Jurnal E-Komtek
Vol. 6, No. 1 (2022) pp. 70-81
https://jurnal.politeknik-kebumen.ac.id/index.php/E-KOMTEK
p-ISSN : 2580-3719 e-ISSN : 2622-3066

Data Mining of The Grouping And Mapping of The Health Centers in Sleman Regency With K-Means Clustering Algorithm

Widy Prasetiyo
Department of Informaticss Engineering, Universitas Mercu Buana Jakarta, Indonesia, 11650

widyprasetiyo27@gmail.com

https://doi.org/10.37339/e-komtek.v6i1.921

Published by Politeknik Piksi Ganesha Indonesia

Artikel Info
Submitted:
26-05-2022
Revised:
12-06-2022
Accepted:
12-06-2022
Online first :
30-06-2022

Abstract

This research which is a grouping and mapping system of Puskesmas (public health center) is one solution to obtain information related to areas that have adequate Puskesmas services. With the help of a geographic information system, it is able to describe the state of the Sleman district based on the availability of health centers per sub-district. In a geographic information system, the information produced is in the form of visual images that make it easier for the public to read the information provided. This study uses data from health workers who work in health centers and were grouped first using the k-means method. The reason for grouping them first was to make it easier to map the existing health centers, while the K-means method was chosen because it has a high accuracy to the size of the object, so this algorithm is relatively more scalable and efficient for processing large numbers of objects. In addition, the K-Means algorithm is not affected by the order of objects.

Keywords: Information systems, Puskesmas, K-Means

Abstrak

Penelitian yang merupakan sistem pengelompokan dan pemetaan puskesmas ini merupakan salah satu solusi untuk mendapatkan informasi terkait wilayah yang memiliki pelayanan puskesmas yang memadai. Dengan bantuan sistem informasi geografis mampu menggambarkan keadaan kabupaten Sleman berdasarkan ketersediaan puskesmas per kecamatan. Dalam sistem informasi geografis, informasi yang dihasilkan berupa gambar visual yang memudahkan masyarakat untuk membaca informasi yang diberikan. Data yang digunakan dalam penelitian ini menggunakan data dari tenaga kesehatan yang bekerja di Puskesmas dan dikelompokan terlebih dahulu menggunakan metode k-means. Alasan pengelompokan terlebih dahulu adalah untuk memudahkan pemetaan Puskesmas yang ada. Sedangkan metode K-means dipilih karena memiliki akurasi yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk memproses objek dalam jumlah besar. Selain itu, algoritma K-Means tidak terpengaruh oleh urutan objek.

***Kata-kata kunci**: Sistem Informasi, Puskesmas, K-Means*

1. Introduction

Public health center (henceforth, Puskesmas) is one of the public health service facilities in an integrated, comprehensive and sustainable manner in an area that is under the direct supervision of the District Office. Adequate facilities and infrastructure, as well as Health Human Resources (HHR) are needed so that health service facilities can function properly. In order to achieve the highest level of health, the government is responsible for the availability of fair and equitable Health Human Resources (HHR).

According to its type, each puskesmas has various HHRs, including General Practitioners, Dentists, Nurses, Midwives, Pharmacies, Community Health Workers, Environmental Health Workers, Nutritionists, Medical Lab Technologists, and Health Support Workers. Health workers are one of the most important aspects of supporting existing services.

Sleman Regency has several puskesmas units spread over 25 sub-districts. In total, there are 25 puskesmas units, with 20 puskesmas capable of providing non-inpatient care and five capable of providing inpatient care. The distribution of puskesmas in all cities of sub-district units in Sleman Regency is undoubtedly even, but this is not necessarily accompanied by an even distribution of the number of health workers. To support the even distribution of HHR in Puskesmas in Sleman district, a grouping method is needed to find out which areas have HHR advantages and which areas lack HHR. One of the most frequently used grouping methods is the K-Means Algorithm.

As an effort should be made in this research to avoid repetition of research, it is necessary to review the previous research studies as a basis and reference. The following are some research papers that are relevant to the research's discussions.

