

**SYSTEMIC: Information System and Informatics Journal**

ISSN: 2460-8092, 2548-6551 (e)

Vol 6 No 1 - Agustus 2020

**Sistem Penunjang Keputusan Pemasaran Produk X Menggunakan Metode K-Means**Ach. Syuhbanul Yaumi<sup>1</sup>, Zainul Zulfikar<sup>2</sup>, Aryo Nugroho<sup>3</sup><sup>1,2,3</sup> Universitas Narotama, Surabaya[ach.syuhbanul@mhs.fasilkom.narotama.ac.id](mailto:ach.syuhbanul@mhs.fasilkom.narotama.ac.id)<sup>1</sup>, [zainul.16@fik.narotama.ac.id](mailto:zainul.16@fik.narotama.ac.id)<sup>2</sup>, [aryo.nugroho@narotama.ac.id](mailto:aryo.nugroho@narotama.ac.id)<sup>3</sup>**Kata Kunci***Data mining, Klasterisasi, K-Means, Pemasaran, Produk X.***Abstrak**

Permasalahan yang sedang ada di toko saat ini adalah kesulitan untuk mengetahui produk X yang saat ini banyak diminati atau paling banyak digunakan oleh konsumen agar persediaan produk x dapat dipenuhi sesuai permintaan pelanggan dan tidak terjadi out of stock. Oleh karena itu, dalam penelitian pengelompokan dengan metode K-Means untuk pemasaran produk x merupakan salah satu cara untuk mengetahui pilihan pelanggan terhadap produk x yang dikonsumsi. Pada penelitian ini melakukan pengelompokan dari data hasil penyebaran angket atau kusioner 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 responden pelanggan toko. Setelah data tersebut diproses menggunakan salah satu metode data mining yaitu 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%.

**Keywords***Clustering, Data mining, Marketing, K-Means, Products X.***Abstract**

The problem that is currently in store is the difficulty to find out which product x is currently in high demand or is most widely used by consumers so that inventory of product x can be met according to customer demand and does not occur out of stock. Therefore, in the research grouping with the K-Means method for marketing product x is one way to determine customer choices for product x consumed. In this study grouping data from questionnaires or questionnaires that are distributed in stores, then the data are grouped into 2 groups using one of the clustering algorithms, K-Means. The data used are data collected by 366 respondents of store customers. After the data is processed using one of the data mining methods, the K-Means algorithm, 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%.

**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.

Beberapa diantara pengkonsumsi produk tipe B adalah usia remaja yang pada dasarnya merupakan masa-masa ingin mencari tahu dan mencoba hal-hal baru tanpa berpikir rasional. Dari hasil penyebaran kusioner di dapat berbagai rentang usia yang telah dibagi menjadi beberapa status dengan jumlah 49,5% dari kalangan mahasiswa, 39% dari kalangan pekerja, dan 11,5% dari kalangan pelajar.

Di era globalisasi, teknologi semakin berkembang bahkan perkembangan teknologi saat ini banyak membantu memudahkan pekerjaan manusia supaya lebih efisien dan efektif dalam berbagai bidang [1]. Klasterisasi adalah salah satu metode *data mining* yang sangat membantu dalam

penemuan pengetahuan [2]. Klusterisasi berarti mengelompokkan objek data ke dalam sebuah kelompok yang memiliki paling sedikit kesamaan dengan anggota *cluster* lainnya [3].

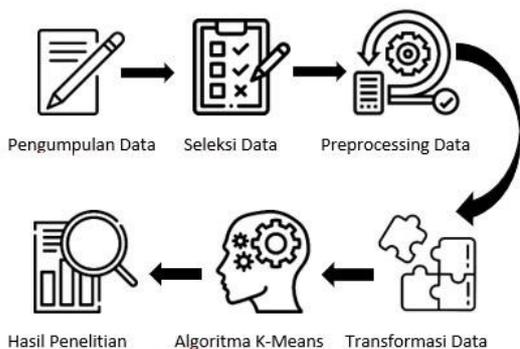
Permasalahan yang sedang ada di toko saat ini adalah kesulitan mengetahui produk x yang saat ini banyak diminati atau paling banyak digunakan oleh konsumen agar persediaan produk x dapat dipenuhi sesuai permintaan pelanggan dan tidak terjadi *out of stock*. Toko ini membutuhkan suatu sistem pendukung keputusan informasi penting yang nantinya bisa bermanfaat untuk penjualan agar dapat memenuhi persediaan produk x yang sesuai permintaan pelanggan. Metode ini digunakan karena mampu menghasilkan kriteria yang telah ditetapkan berdasarkan seleksi keputusan terbaik dari sejumlah keputusan [4].

