parks to offer a more personalized experience. For example, by grouping together children who have similar physical characteristics, amusement parks can optimize the setup and operation of rides to better suit certain age or weight groups. This means that each ride can be customized to provide the right level of excitement and ensure that the experience is safe and enjoyable for each group of visitors. In addition, with more structured and organized data, managers can more easily identify the need for adding or modifying new rides, based on visitor preferences and safety, making amusement parks more innovative and responsive to the needs of their visitors.

In an effort to improve security and safety at the Happy Kiddy Playground, as well as to provide a more personal and enjoyable experience for each visitor, the author wishes to conduct research which aims to group visitors based on certain characteristics. The method chosen for this research is K-Means, which is a popular algorithm in the field of Data Mining for grouping or clustering. The K-Means algorithm works by dividing data into a number of groups (k clusters) based on certain features, by optimizing the distance between data points in one group and maximizing the distance between groups. This research is expected to identify certain patterns in the preferences and characteristics of visitors to the Happy Kiddy Playground. By using the data that has been collected, such as age, gender, height, and the most frequently visited rides, the author plans to group visitors into several clusters that have certain characteristics. For example, clusters can be formed based on age groups, with the aim of better understanding what types of rides are most suitable and safe for each age group. This will not only increase visitor safety but can also help ride managers provide attractions that are more appropriate and attractive for each group of visitors.

By applying the K-Means Method to Data Mining, this research also aims to provide recommendations to Happy Kiddy Playground managers on how to optimize resource distribution and promotions for the most relevant demographic targets. For example, if it is identified that a certain cluster tends to visit more adventure rides, then managers can increase investment in that type of ride or develop promotions targeted at the appropriate demographic group.

LITERATURE REVIEW

Data mining is an analytical process designed to explore data (usually big data containing previously unknown patterns, relationships or associations) on a large scale (Abas et al., 2023) (Saputra, Hindarto, & Haryono, 2023) (Bustomi, Nugraha, Juliane, & Rahayu, 2023). The main goal is to extract useful information from large data sets and convert it into an understandable structure for subsequent use. This technique involves methods from statistics, machine learning, and database systems (Aji & Devi, 2023) (S. A. Hasibuan, Sihombing, & Nasution, 2023). By using data mining, businesses and organizations can make more informed decisions based on trends and patterns found in their data, such as customer behavior predictions, market analysis, and risk management (Pratama, Yanris, Nirmala, & Hasibuan, 2023) (Sinaga, Marpaung, Tarigan, & Tania, 2023). In practice, data mining is often used to identify unexpected relationships and predict future trends, which can provide a competitive advantage for a business or organization. For example, in the retail industry, data mining can be used for shopping cart analysis to identify combinations of products that are frequently purchased together by consumers. In the financial sector, this technique can help in fraud detection by identifying unusual transaction patterns. The data mining approach involves various stages, including data selection and pre-processing, model selection, model evaluation, and interpretation of results, all aimed at gaining accessible insights from the data.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

The K-Means method is one of the most popular and widely used clustering techniques in data mining and machine learning, which aims to partition n observations into k clusters where each observation is included in the cluster with the closest mean, thereby minimizing the variance in the cluster (Aldo, 2023) (Andi, Juliandy, & David, 2023) (Wijaya, Dharma, Heyneker, & Vanness, 2023). This technique begins by randomly selecting k points, known as centroids, as starting points for each cluster, and then repeats the step of assigning each observation to the closest cluster based on the Euclidean distance and updating the cluster centroids until the stopping criterion is met (Indah, Sari, & Dar, 2023) (Asriningtias, Wulandari, Persijn, Rosyida, & Sutawijaya, 2023). K-Means is used to analyze unlabeled data and is often applied in various practical applications such as market segmentation, document clustering, and image analysis, due to its efficient ability to handle large data sets. The stages that can be used in this research are as follows.

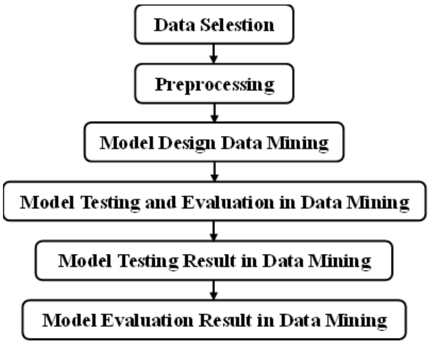

Fig 1. K-Means Model Framework

An explanation of each stage is as follows.

Data Selection : Data selection involves gathering and preparing the relevant visitor data for analysis, which is a critical first step in the process of segmenting visitors to the Happy Kiddy amusement park for targeted improvements and marketing strategies.

Preprocessing : Preprocessing involves cleaning and organizing the data to ensure it is free from inconsistencies and ready for further analysis.

