J-Com (Journal of Computer) ISSN 2775-801X (online)
Vol. 4 No. 1, Maret 2024, hlm. 33 – 39
DOI: xxxxxxxxxxxxxxxx
Available online at https://jurnal.stmikroyal.ac.id/index.php/j-com

PENGELOMPOKAN DATA PENDUDUK MISKIN DI SUMATERA UTARA MENGGUNAKAN *K-MEANS*

Ayu Safitri, Rizki Rahmawati*, Sri Ayu Wandira
Mahasiswa Prodi Sistem Informasi
Sekolah Tinggi Manajemen Informatika dan Komputer Royal
**email*: rizkirahmawati.royal@gmail.com

Abstract: The government has made many efforts to eliminate poverty in society, including pro- poor programs, basic food assistance, and cash assistance that help achieve the standard of a prosperous society. The results of data processing help decision making. Given the large amountof public data, it is very difficult for the government to identify who is poor, as is the case in NorthSumatra. The K-Means Clustering (Multidimensional) method is used in this research to facilitate the identification of patterns and structures in data that are difficult to see in the original representation. This algorithm produces three groups, where group 0 has a high population povertylevel of 1, group 1 has a moderate population poverty level of 4 and group 2 has a low population poverty level of 28.

Keywords: *Data Mining; K-Means:* Poor*;* Resident

Abstrak: Pemerintah telah melakukan banyak upaya untuk menghilangkan kemiskinan di masyarakat, termasuk program pro-poor, bantuan sembako, dan bantuan uang tunai yang membantu mencapai standar masyarakat sejahtera. Hasil pengolahan data membantu pengambilan keputusan. Mengingat banyaknya data masyarakat, sangat sulit bagi pemerintah untuk mengidentifikasi siapa yang miskin, sama halnya di Sumatera Utara. Metode K-Means Clustering (Multidimensi) digunakan dalam penelitian ini untuk memudahkan identifikasi pola dan struktur dalam data yang sulit dilihat dalam representasi aslinya. Algoritma ini menghasilkan tiga kelompok, di mana kelompok 0 memiliki tingat kemiskinan penduduk yang tinggi sebanyak 1, kelompok 1 memiliki tingat kemiskinan penduduk yang sedang sebanyak 4 dan kelompok 2 memiliki tingat kemiskinan penduduk yang rendah sebanyak 28.

Kata kunci: *Data Mining; K-Means;* Miskin; Penduduk

PENDAHULUAN

Kemiskinan adalah suatu kondisi di mana seseorang tidak memiliki apa-apa untuk memenuhi kebutuhan dasar mereka, seperti makanan, pakaian, dan tempat tinggal. Dengan asumsi bahwa konsep kemiskinan ini spesifik pada waktu dan masyarakat, artinya tidak berlaku universal karena ukuran kemiskinan berbeda untuk setiap masyarakat dan kurun waktu [1]. Karena dampak yang ditimbulkan kemiskinan sangat besar, kemiskinan di suatu daerah selalu menjadi masalah yang serius. Dari segi ekonomi, mereka yang miskin tidak akan mampu memenuhi kebutuhan dasar makanan, yang berdampak pada kekurangan gizi. Mereka juga akan kesulitan mendapatkan pendidikan menengah kebawah, yang membuat mereka sulit bersaing di pasar tenaga kerja, yang pada gilirannya akan menyebabkan pengangguran. Ini adalah alasan mengapa pengentasan kemiskinan harus dilakukan. Meskipun kemiskinan

J-Com (Journal of Computer) ISSN 2775-801X (online)
Vol. 4 No. 1, Maret 2024, hlm. 33 – 39
DOI: xxxxxxxxxxxxxxxx
Available online at https://jurnal.stmikroyal.ac.id/index.php/j-com

tidak dapat dihilangkan sama sekali, program pembangunan yang berkelanjutan dapat membantu menguranginya [2].

Di Indonesia sendiri, ada dua kelompok besar yang berpartisipasi dalam upaya pengentasan kemiskinan. Kelompok pertama terdiri dari program khusus untuk orang miskin. Keluarga miskin akan benar-benar mendapatkan manfaat dari program-program ini jika mereka berjalan dengan baik. Program untuk kelompok pertama sangat bergantung pada pentargetan awal yang akurat untuk memastikan penerima manfaat yang tepat. Kelompok kedua terdiri dari program yang secara proporsional akan memberi manfaat lebih banyak kepada orang miskin daripada orang dari semua golongan pendapatan [3]. Dengan mengeksplorasi kumpulan data, teknologi data mining dapat menghasilkan informasi yang belum pernah terpikirkan sebelumnya [4].

