

Algorithm Of K-Medoids Analyzes Personality Types Based On Holland Theory

Haryadi, Ganefri, Fahmi Rizal, Yuyun Yusnida Lase, Yulia Fatmi, B.H.Hayadi, M.Ropianto

Abstract: the difficulty of identifying someone's personality type and wants to prove the Algorithm of K-Medoids in data mining that is used to do quite a lot of data clustering in determining someone's personality type. This algorithm is also known as Partitioning Around Medoids (PAM), which is a variant of the K-Means method. K-Medoids Clustering exists to overcome the weaknesses of K-Means Clustering. K-Medoids uses the partition clustering method to cluster n-objects into a number of k-clusters. This algorithm uses objects in a collection of objects that represent a cluster. The objects that represent a cluster are called Medoids. Clusters are built by calculating the closeness that is owned by between medoids and non-medoids objects. This study uses the algorithm of K-Medoids in determining personality types based on Holland's theory in the Realistic Type, The Investigative Type, the Artistic Type, the Social Type, the Entrepreneur Type, and Routine Type (Conventional Type). Sample data used in this study were 50 (fifty) students obtained from the results of tests conducted. Data samples were clustered into 6 (six) clusters. From the final results of calculations performed that the level of accuracy of the data in conducting clustering is 68% based on the results of the validation of personality tests conducted on 50 (fifty) students. So from the results of this study indicate that the algorithm of K-Medoids can predict student's personality types for future careers according to their personality.

Keywords: Algorithm of K-Medoids, Personality Type, Holland Theory

1 INTRODUCTION

Cluster analysis is one of the types of problems in data mining. Data mining itself according to David Hand, Heikki Mannila, and Padhraic Smyth from MIT in Larose is an analysis of large-sized data) to find clear relationships and infer them that were not previously known in a way that is currently understood and useful for the owner of the data[1], while cluster analysis in data mining (also known as clustering) is a method used to divide a series of data into several groups based on predetermined similarities[2], one of the methods that exist in data mining in the process of clustering quite a lot of data is the Algorithm of K-Medoids. K-Medoids Clustering, also known as Partitioning Around Medoids (PAM), is a variant of the K-Means method. This is based on the use of medoids rather than from observations of the mean held by each cluster, with the aim of reducing the sensitivity of the partition associated with extreme values in the dataset [3]. K-Medoids Clustering exists to overcome the weakness of K-Means Clustering that is sensitive to outliers because an object with a large value may substantially deviate from the data distribution[4]. K-Medoids uses the partition clustering method to cluster a group of n objects into a number of k clusters. This algorithm uses objects in a collection of objects that represent a cluster. The objects that represent a cluster are called medoids. Clusters are built by calculating the closeness that is owned by medoids and non-medoids objects [5]. Associated with research conducted by the author, prior research is very important in order to know the relationship between researchers conducted previously with research conducted at this time. Research conducted by Astri (2017) that discusses the implementation of the algorithm of partitioning around medoid (PAM) to classify high schools in DIY based on the absorption value of national examinations. Based on evaluations using 29 competencies, the value of absorptive capacity can be concluded that this algorithm can be used to classify school data with the given k values. Based on the evaluation using an average of 29 competency scores, it can be concluded that this algorithm can classify the average absorption values into three groups, namely groups with high, medium and low standard deviations [6]. In addition, Yusuf and Novian (2014) also conducted research aimed at designing and implementing fish clustering systems, especially in betta fish using color, shape, and texture features. This research

produces betta fish data clustering using the K-Means and K-Medoids methods to retrieval of images that are able to cluster the database image data in large quantities well. This study also proves that the algorithm of K-Medoids get a more accurate clustering with a faster running time value than using the K-Means method [7]. This study analyzes personality types based on Holland theory using the algorithm of K-Medoids. Personality is a combination of thoughts, emotions and behavior that makes a person unique, different from one another and also how a person sees himself. Personality character prominently distinguishes oneself from others. Personality is an important element in achieving one's success. The experts have formulated various personality theories with various assumptions and backgrounds of different individual environments. Personality conception that emphasizes the interaction between the environment and individuals that is most often used is Holland personality theory. The main focus of Holland's theory is placed on an understanding of vocational behavior to produce practical ways of helping people who are young, adult or even older in their careers both in education and in the work (Louis, 2010). This theory emphasizes the concept of interest as the basis of the formation of one's personality. This theory also emphasizes personal competence, educational behavior, social behavior and personality. The concept of interest concerning work is the result of a combination of a person's life history and overall personality, so that certain interests eventually become a personality trait in the form of self-expression in the field of work, academic studies, core hobbies, various creative activities and many other likes. So it is simply that it can be said that vocational interest is the most important aspect of personality so that the inventory of interests is seen as a personality test [8]. An indication of interest is one's preference for certain activities, while dislike becomes contra indicative. Holland himself developed several tests that can help people get to know themselves, such as: The Vocational Preference Inventory in 1977 and Self-Directed Search in 1979. Based on Holland's theory, a person's characteristic types are divided into six, namely the Realistic Type, the Investigative Type, the Artistic Type, the Social Type, the Enterprising Type), and Routine Type (Conventional Type). The more suitable a person is with one of the six types,

the more it will appear to the characteristics and patterns of behavior that are specific to that type. Each type of personality is a theoretical type or ideal type, which is the result of interactions between internal and external factors.

