

Terakreditasi SINTA Peringkat 4

Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 28/E/KPT/2019
masa berlaku mulai Vol.3 No. 1 tahun 2018 s.d Vol. 7 No. 1 tahun 2022

Terbit online pada laman web jurnal:
<http://publishing-widyagama.ac.id/ejournal-v2/index.php/jointecs>



Vol. 5 No. 3 (2020) 195 - 202

JOINTECS

(Journal of Information Technology and Computer Science)

e-ISSN:2541-6448

p-ISSN:2541-3619

Klasterisasi Karakter Konsumen Terhadap Kecenderungan Pemilihan Produk Menggunakan *K-Means*

Ach. Syuhbanul Yaumi¹, Zainul Zulfiqar², Aryo Nugroho³

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Narotama

¹ach.syuhbanul@mhs.fasilkom.narotama.ac.id, ²zainul.16@fik.narotama.ac.id, ³aryo.nugroho@narotama.ac.id

Abstract

The problem that is in store right now is the difficulty to find out which products are currently in high demand or most used by consumers so that it can be known from each variable the characteristics of consumers are more likely to choose the preferred product. Therefore, in the research grouping with the K-Means method for product selection is one way to determine customer choices for the products consumed. Because the K-Means method has a more accurate division. This research conducted a data mining process to help stores to be able to know the characteristics of consumers towards product selection tendencies by extracting data using one of the data mining methods, namely grouping data. In this study, the grouping of data from questionnaires or questionnaires spread in stores, then the data is grouped into 2 groups using one of the clustering algorithms, K-Means. The data used are data collected as many as 366 store customer response data. After the data is processed using the K-Means algorithm, it shows that cluster 1 is a type A consumer group with a percentage of 33%, while cluster 2 is a type B consumer group with a percentage of 67%.

Keywords: *product; k-means; clustering; data mining; marketing.*

Abstrak

Permasalahan yang sedang ada di toko saat ini adalah kesulitan untuk mengetahui produk yang saat ini banyak diminati atau paling banyak digunakan oleh konsumen agar dapat diketahui dari masing-masing variabel karakteristik konsumen lebih cenderung memilih produk yang disukai. Oleh karena itu, dalam penelitian pengelompokan dengan metode *K-Means* untuk pemilihan produk merupakan salah satu cara untuk mengetahui pilihan pelanggan terhadap produk yang dikonsumsi. Karena metode *K-Means* mempunyai hasil pembagian yang lebih akurat. Penelitian ini melakukan proses *data mining* untuk membantu toko agar dapat mengetahui karakteristik konsumen terhadap kecenderungan pemilihan produk dengan melakukan penggalan data menggunakan salah satu metode data mining yaitu pengelompokan data. Pada penelitian ini melakukan pengelompokan dari data hasil penyebaran angket atau kuesioner yang disebar di toko, kemudian data tersebut di kelompokkan menjadi 2 kelompok menggunakan salah satu algoritma klasterisasi yaitu *K-Means*. Data yang digunakan merupakan data yang dikumpulkan sebanyak 366 data tanggapan pelanggan toko. Setelah data tersebut diproses menggunakan algoritma *K-Means*, menunjukkan bahwa *cluster 1* merupakan kelompok konsumen tipe A dengan presentase sebesar 33%, sedangkan *cluster 2* merupakan kelompok konsumen tipe B dengan presentase sebesar 67%.

Kata kunci: produk; k-means; klasterisasi; data mining; pemasaran.

© 2020 Jurnal JOINTECS

Diterima Redaksi : 30-07-2020 | Selesai Revisi : 27-08-2020 | Diterbitkan Online : 30-09-2020

1. Pendahuluan

Pada umumnya produk x tipe A biasanya terbuat dari bahan tembakau yang digulung oleh kertas. Namun seiring berkembangnya zaman terdapat jenis produk x baru yaitu tipe B. Banyak yang mengira bahwa produk x tipe B lebih aman dibandingkan dengan produk x tipe A. Beberapa alasan dari mereka yang beralih dari tipe A ke ke tipe B juga bermacam-macam, mulai dari aman bagi kesehatan, lebih ramah lingkungan karena menghasilkan uap air bukan asap, mengurangi resiko dari bahaya penyakit, mengurangi kecanduan, dan yang paling utama sebagai *alternative* pengganti produk tipe A untuk perlahan-lahan berhenti kecanduan.

Data mining merupakan suatu istilah yang berfungsi untuk menguraikan penemuan pengetahuan pada database [1]. Teknik data mining juga sering disebut sebagai proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi dan mengekstraksi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [2]. Klasterisasi adalah salah satu metode data mining yang sangat membantu dalam penemuan pengetahuan [3]. Klasterisasi berarti mengelompokkan objek data ke dalam sebuah kelompok yang memiliki paling sedikit kesamaan dengan anggota *cluster* lainnya [4].

Pada penelitian yang pernah dilakukan sebelumnya menggunakan metode *K-means* juga digunakan untuk menentukan kategori stok barang dengan studi kasus data transaksi penjualan dengan memanfaatkan tingkat akurasi yang paling tinggi pada dataset produk menggunakan *K-Means*. Dari hasil penerapan metode *K-Means* ini dapat dinilai lebih akurat dalam melakukan *clustering* pada penentuan stok produk berdasarkan hasil tingkat akurasi yang tinggi pada masing-masing *cluster* [5]. Penerapan metode *K-Means* banyak diterapkan dalam berbagai kasus selain kasus penjualan, salah satunya adalah pengelompokan untuk pelanggan PDAM [6]. Metode *K-Means* merupakan metode dimana dilakukan pengelompokan data (klasterisasi) dengan sistem partisi dan pemodelannya tanpa supervise [7].

