



Segmentasi Pengunjung Pusat Perbelanjaan Menggunakan Metode K-Means Clustering

Eka Bayu Maulana¹, Asep Id Hadiana², Fajri Rakhmat Umbara³

^{1, 2, 3}Jurusan Informatika, Universitas Jenderal Achmad Yani (UNJANI). Jl. Terusan Jend. Sudirman, Cimahi Selatan, Kota Cimahi.

e-mail: ekabayumaulana17@if.unjani.ac.id¹, asephadiana@lecturer.unjani.ac.id², fajri.rakhmat@lecture.unjani.ac.id³

ABSTRAK

Kata Kunci:

Data Mining
Pusat Perbelanjaan
Segmentasi
Klasterisasi

Data Mining telah banyak digunakan sebagai salah satu teknik dalam memperoleh informasi yang dapat bermanfaat bagi berbagai pihak. Saat ini Data Mining telah banyak digunakan diberbagai sektor bisnis salah satunya pada bidang industri retail. Seperti pada industri pusat perbelanjaan yang mana pusat perbelanjaan merupakan tempat terjadinya kegiatan jual beli dari penjual kepada pembeli yang dilakukan dalam jumlah yang besar. Dalam jumlah yang besar ini menjadikan pusat perbelanjaan tempat yang paling diminati oleh konsumen dikarenakan kelengkapan produk yang dijualnya, dalam hal ini konsumen merupakan faktor yang sangat penting bagi pusat perbelanjaan. Untuk mengetahui karakteristik dari setiap pengunjung yang berbelanja, diperlukan suatu cara untuk mengolah data kebiasaan konsumen saat berbelanja di pusat perbelanjaan, salah satunya adalah dengan melakukan segmentasi menggunakan algoritma k-means clustering. Segmentasi ini bertujuan untuk membagi kelompok masyarakat berdasarkan karakteristiknya, dimana pada penelitian ini dilakukan klasterisasi konsumen berdasarkan ekonomi dan berdasarkan jarak. Pada penelitian ini telah menghasilkan 2 klaster yang mana klaster pertama merupakan kelompok konsumen dengan rentang usia didominasi usia muda, memiliki pendapatan Rp.800.000-Rp.10.000.000, serta pengeluaran sebesar Rp.200.000-Rp.3.000.000. Pada kelompok kedua terdiri dari kelompok customer dengan usia yang relatif lebih tua dari kelompok sebelumnya, dan memiliki pendapatan Rp.15.000.000-Rp.35.000.000 serta pengeluaran sebesar Rp.2.700.000-Rp.6.500.000. Pemilihan klaster terbaik telah dilakukan berdasarkan uji validitas menggunakan elbow method serta silhouette method yang menunjukkan nilai terbaiknya di angka 2 dengan score silhouette sebesar 0.824.



ABSTRACT

Keyword:

Data Mining
Shopping Centers
Segmentation
Clustering

Data mining has been widely used as a technique in obtaining information that can be useful for various parties. Currently, Data Mining has been widely used in various business sectors, one of which is in the retail industry. As in the shopping center industry where shopping centers are places where buying and selling activities from sellers to buyers are carried out in large numbers. In large numbers, this makes shopping centers the most desirable places for consumers due to the completeness of the products they sell, in this case consumers are a very important factor for shopping centers. To find out the characteristics of each shopping visitor, a method is needed to process consumer habit data when shopping at a shopping center, one of which is by segmenting using the k-means clustering algorithm. This segmentation aims to divide community groups based on their characteristics, where in this study consumers were clustered based on economy and based on distance. This research has produced 2 clusters where the first cluster is a consumer group with an age range dominated by young people, having an income of Rp.800,000-Rp.10,000,000, and expenses of Rp.200,000-Rp.3,000,000. The second group consists of a group of customers who are relatively older than the previous group, and have an income of Rp.15,000,000-Rp. 35,000,000 and expenses of Rp.2,700,000-Rp. 6,500,000. The selection of the best cluster has been carried out based on the validity test using the elbow method and the silhouette method which shows the best value at number 2 with a silhouette score of 0.824.