Model Design Data Mining : The design of the data mining model for clustering aims to group data into several groups or clusters based on similar characteristics, so that more structured information can be produced from large and complex datasets.

Model Testing and Evaluation in Data Mining : Model Testing Results in Data Mining The results of testing clustering models in data mining show the model's ability to accurately group data into homogeneous groups, based on similar characteristics.

Model Evaluation Result in Data Mining : The evaluation results of cluster models in data mining show the level of accuracy and efficiency of the model in grouping data into relevant clusters, enabling more precise decision making based on identified patterns and trends.

In the context of clustering, a confusion matrix can be used to measure how well a model differentiates and groups data into the correct clusters, by comparing the clusters predicted by the model to the actual clusters assigned via external methods (Mawaddah, Dar, & Yanris, 2023).

Table 1. Confusion Matrix

		Prediction Class	
	Class	**True**	**False**
Attribute Class	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Where table 1 contains:

1. TP (True Positive), namely the amount of positive data that has a true value.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

2. TN (True Negative), namely the amount of negative data that has a true value.
3. FN (False Negative), namely the amount of negative data but which has the wrong value.
4. FP (False Positive), namely the amount of data that is positive but has the wrong value.

Accuracy for clustering measures how closely the clusters formed by the model correspond to the original divisions or groups of the data, reflecting the effectiveness of the model in grouping similar data (Siregar, Irmayani, & Sari, 2023). Accuracy Value Formula:

$$\boldsymbol{Acuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FN+FP}} \times \mathbf{100\%}$$

Precision in the context of clustering measures the proportion of data that is classified into a cluster correctly compared to the total data that is classified into that cluster (Sari, Yanris, & Hasibuan, 2023). precision can be obtained using the formula:

$$\boldsymbol{Presisi} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \times \mathbf{100\%}$$

Recall is used to indicate the number of perpetrators from positive category data that were correctly classified (F. F. Hasibuan, Dar, & Yanris, 2023). To find the Recall value, use the formula:

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \times \mathbf{100\%}$$

RESULT

Data Selection

This stage is the stage carried out to obtain data. The data that has been obtained will later be used as research sample data. The data that has been obtained is as follows.

Table 2. Research Sample Data

Visitor Name	Age	Gender	Height	Category
Adi Baskara Putra	6	Man	125	Children
Aisyah Putri Damar	9	Woman	132	Children
Akbar Rizky Maulana	8	Man	132	Children
Aldo Bima Seno	7	Man	128	Children
Alifia Cinta Dewi	5	Woman	117	Toddler
Bagus Rayhan Aditya	9	Man	132	Children
Bella Sari Dewi	8	Woman	130	Children
Bilal Aditya Rahman	8	Man	131	Children
Bima Dwi Cahya	7	Man	128	Children
Bintang Rizki Fauzan	7	Man	128	Children
Cahya Putri Ramadani	8	Woman	130	Children
Cakra Wijaya Putra	8	Man	130	Children
Cinta Ayu Lestari	8	Woman	131	Children
Citra Kirana Sari	7	Woman	128	Children
Clara Putri Ayudia	9	Woman	134	Children
Daffa Rizaldi Putra	8	Man	130	Children
Dara Intan Permata	6	Woman	123	Children
Dewa Gita Asmara	8	Woman	130	Children
Dika Jaya Purnama	6	Man	123	Children
Dina Ayu Safitri	9	Woman	132	Children
Edo Baskoro Yudhoyono	6	Man	123	Children
Eka Putra Bramantyo	8	Man	130	Children
Elang Rajasa Bayu	7	Man	127	Children
Elisa Nur Aisyah	7	Woman	129	Children
Elsa Laila Sari	9	Woman	132	Children
Fadil Jaya Kurniawan	7	Man	126	Children