Pada penelitian lainnya metode *K-Means* juga dinilai baik dibandingkan metode *K-Medoids* dalam melakukan *clustering* terhadap kelayakan puskesmas di DIY berdasarkan nilai rasio simpangan baku yang lebih kecil yaitu sebesar 54,69% [8]. Penelitian terdahulu lainnya yang berhubungan dengan penjualan menggunakan metode *K-Means* dilakukan untuk memenuhi jumlah barang yang dibutuhkan dalam pemasaran produk aki dengan proses klasterasi didapat hasil akurasi sebesar 81% [9]. Penelitian terkait lainnya tentang klasterisasi yang membandingkan metode *K-Means* dan metode DBSCAN pada pengeleompokan rumah kost mahasiswa di kelurahan tembalang semarang didapat hasil perhitungan indeks *silhouette* menunjukkan

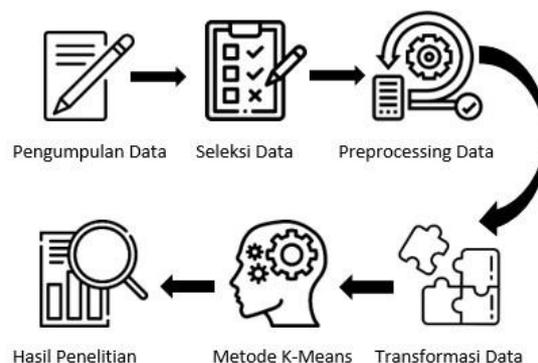
metode *K-Means* menghasilkan nilai 0,463 dan metode DBSCAN menghasilkan nilai 0,281. Sehingga dapat disimpulkan bahwa metode *K-Means* lebih baik dari metode DBSCAN dalam pengelompokan data rumah kost [10].

Permasalahan yang sedang ada di toko saat ini adalah kesulitan mengetahui produk x yang saat ini banyak diminati atau paling banyak digunakan oleh konsumen agar dari hasil penelitian ini toko dapat diketahui dari masing-masing variabel karakteristik konsumen lebih cenderung memilih produk yang disukai. Toko ini membutuhkan suatu informasi penting yang nantinya bisa bermanfaat untuk penjualan agar dapat diketahui produk x mana yang lebih cenderung dipilih oleh pelanggan. Metode ini digunakan karena mampu menghasilkan kriteria yang telah ditetapkan berdasarkan seleksi keputusan terbaik dari sejumlah keputusan [11].

Oleh karena itu dalam penelitian klasterisasi karakter konsumen dengan *K-Means* untuk pemasaran produk x merupakan salah satu cara untuk mengetahui selera atau pilihan pelanggan di toko ini terhadap produk x yang dikonsumsi. Karena metode *K-Means* mempunyai hasil pembagian yang lebih akurat [12]. Keakuratan metode *K-Means* tergantung pada pemilihan data centroid selama proses klasterisasi [13]. Tujuan dari analisis klasterisasi pemasaran produk x adalah untuk merancang strategi dalam penjualan produk x di toko ini pada masa yang akan mendatang berdasarkan dua *cluster* yaitu *cluster* 1 adalah produk x tipe A sedangkan *cluster* 2 adalah produk x tipe B.

2. Metode Penelitian

Dalam sebuah penelitian metode berfungsi sebagai cara untuk memecahkan masalah dan menggambarkan tahapan-tahapan yang akan dilakukan. pada penelitian ini, untuk menganalisis dan menentukan tipe produk x yang digunakan. Langkah-langkah pada penelitian ini ada beberapa langkah, yaitu: pengumpulan data, seleksi data, preprocessing data, transformasi data, metode *K-Means* dan hasil penelitian. Berikut gambaran tahapan-tahapan yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Pengumpulan data berarti mengumpulkan data yang dibutuhkan untuk analisis [14]. Pada tahapan ini bertujuan untuk mencari atau memperoleh informasi bagi pengguna produk x sebagai bahan penelitian yang akan di klusterisasi. Untuk selanjutnya data diproses dan dihitung menggunakan metode *K-Means* yang di dapat dari penyebaran angket kusioner melalui google form atau *hard copy* kepada pembeli atau hanya sekedar bersantai di toko.

2.2. Seleksi Data

Fungsi dari fitur seleksi digunakan untuk meringankan beban pemrosesan dalam data mining [15]. Tujuannya agar mendapatkan hasil data mining yang spesifik dan akurat, hanya beberapa data yang diambil dari data mentah dan akan digunakan sebagai atribut untuk selanjutnya diproses. Ada 7 atribut yang digunakan yaitu status, usia, mulai mengkonsumsi, nama produk, alasan memilih produk, pemakain perhari dan pengeluaran perbulan. Data yang dikumpulkan sebanyak 366 data responden yang merupakan pelanggan di toko.

