Gudang Jurnal Multidisiplin Ilmu
Volume 2 ; Nomor 10 ; November 2024 ; Page 133-139
Doi : https://doi.org/10.59435/gjmi.v2i11.1059
Website : https://gudangjurnal.com/index.php/gjmi
E-ISSN : 2988-5760

Implementasi Algoritma K-Means Clustering Untuk Perokok Usia Di Atas 15 Tahun

Angeli Dwiyanti Nur'azizah[1*], Zaehol Fatah[2]

[1,2] Sistem Informasi, Universitas Ibrahimy Sukorejo Situbondo Jawa Timur
[1*]angelidwiyanti@gmail.com, [2]zaeholfatah@gmail.com

Abstrak

Merokok merupakan aktivitas yang dapat menimbulkan dampak buruk bagi kesehatan, baik untuk diri sendiri maupun untuk orang lain. Hal itu dikarenakan terdapat banyak kandungan yang berbahaya bagi kesehatan. Data mining merupakan bagian dari data analytics dan disiplin ilmu *data science* yang memiliki manfaat luas dan tepat guna. Penelitian ini bertujuan untuk mengetahui pengelompokkan perokok dengan usia lebih dari 15 tahun di setiap Provinsi. Data yang digunakan dalam penelitian ini diambil dari Badan Pusat Statistik (BPS). Metode yang digunakan adalah Clustering dengan algoritma K-Means menggunakan *tools* RapidMiner dan validasinya menggunakan operator *Davies Bouldin Indeks* untuk mencari nilai yang mendekati 0. Pengelompokkan perokok dengan rentan usia lebih dari 15 tahun yang dihasilkan dapat dilihat melalui 3 cluster, Cluster 1 merupakan tingkat perokok tinggi sejumlah 9 provinsi, Cluster 2 merupakan tingkat sedang dengan 17 provinsi dan Cluster 3 merupakan tingkat rendah dengan 8 provinsi.

Kata Kunci: Data Mining, K-Means, Clustering, RapidMiner, Perokok

PENDAHULUAN

Rokok adalah silinder dari kertas berukuran panjang 70 hingga 120 mm dengan diameter 10 mm. didalamnya berisi daun-daun tembakau yang telah dicacah. Untuk menikmatinya salah satu ujung rokok dibakar dan dibiarkan membara agar asapnya dapat dihirup lewat mulut pada ujung lain[1]. Merokok sudah menjadi suatu kebiasaan buruk yang sudah menjadi biasa di Indonesia. Aktivitas merokok tidak hanya dilakukan oleh orang dewasa, tetapi juga dapat dilihat anak usia belasan tahun bahkan wanita melakukan aktivitas merokok.

Indonesia termasuk dalam 10 negara dengan jumlah terbanyak perokok di dunia. Data dari Badan Pusat Statistik (BPS) menyebutkan bahwa di Indonesia pada tahun 2022 persentase merokok pada penduduk umur diatas 15 tahun sebesar 28,26% dan mengalami peningkatan di tahun 2023 menjadi 28,62%.[2] Di Indonesia terlapor sebanyak 63% perokok pria dan 5% perokok Wanita. Faktor-faktor yang mempengaruhi kebiasaan merokok di Indonesia sangat kompleks, mencakup aspek budaya, ekonomi, dan iklan produk tembakau yang mudah diakses. Sedangkan dampak buruk dari rokok sejak usia remaja sangat serius, dampak yang bisa dialami adalah Paru-Paru yang berhenti mengembang, Gejala penyakit jantung dan pembuluh darah, kerusakan gigi, masalah otot dan tulang, serta dapat menyebabkan kanker.[3]

Untuk mendukung penelitian perokok unia 15 tahun ke atas penulis menggunakan teknik penelitian Data Mining. Data Mining adalah suatu proses menganalisis pola data yang tersembunyi menurut berbagai perspektif untuk kategorisasi menjadi informasi yang berguna.[4] klasterisasi adalah suatu teknik atau metode untuk mengelompokkan data.[5] Untuk mendukung proses penelitian ini maka diperlukan suatu teknik pengolahan data, teknik pengolahan data yang digunakan adalah Algoritma K-Means. K-means adalah suatu metode penganalisaan data yang melakukan proses pemodelan tanpa supervisi.[6] Dengan adanya penelitian ini diharapkan dapat mengetahui klaster persentasi merokok pada tingkat tinggi dan rendah sehingga dapat menjadi masukan kepada pemerintah agar dapat mengurangi angka persentase merokok di setiap wilayah.

