Proceeding of International Conference on Digital Advance Tourism, Management and Technology 2023
Volume. 1 No. 1 2023 e-ISSN : 3031-5999, Hal 184-191
DOI: https://doi.org/10.56910/ictmt.v1i1.62

Classterization Of Village Welfare Levels With The K-Means Algorithm For Planning Quality Of Life Improvement Program (Case Study District XYZ)

Yani Prihati [1], Alexander Dharmawan [2], Tri Purwani [3], Yusup [4], Berlian Setya Hanugraheni [5]
[1-5] Universitas AKI, Semarang, Indonesia

Corresponding email: yani.prihati@unaki.ac.id

Abstract. *This research applies the concept of big data and the data mining process and aims to provide clustered data on village welfare levels which can help local governments to properly plan programs to improve the quality of life of village communities. Research case study in District XYZ whose area includes 11 villages. The clustering process is carried out using the K-Means algorithm which groups data based on similar characteristics. The data analysis process was carried out using the RStudio tool and Microsoft Excel as a comparison. The data set used has 10 attributes, namely village name, number of Pre-Prosperous families, economic reasons (Alek) in the Pre-Prosperous column, non-economic reasons (Bulek) in the Pre-Prosperous column, number of Prosperous Families I (KS I), Alek in the KS column I, Bulek in column KS I, number of Prosperous Families (KS II), number of Prosperous Families (KS III) and number of Prosperous Families (KS III) Plus. Data processing with both tools show the same results, namely that there are 2 clusters. Cluster 1 is a cluster with a high level of welfare, consisting of 4 villages and Cluster 2 is a cluster with a low level of welfare, consisting of 7 villages.*

Keywords. *Classterization; welfare levels; K-Means*

INTRODUCTION

The realization of the Happy Prosperous Small Family Norms (Norma Keluarga Kecil Bahagia dan Sejahtera -NKKBS) is the basis for improving the welfare of mothers, as well as the realization of a prosperous society. The quality of human life is generally measured through three stages, namely the fulfillment of basic needs for survival as living creatures, the fulfillment of basic needs for human survival, and the fulfillment of basic needs for choice (Bidari, 2020).

One way the quality of human life is influenced by population (https://kulonprogokab.go.id/v31/detil/4527/perbangun-penbangun-dan-kuulasi-jalan-kita). A large population will require greater sufficiency of food, clothing and shelter. Likewise, facilities and infrastructure for education, health, recreation and so on.

Welfare is the basis of hope and the noble goals of the struggle of the Indonesian people, especially in determining the development of a region. The National Population and Family Planning Agency (BKKBN) divides the criteria for family welfare into five stages, including the Pre-Prosperous Family stage (Keluarga Pra Sejahtera - KPS), the Prosperous Family stage I (Keluarga Sejahtera 1 - KS 1), the Prosperous Family stage II (Keluarga Sejahtera II - KS II), the Prosperous Family stage III (Keluarga Sejahtera III - KS III) and the final stage Prosperous Family III Plus (K eluarga Sejahtera III Plus - KS III Plus). The level of village

Received November 19, 2023; Revised Desember 01 , 2023; Accepted Desember 31, 2023
* Yani Prihati , *yani.prihati@unaki.ac.id*

welfare needs to be clustered to be used as one of the basic data for village development planning in order to improve the quality of life of the community towards prosperity. In District

Data Mining is the process of searching for data to find clear relationships and provide conclusions that can be understood and are useful for the owner of the data. Data mining is defined as mining data or efforts to dig up valuable and useful information in very large databases (Tarigan, 2021). Data mining is a data collection technique to look for hidden patterns in order to produce new knowledge in a collection of data. The clustering method is a data mining analysis method without supervision or what is usually called unsupervised and is a grouping of data using a partition system. K-Means Clustering is an algorithm that groups data into several groups, where the data in one group has the same characteristics as each other and has different characteristics from the data in other groups.

