

PENGELOMPOKAN PROVINSI BERDASARKAN PERSENTASE PEMUDA DAN KEBIASAAN MEROKOK TEMBAKAU SEBULAN TERAKHIR TAHUN 2024 MENGGUNAKAN METODE K-MEANS CLUSTERING

Nola Paulina Hasibuan¹, Noni Oktavia Sihombing², Nur Ahdian Putri³, Vijay Moamar Gosal⁴, Andre Yefi Herlangga⁵, Zurnan Alfian⁶
nolapaulina2@gmail.com¹, nonisihombing201@gmail.com², nurahdian6@gmail.com³,
moamarvijay@gmail.com⁴, andreyefi65@gmail.com⁵
Universitas Pamulang

ABSTRAK

Perilaku merokok di antara remaja adalah salah satu isu kesehatan publik yang perlu ditangani dengan serius. Penelitian ini bertujuan untuk mengelompokkan daerah-daerah di Indonesia berdasarkan proporsi pemuda dan kebiasaan merokok tembakau dalam sebulan terakhir di tahun 2024. Metode yang diterapkan adalah K-Means Clustering dengan variabel seperti frekuensi merokok setiap hari, tidak setiap hari, jumlah rata-rata batang rokok yang dihisap, serta persentase individu yang tidak merokok. Hasil analisis menunjukkan bahwa terbentuk dua kelompok provinsi dengan karakteristik perilaku merokok yang berbeda. Penilaian menggunakan Davies-Bouldin Index menandakan bahwa hasil pengelompokan ini mempunyai kualitas yang cukup baik dan dapat dijadikan dasar untuk rekomendasi kebijakan kesehatan berbasis daerah.

Kata Kunci: Cluster K-Means; Rokok, Kesehatan.

ABSTRACT

Smoking behavior among adolescents is one of the public health issues that needs to be addressed seriously. This study aims to group regions in Indonesia based on the proportion of youth and tobacco smoking habits in the last month in 2024. The method applied is K-Means. Clustering with variables such as smoking frequency every day, not every day, the average number of cigarettes smoked, and the percentage of individuals who do not smoke. The results of the analysis show that two groups of provinces are formed with different smoking behavior characteristics. Assessment using the Davies-Bouldin Index indicates that the results of this grouping have quite good quality and can be used as a basis for recommendations for regional-based health policies.

Keywords: K-Means Cluster; Cigarettes, Health.

PENDAHULUAN

Merokok merupakan aktivitas penting bagi para pemuda yang kecanduan merokok, namun menjadi aktivitas yang tidak disukai banyak orang. Dengan populasi pemuda yang diperkirakan mencapai 64,22 juta jiwa atau sekitar seperlima dari total penduduk Indonesia menurut data Susenas 2024, kebiasaan merokok di kalangan anak muda dapat berdampak signifikan terhadap kesehatan individu dan lingkungan sekitar [1]. Rokok mengandung dua bahan kimia yang dapat mempengaruhi kesehatan dan menyebabkan ketergantungan. Kedua zat tersebut adalah tar dan nikotin. Kedua zat tersebut merupakan kandungan paling berbahaya dalam rokok. Dari sekitar 7.000 bahan kimia dalam asap rokok, 2.000 di antaranya ditemukan dalam asap tembakau. Saat asap rokok dihirup, tar akan membentuk lapisan lengket di paru-paru yang dapat menyumbat dan menghancurkan sel-sel di sana. Hal ini dapat menyebabkan kanker paru-paru, emfisema dan kanker lainnya. Tergantung kondisinya, aspal coklat dapat menodai gigi dan kuku perokok [2]. Rokok yang ada di Indonesia memiliki berbagai macam kandungan nikotin dan tar. Jumlah

nikotin dalam daun tembakau berbeda-beda tergantung pada beberapa faktor seperti jenis tembakau, lokasi daun, dan metode penanaman. Seperti, pemangkasan daun yang tidak tepat bisa meningkatkan kadar nikotin pada daunnya, sementara penggunaan pupuk dengan kadar C1 dan N yang berlebihan juga dapat meningkatkan kandungan nikotin. Indonesia adalah negara ketiga dengan jumlah perokok terbanyak setelah China dan India. Jumlah perokok di Indonesia terus meningkat terutama pada usia anak-anak dan remaja usia 15-19 tahun. Menurut estimasi WHO, jumlah perokok di dunia diperkirakan sebanyak 1,1 miliar dan sepertiganya berumur 15 tahun. Sedangkan di Indonesia, Berdasarkan survei yang dilakukan oleh Syahban tahun 1998 di 14 provinsi, terdapat sekitar 60% penduduk usia 10 tahun ke atas termasuk golongan perokok [3].

