Sains Data Jurnal Studi Matematika dan Teknologi
ISSN 2986-903X

v2i1(2024); p.18-24

Geographic Information System Mapping of Essential Oil Producing Plants in Bangkalan Regency using K-Means Clustering

Achmad Zain Nur [a*], Achmad Jauhari [b]
[a] Departement of Information System Universitas Trunojoyo Madura
[b] Departement of Informatics Engineering Universitas Trunojoyo Madura
Jl. Raya Telang Kamal, Bangkalan 69162, Indonesia
E-mail: [a*] azain.nur@trunojoyo.ac.id, [b] jauhari@trunojoyo.ac.id

DOI : https://doi.org/10.52620/sainsdata.v2i1.41

Abstract
Essential oil is one of the high foreign exchange earning export commodities for Indonesia. Unfortunately, the high world market demand for essential oils has not been fully supplied by Indonesia. In Indonesia, essential oils generally come from farmers who have relatively small land and are then processed using simple distillation equipment. The same thing also happened in Kab. Bangkalan, Madura. This condition is the main reason why Indonesia's essential oil production is not optimal. To find out the distribution potential of essential oil-producing plants in Bangkalan Regency, it is necessary to have a step in building a system that can group the distribution areas of essential oil-producing plants, for this we need a clustering method. This study uses the K-Means Clustering method and uses the Euclidean algorithm to group the distribution areas of essential oil-producing plants, especially in Bangkalan Regency, so that later they can be identified based on clusters and displayed in visual form, namely in the form of digital maps using the Geographic Information System (GIS) application. Therefore, it is hoped that this research will be able to find out the distribution area of essential oil-producing plants in order to make it easier for local governments to monitor the potential of essential oil-producing areas in Bangkalan Regency.

Key words: Essential Oils, Geographic Information Systems, K-Means Clustering

INTRODUCTION

Essential oil is one of the high foreign exchange earning export commodities for Indonesia. The high world demand for essential oils is very profitable for Indonesia, because this high demand is a huge opportunity, because Indonesia is a tropical country that has a very suitable climate for developing essential oil-producing plants with good quality and quantity. This high demand indicates a relatively high price level. These opportunities and conditions should be a trigger for Indonesia to optimize its existing potential. Unfortunately, the high world market demand for essential oils has not been fully supplied by Indonesia. In Indonesia, essential oils generally come from farmers who have relatively small land and are then processed using simple distillation equipment.

The same thing also happened in Bangkalan Madura Regency. This condition is the main reason why Indonesia's essential oil production is not optimal. Bangkalan Regency is rich in other essential oil-producing plants such as nutmeg, cloves, citronella, cinnamon, eucalyptus, etc. To find out the distribution potential of essential oil-producing plants in Bangkalan Regency, it is necessary to develop a system that can classify the distribution areas of essential oil-producing plants.

This work is licensed under a Creative Commons Attribution 4.0 International License ©Author (s)

Geographic Information System (GIS) is a computer system designed to obtain, store, manipulate, analyze and manage geographic data and then present it as geographic information in the form of digital maps [5]. With the mapping of the distribution of essential oil-producing plants, it is certain that it can help interested parties to obtain information in real time about regional conditions in Bangkalan Regency.

K-Means Clustering is a method of grouping or clustering that is easy and does not require a long estimate of time to process large data [6]. The Geographic Information System (GIS) can display the results of grouping the distribution of essential oil-producing plants in Bangkalan Regency in the form of a map of an area clearly and can be stored, modified and updated on a regular basis. In previous studies, research has been conducted to show that K-Means Clustering has an accuracy value of 83.33% compared to Fuzzy C-Means which has an accuracy value of 50% in classifying poverty data in Gunung Kidul Regency. K-Means Clustering is also better at determining membership values compared to Algorithm Hierarchical Clustering (AHC).

This proves that K-Means Clustering can provide good results in the process of grouping data. By using the K-Means Clustering method to classify the distribution areas of essential oil-producing plants in Bangkalan Regency.

