Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1933−1940
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3165

Grouping Biopharmaceutical Plants in Indonesia Using the K-Means Algorithm: Application of Data Mining in Production

Ridha Maya Faza Lubis[1,*], Jen-Peng Huang[2], Mula Sigiro[3], Joel Panjaitan[4]

[1] Department of Business and Management, Southern Taiwan University of Science and Technology, Taiwan
[2] Department of Information Management, Southern Taiwan University of Science and Technology, Taiwan
[3] Department of Physics Education University of HKBP Nommensen, Medan City, Indonesia
[4] Department of Electrical Engineering Academy of Deli Serdang Engineering, Indonesia
Email: [1,*]db01g208@stust.edu.tw, [2]jehuang@stust.edu.tw, [3]mulasigiro@gmail.com, [4]joel.pandjaitan@gmail.com
Correspondence Author Email: db01g208@stust.edu.tw
Submitted: **16/02/2023**; Accepted: **31/03/2023**; Published**: 31/03/2023**

Abstract-There are various types of biopharmaceutical plants or medicinal plants in Indonesia, including ginger, galangal, kencur, turmeric, lempuyang and curcuma aeruginosa whose production is widespread in various provinces in Indonesia. reached 160 million USD annually. Then the application of data mining used to classify biopharmaceutical plant production data in Indonesia from this study resulted in 2 clusters namely cluster 0 which in this cluster is a cluster with low production value of biopharmaceutical plants in Indonesia, namely the Provinces of West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, Central Kalimantan, Southeast Sulawesi, Maluku, West Papua and Papua. While cluster 1 is the cluster where the production rate of biopharmaceutical plants is the highest in Indonesia, namely the Provinces of North Sumatra, West Java, Central Java, East Java and South Sulawesi. From the results that have been obtained, it is hoped that it will be useful for organizations, groups or individuals engaged in the biopharmaceutical plant sector so that they can review the existing deficiencies and can increase the production of biopharmaceutical plants in each province.

Keywords: Data Mining; Biopharmaceutical Plant Production; K-Means Algorithm

1. INTRODUCTION

The majority of people in Indonesia work in the agricultural industry, specifically in horticulture, plantations, fisheries, animal husbandry, forestry, food crops, and agricultural services. Biopharmaceutical plants in horticulture, or Hortus culture, are one of these agricultural sub-sectors. Horticulture is the production of garden crops. Vegetables are one of the basic food necessities that are in great demand daily and there are many different types of horticulture[1]. Fruit plants are divided into two categories: annual fruit plants and seasonal fruit plants. Indonesia also has ornamental plants that are equally as gorgeous as its fruits and veggies. These plants can be utilized as interior and outdoor decorations for spaces. The final one is biopharmaceutical (medicine) plants. This plant is a sort of horticulture that is good for the human body because Indonesians have long been known to outdo it. It can be used as seasoning and cosmetics in addition to being a medical and cook[2].

Indonesia's production of medicinal plants is heavily dependent on imported medicines, which cost 160 million USD yearly and must be replaced with native products. Indonesia can manufacture more high-quality biopharmaceuticals if it wants to avoid falling behind other nations that are actively contributing to the development of global trends, such as the quick return to nature [3]. One application of Indonesian medicinal plant products in the realm of health is in the agricultural revitalization program, which seeks to raise the level of community welfare[4]. The outcomes of the production of biopharmaceutical plants vary depending on the region in Indonesia, including North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, and other areas [5]. The plants that are used to make biopharmaceuticals include ginger, galangal, kencur, turmeric, lempuyang, and others [6]. Therefore, data on the production of biopharmaceutical plants in Indonesia can be pooled using data mining.

Data mining is a process for obtaining information from enormous amounts of data in order to produce significant and necessary patterns[7]. One of the clustering techniques is the K-Means algorithm [8]. The clustering technique K-Means separates each cluster's data item based on how similar they are to one another. The stages of this approach include extracting clusters from data sets and locating centralized values. The purpose of this study is to classify Indonesia's biopharmaceutical plant production data[9].

