

# Implementasi K-Medoids Untuk *Clustering* Penduduk Dengan Pendidikan SMA Menurut Jenis Kelamin

Muhammad Faisal\*<sup>1</sup>, Wiranti Sri Utami<sup>2</sup>, Nila Pratiwi<sup>3</sup>

\*<sup>1</sup>Program Studi Teknik Informatika, Fakultas Sains & Teknologi, Universitas Raharja

<sup>2</sup>Program Studi Teknik Informatika, Fakultas Teknik, Universitas Cendekia Abditama

<sup>3</sup>Program Studi Teknik Elektro Fakultas Teknik, Universitas Cendekia Abditama

E-mail: \*<sup>1</sup>muhammad.faisal@raharja.info, <sup>2</sup>wirantisutami@uca.ac.id,

<sup>3</sup>nilapратиwi@uca.ac.id

## Abstrak

Pendidikan merupakan suatu tahapan dalam pelaksanaan belajar dan mengajar dimana dilaksanakan pada suatu sekolah. Salah satu program pemerintah dalam mendukung Pendidikan adalah Wajib belajar 12 tahun, yang dimaksud adalah generasi bangsa memiliki pendidikan minimal tamatan Sekolah Menengah Atas (SMA) Sederajat. Proses pengelompokan tamatan SMA sangat berguna untuk melihat siswa atau siswi yang lulus SMA di setiap Provinsi. Implementasi Data Mining menggunakan metode K-Medoids ini diharapkan dapat membantu pihak kepala daerah setempat dalam memantau kelompok usia dan jenis kelamin yang lulus SMA di setiap Provinsi masing-masing. Data yang didapat merupakan dataset berdasarkan pencarian data pada laman Badan Pusat Statistik atau BPS, kemudian dari sumber data tersebut dianalisa dengan model K-Medoids. Hasil dari analisa menggunakan metode K-Medoids menampilkan 2 cluster pada tahun 2021 Rata-rata penduduk laki-laki usia 25 tahun lulus SMA 58,2 sampai dengan 32,8. Pada tahun 2022 58,5 sampai 33,9. Sedangkan pada tahun 2021 rata-rata penduduk perempuan usia 25 tahun lulus SMA 52,5 sampai 26,9. Pada tahun 2022 54,0 sampai 29,3.

**Kata Kunci**—Data Mining, K-Medoids, Jenis Kelamin

## Abstract

*Education is a stage in the implementation of learning and teaching which is carried out in a school. One of the government's programs to support education is 12 years of compulsory education, which is meant by the nation's generation having a minimum education of a high school (SMA) equivalent. The process of grouping high school graduates is very useful for seeing students who have graduated from high school in each province. It is hoped that the implementation of Data Mining using the K-Medoids method can help local regional heads monitor the age and gender groups who graduate from high school in each province. The data obtained is a dataset based on data searches on the Central Statistics Agency or BPS website, then the data source is analyzed using the K-Medoids model. The results of the analysis using the K-Medoids method show 2 clusters in 2021. The average male population aged 25 years graduating from high school is 58.2 to 32.8. In 2022 58.5 to 33.9. Meanwhile, in 2021, the average rate for women aged 25 years graduating from high school will be 52.5 to 26.9. In 2022 54.0 to 29.3.*

**Keywords**— Data Mining, K-Medoids, Gender

## 1. PENDAHULUAN

Program wajib belajar awalnya 9 tahun namun seiring perkembangan jaman Pemerintah telah program wajib belajar menjadi 12 tahun. Dengan demikian penduduk usia muda harus mengenyam pendidikan minimal smpa bangku SMA. Kementrian pendidikan, Kebudayaan, Riset, dan Teknologi (KEMENDIKBUDRISTEK) telah membuat peta suatu dokumen tenaga kerja Indonesia minimal berpendidikan formal 12, yang artinya minimal berpendidikan SMA [1]. Pendidikan yang baik dan memiliki mutu yang bagus sangat dibutuhkan demi terciptanya kualitas individu, saat ini setiap negara berlomba-lomba untuk merencanakan pendidikan setiap warganya, demi melahirkan Sumber Daya Manusia (SDM) yang unggul dengan menerapkan system Pendidikan yang baik [2]. Saat ini sudah banyak pelaku usaha menyerap tenaga pekerja lulusan SMA sederajat [3]. Pertumbuhan ekonomi memberikan kesempatan bagi siapa saja untuk bersaing tidak hanya mengandalkan lulusan SMA namun melihat dair *gender*. Tingkat Partisipasi Angkatan Kerja (TPAK) pada data yang diperoleh dari BPS menunjukkan antara Perempuan dan Laki-laki 50,89:82,51 (BPS, 2017) [4]. Dalam hal mendata atau melakukan analisa terkait penduduk dengan pendidikan usia 25 tahun keatas yang lulus SMA berdasarkan jenis kelamin diperlukan suatu metode analisa untuk mengelompokkan data tersebut. Dalam upaya mendukung penelitian ini penulis menggunakan teknik *Data Mining* dengan menerapkan metode K-Medoids. Hal ini dapat memberikan informasi wilayah mana saja yang memiliki Jumlah penduduk dengan minimal Pendidikan SMA menurut gender. Pengelompokkan ini diharapkan dapat membantu Pemerintah terkait dalam hal mendata Penduduk yang minimal Pendidikannya SMA.

