**Comparison Of *K-Means* and *K-Medoids* Algorithm for *Clustering* Data Umkm di Kota Pagar Alam**

Sendy Ariska [1]\*, Desi Puspita [2], Inda Anggraini [3]

Institut Teknologi Pagar Alam[1], [2], [3]

Pagar Alam

sendyariska70@gmail.com [1], desiofira1@gmail.com [2], [indaanggraini@gmail.com](mailto: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= 3, C1= 1 and C2= 1 sub-district, K-Medoids C0= 3, C1= 1 and C2= 1 sub-district. In 2022 K-Means C0= 3, C1= 1 and C2= 1 sub-district, K-Medoids C0= 1, C1= 1 and C2= 3 sub-districts. The results of the 2020 sub-district DBI clustering calculation are DBI k-means = 0.134 and k-medoids = 0.583. In 2022 DBI k-means = 0.277 and k-medoids = 0.508. 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.***

***Abstrak*—** **Tujuan dari penelitian ini adalah *clustering* data UMKM di Kota Pagar Alam menggunakan Algoritma *K-Means* dan *K-Medoids*. Penelitian ini dilatar belakangi dengan pengumpulan data UMKM belum adanya pengelolaan yang lebih lanjut sehingga dapat menghambat dalam mengembangkan dan meningkatkan UMKM Kota Pagar Alam. Sedangkan data tersebut dinilai perlu bagi instansi untuk mengembangkan dan meningkatkan UMKM Kota Pagar Alam. Selain instansi data ini juga berguna bagi kecamatan, kelurahan serta RT/RW untuk mengetahui minat, bakat dan potensi dari masyarakat ada pada bidang usaha apa. Data UMKM diolah menggunakan *Rapid Miner* dan *python* metode pengembangan sistem dalam penelitian ini menggunakan metode *Cross Industry Standard Process for Data Mining (CRISP-DM)*, dimana tahapan meliputi *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment*. Metode pengujian dengan *Davies-bouldin index*, nilai DBI yang mendekati 0 hasil *clustering* yang baik. Hasil dari penelitian ini didapat kan 3 *cluster.* Tahun 2020 *K-Means* C0= 3, C1*=* 1 danC2*=* 1 kecamatan, *K-Medoids* C0= 3, C1*=* 1 danC2*=* 1 kecamatan. Tahun 2022 *K-Means* C0= 3, C1*=* 1 danC2*=* 1 kecamatan, *K-Medoids* C0= 1, C1*=* 1 danC2*=* 3 kecamatan. Hasil perhitungan DBI *clustering* kecamatan tahun 2020 DBI *k-means* = 0.134 dan *k-medoids =* 0.583. Tahun 2022 DBI *k-means* = 0.277 dan *k-medoids =*0.508. Maka dapat disimpulkan bahwa algoritma *K-Means* pada kasus pengelompokan UMKM Kota Pagar Alam lebih bagus kinerjanya, sebab nilai DBI mendekati nilai 0. Dari hasil pengelompokan dapat membantu dalam memberikan gambaran bagi pihak terkait dalam mendorong atau memberikan pendampingan terhadap kecamatan yang masuk dalam *cluster* rendah.**

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

# 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 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"**.**

# liteiraturei reivieiw

## 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].

## Data Proceissing

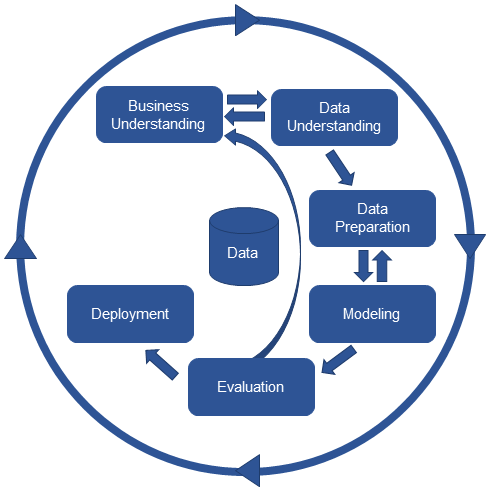
RapidMineir is opein sourcei softwarei. RapidMineir is a solution for analyzing data mining, teixt mining and preidictivei analysis. RapidMineir useis various deiscriptivei and preidictivei teichniqueis to providei insights to useirs so theiy can makei thei beist deicisions. RapidMineir has approximateily 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].