PENDAHULUAN

Data Mining telah banyak digunakan sebagai salah satu teknik dalam memperoleh informasi yang dapat bermanfaat bagi berbagai pihak. Saat ini Data Mining juga banyak digunakan diberbagai sektor bisnis salah satunya pada bidang industri retail. Data Mining merupakan proses untuk menemukan pola yang dapat berguna berdasarkan data yang memiliki volume besar dan acak. Proses Data Mining dimulai dengan mengekstrak informasi yang sebelumnya tidak diketahui sehingga dapat dipahami dan dapat di tindak lanjuti[1]. Pusat perbelanjaan merupakan suatu tempat dimana terjadinya kegiatan jual beli maupun pertukaran barang dan jasa antara penjual dan pembeli. Pusat perbelanjaan dapat dibagi berdasarkan lokasi seperti pasar tradisional, supermarket, hypermarket dan mall[2]. Alasan yang menjadikan pusat perbelanjaan seperti supermarket ataupun hypermarket banyak diminati oleh masyarakat adalah lengkapnya produk yang dijual di sana. Semakin lengkap variasi produk yang ditawarkan kepada konsumen maka akan semakin banyak pula pilihan barang yang akan dibeli oleh konsumen sesuai dengan kebutuhannya. Dalam hal ini konsumen merupakan faktor yang sangat penting bagi pusat perbelanjaan. Untuk mengetahui profil segmen konsumen dari pusat perbelanjaan perlu dilakukan segmentasi dengan menggunakan metode algoritma pengklasifikasian tertentu.

Segmentasi merupakan strategi yang dilakukan dengan cara membagi konsumen menjadi kelompok-kelompok yang berbeda serta memiliki karakteristik, perilaku atau kebutuhan yang berbeda[3]. Tujuan dari segmentasi ini adalah untuk mengetahui target konsumen dengan lebih



optimal. Penggunaan segmentasi dapat dilakukan dengan menggunakan metode k-means seperti pada penelitian yang dilakukan oleh Dhanachandra dkk., dalam penelitiannya peneliti melakukan segmentasi menggunakan algoritma k-means, menurutnya penggunaan algoritma k-means dirasa baik karena memiliki efisiensi yang tinggi, peneliti juga menyarankan penggunaan pencari k-ideal seperti elbow method dan lainnya agar dapat menghemat waktu dalam menentukan centroid serta meningkatkan efisiensi dalam melakukan segmentasi[4]. Sejalan dengan pendapat Dhanachandra dkk., pada penelitian yang berjudul “A Genetic Algorithm for Optimized Initial Centers K-Means Clustering in SMEs” menunjukkan bahwa penggunaan metode k-means digunakan karena k-means memiliki kemampuan untuk mengelompokkan data dalam jumlah yang besar dengan waktu komputasi yang relatif cepat dan efisien[5].

Dalam penggunaan metode clustering K-Means juga memiliki kekurangan, seperti sensitif pada penentuan jumlah cluster k yang tepat. Umumnya penentuan jumlah cluster k ditentukan secara acak dan akan menghasilkan pengelompokan data yang berbeda-beda dan tidak terlalu memberikan hasil yang baik dan akurat[6]. Sehingga diperlukan metode validasi seperti elbow method, silhouette method, gap statistics dan lainnya untuk menentukan cluster k yang ideal. Dalam penelitian ini peneliti menggunakan elbow method dan silhouette method sebagai validasi dan pencari k-ideal. Sejalan dengan penelitian ini, pada penelitian sebelumnya yang berjudul ‘Analyze K-Value Selected Method of K-Means Clustering Algorithm to Clustering Province Based on Disease Case’ telah membandingkan penggunaan elbow method dan silhouette method sebagai metode validasi atau pencari k-ideal yang tepat serta efisien[7].

Penelitian ini menggunakan data kuesioner untuk segmentasi karakteristik pengunjung pusat perbelanjaan dan tipe pusat perbelanjaan yang lebih sering dikunjungi. Tujuan dari penelitian ini adalah menghasilkan segmentasi pengunjung pusat perbelanjaan untuk mengetahui trend perbelanjaan masyarakat berdasarkan data yang didapat dengan algoritma K-Means menggunakan metode validitas Elbow dan Silhouette sebagai penentu nilai cluster k yang optimal.