Landasan dalam hal pengambilan keputusan demi tercapainya akan informasi atau sebuah pengetahuan dengan mendapatkan data yang melimpah dalam sebuah instansi terkait. Hal ini menjadi titik awal perkembangan teknologi *Data Mining* yang mana dengan harapan dapat membantu dalam hal mengelola informasi yang terdapat dalam sebuah data[5]. *Clustering* merupakan suatu pendekatan pada *Data Mining* yang bertujuan dalam hal pengelompokkan Obyek-obyek yang terdapat didalam suatu data kedalam *cluster* [6]. Proses *clustering* berbeda dengan Klasifikasi, klasifikasi tidak memiliki variable target dalam hal pengelompokkan, kualitas hasil dari proses *cluster* tergantung dari metode yang dipilih. *Data Mining* memiliki beberapa cara untuk mengukur nilai kesamaan obyek, salah satunya menggunakan *Euclidean Distance* [7]. Salah satu algoritma yang mendukung proses *Data Mining* adalah K-Medoids, algoritma ini dikenal juga dengan istilah PAM (*Partitioning Around Medoids*), algoritma ini hampir sama dengan metode K-Mean, yang mana pada kedua algoritma tersebut dipisahkan dan dikumpulkan data nya lalu dalam proses tersebut dibagi menjadi beberapa kelompok [8].

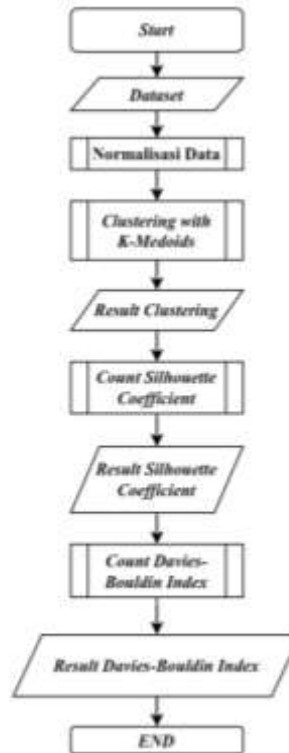
Pembahasan yang sudah dipaparkan diatas, maka penelitian ini ditentukan dengan menggunakan metode K-Medoids untuk *clustering* Penduduk dengan Minimal Pendidikan SMA usai 25 tahun keatas berdasarkan Jenis Kelamin.

## 2. METODE PENELITIAN

Data yang berasal dari BPS (Badan Pusat Statistik) merupakan hasil dari penelitian yang telah dilakukan dengan Persentase terhadap Penduduk disuatu daerah dengan usia Usia 25 Tahun keatas dengan pendidikan SMA menurut jenis kelamin 2021-2022. Pengolahan data dilakukan dengan menggunakan Teknik *Data Mining*, dan menggunakan Metode *Clustering* K-Medoids. Proses penelitian ini menggunakan Bahasa R, Penelitian ini bertujuan untuk mengamati setiap atribut *cluster* yang terbentuk.

### 2.1. *Data Mining*

*Data Mining* merupakan sebuah proses pada suatu pola/ hubungan antar cluster dalam penggunaan sebuah data. Hasil yang didapat dari proses *data mining* tersebut ialah untuk mengambil suatu keputusan [9]. *Data Mining* adalah metode dalam mengamati dan menganalisa data yang memiliki jumlah *record* yang banyak, langkah tersebut dilakukan untuk mengamati hubungan data yang memiliki kesamaan dan meringkas data sehingga dapat dilakukan pengambilan keputusan[10]. Proses penelitian menggunakan algoritma dengan model K-Medoids dapat dilihat pada susunan *flowchart* sebagai berikut.