Research conducted by Normah et al. in 2021 covers sales analysis for hijab fashion figures by applying the K-means algorithm to Helai retailers to identify which garments are selling the most, selling, and selling clothes. This is only one of many studies that may be utilized as research references. selling poorly. gathering stock data for apparel in a sheet shop to use the k-means method. then by initially randomly choosing 3 clusters. The end conclusion, which is that there are 11 best-selling articles, 55 best-selling articles, and 34 less-selling articles, can be seen once the statistics in each cluster remain constant[10]. The application of the K-Means Clustering Method to construct a superior class classification was the result of research done by ari and eko in 2021 that discussed the determination of superior class pupils with the K-Means algorithm from this challenge. To create two clusters for each class, an information system has to be put in place. There were 96 kids who entered the superior class and 96 students who did not, according to the results of a 6-class cluster at SMPN 2 Jati Agung in South Lampung Regency, which included 192 tested data. This application has a 97.56% (very good) user acceptance rate, making it worthwhile to use[11]. The number of attributes Suspect, Probable, Positive Confirmation, Completed Isolation, and Death in each Regency/City

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1933−1940
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3165

that is not the same led to distinct groupings in the research by Zulfa et al. in 2021 that explored the clustering of Covid 19 instances with the K-means method of these problems. According to the test results, there is a difference between calculating the Davies-Bouldin Index (DBI) value manually and utilizing RapidMiner tools. However, the results of both calculations at the end of the month are both very near to 0, indicating that the clusters being evaluated produce good clusters. In this situation, manual calculations yield better results than those obtained with the aid of RapidMiner tools[12]. Using the k-means method to solve the issue and applying the algorithm to the problem, research by gustientiedina, et al. in 2019 examined the clustering of medications in hospitals. Medicines with low use have an average annual demand of under 18,000 units, drugs with moderate use have an average yearly demand of between 18,000 and 70,000 units, and drugs with high use have an average annual demand of over 70,000 units[13]. A grouping pattern for each cluster on only one type of attribute for loading or unloading of goods was produced by research done by Insanul et al. in 2012 that discussed the grouping of loading and unloading transaction data by applying the k-means algorithm to the problem and implementing the algorithm. The dominant cluster obtained the attribute loading type of goods with the most dominant CPO (Crude Palm Oil) transactions. Additionally, it reveals that Buana Listya Tama TBK, PT agents dominate cluster 3 by up to 18.77% with an average of 2847.22 tonnes in transactions. Additionally, Samudera Sarana Karunia, PT agents up to 17.02% with an average of 2827.32 tons in transactions[14].

The authors intend to undertake research on the grouping of biopharmaceutical plant production data in Indonesia using the k-means algorithm in light of the issues raised above and prior studies in order to acquire comparisons based on correct data and provide precise and impartial comparisons.

2. RESEARCH METHODOLOGY

2.1 Phases of Research

There are specific procedures or stages that must be followed when conducting research; by following these steps, research becomes more precise and objective, and research with the right steps will also provide the right outcomes, resulting in more accurate results. Every research project must have a methodical, logical algorithm or stages that detail each step in the research process.

a. Library Research
   Researchers must gather information during the literature study phase by consulting numerous library books or journal articles that can support their work on the issue under study.
b. Automated Analysis
   In one stage of the research process called algorithm analysis, researchers examine themes that have already been decided upon as well as the algorithms that will be applied.
c. Utilizing the Algorithm
   The researcher then put the algorithm into practice after performing the subsequent analysis. where existing methods are used to process the existing data. The K-means algorithm is one of the algorithms used in this investigation.
d. Algorithm Evaluation
   The K-means algorithm must first be implemented, then testing it is the next necessary step. Algorithm testing is the process of evaluating the outcomes of an algorithm's prior application; if the test's outcomes match those of the algorithm's actual use, the test is deemed successful.
e. Final Illustration
   Drawing conclusions from the research that has been done and condensing them into a single conclusion paragraph is the final step at this point.