MATERIAL AND METHODS

K-Means Clustering

Clustering or grouping is a method used in analyzing data to solve data grouping problems to become multiple clusters in data mining operations [14]. One of the clustering methods that is often used is the K-Means Clustering method because this method is easy and does not require a long estimate of time to process large data. This method can also divide data into several clusters, data that has a high level of compatibility between data will be grouped in one cluster while data that has different characteristics will be grouped into another cluster [6]. At the stage of calculating the distance to the cluster, one can use an algorithm, one of which is the Euclidean algorithm. By using this clustering or grouping, it is possible to classify essential oil-producing plants from each sub-district in Bangkalan Regency. The purpose of this clustering is to determine the status of each district.

K-Means Clustering Euclidean

Euclidean is an algorithm that is widely used in determining the value of the shortest distance between two points (straight distance). The following is the K-Means Clustering Eclidean algorithm:

1. Determine the number of existing clusters
2. Choose a cluster center by random method
3. Calculate all distances on each data with the cluster center using the Euclidean formula, namely:

$$d_{in} = \sqrt{\sum_{j=1}^{m} (x_{ij} - k_{nj})^2}$$

Information

din = distance from data to i with cluster center point to n
m = number of features
xij = data to i features to j
knj = cluster center data to n

1. Determine the cluster members, the data will be a member of one of the clusters if it has the smallest distance to other clusters.
2. Group data that has similarities and is a member of each cluster.
3. Update the cluster center value by calculating the average value of each cluster member using the following formula:

$$k_{ij} = \frac{\sum_{h=1}^{n} y_{hj}}{n}$$

Information:

Kij = Center of the ith cluster of the jth feature
n = Sum of all cluster members
$y$hj= data to h features to j in cluster to i

4. Repeat in stages 2-3 so that no data moves to another cluster or reaches the maximum repetition limit that has been determined.

K-Means Clustering Euclidean Flowchart

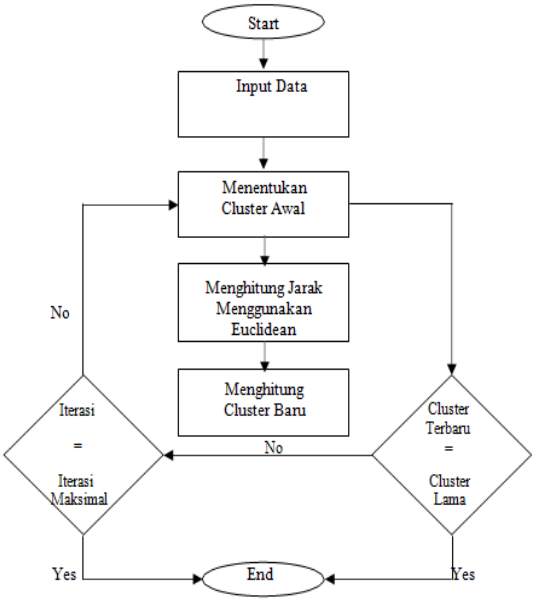

Fig. 1 Flowchart K-Means Clusterung

Figure 1 the Euclidean algorithm shown consists of many stages, namely:

1. Input data on essential oil-producing plants.
2. Determine the initial cluster randomly.
3. Calculating the distance of data to clusters using K-Means Clustering Euclidean.
4. Define a new cluster.
5. Check if the new cluster is the same as the old cluster, the process ends.
6. If the new cluster is not the same as the old cluster, then the iteration is checked until the specified iteration, if yes, the process ends. If during the iteration process there is no then it will return to the process of calculating distances using Euclidean.

Research Methods

This section contains an explanation of the research stages. Figure 2 shows the research design for a geographic information system mapping the distribution area of essential oil-producing plants.