2.3. Preprocessing Data

Data mentah yang telah didapat dari hasil penyebaran angket atau kusioner masih perlu diproses lagi untuk memperbaiki data yang belum sempurna dan data yang masih kurang atau belum lengkap. Tujuan dari *preprocessing* data yaitu pembersihan atau pengurangan atribut data yang tidak diperlukan serta kolom variabel yang tidak memiliki nilai. Selain itu *preprocessing* data juga dapat meningkatkan kualitas data dan mengurangi ukuran file log web [16].

Tabel 1. Data yang digunakan

| S | U | MM | NP | AM | PP | PB |
|---|----|----|----|----|----|----|
| 3 | 22 | 17 | 2 | 1 | 1 | 1 |
| 1 | 17 | 13 | 3 | 2 | 1 | 1 |
| 3 | 21 | 19 | 4 | 3 | 1 | 1 |
| 2 | 21 | 19 | 4 | 3 | 1 | 1 |
| 2 | 19 | 15 | 4 | 1 | 2 | 2 |
| 3 | 22 | 16 | 1 | 1 | 2 | 2 |
| 2 | 20 | 15 | 5 | 4 | 1 | 1 |
| 1 | 17 | 16 | 3 | 2 | 1 | 1 |
| 3 | 24 | 19 | 1 | 4 | 2 | 1 |
| 3 | 21 | 13 | 10 | 3 | 1 | 1 |
| 2 | 20 | 19 | 14 | 2 | 5 | 1 |
| 3 | 29 | 20 | 18 | 3 | 5 | 2 |
| 2 | 22 | 17 | 13 | 8 | 4 | 1 |
| 2 | 19 | 16 | 19 | 7 | 4 | 2 |
| 3 | 30 | 27 | 13 | 5 | 5 | 1 |
| 2 | 19 | 18 | 14 | 7 | 4 | 1 |

Pada Tabel 1 merupakan data yang digunakan untuk tahap pengujian data dan perhitungan iterasi. Data ini memiliki 7 atribut diantaranya status, usia, mulai mengkonsumsi, nama produk, alasan memilih, pemakain perhari dan pengeluaran perbulan. Data yang kurang lengkap dan tidak relevan hanya akan membuat

proses perhitungan mengalami *error*, oleh karena itu diperlukan *preprocessing* data yang akan digunakan untuk kebutuhan data mining.

2.4. Transformasi Data

Transformasi data merupakan upaya yang dilakukan untuk mengubah skala pengukuran data asli menjadi bentuk yang lain. Proses ini bertujuan untuk mengubah data yang dapat diolah dengan algoritma yang akan digunakan dalam bentuk klasifikasi atau *numerik* [17]. Data yang diperoleh dari hasil penyebaran angket atau kusioner belum sesuai format untuk digunakan pemrosesan data. Pemrosesan data diperlukan input data yang sudah dikategorikan dan tipe data numerik, maka jenis data nominal seperti status, usia, mulai mengkonsumsi, nama produk, alasan memilih produk, pemakain perhari, dan pengeluaran perbulan terlebih dahulu diinisialisasikan dalam bentuk angka.

Tabel 2. Pra-proses Data Status

| Tipe Status | Tranformasi |
|-------------|-------------|
| Pelajar | 1 |
| Mahasiswa | 2 |
| Pekerja | 3 |

Pada Tabel 2 merupakan hasil dari proses *transformasi* pada variabel tipe status pengguna dengan perubahan masing-masing *attribute* yang telah diisialisasikan dari angka 1 sampai angka 3. Angka 1 merupakan tipe status pelajar. Angka 2 ialah tipe status mahasiswa. Sedangkan angka 3 merupakan tipe status pekerja.

Tabel 3. Pra-proses Nama Produk

| Nama Produk | Tranformasi |
|---------------------|-------------|
| Surya, Gudang Garam | 1 |
| U Mild | 2 |
| Mild | 3 |
| Malboro | 4 |
| Dunhill | 5 |
| Djarum | 6 |
| Sukun | 7 |
| Marhsband | 8 |
| Classmild | 9 |
| Sampoerna, Samsu | 10 |
| Pro Mild | 11 |
| Tesla | 12 |
| Pod, Smoke | 13 |
| Mod, Hxhom | 14 |
| Vgod | 15 |
| Aegis | 16 |
| Invander 3 | 17 |
| Mekanika | 18 |
| Geekvape | 19 |
| Lswtch | 20 |
| Uwel Caliburn | 21 |
| Battlestar | 22 |

Pada Tabel 3 menunjukkan hasil dari proses transformasi pada variabel nama produk dengan perubahan masing-masing pada attribute yang diinisialisasikan dari angka 1 sampai dengan angka 22. Pada tahap perubahan variabel nama produk merupakan bagian yang penting. Karena pada tahap ini diharapkan mempermudah proses pembelajaran data machine learning agar dapat mengelompokkan data sesuai cluster yang telah ditentukan.

Tabel 4. Pra-proses Alasan Memilih

| Alasan | Tranformasi |
|-----------------|-------------|
| Enak, Suka | 1 |
| Hemat | 2 |
| Simpel, Praktis | 3 |
| Nikmat, Berasa | 4 |
| Aman, Sehat | 5 |
| Lingkungan | 6 |
| Banyak Rasa | 7 |
| Mengurangi | 8 |
| Merokok | |
| Kekinian, Gaya | 9 |

Pada Tabel 4 diatas setelah proses transformasi data didapat hasil perubahan pada variable alasan memilih produk. Terdapat berbagai alasan yang didapat dari hasil kusioner mulai dari alasan suka sampai dengan mengikuti gaya hidup. Dengan masing-masing attribute yang sudah diinisialisasikan dari angka 1 sampai dengan angka 9.