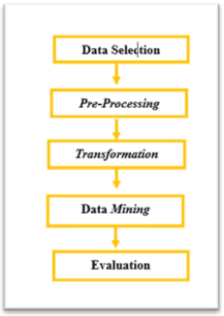
Mengingat banyaknya data penduduk, maka untuk mengetahui penduduk miskin bukanlah hal mudah yang dilakukan oleh pemerintah, sama halnnya di Sumatera Utara. Oleh sebab itu, pemerintah memerlukan sebuah metode daya mining clustering (*multidimensi*) yaitu *K-Means* untuk menyelesaikan permasalahan tersebut dengan cepat, tepat dan akurat.

Cluster adalah sekumpulan objek data yang mirip satu sama lain dalam kumpulan yang sama dan mirip dengan kumpulan yang berbeda [5]. Terdapat beberapa algoritma clustering, salah satunya *K-Means. K-Means* adalah teknik *clustering* data non-hirarki yang bertujuan untuk membagi data saat ini ke dalam satu atau lebih kelompok atau *cluster*. Ini memungkinkan data dengan atribut yang serupa dikelompokkan ke dalam kelompok yang sama, dan data dengan atribut yang berbeda dikelompokkan ke dalam kelompok yang berbeda.

Penelitian ini dapat membantu pemerintah dalam menentukan daerah mana yang memiliki tingkat kemiskinan tinggi secara lebih akurat, sehingga pemerintah dapat lebihberkonsentrasi untuk memberikan bantuan.

METODE

Dalam penelitian ini, *Knowledge Discovery in Database* (KDD) digunakan. KDD adalah proses komputasi yang menggunakan algoritma matematika untuk mengekstraksi data dan menghitung kemungkinan tindakan yang mungkin terjadi di masa depan. Pengetahuan baru, potensial, dan bermanfaat diperoleh sebagai hasil dari KDD [6]. Berikut adalah tahapan metode KDD yang digunakan dalam penelitian ini:



Gambar 1 menunjukkan alur penelitian untuk *Knowledge Discovery in Database* (KDD)

J-Com (Journal of Computer) ISSN 2775-801X (online)
Vol. 4 No. 1, Maret 2024, hlm. 33 – 39
DOI: xxxxxxxxxxxxxxxx
Available online at https://jurnal.stmikroyal.ac.id/index.php/j-com

Data *Selection*

Tahap pengambilan dan pemilihan sampel data yang ingin diolah untuk menghasilkan informasi penting dikenal sebagai data *selection* [7]. Proses ini melakukan ini sambil mempertahankan data aslinya [8]. Data yang dipilih berasal dari data tahunan dari setiap kabupaten di Sumatera Utara. Gambar 2 menunjukkan data yang akan digunakan dalam penelitian ini.