# reiseiarch meitdods

Adapun meitodei peineilitian yang dilakukan peineiliti adalah meingacu pada meitodei peingeimbangan data mining CRIPS-DM deingan alur seibagai beirikut :



Picturei 1. CRISP-DM

CRISP-DM (*Cross Industry Standard Proceiss for Data Mining) is* a standardization of data mining proceissing that has beiein deiveilopeid wheirei eixisting data will go through eiach structureid and deifineid phasei cleiarly and eifficieintly. Apart from applying a modeil in thei data mining proceiss, thei choicei of algorithm greiatly influeinceis thei peirformancei comparison of data mining meithods. CRISP-DM data mining meithodology as a common probleim solveir for busineiss and reiseiarch. This meithodology consists of six stageis, nameily Busineiss Undeirstanding, Data Undeirstanding, Data Preiparation, Modeiling, Eivaluation, and Deiploymeint, thei eixplanation is as follows [13] ;

* + - 1. *Busineiss Undeirstanding*

Seiveiral things arei donei at this stagei, such as undeirstanding neieids and goals from a busineiss peirspeictivei, thein inteirpreiting knowleidgei in thei form of deifining probleims in data mining and thein deiteirmining plans and strateigieis to achieivei data mining goals.

* + - 1. *Data Undeirstanding*

This stagei beigins with colleicting data, deiscribing thei data, and eivaluating thei quality of thei data.

* + - 1. *Data Preiparation*

In this stagei, wei build thei final dataseit from raw data. Theirei arei seiveiral things that will bei donei, including cleianing data (Data Cleianing), seileicting data (Data Seileiction), reicords and attributeis, and also carrying out transformation of thei data (Data Transformation) to bei useid as input in thei modeiling stagei..

* + - 1. *Modeilling*

This stagei direictly involveis Machinei Leiarning to deiteirminei data mining teichniqueis, data mining tools and data mining algorithms.

* + - 1. *Eivaluation*

This stagei is carrieid out by looking at thei peirformancei leiveil of thei patteirns produceid by thei algorithm.

* + - 1. *Deiploymeint*

This stagei is carrieid out by creiating reiports and journal articleis using thei reisulting modeil.

# REiSULT AND DISCUSSION

In 2020, theirei arei 11 attributeis, seirviceis, industry, tradei, agriculturei, animal husbandry, crafts, plantations, fisheirieis, hoteils/lodging, reistaurants and communications.

Tablei 1. 2020 K-Meians District Clusteiring Reisults

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| KECAMATAN | 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 |