METODE

Pengumpulan Data

Pengumpulan data menggunakan *Study Literature* dengan jenis data sekunder. Proses analisis data yang dilakukan diperoleh dari Badan Pusat Statistik (BPS)[2], data yang didapat 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	2021	2022	2023
Aceh	28,3	27,58	28,66

Sumtera Utara	27,24	25,32	26,28
Sumatera Barat	30,5	30,27	30,42
Riau	28,34	26,86	27,76
Jambi	27,47	28,62	28,67
Sumatera Selatan	30,65	30,49	30,91
Bengkulu	33,17	32,16	31,86
Lampung	34,07	33,81	34,08
Kep. Bangka Belitung	28,16	26,84	27,33
Kep. Riau	26,17	23,08	25,49
.....	.....	.....	.....
Papua	24,91	22,22	22,3

Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terikat dari berbagai *database* besar.[4] Data mining ada sejak 1990-an mulai dianggap sebagai cara yang benar dan tepat untuk mengambil informasi terkait dengan hubungan antar data.[7] Data mining membahas penggalian atau pengumpulan informasi yang berguna dari kumpulan data. Informasi yang biasanya dikumpulkan adalah pola-pola tersembunyi pada data, hubungan antar elemen-elemen data, ataupun pembuatan model untuk keperluan peramalan data.[8]

Data mining bertujuan untuk mengubah pandangan data tradisional sehingga bisa menangani jumlah data yang sangat besar, dimensi data yang tinggi, dan data yang heterogen dan berbeda sifat.[9] Data mining memiliki banyak manfaat untuk mengelola data mentah menjadi kumpulan informasi yang dapat dijadikan sebagai penunjang keputusan secara efektif. Secara umum, semua operasi dalam data mining dapat dikelompokkan menjadi dua kategori yaitu metode deskriptif dan metode prediktif. Metode deskriptif bertujuan untuk menemukan pola, relasi, atau anomali dalam data yang mudah dipahami oleh manusia. Metode prediktif bertujuan untuk memperkirakan nilai suatu variabel berdasarkan nilai variabel-variabel lainnya.[8]

Clustering K-Means

Pada dasarnya algoritma K-Means clustering merupakan bidang penelitian dalam analisis dan data mining.[10] *Clustering* adalah suatu teknik atau metode untuk mengelompokkan data. Menurut Tan, 2006 *clustering* adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum.[5] Oleh karena itu clustering sangat berguna dalam menemukan kelompok yang tidak dikenal dalam data.

Metode K-means merupakan metode *clustering* yang paling sederhana dan umum.k-means merupakan salah satu algoritma klastering dengan metode partisi (*partitioning method*) yang berbasis titik pusat (*centroid*) selain algoritma k-Medoids yang berbasis obyek.[5] Langkah-langkah algoritma K-means antara lain[11]:

a. Menentukan k sebagai jumlah klaster yang ingin dibentuk.
b. Mengalokasikan data ke dalam klaster secara acak.
c. Menentukan pusat klaster (*centroid*) dari data yang ada pada masing-masing klaster. dengan pesamaan :

dimana

= pusat klaster ke-k pada variabel ke-j (j=1,2,...,p)

n = banyak data pada klaster ke-k

d. Menentukan jarak setiap objek dengan setiap centroid dengan perhitungan jarak setiap objek dengan setiap centroid menggunakan jarak Euclidean.