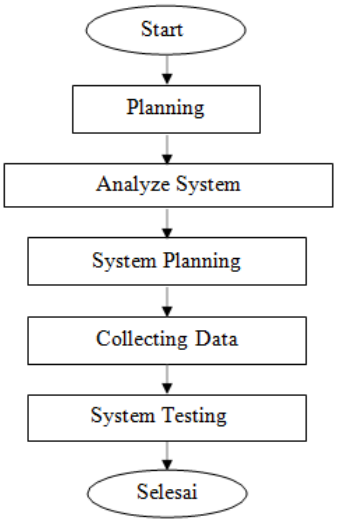

Fig. 2 Research Flow

At this stage a literature search and learning will be carried out from various sources that can help solve the problems found in this research. The collection of literature that will become a reference is in the form of journals, especially those that discuss geographic information systems. For grouping data using the K-Means Clustering method and the Euclidean algorithm.

RESULT AND DISCUSSION

This section will explain the appearance of the system and the testing of the system that has been made. Will be explained about the system wireframe, appearance and features that exist in the system and how it works.

The main page of the system is shown in Figure 3.

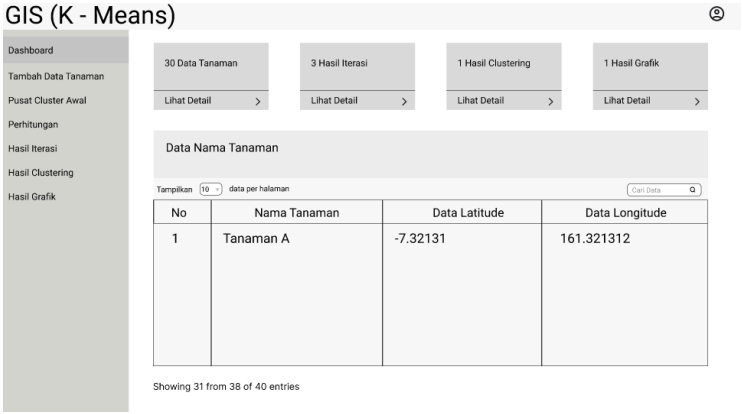

Fig. 3 Dashboard

Figure 3 displays a main page with several menus, namely data pages, calculations, maps and graphs. In addition there is a login feature to enter on the admin login page to manage the system.

The added data page is shown in Figure 4.

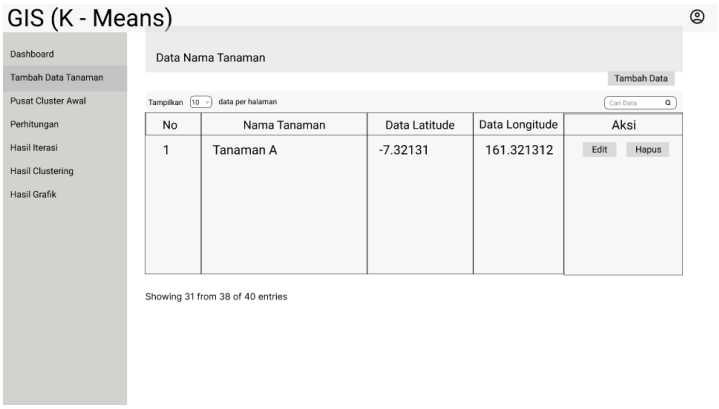

Fig. 4 Add Data Page

Figure 4 displays the menu page for adding plant data. On that page there are several functions, namely adding data, editing data and deleting plant data that can be done by the admin.