Figure 1 below shows images of the study steps based on the preceding explanation:

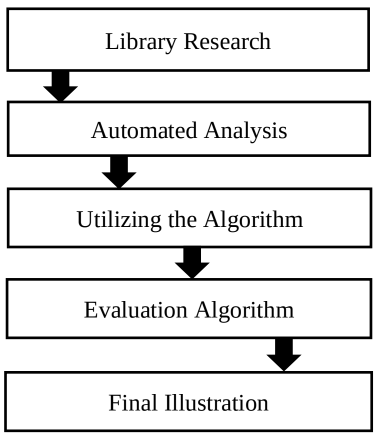

Figure 1. Framework for Research

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1933−1940
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3165

2.2 Data Mining

Data mining is the investigation of a set of data to uncover unanticipated relationships and draw conclusions that are distinct from previous methods. Data mining is the investigation of a set of data to uncover unanticipated relationships and draw conclusions that are distinct from previous methods[15]. Data mining is a procedure or stage that can make use of statistical, artificial intelligence, machine learning, and mathematic methods to find and present knowledge-related information in huge databases[16].

2.3 Biological Plants

One form of a medical plant, known as a biopharmaceutical plant, has the potential to prevent or treat disease as well as be an effective complement to the human body. Ginger, galangal, turmeric, and other plants are examples of biopharmaceutical plants[17].

2.4 K-Means Algorithm

One of the algorithms for data mining clustering that may be used to group data into a partition group is the K-means algorithm[18]. This algorithm solely functions on numeric attributes and uses a distance-based clustering technique that can be separated into multiple clusters[19].
The k-means algorithm goes through the following stages[20]:

a Find the number of clusters, or k, in the data collection.
b calculating the centroid value (center). This value is calculated at random using the available data. By use the subsequent formula:

$$V_{ij}=\frac{1}{N_i}\sum_{k=0}^{N_i} Xkj \quad (1)$$

Information:
V_{ij} = the jth criterion's average centroid for the ith cluster
N_i = the size of the i-th cluster in terms of members
i,k = Cluster index
j = Index of factors
Xkj= The cluster's k-th data value for the j-variable

c Use the Euclidean distance to determine the distance between each object's points and centroid points. Using the following formula:

$$De = \sqrt{(x_i - S_i)^2 + (y_i - t_i)^2} \quad (2)$$

Information:
De = Euclidean Distance
I = The quantity of items
(x, y) = Coordinates of an object
(s, t) = Centroid positions

d Sort items according to their distance from the nearest centroid
e Steps 2 through 4 should be repeated until the centroid is ideal.

3. CONCLUSION AND RESULTS

Ginger, galangal, kencur, turmeric, lempuyang, temulawak, and dindingang are the seven criteria used in the use of the k-means algorithm to categorize the production data of biopharmaceutical plants in Indonesia. There are other 29 alternatives. This information is accurate and was obtained in 2021 from the Indonesian Central Statistics Agency (BPS). Table 1 in the section below contains information on the production of biopharmaceutical plants:

Table 1. Indonesian Biopharmaceutical Plant Production Dataset

No	Province	Ginger	Galangal	Aromatic Ginger	Turmeric	Lempuyang	Curcuma	Curcuma Aeruginosa
1	NORTH SUMATRA	52245300	1099804	1345654	14114104	40779	180468	4412
2	WEST SUMATRA	13772833	3589529	1195254	4354848	2957	132735	3345
3	RIAU	1327441	596679	1249820	1144591	105736	91721	18283
4	JAMBI	2813725	712901	215273	719965	11618	33167	15095
5	SOUTH SUMATRA	1868042	1183599	913895	1741904	20115	74418	28907
6	BENGKULU	16173807	2979195	1730560	6313750	26284	19588	1383
7	LAMPUNG	4085442	918785	5757467	942425	181999	243310	110332
8	KEP. BANGKA BELITUNG	721573	1011835	647319	846065	14480	23995	5000
9	KEP. RIAU	22960	72796	9202	26378	3248	1933	1717
10	DKI JAKARTA	1632	1657	333	600	20	308	50

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1933−1940
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3165