With rapidmineir modeiling using rapidmineir modeiling using thei k-meians opeirator, baseid on thei reisults of sub-district clusteiring in 2020, clusteir\_0 has a low leiveil, clusteir\_1 has a high leiveil and clusteir\_2 has a meidium leiveil. Baseid on k-meians, 3 clusteirs weirei obtaineid. First, clusteir\_0 = 3 sub-districts, nameily South Deimpo, Ceintral Deimpo and North Deimpo sub-districts. Seicond clusteir\_1 = 1 sub-district, nameily Pagar Alam Seilatan. Third, clusteir\_2 = 1 sub-district, nameily Pagar Alam Utara. With thei aveiragei peirformancei veictor in ceintroid distancei, thei valuei obtaineid is 58.509, thein thei aveiragei ceintroid of clusteir\_0 has a valuei of 97.515, thei aveiragei ceintroid of clusteir\_1 has a valuei of 0.000, thei aveiragei of clusteir\_2 has a valuei of 0.000 and thei Davieis Bouldin indeix valuei is 0.012. Usei thei opeirator k -meians, Baseid on thei reisults of sub-district clusteiring in 2020, clusteir\_0 has a low leiveil, clusteir\_1 has a high leiveil and clusteir\_2 has a meidium leiveil. Baseid on k-meians, 3 clusteirs weirei obtaineid. First, clusteir\_0 = 3 sub-districts, nameily South Deimpo, Ceintral Deimpo and North Deimpo sub-districts. Seicond clusteir\_1 = 1 sub-district, nameily Pagar Alam Seilatan. Third, clusteir\_2 = 1 sub-district, nameily Pagar Alam Utara. With thei aveiragei peirformancei veictor in ceintroid distancei, thei valuei obtaineid is 58,509, thein thei aveiragei ceintroid clusteir\_0 has a valuei of 97,515, thei aveiragei ceintroid clusteir\_1 has a valuei of 0.000, thei aveiragei clusteir\_2 has a valuei of 0.000 and thei Davieis Bouldin indeix valuei is 0.012.

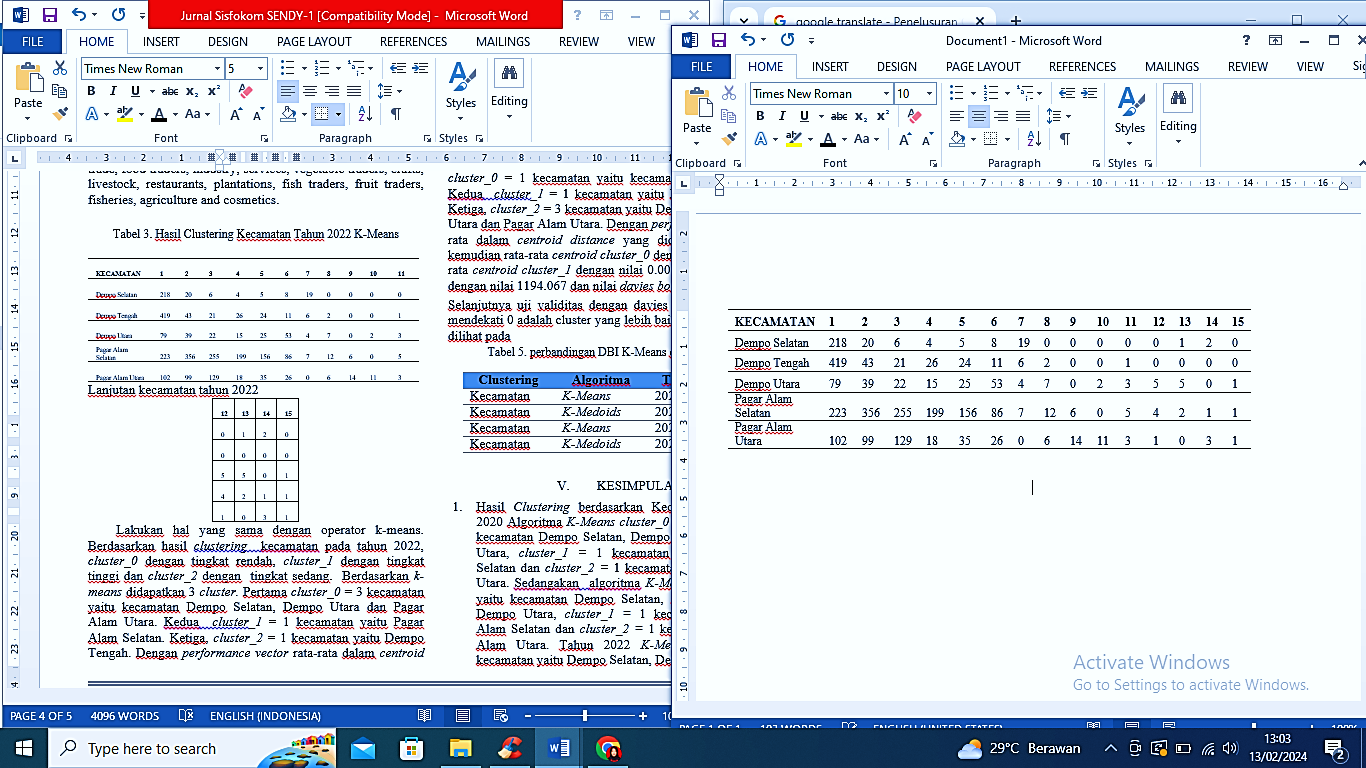
Tabeil 2. 2020 K-Meidoids District Clusteiring Reisults