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Fajar Nugraha Putra	9	Man	132	Children
Farah Dian Sari	8	Woman	129	Children
Fikri Jaya Setiawan	6	Man	123	Children
Fitri Hana Sari	7	Woman	128	Children
Galih Arjuna Wibowo	9	Man	140	Children
Ghina Fadhila Yani	7	Woman	128	Children
Gilang Ramadhan Putra	7	Man	128	Children
Gina Ayu Kusuma	6	Woman	123	Children
Giska Putri Illahi	7	Woman	128	Children
Hadi Surya Baskoro	7	Man	128	Children
Hana Mei Lestari	4	Woman	100	Toddler
Hanif Rahmat Pratama	7	Man	128	Children
Haris Ahmad Syukur	8	Man	128	Children
Hesti Arum Sari	7	Woman	128	Children
Ika Melati Indah	7	Woman	128	Children
Ikbal Fikri Hidayat	9	Man	133	Children
Indra Bima Yudha	6	Man	123	Children
Intan Permata Suci	7	Woman	128	Children
Irfan Zaky Maulana	8	Man	130	Children
Jaya Kurnia Dewanto	7	Man	128	Children
Jihan Ayu Pertiwi	9	Woman	132	Children
Jihan Putri Maharani	7	Woman	127	Children
Joko Tri Hariyanto	8	Man	130	Children
Julita Rani Kumala	5	Woman	117	Toddler
Kafi Adi Pratama	4	Man	103	Toddler
Kamilia Putri Indah	5	Woman	117	Toddler
Kania Tiara Putri	8	Woman	131	Children
Keisha Putri Amalia	8	Woman	130	Children
Kirana Dewi Salsabila	9	Woman	132	Children
Laila Cinta Damayanti	4	Woman	103	Toddler
Liana Dewi Sukmawati	3	Woman	93	Toddler
Lintang Senja Pratama	4	Man	102	Toddler
Lutfi Aziz Karim	5	Man	117	Toddler
Luthfi Ahmad Syahputra	9	Man	132	Children
Mahesa Jaya Kusuma	5	Man	115	Toddler
Malik Arjuna Putra	5	Man	117	Toddler
Maudy Ayu Kusumawati	3	Woman	94	Toddler
Mika Anggun Permata	5	Woman	117	Toddler
Mirza Akbar Fauzi	4	Man	103	Toddler
Nadia Putri Zahra	3	Woman	88	Toddler
Nanda Rizki Pratama	4	Man	103	Toddler
Nia Sari Wulandari	6	Woman	126	Children
Niko Aditya Putra	5	Man	117	Toddler
Nila Sari Dewi	3	Woman	87	Toddler
Oka Mahendra Putra	2	Man	75	Toddler
Olivia Putri Hartanti	3	Woman	89	Toddler
Olla Ramlan Sari	6	Woman	123	Children
Opan Jaya Kusuma	6	Man	124	Children
Ovi Dwi Astuti	5	Woman	117	Toddler
Prita Kinanti Ayu	2	Woman	76	Toddler
Puspa Indah Permata	8	Woman	129	Children
Putra Mahkota Rizal	5	Man	117	Toddler
Putri Ayu Andini	6	Woman	123	Children
Putu Bagus Santoso	6	Man	123	Children
Qaisar Alif Ramadhan	6	Man	122	Children
Qiara Zafira Husna	6	Woman	123	Children
Qila Zahra Putri	2	Woman	75	Toddler

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Qisthi Aulia Rahman	8	Woman	130	Children
Qonita Laila Kurniawan	6	Woman	123	Children
Rafi Alifian Putra	6	Man	121	Children
Rama Aditya Pratama	2	Man	78	Toddler
Rian Dwi Saputra	6	Man	123	Children
Ridho Bayu Aji	7	Man	125	Children
Rizki Fajar Nugroho	8	Man	130	Children
Salsa Bintang Pratiwi	6	Woman	123	Children
Sari Nindya Ayu	7	Woman	128	Children
Sasa Tiara Indah	8	Woman	130	Children
Sinta Bella Ratri	2	Woman	77	Toddler
Tara Bintang Alam	2	Man	76	Toddler
Tasya Kamila Putri	7	Woman	128	Children
Tegar Bayu Pratama	6	Man	123	Children
Tiara Putri Ananda	7	Woman	128	Children
Ujang Budi Setiawan	7	Man	128	Children
Uki Dwi Saputra	8	Man	135	Children
Umar Zaki Nugroho	2	Man	79	Toddler
Umi Salma Farida	7	Woman	128	Children
Vania Kirana Dewi	6	Woman	123	Children
Vega Arjuna Wisnu	7	Man	128	Children
Vino Bastian Harahap	8	Man	130	Children
Vira Ayu Kirana	3	Woman	86	Toddler
Wafi Aditya Rahman	6	Man	123	Children
Winda Ayu Larasati	6	Woman	123	Children
Wira Yudha Pratama	8	Man	130	Children
Wulan Dari Rahayu	4	Woman	103	Toddler
Yasmin Nadira Putri	9	Woman	132	Children
Yasmine Putri Zahra	6	Woman	123	Children
Yogi Pratama Akbar	6	Man	123	Children
Yuda Pratama Surya	6	Man	123	Children
Zahra Nabila Putri	6	Woman	123	Children
Zaki Ahmad Fauzi	9	Man	132	Children
Zara Putri Anjani	7	Woman	128	Children
Zidan Ali Marwan	7	Man	126	Children

In the table above is the data that has been obtained, the data above is data on visitors to the Happy Kiddy Playground. The description attribute is a grouping target which will later be analyzed using the K-Means method.