d( )=

e. Menghitung fungsi objektif dengan formula:

$$J = \sum_{i=1}^{n} \sum_{j=1}^{k} a_{ij} d(x_1, c_{kj})^2$$

f. Mengalokasikan masing-masing data ke centroid/rata-rata terdekat yang dirumuskan sebagai berikut:

$$a_{ij} = \begin{cases} 1, & s = min\{d(x_i, c_{kj})\} \\ 0, & lainnya \end{cases}$$

a_{ij} adalah nilai keanggotaan titik x_i ke pusat klaster c_{kj}, s adalah jarak terpendek dari data x_i ke pusat klaster c_{kj} setelah dibandingkan.

g. Mengulangi kembali langkah 3-6 sampai tidak ada lagi perpindahan objek atau tidak ada perubahan pada fungsi objektifnya.

RapidMiner

RapidMiner adalah *platform* perangkat lunak yang kuat untuk ilmu data dan pembelajaran mesin.[12] Analisis pengolahan data dapat dilakukan dengan menggunakan RapidMiner. Terdapat beberapa teknik yang digunakan dalam RapidMiner, termasuk deskriptif dan prediksi. Bahasa pemograman Java merupakan bahasa pemograman yang digunakan oleh Rapidminer.[13] Perangkat lunak RapidMiner juga dapat digunakan untuk mengolah data mining karena terdapat kecanggihan pada teknologi algoritma yaitu komputasi dan terdapat analisis untuk data yang berbasis komputer.

HASIL DAN PEMBAHASAN

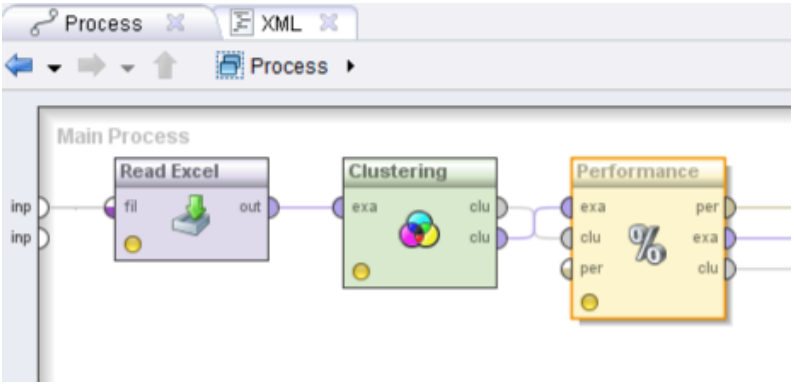
Data yang digunakan dalam penelitian adalah data persentase penduduk 15 tahun ke atas yang merokok tembakau selama sebulan terakhir menurut provinsi dari tahun 2021 – 2023 yang didapatkan dari Badan Pusat Statistik (BPS). Kumpulan data yang diperoleh penulis digunakan sebagai data masukan dalam membuat model aturan menggunakan algoritma K-means Clustering dan menggunakan *software* rapidminer. Analisis data yang digunakan penelitian ini menggunakan data kuantitatif dengan teknik analisis data yang menggunakan jenis *statistic* deskriptif. Data yang diperoleh kemudian diolah dengan *tools* rapidminer menggunakan *performance* yang berfungsi sebagai validasi dan reabilitas data untuk mencari keakuratan data.