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 [4]. Kebiasaan merokok tersebut tidak mutlak terlepas dari iklan, baik visual maupun audiovisual. Iklan sebagai terobosan kemajuan teknologi merupakan bentuk pemasaran dengan komunikasi non-personal mengenai produk, ide, jasa yang dibayar oleh sponsor. Dalam hal ini, intervensi iklan sebagai media promosi dianggap cukup menjanjikan dalam mengubah niat dan perilaku merokok. Iklan tersebut dapat mendorong pola perilaku konsumen, sebagaimana kampanye iklan rokok mempengaruhi perokok aktif. Iklan rokok yang ditayangkan secara terbuka mampu mengubah pandangan dan membuat individu tidak menyadari bahwa merokok dapat merugikan kesehatan. Meskipun berbagai peraturan, pendidikan, dan peringatan tentang risiko merokok telah disebarkan, masyarakat, terutama perokok aktif, tetap melanjutkan konsumsi rokok yang tersedia luas di pasaran [5].

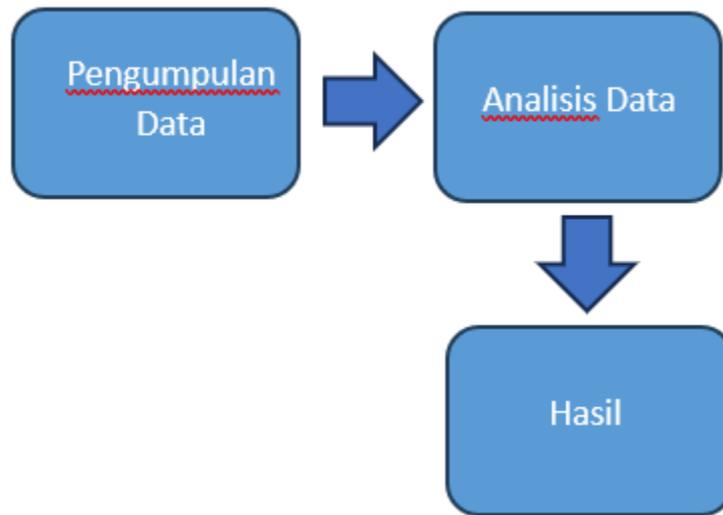
Dalam penelitian ini, penulis memperoleh data dari website resmi pemerintah yaitu Badan Pusat Statistik Indonesia (BPS) khususnya data 2024, mengenai persentase pemuda serta perilaku merokok tembakau dalam sebulan terakhir di seluruh provinsi yang ada di Indonesia. Data tersebut kemudian dikelompokkan menggunakan algoritma K-Means Clustering.

Untuk mendukung penelitian ini, metode K-Means Clustering bisa digunakan sebagai pendekatan analisis data untuk melihat pola tersembunyi dalam kebiasaan merokok antar wilayah di Indonesia. Melalui metode ini, penulis dapat mengelompokkan provinsi-provinsi berdasarkan kesamaan tingkat perokok muda. Diharapkan penelitian ini dapat mengetahui klaster persentase merokok pada tingkat tinggi dan rendah sehingga dapat menjadi masukan kepada pemerintah melalui kegiatan “Hari Tanpa Tembakau Sedunia” dalam merancang kebijakan dan program yang efektif untuk mengurangi tingkat perokok di setiap provinsi di Indonesia [6].

METODE PENELITIAN

Penelitian ini mencakup serangkaian tahapan yang terstruktur untuk mengeksplorasi dan menganalisis data secara mendalam. Metode yang digunakan penulis adalah algoritma K-Means Clustering untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan sejumlah variabel, yaitu kebiasaan merokok setiap hari, kebiasaan merokok tidak setiap hari, tidak merokok, serta rata-rata jumlah batang rokok yang dihisap per hari. Dalam penelitian ini klaster terbentuk terdiri dari dua klaster, yaitu kebiasaan merokok yang paling tinggi dan kelompok dengan kebiasaan merokok yang paling rendah. Hasil dari metode ini kemudian dievaluasi untuk memastikan bahwa hasilnya akurat [13].

Tahapan dari penelitian ini dapat dilihat pada gambar 1



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Data yang digunakan dalam penelitian ini yaitu data sekunder yang diperoleh dari sumber terbuka, yaitu Persentase Pemuda dan Kebiasaan Merokok Tembakau Sebulan Terakhir Tahun 2024 yang berasal dari Badan Pusat Statistik (<https://www.bps.go.id/id/publication/2024/12/31/b2dbaac4542352cea8794590/statistik-pemuda-indonesia-2024.html>)

Data ini mencakup informasi mengenai kebiasaan merokok setiap hari, kebiasaan merokok tidak setiap hari, rata-rata jumlah batang rokok yang dihisap per hari, serta tidak merokok.