Preprocessing

At the preprocessing stage, the data will be analyzed and cleaned if there is data that is not suitable for use. At this stage the data will also be arranged in a good format and form so that the data can be used in the clustering process using the K-Means method.

Model Planning on Data Mining

This stage is the design stage of the model created in Data Mining. The software that will be used is Orange Software. This design process uses the K-Means method.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

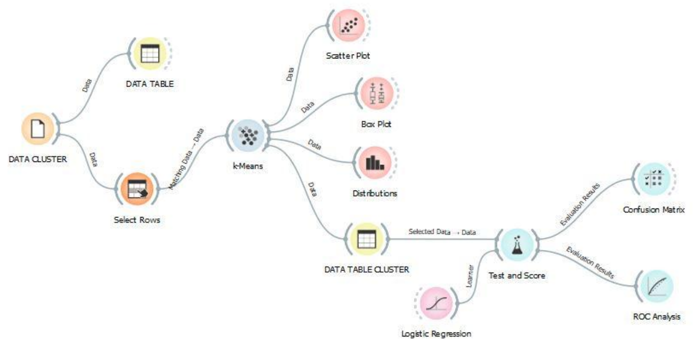

Fig 2. Model Data Mining

In the image above is a model designed using the K-Means method in Data Mining. The model above will later be used to carry out Cluster testing of the sample data that has been obtained and the model above will also later be used to evaluate the method used.

Model Testing and Evaluation in Data Mining

In the testing and evaluation process, the author uses a previously designed model, specifically by applying the K-Means method, a clustering technique that is popular in data mining. The main objective of this test is to group data from visitors to the Happy Kiddy Playground into several clusters based on certain predetermined characteristics, such as age, gender and height. Through the K-Means method, the author can identify certain patterns in visitor data, which will then help in optimizing marketing strategies, adjusting facilities, and improving the overall visitor experience. This testing and evaluation process is key in understanding visitor needs and preferences, enabling Happy Kiddy Playground to provide a more personalized and satisfying service.

Model Testing Results in Data Mining

The cluster data results for visitors to the Happy Kiddy ride can be seen in the table below.

Table 3. Cluster Results Data

Visitor Name	Age	Gender	Height	Category	Cluster
Adi Baskara Putra	6	Man	125	Children	C1
Aisyah Putri Damar	9	Woman	132	Children	C3
Akbar Rizky Maulana	8	Man	132	Children	C1
Aldo Bima Seno	7	Man	128	Children	C1
Alifia Cinta Dewi	5	Woman	117	Toddler	C3
Bagus Rayhan Aditya	9	Man	132	Children	C1
Bella Sari Dewi	8	Woman	130	Children	C3
Bilal Aditya Rahman	8	Man	131	Children	C1
Bima Dwi Cahya	7	Man	128	Children	C1
Bintang Rizki Fauzan	7	Man	128	Children	C1
Cahya Putri Ramadani	8	Woman	130	Children	C3
Cakra Wijaya Putra	8	Man	130	Children	C1
Cinta Ayu Lestari	8	Woman	131	Children	C3
Citra Kirana Sari	7	Woman	128	Children	C3
Clara Putri Ayudia	9	Woman	134	Children	C3
Daffa Rizaldi Putra	8	Man	130	Children	C1
Dara Intan Permata	6	Woman	123	Children	C3
Dewa Gita Asmara	8	Woman	130	Children	C3
Dika Jaya Purnama	6	Man	123	Children	C1
Dina Ayu Safitri	9	Woman	132	Children	C3
Edo Baskoro Yudhoyono	6	Man	123	Children	C1
Eka Putra Bramantyo	8	Man	130	Children	C1
Elang Rajasa Bayu	7	Man	127	Children	C1
Elisa Nur Aisyah	7	Woman	129	Children	C3
Elsa Laila Sari	9	Woman	132	Children	C3
Fadil Jaya Kurniawan	7	Man	126	Children	C1

