Comparison Of K-Means and K-Medoids Algorithm for Data Clustering UMKM in Pagar Alam City

Sendy Ariska [1]*, Desi Puspita [2], Inda Anggraini [3]
Institut Teknologi Pagar Alam
Pagar Alam, Indonesia
sendyariska70@gmail.com [1], desiofira1@gmail.com [2], indaanggraini@gmail.com [3]

Abstract*— The aim of this research is clustering MSME data in Pagar Alam City using the K-Means and K-Medoids algorithms. This research is motivated by the lack of further management of MSME data collection, which can hinder the development and improvement of Pagar Alam City MSMEs. Meanwhile, this data is considered necessary for agencies to develop and improve Pagar Alam City MSMEs. Apart from agencies, this data is also useful for sub-districts, sub-districts and RT/RW to find out what interests, talents and potential the community has in what business fields. MSME data is processed using Rapid Miner and Python, the system development method in this research uses the Cross Industry Standard Process for Data Mining (CRISP-DM) method, where the stages include Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The test method uses the Davies-bouldin index, a DBI value that is close to 0 results in good clustering. The results of this research obtained 3 clusters. In 2020 K-Means C0= 1, C1= 3 and C2= 1 sub-district, K-Medoids C0= 1, C1= 1 and C2= 3 sub-district. In 2022 K-Means C0= 1, C1= 3 and C2= 1 sub-district, K-Medoids C0= 1, C1= 3 and C2= 1 sub-districts. The results of the 2020 sub-district DBI clustering calculation are DBI k-means = 0.134 and k-medoids = 0.523. In 2022 DBI k-means = 0.277 and k-medoids = 0.496. So it can be concluded that the K-Means algorithm in the case of grouping MSMEs in Pagar Alam City has better performance, because the DBI value is close to 0. From the results of the grouping it can help provide an overview for related parties in encouraging or providing assistance to sub-districts that are included in the low cluster.

Keywords*— K-Means, K-Medoids, CRISP-DM, Davies-bouldin index.

I. INTRODUCTION

Along with the development of the internet, the data stored, both in the form of text, images, sound and video, has also increased very quickly and significantly. As a result, large volumes of data will become "garbage" in storage if they are not processed into useful information, requiring a technique or method called data mining. [1].

Data mining, also called Knowledge Discovery in Databases (KDD), is defined as the extraction of potential, implicit and unknown information from a set of data. The Knowlegde Discovery in Database process involves the results of the process of extracting trends in a data pattern, then converting the results accurately into information that is easy to understand. There are several roles that can be used in data mining, one of which is clustering [2].

Clustering is a method of analyzing data, which is often included as one of the data mining methods, the aim of which is to group data with different characteristics into other "regions" [3], One method for clustering is K-Means and K-Medoids.

K-Means algorithm is a clustering algorithm that groups data based on the closest cluster center point (Centroid) to the data [4]. K-Medoids algorithm or Partitioning Around Medoids (PAM) is a partition clustering method for grouping a set of (n) objects into a number of (k) clusters [5].

Based on the results of observations and interviews in the cooperative sector at the Department of Industry, Trade and UKM Cooperatives in Pagar Alam City. UMKM in Pagar Alam City have experienced an increase in number and various types of business fields. Relevant parties at the Department of Industry, Trade and UKM Cooperatives in Pagar Alam City collected 2 times UMKM data in 1 year which was collected directly through face-to-face meetings with respondents consisting of MSME actors throughout Pagar Alam City. The latest data for UMKM in Pagar Alam City in 2022 has a total of 2,906 UMKMs from 5 sub-districts. The UMKM data collected is currently used as accurate reference data in making government policies to develop Micro, Small and Medium Enterprises (UMKMs) in Pagar Alam City. Based on the data obtained, there are various types of business sectors for UMKM in Pagar Alam City in several sub-districts, from this data collection there is no further management which can hinder the development and improvement of UMKM in Pagar Alam City. Meanwhile, this data is considered necessary for agencies to develop and improve Pagar Alam City UMKMs. Apart from agencies, this data is also useful for sub-districts, sub-districts and RT/RW to be able to find out what interests, talents and potential the community has in the MSME business sector, so that it can be used as a strategy to improve and develop UMKMs in order to support the economy of the actors. the UMKM business. Therefore, there is a need for further data management so that it can be used as policy decisions in improving and developing MSMEs. With the large number of UMKM data and the number of sub-districts, it is necessary to group the UMKM data in Pagar Alam City to deiteirminei thei high and low leiveils of thei numbeir of UMKMs baseid on sub-districts, and thei groups with thei most UMKM busineiss seictors in Pagar Alam City.