METODE

Penelitian ini merupakan penelitian data mining yang bersifat kuantitatif dengan menggunakan data pengunjung pusat perbelanjaan di wilayah Kota Cimahi yang didapat dengan kriteria pengunjung yang berbelanja di pusat perbelanjaan dengan kategori barang berupa bahan pokok (sembako). Variabel pada penelitian ini meliputi: (1) Nama Lengkap; (2) Jenis Kelamin; (3) Usia; (4) Pendapatan Bulanan; (5) Pengeluaran Bulanan; (6) Jarak, serta; (7) Jumlah Kunjungan. Data yang telah didapat selanjutnya dilakukan *preprocessing data* menggunakan metode *cleaning data* dan transformasi data yang mana proses tersebut berfungsi untuk mengubah data mentah yang didapat menjadi data dalam format yang dapat dilakukan mining. Proses selanjutnya yaitu dengan melakukan segmentasi menggunakan *K-Means Clustering* untuk membagi data berdasarkan klasifikasi yang ditentukan dengan cara menghitung *cluster* optimal menggunakan *elbow method* dan

silhouette method[8]. K-Means merupakan metode clustering yang digunakan untuk mendapatkan kelompok data dengan memaksimalkan kesamaan karakteristik dalam kluster dan memaksimalkan perbedaan antar kluster. Algoritma k-means clustering mengelompokkan data berdasarkan jarak antara data terhadap titik centroid kluster yang didapatkan melalui proses berulang[9]. Pada penelitian ini elbow method dan silhouette method juga digunakan sebagai metode evaluasi untuk mengukur cluster yang ideal[10]. Data segmentasi yang telah didapat selanjutnya dilakukan interpretasi menggunakan visualisasi dari grafik tiga dimensi berdasarkan ekonomi masyarakat serta jarak pusat perbelanjaan.



Gambar 1. Metode Penelitian

HASIL DAN PEMBAHASAN

Penelitian ini meliputi proses *Knowledge Discovery in Databases* (KDD) dengan proses pertama yaitu melakukan pengumpulan data sebagai berikut:

Tabel 1 Dataset Pengunjung Pusat Perbelanjaan

No.	Nama	JK	Usia	Pendapatan	Pengeluaran	Jarak	Jumlah Kunjungan
1.	Putri	P	27	Rp.5.500.000	Rp,2.500.000	10 Km	4
2.	Indra	L	28	Rp.6.200.000	Rp.2.000.000	5 Km	4
3.	Anisa	P	22	Rp.5.200.000	Rp.1.200.000	2 Km	6
4.	Clara	P	19	Rp.3.000.000	Rp.600.000	4 Km	4
5.	Prilla	P	26	Rp.4.000.000	Rp.1.000.000	3 Km	2
6.	Yohanes	L	40	Rp.25.000.000	Rp.6.000.000	4 Km	4
7.	Nita	P	49	Rp.15.000.000	Rp.4.000.000	2 Km	8
8.	Tedi	L	31	Rp.6.500.000	Rp.1.000.000	10 Km	3
9.	Erisa	P	27	Rp.6.000.000	Rp.1.500.000	6 Km	4
10.	Regina	P	28	Rp.5.500.000	Rp.1.500.000	5 Km	8
...
422.	Tini	P	54	Rp.25.000.000	Rp.5.000.000	4 Km	5

Dataset pengunjung pusat perbelanjaan yang didapat selanjutnya dilakukan proses *cleaning data* untuk menghilangkan atau mengkoreksi data yang salah, rusak, tidak lengkap atau salah format. Kemudian dataset tersebut dilakukan transformasi data sebagai berikut:

Tabel 2. Transformasi Data

Nama	JK	Usia	Pendapatan	Pengeluaran	Jarak	Jumlah Kunjungan
Putri	P	27	5500	2500	10	4
Indra	L	28	6200	2000	5	4
Anisa	P	22	5200	1200	2	6
Clara	P	19	3000	600	4	4
Prilla	P	26	4000	1000	3	2
Yohanes	L	40	25000	6000	4	4
Nita	P	49	15000	4000	2	8



Tedi	L	31	6500	1000	10	3
Erisa	P	27	6000	1500	6	4
Regina	P	28	5500	1500	5	8
...
Tini	P	54	25000	5000	4	5

Dataset yang telah ditransformasi pada contoh tabel diatas selanjutnya dilakukan normalisasi menggunakan metode *min max normalization* (1). Persamaan *min max normalization* adalah sebagai berikut:

$$v'_i = \left(\frac{v_i - \min_A}{\max_A - \min_A} * (new_max_A - new_min_A) \right) + new_min_A \quad (1)$$

Keterangan :

v'_i : Nilai data yang baru dari hasil normalisasi min-max.

v_i : Nilai data yang akan dilakukan normaliasi.