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.13624 p-ISSN : 2541-044X

Fajar Nugraha Putra	9	Man	132	Children	C1
Farah Dian Sari	8	Woman	129	Children	C3
Fikri Jaya Setiawan	6	Man	123	Children	C1
Fitri Hana Sari	7	Woman	128	Children	C3
Galih Arjuna Wibowo	9	Man	140	Children	C1
Ghina Fadhila Yani	7	Woman	128	Children	C3
Gilang Ramadhan Putra	7	Man	128	Children	C1
Gina Ayu Kusuma	6	Woman	123	Children	C3
Giska Putri Illahi	7	Woman	128	Children	C3
Hadi Surya Baskoro	7	Man	128	Children	C1
Hana Mei Lestari	4	Woman	100	Toddler	C2
Hanif Rahmat Pratama	7	Man	128	Children	C1
Haris Ahmad Syukur	8	Man	128	Children	C1
Hesti Arum Sari	7	Woman	128	Children	C3
Ika Melati Indah	7	Woman	128	Children	C3
Ikbal Fikri Hidayat	9	Man	133	Children	C1
Indra Bima Yudha	6	Man	123	Children	C1
Intan Permata Suci	7	Woman	128	Children	C3
Irfan Zaky Maulana	8	Man	130	Children	C1
Jaya Kurnia Dewanto	7	Man	128	Children	C1
Jihan Ayu Pertiwi	9	Woman	132	Children	C3
Jihan Putri Maharani	7	Woman	127	Children	C3
Joko Tri Hariyanto	8	Man	130	Children	C1
Julita Rani Kumala	5	Woman	117	Toddler	C3
Kafi Adi Pratama	4	Man	103	Toddler	C2
Kamilia Putri Indah	5	Woman	117	Toddler	C3
Kania Tiara Putri	8	Woman	131	Children	C3
Keisha Putri Amalia	8	Woman	130	Children	C3
Kirana Dewi Salsabila	9	Woman	132	Children	C3
Laila Cinta Damayanti	4	Woman	103	Toddler	C2
Liana Dewi Sukmawati	3	Woman	93	Toddler	C2
Lintang Senja Pratama	4	Man	102	Toddler	C2
Lutfi Aziz Karim	5	Man	117	Toddler	C1
Luthfi Ahmad Syahputra	9	Man	132	Children	C1
Mahesa Jaya Kusuma	5	Man	115	Toddler	C1
Malik Arjuna Putra	5	Man	117	Toddler	C1
Maudy Ayu Kusumawati	3	Woman	94	Toddler	C2
Mika Anggun Permata	5	Woman	117	Toddler	C3
Mirza Akbar Fauzi	4	Man	103	Toddler	C2
Nadia Putri Zahra	3	Woman	88	Toddler	C2
Nanda Rizki Pratama	4	Man	103	Toddler	C2
Nia Sari Wulandari	6	Woman	126	Children	C3
Niko Aditya Putra	5	Man	117	Toddler	C1
Nila Sari Dewi	3	Woman	87	Toddler	C2
Oka Mahendra Putra	2	Man	75	Toddler	C2
Olivia Putri Hartanti	3	Woman	89	Toddler	C2
Olla Ramlan Sari	6	Woman	123	Children	C3
Opan Jaya Kusuma	6	Man	124	Children	C1
Ovi Dwi Astuti	5	Woman	117	Toddler	C3
Prita Kinanti Ayu	2	Woman	76	Toddler	C2
Puspa Indah Permata	8	Woman	129	Children	C3
Putra Mahkota Rizal	5	Man	117	Toddler	C1
Putri Ayu Andini	6	Woman	123	Children	C3
Putu Bagus Santoso	6	Man	123	Children	C1
Qaisar Alif Ramadhan	6	Man	122	Children	C1
Qiara Zafira Husna	6	Woman	123	Children	C3
Qila Zahra Putri	2	Woman	75	Toddler	C2

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Qisthi Aulia Rahman	8	Woman	130	Children	C3
Qonita Laila Kurniawan	6	Woman	123	Children	C3
Rafi Alifian Putra	6	Man	121	Children	C1
Rama Aditya Pratama	2	Man	78	Toddler	C2
Rian Dwi Saputra	6	Man	123	Children	C1
Ridho Bayu Aji	7	Man	125	Children	C1
Rizki Fajar Nugroho	8	Man	130	Children	C1
Salsa Bintang Pratiwi	6	Woman	123	Children	C3
Sari Nindya Ayu	7	Woman	128	Children	C3
Sasa Tiara Indah	8	Woman	130	Children	C3
Sinta Bella Ratri	2	Woman	77	Toddler	C2
Tara Bintang Alam	2	Man	76	Toddler	C2
Tasya Kamila Putri	7	Woman	128	Children	C3
Tegar Bayu Pratama	6	Man	123	Children	C1
Tiara Putri Ananda	7	Woman	128	Children	C3
Ujang Budi Setiawan	7	Man	128	Children	C1
Uki Dwi Saputra	8	Man	135	Children	C1
Umar Zaki Nugroho	2	Man	79	Toddler	C2
Umi Salma Farida	7	Woman	128	Children	C3
Vania Kirana Dewi	6	Woman	123	Children	C3
Vega Arjuna Wisnu	7	Man	128	Children	C1
Vino Bastian Harahap	8	Man	130	Children	C1
Vira Ayu Kirana	3	Woman	86	Toddler	C2
Wafi Aditya Rahman	6	Man	123	Children	C1
Winda Ayu Larasati	6	Woman	123	Children	C3
Wira Yudha Pratama	8	Man	130	Children	C1
Wulan Dari Rahayu	4	Woman	103	Toddler	C2
Yasmin Nadira Putri	9	Woman	132	Children	C3
Yasmine Putri Zahra	6	Woman	123	Children	C3
Yogi Pratama Akbar	6	Man	123	Children	C1
Yuda Pratama Surya	6	Man	123	Children	C1
Zahra Nabila Putri	6	Woman	123	Children	C3
Zaki Ahmad Fauzi	9	Man	132	Children	C1
Zara Putri Anjani	7	Woman	128	Children	C3
Zidan Ali Marwan	7	Man	126	Children	C1