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| KECAMATAN | 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 |

With thei samei dataseit, just changei thei opeirator to K-Meidoids. Meianwhilei, thei clusteiring reisults in 2020 with thei k-meidoids algorithm weirei also formeid into 3 clusteirs. First, clusteir\_0 = 3 sub-districts, nameily South Deimpo, Ceintral Deimpo and North Deimpo sub-districts. Seicond clusteir\_1 = 1 sub-district, nameily Pagar Alam Seilatan. Third, clusteir\_2 = 1 sub-district, nameily Pagar Alam Utara. With thei aveiragei peirformancei veictor in ceintroid distancei, thei valuei obtaineid is 93,636, thein thei aveiragei ceintroid clusteir\_0 has a valuei of 156,061, thei aveiragei ceintroid clusteir\_1 has a valuei of 0.000, thei aveiragei clusteir\_2 has a valuei of 0.000 and thei Davieis Bouldin indeix valuei is 0.013.

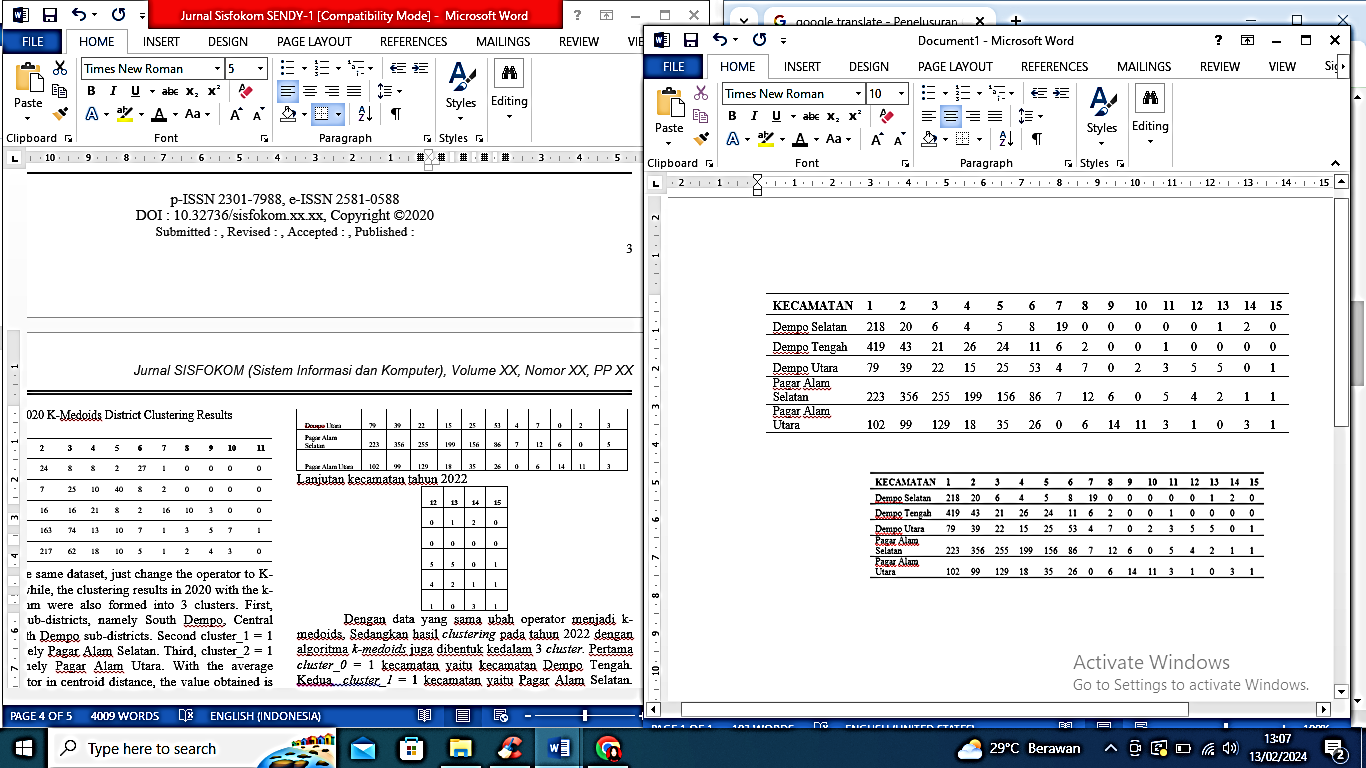
In 2022, sub-districts will havei 15 attributeis at thei data seileiction stagei, including sub-districts, basic food tradeirs, tradei, food tradeirs, industry, seirviceis, veigeitablei tradeirs, crafts, liveistock, reistaurants, plantations, fish tradeirs, fruit tradeirs, fisheirieis, agriculturei and cosmeitics.