The K-means calculation page is shown in Figure 5

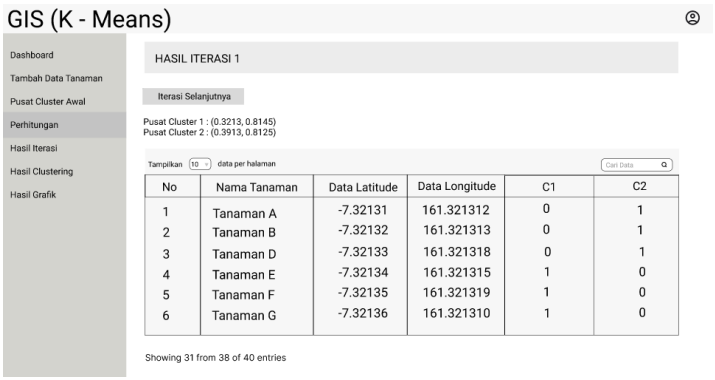

Fig. 5 K-Means Calculation Page

Figure 5 displays the plant data calculation menu. This page displays the iteration results of the calculations, cluster centers and clustering results from applying K-Means to the existing data.

The clustering map is shown in Figure 6.

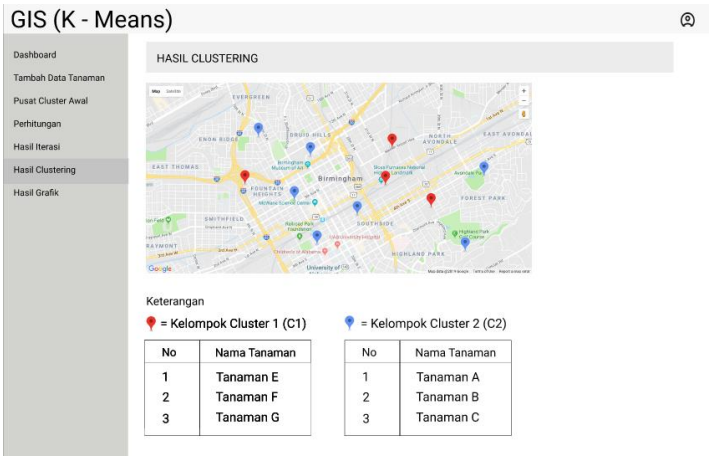

Fig. 6 Clustering Map

Figure 6 shows a map of the distribution of essential oil-producing plants per district in Bangkalan Regency

The clustering chart page is shown in Figure 7

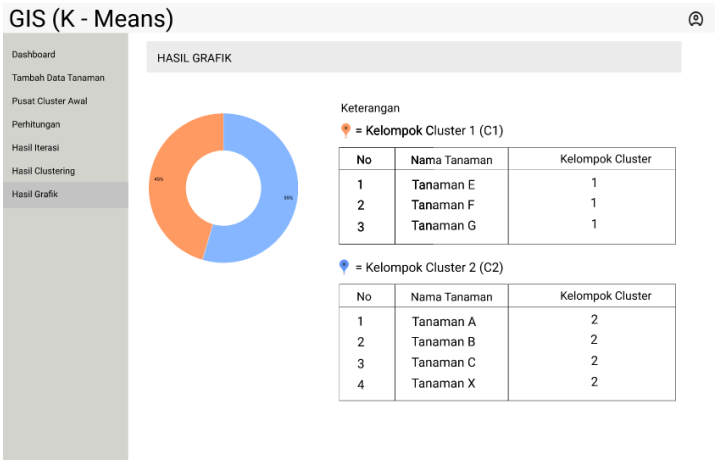

Figure 7 Clustering Graph

Figure 7 shows a graph showing essential oil-producing plant clusters per sub-district in Bangkalan Regency

CONCLUSION

Based on the results of research and field surveys, no specific data was found regarding essential oil-producing plants. As for several alternative discussions, obtained from the Department of Agriculture and Health Office of Bangkalan Regency by utilizing data on family medicinal plants (TOGA) which are then sorted again into data on essential oil-producing plants in Bangkalan Regency. Clustering results show the distribution of medicinal plants in Bangkalan district. The data used is data on the number of essential oil-producing plants scattered in each health center in each sub-district in Bangkalan Regency.

REFERENCES

BPS. Eksport 10 Komoditi Potensial Indonesia, 2017.

