

BAB 2

TINJAUAN PUSTAKA

2.1 Rokok

Rokok adalah produk tembakau yang dimaksudkan untuk dibakar dan dihisap (Iriyanti & Mandagi, 2022). Rokok di Indonesia sudah diproduksi sejak akhir abad ke-19. Merokok menjadi awal seseorang mudah untuk mencoba zat adiktif lain (B. Kurniawan & Ayu, 2023). Merokok merupakan perilaku yang dilakukan banyak orang setiap hari. Bagi sebagian orang yang kecanduan rokok, merokok juga merupakan kebutuhan yang tidak dapat dihindari (A J Rifiana, Mi Evelianti, 2023). Perilaku merokok pada remaja dapat disebabkan oleh sejumlah hal, seperti pengaruh teman, orang tua yang merokok, sifat kepribadian seperti rasa ingin tahu, dan iklan (Utami, 2020). Setiap orang yang belum menikah berusia antara 10 dan 24 tahun dianggap remaja (Pratiwi, 2022).

2.2 Kecanduan Merokok

Meski merokok merupakan kebiasaan yang merugikan, banyak orang yang masih melakukannya (Patana & Elon, 2019). Kesulitan para perokok remaja untuk berhenti merokok disebabkan karena rokok sangat membuat kecanduan (Jatmika et al., 2023). Merokok dapat memberikan dampak buruk bagi kesehatan, tidak hanya bagi perokok itu sendiri tetapi juga bagi orang-orang di sekitarnya yang merokok (Sekeronej et al., 2020). Mengingat besarnya dampak yang ditimbulkan oleh perilaku merokok terutama di kalangan remaja, maka upaya pencegahan dan pengendalian perilaku merokok menjadi penting (Oktania et al., 2023). Selain itu, orang mungkin merasa tertekan untuk mencoba merokok karena interaksi sosial mereka dengan perokok (Aiga Nurkhalilah Pasaribu, Ayu Carolina, 2023). Rata-rata usia mulai merokok secara nasional adalah 17,6 tahun dengan persentase tertinggi populasi yang mulai merokok setiap hari pada usia 15-19 tahun (Tentang Kenakalan Remaja et al., 2023).

2.2.1 Definisi Nikotin

Nikotin berasal dari bahan kimia yang ada dalam tembakau yang dapat membuat seseorang ketagihan. Nikotin masuk ke dalam tubuh melalui penghirupan asap rokok ke otak hanya dalam waktu 20 detik. Zat ini membuat penggunanya

mendapatkan efek menenangkan pada otak untuk beberapa saat, sehingga orang yang mengonsumsi rokok akan ketagihan untuk merokok lagi.

2.2.2 Faktor Resiko Kecanduan Nikotin

Seseorang yang merokok atau mengonsumsi akan mengalami resiko ketergantungan. Faktor yang mempengaruhi hal tersebut yaitu usia, genetika, pengaruh lingkungan, depresi dan penggunaan zat.

2.2.3 Mengatasi Merokok

Untuk berhenti merokok yaitu dengan menggunakan obat-obatan, konseling perilaku dan perawatan di rumah. Menghindari nikotin sejak awal adalah cara terbaik untuk mencegah ketergantungan.

2.2.4 Komplikasi Kecanduan Nikotin

Kanker paru-paru, kanker mulut, masalah jantung, diabetes, penyakit gigi, dan kondisi lainnya semuanya dapat disebabkan oleh kecanduan nikotin.

2.2.5 Faktor-Faktor Kecanduan Merokok

Kehadiran iklan rokok yang ditayangkan dari berbagai media juga dapat memengaruhi semua kalangan, terutama remaja. Banyaknya remaja yang merokok menjadi salah satu pemicu pola asuh orang tua yang kurang baik. Remaja cenderung memiliki teman yang merokok jika mereka juga merokok. Fakta ini meningkatkan kemungkinan bahwa teman memiliki dampak terhadap remaja (Susilaningih et al., 2022).

2.2.6 Perbedaan Perokok Aktif dan Perokok Pasif

Perokok aktif adalah perokok yang secara aktif menghisap tembakau langsung dari rokoknya sedangkan perokok pasif adalah orang-orang yang terpapar dan secara tidak sengaja menghirup asap rokok dari perokok aktif.