**Gambar 1.** Flowchart Algoritma Dengan Model K-Medoids

## 2.2. Clustering

*Clustering* merupakan metode yang digunakan untuk menganalisa data, metode tersebut merupakan salah satu dari metode *data mining*, *Clustering* bertujuan untuk mengelompokkan data yang memiliki karakteristik yang sama kedalam suatu wilayah yang lainnya [11]. *Clustering* biasa disebut dengan segmentasi, Metode tersebut digunakan untuk indentifikasi suatu kelompok dalam suatu penelitian yang berlandaskan terhadap kelompok yang memiliki kesamaan [12].

## 2.3. K-Medoids

Model K-Medoids yang terdapat pada *Data Mining* digunakan agar dapat menjadi solusi dari sebuah permasalahan Clustering atau pengelompokkan. Algoritma ini lebih efisien untuk data dalam jumlah kecil [13]. Proses penerapan algoritma dari K-Medoids adalah dengan menentukan *Cluster* [14]. Pada penelitian ini jumlah *cluster* yang diperoleh 2 *Cluster*.

### 2.3.1. Euclidean Distance

Proses pencarian nilai K dalam penelitian ini menggunakan rumus jarak *Euclidean* yang merupakan generalisasi teorema *Pythagoras* [15]. Rumus dari tersebut sebagai berikut.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \quad (1)$$

## 2.4. Identifikasi Masalah

Penelitian ini memperoleh data dari laman BPS, Data tersebut merupakan presentase Penduduk dengan Pendidikan SMA usia 25 tahun menurut jenis kelamin. Presentase Penduduk dengan Pendidikan SMA usia 25 tahun menurut jenis kelamin (BPS) dapat dilihat pada Tabel 1 dibawah ini.

**Tabel 1. Jumlah Presentase Penduduk Dengan Pendidikan SMA  
Usia Diatas 25 Tahun Menurut Jenis Kelamin (BPS)**

Provinsi/Kabupaten/Kota	Presentase Penduduk Usia Diatas 25 Tahun Pendidikan SMA ke atas Menurut Jenis Kelamin (Persen)			
	Laki-laki		Perempuan	
	2021	2022	2021	2022
ACEH	46,93	47,72	42,05	42,91
SUMATERA UTARA	51,38	54,36	47,5	49,82
SUMATERA BARAT	43,74	45,32	44,97	48,52
RIAU	46,39	45,76	43,04	43,24
JAMBI	39,74	40,53	33,73	35,28
SUMATERA SELATAN	37,94	38,77	32,99	35,75
BENGKULU	41,64	42,62	37,43	38,72
LAMPUNG	33,72	34,88	28,77	30,44
KEP. BANGKA BELITUNG	37,52	40,76	33,58	36,08
KEPULAUAN RIAU	60,43	60,73	59,2	58,4
DKI JAKARTA	69,07	72,68	59,76	64,51
JAWA BARAT	40,99	41,24	33,12	35,36
JAWA TENGAH	32,26	34,14	26,82	28,79
D I YOGYAKARTA	52,14	52,33	47,37	47,12
JAWA TIMUR	36,65	36,5	29,61	30,95
BANTEN	46,65	47,38	36,88	40,14
BALI	52,5	53,52	40,07	43,18
NUSA TENGGARA BARAT	38,17	36,78	27,21	29,68
NUSA TENGGARA TIMUR	33,47	31,15	29,14	28,45
KALIMANTAN BARAT	32,76	32,82	26,99	29,03
KALIMANTAN TENGAH	33,93	34,54	29,37	30,3
KALIMANTAN SELATAN	36,72	36,72	31,17	31,02
KALIMANTAN TIMUR	55,6	55,46	46,36	50,12
KALIMANTAN UTARA	46,07	47,27	38,66	41,48
SULAWESI UTARA	45,8	46,22	45,72	46,79
SULAWESI TENGAH	37,55	36,58	33,54	34,48
SULAWESI SELATAN	40,63	41,53	35,35	38,38
SULAWESI TENGGARA	43,1	42,12	36,18	38,11
GORONTALO	29,75	31,02	32,99	34,65
...	...	...	...	...
PAPUA	36,7	38,8	26,44	30,8