Research on clustering the number of human resources for health centers to support equity in Central Java province using K-Means algorithm, "Pengelompokan Jumlah Sumber Daya Manusia Kesehatan Puskesmas untuk Menunjang Pemerataan pada Provinsi Jawa Tengah Menggunakan Algoritma K-Means". The problem in this research is that the puskesmas in the area were still short of health workers. This was due to the unequal distribution of health workers; some areas had an excess of health workers, while some others lacked health workers. The research concluded that the K-Means algorithm can group data on the number of HHR Puskesmas in Central Java Province into three clusters: High scores (excess HHR), Medium (adequate HHR), and Low (lack HHR) [1].

Research on building a geographical information system for clustering the crime levels in Malang City using the K-Means method, "Sistem Informasi Geografis Pengelompokan Tingkat Kriminalitas Kota Malang Menggunakan Metode K-Means". The problem of this research is that there was no information that shows the crime-prone areas in Malang City. This research produced a system that can notify the public in Malang City about the sub-district areas that are quite safe, quite vulnerable, vulnerable, and very prone to crimes, using K-Means with the help of Qgis tools [2].

Research on clustering and mapping the health degrees of Bengkulu using the K-Means clustering method, "Pengelompokan Dan Pemetaan Derajat Kesehatan Kota Bengkulu Dengan Metode K-Means Clustering". The problem of this study is that there was no explanation regarding the general health status grouping of each region based on the value of the indicators. This study concluded that K-Means Clustering gives variable results because the initial centroids were formed randomly, so it is recommended to add a method to overcome the determination of the optimal initial centroid [3].

Research on building a geographic information system for mapping high schools/equivalents in Surakarta City using website-based Javascript Library Leaflets, "Sistem Informasi Geografis Pemetaan Sekolah Menengah Atas/Sederajat di Kota Surakarta Menggunakan Leaflet Javascript Library Berbasis Website". This study concluded that the system created using the leaflet library can facilitate researchers in making maps, and some features can be made using the plugins that are already available on the internet and are open source. The Usability test obtained a SUS score of 79.4, which indicates that the system can be accepted by the community and shows a good system classification [4]. The problem and obstacle in this research are that the Geographic Information System (GIS) features that are used and presented on the website are not yet complete, so users or the public found it difficult to find information related to the availability of the number of school capacities that meet the criteria.

Research on building a geographic information system for mapping the laundry application partner users at PT Tenten Digital Indonesia based on web and mobile, "Sistem Informasi Geografis Pemetaan Mitra Pengguna Aplikasi Laundry di PT Tenten Digital Indonesia Berbasis Web Dan Mobile". In this study, the system was created to help PT Tenten Digital Indonesia marketing provide a database of laundry business information that has the potential to become partners. The system can be accessed through PCs, laptops, or mobile devices (smartphones or tablets). The testing of the system functionality using black-box testing did not find any bugs or errors when it was running on a web device (PC or laptop) or mobile (smartphone or tablet) [5]. The problem of this research is the security system on the website that is not strong enough, so the data from the partner users of the application is not protected safely and tends to be easy to hack.

2. Method

a. Material

1) K-Means Algorithm

K-means algorithm is an algorithm that partitions data into clusters so that data that have similarities are in the same cluster and data that have dissimilarities are in other clusters [6].

2) Clustering

One of the techniques known in data mining is clustering. Clustering in data mining means grouping a number of data or objects into clusters (groups) so that one cluster contains data with similarities; and are different from those of others. Until now, scientists are still making various efforts to improve the cluster model and calculate the optimal number of clusters; so the best clusters can be produced. There are two clustering methods that we are familiar with, namely hierarchical clustering and partitioning. The hierarchical clustering method consists of complete linkage clustering, single linkage clustering, average linkage clustering, and centroid linkage clustering, while the partitioning method consists of k-means and fuzzy k-means **[7]**.