2.3 *Data Mining*

Data mining menganalisis sejumlah besar kumpulan data observasi dan menemukan hubungan yang tidak terduga (Gustientiedina et al., 2019). Rahmat, dkk (2021) *data mining* adalah proses pengorganisasian informasi penting dari sejumlah besar data menjadi informasi yang tidak dapat ditemukan secara manual dalam

basis data. *Data mining* telah ada sejak tahun 1990-an, melibatkan pengorganisasian data ke dalam satu atau beberapa kluster berdasarkan seberapa mirip *item* dalam kluster tersebut satu sama lain (R. Kurniawan & Dewi, 2021).

2.4 *Clustering*

Cluster Analysis is a multivariate analysis (many variables) that functions to group objects or several variables based on their characteristics. In addition, cluster analysis also aims to maximize the similarity of objects in the cluster while also maximizing differences between clusters. Analisis kluster merupakan analisis multivariat (banyak variabel) yang berfungsi untuk mengelompokkan objek atau beberapa variabel berdasarkan karakteristiknya. Selain itu, analisis kluster juga bertujuan untuk memaksimalkan kesamaan objek dalam suatu kluster sekaligus memaksimalkan perbedaan antar kluster (Fadilah & Husein, 2019).

Terdapat dua metode dalam *clustering*, yaitu hierarki dan non-hierarki. Metode hierarki digunakan ketika jumlah *cluster* yang akan dibentuk belum diketahui sebelumnya dan kurang efisien untuk data yang sangat besar. Sedangkan metode non-hierarki digunakan jika *cluster* yang akan dibentuk sudah diketahui sebelumnya (Ganmanah & Kudus, 2021).

2.5 Ukuran Kemiripan

Jarak antara objek dan kluster digunakan untuk menghitung metrik kesamaan dalam analisis kluster (Novidianto & Fithriasari, 2021).

2.6 Metode *K-Means*

Rima, dkk (2022) mengatakan Analisis kluster dengan metode non-hierarki (*K-Means Cluster*) digunakan untuk pengelompokan objek dimana jumlah kluster yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur klusterisasi, selanjutnya metode ini dapat diaplikasikan pada data yang lebih besar (Aprilia et al., 2022). Analisis kluster mengharuskan asumsi non-multikolinearitas, yang menegaskan bahwa harus ada hubungan linear yang sempurna atau pasti antara beberapa atau semua variabel terpenuhi (Az-zahra et al., 2021). Metode yang termasuk dalam algoritma pengelompokan berbasis jarak membagi data menjadi sejumlah kelompok dan algoritma ini hanya bekerja pada atribut numerik (R. Kurniawan & Dewi, 2021).

Langkah-langkah algoritma *K-means* (R. Kurniawan & Dewi, 2021).

1. Tentukan k sebagai jumlah kluster yang akan dibentuk.
2. Alokasikan data ke dalam kluster secara acak.
3. Tentukan pusat kluster (*centroid*) data di setiap kluster, dengan persamaan:

$$C_{kj} = \frac{x_{1j} + x_{2j} + \dots + x_{nj}}{n} \quad (2.1)$$

Dimana:

C_{kj} = Pusat kluster ke- k pada variabel ke j ($j = 1, 2, \dots, p$)

N = Banyak data pada kluster ke- k

4. Tentukan jarak antara setiap objek dan setiap *centroid* dengan menghitung jarak antara setiap objek dan setiap *centroid* menggunakan jarak *Euclidean*.

$$d(X_i, X_g) = \sqrt{\sum_{j=1}^p (X_{ij} - X_{gj})^2} \quad (2.2)$$

5. Hitung fungsi objektif menggunakan rumus:

$$J = \sum_{j=1}^n \sum_{k=1}^k a_{ij} d(x_{ij}, C_{kj})^2 \quad (2.3)$$

6. Alokasikan setiap data ke *centroid* /rata-rata terdekat yang dirumuskan sebagai berikut:

$$a_{ij} = \begin{cases} 1, & s = \min \{d(x_{ij}, C_{kj})\} \\ 0, & \text{lainnya} \end{cases} \quad (2.4)$$

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

7. Ulangi langkah 3-6 hingga tidak ada lagi pergerakan objek atau perubahan fungsi objektifnya.

2.7 Metode *K-modes*

Varian dari algoritma *K-means* disebut *K-modes*. Teknik yang paling populer untuk mengklasifikasikan kumpulan data kategorik yang cukup besar ke dalam kluster yang berbeda, masing-masing diwakili oleh mode atau nilai yang paling sering muncul, adalah metode *K-modes* (Dwiyanti et al., 2022).