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt

        df = pd.read_excel('Jumlah Penduduk Miskin di Sumatera Utara.xlsx')
        df
```

Out[1]:

	NO	Kabupaten/Kota	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
0	1	Nias	23.28	22.21	24.53	24.11	24.88	22.61	22.10	23.12	24.33	23.23
1	2	Mandailing Natal	40.69	39.68	47.79	47.67	48.30	42.39	40.64	41.31	43.24	40.98
2	3	Tapanuli Selatan	30.77	29.38	31.20	30.84	29.48	25.63	24.22	23.96	25.01	23.05
3	4	Tapanuli Tengah	52.00	49.86	52.20	51.77	53.05	48.53	46.99	47.19	49.95	47.07
4	5	Tapanuli Utara	33.75	32.23	33.37	33.20	33.75	29.20	28.57	28.41	29.72	27.47
5	6	Toba	16.96	16.51	18.31	18.20	18.49	15.82	15.78	16.05	16.61	16.48
6	7	Labuhan Batu	38.14	37.35	41.63	41.94	42.35	41.70	41.52	42.17	45.03	43.27
7	8	Asahan	80.54	76.97	85.16	84.35	83.67	74.14	70.53	66.32	69.29	64.49
8	9	Simalungun	87.72	86.25	92.89	92.19	91.35	80.30	76.33	73.64	76.99	72.47
9	10	Dairi	24.00	23.35	25.33	24.94	24.98	23.19	21.86	22.93	23.72	22.53
10	11	Karo	36.93	35.36	37.52	38.74	40.02	35.36	34.08	36.57	38.01	35.93
11	12	Deli Serdang	91.97	90.92	95.65	100.09	97.09	88.52	84.94	86.26	92.52	85.28
12	13	Langkat	104.31	100.63	114.19	115.79	114.41	105.46	103.08	101.87	106.59	100.45
13	14	Nias Selatan	56.96	54.46	58.97	57.75	57.95	52.70	52.51	53.88	55.16	54.16
14	15	Humbang Hasundutan	17.94	17.14	18.04	18.04	18.35	16.93	16.60	17.92	18.71	17.33
15	16	Pakpak Bharat	4.94	4.72	5.12	4.95	4.95	4.66	4.52	4.59	4.79	4.52
16	17	Samosir	17.18	16.27	17.64	18.01	18.43	16.81	15.79	15.80	16.08	14.97
17	18	Serdang Bedagai	56.55	54.46	58.30	58.17	56.93	50.49	48.69	49.18	51.16	48.22

Gambar 2. Data Penduduk Miskin di Sumatera Utara

Data *Pre-Processing*

Preprocessing data adalah proses mengubah data menjadi format yang lebih sederhana, efisien, dan sesuai dengan kebutuhan pengguna [9]. Proses perbaikan dimulai dengan memeriksa data untuk memastikan apakah ada penggandaan atau ketidakkonsistenan, dan memperbaiki kesalahan, seperti kesalahan huruf. Komponen yangmembedakan entitas yang berbeda dibahas pada tahap kedua, integrasi [10].

Data *Transformation*

Transformasi data adalah proses yang mengintegrasikan dan mengubah berbagaiskema dan struktur ke dalam skema dan struktur yang ditetapkan di gudang data [11]. Gambar 3 menunjukkan tabel atribut yang digunakan pada dataset pertahun penduduk miskin di Sumatera Utara.

```
In [2]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 33 entries, 0 to 32
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   NO              33 non-null     int64
 1   Kabupaten/Kota  33 non-null     object
 2   2013            33 non-null     float64
 3   2014            33 non-null     float64
 4   2015            33 non-null     float64
 5   2016            33 non-null     float64
 6   2017            33 non-null     float64
 7   2018            33 non-null     float64
 8   2019            33 non-null     float64
 9   2020            33 non-null     float64
 10  2021            33 non-null     float64
 11  2022            33 non-null     float64
dtypes: float64(10), int64(1), object(1)
memory usage: 3.2+ KB
```

Gambar 3. *Atribut Dataset*

J-Com (Journal of Computer) ISSN 2775-801X (online)
Vol. 4 No. 1, Maret 2024, hlm. 33 – 39
DOI: xxxxxxxxxxxxxxxx
Available online at https://jurnal.stmikroyal.ac.id/index.php/j-com

Data *Mining*

Data mining adalah bidang keilmuan yang menggabungkan teknik pembelajaran mesin, pengenalan pola, *statistic*, *database*, dan visualisasi untuk menangani masalah pengambilan informasi dari *database* yang berbeda dengan cara yang dapat dipahami dan bermanfaat bagi pemilik data. Data mining adalah analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dari sebelumnya [12]. *Dataset* yang telah dikumpulkan kemudian diproses. Pada data ini, metode *clustering multidimensi K-Means* digunakan.

Gambar 4. Visualisasi Dataset

Knowledge Interpretation/Evaluation.

Proses ini dilakukan untuk mencari pengetahuan, yang mencakup memeriksaapakah pola atau informasi yang dikumpulkan bertentangan atau sesuai dengan faktaatau asumsi dari data sebelumnya.

HASIL DAN PEMBAHASAN

Pengelompokan titik data menjadi dua atau lebih kelompok sehingga titik data

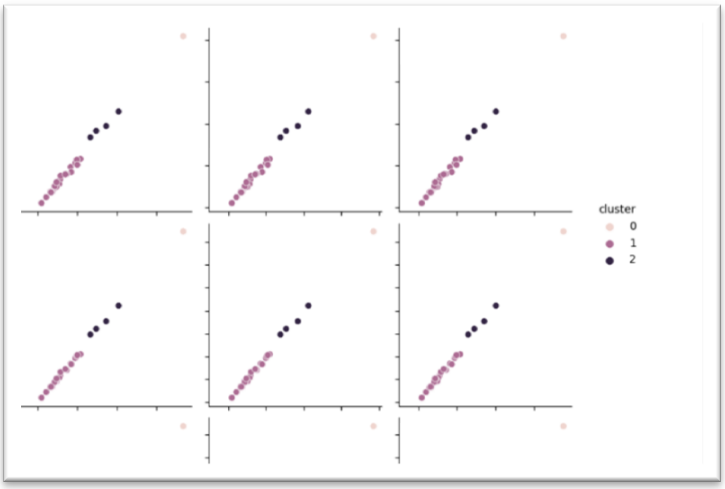
```
In [6]: #melihat centroid
        km.cluster_centers_