### 3. HASIL DAN PEMBAHASAN

Data yang akan diolah menggunakan data yang diperoleh dari BPS, dengan data Penduduk minimal Pendidikan SMA usia 25 tahun menurut jenis kelamin tahun 2021 dan 2022. Atribut yang digunakan adalah Laki-laki dan Perempuan. Tahap pertama adalah memasukan *dataset* kedalam *tools* R Studio dan penulis menggunakan Posit Cloud. Tahap awal dalam memasukan data tersebut dapat dilihat pada gambar 2 sebagai berikut.

```
> dataset <- read_excel("Persentase penduduk usia 25 Tahun keatas dengan pendidikan SMA ke
atas menurut jenis kelamin, 2021-2022.xlsx")

> dataset
# A tibble: 548 x 5
  `Provinsi/Kabupaten/Kota` `Laki-laki 2021` `Laki-laki 2022` `Perempuan 2021`
  <chr>                       <dbl>           <dbl>           <dbl>
1 ACEH                        46.9            47.7            42.0
2 Simeulue                    52.0            51.0            40
3 Aceh Singkil                42.2            38.4            38.8
4 Aceh Selatan                38.2            37.5            32.6
5 Aceh Tenggara               54.4            55.5            43.3
6 Aceh Timur                  33.5            32.2            28.8
7 Aceh Tengah                 48.5            48.7            45.2
8 Aceh Barat                  49.5            51.3            45.9
9 Aceh Besar                  53.7            52.1            51.6
10 Pidie                      40.1            42.4            33.0
# i 538 more rows
# i 1 more variable: `Perempuan 2022` <dbl>
# i Use `print(n = ...)` to see more rows
> |
```

**Gambar 2. Dataset Penelitian**

Setelah tahap awal dilakukan, proses selanjutnya adalah *preprocessing*. *Preprocessing* adalah langkah untuk pembersihan data. Proses tersebut digunakan untuk melakukan *review* terhadap data *missing value*, tahapan pada *missing value* sesuai dengan gambar 3 berikut ini.

```
> anyNA(dataset)
[1] FALSE
>
```