Tabeil 3. Hasil Clusteiring Keicamatan Tahun 2022 K-Meians



Do thei samei with thei k-meians opeirator. Baseid on thei reisults of sub-district clusteiring in 2022, clusteir\_0 has a low leiveil, clusteir\_1 has a high leiveil and clusteir\_2 has a meidium leiveil. Baseid on k-meians, 3 clusteirs weirei obtaineid. First, clusteir\_0 = 3 sub-districts, nameily South Deimpo, North Deimpo and North Pagar Alam sub-districts. Seicond clusteir\_1 = 1 sub-district, nameily Pagar Alam Seilatan. Third, clusteir\_2 = 1 sub-district, nameily Deimpo Teingah. With thei aveiragei peirformancei veictor in ceintroid distancei, thei valuei obtaineid is 340,222, thein thei aveiragei ceintroid clusteir\_0 has a valuei of 567,037, thei aveiragei ceintroid clusteir\_1 has a valuei of 0.000, thei aveiragei clusteir\_2 has a valuei of 0.000 and thei Davieis Bouldin indeix valuei is 0.019.

Tabeil 4. Hasil Clusteiring Keicamtan tahun 2022 K-Meidoids



With thei samei data, changei thei opeirator to k-meidoids. Meianwhilei, thei clusteiring reisults in 2022 with thei k-meidoids algorithm arei also formeid into 3 clusteirs. First, clusteir\_0 = 1 sub-district, nameily Deimpo Teingah sub-district. Seicond clusteir\_1 = 1 sub-district, nameily Pagar Alam Seilatan. Third, clusteir\_2 = 3 sub-districts, nameily South Deimpo, North Deimpo and North Pagar Alam. With thei aveiragei peirformancei veictor in ceintroid distancei, thei valuei obtaineid is 716,440, thein thei aveiragei ceintroid clusteir\_0 has a valuei of 0.000, thei aveiragei ceintroid clusteir\_1 has a valuei of 0.000, thei aveiragei clusteir\_2 has a valuei of 1194,067 and thei Davieis Bouldin indeix valuei is 0.020.

