

Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Clustering Perokok Usia Lebih dari 15 Tahun

Suharmanto¹, Wiranti Sri Utami², Nila Pratiwi², Muhammad Faisal^{*3}

¹Fakultas Teknologi Informatika, Teknik Informatika, STIE Ganesha, Kota Tangerang Selatan, Indonesia

²Fakultas Teknik, Teknik Informatika, Universitas Cendekia Abditama, Kabupaten Tangerang, Indonesia

³Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Raharja, Kota Tangerang, Indonesia

Email: suharmanto.java@gmail.com , wirantisutami@uca.ac.id , nilapратиwi@uca.ac.id , muhhammad.faisal@raharja.info.com

(* : coresponding author: *4muhhammad.faisal@raharja.info.com)

Abstrak - Merokok merupakan aktivitas yang dapat menimbulkan dampak buruk bagi kesehatan, baik untuk diri sendiri maupun untuk orang lain. Hal tersebut dikarenakan terdapat berbagai macam kandungan yang berbahaya bagi kesehatan, terdapat 4000 jenis bahan kimia yang terdapat pada rokok mulai dari nikotin dan zat lainnya yang dapat menyebabkan kanker bagi manusia. Pemanfaatan dataset yang didapatkan dari BPS (Badan Pusat Statistik) dapat dimanfaatkan dengan menggunakan Clustering K-Means untuk mengetahui pengelompokan perokok dengan usia lebih dari 15 tahun di setiap daerah. Pengelompokan perokok dengan rentan usia lebih dari 15 tahun dapat dilihat melalui 3 cluster, Cluster 1 merupakan tingkat perokok paling tinggi 10 Provinsi dari 34 Provinsi yang diteliti, Cluster 2 merupakan tingkat sedang dengan 15 Provinsi dari 34 Provinsi yang diteliti, dan Cluster 3 merupakan tingkat rendah dengan 9 Provinsi.

Kata Kunci: Data Mining; Cluster K-Means; Rokok

Abstract - Smoking is an activity that can have a bad impact on health, both for yourself and for others. This is because there are various kinds of ingredients that are dangerous to health, there are 4000 types of chemicals found in cigarettes, starting from nicotine and other substances that can cause cancer in humans. The dataset obtained from BPS (Central Statistics Agency) can be utilized by using K-Means Clustering to determine the grouping of smokers aged more than 15 years in each region. The grouping of smokers aged over 15 years can be seen through 3 clusters, Cluster 1 is the highest level of smoking in 10 provinces out of 34 provinces studied, Cluster 2 is the medium level with 15 provinces out of 34 provinces studied, and Cluster 3 is the lowest level. with 9 Provinces.

Keywords: Data Mining; K-Means Cluster; Cigarette

1. PENDAHULUAN

Aktivitas merokok menjadi pemandangan umum yang sering dilihat, kebiasaan tersebut dapat memberikan dampak yang buruk bagi perokok dan orang yang terkena paparan asap rokok. Merokok menjadi masalah kesehatan utama disetiap negara maju maupun negara berkembang. Aktivitas Perokok tidak hanya dilakukan oleh orang dewasa tetapi juga bisa kita lihat anak usia belasan tahun bahkan wanita melakukan aktivitas merokok. WHO mencatat 3 juta orang mengalami kematian dini setiap tahunnya dikarenakan mengkonsumsi tembakau dan menimbulkan penyakit kardiovaskular seperti serangan jantung dan stroke [1]. Dikehidupan sehari-hari dapat ditemukan orang yang sedang merokok dimana saja, dikantor, dirumah, dipasar, dan dimana saja, umumnya kebiasaan merokok dimulai saat usia remaja, Kebiasaan merokok juga dapat dipengaruhi oleh lingkungan sekitar hal tersebut menjadi perilaku yang determinan dalam memulai kebiasaan merokok [2]. Kebiasaan mereka sejak usia dini dapat mempengaruhi perilaku remaja di lingkungan sekolah, Hal tersebut dapat mempengaruhi prestasi siswa dapat menjadikan malas belajar, umumnya pengetahuan remaja akan bahaya rokok dipengaruhi Karena faktor inteligensi dan pengalaman [3]. Data Kementerian Kesehatan RI menerangkan terdapat 80% Perokok di Indonesia dimulai sejak usia 19 tahun, Kelompok tersebut terjadi antara rentang usia 15 sampai 19 tahun. Sedangkan dampak buruk dari Rokok sejak usia remaja sangat serius, dampak yang bisa dialami seorang perokok sejak usia remaja adalah Paru-paru yang berhenti mengembang, Gejala penyakit jantung dan pembuluh darah, Kerusakan gigi, Masalah otot dan tulang, serta dapat menyebabkan Kanker [4]. Merokok sudah menjadi *trend* dikalangan remaja dan terus mengalami peningkatan setiap tahunnya, *Global Youth Tobacco Survey* mencatat perokok dengan usia 13-15 tahun mengalami lonjakan dalam 5 tahun terakhir yaitu 18,3% ditahun 2014 menjadi 19,2% ditahun 2019 [5].