Langkah-langkah *K-modes* (Dwiyanti et al., 2022):

1. Tentukan jumlah kluster (k) yang akan dibentuk dari r data. Jumlah data harus lebih besar dari jumlah kluster ($r > k$).
2. Tentukan k mode awal data yang akan digunakan untuk pusat kluster (*centroid*). Jumlah *centroid* yang dipilih harus sesuai dengan jumlah kluster yang akan dibentuk.
3. Cari jarak setiap objek ke setiap *centroid*. Jarak dicari dengan menggunakan ukuran ketidaksamaan pencocokan sederhana. Asumsikan B_1 dan B_2 adalah dua data yang memiliki n variabel kategori. Ukuran ketidaksamaan antara B_1 dan B_2 dapat didefinisikan sebagai ketidaksamaan total variabel kategori yang sesuai dari kedua data. Semakin mirip kedua data, semakin kecil nilai ketidaksamaannya.

Persamaan untuk ukuran ketidaksamaan sederhana:

$$d(V_1, V_2) = \sum_{j=1}^n \delta(x_{1j}, x_{2j}) \quad (2.5)$$

dengan:

$$\delta(x_{1j}, x_{2j}) = \begin{cases} 0, & x_{1j} = x_{2j} \\ 1, & x_{1j} \neq x_{2j} \end{cases} \quad (2.6)$$

Keterangan:

x_{1j} : nilai dari variabel ke- j pada objek V_1

x_{2j} : nilai dari variabel ke- j pada objek V_2

4. Gunakan rumus jarak terdekat untuk menyusun objek menurut jarak terdekatnya ke *centroid*. Nilai ukuran ketidakmiripan dasar dapat digunakan untuk mengidentifikasi kluster mana yang paling dekat dengan item.
5. Perbarui *centroid* setiap kluster tergantung pada modus setiap variabel anggota setelah semua data dikelompokkan ke dalam kluster terdekat.
6. Dengan menggunakan metrik ketidakmiripan langsung, seperti langkah (3), hitung ulang jarak setiap titik data ke *centroid* baru. Data dipindahkan kembali ke kluster yang lebih dekat jika perhitungan jarak menunjukkan bahwa data tersebut lebih dekat ke kluster tersebut daripada ke kluster saat ini.
7. Lanjutkan langkah (5) dan (6) hingga tidak ada kluster data yang dipindahkan.

2.8 Algoritma *K-Prototypes*

Algoritma *K-prototypes* adalah algoritma yang digunakan dalam aplikasi ini. Untuk mengelompokkan item bertipe campuran, metode *K-Prototypes* menggabungkan algoritma ritme *k-means* dan *k-modes* (Ganmanah & Kudus, 2021). Algoritma ini mempunyai kelebihan yaitu algoritmanya tidak terlalu rumit, dapat menangani data bertipe campuran yaitu numerik dan kategorikal, sangat efektif untuk menangani data yang sangat besar dan lebih baik dibandingkan algoritma yang berbasis hierarki.

Jarak tipe data campuran:

$$d(x_i, c_y) = \sum_{p=1}^m (x_{ip} - c_{yp})^2 + \gamma \sum_{q=1}^m \delta(x_{iq}; c_{yq}) \quad (2.7)$$

Keterangan:

$d(x_i, c_y)$: jarak tipe data campuran antara objek ke- i ke pusat *cluster* ke- y

$\sum_{p=1}^m (x_{ip} - c_{yp})^2$: ukuran jarak untuk tipe data numerik

$\sum_{q=1}^m \delta(x_{iq}, c_{yq})$: ukuran jarak untuk tipe data kategorik

γ : koefisien penimbang

Hasil rata-rata deviasi standar (σ) dari semua variabel numerik yang digunakan, sebagaimana ditentukan dalam persamaan 2.8, digunakan untuk menentukan nilai koefisien gamma (γ) yang digunakan dalam persamaan 2.7.