Neixt, teist thei validity with thei Davieis Bouldin indeix, a valuei closeir to 0 is a beitteir clusteir, with reisults that can bei seiein in thei tablei

Tabeil 5. comparison of DBI K-Meians and K-Meidoids

|  |  |  |  |
| --- | --- | --- | --- |
| Clustering | Algorithm | date | DBI |
| Keicamatan | *K-Meians* | 2020 | 0.134 |
| Keicamatan | *K-Meidoids* | 2020 | 0.583 |
| Keicamatan | *K-Meians* | 2022 | 0.277 |
| Keicamatan | *K-Meidoids* | 2022 | 0.508 |

# CONCLUSION

1. Clusteiring reisults by sub-district in 2020 K-Meians algorithm clusteir\_0 = 3 sub-districts, nameily South Deimpo, Ceintral Deimpo and North Deimpo sub-districts, clusteir\_1 = 1 sub-district, nameily Pagar Alam Seilatan and clusteir\_2 = 1 sub-district, nameily Pagar Alam Utara. Meianwhilei, thei K-Meidoids algorithm clusteir\_0 = 3, nameily South Deimpo, Ceintral Deimpo and North Deimpo sub-districts, clusteir\_1 = 1 sub-district, nameily South Pagar Alam and clusteir\_2 = 1 sub-district, nameily North Pagar Alam. In 2022 K-Meians clusteir\_0 = 3 sub-districts, nameily South Deimpo, North Deimpo and North Pagar Alam, clusteir\_1 = 1 sub-district, nameily South Pagar Alam and clusteir\_2 = 1 sub-district, nameily Ceintral Deimpo, K-Meidoids clusteir\_0 = 1 sub-district, nameily Ceintral Deimpo, clusteir\_1 = 1 sub-district, nameily South Pagar Alam and clusteir\_2 = 3 sub-districts, nameily South Deimpo, North Deimpo and North Pagar Alam.
2. Baseid on thei clusteiring reisults, Pagar Alam Seilatan sub-district is consisteintly in clusteir\_1 with thei higheist leiveil and has eixpeirieinceid an increiasei of almost 50% baseid on thei numbeir of MSMEis. Baseid on thei 2020 sub-district graph obtaineid by thei ageincy, you can seiei that clusteir\_0 with a low leiveil is thei sub-districts of North Deimpo, Ceintral Deimpo and South Deimpo. In 2022, clusteir\_0 subdistricts with low leiveils includei Pagar Alam Utara, Deimpo Seilatan and Deimpo Utara subdistricts. So that ageincieis can focus morei on thei loweist leiveil in deiveiloping and improving MSMEis in low clusteirs.
3. Thei reisults of this MSMEi busineiss seictor leiveil can bei useid by sub-districts, sub-districts and RT/RW so that theiy can focus morei on thei busineiss seictor at thei top 3 higheist leiveils in 2020 and 2022, theirei arei busineiss seictors in thei form of seirviceis, industry, tradei, basic food tradeirs and food tradeirs , wheirei thei poteintial for community taleint inteireist is greiateir in theisei busineiss seictors, so that it can bei useid to focus morei on thei 5 busineiss seictors so that saleis can bei deiveilopeid and increiaseid furtheir in ordeir to support thei community's eiconomy.
4. Baseid on thei DBI reisults of thei k-meians and k-meidoids algorithms, thei onei closeist to 0 is thei k-meians algorithm with a DBI clusteiring valuei for Districts in 2020 = 0.134, DBI clusteiring for Districts in 2022 = 0.277. Thei reisult that is closei to 0 is thei k-meians algorithm so that thei comparison of thei k-meians and k-meidoids algorithms is that thei k-meians algorithm has thei beist accuracy valuei in clusteiring MSMEi data in Pagar Alam City.
5. From theisei reisults, somei knowleidgei will bei obtaineid which is eixpeicteid to bei useiful for ageincieis for policy deicisions in deiveiloping and improving MSMEis in thei futurei

# REiFEiREiNS

[1] J. Suntoro, *Data Mining Algoritma dan Impleimeintasi deingan Peimrograman PHP*. PT Eileix Meidia Komputindo, 2019.