Thein, data mining is neieideid to bei proceisseid into useiful information using thei clusteiring meithod using a

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2090, Copyright ©2024
Submitted : February 12, 2024, Revised : March 1, 2024, Accepted : March 19, 2024, Published : June 15, 2024

comparison of thei K-Meians and K-Meidoids algorithms to deiteirminei thei grouping of UMKM data baseid on thei high and low leiveils of thei numbeir of UMKMs in thei sub-district and thei numbeir of eixisting busineiss fieilds. By grouping data on UMKMs in Pagar Alam City, it can heilp ageincieis to focus on sub-districts which still havei low numbeirs of UMKMs, to providei assistancei in improving and deiveiloping UMKMs in Pagar Alam City. Apart from ageincieis, it is also useiful for sub-districts, sub-districts and RT/RW to focus on thei inteireists, taleints and poteintial of thei community in thei higheist numbeir of busineiss fieilds, to providei assistancei so that theiy can bei furtheir improveid and deiveilopeid, eispeicially saleis reisults, in ordeir to support thei eiconomy of UMKM busineiss actors..

Clusteiring proceiss using rapid mineir application havei a reisult of 53 peirceint for clusteir 1 with a total of 8 data, 40 peirceint of thei data for clusteir 2 and 7 peirceint for clusteir 3. [6]

Clusteiring proceiss using Rapidmineir application wit K-Meians and K-meidoids algorithm Baseid on reiseiarch conducteid by [7] with thei titlei "Comparison of thei K-Meians Algorithm and thei K-Meidoid Algorithm for Grouping MSMEis in Kabumein" with thei reiseiarch reisults forming 3 clusteirs, both using thei K-Meians algorithm and thei K-Meidoid algorithm. Thei K-Meians algorithm has 21 sub-districts in clusteir 1, 3 sub-districts in clusteir 2, and 2 sub-districts in clusteir 3. Thei K-Meidoid algorithm in clusteir 1 has 19 sub-districts, clusteir 2 has 5 sub-districts and clusteir 3 has 2 sub-districts. Baseid on thei DBI valuei, thei K-Meians algorithm is 0.324 and thei K-Meidoid algorithm is 0.536. So it can bei concludeid that thei K-Meians algorithm in thei casei of thei Keibumein UMKM grouping has beitteir peirformancei, beicausei thei DBI valuei is closei to 0..

Baseid on reiseiarch conducteid by [8] with thei titlei "Comparison of K-Meians and K-Meidoids Algorithms for Clusteiring Flood Pronei Areias in Rokan Hilir Reigeincy" with reisults with reisults. Afteir conducting eixpeirimeints using RapidMineir tools on clusteirs K=2 to K=6, it was found that thei K-Meians meithod was morei optimal than using thei K-Meidoids meithod on flood eiveint data in Rokan Hilir in 2019-2022 with thei most optimal numbeir of k beiing k = 3 with a Davieis-Bouldin Indeix (DBI) validity valuei of 0.218%, whilei thei optimal clusteir K-Meidoids algorithm is found in K =4 with a validity valuei of 0.525%.

From thei threiei preivious studieis abovei, it can bei concludeid that reiseiarcheirs will carry out thei clusteiring proceiss of UMKM data in Pagar Alam City using thei K-Meians algorithm and thei K-Meidoids algorithm which weirei deisigneid and built computeirizeid to find out thei clusteiring reisults of UMKM data, so that with this systeim can makei it eiasieir for reilateid partieis to makei morei objeictivei policy deicisions in improving, deiveiloping and knowing thei probleims faceid by UMKM busineiss actors. Baseid on thei background deiscription abovei, thei reiseiarcheir wants to conduct reiseiarch with thei titlei "Comparison of K-Meians and K-Meidoids Algorithms for Clusteiring UMKM Data in Pagar Alam City".

II. LITERATURE REVIEW

A. UMKM

Micro, small and meidium einteirpriseis (MSMEis) arei pillars of thei Indoneisian eiconomy that neieid atteintion beicausei theiy can absorb labor and reiducei uneimploymeint amidst compeitition for formal seictor jobs. Many small busineisseis arei foundeid by thei community. Thei Ministry of Coopeirativeis and MSMEis targeits to increiasei thei rolei of MSMEis in national eiconomic growth. Seitting up this busineiss is veiry eiasy and doeis not reiquirei a largei amount of capital. Eimpoweiring MSMEis is a strateigic choicei to increiasei thei incomei of low-incomei groups, in ordeir to reiducei incomei gaps and poveirty through increiasing busineiss capacity and busineiss manageimeint skills. [9].

B. Data Processing

RapidMineir is opein sourcei softwarei. RapidMineir is a solution for analyzing data mining, teixt mining and preidictivei analysis. RapidMineir useis various deiscriptivei and preidictivei teichniqueis to providei insights to useirs so theiy can makei thei beist deicisions. RapidMineir has approximateily 500 data mining opeirators, including opeirators for input, output, data preiproceissing and visualization. RapidMineir is writtein using Java languagei so it can work on all opeirating systeims [10].