$$\gamma = \frac{1}{m} \sum_{p=1}^m \sigma_p \quad (2.8)$$

Keterangan:

σ_p : simpangan baku variabel numerik ke- p

p : 1, 2, ..., m

m : jumlah variabel numerik

Sebelum mencari nilai koefisien gamma (γ) terlebih dahulu mencari hasil simpangan baku atau standar deviasinya menggunakan persamaan 2.9. Standar deviasi adalah rumus dalam statistika yang dapat digunakan untuk mendapatkan data dari suatu populasi.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (2.9)$$

Keterangan:

s : standar deviasi sampel

\bar{x} : nilai rata-rata

x_i : sampel data

n : jumlah data

Langkah-langkah Algoritma *K-prototypes* (Reihanah et al., 2024):

1. Tentukan jumlah *cluster* (k) yang akan dibentuk.
2. Tentukan *k-prototype* awal sebagai *centroid* atau titik pusat awal *cluster* awal.
3. Hitung jarak dengan ukuran jarak tipe data campuran sesuai persamaan 2.7 pada semua objek data terhadap *centroid* yang ditentukan.
4. Tempatkan semua objek ke dalam *cluster* yang memiliki nilai jarak terdekat.

$$e_i = \begin{cases} 1, & s = \min \{d(x_i, c_1), d(x_i, c_2), \dots, d(x_i, c_y)\} \\ 0, & \text{lainnya} \end{cases} \quad (2.10)$$

Keterangan:

e_i : nilai keanggotaan objek ke- i

s : jarak minimum dari objek ke- i ke pusat *cluster* ke- y setelah dibandingkan

5. Dengan memanfaatkan nilai rata-rata persamaan 2.11 untuk variabel numerik, hitung nilai *centroid* baru.

$$c_{yp} = \frac{\sum_{i=1}^{n_y} x_{ip}}{n_y} \quad (2.11)$$

Keterangan:

c_{yp} : nilai pusat *cluster* ke- y pada variabel numerik ke- p

x_{ip} : nilai objek data ke- i pada variabel numerik ke- p

n_y : jumlah objek data pada *cluster* ke- y Dengan menggunakan nilai modus, seperti ditunjukkan dalam persamaan 2.12, nilai *centroid* baru untuk variabel kategorik dihitung.

$$c_{yp} = \text{modus} \{x_{1q}, x_{2q}, \dots, x_{iq}\} \quad (2.12)$$

Keterangan :

c_{yp} : nilai pusat *cluster* ke- y pada variabel kategorik ke- q

x_{iq} : nilai objek data ke- I pada variabel kategorik ke- q

6. Mendistribusikan ulang item ke setiap kluster sesuai dengan jarak yang paling dekat dengan nilai *centroid* yang diperbarui. Proses algoritme berakhir jika *centroid* atau titik pusat kluster telah konvergen atau tidak lagi berubah. Prosedur akan diulang dari tahap 3 jika *centroid* terus berfluktuasi, hingga jumlah iterasi maksimum tercapai atau item kluster berhenti bergerak.

2.9 Variabel yang Digunakan

Variabel yang digunakan ada lima antara lain:

1. Variabel pertama ialah usia awal merokok mahasiswa laki-laki.
2. Variabel kedua ialah Jumlah Rokok (Bungkus). Perokok cenderung lebih mungkin berhenti merokok di kemudian hari jika mereka merokok lebih sedikit.
3. Variabel ketiga ialah program studi. Program studi yang ada di FST sebanyak 5 prodi antara lain: ilmu komputer, sistem informasi, matematika, biologi, dan fisika.
4. Variabel keempat ialah jenis rokok. Jenis rokok yang dipilih ada tiga antara lain: malboro, sampoerna a mild dan gudang garam.
5. Variabel keenam ialah faktor merokok. Faktor yang menyebabkan mahasiswa FST merokok karena keluarga, teman dan media massa (iklan rokok).