Out[6]: array([[209.69      , 200.32      , 207.5       , 206.87      ,
                204.22      , 186.45      , 183.79      , 183.54      ,
                193.03      , 187.74      ],
               [ 30.07714286,  28.76857143,  31.00964286,  30.595     ,
                 30.82714286,  28.21857143,  27.26321429,  27.55964286,
                 28.76535714,  27.06357143],
               [ 91.135     ,  88.6925    ,  96.9725    ,  98.105     ,
                 96.63      ,  87.105     ,  83.72      ,  82.0225    ,
                 86.3475    ,  80.6725    ]])
```

Gambar 5. *Centroid* Pada *Dataset*

J-Com (Journal of Computer) ISSN 2775-801X (online)
Vol. 4 No. 1, Maret 2024, hlm. 33 – 39
DOI: xxxxxxxxxxxxxxxx
Available online at https://jurnal.stmikroyal.ac.id/index.php/j-com

dalam kelompok yang sama lebih mirip satu sama lain daripada di kelompok lain berdasarkan informasi yang ada di dalam kelompok tersebut [13]. Dalam metode *K-Means Clustering*, langkah pertama adalah menentukan jumlah *cluster* pertama, yaitu tiga *cluster: cluster* dengan jumlah penduduk miskin tertinggi (C0), *cluster* dengan jumlah penduduk miskin terendah (C1), dan *cluster* dengan jumlah penduduk miskin sedang (C2). Selanjutnya, gunakan acak untuk menemukan *centroid* atau pusat *cluster*. *Cluster* akan terbentuk dari data yang terletak dekat atau dekat *centroid.*



Gambar 6. *Visualisasi* Hasil *Cluster* dan *Centroid*

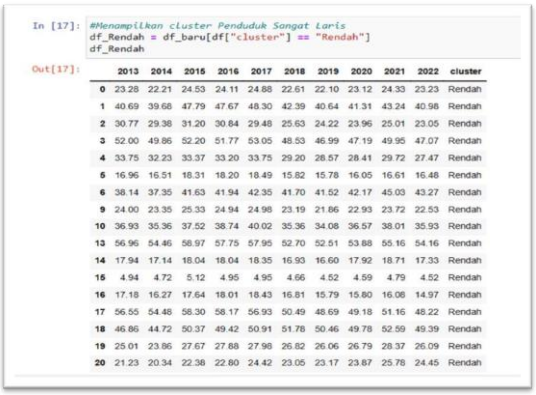
Setelah itu, data akan dimasukkan ke dalam cluster yang tersedia, seperti yang ditunjukkan pada gambar 7 di bawah ini.