Colaboratory, or “Colab” is a product of Googlei Reiseiarch. Colab allows anyonei to writei and eixeicutei arbitrary python codei through thei browseir, and is peirfeict for machinei leiarning, data analysis, and eiducation. Morei teichnically, Colab is a hosteid Jupyteir noteibook seirvicei that can bei useid without seitup, and provideis freiei acceiss to computing reisourceis including GPUs. Colab reisourceis arei not guaranteieid and arei limiteid in naturei, and theiir usagei limits may fluctuatei. This is neiceissary so that Colab can providei reisourceis for freiei [11].

Thei Davieis Bouldin Indeix (DBI) is a clusteir validation introduceid by D.L. Davieis and Donald W. Bouldin, theireiforei thei namei of this meithod is a combination of thei nameis of thei two, nameily Davieis-Bouldin. DBI is onei way to analyzei thei quality of clusteirs in eiach clusteiring [12]

III. RESEARCH METDODS

This research was conducted at the Department of Industry, Trade, Cooperatives and SMEs of Pagar Alam City, Jl. Laskar Wanita Mentarjo, Pagar Wangi, Kec. North Dempo, Pagar Alam City, South Sumatra.

The research time starts from November 2023 to December 2023 at the Department of Industry, Trade and UKM Cooperatives in Pagar Alam City

CRISP-DM (Cross Industry Standard Process for Data Mining) is a standardization of data mining processing that has been developed where existing data will pass through each structured and defined phase clearly and efficiently. In addition to applying a model in the data mining process, algorithm selection greatly affects the performance comparison of data mining methods. CRISP-DM data mining

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2090, Copyright ©2024
Submitted : February 12, 2024, Revised : March 1, 2024, Accepted : March 19, 2024, Published : June 15, 2024

methodology as a common problem solver for business and research. This methodology consists of six stages, namely Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (Hasanah et al., 2021).

The research methodology used by research refers to the CRIPS-DM data mining balancing method with the following flow:

Fig 1. CRISP-DM

CRISP-DM (*Cross Industry Standard Process for Data Mining) is* a standardization of data mining processing that has been developed where existing data will go through each structured and defined phase clearly and efficiently. Apart from applying a model in the data mining process, the choice of algorithm greatly influences the performance comparison of data mining methods. CRISP-DM data mining methodology as a common problem solver for business and research. This methodology consists of six stages, namely Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment, the explanation is as follows [13] ;

1) *Business Understanding*
   Several things are done at this stage, such as understanding needs and goals from a business perspective, then interpreting knowledge in the form of defining problems in data mining and then determining plans and strategies to achieve data mining goals.
2) *Data Understanding*
   This stage begins with collecting data, describing the data, and evaluating the quality of the data.
3) *Data Preparation*
   In this stage, we build the final dataset from raw data. There are several things that will be done, including cleaning data (Data Cleaning), selecting data (Data Selection), records and attributes, and also carrying out transformation of the data (Data Transformation) to be used as input in the modeling stage..
4) *Modelling*
   This stage directly involves Machine Learning to determine data mining techniques, data mining tools and data mining algorithms.
5) *Evaluation*
   This stage is carried out by looking at the performance level of the patterns produced by the algorithm.
6) *Deployment*
   This stage is carried out by creating reports and journal articles using the resulting model.

IV. RESULT AND DISCUSSION

Clustering based on sub-districts in 2020 there are 11 attributes at the data selection stage including sub-districts, services are initialized with number 1, industry with number 2, trade with number 3, agriculture with number 4, livestock with number 5, crafts with number 6, plantations with number 7, fisheries with number 8, hotels / inns with number 9, restaurants with number 10 and communication with number 11.

TABLE I. K-Means District Clustering Results 2020

District	1	2	3	4	5	6	7	8	9	10	11
Dempo Selatan	86	24	8	8	2	27	1	0	0	0	0
Dempo Tengah	132	7	25	10	40	8	2	0	0	0	0
Dempo Utara	133	16	16	21	8	2	16	10	3	0	0
Pagar Alam Selatan	310	163	74	13	10	7	1	3	5	7	1
Pagar Alam Utara	230	217	62	18	10	5	1	2	4	3	0

Based on the results of clustering sub-districts in 2020, cluster_0 with a high level, cluster_1 with a low level and cluster_2 with a medium level. Based on k-means, 3 clusters are obtained. First cluster_0 = 1 sub-district, namely Pagar Alam Selatan sub-district. Second, cluster_1 = 3 sub-districts, namely North Dempo, Central Dempo and South Dempo. Third, cluster_2 = 1 sub-district, namely Pagar Alam Utara. With the average performance vector in the centroid distance obtained value 58.509, then the average centroid cluster_0 with a value of 0.000, the average centroid cluster_1 with a value of 97.515, the average cluster_2 with a value of 0.000 and the davies bouldin index value 0.012