Tabel 5. Pra-proses Pemakaian Perhari

| Pemakaian Perhari | Tranformasi |
|-------------------|-------------|
| 1-10 Batang | 1 |
| 10-20 Batang | 2 |
| 20-30 Batang | 3 |
| 1-10 ml | 4 |
| 10-20 ml | 5 |
| 20-30 ml | 6 |

Pada Tabel 5 hasil proses transformasi variabel pemakaian perhari dengan perubahan attribute yang sudah diinisialisasikan dari angka 1 sampai dengan angka 6. Untuk pemakaian pada jenis batang diubah dari angka 1 sampai angka 3. Sedangkan untuk pemakaian jenis ml diubah dari angka 4 sampai angka 6.

Tabel 6. Pra-proses Pengeluaran Perbulan

| Kisaran Nominal | Tranformasi |
|-----------------|-------------|
| <500 ribu | 1 |
| >500 ribu | 2 |

Pada Tabel 6 diatas merupakan hasil transformasi variabel pengeluaran perbulan dengan masing-masing attribute yang telah diinisialisasikan dengan angka 1 dan 2. Angka 1 merupakan pengeluaran perbulan dibawah

500 ribu. Sedangkan angka 2 untuk pengeluaran perbulan diatas 500 ribu.

2.5. Perhitungan data dengan K-Means

Metode K-Means merupakan salah satu metode clustering yang paling populer [18]. Tujuan dari metode ini adalah untuk membagi data menjadi beberapa kelompok [19]. Metode K-Means ini akan menerima masukan berupa data tanpa label kelas. Pada penelitian ini perhitungan K-Means menggunakan tools cluster Data yang dipilih dari data mentah melalui proses seleksi data, processing data dan tranformasi data akan dimasukkan lalu membaca data training dan diproses, kemudian menilai hasil dari proses tersebut. Untuk menghitung iterasi jarak terdekat antar cluster menggunakan rumus Euclidean Distance. Dengan menentukan nilai centroid awal (nilai tengah) kemudian partisi data menggunakan nilai minimum. Lakukan iterasi sampai nilai atau posisi data tidak mengalami perubahan. Jika data sudah tidak berubah lagi maka hentikan iterasi. Maka hasil dari perhitungan iterasi merupakan centroid akhir dari data yang sudah dipartisi sebelumnya. Berikut adalah flowchart diagram metode K-Means.



Gambar 2. Flowchart K-Means

Pada Gambar 2 dapat dijelaskan bahwa sebelum memulai perhitungan K-Means terlebih dahulu menentukan nilai *cluster* lalu menentukan nilai centroid atau nilai titik tengah pada klaster. Setelah menentukan nilai centroid selanjutnya adalah menghitung jarak objek ke centroid dengan rumus *Euclidean Distance* kemudian hasilnya dikelompokkan objek ke dalam *cluster* berdasarkan jarak terdekat. Jika hasil nilai centroid baru tidak sama dengan centroid sebelumnya, maka proses perhitungan tetap dilanjutkan sampai nilai centroid baru yang dihitung sama dengan centroid sebelumnya dan didapatkan hasil nilai centroid akhir dapat dinyatakan proses perhitungan *K-Means* telah selesai.

Langkah-langkah klasterisasi dengan metode *K-Means* dapat dijelaskan sebagai berikut:

Program Jurnal

Input: K (jumlah cluster), $X_i, i= 1, 2, \dots, N\}$. (training set)

Initialization $C_k, k= 1, 2, \dots, K\}$ (pilih K secara acak)

Repeat: {

a. langkah cluster

Untuk setiap contoh $\{X_i\}$

Hitung jarak $K // |X_i - C_k|, k= 1, 2, \dots, K.$

Atur label kelas $c_i :=$ indeks k (dari 1 ke K) dari centroid cluster yang memiliki jarak minimum ($c_i \in \{1, 2, \dots, K\}$).

b. langkah centroid

for $k= 1$ to K (untuk setiap kelas)

$C_k :=$ mean yang ditugaskan ke cluster k

}

end

2.6. Hasil Penelitian

Pada tahap ini bertujuan untuk memvisualisasikan hasil proses data mining berupa tabel atau gambar serta menganalisa dan mengevaluasi agar mudah dipahami oleh pembaca. Dengan hasil klasterisasi data akan dikelompokkan sesuai kelas yang telah dihitung. Akurasi merupakan salah satu indikator dalam menilai sebuah algoritma [20]. Dalam menentukan tingkat akurasi perlu diketahui nilai sebenarnya dari besaran yang diukur dan dapat diketahui seberapa besar tingkat akurasinya.

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (1)$$

Keterangan untuk rumus 1 akan dijelaskan pada paragraf ini. Nilai TP (True Positif) merupakan data benar yang terklasifikasikan menjadi data benar, nilai FP (False Positif) yaitu data benar yang terklasifikasikan benar. Sedangkan nilai FN (False Negatif) adalah data salah yang tidak di klasifikasikan salah, dan nilai TN (True Negatif) merupakan data yang tidak di klasifikasikan benar[20].