Kebiasaan merokok yang dimulai sejak usia remaja dapat menimbulkan penyakit bagi masyarakat, untuk itu diperlukan suatu informasi kesehatan mengenai bahaya rokok [6]. Untuk mendukung penelitian rentang usia remaja 15 tahun keatas yang sudah merokok penulis menggunakan teknik penelitian Data Mining. Data Mining merupakan proses penggalian informasi yang tersembunyi dari sebuah data dan merupakan bagian dari *Knowledge Discovery in Database (KDD)* tujuannya adalah menemukan informasi yang berguna didalam sebuah basis data [7]. Klasterisasi adalah pengaturan atau pemetaan yang disesuaikan dengan penempatan atau tata letak untuk penempatan sesuai dengan standar [8]. Untuk mendukung proses penelitian ini maka diperlukan suatu teknik pengolahan data, teknik pengolahan data yang digunakan adalah Algoritma K-Means. K-Means adalah salah satu teknik *clustering* yang awal pendekatan iterative dalam mengelompokkan data, Algoritma ini memiliki keunggulan karena sederhana dan aksesibilitas [9]. Penerapan *data mining* pada sauatu kasus pemberian imunisasi campak terhadap balita menjadi lebih efektif dan efisien karena proses tersebut dapat dipantau wilayah dengan intensitas pemberian imunisasi campak yang tinggi maupun rendah [10]. Metode K-means dapat digunakan untuk penelitian disegala bidang, Salah satu penelitian menggunakan metode K-Means adalah untuk Pengelompokan Bayi penerima ASI Eksklusif. Proses pemberian ASI terhadap bayi usia 0 sampai dengan 6 bulan dapat

diamati oleh pemerintah terkait dengan menggunakan teknik *data mining* [11]. Metode K-Means dapat juga digunakan untuk analisis data penjualan, penerapan K-Means digunakan untuk mengelompokkan data penjualan sepeda motor. Proses tersebut berupa *clustering* untuk memisahkan data menjadi 3 bagian yaitu murah, standar, dan mahal [12]. Metode K-Means juga dapat digunakan untuk membagi suatu Pasar menjadi kelompok-kelompok pembeli, Algoritma *Bisecting K-Means* dapat digunakan juga untuk pengelompokkan. Algoritma tersebut merupakan variasi dari algoritma K-Means, proses tersebut digunakan untuk mengelompokkan data kedalam suatu *cluster* sesuai dengan titik *centeroid*, Algoritma ini telah digunakan untuk mengetahui segmentasi pelanggan [13]. Metode K-Means pada suatu penelitian digunakan untuk analisis sentimen terhadap keluhan pengguna *provider* pada media social twitter yang kemudian akan di *clustering* menggunakan K-Means [14].

Dengan dilakukannya penelitian menggunakan data mining dan Algoritma K-Means diharapkan dapat memberikan informasi setiap *cluster* Perokok usia lebih dari 15 tahun di setiap Provinsi.