Tabel 1. Data yang digunakan

No	Provinsi	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok
1	Aceh	20,20	2,70	12,34	77,10
2	Sumatera Utara	19,08	1,71	12,56	79,21
3	Sumatera Barat	25,07	1,48	13,96	73,45
4	Riau	20,59	1,41	16,96	78,00
5	Jambi	24,53	1,35	17,18	74,13
6	Sumatera Selatan	25,67	1,83	13,95	72,50
7	Bengkulu	26,86	1,79	15,06	71,35
8	Lampung	27,52	1,67	13,62	70,81

9	Kep.Bangka Belitung	23,25	1,18	15,01	75,57
10	Kepulauan Riau	19,86	0,68	15,33	79,48
11	DKI Jakarta	16,98	1,66	9,54	81,36
12	Jawa Barat	27,13	1,83	10,29	71,04
13	Jawa Tengah	25,21	2,20	10,52	72,59
14	DI Yogyakarta	23,12	2,65	9,08	74,23
15	Jawa Timur	23,54	1,91	11,03	74,55
16	Banten	25,83	1,41	12,93	72,75
17	Bali	14,53	0,89	11,74	84,58
18	Nusa Tenggara Barat	27,41	2,63	11,63	69,96
19	Nusa Tenggara Timur	18,84	4,87	8,01	76,28
20	Kalimantan Barat	22,81	1,06	15,16	76,13
21	Kalimantan Tengah	21,19	2,09	16,46	76,72
22	Kalimantan Selatan	18,63	1,87	14,65	79,50
23	Kalimantan Timur	16,46	1,79	14,56	81,75
24	Kalimantan Utara	19,06	1,45	16,20	79,49
25	Sulawesi Utara	17,72	3,41	11,77	78,86
26	Sulawesi Tengah	20,14	2,25	14,98	77,61
27	Sulawesi Selatan	18,91	1,81	14,80	79,28
28	Sulawesi Tenggara	17,44	2,39	13,51	80,17

29	Gorontalo	21,67	3,86	10,94	74,47
30	Sulawesi Barat	21,89	1,97	18,02	76,14
31	Maluku	19,35	4,80	7,56	75,84
32	Maluku Utara	18,92	3,36	11,25	77,73
33	Papua Barat	12,72	6,51	10,54	80,78
34	Papua Barat Daya	14,61	4,72	10,68	80,67
35	Papua	10,99	4,74	9,78	84,27
36	Papua Selatan	20,52	7,14	8,39	72,34
37	Papua Tengah	9,00	3,51	8,67	87,49
38	Papua Pegunungan	6,65	5,31	5,00	88,03

Analisis Data

Analisis data dilakukan dengan melakukan beberapa proses diantaranya yaitu:

1. Pra- Pemrosesan data : Tahapan untuk menghilangkan beberapa permasalahan yang bisa mengganggu saat pemrosesan data. Berdasarkan tabel 1, data tidak perlu melanjutkan pra pemrosesan lebih lanjut, dikarenakan data tersebut sudah sesuai.
2. Analisis klaster (K-Means): K-Means digunakan untuk pengelompokan data pada setiap provinsi berdasarkan kebiasaan merokok setiap hari, kebiasaan merokok tidak setiap hari, rata-rata jumlah batang rokok yang dihisap per hari, serta tidak merokok. Dalam melakukan analisis cluster dilakukan beberapa tahap yaitu sebagai berikut:
 - a. Tentukan jumlah cluster yang akan dibentuk
 - b. Tetapkan titik K secara random untuk menjadikan pusat cluster pertama
 - c. Menghitung jarak Euclidian Distance antara setiap titik data dengan menggunakan masing-masing centroid. Untuk menghitung jarak semua data kesetiap titik pusat cluster dapat menggunakan Rumus Euclidian Distan

$$D(i,j) = \frac{\sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}}{(1)}$$

Ket :

D(Ij) : Jarak data ke i ke pusat cluster j

Xki = Data ke i pada atribut data ke k

Xkj = Titik pusat ke j pada atribut data k

- d. Pembaruan Pusat: Hitung ulang pusat setiap kelompok dengan mengambil rata-rata dari semua titik data yang termasuk dalam kelompok tersebut.
- e. Iterasi: Langkah 2 dan 3 diulangi hingga tidak ada perubahan yang signifikan dalam posisi pusat kelompok atau titik-titik data tidak berpindah kelompok.e.Penyelesaian: Proses dihentikan ketika kriteria penghentian telah terpenuhi

Interprestasi Hasil

Hasil dari algoritma K-Means adalah sebuah model yang dapat digunakan untuk

memprediksi kelompok mana suatu data baru akan termasuk, serta posisi pusat dari masing-masing kelompok. Dengan hasil analisis cluster yang telah dijabarkan maka akan mendapatkan hasil untuk mengetahui Presentase pemuda provinsi yang tertinggi dan terendah di provinsi Indonesia.