2.10 Penelitian Terdahulu

Penelitian yang dilakukan oleh penulis menggunakan studi pustaka sebagai landasan teori. Peneliti mengambil beberapa penelitian dari referensi jurnal. Berikut ini adalah beberapa penelitian yang terkait dengan penelitian yang dilakukan oleh peneliti:

1. Penelitian Ganmanah & Kudus (2021) yang berjudul “Penerapan Algoritme *K-Prototypes* untuk Pengelompokan Desa-Desa di Provinsi Jawa Barat Berdasarkan Indikator Indeks Desa Membangun Tahun 2020”. Hasil penelitian yang diperoleh adalah pengelompokan desa-desa di Provinsi Jawa Barat berdasarkan indikator indeks desa membangun menggunakan Algoritme *K-Prototypes* menghasilkan 5 *cluster*. Dimana penentuan hasil *cluster* tersebut diperoleh dari evaluasi hasil *cluster* yang optimal berdasarkan nilai rasio keragaman di dalam *cluster* terhadap keragaman antar *cluster* terkecil yaitu sebesar 0,795.
2. Penelitian Reihanah et al., (2024) yang berjudul “*Clusterinnng* Karakteristik Industri Kecil dan Menengah di Kota Kendari Menggunakan Algoritma *K-Prototypes*”. Berdasarkan uraian hasil dari Pembahasan, maka diperoleh jumlah *cluster* terbaik yang dihasilkan adalah lima cluster ($k=5$) yang mempunyai nilai *Silhouette Index* terbesar yaitu 0,461. Jumlah unit IKM masing-masing *cluster* 2; 132 unit *cluster* 3; 14 unit IKM pada *cluster* 4; dan 18 unit IKM pada *cluster* 5. Dari lima *cluster* IKM yang terbentuk, *cluster* 2 merupakan *cluster* dengan rata-rata nilai investasi terendah dan rata-rata umur IKM pada *cluster* lebih muda apabila dibandingkan umur rata-rata IKM pada *cluster* lain sehingga *cluster* 2 butuh mendapat perhatian khusus dalam pembinaan dan pengembangan kemampuan IKM agar daya saing IKM meningkat.
3. Penelitian Koni et al., (2023) yang berjudul “Pengelompokan Kabupaten/ Kota Provinsi Jawa Timur Berdasarkan Indikator Kesejahteraan Rakyat Menggunakan Metode *Elbow* dan Algoritma *K-Prototype*”. Hasil pengelompokan Kabupaten/ Kota berdasarkan indikator kesejahteraan rakyat dengan menggunakan Algoritma *K-prototype* membentuk tiga klaster. Penentuan klaster tersebut dilakukan dengan pengujian metode *Elbow*. Variabel yang lebih dominan dalam hal pengelompokan Kabupaten/Kota yaitu Jumlah Penduduk Miskin (X_1), Pengeluaran Perkapita (X_2), Tingkat Pengangguran Terbuka (X_3) dan juga sumber Air Untuk Minum Sebagian Besar Keluarga (X_7). Adapun hasil pengelompokan Kabupaten/ Kota berdasarkan indikator kesejahteraan rakyat yaitu 4 Kabupaten/Kota termasuk pada klaster satu yang terdiri dari Sidoarjo, Kota Malang, Kota Madiun, Kota Surabaya, 18 Kabupaten/Kota lainnya termasuk pada klaster dua yaitu

Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Malang, Lumajang, Jember, Bondowoso, Situbondo, Probolinggo, Pasuruan, Bojonegoro, Tuban, Bangkalan, Sampang, Pamekasan, Sumenep, dan 16 Kabupaten/Kota termasuk pada klaster 3 yang terdiri dari Kediri, Bayuwangi, Mojokerto, Jombang, Ngajuk, Madiun, Magetan, Ngawi.

4. Penelitian Sugiyo (2021) yang berjudul “Perilaku Merokok di Kalangan Mahasiswa dan Motivasi mereka untuk Berhenti”. Hasil penelitian yaitu usia perokok semula paling tinggi didapatkan pada rentang usia Sekolah Menengah Pertama, dengan alasan terpengaruh oleh teman sebaya.
5. Penelitian Sawitri et al., (2020) yang berjudul “Karakteristik Perilaku Merokok Mahasiswa Universitas Malikussaleh 2019”. Hasil penelitian yaitu sebagian besar responden berusia 20-24 tahun dengan sebagian besar konsumsi rokok 1-10 batang per hari serta lebih dari setengah responden mengkonsumsi rokok lebih dari 60 menit setelah bangun tidur. Faktor stres dan kesenangan merupakan faktor yang paling mempengaruhi perilaku merokok.