3) Data Mining

Data mining is the process of obtaining information by searching for patterns and relationships hidden in large piles of data. Data mining, or often referred to as knowledge discovery in database (KDD), is an activity that collects and uses historical data to find regularities, patterns, or relationships in large data. Data Mining output can be used to assist future decision making. The development of KDD has reduced pattern recognition because it has become a part of Data Mining **[8]**.

4) Geographic Information System

Geographic Information System refers to spatial data that are processed by a system, and to process this spatial data requires a software or tool of many kinds. Geographic information is information which provides information about the geographical location of a location on a map of the earth **[9]**. The benefit of GIS is that it makes it easy for users or decision makers to determine the policies to make, especially those related to spatial aspects. With this technology, it will be easier in terms of land mapping **[10].** The advantage of GIS is to produce information that is close to real-world conditions, and can predict a result and strategic planning **[11].**

5) Unified Modelling Language

Unified Modeling Language (UML) is a visual design method for object-oriented systems. UML defines in the form of diagrams, including Use Case Diagrams, Class Diagrams, Sequence Diagrams, State Chart Diagrams, Activity Diagrams, Collaboration Diagrams, Component Diagrams, and Deployment Diagrams. System Development Requirement: (a) framework codeigniter v3.1.11, (b) editor text, (c) design graphic software, (d) XAMPP, (e) hosting, (f) sub domain, dan (g) computer and smartphone.

6) Blackbox Testing

Black box testing is a testing method where the tester only focuses on what the system should do. A test is successful when a system can process data; and the results meet the expectation. When using the black box method, the tester does not need to know the structure and design of the data in the system. They only see whether the system has bugs or not **[12]**.

b. Methods

The description of the research stages in this study is shown in **Figure 1**.