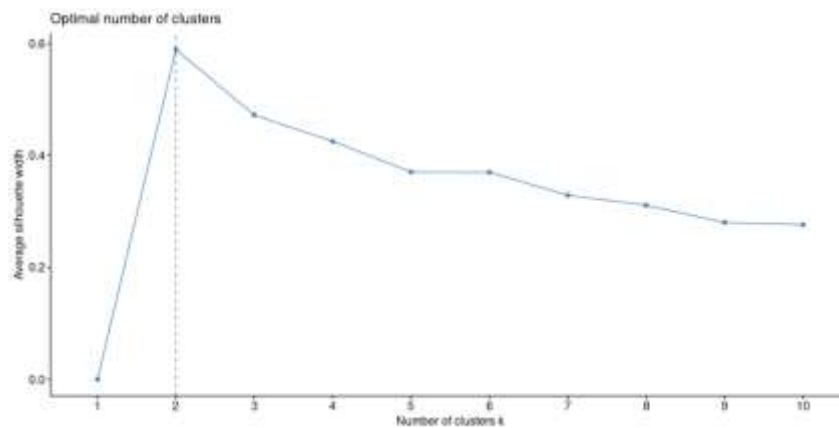
**Gambar 3. Missing Value**

Pada gambar 3 memperoleh hasil FALSE yang memiliki arti tidak ada data *missing value*. Proses selanjutnya adalah *transformation*. *Transformation* merupakan proses untuk normalisasi data, proses tersebut dilakukan untuk membuat atribut numerik skala terhadap jarak data lebih kecil yang bertujuan untuk proses skala data tidak terlalu jauh. Dalam tahapan normalisasi suatu data ditampilkan dalam gambar 4 sebagai berikut.

```
> datasetnorm <- scale(dataset[,2:5])
> datasetnorm
      Laki-laki 2021 Laki-laki 2022 Perempuan 2021 Perempuan 2022
[1,] 0.456424749 4.630208e-01 0.513152851 0.441184384
[2,] 0.819938313 6.995850e-01 0.370332821 0.263505219
[3,] 0.119180719 -2.169198e-01 0.284640803 0.043555608
[4,] -0.163394279 -2.815033e-01 -0.145212653 -0.543931952
[5,] 0.983945410 1.026131e+00 0.598148186 0.695523511
[6,] -0.499218334 -6.631989e-01 -0.406468806 -0.545364849
[7,] 0.567892775 5.334096e-01 0.729821287 0.646088582
[8,] 0.639601506 7.206291e-01 0.779982468 0.628893824
[9,] 0.936376252 7.794073e-01 1.180575236 1.232143246
[10,] -0.027076692 7.769695e-02 -0.119435380 0.099438571
[11,] -0.129314882 8.059957e-02 0.278370655 0.184695912
[12,] -0.147064568 -2.607201e-02 -0.045586974 -0.202186140
[13,] -0.182563940 -2.270790e-01 0.194768687 0.030659540
[14,] 0.171009801 4.405586e-03 -0.300572979 -0.447211439
[15,] 0.060251762 4.173044e-01 0.281157388 0.461244935
[16,] 0.303067463 4.477820e-01 0.148090921 0.171799844
[17,] -0.159844342 -3.105296e-01 -0.221151108 -0.272398068
[18,] 0.858987622 3.490927e-01 0.874034683 0.600235894
[19,] 0.475594409 8.127875e-01 0.165507997 0.178247878
[20,] 2.863992129 3.084820e+00 3.236486986 2.970246686
[21,] 1.843030202 1.479666e+00 2.017988290 1.602546988
[22,] 1.989287613 1.888211e+00 1.786689510 1.770195877
[23,] 1.843030202 1.482569e+00 2.194945791 1.596098954
```

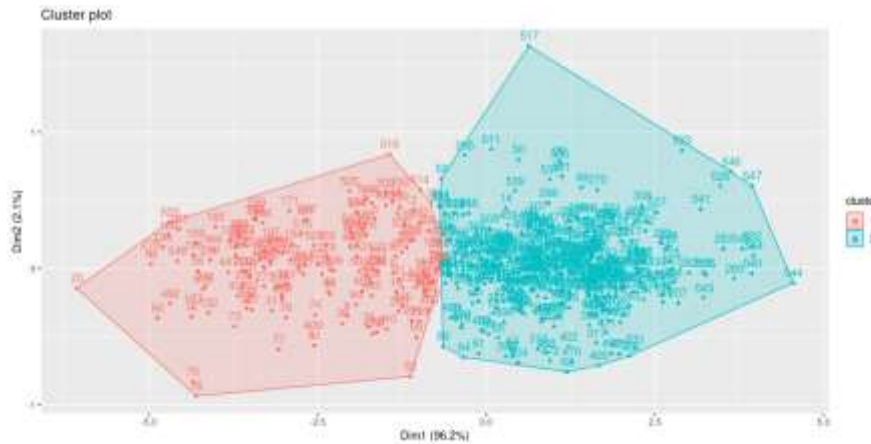
Gambar 4. Tahapan Normalisasi Suatu Data

Proses berikutnya masuk kedalam proses yang dinamakan *data mining*, dimana memiliki tujuan dalam menunjukkan adanya suatu model terhadap penggalian dari data yang ada untuk mendapatkan informasi dengan tepat menggunakan algoritma. K-Medoids adalah Algoritma yang dipergunakan dalam penelitian ini, tahapan awal dari algoritma K-Medoids menentukan jumlah *cluster* menggunakan metode *silhouette coefficient*. Berikut ini merupakan gambar dari hasil dari metode *silhouette coefficient*.



Gambar 5. Tahapan *silhouette coefficient*

Gambar di atas menunjukkan hasil *cluster* yang optimal terhadap *dataset* dari penelitian ini adalah 2 *cluster*, Hasil tersebut digunakan untuk proses *clustering* terhadap suatu algoritma yaitu dengan model K-Medoids. Proses mengelompokkan suatu data ditampilkan dalam bentuk gambar tahapan *clustering* sebagai berikut.