3. Hasil dan Pembahasan

Pada penelitian ini proses data mining menggunakan *tools cluster*, tools tersebut berfungsi untuk membantu proses pengelompokan data dengan menggunakan salah satu metode algoritma *K-Means* dari *cluster* itu sendiri. Sampel data yang akan digunakan dalam penelitian ini sebanyak 366 data yang sudah ditata sesuai atribut yang diperlukan untuk proses data mining, setelah itu akan dikelompokkan menjadi 2 *cluster* yaitu pengguna produk x tipe A atau produk x tipe B sesuai atribut-atribut yang telah didapat sebelumnya. Berikut hasil dari proses data mining.

3.1. Hasil Pengujian

Pada tahap hasil pengujian ini dilakukan langkah awal dengan *tools cluster* untuk *training* data menggunakan fitur *Use Data Training Set* dengan keseluruhan 366 data. Dengan memasukkan data yang telah diseleksi berdasarkan 7 atribut yaitu status, usia, mulai mengkonsumsi, nama produk, alasan memilih produk, pemakaian perhari, dan pengeluaran perbulan. Hasil dari proses *training* dapat dilihat pada Tabel 7.

Tabel 7. Hasil Use Data Training Set

| Attribute | Full Data (366) | Cluster 0 (122) | Cluster 1 (244) |
|--------------------|--------------------|--------------------|--------------------|
| Status | 2.224 | 2.2213 | 2.2254 |
| Usia | 22.4863 | 22.5246 | 22.4672 |
| Mulai Mengkonsumsi | 17.3497 | 15.7131 | 18.168 |
| Nama Produk | 11.9481 | 4.5 | 15.6721 |
| Alasan | 4.2814 | 2.8361 | 5.0041 |
| Pemakaian Perhari | 3.3169 | 1.1885 | 4.3811 |
| Perbulan | 1.5055 | 1.6311 | 1.4426 |

Pada Tabel 7 menampilkan hasil dari *Use Data Training Set* proses pengetesan data ini dilakukan dengan menggunakan seluruh data *training* itu sendiri. Dari proses sebanyak 366 data diperoleh hasil nilai cluster 0 sebanyak 33% dengan jumlah data sebesar 122 data dan cluster 1 sebanyak 67% dengan jumlah data sebesar 244 data. Pada proses ini juga diperoleh hasil perhitungan iterasi sebanyak 3 kali dengan pemilihan centroid awal cluster 0 dengan nilai (3,23,17,6,2,1,2) dan cluster 1 dengan nilai (3,29,20,18,3,5,2).

Tabel 8. Hasil Percentage Split 85%

| Attribute | Full Data (311) | Cluster 0 (159) | Cluster 1 (152) |
|--------------------|--------------------|--------------------|--------------------|
| Status | 2.2669 | 2.4088 | 2.1184 |
| Usia | 22.6785 | 23.956 | 21.3421 |
| Mulai Mengkonsumsi | 17.4695 | 17.2264 | 17.7237 |
| Nama Produk | 12.0579 | 11.2453 | 12.9079 |
| Alasan | 4.2926 | 4.3711 | 4.2105 |
| Pemakaian Perhari | 3.3312 | 3.1887 | 3.4803 |
| Perbulan | 1.5113 | 2 | 1 |

Tabel 8 adalah hasil dari proses *percentage split* dimana hasil klasterisasi dipes menggunakan nilai $k\%$ dari data yang telah diinput, nilai k adalah nilai masukan yang ditentukan oleh *user*. Dalam proses uji tes ini menggunakan nilai k sebesar 85%. Artinya nilai 85% datanya akan menjadi data *training set* sedangkan sisanya yaitu 15% menjadi *test set*. Dari proses sebanyak 311 data didapatkan hasil *cluster 0* sebanyak 47% dengan jumlah data sebesar 159 data dan *cluster 1* sebanyak 53% dengan jumlah data sebesar 152 data dengan hasil perhitungan iterasi sebanyak 3 kali.

Sedangkan proses perhitungan klasterisasi secara *manual* dengan rumus *Euclidean Distance* diawali dengan menentukan nilai *cluster*. Jumlah *cluster* yang akan digunakan pada penelitian ini adalah 2 buah *cluster*, sehingga nilai $k=2$ yaitu *cluster 0* dan *cluster 1*. Kemudian tentukan nilai titik tengah *cluster* (centroid). Pada tahap ini nilai centroid dapat dipilih secara acak dari *dataset* yang akan digunakan. Hasil penentuan nilai centroid awal dapat dilihat pada Tabel 9.

Tabel 9. Centroid Awal

| Cluster 0 | Cluster 1 |
|-----------|-----------|
| 1 | 1 |
| 2 | 2 |
| 21 | 19 |
| 19 | 18 |
| 4 | 14 |
| 3 | 7 |
| 1 | 4 |
| 1 | 1 |

Pada Tabel 9 Nilai centroid awal merupakan nilai yang digunakan untuk memulai perhitungan iterasi. Nilai ini ditentukan secara acak, diambil dari data mentah yang telah diseleksi dan diubah menjadi data numerik. Pada proses ini juga ditentukan pemilihan centroid awal cluster 0 kategori konsumen produk x tipe A dengan nilai (1,2,21,19,4,3,1,1) dan cluster 1 kategori konsumen produk x tipe B dengan nilai (1,2,19,18,14,7,4,1). Untuk proses selanjutnya dilakukan pengulangan ke- 1 sampai selesai. Jika centroid baru tidak sama dengan centroid sebelumnya, maka proses pengulangan tetap dilanjutkan, tetapi jika centroid baru yang dihitung sama dengan centroid sebelumnya, maka proses klasterisasi telah selesai.