2. METODOLOGI PENELITIAN

2.1 Bahan Penelitian

Proses analisis data yang dilakukan diperoleh dari Badan Pusat Statistik (BPS), data yang didapatkan akan di proses menggunakan teknik Data Mining untuk menggali dan menemukan informasi terhadap data perokok dengan rentang usia 15 tahun keatas. Tahapan awal yang akan dilakukan adalah pengelompokkan data Provinsi dan menentukan atribut untuk mempermudah proses penelitian sehingga dapat berjalan dengan baik serta efisien. Data tersebut dapat dilihat pada tabel di bawah ini

Tabel 1. Data Perokok Usia Lebih dari 15 Tahun

Provinsi	2020	2021	2022
Aceh	28,06	28,3	27,58
Sumatera Utara	27,28	27,24	25,32
Sumatera Barat	30,08	30,5	30,27
Riau	28,06	28,34	26,86
Jambi	28,01	27,47	28,62
Sumatera Selatan	30,56	30,65	30,49
Bengkulu	32,31	33,17	32,16
Lampung	33,43	34,07	33,81
Kep. Bangka Belitung	28,23	28,16	26,84
Kep. Riau	26,16	26,17	23,08
DKI Jakarta	25,75	24,44	21,25
Jawa Barat	32,55	32,68	32,07
Jawa Tengah	27,7	28,24	28,72
DI Yogyakarta	22,64	24,54	23,97
Jawa Timur	27,78	28,53	28,51
...
Papua	26,97	24,91	22,22

2.2 Data Mining

Data dapat di definisikan sebagai nilai yang menggambarkan deskripsi dari suatu obyek dan kejadian, Proses pengolahan data kedalam suatu bentuk visualisasi lebih berguna dan lebih mudah dipahami [15]. Penelitian yang dilakukan ini menggunakan metode data mining, Data mining adalah merupakan teknik penambangan data yang digunakan untuk menggali informasi didalam basis data. Data mining dapat direpresentasikan kedalam berbagai teknik penelitian seperti kecerdasan buatan, *machine learning*, dan statistik [16].

2.3 Clustering K-Means

Clustering merupakan sebuah proses dalam mengelompokkan data kedalam beberapa *cluster* atau kelompok sehingga didalam *cluster* mempunyai kemiripan yang maksimum dan minimum [17]. Algoritma K-Means *Cluster* dapat diterapkan dalam permasalahan Analisis data Perokok dengan rentang usia 15 tahun keatas, dapat mengidentifikasi

peluang baru dari perokok di suatu provinsi, Algoritma ini dapat meringkas data dengan jumlah besar sehingga dapat diamati karakteristik dari kelompok masing-masing [18].

Dalam proses melakukan analisa penelitian ini menggunakan Bahasa pemrograman R. Bahasa R adalah pemrograman yang digunakan untuk komputasi statistik dan grafik [19]. Untuk mendukung jumlah *cluster* yang akan dibentuk maka penulis menggunakan metode *Elbow criterion* dengan rumus sebagai berikut: [20].

$$SSE = \sum_k^K = 1 \sum x_i = S_k || N_i - C_k ||$$

Sedangkan untuk menentukan titik pusat *cluster* awal diambil sampel secara *random*. Rumus yang digunakan sebagai berikut:

$$v = \frac{\sum_i^n = X_1}{n} : i = 1,2,3, \dots n$$

Kemudian untuk menghitung jarak dari setiap obyek ke pusat *centroid* dari *cluster* menggunakan *Euclidean distance* dengan rumus sebagai berikut:

$$d(x, y) = ||x - y|| = \sqrt{\sum_i^n = 1 (x_i - y_i)^2} : i = 1,2,3, \dots n$$

Untuk menentukan nilai K dari data *centroid* diambil *sample* secara acak, perhitungan jarak menggunakan *Euclidean distance* dari seluruh *dataset cluster* yang ada. *Centroid* awal dapat dilihat pada tabel dibawah ini.