**Gambar 6. Cluster Plot Cluster K-Medoids**

Hasil dari *cluster* diatas menunjukkan 2 *cluster* optimal dari algoritma K-Medoids, kemudian dilakukan proses *cluster* vector untuk mengetahui hasil dari anggota masing-masing. Proses tersebut ditampilkan dalam bentuk gambar sebagai berikut.

```
> pan.result <- pan(dataset[,2:5],2)
> print(pan.result)
Medoids:
  ID Laki-laki 2021 Laki-laki 2022 Perempuan 2021 Perempuan 2022
[1,] 528      58.18      58.73      52.07      54.08
[2,] 200      32.26      34.14      26.82      28.79
Clustering vector:
[1] 1 1 2 2 1 2 1 1 1 1 2 2 2 2 2 2 2 2 1 1 1 1 1 1 2 1 2 2 2 2 1 1 1 1 2 2 1 1 1 1 2 2 1 1 1 1 2 2
[40] 1 1 1 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[79] 1 2 2 2 2 1 2 2 1 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[118] 1 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[157] 1 2 1 1 2 2 1 1 1 2 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[196] 1 1 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[235] 1 1 1 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[274] 1 1 1 1 1 1 1 2 2 2 2 2 2 1 1 2 1 1 2 1 1 1 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[313] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[352] 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[391] 2 2 2 1 1 2 1 2 1 2 1 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[430] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[469] 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[508] 2 1 1 2 2 2 1 1 1 2 1 2 1 2 1 1 1 1 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[547] 2 1
Objective function:
  build  swap
14.54795 13.86030
```

**Gambar 7. Hasil pengelompokkan data (*cluster*) vector**

Hasil dari *cluster* algoritma K-Medoids diatas menunjukkan pada tahun 2021 Penduduk laki-laki dengan usia 25 tahun dengan minimal pendidikan SMA terdiri dari 58,18% dan 32,26%, dan ditahun 2022 58,73% dan 34,14%. Sedangkan pada tahun 2021 Penduduk Perempuan dengan usia 25 tahun keatas dengan minimal pendidikan SMA terdiri dari 52,07% dan 26,82%, dan pada tahun 2022 54,08% dan 28,79%.

Proses terakhir adalah evaluasi hasil *cluster* sebagai tahap akhir KDD, proses tersebut menggunakan suatu metode yang disebut dengan *Davies-Bouldin Index*. Metode *Davies-Bouldin Index* adalah suatu teknik dimana dipergunakan dalam mengevaluasi hasil *clustering* dengan memaksimalkan hasil jarak setiap obyek *cluster*. Hasil dari proses *Davies-Bouldin Index* dapat diketahui pada gambar 8 berikut ini.

```
> library(clusterSim)
> (index.DB(data.cluster[,2:5],
+          pam.result$clustering,
+          centrotypes = 'medoids',
+          d = dist(data.cluster[,2:5])))
$DB
[1] 0.6562554
```

**Gambar 8. Hasil Tahapan *Davies-Bouldin Index***

Hasil yang diperoleh dari metode DBI sebesar 0,6562554. Hasil *cluster* yang optimal jika menggunakan DBI hampir 0. Sehingga dapat dikatakan hasil *clustering* menggunakan 2 *cluster* mempunyai kesamaan data nilai disetiap anggota *cluster*.

Dengan dihasilkannya *cluster* data Penduduk dengan rentang usia diatas 25 tahun yang memiliki latar belakang pendidikan minimal Sekolah Menengah Atas, berdasarkan pada jenis kelamin dapat membantu pemerintah terkait dalam mendata penduduk di setiap Provinsi.

#### 4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat diambil kesimpulan:

1. *Data mining* dengan Algoritma K-Medoids dapat digunakan untuk analisis Penduduk dengan Pendidikan SMA disetiap Provinsi dengan rata-rata usia 25 tahun keatas.
2. Kualitas *cluster* yang diperoleh dari pengujian didapatkan hasil nilai 0,6562554 menggunakan metode *Davies-Boulding Index*, dengan jumlah *cluster* 2 dan total data 547. Hasil dari *cluster* tersebut dapat disimpulkan jumlah *cluster* dan jumlah data dapat mempengaruhi hasil dari kualitas *cluster*.
3. Hasil dari penelitian ini menunjukkan tingkat kelulusan Penduduk dengan Pendidikan SMA dan rata-rata usia 25 tahun, Penduduk Pria pada tahun 2021 sebanyak 58,18% dan 32,26%, dan ditahun 2022 58,73% dan 34,14%. Sedangkan penduduk Perempuan pada tahun 2021 sebanyak 52,07% dan 26,82% dan ditahun 2022 54,08% dan 28,79%. Dari hasil tersebut menunjukkan Penduduk Pria dengan minimal Pendidikan SMA memiliki presentase lebih besar.
4. Diharapkan dengan hasil penelitian ini dapat memberikan informasi terhadap pihak terkait Provinsi ataupun Pemerintah mengenai indeks rata-rata Penduduk dengan Pendidikan SMA minimal usia 25 tahun berdasarkan jenis kelamin, dan dapat memberikan suatu kebijakan khusus agar penduduk setiap Provinsi memiliki rata-rata Pendidikan SMA lebih tinggi ditahun berikutnya

#### 5. SARAN

Dengan adanya penelitian ini penulis memiliki harapan bahwa informasi yang disampaikan dari penelitian ini dapat membantu Pemerintah terkait dalam mengambil suatu keputusan. Penulis juga berharap akan adanya penelitian lebih lanjut yang membahas penelitian ini menggunakan metode lainnya.