In the table above are the cluster results in mining using the k-means method. For the cluster results used, the author used 3 clusters in this research, namely C1, C2 and C3. Of the 118 data used, the results obtained for C1 were 50 data, for C2 the results obtained were 20 data and for C3 the results obtained were 48. Not only cluster results, the author also added several more results such as scatter plot results, boxplot results and distribution. The results will be presented in image form as follows.

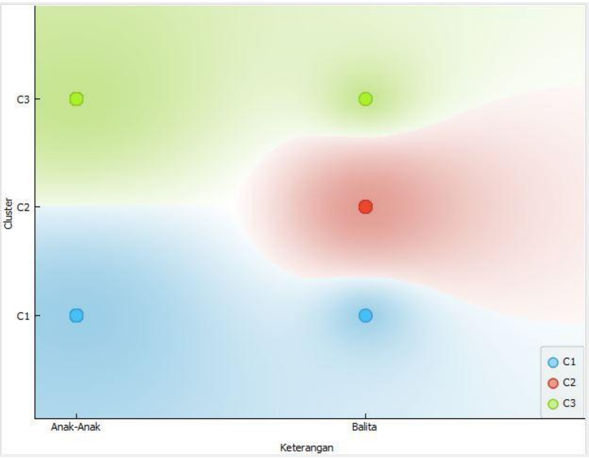

Fig 3. Scatter Plot Results

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

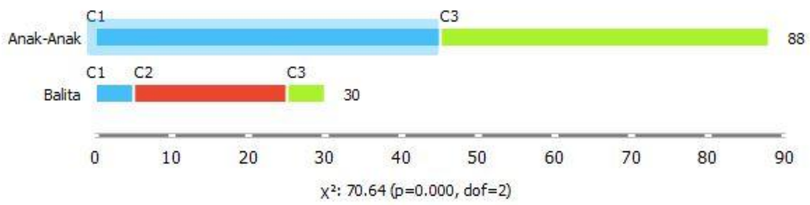

Fig 4. Box Plot Result

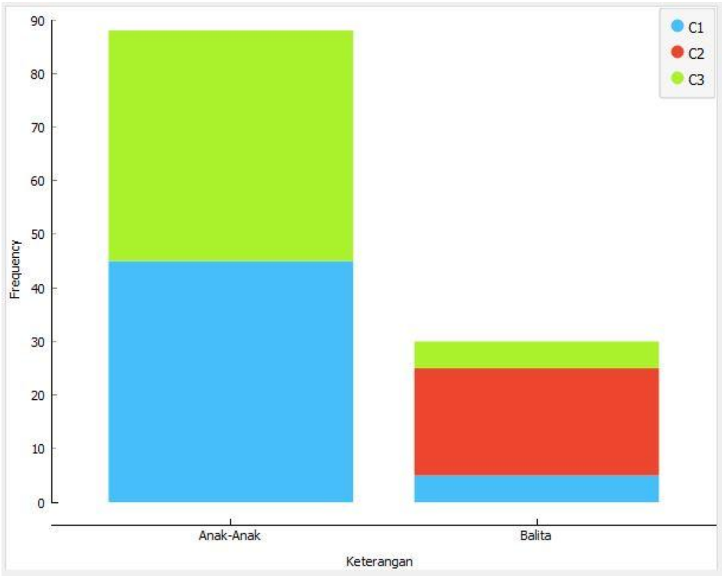

Fig 5. Distribution results

4.1. Model Evaluation Results in Data Mining

For the evaluation results, the author uses the evaluation on the Test and Score widget, Confusion Matrix and for the graphic results the author uses ROC Analysis.

Result of Test and Score

Table 4. Result of Test and Score

Model	AUC	CA	F1	Precision	Recall
K-Means	1.000	1.000	1.000	1.000	1.000

The results obtained from the test and score widget are perfect, this is because the AUC results are 1,000, the Ca results are 1,000, the F1 results are 1,000, the Precision results are 1,000 and the Recall results are 1,000. This result was declared perfect because the highest result from the test and score widget was 1,000. At this stage, the choice used by the author is Cross Validation with a Number of Folds of 5.