This research applies the concept of big data and the data mining process and aims to provide clustered data on village welfare levels which can help local governments to properly plan programs to improve the quality of life of village communities. The case study for this research is in XYZ District, whose area includes 11 villages

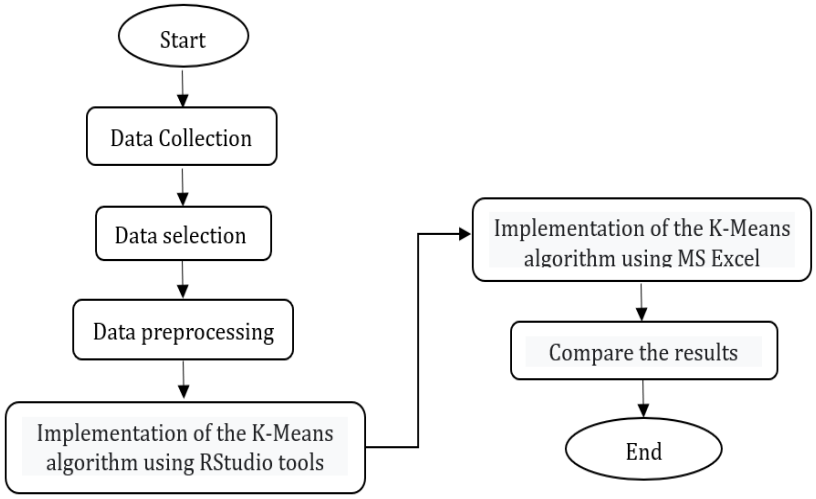
METHOD

In order for the research objectives to be achieved, initially an analysis of the research object must be carried out to find out all the issues and problems experienced by the research object. The research object must be specific so that in detail it can be determined what will be researched and what data should be collected. In this research, data was collected by interviews, observation and literature study. The data obtained and used as a dataset is in the form of data on the achievement of the prosperous family stages of 11 villages included in the XYZ District area in 2022. The 11 villages are coded as follows: Taja, Bumi, Cikl, Kedu, Tari, Reja, Sida, Kali, Jati, Bang and Bojo. The data obtained is then carried out data preprocessing consisting of data cleaning and data transformation so that it is ready for clustering.

The clustering process in research will be carried out according to the K-Means algorithm. The K-Means algorithm is a clustering algorithm that groups data based on the closest cluster center point (centroid) to the data which aims to group data by maximizing the similarity of data in one cluster and minimizing the similarity of data between clusters (Prihati, Suwarno, Dharmawan, 2021). The K-Means algorithm is a non-hierarchical method that partitions data into one or more clusters, so that data that has the same characteristics is grouped in the same cluster and data that has different characteristics is grouped into another cluster.

Irahman, 2017). So that data processing can be compared, it is done with 2 tools, namely RStudio and Microsoft Excel. The results of data processing are a list of villages included in each cluster. The research process was carried out according to the flowchart in Figure 1.

Figure 1. Research Process



RESULTS AND DISCUSSION

Initially, the dataset used had 10 attributes, namely: village name, economic reasons (alek) pre-prosperous, not economic reasons (bulek) pre-prosperous, number of pre-prosperous, alek KS I, bulek KS I, number of prosperous families (KS) I , number of KS II, number of KS III and number of KS III Plus. Data for these ten attributes is available for January – May 2022, so there are a total of 50 attributes.

Data cleaning is carried out to ensure the quality of the data to be processed. Data cleaning is a process used to detect, repair or delete corrupt or inaccurate datasets, tables and databases (Widiari, 2020). Data Cleaning is carried out by deleting data if there is duplicate data or data that is not needed, editing structural errors such as if there are typos, there are upper and lower case letters, or the number zero has changed to N/A. In the dataset, there was no empty data and no double or duplicate data was found. In this case, the data can be processed through the data transformation stages. Data transformation produces a new dataset consisting of 11 objects and 26 attributes, namely: Village, Jan-KPS, Jan-KS I, Jan-KS II, Jan-KS III, Jan-KS III Plus, Feb-KPS, Feb-KS I , Feb-KS II, Feb-KS III, Feb-KS III Plus, Mar-KPS, Mar- KS I, Mar- KS II, Mar- KS III, Mar- KS III Plus, Apr-KPS, Apr-KS I , Apr-KS II, Apr-KS III, Apr-KS III Plus, Mei-KPS, Mei-KS I, Mei-KS II, Mei-KS III, Mei-KS III Plus. The results of data transformation are presented in Figure 2.