Tabel 2. Data Perokok dari 34 Provinsi

Provinsi	2021	2022	2023
Aceh	28,3	27,58	28,66
Sumtera Utara	27,24	25,32	26,28
Sumatera Barat	30,5	30,27	30,42
Riau	28,34	26,86	27,76
Jambi	27,47	28,62	28,67
Sumatera Selatan	30,65	30,49	30,91
Bengkulu	33,17	32,16	31,86
Lampung	34,07	33,81	34,08
Kep. Bangka Belitung	28,16	26,84	27,33
Kep. Riau	26,17	23,08	25,49
DKI Jakarta	24,44	21,25	22,6
Jawa Barat	32,68	32,07	32,78
Jawa Tengah	28,24	28,72	28,55
DI Yogyakarta	24,54	23,97	24,82
Jawa Timur	28,53	28,51	28,83
Banten	31,76	31,21	29,34
Bali	19,58	17,91	18,9
NTB	32,71	33,2	32,79
NTT	27,22	26,76	26,64
Kalimantan Barat	27,93	26,64	26,96
Kalimantan Tengah	29,33	26,54	27,24
Kalimantan Selatan	24,51	21,89	22,24
Kalimantan Timur	23,37	22,21	22,97
Kalimantan Utara	27,46	24,23	25,36
Sulawesi Utara	27,87	25,29	26,96
Sulawesi Tengah	29,77	29,04	28,28
Sulawesi Selatan	24,91	23,76	24,24
Sulawesi Tenggara	25,85	23,35	24,66
Gorontalo	30,5	30,38	30,69
Sulawesi Barat	27,17	25,36	25,3

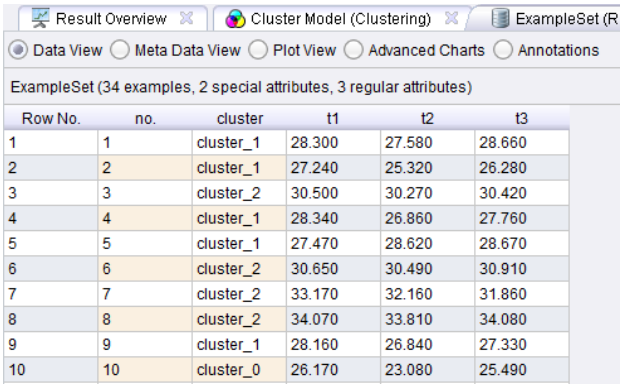
Maluku	27,9	26,8	28,04
Maluku Utara	29,84	28,82	28,82
Papua Barat	27,07	24,8	25,3
Papua	24,91	22,22	22,3

Data yang telah didapatkan lalu di preprocessing dengan tidak memasukkan *recovery case dan case fatality rate* supaya pengolahan menggunakan *tools* RapidMiner berjalan dengan sesuai keinginan. Data yang telah tersusun dimasukkan ke dalam aplikasi RapidMiner untuk melakukan pengujian.



Gambar 1. *Clustering* K-Means menggunakan *tools* RapidMiner

Berdasarkan evaluasi hasil dari pengujian dan eksperimen model, *cluster* yang digunakan pada penelitian ilmiah ini adalah 3 parameter *cluster*.


Result Overview | Cluster Model (Clustering) | ExampleSet (R

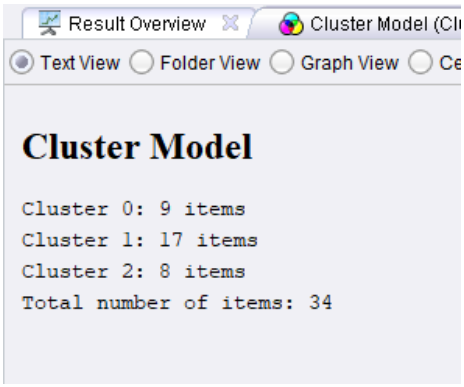
Data View | Meta Data View | Plot View | Advanced Charts | Annotations

ExampleSet (34 examples, 2 special attributes, 3 regular attributes)

Row No.	no.	cluster	t1	t2	t3
1	1	cluster_1	28.300	27.580	28.660
2	2	cluster_1	27.240	25.320	26.280
3	3	cluster_2	30.500	30.270	30.420
4	4	cluster_1	28.340	26.860	27.760
5	5	cluster_1	27.470	28.620	28.670
6	6	cluster_2	30.650	30.490	30.910
7	7	cluster_2	33.170	32.160	31.860
8	8	cluster_2	34.070	33.810	34.080
9	9	cluster_1	28.160	26.840	27.330
10	10	cluster_0	26.170	23.080	25.490