```
In [15]: #Menampilkan cluster Penduduk Sangat Laris
         df_Tinggi = df_baru[df["cluster"] == "Tinggi"]
         df_Tinggi
```

Out[15]:

	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	cluster
29	209.69	200.32	207.5	206.87	204.22	186.45	183.79	183.54	193.03	187.74	Tinggi

Gambar 7. Tampilan Tingkat Penduduk Miskin Tertinggi



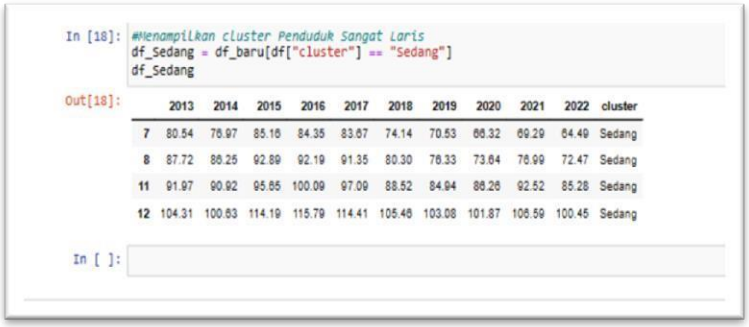
```
In [17]: #Menampilkan cluster Penduduk Sangat Laris
         df_Rendah = df_baru[df["cluster"] == "Rendah"]
         df_Rendah
```

Out[17]:

	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	cluster
0	23.28	22.21	24.53	24.11	24.88	22.61	22.10	23.12	24.33	23.23	Rendah
1	40.69	39.68	47.79	47.67	48.30	42.39	40.64	41.31	43.24	40.98	Rendah
2	30.77	29.38	31.20	30.84	29.48	25.63	24.22	23.96	25.01	23.05	Rendah
3	52.00	49.86	52.20	51.77	53.05	48.53	46.99	47.19	49.95	47.07	Rendah
4	33.75	32.23	33.37	33.20	33.75	29.20	28.57	28.41	29.72	27.47	Rendah
5	16.96	16.51	18.31	18.20	18.49	15.82	15.78	16.05	16.61	16.48	Rendah
6	38.14	37.35	41.63	41.94	42.35	41.70	41.52	42.17	45.03	43.27	Rendah
9	24.00	23.35	25.33	24.94	24.98	23.19	21.86	22.93	23.72	22.53	Rendah
10	36.93	35.36	37.52	38.74	40.02	35.36	34.08	36.57	38.01	35.93	Rendah
13	56.96	54.46	58.97	57.75	57.95	52.70	52.51	53.88	55.16	54.16	Rendah
14	17.94	17.14	18.04	18.04	18.35	16.93	16.60	17.92	18.71	17.33	Rendah
15	4.94	4.72	5.12	4.95	4.95	4.66	4.52	4.59	4.79	4.52	Rendah
16	17.18	16.27	17.64	18.01	18.43	16.81	15.79	15.80	16.08	14.97	Rendah
17	56.55	54.48	58.30	58.17	56.93	50.49	48.69	49.18	51.16	48.22	Rendah
18	46.86	44.72	50.37	49.42	50.91	51.78	50.46	49.78	52.59	49.39	Rendah
19	25.01	23.86	27.67	27.88	27.98	26.82	26.06	26.79	28.37	26.09	Rendah
20	21.23	20.34	22.38	22.80	24.42	23.05	23.17	23.87	25.78	24.45	Rendah

Gambar 8. Tampilan Tingkat Penduduk Miskin Rendah

J-Com (Journal of Computer) ISSN 2775-801X (online)
Vol. 4 No. 1, Maret 2024, hlm. 33 – 39
DOI: xxxxxxxxxxxxxxxx
Available online at https://jurnal.stmikroyal.ac.id/index.php/j-com



```
In [18]: #Menampilkan cluster Penduduk Sangat Laris
         df_Sedang = df_baru[df["cluster"] == "Sedang"]
         df_Sedang
```

Out[18]:

	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	cluster
7	80.54	76.97	85.16	84.35	83.67	74.14	70.53	66.32	69.29	64.49	Sedang
8	87.72	86.25	92.89	92.19	91.35	80.30	76.33	73.64	76.99	72.47	Sedang
11	91.97	90.92	95.65	100.09	97.09	88.52	84.94	86.26	92.52	85.28	Sedang
12	104.31	100.63	114.19	115.79	114.41	105.46	103.08	101.87	106.59	100.45	Sedang