Figure 2. The Results Of Data Transformation

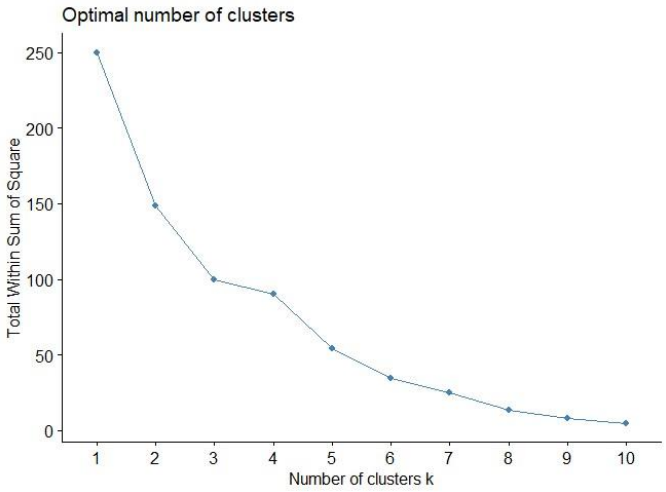
```
> data
        Desa Jan..KPS Jan..KS.I Jan..KS.II Jan...KS.III Jan..KS.III.Plus Feb..KPS Feb..KS.I
1  Tamb.          1094       389        344           62                17     1095       390
2    Bumi         1260       422        213           71                14     1260       423
3     Cikl        1365       842        510          323                87     1367       844
4  Kedu           911       807        424          399                35      913       807
5  Tamb           940       851        795          243                93      941       851
6   Reja.        1161       479        515           72                36     1165       480
7  Sida           1051       759        481          242                31     1052       759
8   Kali          1045       378        358          327                17     1047       381
9    Jati         1141       661        345          204                37     1141       661
10 Bangi          932       608        267          221                24      933       608
11 Bojo           936       714        332           89                19      938       714
   Feb..KS.II Feb...KS.III Feb..KS.III.Plus Mar..KPS Mar..KS.I Mar..KS.II Mar...KS.III
1         344           62                17     1096       391        344           62
2         213           71                14     1264       424        213           71
3         511          323                87     1373       845        511          323
4         424          399                35      914       808        424          399
5         795          243                93      945       854        795          243
6         515           72                36     1165       480        515           72
7         481          242                31     1054       760        481          242
8         358          327                17     1050       381        358          327
9         345          204                37     1143       662        345          204
10        267          221                24      935       609        267          221
11        332           89                19      940       715        332           89

   Mar..KS.III.Plus Apr..KPS Apr..KS.I Apr..KS.II Apr...KS.III Apr..KS.III.Plus Mei..KPS Mei..KS.I
1                17     1097       391        344           62                17     1100       392
2                14     1265       425        213           71                14     1266       425
3                87     1375       845        511          323                87     1379       847
4                35      915       808        424          399                35      916       808
5                93      947       855        795          243                93      951       858
6                36     1167       481        515           72                36     1170       483
7                31     1056       760        481          242                31     1058       760
8                17     1050       382        358          327                17     1053       383
9                37     1144       662        345          204                37     1147       663
10               24      936       609        267          221                24      939       610
11               19      942       715        332           89                19      942       715
   Mei..KS.II Mei...KS.III Mei..KS.III.Plus
1         344           62                17
2         213           71                14
3         511          323                87
4         424          399                35
5         795          243                93
6         515           72                36
7         481          242                31
8         358          327                17
9         345          204                37
10        267          221                24
11        332           89                19
```

Data Processing with RStudio

Before the K-Means algorithm is implemented, you must import the library, import the dataset, and normalize the data in RStudio. To determine the number of clusters, the Elbow method is used. The aim of the elbow method is to choose a k value that is small and still has a low withinss value. The plot of the Elbow method results is presented in Figure 3.

Figure 3. The Plot of the Elbow method results



It can be seen that the angled shape of the curve is at number 3 (three) but the drastic decrease in value is at number two where there is a decrease of one hundred while at number three there is a decrease of only fifty. So it is determined that the optimal number of clusters is two. The results of clustering with RStudio are presented in Figure 4. Clustering with RStudio produces output. Cluster 1 is a high level of prosperity consisting of four villages, namely Cikl, Kedu, Tari and Sida, while Cluster 2 is a low level of prosperity consisting of seven villages, namely Taja, Bumi, Reja, Kali, Jati, Bang and Bojo