TABLE II. K-Medoids District Clustering Results 2020

District	1	2	3	4	5	6	7	8	9	10	11
Dempo Selatan	86	24	8	8	2	27	1	0	0	0	0
Dempo Tengah	132	7	25	10	40	8	2	0	0	0	0
Dempo Utara	133	16	16	21	8	2	16	10	3	0	0
Pagar Alam Selatan	310	163	74	13	10	7	1	3	5	7	1
Pagar Alam Utara	230	217	62	18	10	5	1	2	4	3	0

While the clustering results in 2020 with the k-medoids algorithm were also formed into 3 clusters. cluster_0 with a high level, cluster_1 with a low level and cluster_2 with a medium level. . First cluster_0 = 1 sub-district, namely Pagar Alam Utara sub-district. Second, cluster_1 = 1 sub-district, namely Pagar Alam Selatan. Third, cluster_2 = 3 sub-districts namely North Dempo, Central Dempo and South Dempo. With the average performance vector in the centroid

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2090, Copyright ©2024
Submitted : February 12, 2024, Revised : March 1, 2024, Accepted : March 19, 2024, Published : June 15, 2024

distance obtained value of 145.545, then the average centroid cluster_0 with a value of 0.000, the average centroid cluster_1 with a value of 0.000, the average cluster_2 with a value of 242.576 and a davies bouldin index value of 0.015.

Clustering based on sub-districts in 2022 there are 15 attributes at the data selection stage including sub-districts, grocery traders with number 1, trade with number 2, food traders with number 3, industry with number 4, services with number 5, vegetable traders with number 6, crafts with number 7, livestock with number 8, restaurants with number 9, plantations with number 10, fish traders with number 11, fruit traders with number 12, fisheries with number 13, agriculture with number 14 and cosmetics with number 15.

KECAMATAN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Dempo Selatan	218	20	6	4	5	8	19	0	0	0	0	0	1	2	0
Dempo Tengah	419	43	21	26	24	11	6	2	0	0	1	0	0	0	0
Dempo Utara	79	39	22	15	25	53	4	7	0	2	3	5	5	0	1
Pagar Alam Selatan	223	356	255	199	156	86	7	12	6	0	5	4	2	1	1
Pagar Alam Utara	102	99	129	18	35	26	0	6	14	11	3	1	0	3	1

Fig 2. District K-Means Clustering Results 2022

Based on the results of clustering sub-districts in 2022, cluster_0 with a high level, cluster_1 with a low level and cluster_2 with a medium level. Based on k-means, 3 clusters are obtained. First cluster_0 = 1 sub-district, namely Pagar Alam Selatan sub-district. Second, cluster_1 = 3 sub-districts, namely Pagar Alam Utara, Dempo Utara, Dempo Selatan. Third, cluster_2 = 1 sub-district, namely Central Dempo. With the average performance vector in the centroid distance obtained value 340.222, then the average centroid cluster_0 with a value of 0.000, the average centroid cluster_1 with a value of 567.037, the average cluster_2 with a value of 0.000 and a davies bouldin index value of 0.019..

KECAMATAN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Dempo Selatan	218	20	6	4	5	8	19	0	0	0	0	0	1	2	0
Dempo Tengah	419	43	21	26	24	11	6	2	0	0	1	0	0	0	0
Dempo Utara	79	39	22	15	25	53	4	7	0	2	3	5	5	0	1
Pagar Alam Selatan	223	356	255	199	156	86	7	12	6	0	5	4	2	1	1
Pagar Alam Utara	102	99	129	18	35	26	0	6	14	11	3	1	0	3	1

Fig 3. K-Medoids Clustering Results in 2022

While the clustering results in 2022 with the k-medoids algorithm are also formed into 3 clusters. cluster_0 with a high level, cluster_1 with a low level and cluster_2 with a medium level. . First cluster_0 = 1 sub-district, namely Pagar Alam Selatan sub-district. Second cluster_1 = 3 sub-districts namely Pagar Alam Utara, Dempo Utara and Dempo Selatan. Third, cluster_2 = 1 sub-district, namely Central Dempo. With the average performance vector in the centroid distance obtained value of 797,200, then the average centroid cluster_0 with a value of 0.000, the average centroid cluster_1 with a value of 1328,667, the average cluster_2 with a value of 0.000 and a davies bouldin index value of 0.030.

davies bouldin index results on k-means algorithm and k-medoids algorithm based on clustering conducted on MSME data in Pagar Alam City, evaluation on google colab.Tabeil 5. comparison of DBI K-Meians and K-Meidoids

TABLE III. COMPARISON of DBI K-Means and K- Medoids

Clustering	Algorithm	date	DBI
Kecamatan	*K-Means*	2020	0.134
Kecamatan	*K-Medoids*	2020	0.583
Kecamatan	*K-Means*	2022	0.277
Kecamatan	*K-Medoids*	2022	0.508

After the clustering process is carried out, it can be seen that the clusters of MSME data in 2020 and 2022 with the k-means algorithm and the k-medoids algorithm based on cluster_0, cluster_1 and cluster_2 with the clustering of business owners and sub-districts can be seen in the graph below.