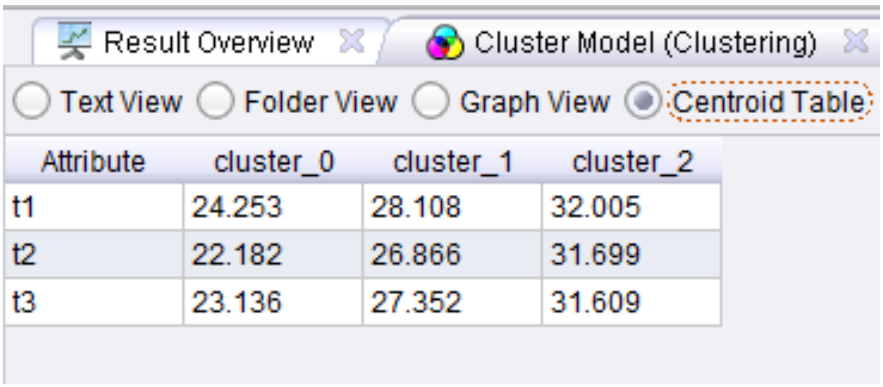
Gambar 2. *ExampleSet Result*

Pada gambar 2. Menampilkan hasil dari *clustering* data, label cluster terbagi menjadi 3 kelompok yaitu cluster_0, cluster_1, dan cluster_3. Pembagian ini berdasarkan hasil dari *centroid* masing-masing data dengan jarak terdekat.



Gambar 3. Cluster model

Pada gambar 3. Bisa dilihat hasil pembagian data terhadap tiap *cluster*, cluster_0 terdapat 9 provinsi, cluster_1 terdapat 17 provinsi dan cluster_2 terdapat 8 provinsi. Total dari dataset yang diuji adalah 34 provinsi.



Attribute	cluster_0	cluster_1	cluster_2
t1	24.253	28.108	32.005
t2	22.182	26.866	31.699
t3	23.136	27.352	31.609

Gambar 4. *Centroid Table*

Centroid Table menampilkan nilai centroid dari masing-masing atribut pada tiap cluster. Nilai tersebut menjadi pusat perhitungan dataset dengan cara menghitung dan mengukur kedekatan nilai dengan masing-masing titik pusat cluster. Jadi, berdasarkan hasil yang telah diuji pada proses *clustering* K-Means, cluster_0 terdapat 9 provinsi yakni cluster dengan persentase rendah, cluster_1 terdapat 17 item yakni cluster dengan persentase sedang dan cluster_2 terdapat 8 provinsi yakni cluster dengan persentase tinggi. Berikut *ExampleSet* dari hasil pengujian *Clustering* K-Means.

Tabel 3. *ExampleSet* Hasil pengujian *Clustering*

Provinsi	2021	2022	2023	Cluster
Aceh	28,3	27,58	28,66	Cluster_1
Sumtera Utara	27,24	25,32	26,28	Cluster_1
Sumatera Barat	30,5	30,27	30,42	Cluster_2
Riau	28,34	26,86	27,76	Cluster_1
Jambi	27,47	28,62	28,67	Cluster_1
Sumatera Selatan	30,65	30,49	30,91	Cluster_2
Bengkulu	33,17	32,16	31,86	Cluster_2
Lampung	34,07	33,81	34,08	Cluster_2
Kep. Bangka Belitung	28,16	26,84	27,33	Cluster_1
Kep. Riau	26,17	23,08	25,49	Cluster_0
DKI Jakarta	24,44	21,25	22,6	Cluster_0
Jawa Barat	32,68	32,07	32,78	Cluster_2
Jawa Tengah	28,24	28,72	28,55	Cluster_1
DI Yogyakarta	24,54	23,97	24,82	Cluster_0
Jawa Timur	28,53	28,51	28,83	Cluster_1
Banten	31,76	31,21	29,34	Cluster_2
Bali	19,58	17,91	18,9	Cluster_0
NTB	32,71	33,2	32,79	Cluster_2
NTT	27,22	26,76	26,64	Cluster_1
Kalimantan Barat	27,93	26,64	26,96	Cluster_1
Kalimantan Tengah	29,33	26,54	27,24	Cluster_1
Kalimantan Selatan	24,51	21,89	22,24	Cluster_0
Kalimantan Timur	23,37	22,21	22,97	Cluster_0
Kalimantan Utara	27,46	24,23	25,36	Cluster_1
Sulawesi Utara	27,87	25,29	26,96	Cluster_1
Sulawesi Tengah	29,77	29,04	28,28	Cluster_1
Sulawesi Selatan	24,91	23,76	24,24	Cluster_0
Sulawesi Tenggara	25,85	23,35	24,66	Cluster_0
Gorontalo	30,5	30,38	30,69	Cluster_2
Sulawesi Barat	27,17	25,36	25,3	Cluster_1
Maluku	27,9	26,8	28,04	Cluster_1