Tabel 2. Data *Centroid* awal

Provinsi	<i>Centroid</i>			<i>Cluster</i>
Sumatera Utara	27,28	27,24	25,32	C1
Sumatera Barat	30,08	30,5	30,27	C2
Sumatera Selatan	30,56	30,65	30,49	C3

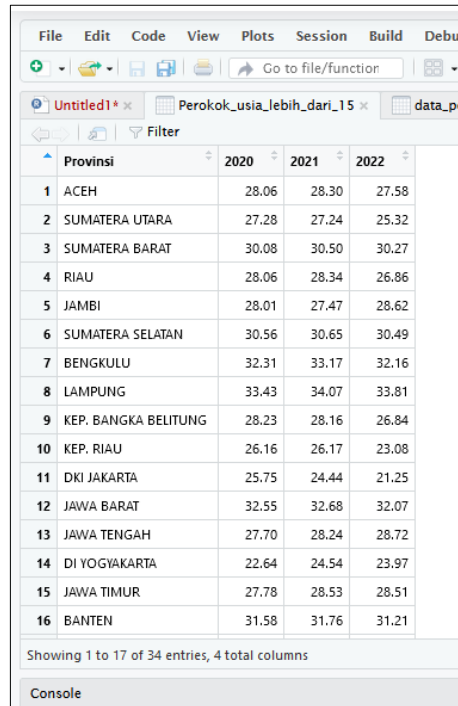
Selanjutnya melakukan proses perhitungan jarak masing-masing *cluster* menggunakan persamaan *Euclidean distance*. Proses tersebut dapat dilihat pada tabel dibawah ini.

Tabel 3. Hasil *Clustering*

Provinsi	C1	C2	C3
Aceh	2,615263	4,019515	4,498955
Sumatera Utara	0	6,555158	7,008238
Sumatera Barat	6,555158	0	0,548908
Riau	2,046949	4,513768	4,976244
Jambi	3,387595	4,023469	4,484618
Sumatera Selatan	7,008238	0,548908	0
Bengkulu	10,35623	3,959028	3,493107
Lampung	12,51205	6,04144	5,563785
Kep. Bangka Belitung	2,01477	4,545657	4,995148
Kep. Riau	2,723399	9,263444	9,712801
Dki Jakarta	5,171634	11,69756	12,12756
Jawa Barat	10,14539	3,754104	3,252291
Jawa Tengah	3,568809	3,629669	4,137705
Di Yogyakarta	5,535531	11,42651	11,94022
Jawa Timur	3,477096	3,502642	4,01786
...
Papua	3,890373	10,28215	10,68778

3. HASIL DAN PEMBAHASAN

Data yang didapatkan dari Badan Pusat Statistik di *import* kedalam *posit.cloud*, data yang di *import* tersebut dapat dilihat dibawah ini.



Provinsi	2020	2021	2022
1 ACEH	28.06	28.30	27.58
2 SUMATERA UTARA	27.28	27.24	25.32
3 SUMATERA BARAT	30.08	30.50	30.27
4 RIAU	28.06	28.34	26.86
5 JAMBI	28.01	27.47	28.62
6 SUMATERA SELATAN	30.56	30.65	30.49
7 BENGKULU	32.31	33.17	32.16
8 LAMPUNG	33.43	34.07	33.81
9 KEP. BANGKA BELITUNG	28.23	28.16	26.84
10 KEP. RIAU	26.16	26.17	23.08
11 DKI JAKARTA	25.75	24.44	21.25
12 JAWA BARAT	32.55	32.68	32.07
13 JAWA TENGAH	27.70	28.24	28.72
14 DI YOGYAKARTA	22.64	24.54	23.97
15 JAWA TIMUR	27.78	28.53	28.51
16 BANTEN	31.58	31.76	31.21