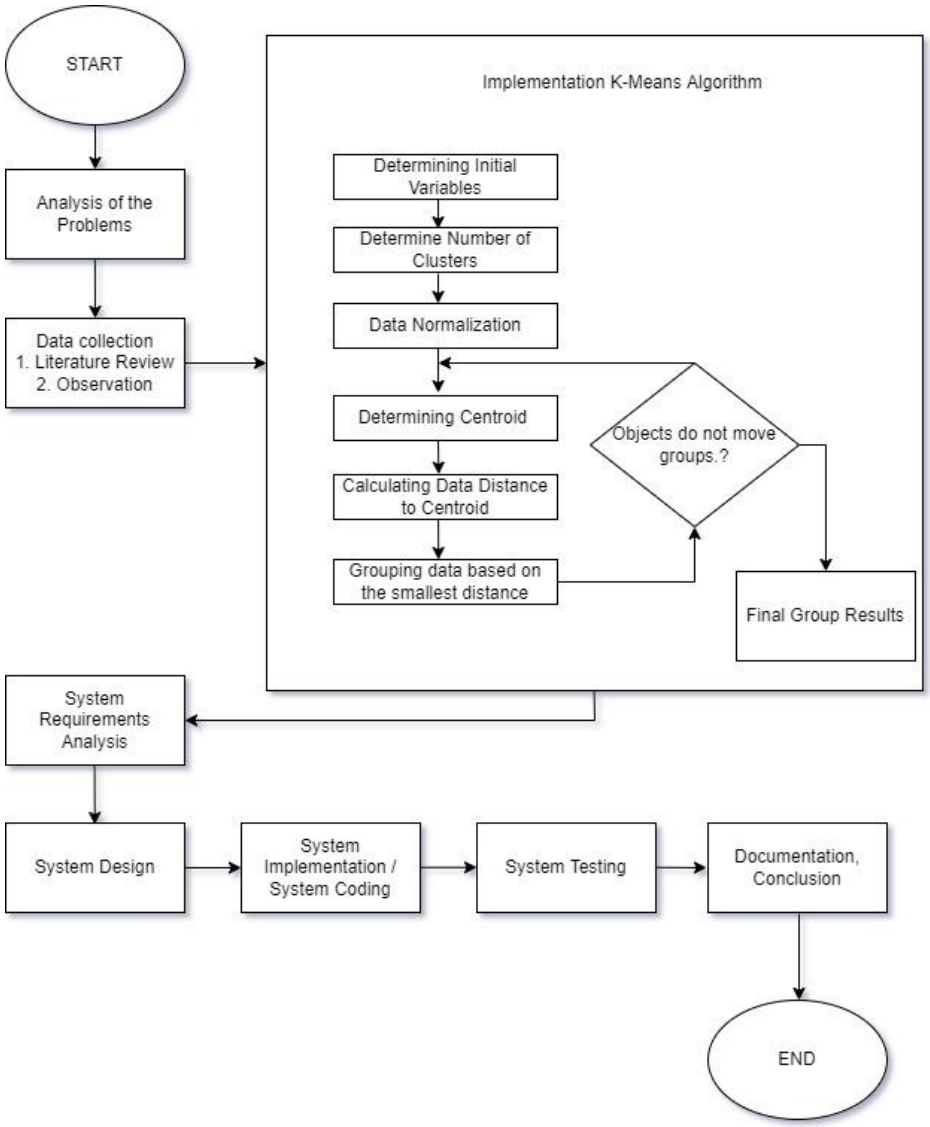

Figure 1. Research Stages

The research stages listed in Figure 1 are elaborated as follows.

a. Formulation of the problem

This stage is the beginning of the research. The problems obtained were then formulated as a research reference.

b. Data Collection

In this stage, the researcher collected the data that are relevant and useful for the research. The data collection was accompanied by a literature study by looking for related research. The data collected were in the form of data on health centers in Sleman district along with the number of health workers obtained from the website of the ministry of health for the number of personnel and the distribution of health centers.

c. The Implementation of the K-Means Algorithm

In this stage, the calculation process and implementation of the K-means algorithm begins with the selection of variables that are used as the initial stages, followed by determining the number of groups, which consist of 3 groups, namely low, sufficient, and high. The next step was to normalize the data and determine the centroid point. The step following it was to calculate the distance of the data to the centroid point, then calculate and based on the smallest distance to be included in the group. The final stage will be checked for the initial group and after that whether it is the same; if it is, the process will stop, and if not, the group calculation will start with the new centroid point value.

d. System Requirement Analysis

In this stage, the system requirements are described for the sake of system design. System requirement analysis includes system features to be built along with the hardware and software components to use.

e. System Design

System design stage is the system design process starting with the system flow described with Data Flow Diagrams, database design, and system display design.

f. System Implementation

In this stage, the researcher created a system based on a predetermined system design. The system were built on a web-based basis with the PHP programming language and MySQL database. The map creation was assisted by the Google Maps API with java script technology.

g. System Test

The system testing is carried out on the system that has been made, with black box testing based on functional requirements that have been described whether it is running properly or there are still obstacles.

h. Documentation and Conclusion

In this stage, the researcher made documentation of the results of research carried out accompanied by conclusions from the research conducted.

3. Results and Discussion

a. System Design

The system design was carried out with the materials and data needed in this study, namely data on health facilities in Sleman Regency taken from the website page of the Ministry of Health of the Republic of Indonesia. In addition to this, location data was searched with the help of Google Maps Api from the address that has been obtained from the Ministry of Health.

b. Application of the K-Means Clustering Algorithm

The K-Means algorithm stage process was carried out with the following stages:

1) Determining the number of clusters. The number of clusters refers to the number of data groups to be created or generated. In this study, the number of clusters to be created is three.

2) Generating the initial centroid. The initial centroid was obtained randomly, and the number of centroids is as many as clusters to be created. The initial centroid is the first cluster center point or the beginning of the cluster center. The initial controid can be calculated by the following equation:

$$ci = \min + \frac{(i-1) * (\max - \min)}{n} + \frac{(\max - \min)}{2 * n}$$

In this case, we try to calculate Centroid 1 in the *puskesmas* data by grouping a number of three clusters.

C [Dokter][1] =1 + $\frac{(1-1)*(7-1)}{3} + \frac{(7-1)}{2*3}$

C [Dokter][1] = 2

From the results of calculations with the above formula, it was found that the centroid value in the three clusters is as it can be seen in **Table 1**.