Sistem Pendukung Keputusan merupakan suatu sistem informasi interaktif yang menyediakan pemodelan, manipulasi data dan informasi [5]. Dalam proses pengambilan keputusan pengetahuan dan pengalaman merupakan peran penting [6].

Oleh karena itu dalam penelitian Sistem Pendukung Keputusan pengelompokkan 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 [5]. Keakuratan metode K-Means tergantung pada pemilihan data centroid selama inisialisasi [7]. Tujuan dari analisis pengelompokkan 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. 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[8]. Pada

tahap ini bertujuan untuk mencari atau memperoleh informasi bagi pengguna produk x sebagai bahan penelitian yang akan di klusterisasi 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*[9]. 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, merk produk, alasan memilih produk x, 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 meningkatkan kualitas data dan mengurangi ukuran file log web[10]. Data yang kurang lengkap dan tidak relevan hanya akan membuat proses perhitungan pada weka mengalami *error*, oleh karena itu diperlukan *preprocessing* data yang akan digunakan untuk kebutuhan *data mining*.

Tabel 1 Data yang digunakan

No	JK	S	U	MM	MP	AM	PP	PB
1	1	3	22	17	2	1	1	1
2	1	1	17	13	3	2	1	1
3	1	3	21	19	4	3	1	1
4	1	2	21	19	4	3	1	1
5	1	2	19	15	4	1	2	2
6	1	3	22	16	1	1	2	2
7	1	2	20	15	5	4	1	1
8	1	1	17	16	3	2	1	1
9	1	3	24	19	1	4	2	1
10	1	3	21	13	10	3	1	1
11	1	2	26	20	1	2	3	1
12	1	1	17	12	1	2	1	1
13	1	1	18	16	6	4	1	1
14	1	2	18	15	1	3	1	1

Berikut keterangan dari variable Tabel 1:

- JK = Jenis Kelamin
- S = Status
- U = Usia
- MM = Mulai Mengonsumsi
- MP = Merk Produk
- AM = Alasan Memilih
- PP = Pengeluaran Perhari
- PB = Pengeluaran Bulanan

## 2.4 Transformasi Data

Proses ini bertujuan untuk mengubah data yang dapat diolah dengan algoritma yang akan digunakan dalam bentuk klasifikasi atau numerik[11]. 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 mengonsumsi, merk produk x, alasan memilih produk x, pemakain perhari, dan pengeluaran perbulan terlebih dahulu diinisialisasikan dalam bentuk angka.

Tabel 2 Pra-Proses data Status

Tipe Status	Transformasi
Pelajar	1
Mahasiswa	2
Pekerja	3

Tabel 3 Pra-Proses data Merk Produk

Merk Produk	Transformasi
Surya, Gudang Garam	1
U Mild	2
Mild	3
Malboro	4
Dunhill	5
Djarum	6
Sukun	7
Marshband	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

Tabel 4 Pra-Proses data Alasan Memilih

Alasan	Transformasi
Lebih Enak,Enak,Suka	1
Hemat,Lebih Hemat	2
Lebih Sempel,Praktis	3
Lebih Nikmat, Lebih Berasa	4
Lebih Aman,Lebih Sehat	5
Faktor Lingkungan	6
Karena Rasa,Banyak Rasa	7
Mengurangi Merokok	8
Gaya,Kekinian,Coba-Coba	9

Tabel 5 Pra-Proses data Pemakain Perhari

Pemakain Perhari	Transformasi
1-10 Batang	1
10-20 Batang	2
20-30 Batang	3
1-10 ml	4
10-20 ml	5
20-30 ml	6