11	WEST JAVA	43833254	19385851	9791218	20047217	1197281	89692	21558
12	CENTRAL JAVA	39087220	13035571	15612280	20272747	958748	5512125	1134815
13	DI YOGYAKARTA	5120703	1225027	1874687	3374060	585421	1485949	766561
14	EAST JAVA	27595251	16303036	3573890	82988205	5096901	23111308	4163118
15	BANTEN	1926635	2635336	1748971	667836	37451	24867	7329
16	BALI	2903556	217370	665072	5228218	500	7098	4
17	WEST NUSA TENGGARA	932036	897798	59277	1401294	41435	51779	3175
18	EAST NUSA TENGGARA	1625741	1013628	216059	1366902	8093	201860	69400
19	WEST KALIMANTAN	2772531	447500	335730	816783	14129	65835	21707
20	CENTRAL KALIMANTAN	592731	339305	110430	220943	31165	18012	5681
21	SOUTH KALIMANTAN	7139996	913078	5550994	2162178	1744	390023	112
22	EAST KALIMANTAN	2441371	274326	136080	525390	6864	75920	80839
23	NORTH KALIMANTAN	657717	1691682	47549	134099	1193	54788	6904
24	CENTRAL SULAWESI	643341	306776	511939	372073	1623	34049	4251
25	SOUTH SULAWESI	60793383	3293700	251133	9808610	13140	45182	773
26	SOUTHEAST SULAWESI	922195	196032	16626	107851	4545	32774	14955
27	MALUKU	665819	257758	57435	362287	1637	987	2679
28	WEST PAPUA	203323	74862	64764	807076	1298	36728	470
29	PAPUA	113016	84069	35413	62437	9319	13280	8876

3.1 Application of the K-Means Algorithm

In the biopharmaceutical plant production data set, there are 29 provinces in Indonesia which annually produce medicinal plant production, and the production results vary each year in each of these provinces. Then the above data processing can be implemented in Rapidminer.

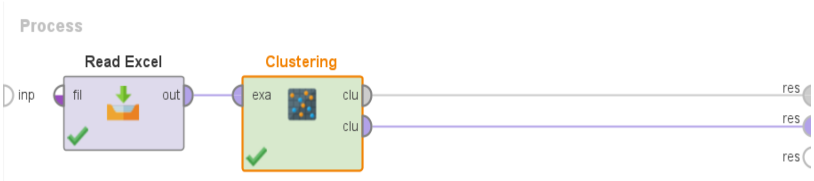

Figure 2. Implementation of biopharmaceutical plant production datasets in Indonesia

Cluster data is generated using the Cluster model, from the data in table 1, there are 2 clusters of them:
Cluster 0 = Which consists of 24 items.
Cluster 1 = Which consists of 5 items.

Cluster Model

```
Cluster 0: 24 items
Cluster 1: 5 items
Total number of items: 29
```

Figure 3. Cluster data using the Cluster Model

Grouping data based on the cluster, from the data in table 1, it is grouped into 2 clusters as can be seen in Figure 4 below:

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1933−1940
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3165

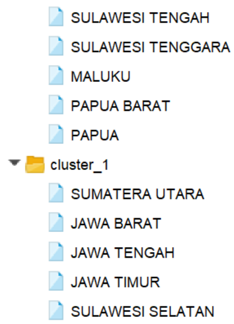

Figure 4. Grouping data based on clusters

The data generated based on the initial data are then grouped into 2 clusters of them:

Cluster 0= consisting of the Provinces of WEST SUMATERA, RIAU, JAMBI, SOUTH SUMATRA, BENGKULU, LAMPUNG, KEP. BANGKA BELITUNG, KEP. RIAU, DKI JAKARTA, IN YOGYAKARTA, BANTEN, BALI, WEST NUSA TENGGARA, EAST NUSA TENGGARA, WEST KALIMANTAN, CENTRAL KALIMANTAN, SOUTH KALIMANTAN, EAST KALIMANTAN, NORTH KALIMANTAN, CENTRAL SULAWESI, SOUTHEAST SULAWESI, MALUKU, WEST PAPUA and PAPUA.

Cluster 1= Consisting of the Provinces of NORTH SUMATRA, WEST JAVA, CENTRAL JAVA, EAST JAVA and SOUTH SULAWESI.