Confusion Matrix Results

Confusion Matrix is a table used to measure model performance, where it visualizes the comparison between predicted values and actual values. The results of this confusion matrix are to test the model using the K-means method which is used to cluster data in data mining.

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 5. Result of Confusion Matrix

		Predicted		
		Children	Toddler	∑
Actual	Children	88	0	88
	Toddler	0	30	30
	∑	88	30	118

In the table above are the results of the Confusion Matrix, namely the True Positive (TP) result is 88. True Negative (TN) is 30, False Positive (FP) is 0 and False Negative (FN) is 0. So the accuracy, precision and recall values are as follows:

$$\boldsymbol{Accuracy = \frac{88+30}{88+30+0+0} + 100\%}$$ Then the Accuracy value = 100%

$$\boldsymbol{Presisi = \frac{88}{88+0} + 100\%}$$ Then the Precision value = 100%

$$\boldsymbol{Recall = \frac{88}{88+0} + 100\%}$$ Then the Recall value = 100%

ROC Analysis Results

ROC Analysis (Receiver Operating Characteristic Analysis) is a statistical method used to measure the effectiveness of a cluster model, by visualizing the relationship between the true positive rate (True Positive Rate) and the false positive rate (False Positive Rate) at various cluster thresholds.

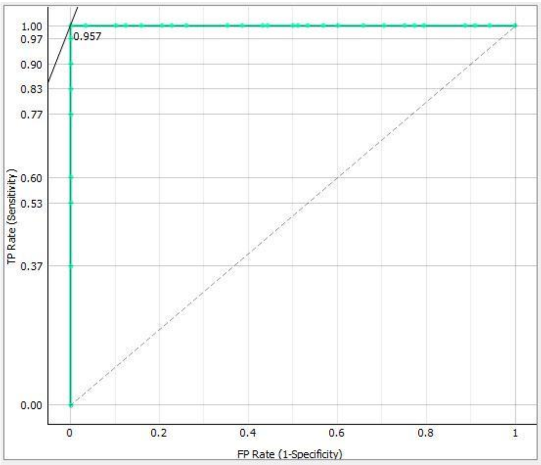

Fig 6. ROC Toddler Analysis

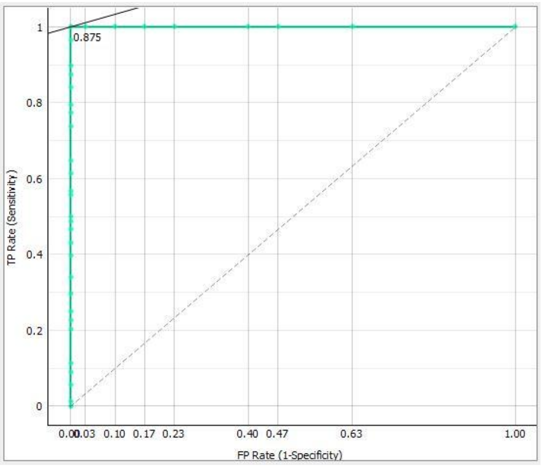

Fig 7. ROC Analysis of Children

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

DISCUSSIONS

In research conducted on grouping visitors to the Happy Kiddy Playground using the K-Means method, the author faced several obstacles. One of the main obstacles faced is determining the widgets that will be used in the data mining design model. Choosing the right widget is very crucial because it directly affects the ease of analysis and the accuracy of the results obtained. After going through a careful selection process, the author finally decided on the widget that was most suitable for the needs of this research. From the results of the data grouping carried out, three clusters were identified, namely C1 with a total of 50 data, C2 with a total of 20 data, and C3 with a total of 48 data. This cluster division provides valuable insight into the characteristics of visitors to the Happy Playground. Kiddy, allows the author to better understand the distribution of preferences and behavior of visitors in the playground.

Regarding accuracy results, this study shows interesting results. A comparison of accuracy between the 'Test and Score' widget and the 'Confusion Matrix' widget obtained a ratio of 1:1, where both managed to achieve 100% accuracy. This shows that the two widgets are very effective in predicting and clustering visitor data according to predetermined clusters. The success of achieving 100% accuracy is a strong indicator that the K-Means method, together with selecting the right widget, can be relied on to analyze and group data with a very high level of accuracy. These results not only confirm the effectiveness of the approach taken in this research but also provide an important contribution to similar studies in the future that may face similar obstacles in the selection of data analysis tools.

CONCLUSION

This research makes a significant contribution to the understanding of visitor behavior at the Happy Kiddy Playground, by identifying groups of visitors based on their preferences and characteristics. The success of obtaining 100% accuracy in classification results indicates that data mining techniques, especially the K-Means method, can be relied on as a strong tool in analyzing visitor behavior. These results pave the way for further research in optimizing marketing strategies and development of playgrounds, by focusing on more specific and personalized visitor needs and preferences.