N. Mona, “Konsep Isolasi Dalam Jaringan Sosial Untuk Meminimalisasi Efek Contagious(Kasus Penyebaran Virus Corona Di Indonesia),” *J. Sos. Hum. Terap.*, vol. 2, no. 2, pp. 117–125, 2020.

A. N. D. Santoso, “Sistem Informasi Geografis Pengenalan Pariwisata Kabupaten PacitanBerbasis Web Universitas Muhammadiyah Surakarta,” pp. 1–21, 2016.

K. M. Wibowo, K. Indra, and J. Jumadi, “Sistem Informasi Geografis (SIG) MenentukanLokasi Pertambangan Batu Bara di Provinsi Bengkulu Berbasis Website,” *J. Media Infotama*, vol. 11, no. 1, pp. 51–60, 2015, [Online]. Available: https://jurnal.unived.ac.id/index.php/jmi/article/view/252/231.

N. Butarbutar, A. Perdana Windarto, D. Hartama, and Solikhun, “Komparasi Kinerja Algoritma Fuzzy C-Means dan K-Means dalam Pengelompokan Data Siswa BerdasarkanPrestasi Nilai Akademik Siswa (Studi Kasus : SMP Negeri 2 Pematangsiantar),” *J. Ris. Inf. Tek. Inform.*, vol. 1, no. 1, 2016.

R. Buaton, Y. Sundari, and Y. Maulita, “Clustering Tindak Kekerasan Pada Anak Menggunakan

Algoritma K-Means Dengan Perbandingan Jarak Kedekatan Manhattan City Dan Euclidean," *MEANS (Media Inf. Anal. dan Sist.*, vol. 1, no. 2, pp. 47–53, 2016.

S. Yuliani, B. Sudarsono, and A. Wijaya, "Aplikasi Sistem Informasi Geografis (Sig) Untuk Pemetaan Pasar Tradisional Di Kota Semarang Berbasis Web," *J. Geod. Undip*,vol. 5, no. 2, pp. 208–2016, 2016.

K. D. Kemiskinan, "Analisis Kinerja Algoritma Fuzzy C-Means dan K -Means pada Data Kemiskinan," vol. 1, no. 2, 2015.

W. J. Barat, P. Minggu, D. Penelitian, S. Utara, and S. Barat, "Komparasi algoritma kmean dan ahc untuk klasifikasi curah hujan di indonesia," 2017, [Online]. Available: https://media.neliti.com/media/publications/226357-komparasi-algoritma-kmean-dan-ahc-untuk-597b4744.pdf.

E. Fernando, "Sistem Informasi Geografis Untuk Pemetaan Tempat Kesehatan Di KotaJambi," *Tek. Inform.*, no. September 2012, pp. 1–7, 2013, doi: 10.13140/RG.2.1.1476.0405.

M. Destiningrum and Q. J. Adrian, "Sistem Informasi Penjadwalan Dokter Berbassis WebDengan Menggunakan Framework Codeigniter (Studi Kasus: Rumah Sakit Yukum Medical Centre)," *J. Teknoinfo*, vol. 11, no. 2, p. 30, 2017, doi: 10.33365/jti.v11i2.24.

W. N. M. Dj and A. B. Cahyono, "Perancangan Sistem Informasi Geografis Zona NilaiTanah Berbasis Web Menggunakan Leaflet Javascript Library," *Jurnalteknik Its*, vol. 5,no. 2, 2016.

A. K. Wardhani, "K-Means Algorithm Implementation for Clustering of Patients Diseasein Kajen Clinic of Pekalongan," *J. Transform.*, vol. 14, no. 1, p. 30, 2016, doi: 10.26623/transformatika.v14i1.387.

A. H. Ardiansyah, W. Nugroho, N. H. Alfiyah, R. A. Handoko, and M. A. Bakhtiar, "Penerapan Data Mining Menggunakan Metode Clustering untuk Menentukan StatusProvinsi di Indonesia 2020," *Semin. Nas. Inov. Teknol.*, pp. 329–333, 2020.