3.2. Hasil Perhitungan Iterasi

Pada tahap ini dilakukan perhitungan iterasi, langkah awal yaitu menentukan nilai centroid awal. Kemudian dilakukan perhitungan hingga nilai centroid baru tidak sama dengan centroid sebelumnya. maka proses perhitungan tetap dilanjutkan sampai nilai centroid baru yang dihitung sama dengan centroid sebelumnya dan didapatkan hasil nilai centroid akhir. Selanjutnya nilai yang dihasilkan dapat dikelompokkan menjadi sesuai jumlah *cluster* yang sudah ditentukan berdasarkan jarak minimum dari nilai centroid dengan nilai setiap objek.

Jumlah pengulangan iterasi untuk mencapai cluster centroid akan dipengaruhi oleh calon *cluster* centroid awal secara acak dimana jika posisi centroid baru tidak berubah. Setelah proses perhitungan iterasi selesai maka akan didapatkan nilai minimal jarak terdekat dengan mengambil nilai terkecil dari hasil iterasi. Hasil proses perhitungan iterasi dapat dilihat pada Tabel 10, Tabel 11, Tabel 12, dan Tabel 13.

Tabel 10. Hasil Iterasi ke-1

| C0 | C1 | Minimal | Cluster |
|-------|-------|---------|---------|
| 3.74 | 14.14 | 3.74 | 1 |
| 7.42 | 13.60 | 7.42 | 1 |
| 1.00 | 11.45 | 1.00 | 1 |
| 0.00 | 11.40 | 0.00 | 1 |
| 5.10 | 12.25 | 5.10 | 1 |
| 5.10 | 14.97 | 5.10 | 1 |
| 4.36 | 10.44 | 4.36 | 1 |
| 5.29 | 12.81 | 5.29 | 1 |
| 4.58 | 14.46 | 4.58 | 1 |
| 8.54 | 8.43 | 8.43 | 2 |
| 10.86 | 5.29 | 5.29 | 2 |
| 16.70 | 11.79 | 11.79 | 2 |
| 10.95 | 3.46 | 3.46 | 2 |
| 16.25 | 5.48 | 5.48 | 2 |
| 15.72 | 14.46 | 14.46 | 2 |

Tabel 11. Hasil Iterasi ke-2

| C0 | C1 | Minimal | Cluster |
|-------|-------|---------|---------|
| 2.55 | 13.57 | 2.55 | 1 |
| 5.05 | 15.00 | 5.05 | 1 |
| 2.94 | 11.35 | 2.94 | 1 |
| 2.84 | 11.35 | 2.84 | 1 |
| 2.84 | 12.73 | 2.84 | 1 |
| 3.25 | 14.48 | 3.25 | 1 |
| 3.09 | 11.22 | 3.09 | 1 |
| 3.63 | 14.15 | 3.63 | 1 |
| 5.23 | 13.69 | 5.23 | 1 |
| 7.96 | 8.41 | 7.96 | 1 |
| 11.86 | 4.59 | 4.59 | 2 |
| 18.09 | 6.98 | 6.98 | 2 |
| 11.93 | 4.33 | 4.33 | 2 |
| 16.96 | 7.33 | 7.33 | 2 |
| 17.99 | 10.74 | 10.74 | 2 |

Tabel 12. Hasil Iterasi ke-3

| C0 | C1 | Minimal | Cluster |
|-------|-------|---------|---------|
| 2.94 | 14.86 | 2.94 | 1 |
| 4.93 | 16.53 | 4.93 | 1 |
| 3.04 | 12.53 | 3.04 | 1 |
| 2.98 | 12.53 | 2.98 | 1 |
| 2.58 | 14.19 | 2.58 | 1 |
| 3.67 | 15.79 | 3.67 | 1 |
| 2.46 | 12.71 | 2.46 | 1 |
| 3.75 | 15.53 | 3.75 | 1 |
| 5.63 | 14.71 | 5.63 | 1 |
| 7.16 | 10.09 | 7.16 | 1 |
| 11.32 | 5.30 | 5.30 | 2 |
| 17.55 | 6.06 | 6.06 | 2 |
| 11.33 | 5.22 | 5.22 | 2 |
| 16.29 | 7.57 | 7.57 | 2 |
| 17.78 | 9.71 | 9.71 | 2 |