Gambar 1. Tampilan data yang sudah di *import* pada *posit.cloud*

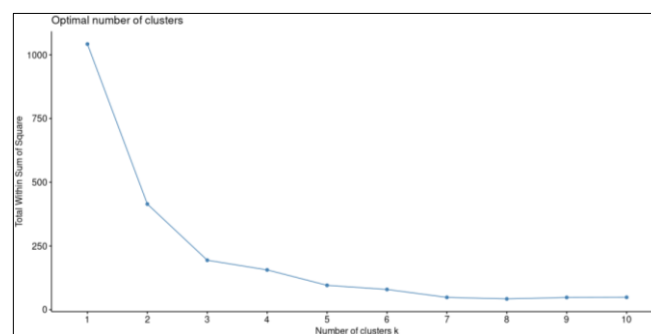
Setelah data di *import* proses selanjutnya adalah melakukan standarisasi data. Sebelum dilakukan standarisasi data, data diproses terlebih dahulu agar memiliki rentang nilai atau *scale* yang sama.

```
> head(datastd)
      X2020      X2021      X2022
1  0.11447521  0.09638472  0.21860503
2 -0.15731373 -0.24960186 -0.38649621
3  0.81833887  0.81447006  0.93883615
4  0.11447521  0.10944081  0.02582941
5  0.09705284 -0.17452930  0.49705870
6  0.98559360  0.86343043  0.99773982
```

Gambar 2. Proses standarisasi data

3.1 Algoritma K-Means

Proses awal untuk melakukan *clustering* K-Means pada Bahasa R adalah menentukan jarak matriks dengan metode *Euclidean distance*. Kemudian akan dicoba K Optimal menggunakan *clustering* dengan metode Elbow.



Gambar 3. Jumlah Cluster Optimal

Berdasarkan hasil diatas pergerakan grafik yang landai terdapat pada $K = 3$. Banyaknya *cluster* optimal terbentuk adalah 3, maka diperoleh pengelompokkan dengan K Optimal 4.

Clustering vector:

[1] 2 2 1 2 2 1 1 1 2 3 3 1 2 3 2 1 3 1 2 2 2 3 3 2 2 1 3 3 1 2 2 1 2 3

Gambar 4. Clustering K-Means

Langkah selanjutnya menggunakan metode *silhouette*, gambar dibawah ini menunjukkan diagram baris hasil dari metode *silhouette*. *Cluster* yang dihasilkan menggunakan metode *silhouette* dapat dilihat pada gambar dibawah ini.