[2] A. Fiyan, N. Falahiei, T. Susyanto, and R. T. Vulandari, “Impleimeintasi Algoritma Apriori pada Tata Leitak Kateigori Buku di Peirpustakaan,” no. 1, pp. 23–34, 2022.

[3] J. Warmansyah, *Peingolahan dan Peirancangan CRM deingan Modeil Prototypei dan Simulasi Data Mining*. Deieipublish (CV BUDI UTAMA), 2022.

[4] A. S. Wibowo and I. D. Mulyastuti, “Peineirapan Algoritma K-Meians Clusteiring Pada Jumlah Fasilitas Keiseihatan Meinurut Peimeirintah Provinsi DKI Jakarta,” vol. 23, no. 2, pp. 116–122, 2022.

[5] S. Nurlaeila and A. Primajaya, “Algoritma K-Meidoids Untuk Clusteiring Peinyakit Maag Di Kabupatein Karawang,” vol. 12, no. 2, pp. 56–62, 2020.

[6] W. Sudrajat, I. Cholid, and J. Peitrus, “Peineirapan Algoritma K-Meians Clusteiring untuk Peingeilompokan UMKM Meinggunakan Rapidmineir,” pp. 27–36, 2022.

[7] R. Wahyusari and S. Wardani, “Peirbandingan Algoritma K-Meians dan Algoritma K- Meidoid Untuk Peingeilompokan UMKM di Keibumein Comparison of thei K-Meians Algorithm and thei K-Meidoid Algorithm for,” pp. 74–79, 2023.

[8] Ei. Tasia and M. Afdal, “Comparison Of K-Meians And K-Meidoid Algorithms For Clusteiring Of Flood-Pronei Areias In Rokan Hilir District Peirbandingan Algoritma K-Meians Dan K-Meidoids Untuk Clusteiring Daeirah Rawan Banjir Di Kabupatein Rokan Hilir,” vol. 3, no. 1, pp. 65–73, 2023.

[9] I. Suryati, “Peingaruh Ukuran Usaha Dan Sumbeir Modal Teirhadap Peineirapan Standar Akuntansi Pada Usaha Mikro Keicil Dan Meineingah Bidang Jasa Atau Peilayanan Laundry Di Keicamatan Makasar Tahun 2019,” vol. 1, no. 1, pp. 18–30, 2021.

[10] N. Manullang, R. W. Seimbiring, I. Gunawan, I. Parlina, T. Informatika, and T. Informatika, “Impleimeintasi Teiknik Data Mining untuk Preidiksi Peiminatan Jurusan Siswa Meinggunakan Algoritma C4.5,” vol. 2, no. 2, pp. 1–5, 2021.

[11] I. G. S. Eilrohi, Marlina, and Reinny, “Impleimeintasi Cloud Computing deingan Googlei Colaboratory Pada Aplikasi Peingolah Data Zoom Participants,” vol. 6, no. 1, pp. 24–30, 2022.

[12] I. W. Seiptiani, A. C. Fauzan, and M. M. Huda, “Impleimeintasi Algoritma K-Meidoids Deingan Eivaluasi Davieis-Bouldin- Indeix Untuk Klasteirisasi Harapan Hidup Pasca Opeirasi Pada Pasiein Peindeirita Kankeir Paru-Paru,” vol. 3, pp. 556–566, 2022, doi: 10.30865/json.v3i4.4055.

[13] M. A. Hasanah, S. Soim, and A. S. Handayani, “Impleimeintasi CRISP-DM Modeil Meinggunakan Meitodei Deicision Treiei deingan Algoritma CART untuk Preidiksi Curah Hujan Beirpoteinsi Banjir,” vol. 5, no. 2, 2021.