Tabel 13. Hasil Iterasi ke-4

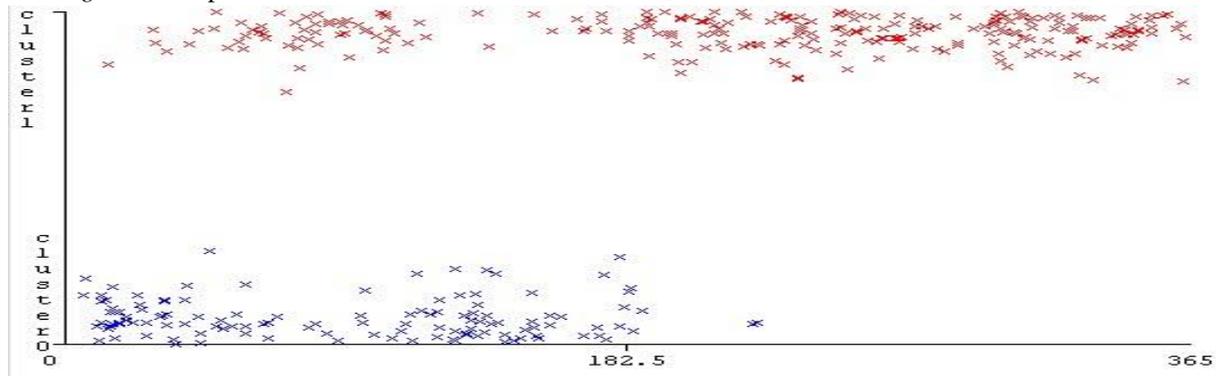
| C0 | C1 | Minimal | Cluster |
|-------|-------|---------|---------|
| 2.94 | 14.86 | 2.94 | 1 |
| 4.93 | 16.53 | 4.93 | 1 |
| 3.04 | 12.53 | 3.04 | 1 |
| 2.98 | 12.53 | 2.98 | 1 |
| 2.58 | 14.19 | 2.58 | 1 |
| 3.67 | 15.79 | 3.67 | 1 |
| 2.46 | 12.71 | 2.46 | 1 |
| 3.75 | 15.53 | 3.75 | 1 |
| 5.63 | 14.71 | 5.63 | 1 |
| 7.16 | 10.09 | 7.16 | 1 |
| 11.32 | 5.30 | 5.30 | 2 |
| 17.55 | 6.06 | 6.06 | 2 |
| 11.33 | 5.22 | 5.22 | 2 |
| 16.29 | 7.57 | 7.57 | 2 |
| 17.78 | 9.71 | 9.71 | 2 |

Setelah dilakukan perhitungan iterasi sebanyak 4 kali dapat dilihat pada Tabel 10, Tabel 11, Tabel 12 dan Tabel 13. Dari perhitungan menggunakan rumus *Euclidean Distance* didapatkan hasil nilai centroid akhir pada iterasi ke-4 dimana nilai centroid tidak berubah dari hasil perhitungan sebelumnya. Di dapatkan hasil nilai centroid akhir dari perhitungan iterasi dapat dilihat sebagai berikut.

Tabel 14. Centroid Akhir

| Cluster 0 | Cluster 1 |
|-----------|-----------|
| 1 | 1 |
| 2.3 | 2.4 |
| 20.4 | 24 |
| 16.2 | 19.8 |
| 3.7 | 15.4 |
| 2.4 | 5 |
| 1.3 | 4.6 |
| 1.2 | 1.4 |

Karena hasil dari perhitungan iterasi yang ke-4 pada Tabel 14 nilai centroid tidak berubah atau sama dengan nilai centroid yang sebelumnya maka proses klasterisasi telah selesai. Hasil klasterisasi dari *tools cluster* melakukan iterasi sebanyak 3 kali dengan dengan keseleruhan data sebanyak 366 data dengan fitur *use training set* mendapat hasil *visualisasi* Gambar 3 berikut.



Gambar 3. Grafik Hasil Visual Klasterisasi

Dapat dilihat pada Gambar 3 hasil dari perluasan data *cluster* berdasarkan anggota yang telah dihitung sesuai jarak terdekat. Pada sumbu Y menampilkan jumlah *cluster* dan pada sumbu X menampilkan total data yang diklaster. *cluster* 0 berwarna biru merupakan kelompok data yang memiliki jumlah data sebanyak 122 data dengan kategori produk x tipe A diperoleh jarak terdekat dengan nilai 4.5, sedangkan. *cluster* 1 berwarna merah merupakan kelompok data yang memiliki jumlah data sebanyak 244 data dengan kategori produk x tipe B diperoleh jarak terdekat dengan nilai 15.6721 dari atribut nama produk yang sudah diinisialisasikan.

Tabel 15. Hasil Cluster

| Cluster | Jumlah | Persentase |
|---------|--------|------------|
| 0 | 122 | 33% |
| 1 | 244 | 67% |

Pada Tabel 15 menunjukkan jumlah data responden yang sudah dibagi dalam 2 kelompok yaitu *cluster* 0 kategori konsumen produk x tipe A dan *cluster* 1 kategori konsumen produk x tipe B, kemudian persentase jumlah data responden dapat dilihat di setiap *cluster*. Dari hasil klasterisasi pada *cluster* 0 mendapatkan persentase sebesar 33%. Sedangkan pada *cluster* 1 mendapatkan persentase sebesar 67%. Presentase *cluster* 1 lebih besar daripada *cluster* 0 karena dari hasil klasterisasi jumlah konsumen produk x tipe B lebih unggul sebanyak 244 data sedangkan jumlah konsumen produk x tipe A hanya 122 data.

Hasil dari klasterisasi tersebut jika dalam keadaan normal diluar adanya keadaan yang tidak diharapkan seperti terjadinya pandemi. Kiranya klasterisasi juga dapat meleset jika terjadi pandemi, karena data tersebut diambil diluar keadaan pandemi. Karena pada saat kondisi pandemi toko ditutup demi keamanan bersama dan akan dibuka kembali setelah adanya pemberitahuan dari pemerintah daerah setempat dengan menerapkan aturan sesuai dengan protokol kesehatan.