Figure 4. Clustering Result With RStudio

```
> clustering = kmeans(datascale, centers = 2, nstart = 30)
> clustering
K-means clustering with 2 clusters of sizes 4, 7

Cluster means:
     Jan..KPS  Jan..KS.I Jan..KS.II Jan...KS.III Jan..KS.III.Plus     Feb..KPS Feb..KS.I Feb..KS.II
1 -0.06282292  1.0149929  0.8576458     0.8219768         0.8847622 -0.06252420  1.016505  0.8583298
2  0.03589881 -0.5799959 -0.4900833    -0.4697010        -0.5055784  0.03572811 -0.580860 -0.4904742
  Feb...KS.III Feb..KS.III.Plus     Mar..KPS  Mar..KS.I Mar..KS.II Mar...KS.III Mar..KS.III.Plus
1    0.8219768         0.8847622 -0.05680562  1.0166823  0.8583298    0.8219768         0.8847622
2   -0.4697010        -0.5055784  0.03246035 -0.5809613 -0.4904742   -0.4697010        -0.5055784
     Apr..KPS  Apr..KS.I Apr..KS.II Apr...KS.III Apr..KS.III.Plus     Mei..KPS  Mei..KS.I Mei..KS.II
1 -0.05416853  1.0172072  0.8583298     0.8219768         0.8847622 -0.05203720  1.0175641  0.8583298
2  0.03095345 -0.5812613 -0.4904742    -0.4697010        -0.5055784  0.02973554 -0.5814652 -0.4904742
  Mei...KS.III Mei..KS.III.Plus
1    0.8219768         0.8847622
2   -0.4697010        -0.5055784

Clustering vector:
 [1] 2 2 1 1 1 2 1 2 2 2 2

Within cluster sum of squares by cluster:
[1] 75.19452 73.23629
 (between_SS / total_SS =  40.6 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"
[7] "size"         "iter"         "ifault"
```

Data Processing with MS Excel

Repeated iterations were carried out to ensure that no cluster points moved. If during the repetition process the Cluster points remain and do not move, then the K-Means Clustering process is declared complete. In the second and subsequent iterations, determining the centroid center point is done by adding up the data from the cluster and dividing by the number of data. From the new centroid point in the second iteration, cluster group data is produced as in Table 1.

Tabel 1. Second Iteration Cluster Group Data

No	Village	C1	C2	Closest Distance	Cluster
1	Taja	1193,7	356,28	356,28	C2
2	Bumi	1345,1	565,00	565 ,00	C2
3	Cikl	685,74	1117,07	685,74	C1
4	Kedu	507,76	948,04	507,76	C1
5	Tari	634,5	1323,53	634,5	C1
6	Reja	939,52	476,71	476,71	C2
7	Sida	256,48	654,36	256,48	C1
8	Kali	1073,8	516,55	516,55	C2
9	Jati	641,83	359,83	359,83	C2
10	Bang	869,45	447,48	447,48	C2
11	Bojo	785,65	554,22	554,22	C2

The second iteration process uses the same methods and functions as in the first iteration. The results of the second iteration have the same results as the first, namely Cluster one consists of four villages, namely Cikl, Kedu, Tar, Sida Villages, while Cluster two consists of seven villages, namely Taja, Bumi, Reja, Kali, Jati, Bang, Bojo Villages. Because the results are the same, the calculation process is stopped and declared complete. Implementation of the K-Means algorithm using RStudio and MS Excel shows the same results, which can be seen in Table 2.

Table 2. Final Result

Cluster	RStudio	MS.Excel
C1	Cikl	Cikl
	Kedu	Kedu
	Tari	Tari
	Sida	Sida
C2	Taja	Taja
	Bumi	Bumi
	Reja	Reja
	Kali	Kali
	Jati	Jati
	Bang	Bang
	Bojo	Bojo

The results of implementing the K-Means algorithm with these two tools will later be used as information for the government and BPKB office staff as a form of evaluation and consideration in realizing the program planning objectives to improve the quality of life of village communities in XYZ District. The expansion of social assistance programs is the government's commitment to accelerate poverty reduction. By paying attention to the characteristics of each cluster, the programs that can be implemented are as follows. In villages with a high level of welfare, the appropriate program is (i) forming a village government that is professional, efficient and effective, open and responsible, (ii) strengthening government capacity and dynamic interaction between citizen organizations in administering government, (iii) building a responsive and participatory village planning and budgeting system, (iv) building independent and productive local economic institutions, (v) increasing the socio-cultural resilience of village communities in order to create village communities that are able to maintain social unity as part of national resilience and (vi) strengthening village communities as subjects of development