HASIL DAN PEMBAHASAN

1. Menentukan Titik Pusat Cluster

Menentukan titik pusat kluster awal dalam metode Data Mining dengan menggunakan algoritma K-Means Clustering dilakukan dengan memilih beberapa titik secara acak dari data. Titik pusat awal ini dipilih secara acak oleh penulis sebagai langkah awal untuk proses inisialisasi kluster. Titik tersebut selanjutnya berfungsi sebagai referensi untuk menggolongkan data ke dalam kluster berdasarkan kedekatan jaraknya dengan titik pusat kluster. Tabel berikut menunjukkan daftar titik pusat kluster yang sudah dipilih:

Tabel 2. Titik pusat awal cluster

Provinsi	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok
Gorontalo	21,67	3,86	10,94	74,47
Sulawesi Barat	21,89	1,97	18,02	76,14

2. Algoritma K-Means Clustering

a. Menghitung Jarak Data Ke Pusat Cluster

Pada algoritma K-Means Clustering, proses menghitung jarak antara setiap data dengan pusat cluster dilakukan dengan tujuan untuk menentukan pembaruan posisi titik pusat cluster yang baru pada setiap iterasi. Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid, sehingga data dapat dikelompokkan ke dalam cluster yang paling dekat jaraknya.

$$D(ik) = \sqrt{\sum_j^m (Cij - Ckj)^2}$$

Dimana: Cij = Pusat cluster
Ckj = Data

Setelah melakukan perhitungan maka didapatkan hasil seperti tabel berikut:

Tabel 3. Hasil Iterasi 1

Provinsi	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok	C1	C2	Jarak terdekat	Pengelompokan data
Aceh	20,20	2,7	12,34	77,1	11,6925	35,407	11,6925	Cluster 1
Sumatera Utara	19,08	1,71	12,56	79,21	32,3045	42,114	32,3045	Cluster 1
Sumatera Barat	25,07	1,48	13,96	73,45	19,2252	27,14	19,2252	Cluster 1
Riau	20,59	1,41	16,96	78	55,7838	6,1968	6,1968	Cluster2
Jambi	24,53	1,35	17,18	74,13	48,2133	7,7701	7,7701	Cluster2
Sumatera Selatan	25,67	1,83	13,95	72,5	21,0619	33,614	21,0619	Cluster 1

Bengkulu	26,86	1,79	15,06	71,35	36,1837	36,708	36,1837	Cluster 1
Lampung	27,52	1,67	13,62	70,81	31,2241	53,489	31,2241	Cluster 1
Kep. Bangka Belitung	23,25	1,18	15,01	75,57	26,5373	11,369	11,3691	Cluster2
Kepulauan Riau	19,86	0,68	15,33	79,48	56,2946	22,086	22,0858	Cluster2
DKI Jakarta	16,98	1,66	9,54	81,36	58,9621	104,16	58,9621	Cluster 1
Jawa Barat	27,13	1,83	10,29	71,04	21,7683	91,023	21,7683	Cluster 1
Jawa Tengah	25,21	2,2	10,52	72,59	10,0064	72,225	10,0064	Cluster 1
DI Yogyakarta	23,12	2,65	9,08	74,23	6,4313	85,264	6,4313	Cluster 1
Jawa Timur	23,54	1,91	11,03	74,55	5,687	53,042	5,687	Cluster 1
Banten	25,83	1,41	12,93	72,75	17,081	41,654	17,081	Cluster 1
Bali	14,53	0,89	11,74	84,58	118,813	119,2	118,813	Cluster 1
Nusa Tenggara Barat	27,41	2,63	11,63	69,96	28,0691	84,98	28,0691	Cluster 1
Nusa Tenggara Timur	18,84	4,87	8,01	76,28	15,7111	111,68	15,7111	Cluster 1
Kalimantan Barat	22,81	1,06	15,16	76,13	29,544	9,9278	9,9278	Cluster2
Kalimantan Tengah	21,19	2,09	16,46	76,72	39,1458	3,4844	3,4844	Cluster2
Kalimantan Selatan	18,63	1,87	14,65	79,5	46,0651	25,917	25,9165	Cluster2
Kalimantan Timur	16,46	1,79	14,56	81,75	75,5977	48,906	48,9061	Cluster2
Kalimantan Utara	19,06	1,45	16,2	79,49	61,2861	17,635	17,6353	Cluster2
Sulawesi Utara	17,72	3,41	11,77	78,86	24,1135	52,705	24,1135	Cluster 1
Sulawesi Tengah	20,14	2,25	14,98	77,61	30,3033	13,231	13,2309	Cluster2
Sulawesi Selatan	18,91	1,81	14,8	79,28	44,9982	23,234	23,2336	Cluster2
Sulawesi Tenggara	17,44	2,39	13,51	80,17	45,4858	41,207	41,2074	Cluster2
Gorontalo	21,67	3,86	10,94	74,47	0	56,707	0	Cluster 1