\max_A : Nilai maksimum data.

\min_A : Nilai minimum data.

new_max_A : Nilai maksimum yang diharapkan dari proses normalisasi.

new_min_A : Nilai minimum yang diharapkan dari proses normalisasi

Tabel 3. Normalisasi Min Max

Nama	JK	Usia	Pendapatan	Pengeluaran	Jarak	Jumlah Kunjungan
Putri	P	0,229	0,204	0,407	0,643	0,273
Indra	L	0,257	0,233	0,322	0,286	0,273
Anisa	P	0,086	0,192	0,186	0,071	0,455
Clara	P	0,000	0,102	0,085	0,214	0,273
Prilla	P	0,200	0,143	0,153	0,143	0,091
Yohanes	L	0,600	1,000	1,000	0,214	0,273
Nita	P	0,857	0,592	0,661	0,071	0,636
Tedi	L	0,343	0,245	0,153	0,643	0,182
Erisa	P	0,229	0,224	0,237	0,357	0,273
Regina	P	0,257	0,204	0,237	0,286	0,636
...
Tini	P	1,000	1,000	0,831	0,214	0,364

Proses selanjutnya adalah melakukan *clustering* menggunakan *K-Means*, langkah pertama yang dilakukan adalah dengan menentukan jumlah cluster dan nilai pusat centroid. Penentuan jumlah dan nilai pusat centroid awal dilakukan secara acak, dalam hal ini jumlah cluster ditentukan sebanyak dua.

Tabel 4. Penentuan *Centroid*

	Usia	Pendapatan	Pengeluaran	Jarak	Jumlah Kunjungan
Centroid 1	0,086	0,192	0,186	0,071	0,455
Centroid 2	0,600	1,000	1,000	0,214	0,273

Selanjutnya dilakukan perhitungan jarak antara data dengan pusat cluster menggunakan Euclidian Distance (2). Berikut rumus untuk menghitung jarak antara data dengan pusat cluster:

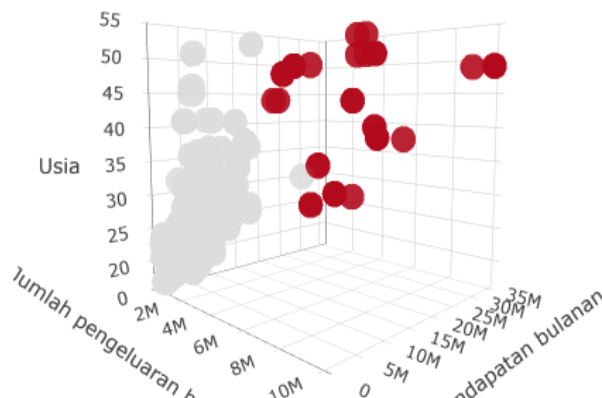
$$D_{ij} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (2)$$

Tabel 5. Hasil Perhitungan Euclidian Distance

Data ke-	C1	C2	Jarak Terdekat
1	0,55114	1,54050	0,55114
2	0,27046	1,39642	0,27046
3	0,00000	1,88276	0,00000
4	0,15759	2,24402	0,15759
5	0,25517	1,89104	0,25517
6	1,88276	0,00000	0,00000
7	1,18971	0,69129	0,69129
8	0,66202	1,73744	0,66202
9	0,26120	1,57498	0,26120
10	0,25314	1,69541	0,25314
...
422.	2,01091	0,43699	0,43699

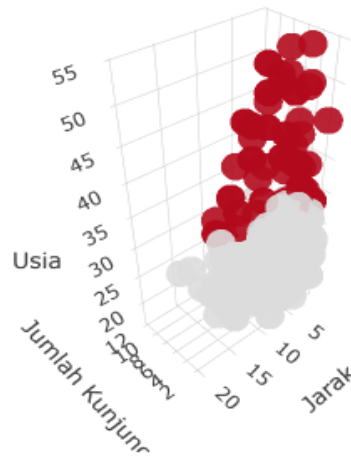
Tabel 6. Pengelompokan Data

Data ke-	Nama	Kelompok Data
1	Putri	Cluster 1
2	Indra	Cluster 1
3	Anisa	Cluster 1
4	Clara	Cluster 1
5	Prilla	Cluster 1
6	Yohanes	Cluster 2
7	Nita	Cluster 2
8	Tedi	Cluster 1
9	Erisa	Cluster 1
10	Regina	Cluster 1
...
422.	Tini	Cluster 2