In villages with the low welfare level can be programmed: (i) Smart Indonesia Program (Program Indonesia Pintar – Kartu Indonesia Pintar -KIP), (ii) National Health Insurance Program (Jaminan Kesehatan Nasional - JKNKIS) and Family Hope Program (Program

Keluarga Harapan - PKH), (iii)) Improving the quality of the health sector, (iv) Improving infrastructure, (v) Economic empowerment carried out by BKKBN through the Prosperous Family Income Increasing Business (Usaha Peningkatan Pendapatan Keluarga Sejahtera - UPPKS) group and carrying out family development efforts through the Toddler Family Development (Bina Keluarga Balita - BKB).

CONCLUSION

Proses klasterisasi dilakukan menggunakan algoritma K-Means yang akan mengelompokkan data berdasarkan kesamaan karakteristiknya. Proses analisa data dilakukan menggunakan Tool Rstudio dan Microsoft Excel sebagai pembanding. Jumlah k optimal pada proses K-Means Clustering bernilai 2 (dua). Proses penghitungannya membutuhkan 2 (dua) kali iterasi. Karena sudah tidak terjadi perubahan Cluster pada iterasi kedua, maka langkah iterasi dinyatakan selesai. Hasil pengolahan data dengan kedua *tools* menunjukkan hasil yang sama yaitu terdapat 2 klaster. Klaster 1 merupakan klaster dengan tingkat kesejahteraan tinggi, terdiri dari 4 desa dan Klaster 2 merupakan klaster dengan tingkat kesejahteraan rendah, terdiri dari 7 desa. Berdasarkan hasil klasterisasi telah dapat disiapkan program-program yang diharapkan dapat meningkatkan kualitas hidup masyarakat.

The clustering process is carried out using the K-Means algorithm which groups data based on similar characteristics. The data analysis process was carried out using the RStudio tool and Microsoft Excel as a comparison. The optimal number of k in the K-Means Clustering process is two. The calculation process requires two iterations. Because there are no cluster changes in the second iteration, the iteration step is declared complete. The results of data processing with both tools show the same results, namely that there are two clusters. Cluster 1 is a cluster with a high level of welfare, consisting of four villages and Cluster 2 is a cluster with a low level of welfare, consisting of seven villages. Based on the results of the clustering, programs can be prepared which are expected to improve the quality of life of the community.

REFERENCES

Agustina Bidari, Teori Kependudukan, Bogor, Penerbit Lindan Bestari, 2020

[PemKab Kulonprogo] Pemerintah Kabupaten Kulonprogo, 2016, Pertumbuhan Penduduk dan Kualitas Hidup Kita, https://kulonprogokab.go.id/v31/detil/4527/pertumbuhan-penduduk-dan-kualitas-hidup-kita

Tarigan, P.M (2022). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus: Toko Sinar Harahap), *Jurnal Sistem Informasi, Teknologi Informasi dan Komputer,* 12 (2), 51-61

Prihati, Y (2021), Implementasi Algoritma K-Means Untuk Pemetaan Prestasi Akademik Siswa Disekolah Dasar Terang Bagi Bangsa Pati, *Jurnal Komputaki*, 7(1)

Rahman, A.T (2017), Coal Trade Data Clusterung Using K-Means (Case Study PT. Global Bangkit Utama), *ITSMART:Jurnal Ilmiah Teknologi dan Informasi, 6(1), 24-31*

Widiari, N.P.A. (2020). Teknik Data Cleaning Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali, *Jurnal Ilmiah Merpati*, 8(2)

Syaripul, A. N (2016), Visualisasi Data Interaktif Data Terbuka Pemerintah Provinsi DKI Jakarta: Topik Ekonomi Dan Keuangan Daerah, *Jurnal Sistem Informasi (Journal of Information Systems)*, 12(2), 82-89

[Pemerintah Republik Indonesia] Undang-Undang Republik Indonesia Nomor 52 Tahun 2009 tentang Perkembangan Kependudukan dan Pembangunan Keluarga. Lembaran Negara RI Tahun 2009 BAB II. Sekretariat Negara. Jakarta

R Core Team (2022). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing

Zhou Hong (2020). *Learn Data Mining Through Excel: A Step By Step for Understande Machine Learning Methods*. Springer