Sulawesi Barat	21,89	1,97	18,02	76,14	56,7074	0	0	Cluster2
Maluku	19,35	4,8	7,56	75,84	16,5049	120,05	16,5049	Cluster 1
Maluku Utara	18,92	3,36	11,25	77,73	13,7237	53,263	13,7237	Cluster 1
Papua Barat	12,72	6,51	10,54	80,78	55,9486	107,26	55,9486	Cluster 1
Papua Barat Daya	14,61	4,72	10,68	80,67	46,3072	89,239	46,3072	Cluster 1
Papua	10,99	4,74	9,78	84,27	108,84	152,57	108,84	Cluster 1
Papua Selatan	20,52	7,14	8,39	72,34	22,9478	135,28	22,9478	Cluster 1
Papua Tengah	9	3,51	8,67	87,49	187,466	231,51	187,4658	Cluster 1
Papua Pegunungan	6,65	5,31	5	88,03	236,28	337,29	236,2797	Cluster 1

Berdasarkan tabel hasil iterasi 1, maka diperoleh hasil pengelompokan seperti yang terlihat pada tabel dibawah:

Tabel 4. Hasil pengelompokan Iterasi 1

Kelompok Cluster	Anggota Cluster	Jumlah
C1	1,2,3,6,7,8,11,12,13,14,15,16,17,18,19,25,29,31,32,33,34,35,36,37,38	25
C2	4,5,9,10,20,21,22,23,24,26,27,28,30	13

Tahapan selanjutnya adalah menentukan titik pusat cluster baru yakni dengan mencari rata-rata semua anggota cluster dengan cara menjumlahkan seluruh value anggota cluster dibagi jumlah anggota.

Tabel 5. Titik pusat cluster baru

Cluster	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok
Cluster 1	19,9656	3,1436	10,8336	76,8896
Cluster2	20,36615	1,638462	15,60154	77,99769

b. Perhitungan Iterasi 2

Setelah melakukan perhitungan dengan menggunakan titik pusat cluster yang baru maka ditemukan hasil sebagai berikut:

Kelompok Cluster	Anggota Cluster	Jumlah
C1	1,2,6,8,11,12,13,14,15,16,18,19,25,29,31,32,33,34,35,36,37,38	22
C2	4,5,9,10,20,21,22,23,24,26,27,28,30	16

Tabel 6. Hasil pengelompokan Iterasi 2

Dengan data yang telah diperoleh, dilakukan perhitungan untuk menentukan pusat cluster baru dan diperoleh hasil sebagai berikut:

Tabel 7. Titik pusat cluster baru

Cluster	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok
Cluster1	19,66727	3,383182	10,45818	76,94818
Cluster2	20,70125	1,59125	15,22375	77,70938

c. Perhitungan Iterasi 3

Setelah menghitung jarak antara setiap data dengan pusat cluster baru, maka diperoleh data sebagai berikut:

Tabel 8. Hasil pengelompokan Iterasi 3

Kelompok Cluster	Anggota Cluster	Jumlah
C1	1,8,11,12,13,14,15,16,18,19,25,29,31,32,33,34,35,36,37,38	20
C2	2,3,4,5,6,7,9,10,17,20,21,22,23,24,26,27,28,30,	18

d. Perhitungan Iterasi 4

Setelah melakukan perhitungan dengan menggunakan titik pusat cluster yang baru maka ditemukan hasil sebagai berikut:

Tabel 10. Hasil pengelompokan Iterasi 4

Kelompok Cluster	Anggota Cluster	Jumlah
C1	1,11,12,13,14,15,18,19,25,29,31,32,33,34,35,36,37,38	18
C2	2,3,4,5,6,7,8,9,10,16,17,20,21,22,23,24,26,27,28,30,	20