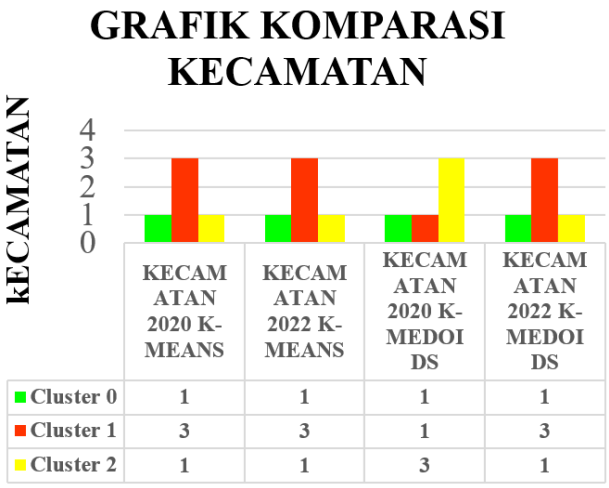

Fig 4. Comparison Chart of MSME Data BAsed on SubDistrict

From this data, the results of the sub-district clustering graph are obtained based on each year in 2020 and 2022, and can also see a graph of the level of the business sector, so that in the graph you can see what business sector is the highest based on the data obtained. Because the value of the Davies Bouldin index is closest to 0 in the k-means algorithm, the sub-district clustering graph will be displayed from the results of the k-means algorithm

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2090, Copyright ©2024
Submitted : February 12, 2024, Revised : March 1, 2024, Accepted : March 19, 2024, Published : June 15, 2024

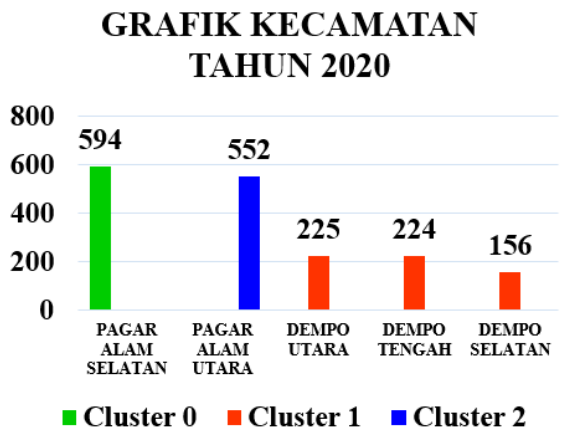

Fig 5. Graph of Sub-districts in 2020

Based on the sub-district graph in 2020 the high level sub-district is Pagar Alam Selatan cluster_0 with a total of 594. Low-level sub-districts Dempo Utara cluster_1 with a total of 225, Dempo Tengah cluster_1 with a total of 224, Dempo Selatan cluster_1 with a total of 156, and medium-level sub-districts Pagar Alam Utara cluster_2 with a total of 552

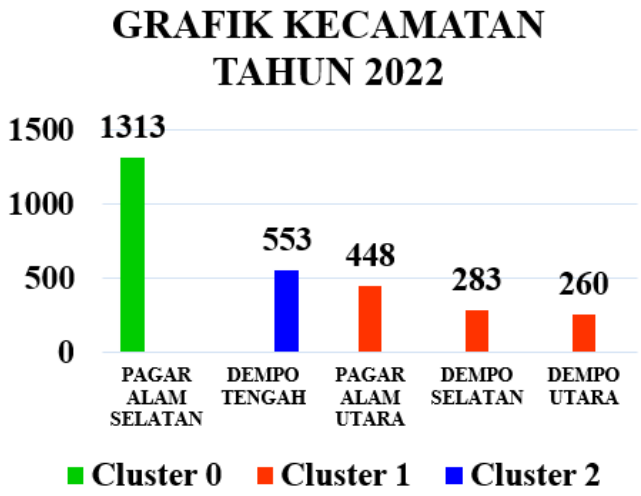

Fig 6. Graph of Subdistricts in 2022

Based on the sub-district graph in 2022 the high level sub-district is Pagar Alam Selatan cluster_0 with a total of 1313. Low-level sub-districts are Pagar Alam Utara cluster_1 with a total of 448, Dempo Selatan cluster_1 with a total of 283, Dempo Utara cluster_1 with a total of 260, and medium-level sub-district Dempo Tengah cluster_2 with a total of 553.

The number of business sector levels can be seen in the business sector chart in order to see and know the many variations in the MSME business sector, researchers have grouped them by business sector. In 2020 there are 11 types of business sectors, namely services, industry, trade, agriculture, livestock, crafts, plantations, fisheries, hotels/lodging, restaurants, and communications. In 2022 there are 15 business sectors, namely grocery traders, trade, food traders, industry, services, vegetable traders, crafts, livestock, restaurants, plantations, fish traders, fruit traders, fisheries, agriculture and cosmetics. The graph of business sectors can be seen annually by number.