The resulting data plot is based on the initial dataset, which can be seen in the image below:

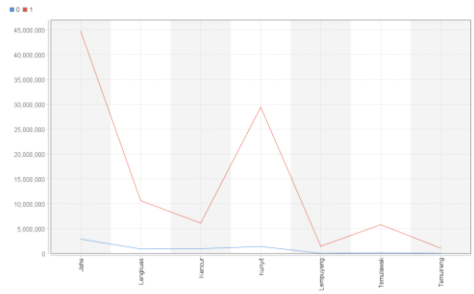

Figure 5. Data plot based on Clusters 0 and 1

The resulting data visualization based on the initial dataset, can be seen in the image below:

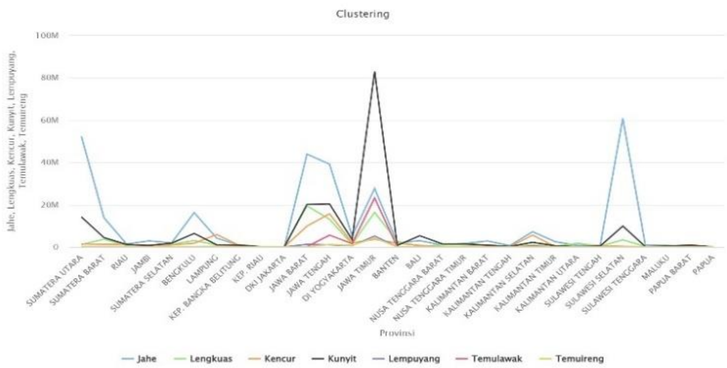

Figure 6. Data visualization

Initial data visualization of biopharmaceutical plant production in Indonesia based on 2 clusters, namely 0 and 1:

1. NORTH SUMATRA Province is included in cluster 1 with production of Ginger 52245300 (tons/year), Galangal 1099804 (tons/year), Kencur 1345654 (tons/year), Turmeric 14114104 (tons/year), Lempuyang 40779 (tons/year), Temulawak 180468 (tons/year) and Temuireng 4412 (tons/year).
2. WEST SUMATERA Province is included in cluster 0 with production of Ginger 13772833 (tons/year), Galangal 3589529 (tons/year), Kencur 1195254 (tons/year), Turmeric 4354848 (tons/year), Lempuyang 2957 (tons/year), Temulawak 132735 (tons/year) and Temuireng 3345 (tons/year).
3. RIAU Province is included in cluster 0 with production of Ginger 1327441 (tons/year), Galangal 596679 (tons/year), Kencur 1249820 (tons/year), Turmeric 1144591 (tons/year), Lempuyang 105736 (tons/year), Temulawak 91721 (tonnes/year) and Temuireng 18283 (tonnes/year).

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1933−1940
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3165