2. METHODOLOGY

The research method used is the study of literature and data samples. The deepening of the concept of a proposition is to collect literature relating to the algorithm using basic research types, the performance of the algorithm of K-Medoids in sequence is as follows:

1. Normalizing the data to be used in the calculation of the K-Medoids.
2. Initializing the cluster center randomly, then calculate the distance of the data (object) with the center of the cluster with Euclidean Distance.
3. Calculating the total distance of all data in the cluster.
4. Initializing the new cluster center randomly then calculate the distance of the data (object) with the cluster center and Euclidean Distance.
5. Calculating the difference in the total distance by reducing the total new distance - the total old distance.
6. repeating steps 3 to 5, until there is no change in the medoid, then the cluster and its cluster members are obtained.
7. K-Medoids results.

3. RESULT AND DISCUSSION

K-Medoids or Partitioning Around Medoids (PAM) is a clustering algorithm similar to K-Means. The difference between these two algorithms is the K-Medoids or PAM algorithm that uses objects as representatives (medoid) as the center of the cluster for each cluster, while K-Means uses the mean as the center of the cluster [9] [Kaur, et al. , 2014]. Algorithm of K-Medoids has advantages to overcome weaknesses in the algorithm of K-Means which are sensitive to noise and outliers, where objects with large values that allow deviations from the data distribution. Another plus is that the results of the clustering process do not depend on the order of entry in the dataset [10][11][12][13]. The algorithm steps of K-Medoids are:

1. Initializing the center of the cluster by k (number of clusters)
2. Allocate each data (object) to the nearest cluster is to use the Euclidian Distance distance measurement equation with the following equation:

$$d(x, y) = |x - y|$$

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; 1,2,3, \dots n$$

3. Randomly selecting the object in each cluster as a new medoid candidate.
4. Calculating the distance of each object in each cluster with the new medoid candidate.
5. Calculating the total deviation (S) by calculating the new total distance value - the old total distance. If S < 0, then exchanging objects with data clusters to form a new set of k objects is medoid.

Repeating steps 3 to 5, so that there is no change in the medoid, then a cluster and its cluster members are obtained.

The data sample of 50 (fifty) people are consisting of men and women for which the data taken is the result of the test for each sample. From these results the data obtained are as follows :

Data	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Anak Ke-1	45	80	70	65	75	78
Anak Ke-2	78	66	64	77	56	75
Anak Ke-3	67	50	45	78	67	67
Anak Ke-4	70	67	57	45	67	68
Anak Ke-5	67	89	88	87	66	45
Anak Ke-6	45	77	65	67	56	89
Anak Ke-7	50	60	64	66	78	65
Anak Ke-8	70	66	75	78	60	55
Anak Ke-9	50	50	75	67	65	50
Anak Ke-10	30	67	66	78	40	45
Anak Ke-11	78	78	60	70	50	55
Anak Ke-12	67	60	78	75	60	65
Anak Ke-13	60	45	67	60	65	55
Anak Ke-14	66	60	77	65	60	60
Anak Ke-15	56	77	70	68	55	50
Anak Ke-16	88	45	70	70	50	55
Anak Ke-17	45	35	67	65	60	45
Anak Ke-18	65	67	70	60	55	50
Anak Ke-19	34	66	70	68	50	55
Anak Ke-20	67	78	65	70	45	40
Anak Ke-21	78	56	66	75	65	45
Anak Ke-22	73	50	65	66	60	50
Anak Ke-23	75	67	50	65	60	50
Anak Ke-24	66	66	55	70	65	55
Anak Ke-25	60	76	67	75	50	55
Anak Ke-26	62	78	65	70	55	55
Anak Ke-27	56	58	67	75	60	50
Anak Ke-28	78	82	68	70	45	54
Anak Ke-29	79	56	78	78	50	65
Anak Ke-30	65	50	80	75	55	60
Anak Ke-31	56	45	75	60	65	40
Anak Ke-32	45	56	65	65	70	50
Anak Ke-33	36	55	67	60	66	65
Anak Ke-34	79	60	60	55	60	50
Anak Ke-35	80	66	77	65	55	50
Anak Ke-36	64	50	68	60	60	55
Anak Ke-37	74	45	65	70	60	60
Anak Ke-38	60	67	72	75	70	55
Anak Ke-39	56	70	65	50	67	45
Anak Ke-40	45	60	65	65	65	45
Anak Ke-41	65	55	77	60	55	50
Anak Ke-42	60	56	70	60	60	65
Anak Ke-43	63	60	70	65	78	55
Anak Ke-44	56	60	60	70	60	50
Anak Ke-45	55	65	65	65	78	65
Anak Ke-46	50	55	67	70	60	60
Anak Ke-47	45	60	67	75	45	50
Anak Ke-48	77	60	55	77	75	45
Anak Ke-49	66	50	65	73	55	50
Anak Ke-50	60	65	68	68	50	70

The data obtained from the test results is a calculation process using the algorithm of K-Medoids. The steps are used as described previously. From the results of these calculations then the results are obtained as the table below.