Maluku Utara	29,84	28,82	28,82	Cluster_1
Papua Barat	27,07	24,8	25,3	Cluster_1
Papua	24,91	22,22	22,3	Cluster_0

Proses *clustering* K-Means ini menghasilkan 3 cluster Dimana cluster_0 (c1) mempunyai nilai *centroid* pada tahun 2021 yakni 24,253, tahun 2022 22,182 dan tahun 2023 23,136 dengan status klaster rendah yang menunjukkan bahwa provinsi dengan klaster tersebut merupakan provinsi dengan tingkat perokok rendah. cluster_1 (c2) mempunyai nilai *centroid* pada tahun 2021 yakni 28,108, tahun 2022 26,866 dan tahun 2023 27,352 dengan status klaster sedang yang menunjukkan bahwa provinsi dengan klaster tersebut merupakan provinsi dengan tingkat perokok sedang. cluster_2 (c3) mempunyai nilai *centroid* pada tahun 2021 yakni 32,005, tahun 2022 31,699 dan tahun 2023 31,609 dengan status klaster tinggi yang menunjukkan bahwa provinsi dengan klaster tersebut merupakan provinsi dengan tingkat perokok tinggi.

Hasil uji *performance* dilakukan menggunakan operator *cluster distance performance*. Operator ini digunakan untuk mengevaluasi hasil kinerja metode *clustering* berbasis *centroid*. Operator ini juga memberikan daftar nilai kriteria kinerja berdasarkan *centroid*. *Cluster distance performance* yang dimaksud adalah *performance Davies Bouldin Index* (DBI). Dimana pada penelitian ini nilai *Davies Bouldin Index* sangat optimal yakni 0,162.

KESIMPULAN

Proses clustering K-Means ini menghasilkan 3 cluster Yakni Cluster_0 (klaster 1), Cluster_1 (klaster 2) dan Cluster_2 (klaster 3). Pada klaster 1 yakni klaster rendah didapati 9 Provinsi , diantaranya Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Bali, Kalimantan Selatan, Kalimantan Timur, Sulawesi Selatan, Sulawesi Tenggara dan Papua. Pada klaster 2 didapati sebanyak 17 Provinsi diantaranya Aceh, Sumatera Utara, Riau, Jambi, Kepulauan Bangka Belitung, Jawa Tengah, Jawa Timur, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Utara, Sulawesi Utara, Sulawesi tengah, Sulawesi Barat, Maluku, Maluku Utara dan Papua Barat. Klaster 3 merupakan klaster dengan tingkat perokok tinggi yang terdapat pada 8 Provinsi diantaranya Sumatera Barat, Sumatera Selatan, Bengkulu, Lampung, Jawa Barat, Banten, NTB, dan Gorontalo. Dengan hasil meyakinkan pada cluster distance performance yaitu performance Davies Bouldin Indeks (DBI) dengan nilai sangat optimal yakni 0,162. Diharapkan pemerintah bisa lebih fokus kepada 8 provinsi ini untuk mengurangi tingkat perokok tembakau dengan mengadakan penyuluhan bahaya rokok kepada sekolah-sekolah, menghilangkan iklan-iklan rokok dan sebagainya.