Tabel 6 Pra-Proses Pengeluaran Perbulan

Kisaran Nominal	Transformasi
<500 ribu	1
>500 ribu	2

## 2.5 Perhitungan data dengan K-Means

Algoritma *K-Means* merupakan salah satu metode *clustering* yang paling populer[12]. Tujuan dari algoritma ini adalah untuk membagi data menjadi beberapa kelompok[13]. Algoritma *K-Means* ini akan menerima masukan berupa data tanpa label kelas. Pada penelitian ini perhitungan *K-Means* menggunakan aplikasi weka, Data yang dipilih dari data mentah melalui proses seleksi data, *processing* data dan *transformasi* data akan dimasukkan lalu membaca data *training* dan diproses, kemudian menilai hasil dari proses tersebut. Untuk menghitung iterasi jarak terdekat antar *cluster* menggunakan Persamaan (1) *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[14][15]. 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 algoritma *K-Means* :



Gambar 2 Flowchart Algoritma K-Means

Algoritma K-Means Klasterisasi dapat dijelaskan sebagai berikut :

1. Tentukan *k* buah *cluster*
2. Pilih sejumlah *k* buah objek secara acak yang akan dijadikan sebagai titik centroid *cluster*
3. Tentukan *k* buah centroid (titik tengah)
4. Kelompokkan objek ke centroid *cluster* terdekat berdasarkan Euclidean Distance. Berikut dibawah ini rumus *Euclidean Distance*:
- 5.

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (1)$$

Notasi Persamaan 1 dapat dijelaskan sebagai berikut :

- D = jarak antara x dan y
- x = data pada atribut
- y = nilai pusat
- j = setiap data
- n = jumlah data
- xj = nilai pusat ke j
- yj = objek data ke j

6. Hitung kembali semua nilai titik centroid Ulangi langkah 3 sampai 5 hingga nilai titik centroid tidak lagi berubah

### 2.6 Hasil Penelitian

Pada tahap terakhir ini bertujuan untuk

memvisualisasikan hasil dari proses *data mining* berupa Tabel atau Gambar serta menganalisa dan mengevaluasi apakah dapat digunakan untuk menyelesaikan masalah pada penelitian dengan hasil *data mining*.

### 3. Hasil Dan Pembahasan

Pada penelitian ini proses *data mining* menggunakan aplikasi weka , aplikasi tersebut berfungsi untuk membantu proses pengelompokan data dengan tool cluster dan 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*.

Tabel 7 Hasil Use Data Training Set

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

Pada proses pengetesan 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.

Tabel 8 Hasil Percentage Split 85%

Attribute	Full Data	Cluster 0	Cluster 1
	(311)	(159)	(152)
Status	2.2669	2.4088	2.1184
Usia	22.6785	23.956	21.3421
Mulai konsumsi	17.4695	17.2264	17.7237
Merk Produk	12.0579	11.2453	12.9079
Alasan Memilih	4.2926	4.3711	4.2105
Pemakaian Perhari	3.3312	3.1887	3.4803
Pengeluaran Bulanan	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 *training set* sedangkan sisanya yaitu 15% menjadi *test set*. Dan didapatkan hasil *cluster 0* sebanyak 47% dan *cluster 1* sebanyak 53%.

Sedangkan proses pengelompokan data 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 1* dan *cluster 2*. 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 1	Cluster 2
1	1
2	2
21	19
19	18
4	14
3	7
1	4
1	1

Pada Tabel 9 Nilai centroid awal ditentukan secara acak, diambil dari data mentah yang sudah diubah menjadi data *numerik*. 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.