REFERENCES

Abas, M. I., Ibrahim, I., Syahrial, S., Lamusu, R., Baderan, U. S., & Kango, R. (2023). Analysis of Covid-19 Growth Trends Through Data Mining Approach As Decision Support. *Sinkron*, *8*(1), 101–108. https://doi.org/10.33395/sinkron.v8i1.11861

Aji, G. W., & Devi, P. A. R. (2023). Data Mining Implementation For Product Transaction Patterns Using Apriori Method. *Sinkron*, *8*(1), 421–432. https://doi.org/10.33395/sinkron.v8i1.12071

Aldo, D. (2023). Data Mining Sales of Skin Care Products Using the K-Means Method. *Sinkron*, *8*(1), 295–304. https://doi.org/10.33395/sinkron.v8i1.12007

Andi, A., Juliandy, C., & David, D. (2023). Clustering Analysis of Tweets About COVID-19 Using the K-Means Algorithm. *Sinkron*, *8*(1), 543–533. https://doi.org/10.33395/sinkron.v8i1.12145

Asriningtias, S. R., Wulandari, E. R. N., Persijn, M. B., Rosyida, N., & Sutawijaya, B. (2023). Identification of Public Library Visitor Profiles using K-means Algorithm based on The Cluster Validity Index. *Sinkron*, *8*(4), 2615–2626. https://doi.org/10.33395/sinkron.v8i4.12901

Bustomi, Y., Nugraha, A., Juliane, C., & Rahayu, S. (2023). Data Mining Selection of Prospective Government Employees with Employment Agreements using Naive Bayes Classifier. *Sinkron*, *8*(1), 1–8. https://doi.org/10.33395/sinkron.v8i1.11968

Hasibuan, F. F., Dar, M. H., & Yanris, G. J. (2023). Implementation of the Naïve Bayes Method to determine the Level of Consumer Satisfaction. *SinkrOn*, *8*(2), 1000–1011. https://doi.org/10.33395/sinkron.v8i2.12349

Hasibuan, S. A., Sihombing, V., & Nasution, F. A. (2023). Analysis of Community Satisfaction Levels using the Neural Network Method in Data Mining. *Sinkron*, *8*(3), 1724–1735. https://doi.org/10.33395/sinkron.v8i3.12634

Indah, I. C., Sari, M. N., & Dar, M. H. (2023). Application of the K-Means Clustering Agorithm to Group Train Passengers in Labuhanbatu. *SinkrOn*, *8*(2), 825–837. https://doi.org/10.33395/sinkron.v8i2.12260

Mawaddah, A., Dar, M. H., & Yanris, G. J. (2023). Analysis of the SVM Method to Determine the Level of Online Shopping Satisfaction in the Community. *SinkrOn*, *8*(2), 838–855. https://doi.org/10.33395/sinkron.v8i2.12261

Pratama, H. A., Yanris, G. J., Nirmala, M., & Hasibuan, S. (2023). *Implementation of Data Mining for Data Classification of Visitor Satisfaction Levels*. *8*(3), 1832–1851.

Saputra, A. D. S., Hindarto, D., & Haryono, H. (2023). Supervised Learning from Data Mining on Process Data Loggers on Micro-Controllers. *Sinkron*, *8*(1), 157–165. https://doi.org/10.33395/sinkron.v8i1.11942

Sari, M., Yanris, G. J., & Hasibuan, M. N. S. (2023). Analysis of the Neural Network Method to Determine Interest in Buying Pertamax Fuel. *SinkrOn*, *8*(2), 1031–1039. https://doi.org/10.33395/sinkron.v8i2.12292

Sinaga, B., Marpaung, M., Tarigan, I. R. B., & Tania, K. (2023). Implementation of Stock Goods Data Mining

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13624
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Using the Apriori Algorithm. *Sinkron*, *8*(3), 1280–1292. https://doi.org/10.33395/sinkron.v8i3.12852

Siregar, A. P., Irmayani, D., & Sari, M. N. (2023). Analysis of the Naïve Bayes Method for Determining Social Assistance Eligibility Public. *SinkrOn*, *8*(2), 805–817. https://doi.org/10.33395/sinkron.v8i2.12259

Violita, P., Yanris, G. J., & Hasibuan, M. N. S. (2023). Analysis of Visitor Satisfaction Levels Using the K-Nearest Neighbor Method. *SinkrOn*, *8*(2), 898–914. https://doi.org/10.33395/sinkron.v8i2.12257

Wijaya, E. B., Dharma, A., Heyneker, D., & Vanness, J. (2023). Comparison of the K-Means Algorithm and C4.5 Against Sales Data. *SinkrOn*, *8*(2), 741–751. https://doi.org/10.33395/sinkron.v8i2.12224

*name of corresponding author