Dengan data yang telah diperoleh, dilakukan perhitungan untuk menentukan pusat cluster baru dan diperoleh hasil sebagai berikut:

Tabel 11. Titik pusat cluster baru

Cluster	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok
Cluster1	18,58778	3,767222	9,834444	77,64389
Cluster2	21,466	1,604	14,832	76,931

e. Perhitungan Iterasi 5

Setelah melakukan perhitungan dengan menggunakan titik pusat cluster yang baru maka ditemukan hasil sebagai berikut:

Tabel 12. Hasil pengelompokan Iterasi 5

Kelompok Cluster	Anggota Cluster	Jumlah
C1	11,12,13,14,15,17,19,25,29,31,32,33,34,35,36,37,38	17
C2	1,2,3,4,5,6,7,8,9,10,16,18,20,21,22,23,24,26,27,28,30,	21

Dengan data yang telah diperoleh, dilakukan perhitungan untuk menentukan pusat cluster baru dan diperoleh hasil sebagai berikut:

Tabel 13. Hasil pengelompokan Iterasi 5

Cluster	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok
Cluster 1	17,73529	3,727647	9,693529	78,53588
Cluster2	21,07333	1,706667	13,9781	72,4581

f. Perhitungan Iterasi 6

Setelah melakukan perhitungan dengan menggunakan titik pusat cluster yang baru maka ditemukan hasil sebagai berikut:

Tabel 14. Hasil pengelompokan Iterasi 6

Kelompok Cluster	Anggota cluster	Jumlah
C1	1,2,10,11,14,17,19,22,23,24,25,27,28,31,32,33,34,35,36,37,38	21
C2	3,4,5,6,7,8,9,12,13,15,16,18,20,21,26,29,30	17

Dengan data yang telah diperoleh, dilakukan perhitungan untuk menentukan pusat cluster baru dan diperoleh hasil sebagai berikut:

Tabel 15. Titik pusat cluster baru

Cluster	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok
Cluster1	16,83762	3,236667	11,23619	79,92571
Cluster 2	24,13588	1,877647	13,98235	73,98647

g. Perhitungan iterasi 7

Setelah melakukan perhitungan dengan menggunakan titik pusat cluster yang baru maka ditemukan hasil sebagai berikut:

Tabel 16. Hasil pengelompokan Iterasi 7

Kelompok Cluster	Anggota cluster	jumlah
C1	1,2,10,11,17,19,22,23,24,25,27,28,31,32,33,34,35,37,38	19
C2	3,4,5,6,7,8,9,12,13,14,15,16,18,20,21,26,29,30,26	19

Dengan data yang telah diperoleh, dilakukan perhitungan untuk menentukan pusat cluster baru dan diperoleh hasil sebagai berikut:

Tabel 17. Titik pusat cluster baru

Cluster	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok
Cluster1	16,31316	3,062105	11,49947	80,62474
Cluster2	23,89211	2,195263	13,43	73,91263

h. perhitungan iterasi 8

Setelah melakukan perhitungan dengan menggunakan titik pusat cluster yang baru maka ditemukan hasil sebagai berikut:

Tabel 18. Hasil pengelompokan Iterasi 8

Kelompok Cluster	Anggota cluster	jumlah
C1	2,10,11,17,19,22,23,24,25,27,28,31,32,33,34,35,37,38	18
C2	1,3,4,5,6,7,8,9,12,13,14,15,16,18,20,21,26,29,30,36	20

Dengan data yang telah diperoleh, dilakukan perhitungan untuk menentukan pusat cluster baru dan diperoleh hasil sebagai berikut:

Tabel 19. Titik pusat cluster baru

Cluster	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok
Cluster1	16,09722	3,082222	11,45278	80,82056
Cluster2	23,7075	2,2205	13,3755	74,072

i. Perhitungan iterasi 9

Setelah melakukan perhitungan dengan menggunakan titik pusat cluster yang baru maka ditemukan hasil sebagai berikut:

Tabel 20. Hasil pengelompokan Iterasi 9

Kelompok Cluster	Anggota cluster	jumlah
C1	2,10,11,17,19,22,23,24,25,27,28,31,32,33,34,35,37,38	18
C2	1,3,4,5,6,7,8,9,12,13,14,15,16,18,20,21,26,29,30,36	20

Dengan data yang telah diperoleh, dilakukan perhitungan untuk menentukan pusat cluster baru dan diperoleh hasil sebagai berikut:

Tabel 21. Titik pusat cluster baru

Cluster	Setiap Hari	Tidak Setiap Hari	Rata-Rata Jumlah Batang Rokok Yang Dihisap Per Hari	Tidak Merokok
Cluster1	16,09722	3,082222	11,45278	80,82056
Cluster2	23,7075	2,2205	13,3755	74,072

Iterasi berhenti pada Iterasi ke 9 karena tidak ada perubahan cluster yang terjadi dengan 38 provinsi di cluster 1 dan 18 provinsi di cluster 20.