Tabel 10 Hasil Iterasi ke-1

data ke -I	c1	c2	Min	cluster
1	3.74	14.14	3.74	1
2	7.42	13.60	7.42	1
3	1.00	11.45	1.00	1
4	0.00	11.40	0.00	1
5	5.10	12.25	5.10	1
6	5.10	14.97	5.10	1
7	4.36	10.44	4.36	1
8	5.29	12.81	5.29	1
9	4.58	14.46	4.58	1
10	8.54	8.43	8.43	2
11	10.86	5.29	5.29	2
12	16.70	11.79	11.79	2
13	10.95	3.46	3.46	2
14	16.25	5.48	5.48	2
15	15.72	14.46	14.46	2

Tabel 11 Hasil Iterasi ke-2

data ke -I	c1	c2	Min	cluster
1	2.55	13.57	2.55	1
2	5.05	15.00	5.05	1
3	2.94	11.35	2.94	1
4	2.84	11.35	2.84	1
5	2.84	12.73	2.84	1
6	3.25	14.48	3.25	1
7	3.09	11.22	3.09	1
8	3.63	14.15	3.63	1
9	5.23	13.69	5.23	1
10	7.96	8.41	7.96	1
11	11.86	4.59	4.59	2
12	18.09	6.98	6.98	2
13	11.93	4.33	4.33	2
14	16.96	7.33	7.33	2
15	17.99	10.74	10.74	2

Tabel 12 Hasil Iterasi ke-3

data ke-I	c1	c2	min	Cluster
1	2.94	14.86	2.94	1
2	4.93	16.53	4.93	1
3	3.04	12.53	3.04	1
4	2.98	12.53	2.98	1
5	2.58	14.19	2.58	1
6	3.67	15.79	3.67	1
7	2.46	12.71	2.46	1
8	3.75	15.53	3.75	1
9	5.63	14.71	5.63	1
10	7.16	10.09	7.16	1
11	11.32	5.30	5.30	2
12	17.55	6.06	6.06	2
13	11.33	5.22	5.22	2
14	16.29	7.57	7.57	2
15	17.78	9.71	9.71	2

Setelah dilakukan perhitungan iterasi sebanyak 4 kali dapat dilihat pada Tabel 10,11,12 dan 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 13 Hasil Iterasi ke-4

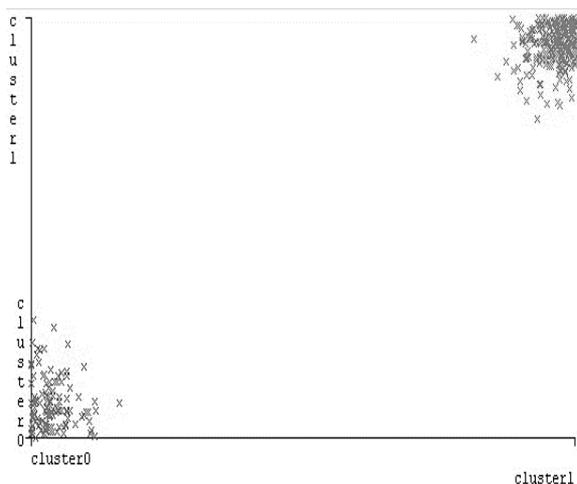
data ke-I	c1	c2	min	cluster
1	2.94	14.86	2.94	1
2	4.93	16.53	4.93	1
3	3.04	12.53	3.04	1
4	2.98	12.53	2.98	1
5	2.58	14.19	2.58	1
6	3.67	15.79	3.67	1
7	2.46	12.71	2.46	1
8	3.75	15.53	3.75	1
9	5.63	14.71	5.63	1
10	7.16	10.09	7.16	1
11	11.32	5.30	5.30	2
12	17.55	6.06	6.06	2
13	11.33	5.22	5.22	2
14	16.29	7.57	7.57	2
15	17.78	9.71	9.71	2

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

Tabel 14 Centroid Akhir

Cluster 1	Cluster 2
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

Dapat dilihat pada Gambar 3 hasil dari perluasan data *cluster* berdasarkan anggota yang telah dihitung sesuai jarak data terdekat. Pada Gambar 3 memperlihatkan bahwa *cluster 0* berada di posisi bawah merupakan kelompok data yang memiliki nilai rendah dengan kategori produk x tipe A diperoleh jarak terdekat dengan nilai 4.5, sedangkan *cluster 1* yang berada di posisi atas merupakan kelompok data yang memiliki nilai tinggi dengan kategori produk x tipe B diperoleh jarak terdekat dengan nilai 15.6721 dari atribut merk produk yang sudah diinisialisasikan.