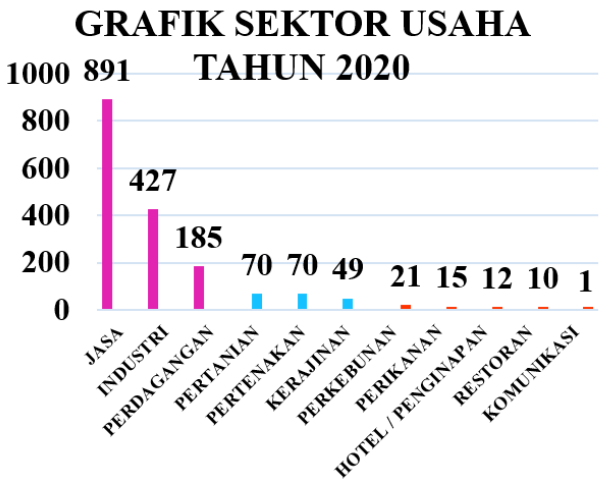

Fig 7. 2020 Business Sector Chart

Based on the 2020 business sector graph above, services with a total of 891 business sectors, industry with a total of 427 business sectors, trade with a total of 185 business sectors, agriculture with a total of 70 business sectors, livestock with a total of 70 business sectors, crafts with a total of 49 business sectors, plantations with a total of 21 business sectors, fisheries with a total of 15 business sectors, hotels / inns with a total of 12 business sectors, restaurants with a total of 10 business sectors, communication with a total of 1 business sector.

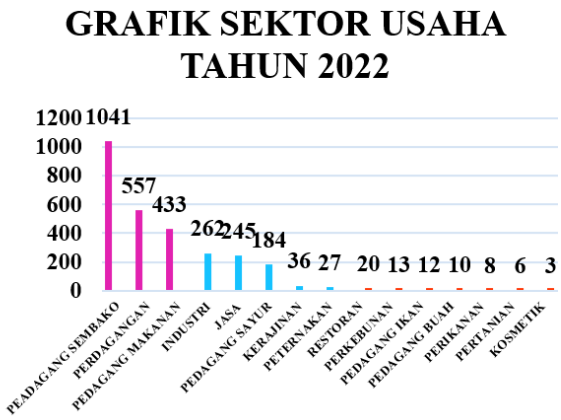

Fig 8. Graph of Business Sector in 2022

Based on the business sector graph in 2022 above, grocery traders with a total of 1041 business sectors, trade with a total of 557 business sectors, food traders with a total of 433 business sectors, industry with a total of 262 business sectors, services with a total of 245 business sectors, vegetable traders with a total of 184 business sectors, crafts with a total of 36 business sectors, livestock with a total of 27 business sectors, restaurants with a total of 20 business sectors, plantations with a total of 13 business sectors, fish traders with a total of 12

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2090, Copyright ©2024
Submitted : February 12, 2024, Revised : March 1, 2024, Accepted : March 19, 2024, Published : June 15, 2024

business sectors, fruit traders with a total of 10 business sectors, fisheries with a total of 8 business sectors, agriculture with a total of 6 business sectors, cosmetics with a total of 3 business sectors.

The results of this study are clustering MSME data in Pagar Alam City with the k-means model algorithm and k-medoids model which can be used as an ideal cluster recommendation to see MSME groups based on the highest sub-districts and business sectors or business fields for 2 years, namely 2020 with 2038 data and 2022 with 2906 data. The cluster results are obtained through the calculation of the k-means and k-medoids algorithms.

The k-means algorithm starts from determining the number of clusters (k), calculating the distance using Euclidean distance, then determining the centroid point randomly, fixing the centroid value. While the K-Medoids algorithm starts from selecting k objects in a set of n objects as medoids, calculating the medoid distance using Euclidean, calculating the total distance of all objects to their cluster centers, calculating the total deviation. Before clustering the data, researchers used the help of the rapidminer application first to generate clusters. MSME data in 2020 and 2022 total records from 4,944 after cleaning processing to 4,608 data. The results of cleaning will be clustered by year with UMKM records of 1,751 data in 2020 and 2,857 data in 2022. After that, researchers carried out the process of analyzing MSME clustering based on sub-districts and business sectors which will become a datasheet per year which is processed into the rapid miner.