- Cluster 1 : Sumatera utara, kepulauan Riau, Dki Jakarta, Bali, Nusa Tenggara Timur, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara , Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Maluku, maluku Utara, Papua Barat, Papua Barat Daya, Papua, Papua Tengah, Papua pegunungan
- Cluster 2 : Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep.Bangka Belitung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tengah, Gorontalo, Sulawesi Barat, Papua Selatan.

Hasil klasterisasi berdasarkan persentase pemuda dan kebiasaan merokok tembakau sebulan terakhir di berbagai provinsi Indonesia tahun 2024 menghasilkan dua klaster utama yang mencerminkan karakteristik distribusi perokok di tiap wilayah. Klaster pertama menunjukkan provinsi dengan jumlah persentase pemuda terendah yang merokok tembakau dalam sebulan terakhir, klaster kedua menunjukkan provinsi dengan persentase pemuda tertinggi yang merokok tembakau dalam sebulan terakhir.

Perbedaan yang signifikan antara kedua kluster ini dipengaruhi oleh berbagai aspek seperti tingkat pendidikan, pendapatan daerah, dampak lingkungan sosial, serta kebijakan pengendalian rokok di masing-masing provinsi. Secara keseluruhan, hasil pengelompokan ini memberikan wawasan yang jelas mengenai pembagian provinsi berdasarkan persentase pemuda dan kebiasaan merokok tembakau. Dengan adanya penelitian ini, diharapkan hasil ini dapat menjadi acuan dalam merancang kebijakan pemerintah dalam pengendalian rokok yang lebih tepat dan efisien, sesuai dengan karakteristik dan kebutuhan setiap daerah [14].

KESIMPULAN

Penelitian ini berawal dari tingginya angka pemuda yang merokok di Indonesia dan dampak buruknya terhadap kesehatan masyarakat. Fokus utama dari studi ini adalah mengelompokkan provinsi-provinsi di Indonesia berdasarkan proporsi pemuda dan kebiasaan merokok tembakau dalam sebulan terakhir dengan menggunakan metode K-Means Clustering, demi memberikan saran kebijakan yang lebih efektif untuk mengendalikan konsumsi rokok.

Dalam penelitian ini, metode K-Means Clustering diterapkan untuk menganalisis dan mengelompokkan provinsi-provinsi di Indonesia sesuai dengan persentase pemuda dan perilaku merokok tembakau. K-Means Clustering adalah algoritma pembelajaran mesin yang berfungsi untuk mengelompokkan data ke dalam beberapa klaster berdasarkan kesamaan karakteristik. Dalam penelitian ini, algoritma tersebut digunakan untuk mengenali pola distribusi kebiasaan merokok di kalangan pemuda secara geografis di seluruh provinsi di Indonesia.

Penelitian ini membagi provinsi-provinsi di Indonesia menjadi dua klaster berdasarkan kebiasaan merokok pemuda. Klaster pertama mencakup provinsi dengan tingkat merokok rendah, seperti DKI Jakarta, Bali, dan Papua, yang memiliki persentase perokok harian di bawah 20% dan konsumsi rokok kurang dari 10 batang per hari. Klaster kedua terdiri dari provinsi dengan kebiasaan merokok tinggi, seperti Jawa Barat, Sumatera Barat, dan Lampung, dengan perokok harian di atas 25% dan konsumsi lebih dari 12 batang per hari. Pengelompokan ini dilakukan menggunakan metode K-Means hingga posisi centroid stabil. Hasil penelitian ini bisa menjadi dasar untuk kebijakan kesehatan yang lebih tepat.

Temuan ini memberikan pemahaman yang jelas mengenai perbedaan kebiasaan merokok di kalangan pemuda antar provinsi di Indonesia. Provinsi yang tergabung dalam klaster dengan kebiasaan merokok tinggi, terutama yang berada di daerah Sumatera dan Jawa, membutuhkan intervensi kebijakan yang lebih kuat seperti program edukasi kesehatan yang lebih intensif dan pembatasan akses terhadap produk tembakau. Sementara itu, provinsi yang tergolong dalam klaster dengan kebiasaan merokok rendah disarankan untuk terus menjaga pencapaian tersebut dengan menguatkan regulasi Kawasan Tanpa Rokok dan upaya pencegahan.