Table 1. Initial Centroid Value

Cluster	Doctor	Nurse	Midwife
Centroid 1	2,00	4,17	4,50
Centroid 2	4,00	8,50	9,50
Centroid 3	6,00	12,83	14,50

c. Input System

The system input is the data needed for the Geographic Information System for Mapping *Puskesmas* in Sleman district. From the data entered into the system, the system will process the data into information. Enter the required data, namely: district data, health center data, human resources data, and data on the number of human resources in health facilities.

d. User Level Output

The output of the Geographic Information System for Mapping Health Facilities in Sleman Regency is a visualization of health facility data in the form of sub-district mapping; besides, it can group data with the K-Means algorithm. The output picture is presented in **Figure 2**.

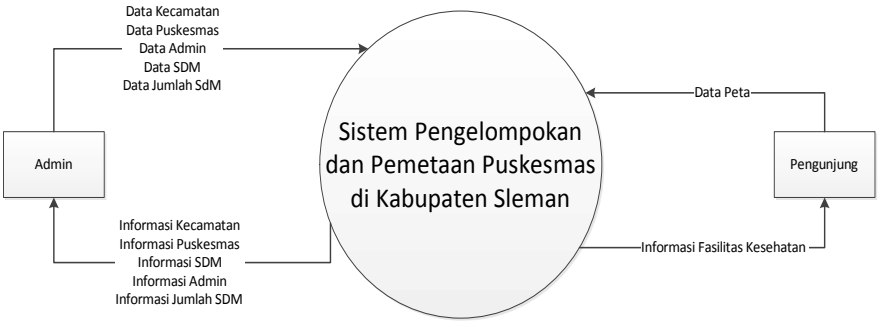

Figure 2. Context Diagram

e. Functional Design

The functional design is built using ERD on data storage in a system to produce good information and obtain a good database. A good database requires a model or a good database design technique as well. The design of the ERD to be built has five main entities, namely sub-district data, admin, facilities, human resources, and the number of human resources. The ERD design can be seen in **Figure 3**.

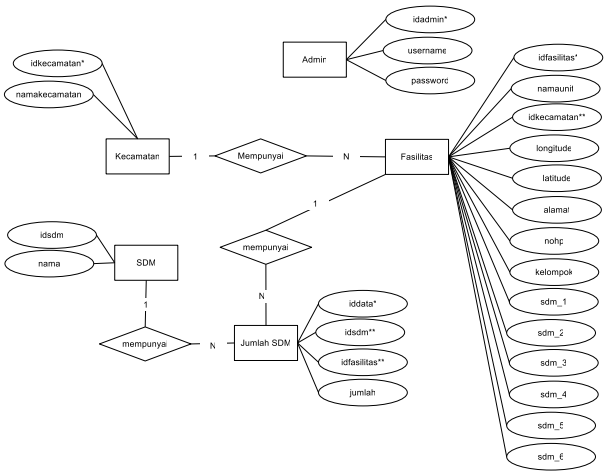

Figure 3. DFD Level 2 Manage Facilities

f. Table Implementation Design

The design of relations between tables is the result of the implementation of the results of the ERD design which is visualized in the form of storage media. The design of relations between tables can be seen in **Figure 4**.