Hasil cluster yang terbentuk dapat dikembangkan dengan metode lain, supaya mendapatkan informasi lain yang belum diketahui atau dicoba. Penelitian ini bisa dikembangkan lagi menjadi induk atau basis pengetahuan untuk sistem keputusan perokok berdasarkan provinsi di Indonesia.

DAFTAR PUSTAKA

[1] R. Fajar and P. T. B. Pustaka, *Bahaya Merokok*. PT Balai Pustaka (Persero), 2011. [Online]. Available: https://books.google.co.id/books?id=HYY2DwAAQBAJ

[2] "Persentase Penduduk Berumur 15 Tahun ke Atas yang Merokok Tembakau selama Sebulan Terakhir Menurut Provinsi (Persen), 2021-2023," *BPS*. https://www.bps.go.id/id/statistics-table/2/MTQzNSMy/persentase-penduduk-berumur-15-tahun-ke-atas-yang-merokok-tembakau-selama-sebulan-terakhir-menurut-provinsi.html (accessed Oct. 30, 2024).

[3] T. S. A, *MENGENAL ROKOK DAN BAHAYANYA*. BE CHAMPION. [Online]. Available: https://books.google.co.id/books?id=9AdrCwAAQBAJ

[4] S. S. M. K. Muhammad Arhami and S. T. M. T. Muhammad Nasir, *Data Mining - Algoritma dan Implementasi*. Andi Offset, 2020. [Online]. Available: https://books.google.co.id/books?id=AtcCEAAAQBAJ

[5] E. Irwansyah and M. Faisal, *Advanced Clustering: Teori dan Aplikasi*. DeePublish, 2015. [Online]. Available: https://books.google.co.id/books?id=8y80BgAAQBAJ

[6] *PEMODELAN K- MEANS ALGORITMA DAN BIG DATA ANALYSIS (PEMETAAN DATA MUSTAHIQ)*. Pascal Books, 2022. [Online]. Available: https://books.google.co.id/books?id=_bJmEAAAQBAJ

[7] zaehol fatah zainur rohman, ahmad homaidi, "G-Tech : Jurnal Teknologi Terapan," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 186–195, 2024, [Online]. Available: https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229

[8] S. Adinugroho and Y. A. Sari, *Implementasi Data Mining Menggunakan Weka*. Universitas Brawijaya Press, 2018. [Online]. Available: https://books.google.co.id/books?id=p91qDwAAQBAJ

[9] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess.v5i2.18519.

[10] M. S. Iskandar and Z. Fatah, "Gudang Jurnal Multidisiplin Ilmu Implementasi Metode Algoritma K-Means Clustering Untuk Menentukan Penerima Program Indonesia Pintar (PIP)," vol. 2, no. November, pp. 1–8, 2024.

[11] R. Kurniawan and R. Dewi, "Penerapan Algoritma K-Means Clustering Dalam Persentase Merokok Pada Penduduk Umur Di Atas 15 Tahun Menurut Provinsi," *J. Sist. Komput. dan Inform. Hal*, vol. 2, no. 2, pp. 178–

186, 2021, doi: 10.30865/json.v2i2.2770.
[12] S. K. M. K. D. A. N. A. P. S. K. M. K. Amril Mutoi Siregar, *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group. [Online]. Available: https://books.google.co.id/books?id=rTlmDwAAQBAJ
[13] Ainurrohma, "Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka," *Prism. Pros. Semin. Nas. Mat.*, vol. 4, pp. 493–499, 2021, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

Mardalius, M. (2018). Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, *4*(2), 123-132.

Mardalius, M., & Christy, T. (2020). Mapping of Potential Customers As a Clothing Promotion Strategy Using K-Means Clustering Algorithm. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, *6*(1), 67-72.