4. JAMBI Province is included in cluster 0 with production of Ginger 2813725 (tons/year), Galangal 712901 (tons/year), Kencur 215273 (tons/year), Turmeric 719965 (tons/year), Lempuyang 11618 (tons/year), Temulawak 33167 (tons/year) and Temuireng 15095 (tons/year).
5. SELATAN SUMATERA Province is included in cluster 0 with production of Ginger 1868042 (tons/year), Galangal 1183599 (tons/year), Kencur 913895 (tons/year), Turmeric 1741904 (tons/year), Lempuyang 20115 (tons/year), Temulawak 74418 (tons/year) and Temuireng 28907 (tons/year).
6. BENGKULU Province is included in cluster 0 with production of Ginger 16173807 (tons/year), Galangal 2979195 (tons/year), Kencur 1730560 (tons/year), Turmeric 6313750 (tons/year), Lempuyang 26284 (tons/year), Temulawak 19588 (tonnes/year) and Temuireng 1383 (tonnes/year).
7. LAMPUNG Province is included in cluster 0 with production of Ginger 4085442 (tons/year), Galangal 918785 (tons/year), Kencur 5757467 (tons/year), Turmeric 942425 (tons/year), Lempuyang 181999 (tons/year), Temulawak 243310 (tons/year) and Temuireng 110332 (tons/year).
8. KEP Province. BANGKA BELITUNG is included in cluster 0 with production of Ginger 721573 (tons/year), Galangal 1011835 (tons/year), Kencur 647319 (tons/year), Turmeric 846065 (tons/year), Lempuyang 14480 (tons/year), Temulawak 23995 (tons/year) and Temuireng 5000 (tons/year).
9. KEP Province. RIAU is included in cluster 0 with production of Ginger 22960 (tons/year), Galangal 72796 (tons/year), Kencur 9202 (tons/year), Turmeric 26378 (tons/year), Lempuyang 3248 (tons/year), Temulawak 1933 (ton/year) and Temuireng 1717 (ton/year).
10. DKI JAKARTA Province is included in cluster 0 with production of Ginger 1632 (tons/year), Galangal 1657 (tons/year), Kencur 333 (tons/year), Turmeric 600 (tons/year), Lempuyang 20 (tons/year), Temulawak 308 (tonnes/year) and Temuireng 50 (tons/year).
11. WEST JAVA Province is included in cluster 1 with production of Ginger 43833254 (tons/year), Galangal 19385851 (tons/year), Kencur 9791218 (tons/year), Turmeric 20047217 (tons/year), Lempuyang 1197281 (tons/year), Temulawak 89692 (tons/year) and Temuireng 21558 (tons/year).
12. CENTRAL JAVA Province is included in cluster 1 with production of Ginger 39087220 (tons/year), Galangal 13035571 (tons/year), Kencur 15612280 (tons/year), Turmeric 20272747 (tons/year), Lempuyang 958748 (tons/year), Temulawak 5512125 (tons/year) and Temuireng 1134815 (tons/year).
13. The Province of DI YOGYAKARTA is included in cluster 0 with production of Ginger 5120703 (tons/year), Galangal 1225207 (tons/year), Kencur 1874687 (tons/year), Turmeric 3374060 (tons/year), Lempuyang 585421 (tons/year), Temulawak 1485949 (tons/year) and Temuireng 766561 (tons/year).
14. EAST JAVA Province is included in cluster 1 with production of Ginger 27595251 (tons/year), Galangal 16303036 (tons/year), Kencur 3573890 (tons/year), Turmeric 82988205 (tons/year), Lempuyang 5096501 (tons/year), Temulawak 23111308 (tons/year) and Temuireng 4163118 (tons/year).
15. BANTEN Province is included in cluster 1 with production of Ginger 1926635 (tons/year), Galangal 2635336 (tons/year), Kencur 1748971 (tons/year), Turmeric 667836 (tons/year), Lempuyang 37451 (tons/year), Temulawak 24867 (tonnes/year) and Temuireng 7329 (tonnes/year).
16. BALI Province is included in cluster 0 with production of Ginger 2903556 (tons/year), Galangal 217370 (tons/year), Kencur 665072 (tons/year), Turmeric 5228218 (tons/year), Lempuyang 500 (tons/year), Temulawak 7098 (tons/year) and Temuireng 4 (tons/year).
17. The province of NUSA TENGGARA BARAT is included in cluster 0 with production of Ginger 932036 (tons/year), Galangal 897798 (tons/year), Kencur 59277 (tons/year), Turmeric 1401294 (tons/year), Lempuyang 41435 (tons/year), Temulawak 51779 (tons/year) and Temuireng 3175 (tons/year).
18. The Province of NUSA TENGGARA TIMUR is included in cluster 0 with production of Ginger 1625741 (tons/year), Galangal 1013628 (tons/year), Kencur 216059 (tons/year), Turmeric 1366902 (tons/year), Lempuyang 8093 (tons/year), Temulawak 201860 (tons/year) and Temuireng 69400 (tons/year).
19. West Kalimantan Province is included in cluster 0 with production of Ginger 2772531 (tons/year), Galangal 447500 (tons/year), Kencur 335730 (tons/year), Turmeric 816783 (tons/year), Lempuyang 14129 (tons/year), Temulawak 65835 (tons/year) and Temuireng 21707 (tons/year).
20. CENTRAL KALIMANTAN Province is included in cluster 0 with production of Ginger 592731 (tons/year), Galangal 339305 (tons/year), Kencur 110430 (tons/year), Turmeric 220943 (tons/year), Lempuyang 31165 (tons/year), Temulawak 18012 (tons/year) and Temuireng 5681 (tons/year).
21. SELATAN KALIMANTAN Province is included in cluster 0 with production of Ginger 7139996 (tons/year), Galangal 913078 (tons/year), Kencur 5550994 (tons/year), Turmeric 2162178 (tons/year), Lempuyang 1744 (tons/year), Temulawak 390023 (tons/year) and Temuireng 112 (tons/year).
22. North Kalimantan Province is included in cluster 0 with production of Ginger 2441371 (tons/year), Galangal 274326 (tons/year), Kencur 136080 (tons/year), Turmeric 535390 (tons/year), Lempuyang 6864 (tons/year), Temulawak 75920 (tons/year) and Temuireng 80839 (tons/year).
23. North Kalimantan Province is included in cluster 0 with production of Ginger 657717 (tons/year), Galangal 1691682 (tons/year), Kencur 47549 (tons/year), Turmeric 134099 (tons/year), Lempuyang 1193 (tons/year), Temulawak 54788 (tons/year) and Temuireng 6904 (tons/year).