DAFTAR PUSTAKA

- I. Wafa, "GoodStats," 3 Januari 2025. [Online]. Available: <https://data.goodstats.id/statistic/23-pemuda-indonesia-merokok-rata-rata-12-batang-per-hari-qCPDM>. [Accessed 22 Juni 2025].
- E. Hariani, "FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN PERILAKU MEROKOK PADA REMAJA SMA DI MENSIKU TAHUN 2024," *Multidisciplinary Indonesian Center Journal*, vol. Vol.2. No 1 Edisi Januari 2025, pp. pp 1-8, 2025.
- R. Florentika and K. Widyanto, "22ERUDITIO Vol. 2, No. 2, Juni 2022, pp. 22-32 P-ISSN: 2580-7722 | E-ISSN: 2807-6222Relita Florentika, Widiyanto Kumiawan Analisis Kuantitatif Tar dan Nikotin Terhadap Rokok Kretek yang Beredar di IndonesiaERUDITIO Vol. 2, No. 2, Juni 2022, pp. 22-32 P-I," *Eruditio : Indonesia Journal Of Food and Drug Safety*, Vols. Vol.2, No 2, Juni 2022, pp. pp.22-32, 2022.
- A. D. Nur'azizah and Z. Fatah, "Implementasi Algoritma K-Means Clustering Untuk Perokok Usia Di Atas 15 Tahun," *Gudang Jurnal Multidisiplin Ilmu(GJMI)*, vol. Vol 2 ; Nomor 10 ; November 2024, pp. pp 133-139, 2024.
- F. D. ulya, "AANALISIS TINGKAT INTENSITAS KONSUMSI ROKOK PADA PEROKOK AKTIF DI KOTA MAGELANG SETELAH MELIHAT IKLAN ROKOK," *Jurnal Tata Kelola dan Administrasi Publik*, Vols. Vol 2, No 1 Desember 2024, pp. pp 1- 13, 24 - 12 - 2024.
- R. Kumiawan, "Penerapan Algoritma K-Means Clustering Dalam Persentase Merokok Pada Penduduk Umur Di Atas 15 Tahun Menurut Provinsi," *Jurnal Sistem Komputer dan Informatika (JSON)*, Vols. Vol 2, No 2 (2021), pp. pp 178-186, 24 Januari 2021.
- V. W. Eda Kavlakoglu, "Apa itu K-means clustering?," *IBM*, 26 06 2024. [Online]. Available: <https://www.ibm.com/id-id/topics/k-means-clustering>.
- Dosbing, "Implementasi dan Perhitungan Manual Algoritma K-Means dalam Clustering Data," *dosbung.id*, 24 06 2024. [Online]. Available: <https://dosbung.id/2024/06/24/implementasi-dan-perhitungan-manual-algoritma-k-means-dalam-clustering-data/>.
- Y. m. m. n. &. xeratic, "K-Means Clustering, Salah Satu Contoh Teknik Analisis Data Populer," *dQLab*, 12 08 2022. [Online]. Available: <https://dqlab.id/k-means-clustering-salah-satu-contoh-teknik-analisis-data-populer>.
- D. F. Deni Triyansyah, "Analisis Data Mining Menggunakan Algoritma K-means clustering untuk menemukan strategi marketing," *Telekomunikasi dan komputer*, vol. 8, no. 3, pp. 163-164, 2018.
- N. N. Salsabila, N. Indraswari and B. Sujatmiko, "Gambaran Kebiasaan Merokok Di Indonesia Berdasarkan Indonesia Family Life Survey 5 (IFLS 5)," *Jurnal Ekonomi Kesehatan Indonesia*, vol. 7, no. 1, p. 14, 2022.
- R. Isra and U. , "Presepsi Bahaya merokok Terhadap Kesehatan," *Merokok, Kesehatan*, p. 240, 2018.
- N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal ilmiah Informatika Dan Ilmu Komputer(ILMA - ILKOM)*, Vols. Volume 3, Nomor 1, Maret 2024, pp. page 46-56, 2024-03-22.
- D. M. Karina and A. P. Sanusi, "Visualisasi Data Persebaran Perokok di Indonesia menggunakan K-Means dan Tableau," *Digital Transformation Technology (Digitech)*, Vols. Volume4, Number2,September 2024, 14 Januari 2025.
- YangJjie, "Adaptive Initialization Method for K-Means Algorithm," *Metode Inisialisasi Adaptif untuk Algoritma K-Means*, 25 November 2021. [Online]. Available: <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2021.740817/full>.