Gambar 5. Grafik Silhouette

Nilai rata-rata yang dihasilkan menggunakan metode *silhouette*

cluster	size	ave.sil.width
1	10	0.52
2	15	0.58
3	9	0.36

Gambar 6. Rata-rata nilai dari Metode Silhouette

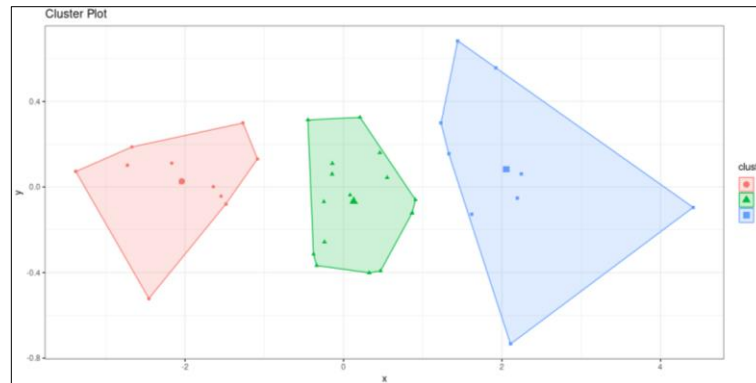
3.1 Hasil Cluster K-Means

Hasil dari *cluster* K-Means Perokok dengan usia 15 tahun keatas dapat dilihat pada gambar dibawah ini.

```
> data_cluster <- cbind(data_perokok, cluster = result1$cluster)
> data_cluster
  Provinsi 2020 2021 2022 cluster
1 ACEH 28.06 28.30 27.58 2
2 SUMATERA UTARA 27.28 27.24 25.32 2
3 SUMATERA BARAT 30.08 30.50 30.27 1
4 RIAU 28.06 28.34 26.06 2
5 JAMBI 28.01 27.47 28.62 2
6 SUMATERA SELATAN 30.56 30.65 30.49 1
7 BENGKULU 32.31 33.17 32.16 1
8 LAMPUNG 33.43 34.07 33.81 1
9 KEP. BANGKA BELITUNG 28.23 28.16 26.84 2
10 KEP. RIAU 26.16 26.17 23.08 3
11 DKI JAKARTA 25.75 24.44 21.25 3
12 JAWA BARAT 32.55 32.68 32.07 1
13 JAWA TENGAH 27.70 28.24 28.72 2
14 DI YOGYAKARTA 22.64 24.54 23.97 3
15 JAWA TIMUR 27.78 28.53 28.51 2
16 BANTEN 31.58 31.76 31.21 1
17 BALI 20.50 19.58 17.91 3
18 NUSA TENGGARA BARAT 30.58 32.71 33.20 1
19 NUSA TENGGARA TIMUR 26.14 27.22 26.76 2
20 KALIMANTAN BARAT 27.49 27.93 26.64 2
21 KALIMANTAN TENGAH 28.89 29.33 26.54 2
22 KALIMANTAN SELATAN 23.83 24.51 21.89 3
23 KALIMANTAN TIMUR 24.42 23.37 22.21 3
24 KALIMANTAN UTARA 25.66 27.46 24.23 2
25 SULAWESI UTARA 27.95 27.87 25.29 2
26 SULAWESI TENGAH 30.64 29.77 29.04 1
27 SULAWESI SELATAN 24.89 24.91 23.76 3
28 SULAWESI TENGGARA 25.77 25.85 23.35 3
29 GORONTALO 30.30 30.50 30.38 1
30 SULAWESI BARAT 26.85 27.17 25.36 2
31 MALUKU 26.18 27.90 26.00 2
32 MALUKU UTARA 29.83 29.84 28.82 1
33 PAPUA BARAT 25.80 27.07 24.80 2
34 PAPUA 26.97 24.91 22.22 3
```

Gambar 7. Hasil akhir Cluster Perokok dengan Usia 15 tahun keatas

Gambar dibawah ini merupakan proses *cluster plot* menggunakan metode K-Means, gambar tersebut bisa dilihat dibawah ini.



Gambar 8. Diagram *cluster plot*

4. KESIMPULAN

Penerapan metode K-Means dalam mengelompokkan usia perokok pada setiap provinsi Republik Indonesia, dalam menentukan titik pusat atau centeroid memberikan dampak yang cukup besar untuk menentukan cluster yang sudah diproses dalam menganalisa Perokok dengan usai 15 tahun keatas. Hasil dari penelitian ini diperoleh tiga cluster, C1 dengan kategori tinggi sebanyak 10 cluster, C2 dengan kategori rendah sebanyak 15 cluster, dan C3 dengan kategori rendah sebanyak 9 cluster.