24. CENTRAL SULAWESI Province is included in cluster 0 with production of Ginger 643341 (tons/year), Galangal 306776 (tons/year), Kencur 511939 (tons/year), Turmeric 372073 (tons/year), Lempuyang 1623 (tons/year), Temulawak 34049 (tonnes/year) and Temuireng 4251 (tonnes/year).
25. SELATAN SULAWESI Province is included in cluster 1 with production of Ginger 60793383 (tons/year), Galangal 3293700 (tons/year), Kencur 251133 (tons/year), Turmeric 9808610 (tons/year), Lempuyang 13140 (tons/year), Temulawak 45182 (tons/year) and Temuireng 773 (tons/year).
26. The Province of SOUTHEAST SULAWESI is included in cluster 0 with production of Ginger 922195 (tons/year), Galangal 196032 (tons/year), Kencur 16626 (tons/year), Turmeric 107851 (tons/year), Lempuyang 4545 (tons/year), Temulawak 32774 (tons/year) and Temuireng 14955 (tons/year).
27. MALUKU Province is included in cluster 0 with production of Ginger 665819 (tons/year), Galangal 257758 (tons/year), Kencur 57435 (tons/year), Turmeric 362287 (tons/year), Lempuyang 1637 (tons/year), Temulawak 987 (tonnes/year) and Temuireng 2697 (tonnes/year).
28. WEST PAPUA Province is included in cluster 0 with production of Ginger 203323 (tons/year), Galangal 74862 (tons/year), Kencur 64764 (tons/year), Turmeric 807076 (tons/year), Lempuyang 1298 (tons/year), Temulawak 36728 (tons/year) and Temuireng 470 (tons/year).
29. PAPUA Province is included in cluster 0 with production of Ginger 113016 (tons/year), Galangal 84069 (tons/year), Kencur 35413 (tons/year), Turmeric 62437 (tons/year), Lempuyang 9319 (tons/year), Temulawak 13280 (tonnes/year) and Temuireng 8867 (tonnes/year).

By applying data mining using the K-means algorithm, it can determine which provinces in Indonesia have the highest and lowest production of biopharmaceutical plants. Cluster 0 is the province with the lowest production rate of biopharmaceutical plants while cluster 1 is the province with the highest production rate of biopharmaceutical plants in Indonesia.

4. CONCLUSION

Based on research that has been carried out using accurate data on biopharmaceutical production in Indonesia as well as the correct stages of the k-means data mining algorithm and by testing using Rapidminer Studio, 2 clusters are obtained, namely cluster 0 which in this cluster is a cluster with a production value Low biopharmaceutical plants in Indonesia, namely the provinces of West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, Central Kalimantan, Southeast Sulawesi, Maluku, West Papua and Papua. While cluster 1 is the cluster where the production rate of biopharmaceutical plants is the highest in Indonesia, namely the Provinces of North Sumatra, West Java, Central Java, East Java and South Sulawesi. From the results that have been obtained, hopefully it can be useful for organizations, groups or individuals engaged in the biopharmaceutical plant sector so that they can review the existing deficiencies and can increase the production of biopharmaceutical plants in each province.