## DAFTAR PUSTAKA

- [1] S. J. D. R. "Dewan Perwakilan Rakyat Republik Indonesia," DPR RI, 24 Januari 2023. [Online]. Available: <https://www.dpr.go.id/berita/detail/id/42889/t/Harus%20Ada%20%27Political%20Will%27%20Wujudkan%20Wajib%20Belajar%2012%20Tahun>. [Accessed 11 Desember 2023].
- [2] I. Margiyanti and S. T. Maulia, " Kebijakan Pendidikan Implementasi Program Wajib Belajar 12 Tahun," *Pendidikan dan Sastra Inggris*, vol. 3, no. 1, pp. 200-208, 2023.
- [3] S. Hidayat and M. Saleh, "Kemampuan Kerja antara Pekerja Lulusan SMA dan SMK di Industri Manufaktur: Studi Komparasi," *Valid Jurnal Ilmiah*, vol. 17, no. 1, pp. 16-21, 2020.
- [4] Y. Nuraeni and I. L. Suryono, "Analisis Kesetaraan Gender dalam Bidang Ketenagakerjaan di Indonesia," *Nahkoda : Jurnal Ilmu Pemerintahan*, vol. 20, no. 1, pp. 68-79, 2021.
- [5] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *Informatika*, vol. 7, no. 1, pp. 1-7, 2020.
- [6] E. Rahmad, E. Haerani, A. Nazir and S. Ramadhani, "Penerapan Algoritma K-Medoids Clustering Untuk Menentukan Strategi Promosi Pada Data Mahasiswa (Studi Kasus : Stikes Perintis Padang)," *JNKTI*, vol. 5, no. 3, pp. 556-564, 2022.
- [7] B. Wira, A. E. Budianto and A. S. Wiguna, "Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang," *RAINSTEK*, vol. 1, no. 3, pp. 54-69, 2019.
- [8] M. Arifandi, A. Hermawan and D. Avianto, "Implementasi Algoritma K-Medoids Untuk Clustering Wilayah Terinfeksi Kasus Covid19 Di Dki Jakarta," *JTT*, vol. 7, no. 2, pp. 120-128, 2021.
- [9] P. N. Harahap and S. , "Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT.Arma Anugerah Abadi Cabang Sei Rampah)," *MATICS*, vol. 11, no. 2, pp. 46-50, 2019.
- [10] M. Faisal, S. and D. Prastyo, "Implementasi Data Mining Menggunakan Metode K-Means Dalam Pengelompokan Bayi Penerima ASI Eksklusif," *ICIT*, vol. 9, no. 2, pp. 198-207, 2023.
- [11] L. Tanti, "Metode Data Mining Clustering," Binus University School of Information Systems, 29 Oktober 2021. [Online]. Available: <https://sis.binus.ac.id/2021/10/29/metode-data-mining-clustering/>. [Accessed 13 Desember 2023].
- [12] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means," *SIMETRIS*, vol. 11, no. 2, pp. 690-703, 2020.
- [13] S. Bahri and D. M. Midyanti, "PENERAPAN METODE K-MEDOIDSUNTUK PENGELOMPOKAN MAHASISWA BERPOTENSI DROP OUT," *JTHIK*, vol. 10, no. 1, pp. 165-172, 2023.
- [14] W. Kurniawan, A. Rifai, W. Gata and D. Gunawan, "Analisis Algoritma K-Medoids Clustering Dalam Menentukan Pemesanan Hotel," *SWABUMI*, vol. 8, no. 2, pp. 182-187, 2020.
- [15] M. Jannah and N. Humaira, "Implementasi Metode Euclidean Distance Untuk Ekstraksi Fitur Jarak Pada Citra Skeleton," *Ilmiah Informatika Komputer*, vol. 24, no. 2, pp. 134-139, 2019.