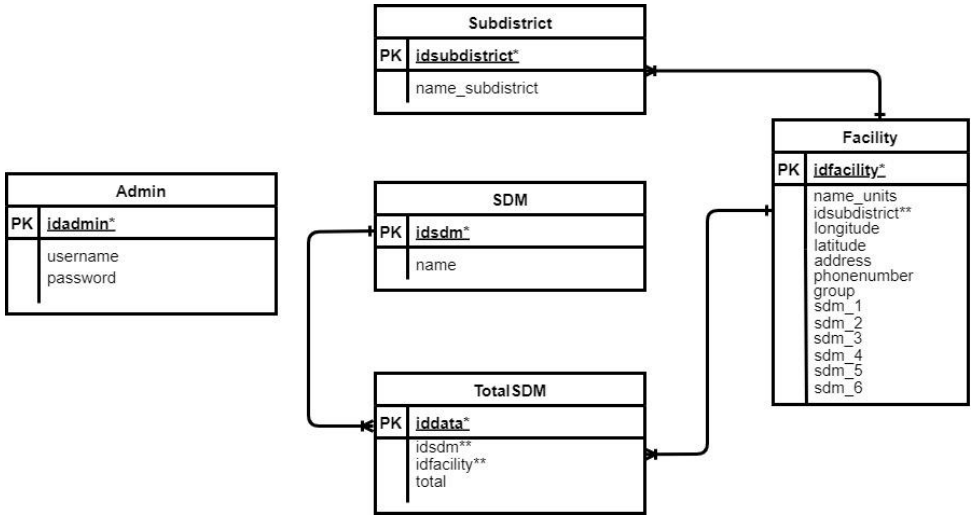

Figure 4. Relation Table

Admin table serves to store admin data. The admin table has three columns. In the admin table, idadmin serves as the primary key.

g. Information System Implementation

System implementation is to prepare all system implementation activities according to a predetermined design. After that, the system testing stage is carried out to minimize all possible errors that occur.

1) Login Page Display

Lgin page display is presented in **Figure 5**.

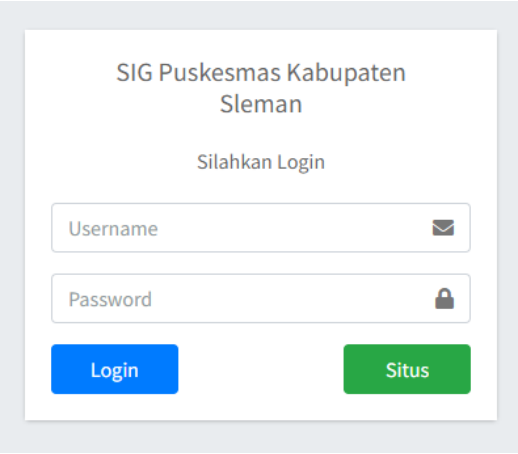

Figure 5. Login Page Display

2) Sub-district Data Page

The sub-district page is a page that serves to display sub-district data. This page is able to display sub-district information. In the sub-district page, the admin can search data and is only able to process changes to sub-district data. District data display is presented in **Figure 6**.

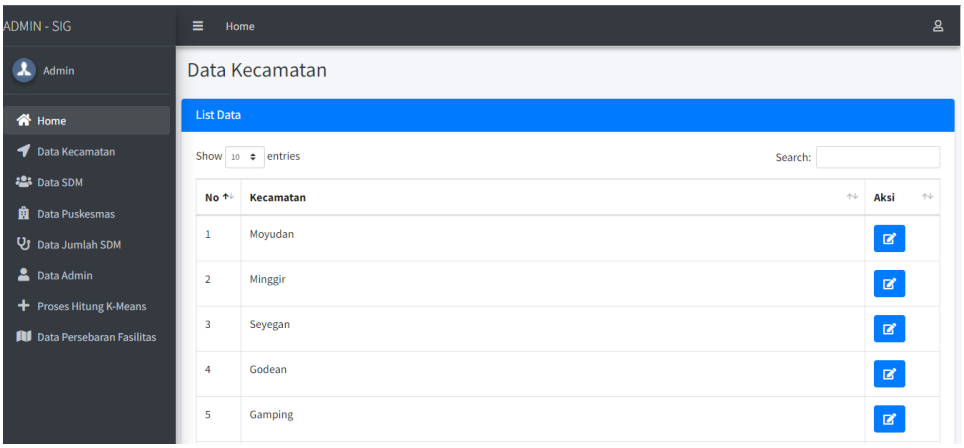

Figure 6. District Data Display

3) Data & Human Resources Page View

Human Resources data page is a page used for processing Human Resources data. In this page, the admin can add new Human Resources data and process data in the form of changing and deleting Human Resources data. The Human Resources page is also equipped with a data search facility. Human resources data page view is presented in **Figure 7**.