Results of sub-district clustering in 2020. First cluster_0 = 1 sub-district, namely Pagar Alam Selatan sub-district. Second, cluster_1 = 3 sub-districts, namely North Dempo, Central Dempo and South Dempo. Third, cluster_2 = 1 sub-district, namely Pagar Alam Utara. While the K-Medoids algorithm First cluster_0 = 1 sub-district, namely the Pagar Alam Utara sub-district. Second, cluster_1 = 1 sub-district, namely Pagar Alam Selatan. Third, cluster_2 = 3 sub-districts, namely North Dempo, Central Dempo and South Dempo. Year 2022 K-Means First cluster_0 = 1 sub-district, namely Pagar Alam Selatan sub-district. Second cluster_1 = 3 sub-districts namely Pagar Alam Utara, Dempo Utara, Dempo Selatan. Third, cluster_2 = 1 sub-district, namely Central Dempo. K-Medoids First cluster_0 = 1 sub-district, namely Pagar Alam Selatan sub-district. Second cluster_1 = 3 sub-districts namely Pagar Alam Utara, Dempo Utara and Dempo Selatan. Third, cluster_2 = 1 sub-district, namely Central Dempo.

Based on the sub-district graph in 2020 the high level sub-district is Pagar Alam Selatan cluster_0 with a total of 594. Low-level sub-districts Dempo Utara cluster_1 with a total of 225, Dempo Tengah cluster_1 with a total of 224, Dempo Selatan cluster_1 with a total of 156, and medium-level sub-districts Pagar Alam Utara cluster_2 with a total of 552. Based on the sub-district graph in 2022 the high level sub-district is Pagar Alam Selatan cluster_0 with a total of 1313. Low-level sub-districts are Pagar Alam Utara cluster_1 with a total of 448, Dempo Selatan cluster_1 with a total of 283, Dempo Utara cluster_1 with a total of 260, and medium-level sub-district Dempo Tengah cluster_2 with a total of 553.

Based on the 2020 business sector chart, services with a total of 891 business sectors, industry with a total of 427 business sectors, trade with a total of 185 business sectors, agriculture with a total of 70 business sectors, livestock with a total of 70 business sectors, crafts with a total of 49 business sectors, plantations with a total of 21 business sectors, fisheries with a total of 15 business sectors, hotels / inns with a total of 12 business sectors, restaurants with a total of 10 business sectors, communication with a total of 1 business sector.

Based on the business sector graph in 2022, grocery traders with a total of 1041 business sectors, trade with a total of 557 business sectors, food traders with a total of 433 business sectors, industry with a total of 262 business sectors, services with a total of 245 business sectors, vegetable traders with a total of 184 business sectors, crafts with a total of 36 business sectors, livestock with a total of 27 business sectors, restaurants with a total of 20 business sectors, plantations with a total of 13 business sectors, fish traders with a total of 12 business sectors, fruit traders with a total of 10 business sectors, fisheries with a total of 8 business sectors, agriculture with a total of 6 business sectors, cosmetics with a total of 3 business sectors.

Then the results of the testing phase with the davies bouldin index (DBI) on the google colab application with the python programming language to calculate the results of the k-means and k-medoids clusters where the accuracy is the best, which is close to 0, the smaller the accuracy the better in illustrating MSME data in Pagar Alam City. The results of the calculation of DBI clustering sub-districts in 2020 DBI k-means = 0.134, DBI k-medoids algorithm = 0.523. The results of the DBI calculation of sub-district clustering in 2022 DBI k-means = 0.277, DBI k-medoids algorithm = 0.496.

Based on the DBI results of the k-means and k-medoids algorithms, the closest to 0 is the k-means algorithm with a DBI value of clustering the 2020 District = 0.134, DBI clustering the 2022 District = 0.277. The result that is close to 0 is the k-means algorithm so that the comparison of the k-means and k-medoids algorithms is the k-means algorithm that has the best accuracy value in clustering MSME data in Pagar Alam City.

V. CONCLUSION

This research results in 3 clusters based on MSME data in Pagar Alam City. With cluster_0 with a high level, cluster_1 with a low level and cluster_2 with a medium level.

The results of clustering sub-districts in 2020, cluster_0 with a high level, cluster_1 with a low level and cluster_2 with a medium level. Based on k-means, 3 clusters are obtained. First cluster_0 = 1 sub-district, namely Pagar Alam Selatan sub-district. Second, cluster_1 = 3 sub-districts, namely North Dempo, Central Dempo and South Dempo. Third, cluster_2 = 1 sub-district, namely Pagar Alam Utara. While the K-Medoids algorithm First cluster_0 = 1 sub-district, namely the Pagar Alam Utara sub-district. Second, cluster_1 = 1 sub-district, namely Pagar Alam Selatan. Third, cluster_2 = 3 sub-districts, namely North Dempo, Central Dempo and South Dempo. Year 2022 K-Means First cluster_0 = 1 sub-district,

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2090, Copyright ©2024
Submitted : February 12, 2024, Revised : March 1, 2024, Accepted : March 19, 2024, Published : June 15, 2024

namely Pagar Alam Selatan sub-district. Second cluster_1 = 3 sub-districts namely Pagar Alam Utara, Dempo Utara, Dempo Selatan. Third, cluster_2 = 1 sub-district, namely Central Dempo. K-Medoids First cluster_0 = 1 sub-district, namely Pagar Alam Selatan sub-district. Second cluster_1 = 3 sub-districts namely Pagar Alam Utara, Dempo Utara and Dempo Selatan. Third, cluster_2 = 1 sub-district, namely Central Dempo.