Gambar 3 Grafik Visual Hasil Cluster

Pada Tabel 15 menunjukkan jumlah data responden yang sudah dibagi dalam 2 kelompok yaitu *cluster 0* dan *cluster 1*, kemudian *presentase* jumlah data responden dapat dilihat di setiap *cluster*.

Tabel 15 Hasil Klaster

Cluster	Jumlah	Persen
0	122	33%
1	244	67%

#### 4. Kesimpulan

Berdasarkan hasil dari penelitian, metode *K-Means* dapat digunakan untuk melakukan proses klusterisasi dengan mengubah data menjadi 2 kelompok kemudian kelompok tersebut dapat digunakan untuk mengetahui jenis produk x yang banyak digunakan oleh konsumen. Dari hasil klusterisasi, toko tersebut dapat melakukan pengendalian stok produk x yang banyak dikonsumsi oleh pelanggan agar tidak terjadi *out of stock* 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. Saran pengembangan untuk penelitian selanjutnya dapat dilakukan dengan melakukan perbandingan menggunakan metode klusterisasi yang lain sebagai pembandingan tingkat akurasi yang lebih akurat dan bisa bermanfaat bagi pihak yang memerlukan.

#### Daftar Pustaka

- [1] R. Andrean, S. Fendy, and A. Nugroho, "Klusterisasi Pengendalian Persediaan Aki Menggunakan Metode K-Means," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 4, no. 1, pp. 5–12, 2019.
- [2] N. Sapkota, A. Alsadoon, P. W. C. Prasad, A. Elchouemi, and A. K. Singh, "Data Summarization Using Clustering and Classification: Spectral Clustering Combined with k-Means Using NFPH," in *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects, COMITCon 2019*, 2019, pp. 146–151.
- [3] R. Molaei and S. Alizadeh, "Checking and Analysing Customers Buying Behavior with Clustering Algorithm," *Palma J.*, vol. 16, no. 2, pp. 486–492, 2017.
- [4] 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.
- [5] A. Masruro, Kusriani, and E. T. Luthfi, "SISTEM PENUNJANG KEPUTUSAN PENENTUAN LOKASI WISATA MENGGUNAKAN K-MEANS CLUSTERING DAN TOPSIS," *J. Ilm. DASi*, vol. 15, no. 04, pp. 1–5, 2014.
- [6] M. E. Corapcioglu and H. Ogul, "A medical decision support system proposal supported by genomic analysis," in *2017 25th Signal Processing and Communications Applications*

Conference, SIU 2017, 2017.

- [7] R. M. Esteves, T. Hacker, and C. Rong, "Competitive K-means: A new accurate and distributed K-means algorithm for large datasets," in *Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom*, 2013, vol. 1, pp. 17–24.
- [8] S. Sharma and A. Bhagat, "Data preprocessing algorithm for Web Structure Mining," in *Proceedings on 5th International Conference on Eco-Friendly Computing and Communication Systems, ICECCS 2016*, 2017, pp. 94–98.
- [9] 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.
- [10] S. K. Dwivedi and B. Rawat, "A review paper on data preprocessing: A critical phase in web usage mining process," in *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, 2016, pp. 506–510.
- [11] 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.
- [12] 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.
- [13] R. Primartha, *BELAJAR MACHINE LEARNING TEORI DAN PRAKTIK*. Informatika Bandung, 2018.
- [14] A. Nugroho and T. M. Fahrudin, "Analisis dan Pemetaan Jumlah Penumpang Kereta Api di Indonesia Menggunakan Metode Statistik Deskriptif dan K-means Clustering," *J. Mantik*, vol. 3, no. 9, pp. 1–9, 2019.
- [15] S. Kapil, M. Chawla, and M. D. Ansari, "On K-means data clustering algorithm with genetic algorithm," in *2016 4th International Conference on Parallel, Distributed and Grid Computing, PDGC 2016*, 2016, pp. 202–206.