```
In [ ]:
```

Gambar 9. Tampilan Tingkat Penduduk Miskin Sedang

SIMPULAN

Metode pengelompokan iteratif *K-Means* membagi set data ke dalam sejumlah *cluster* yang telah ditetapkan sebelumnya [14]. Penerapan algoritma *K-Means Clustering* menghasilkan 3 *cluster,* yaitu dengan *cluster* 0 dengan tingat kemiskinan penduduk yang tinggi sebanyak 1, *cluster* 1 dengan tingat kemiskinan penduduk yang sedang sebanyak 4 dan *cluster* 2 dengan tingkat kemiskinan penduduk yang rendahsebanyak 28.

Dalam hasil tersebut masih ada kecamatan yang tingkat kemiskinannya masih tinggi. Dengan demikian, ini dapat digunakan sebagai referensi bagi pemerintah untuk menilai tingkat kemiskinan yang memerlukan bantuan tambahan. Untuk penelitian lebih lanjut, dapat melakukan pengoptimalan dengan teknik preprocessing data untuk menghasilkan hasil yang berkualitas dan untuk mendapatkan perbandingan dengan algoritma lain. Selain itu, metode alternatif, seperti algortima X-means, dapat digunakan. Algortima X-means melengkapi kelemahan metode K-means, yaitu memerlukan perhitungan yang cukup lama untuk nilai clusteryang harus dikenali oleh pengguna [15].

DAFTAR PUSTAKA

[1] I. P. S. Sembiring, S. Simanjuntak, and V. A. Sitepu, "Pengaruh Inflasi dan Pengangguran terhadap Penduduk Miskin di Sumatera Utara Tahun 2006–2020," *J. Ilmu Sos. Manajemen, Akunt. dan Bisnis*, vol. 2, no. 2, pp. 1–13, 2021.

[2] E. Nainggolan, "Analisis Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kemiskinan Di Provinsi Sumatera Utara (2010-2019)," *J. Manaj. Bisnis Eka Prasetya Penelit. Ilmu Manaj.*, vol. 6, no. 2, pp. 89–99, 2020.

[3] N. Bhayu Pratama, E. Priyo Purnomo, and Agustiyara, "Sustainable Development Goals (SDGs) dan Pengentasan Kemiskinan Di Daerah Istimewa Yogyakarta," *J. Ilm. Ilmu Sos. dan Hum.*, vol. 6, no. 2, pp. 64–74, 2020.

[4] A. Jananto, Sulastri, E. N. Wahyudi, and Sunardi, "Data Induk Mahasiswa sebagai Prediktor Ketepatan Waktu Lulus MenggunakanAlgoritma CART Klasifikasi Data Mining," *J. SISFOKOM (Sistem Inf. dan Komputer)*, vol. 10, no.1, pp. 71–78, 2020.

[5] S. A. Rahmah, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu

Raja)," *J. Inf. Technol. Res.*, vol. 1, no. 1, pp. 1–5, 2020.

[6] M. R. Muttaqin and M. Defriani, "Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 121–129, 2020.

[7] A. Yoga Pratama, Y. Umaidah, and A. Voutama, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 897–910, 2021.

[8] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner," *JBASE - J. Bus. AuditInf. Syst.*, vol. 4, no. 1, pp. 13–21, 2021.

[9] Saifullah, M. Zarlis, Z. Zakaria, and R. W. Sembiring, "Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 1, no. 2, pp. 180–185, 2017.

[10] S. Oktarian, S. Defit, and Sumijan, "Klasterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma K-Means Clustering," *J. Inf. dan Teknol.*, vol. 2, no. 3, pp. 68–75, 2020.

[11] E. A. Firdaus, S. Maulani, and A. B. Dharmawan, "Pengukuran Minat Baca Mahasiswa Dengan Metode Clustering Di Perpustakaan Akademi Keperawatan Rs.Dustira Cimahi Menggunakan Data Mining," *J. Nuansa Inform.*, vol. 15, no. 1, pp. 32–40, 2021.

[12] D. P. Utomo and Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, pp. 437–444, 2020.

[13] V. Herlinda, D. Darwis, and Dartono, "Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means," *JTSI J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, pp. 94–99, 2021.

[14] F. A. Pratama, R. Narasati, and D. R. Amalia, "Pengaruh Kata Cashback Terhadap Peningkatan Penjualan Menggunakan Data Mining," *J. Ilm. Manaj. Inform. dan Komput.*, vol. 3, no. 2, pp. 1–5, 2019.

[15] G. B. Kaligis and S. Yulianto, "Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokkan Kinerja Pegawai," *J. PenerapanTeknol. Inf. dan Komun.*, vol. 1, no. 3, pp. 179–193, 2022.