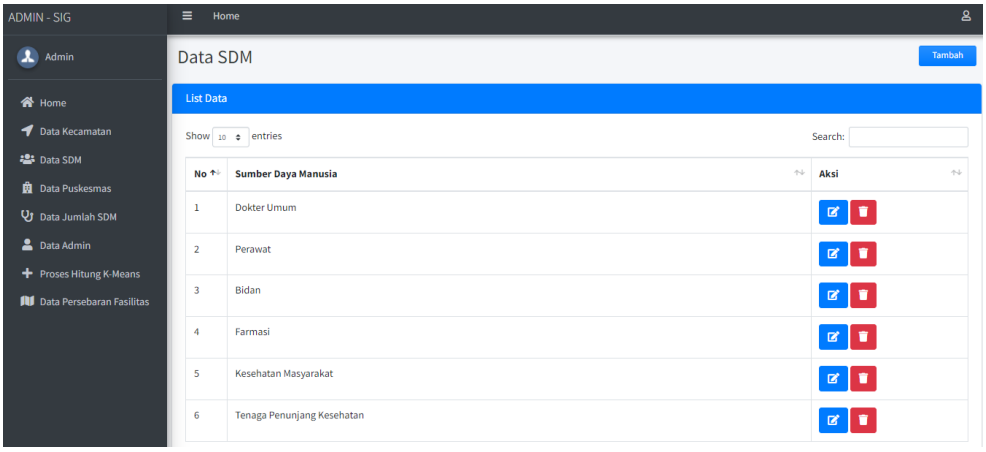

Figure 7. Human Resources Data Page View

4) Health Center Distribution Map Page View

The distribution map of the *puskesmas* is the page that is used to display map information. In this page, the map displayed per sub-district has a different color, informing the group for each sub-district. This page shows the location of the health center facilities. The

facility distribution page is equipped with legend information for the resulting color description. Health center distribution map page view is presented in **Figure 8**.

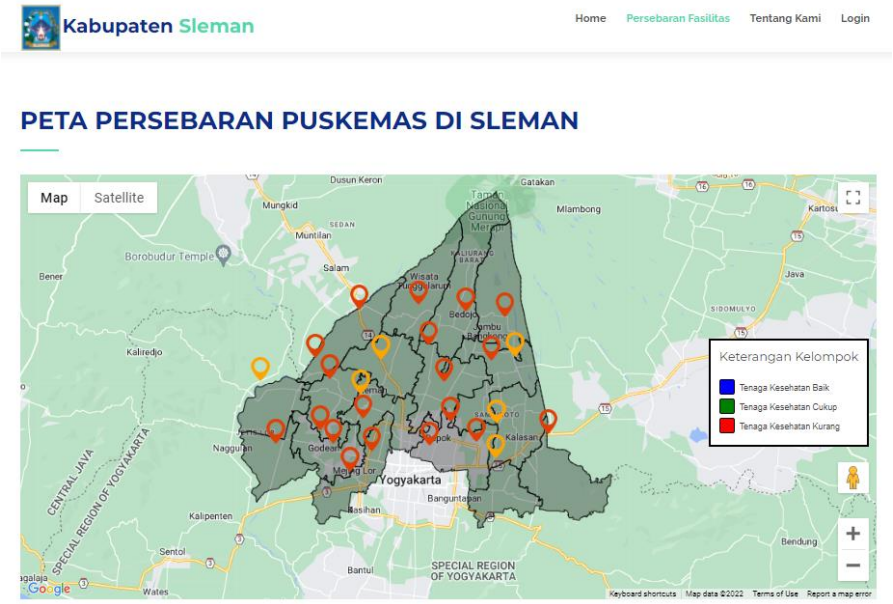

Figure 8. Health Center Distribution Map Page View

4. Conclusion

Based on the research, the following conclusions were drawn:

a. The Geographic Information System for Mapping Health Facilities in Sleman Regency was built using the K-Means Algorithm to determine the group of each puskesmas based on the number of the existing workers. The system was built in the PHP programming language with a MySQL database and used the Google Maps Api library and visual maps using GeoJSON data for the regional borders of each sub-district.

b. Based on the system testing, this system could run all blackbox test units with test items that have been made in functional requirements analysis, starting from data processing and generating data information in the form of maps.

c. The system that was built displays detailed information on facilities in the form of the number of health workers in certain facilities and displays points based on the category of health centers with different icons. The information displayed has links to map points that can be displayed in Google Maps.