Table 3. Final Calculation Results of Algorithm of K-Medoids.....

The calculation above with the algorithm of K-Medoids is stopped at the first iteration because the value $S > 0$.

$S = \text{new total cost} - \text{old total cost}$

$$S = 181,64275$$

Because the value of $S > 0$, then it iterates and is stopped. The final results of calculations with K-Medoids, the clustering results are obtained as follows:

1. Children who have realistic types are 10 (ten) children, namely the 2nd child, 4th child, 16th child, 21st child, 23rd child, 29th child, 34th child, ninth child 35th and 48th child, and the accuracy of K-Medoids calculations in determining realistic personality types is 100%.
2. There are 7 (seven) children who have investigative type, namely 5th child, 11th child, 15th child, 20th child, 25th child, 26th child and 28th child, and accuracy K-Medoids calculation in determining investigative personality type is 77%.
3. There are 10 children who have artistic types, namely 9th child, 12th child, 13th child, 17th child, 30th child, 31st child, 36th child, 9th child 41, 42nd child and 49th child, and K-Medoids calculation accuracy in determining personality type is 71%.
4. Children who have Social type are 2 (two) children, namely the 10th child and 47th child, and the accuracy of K-Medoids calculation in determining the social personality type is 22%.
5. There are 5 (five) children who have enterprising type, namely 7th child, 32nd child, 40th child, 43rd child and 45th child, and the accuracy of K-Medoids calculation in determining the enterprising personality type is 83 %.
6. And no child has a conventional type and the accuracy of K-Medoids calculations in determining conventional personality types is 0%.
7. The accuracy of the algorithm of K-Medoids in determining personality types based on Holland theory is 68%.

4. CONCLUSION

From the results of calculations carried out in determining personality types based on Holland theory it can be concluded that:

1. The process of calculating the algorithm of K-Medoids in determining personality types based on Holland theory stops at iteration 1 (first) because the value is $S > 0$.
2. From the results of calculations performed on 50 (fifty) children with the algorithm of K-Medoids, it is found that children who have the realistic personality types are 10 (ten) children, investigative types are 7 (seven) children, artistic types are 10 (ten) children, social type are 2 (two) children, enterprising type are 5 (five) children, and no children who have conventional type.
3. The accuracy of the algorithm of K-Medoids in determining personality types based on Holland theory is 68%.

5. REFERENCES

- [1]. "Larose, Daniel T. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. Hoboken: John Wiley and Sons, Inc."
- [2]. "Sitepu, R., Irmeilyana, dan Gultom, B. 2011. Analisis Cluster terhadap Tingkat Pencemaran Udara pada Sektor Industri di

- Sumatera Selatan. Jurnal Penelitian Sains. Vol. 14 No. 3A Juli 2011 11 – 17."
- [3]. "Vercillis, Carlo. 2009. Business Intelligence: Data Mining and Optimization for Decision Making. Milan: WILEY."
- [4]. "Han, J., Kamber, M. 2006. Data Mining: Concept and Techniques. Waltham: Morgan Kaufman Publisher."
- [5]. "Setyawati, Astri Widiastuti. 2017. Implementasi Algoritma Partitioning Around Medoid (PAM) untuk Pengelompokan Sekolah Menengah Atas di DIY Berdasarkan Nilai Daya Serap Ujian Nasional. Skripsi. Yogyakarta: Fakultas Sains dan Teknologi Universitas Sanata D."
- [6]. "Sharf, R.S. 2006. Applying Career Development Theory To counseling. Canada."
- [7]. "Yusuf, Rampi dan Novian, D. 2014. Aplikasi Image Retrieval pada Varian Ikan Cupang Menggunakan K-Means dan K-Medoids Algoritm. Gorontalo: Jurnal Teknik. Vol. 12 No.2 Desember 2014 111 – 122."
- [8]. "Holland, J. L. (1997). Making Vocational Choice: A Theory of Vocational Personalities and Work Environments (3rd/17 Edition). New Jersey: Prentice-hall. Inc."
- [9]. "Pramesti, Dyang Falila., Furqon, M. Tanzil., Candra Dewi. 2017. Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). Jurnal Pengembangan Teknologi Informasi dan Ilmu ."
- [10]. "Pervin, L.A., John, OP. 2001. Personality: Theory and Research 8ed. John Wiley & Sons, Inc, New York."
- [11]. "Nindyati, Dewi Ayu. 2006. Kepribadian dan Motivasi Berprestasi Kajian Big Five Personality. Jurnal Psikodinamik."
- [12]. R. Pi, "The Implementation of Calendar Academic Monitoring System in University Using," vol. 11, pp. 497–500, 2019.
- [13]. A. Lubis and B. H. Hayadi, "Designing Architecture of Information Dashboard System to Monitor Implementation Performance of Economic Census 2016 in Statistics Indonesia," 2016 4th Int. Conf. Inf. Commun. Technol., vol. 4, no. c, pp. 1–5, 2016.