REFERENCES

[1] R. E. PRANG, F. TULUSAN, and V. LONDA, "IMPLEMENTASI PROGRAM SENTRA HORTIKULTURA DI DESA WULURMAATUS KECAMATAN MODOINDING KABUPATEN MINAHASA SELATAN," vol. IX, no. 127, pp. 80–88, 2023.
[2] R. Khalida, "Analisa Komparasi Tiga Metode Data Mining dalam Prediksi Impor Komoditas Tanaman Biofarmaka," vol. 19, pp. 145–154, 2020.
[3] Y. Yanti, H. Hamid, D. Suhendra, Juniarti, and I. R. Putra, "Prosiding seminar nasional penelitian dan pengabdian kepada masyarakat (snpp) tahun 2022 "," pp. 39–46, 2022.
[4] M. Marini, I. Maesaroh, A. Darotulmutmainnah, and N. Oktavia, "Pendampingan Penanaman dan Pemanfaatan Tanaman Obat Keluarga (TOGA) Dalam Program Peningkatan Peranan Wanita menuju Keluarga Sehat Sejahtera (P2WKSS) di Desa Sakerta Timur Kabupaten Kuningan Assistance in Planting and Utilizing Family Medicinal Plants (TOGA) in the program to Increase the Role of Women to wards Healthy Prosperous Families (P2WKSS) in East Sakerta Village , Kuningan Regency," vol. 2, pp. 9–16, 2023.
[5] B. Moray, Y. D. Y. Rindengan, and A. S. M. Lumenta, "Development Of Horticiltural Commodity Price Website," pp. 1–5.
[6] T. Trsnawati, "Produk tanaman biofarmaka," 2021.
[7] H. Santoso, I. P. Hariyadi, and Prayitno, "Data Mining Analisa Pola Pembelian Produk," *Tek. Inform.*, no. 1, pp. 19–24, 2016.
[8] I. W. S. A. Nugraha and Mahasiswa, "Clustering Pemetaan Tingkat Kemiskinan di Provinsi Jawa Barat Menggunakan Algoritma K-Means," vol. 9, no. 2, pp. 234–244, 2023.
[9] N. Sirait, "IMPLEMENTASI K-MEANS CLUSTERING PADA PENGELOMPOKAN MUTU BIJI SAWIT (Studi Kasus : PT . Multimas Nabati Asahan)," *J. Pelita Inform.*, vol. 6, no. 2, pp. 170–174, 2017.
[10] N. Normah, S. Nurajizah, and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," *J. Tek. Komput. AMIK BSI*, vol. 7, no. 2, pp. 158–163, 2021.
[11] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan," vol. 15, no. 2, pp. 25–36, 2020.
[12] A. Zulfandi, "Penerapan Kombinasi Metode Entropy dan Extended Promethee II Dalam Menentukan Kepala Jurusan Pada SMK," vol. 3, no. 1, 2021, doi: 10.47065/josyc.v3i1.821.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1933−1940
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3165

[13] Gustientiedina, M. H. Adiya, and Y. Desnelita, "Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," vol. 01, pp. 17–24, 2019.
[14] I. Kamila *et al.*, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," vol. 5, no. 1, pp. 119–125, 2019.
[15] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.
[16] F. R. Pare, O. Wati, L. P. Taran, and L. M. Arsai, "G-Tech : Jurnal Teknologi Terapan," vol. 7, no. 1, pp. 255–261, 2023.
[17] R. Satriani, A. Hartati, and S. G. Sukmaya, "PENGUATAN PASAR DAN DIVERSIFIKASI OLAHAN MINUMAN BERBAHAN DASAR PRODUK BIOFARMAKA PADA KELOMPOK TANI SIDO JODHO KABUPATEN BANJANEGARA," vol. 2, no. 9, pp. 6261–6266, 2023.
[18] N. Pulungan, S. Suhada, and D. Suhendro, "Penerapan Algoritma K-Medoids Untuk Mengelompokkan Penduduk 15 Tahun Keatas Menurut Lapangan Pekerjaan Utama," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, 2019.
[19] S. W. Harjono, N. W. Utami, I. G. Agung, and P. Dwi, "Klasterisasi Tingkat Penjualan pada Startup Panak . id dengan Algoritma K-Means," vol. 17, no. 1, pp. 55–66, 2023.
[20] M. Rosadi *et al.*, "Clustering Panjang Ruas Jalan di BBPJN Sumut Menggunakan Algoritma K-Means," vol. 02, no. 1, pp. 29–38, 2023.