4. Kesimpulan

Berdasarkan hasil dari penelitian, metode *K-Means* dapat digunakan untuk melakukan proses klasterisasi dengan mengubah data menjadi 2 kelompok kemudian

kelompok tersebut dapat digunakan untuk mengetahui jenis produk x yang banyak digunakan oleh konsumen. Dari hasil klasterisasi, toko tersebut dapat melakukan evaluasi pada karakteristik konsumen terhadap kecenderungan pemilihan produk x berdasarkan kelompok produk x seperti pembahasan dan penjelasan hasil diatas yaitu presentase produk x *cluster* 0 di dapat sebesar 33%, sedangkan presentase produk x *cluster* 1 di dapat sebesar 67%. Jadi dapat disimpulkan bahwa pengguna produk x tipe B lebih banyak daripada pengguna produk x tipe A dilihat dari jumlah data hasil klasterisasi konsumen produk x tipe B lebih unggul sebanyak 244 data sedangkan konsumen produk x tipe A hanya 122 data. Diharapkan penelitian selanjutnya dapat melakukan perbandingan menggunakan metode klasterisasi yang lain sebagai pembanding tingkat akurasi yang lebih akurat dan bisa bermanfaat bagi pihak yang memerlukan.

Daftar Pustaka

- [1] R. R. Agustin, M. Akbar, and S. Sa'uda, "Penerapan Data Mining Untuk Pengelompokan Minat Konsumen Terhadap Produk Lg Mobile Palembang," *Univ. Bina Darma*, pp. 1–11, 2015.
- [2] A. Sani, "Penerapan metode k-means clustering pada perusahaan," *J. Ilm. Teknol. Inf.*, no. 353, pp. 1–7, 2018.
- [3] C. Slamet, A. Rahman, M. A. Ramdhani, and W. Darmalaksana, "Clustering the Verses of the Holy Qur'an using K-Means Algorithm," *Asian J. Inf. Technol.*, vol. 15, no. 24, pp. 5159–5162, 2016.
- [4] R. Molaei and S. Alizadeh, "Checking and Analysing Customers Buying Behavior with Clustering Algorithm," *Palma J.*, vol. 16, no. 2, pp. 486–492, 2017.
- [5] E. Muningsih and S. Kiswati, "Penerapan Metode K-Means untuk Clustering Produk Online Shop dalam Penentuan Stok Barang," *J. Bianglala Inform.*, vol. 3, no. 1, pp. 10–17, 2015.
- [6] S. T. Siska, "Analisa Dan Penerapan Data Mining Untuk Menentukan Kubikasi Air Terjual Berdasarkan Pengelompokan Pelanggan Menggunakan Algoritma K-Means Clustering," *J. Teknol. Inf. Pendidik.*, vol. 9, no. 1, pp. 86–93, 2016.
- [7] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA J. Inform.*, vol. 7, no. 2, pp. 109–113, 2019.
- [8] I. Kamila, U. Khairunnisa, and Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, pp. 119–125, 2019.
- [9] R. Andrean, S. Fendy, and A. Nugroho, "Klasterisasi Pengendalian Persediaan Aki Menggunakan Metode K-Means," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 4, no. 1, pp. 5–12, 2019.
- [10] S. Budiman, D. Safitri, and D. Ispriyanti, "Perbandingan Metode K-Means Dan Metode DBSCAN Pada Pengelompokan Rumah Kost Mahasiswa Di Kelurahan Tembalang Semarang," *J. GAUSSIAN*, vol. 5, no. 4, pp. 757–762, 2016.
- [11] R. A. Siregar, "Seleksi Penyerang Utama Menggunakan K-Means Clustering Dan Sistem Pendukung Keputusan Metode Topsis," *Technomedia J.*, vol. 2, no. 1, pp. 37–48, 2017.
- [12] S. Rustam, "Analisa Clustering Phising Dengan K-Means Dalam Meningkatkan Keamanan Komputer," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 175–181, 2018.
- [13] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, 2017.
- [14] D. Zhou, Z. Yan, Y. Fu, and Z. Yao, "A Survey on Network Data Collection," *Journal of Network and Computer Applications*, vol. 116, Academic Press, pp. 9–23, 15-Aug-2018.
- [15] S. Fong, R. Wong, and A. V. Vasilakos, "Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data," *IEEE Trans. Serv. Comput.*, vol. 9, no. 1, pp. 33–45, Jan. 2016.
- [16] J. Mehra and R. S. Thakur, "An Effective method for Web Log Preprocessing and Page Access Frequency using Web Usage Mining," *Int. J. Appl. Eng. Res.*, vol. 13, no. 2, pp. 1227–1232, 2018.
- [17] D. Triyansyah and D. Fitriana, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.*, vol. 8, no. 3, pp. 163–182, 2018.
- [18] M. Capo, A. Perez, and J. A. A. Lozano, "An efficient Split-Merge re-start for the K-means algorithm," *IEEE Trans. Knowl. Data Eng.*, pp. 1–1, 2020.
- [19] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, pp. 167–174, 2018.
- [20] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics J.*, vol. 5, no. 1, pp. 10–17, 2020.