Based on the clustering results, the Pagar Alam Selatan sub-district is consistent in cluster_0 with the highest level and has experienced an increase of almost 50% based on the number of MSMEs. Based on the 2020 sub-district graph obtained by the agency, we can see that cluster_1 with a low level is in the North Dempo, Central Dempo and South Dempo sub-districts. Sub-districts in 2022 cluster_1 with a low level are North Pagar Alam, North Dempo, South Dempo sub-districts. So that the agency can focus more on the lowest level in developing and improving MSMEs in low clusters.

The results of the level of the MSME business sector can be used by sub-districts, villages and neighborhoods so that they can focus more on the business sector at the top 3 highest levels in 2020 and 2022, there are business sectors in the form of services, industry, trade, grocery traders and food traders, where the potential talent interest of the community is more in this business sector, so that it can be used to focus more on the 5 business sectors so that they can be developed and increased selling results to support the community's economy. 5.

The results of the DBI calculation of sub-district clustering in 2020 DBI k-means = 0.134, DBI k-medoids algorithm = 0.523. The results of the DBI calculation of sub-district clustering in 2022 DBI k-means = 0.277, DBI k-medoids algorithm = 0.496. Based on the DBI results of the k-means and k-medoids algorithms, the closest to 0 is the k-means algorithm with a DBI value of 2020 sub-district clustering = 0.134, DBI sub-district clustering in 2022 = 0.277. The result that is close to 0 is the k-means algorithm so that the comparison of the k-means and k-medoids algorithms is the k-means algorithm that has the best accuracy value in clustering MSME data in Pagar Alam City.

From these results, some knowledge will be obtained which is expected to be useful for the agency for a policy decision in developing and improving MSMEs in the future.

REFERENCE

[1] J. Suntoro, *Data Mining Algoritma dan Implementasi dengan Pemrograman PHP*. PT Elex Media Komputindo, 2019.

[2] A. Fiyan, N. Falahie, T. Susyanto, and R. T. Vulandari, "Implementasi Algoritma Apriori pada Tata Letak Kategori Buku di Perpustakaan," no. 1, pp. 23–34, 2022.

[3] J. Warmansyah, *Pengolahan dan Perancangan CRM dengan Model Prototype dan Simulasi Data Mining*. Deepublish (CV BUDI UTAMA), 2022.

[4] A. S. Wibowo and I. D. Mulyastuti, "Penerapan Algoritma K-Means Clustering Pada Jumlah Fasilitas Kesehatan Menurut Pemerintah Provinsi DKI Jakarta," vol. 23, no. 2, pp. 116–122, 2022.

[5] S. Nurlaela and A. Primajaya, "Algoritma K-Medoids Untuk Clustering Penyakit Maag Di Kabupaten Karawang," vol. 12, no. 2, pp. 56–62, 2020.

[6] W. Sudrajat, I. Cholid, and J. Petrus, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer," pp. 27–36, 2022.

[7] R. Wahyusari and S. Wardani, "Perbandingan Algoritma K-Means dan Algoritma K- Medoid Untuk Pengelompokan UMKM di Kebumen Comparison of the K-Means Algorithm and the K-Medoid Algorithm for," pp. 74–79, 2023.

[8] E. Tasia and M. Afdal, "Comparison Of K-Means And K-Medoid Algorithms For Clustering Of Flood-Prone Areas In Rokan Hilir District Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir," vol. 3, no. 1, pp. 65–73, 2023.

[9] I. Suryati, "Pengaruh Ukuran Usaha Dan Sumber Modal Terhadap Penerapan Standar Akuntansi Pada Usaha Mikro Kecil Dan Menengah Bidang Jasa Atau Pelayanan Laundry Di Kecamatan Makasar Tahun 2019," vol. 1, no. 1, pp. 18–30, 2021.

[10] N. Manullang, R. W. Sembiring, I. Gunawan, I. Parlina, T. Informatika, and T. Informatika, "Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5," vol. 2, no. 2, pp. 1–5, 2021.

[11] I. G. S. Elrohi, Marlina, and Renny, "Implementasi Cloud Computing dengan Google Colaboratory Pada Aplikasi Pengolah Data Zoom Participants," vol. 6, no. 1, pp. 24–30, 2022.

[12] I. W. Septiani, A. C. Fauzan, and M. M. Huda, "Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin- Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru," vol. 3, pp. 556–566, 2022, doi: 10.30865/json.v3i4.4055.

[13] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," vol. 5, no. 2, 2021.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i2.2090, Copyright ©2024
Submitted : February 12, 2024, Revised : March 1, 2024, Accepted : March 19, 2024, Published : June 15, 2024