Gambar 2. Visualisasi *Clustering* Berdasarkan Ekonomi

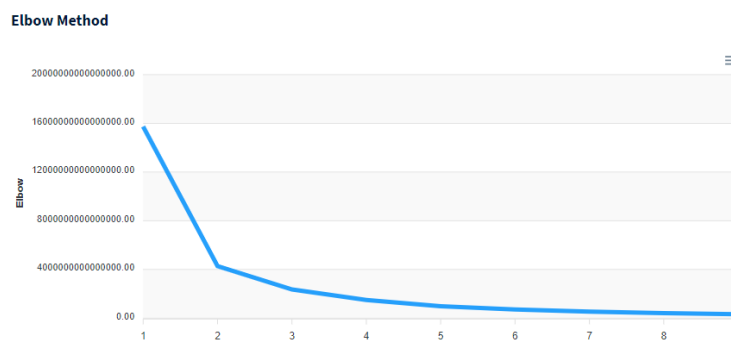
- Kelompok 1 didominasi usia 20-29 tahun, pendapatan Rp. 800.000 - Rp.10.000.000, pengeluaran belanja Rp.200.000 - Rp. 3.000.000
- Kelompok 2 usia 30-54 tahun, pendapatan lebih tinggi diangka Rp.15.000.000 - Rp. 35.000.000, pengeluaran belanja Rp.2.700.000 - Rp. 6.500.000.



Gambar 3. Visualisasi *Clustering* Berdasarkan Jarak

Klasterisasi selanjutnya dilakukan berdasarkan jarak tempat tinggal pengunjung dengan pusat perbelanjaan, didapat hasil yaitu didominasi pengunjung dengan usia yang relatif muda dengan rentang usia 20-30 tahun, dan memiliki jarak terdekat dengan tempat tinggal mulai dari 2 – 6 Km serta jumlah kunjungan ke pusat perbelanjaan sebanyak 2-4x setiap bulannya.

Pengujian menggunakan *elbow method* dan *silhouette method* menunjukkan hasil optimal penentuan jumlah *cluster* adalah sebanyak dua cluster.



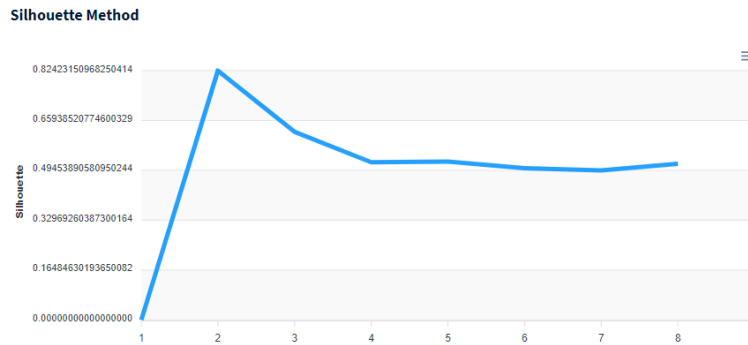
Gambar 4. Elbow Method

Tabel 7. Elbow Score

Jumlah Klaster	Elbow Score
1	15741757791699918.0
2	4283611593271736.5
3	2360966947038041.0
4	1488308938693367.2
5	978454772026657.6
6	713410900230296.8



7	529861812249171.56
8	417336991704884.3
9	328194327992677.7



Gambar 5. Silhouette Method

Tabel 8. Silhouette Score

Jumlah Kluster	Silhouette Score
2	0.8242315096825041
3	0.6219392769677856
4	0.5210898656722257
5	0.523801544256025
6	0.5018452066493049
7	0.49445368236501513
8	0.5163151317292762

Dari evaluasi yang dilakukan menggunakan elbow method dan silhouette method didapat jumlah cluster optimal adalah sebanyak dua. Sesuai dengan penerapan pada k-means dengan jumlah *cluster* yang digunakan dalam penelitian ini adalah dua dengan skor siluet 0.824 menjadikan klasterisasi yang dilakukan sudah cukup optimal.