REFERENCES

- [1] P. K. RI, "WHO: Rokok Tetap Jadi Sebab Utama Kematian dan Penyakit," Kementerian Kesehatan Republik Indonesia, 18 June 2018. [Online]. Available: <https://p2ptm.kemkes.go.id/kegiatan-p2ptm/pusat-/who-rokok-tetap-jadi-sebab-utama-kematian-dan-penyakit>. [Accessed 28 November 2023].
- [2] S. H. Wati, Bahtiar and D. Anggraini, "DAMPAK MEROKOK TERHADAP KEHIDUPAN SOSIAL REMAJA (Studi di Desa Mabodo Kecamatan Kontunaga Kabupaten Muna)," *Neo Societal*, vol. 3, no. 2, pp. 503-509, 2018.
- [3] A. J. Rifiana, M. Evelianti and V. G. Pratiwi, "ANALISIS KEBIASAAN MEROKOK PADA REMAJA DI KAMPUNG," *Penelitian Perawatan Profesional*, vol. 5, no. 1, pp. 315-322, 2023.
- [4] d. D. Bachtiar, Sp.P, "Masih Remaja Sudah Mulai Merokok? Ini Dampak dan Bahayanya," EMC Healthcare, 1 Juni 2020. [Online]. Available: <https://www.emc.id/id/care-plus/masih-remaja-sudah-mulai-merokok-ini-dampak-dan-bahayanya>. [Accessed 5 Desember 2023].
- [5] N. A. Muslim, S. Adi, S. P. Ratih and N. H. Ulfah, "Determinan Perilaku Merokok Remaja SMA/Sederajat di Kecamatan Lowokwaru Kota Malang," *Indonesia Journal of Health Promotion and Behavior*, vol. 5, no. 1, pp. 20-28, 2023.
- [6] Y. N. Iriyanti and A. M. Mandagi, "Pengetahuan Mengenai Bahaya Merokok dengan Keinginan Berhenti Merokok Masyarakat Desa Pakel," *PREVENTIF*, vol. 13, no. 1, pp. 15-24, 2022.
- [7] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Saffi and R. Winanjaya, "IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA APRIORI DALAM MENENTUKAN PERSEDIAAN BARANG (STUDI KASUS : TOKO SINAR HARAHAHAP)," *JUST IT*, vol. 12, no. 2, pp. 51-61, 2022.
- [8] N. Suwaryo, A. Rahman, D. M. U. Atmaja and A. Basri, "Klasterisasi Stok Produk Retail Untuk Menentukan Pergerakan Kebutuhan Konsumen Dengan Algoritma K-Means," *BIT*, vol. 4, no. 2, pp. 306-312, 2023.
- [9] K. R. Watrionthos and S. Suryadi, "Distribusi Spasial Unmet Need Pelayanan Kesehatan dengan Algoritma K-Means Untuk Pemetaan Provinsi di Indonesia," *BIT*, vol. 4, no. 1, pp. 361-368, 2023.
- [10] M. Faisal and W. S. Utami, "Penerapan Data Mining Untuk Klastering Balita Penerima Imunisasi Campak Menggunakan Metode K-Means," *ICIT*, vol. 8, no. 1, pp. 49-55, 2022.
- [11] 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.

- [12] R. K. Dinata, S. N. Hasdyana and N. Azizah, "Analisis K-Means Clustering Pada Data Sepeda Motor," *Informatics*, vol. 5, no. 1, pp. 10-17, 2020.
- [13] N. Puspitasari, J. A. Widiyans and N. B. Setiawan, "Segmentasi pelanggan menggunakan algoritme bisecting k-means berdasarkan model recency, frequency, dan monetary (RFM)," *Teknologi dan Sistem Komputer*, vol. 8, no. 2, pp. 78-83, 2020.
- [14] T. I. Saputra and R. Arianty, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA ANALISIS SENTIMEN KELUHAN PENGGUNA INDOSAT," *Ilmiah Informatika Komputer*, vol. 24, no. 3, pp. 191-198, 2019.
- [15] Amna, W. S, T. A. E, I. G. I. Sudipa, A. J. Wahidin, W. A. Syukrilla, A. K. Wardhani, N. Heryana, T. Indriyani and L. W. Santoso, *Data Mining, Sumatera Barat: PT GLOBAL EKSEKUTIF TEKNOLOGI*, 2023.
- [16] D. N. Yoliadi, "Data mining Dalam Analisis Tingkat Penjualan Barang Elektronik Menggunakan Algoritma K-Means," *Insearch*, vol. 3, no. 1, pp. 43-53, 2023.
- [17] E. Irwansyah, "Clustering," Binus University, 9 May 2017. [Online]. Available: <https://socs.binus.ac.id/2017/03/09/clustering/>. [Accessed 5 Desember 2023].
- [18] H. E. Widodo and W. Hadikristanto, "Pengelompokan Untuk Penjualan Obat Dengan Menggunakan Algoritma K-Means," *BIT*, vol. 4, no. 2, pp. 408-413, 2023.
- [19] K. F. Hidayati, "Bahasa R: Apa Itu, Kelebihan, dan Kekurangannya," *glints*, 19 Mei 2022. [Online]. Available: <https://glints.com/id/lowongan/bahasa-r/>. [Accessed 6 Desember 2023].
- [20] G. M. H. Aditya and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-obatan Pada RSUD Pekanbaru," *Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17-24, 2019.