References

[1] S. Candra Dewi, A. Mutoi Siregar, and D. Sulistya Kusumaningrum, "Pengelompokan Jumlah Sumber Daya Manusia Kesehatan Puskesmas untuk Menunjang Pemerataan pada Provinsi Jawa Tengah Menggunakan Algoritma K-Means," vol. 1, no. 2, p. 86, 2020.

[2] S. Candra Dewi, A. Mutoi Siregar, and D. Sulistya Kusumaningrum, "Pengelompokan Jumlah Sumber Daya Manusia Kesehatan Puskesmas untuk Menunjang Pemerataan pada Provinsi Jawa Tengah Menggunakan Algoritma K-Means," vol. 1, no. 2, p. 86, 2020.

[3] R. Efendi, F. Farady Coastera, and F. Rizky Tanjung, "Pengelompokan dan Pemetaan Derajat Kesehatan Kota Bengkulu Dengan Metode K-Means Clustering," 2019.[Online].Available: http://ejournal.unib.ac.id/index

[4] Dimas A,A., Ridwan,R.(2020).Sistem Informasi Geografis Pemetaan Sekolah Menengah Atas / Sederajat di Kota Surakarta Menggunakan Leaflet Javascript Library Berbasis Website. Jurnal Teknik Elektro,Vol. 20 No. 02 September 2020.

[5] H. Santoso, "Sistem Informasi Geografis Pemetaan Mitra Pengguna Aplikasi Laundry di PT Tenten Digital Indonesia Berbasis Web dan Mobile," 2020.

[6] N. W. Rohmawati, S. Defiyanti, M. Jajuli, and T. H. Informatika Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang Jl Ronggowaluyo Telukjambe Timur Karawang, "Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa," 2015.

[7] F. Yunita, "Penerapan Data Mining Menggunkan Algoritma Kmeans Clustring pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri)," 2018.

[8] Fadlina, F, "Data Mining Untuk Analisa Tingkat Kejahatan Jalanan Dengan Algoritma Association Rule Metode Apriori (Studi Kasus Di Polsekta Medan Sunggal)", Informasi dan Teknologi Ilmiah (INTI), 2014.

[9] A. Nugroho and W. A. Kusuma, "Sistem Informasi Geografis Pemetaan Lokasi Bird Contest Kota Malang Berbasis Android," 2018.

[10] K. M. W. Wibowo, I. Kanedi, and J. Jumadi, "Sistem Informasi Geografis (SIG) Menentukan Lokasi Pertambangan Batu Bara di Provinsi Bengkulu Berbasis Website", jmi, vol. 11, no. 1, Jun. 2016.

[11] J. Hutagalung, D. A. Program Studi Sistem Informasi STMIK Triguna Dharma Medan Jl Jenderal Besar Nasution No, K. Bekala, M. Johor, and K. Medan, "Sistem Informasi Geografis Pemetaan Penyaluran Dana Bantuan Operasional (Studi Kasus di SMA Negeri Kota Pematangsiantar)," 2018. [Online]. Available: http://tunasbangsa.ac.id/ejurnal/index.php/jsaktiGISPemetaanPenyaluranDanaBantuanOperasional

[12] I. Safarina, I. K. Raharjana, and E. Purwanti, "Perencanaan Arsitektur Perusahaan untuk Pengelolaan Aset menggunakan Kerangka Kerja Zachman," J. Inf. Syst. Eng. Bus. Intell., vol. 1, no. 2, p. 59, 2015.