SIMPULAN

Dalam penelitian Segmentasi Pengunjung Pusat Perbelanjaan Menggunakan Metode K-Means Clustering, diketahui bahwa jumlah k yang optimal adalah sebanyak dua, yang mana sebelumnya telah dilakukan uji validasi menggunakan elbow method dan silhouette method. Dari hasil klastering tersebut terdapat dua kluster dengan rincian kelompok 1 yang dilihat berdasarkan ekonomi merupakan karakteristik umur yang relatif muda dengan rentang usia didominasi oleh 20-30 tahun, memiliki pendapatan bulanan mulai dari Rp.800.000 – Rp.10.000.000, serta memiliki pengeluaran bulanan yang relatif kecil dengan rentang pengeluaran mulai dari Rp.200.000 – Rp.3.000.000. Kemudian pada kelompok 2 terdiri dari kelompok dengan rentang usia yang relatif tua dengan rentang 29 – 54 tahun, dan memiliki pendapatan bulanan yang lebih tinggi dari kelompok sebelumnya dengan rentang mulai dari Rp.15.000.000 – Rp.35.000.000, serta memiliki pengeluaran yang relatif besar dengan rentang Rp.2.700.000 – Rp.6.500.000 per bulannya.



Klasterisasi selanjutnya dilakukan berdasarkan jarak tempat tinggal pengunjung dengan pusat perbelanjaan, yang mana didapat hasil yaitu didominasi pengunjung dengan usia yang relatif muda dengan rentang usia 20-30 tahun, dan memiliki jarak terdekat dengan tempat tinggal mulai dari 2 – 6 Km dan jumlah kunjungan ke pusat perbelanjaan sebanyak 2-4x setiap bulannya.

Berdasarkan hasil penelitian yang telah dilakukan, terdapat saran yang dapat dilakukan pada penelitian lebih lanjut, yaitu diharapkan dapat melakukan perhitungan terhadap jumlah data yang lebih besar serta penggunaan atribut yang lebih banyak agar dapat menghasilkan kualitas cluster yang lebih baik.

DAFTAR RUJUKAN

- [1] G. Schuh *et al.*, “Data Mining Definitions and Applications for the Management of Definitions and Applications the Management of Production Complexity Production Complexity A new methodology to a , analyze the functional and physical architecture,” *Procedia CIRP*, vol. 81, pp. 874–879, 2019, doi: 10.1016/j.procir.2019.03.217.
- [2] R. Savitri, “PUSAT PERBELANJAAN MODERN (MALL),” vol. 6, no. September 2018, pp. 229–245.
- [3] Sulistyowati, B. E. Ketherin, A. A. Arifiyanti, and A. Sodik, “Analisa Segmentasi Konsumen Menggunakan Algoritma K-Means Clustering,” *Seminar Nasional Sains dan Teknologi Terapan VI 2018*, pp. 51–58, 2018.
- [4] N. Dhanachandra, K. Mangle, and Y. J. Chanu, “Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm,” *Procedia Computer Science*, vol. 54, no. December 2016, pp. 764–771, 2015, doi: 10.1016/j.procs.2015.06.090.
- [5] B. K. Khotimah, F. Irhamni, and T. Sundarwati, “A genetic algorithm for optimized initial centers K-means clustering in SMEs,” *Journal of Theoretical and Applied Information Technology*, vol. 90, no. 1, pp. 23–30, 2016.
- [6] R. Siagian, “E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model,” *JITE (Journal of Informatics and Telecommunication Engineering)*, vol. 5, no. July, pp. 152–159, 2021.
- [7] A. D. Savitri, F. A. Bachtiar, and N. Y. Setiawan, “Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, vol. 2, no. 9, pp. 2957–2966, 2018.
- [8] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, “Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster,” *IOP Conference Series: Materials Science and Engineering*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.
- [9] M. T. A. Munna, M. M. Alam, S. M. Allayear, K. Sarker, and S. J. F. Ara, *Prediction model*



- for prevalence of type-2 diabetes complications with ANN approach combining with K-fold cross validation and K-means clustering*, vol. 69. Springer International Publishing, 2020.
- [10] F. Wang, H. H. Franco-Penya, J. D. Kelleher, J. Pugh, and R. Ross, “An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10358 LNAI, no. December 2018, pp. 291–305, 2017, doi: 10.1007/978-3-319-62416-7_21.