

ANALISIS SEGMENTASI PERILAKU BELANJA PELANGGAN DENGAN ALGORITMA K-MEANS CLUSTERING

Dian Ayu Fauziah

Program Studi Sains Data, FMIPA, Universitas Negeri Surabaya, Surabaya, Indonesia e-mail:

dian.22011@mhs.unesa.ac.id

ABSTRAK

Dunia perbelanjaan memang suatu hal yang sangat digemari masyarakat sedari dulu, tidak hanya kalangan ibu-ibu atau wanita saja tetapi mulai dari anak-anak, remaja, bapak-bapak, ibu-ibu, bahkan lansia pun masih melakukan kegiatan perbelanjaan. Tujuan dari artikel ini adalah melakukan analisis dan segmentasi perilaku belanja pelanggan dengan memanfaatkan algoritma K-Means Clustering. Di tengah perkembangan pesatnya e-commerce, pemahaman yang komprehensif terhadap pola pembelian pelanggan menjadi krusial untuk meningkatkan pengalaman berbelanja dan merancang strategi pemasaran yang efisien. Algoritma K-Means Clustering diadopsi untuk mengelompokkan pelanggan berdasarkan ciri-ciri pembelian yang mereka tunjukkan. Metode artikel ini mencakup langkah-langkah seperti pemrosesan data, pemilihan fitur yang sesuai, penerapan algoritma K-Means, dan analisis hasil segmentasi. Penilaian kinerja model dan interpretasi hasil segmentasi menjadi fokus utama untuk memastikan relevansi dan efektivitas dari strategi pemasaran yang dihasilkan. Melalui penerapan pendekatan ini, diharapkan proyek ini dapat memberikan wawasan berharga kepada perusahaan dalam memahami dan merespons kebutuhan khusus dari berbagai segmen pelanggan dalam konteks perilaku belanja mereka.

Kata kunci : Segmentasi, Clustering, K-Means Clustering

PENDAHULUAN

Selama bertahun-tahun, persaingan antar bisnis semakin ketat meningkat dan data historis besar yang tersedia telah menyebabkan meluasnya penggunaan data mining teknik dalam mengekstraksi yang bermakna dan strategis informasi dari database organisasi (Rahmanto & Fernando, 2019); (Alita et al., 2020); (Aldino et al., 2021b). Dunia perbelanjaan tidak akan pernah mati, dikarenakan tren belanja setiap tahunnya terus meningkat. Manajemen dan pengelolaan hubungan pelanggan mempunyai peran yang sangat penting dalam menyediakan Business Intelligent untuk mengorganisir, membangun, mengelola dan mengembangkan hubungan customer yang bernilai dalam jangka panjang. Pentingnya memperlakukan pelanggan sebagai aset utama suatu organisasi semakin meningkat saat ini. Organisasi tertarik untuk berinvestasi dalam pengembangan strategi akuisisi, retensi, dan pengembangan pelanggan. Intelijen bisnis memainkan peran penting dalam memungkinkan perusahaan menggunakan pengetahuan teknis untuk mendorong hasil melalui wawasan dan program pelanggan yang lebih

baik. Teknik clustering seperti k-means mengelompokkan pelanggan dengan nilai mean yang sama secara bersamaan. Segmentasi pelanggan membantu tim pemasaran mengidentifikasi dan menemukan segmen pelanggan yang memikirkan dan mengikuti strategi pembelian yang berbeda. [1] Segmentasi pelanggan membantu menemukan pelanggan yang berbeda dalam hal preferensi, harapan, keinginan dan karakteristik. Tujuan utama dari segmentasi pelanggan adalah untuk mengelompokkan orang-orang dengan minat yang sama sehingga tim pemasaran dapat menyusun rencana pemasaran yang efektif. [2]

Clustering adalah proses berulang untuk menemukan informasi dari sejumlah besar data mentah yang tidak terstruktur. Clustering adalah jenis eksplorasi data yang digunakan dalam banyak aplikasi seperti pembelajaran mesin, klasifikasi, dan pengenalan pola. Sedangkan tujuan dari segmentasi pelanggan pada project ini, yaitu untuk mengetahui perilaku belanja pelanggan berdasarkan pendapatan dan juga skor pengeluaran yang dilakukan oleh pelanggan itu sendiri. [3]

KAJIAN PUSTAKA

Customer Segmentation

Pelanggan segmentasi adalah strategi membagi pasar menjadi kelompok homogen. Data yang digunakan di pelanggan teknik segmentasi yang membagi pelanggan menjadi kelompok bergantung pada berbagai faktor seperti, data kondisi geografis, kondisi ekonomi, kondisi demografis serta perilaku pola (Susanto & Puspaningrum, 2019);(Budiman et al., 2021);(Hendrastuty et al., 2021).

Teknik segmentasi pelanggan memungkinkan bisnis untuk memanfaatkan pemasaran mereka dengan lebih baik anggaran, mendapatkan keunggulan kompetitif atas saingan mereka perusahaan, menunjukkan pengetahuan yang lebih baik tentang kebutuhan pelanggan (Wahyudi et al., 2021);(Nurkholis, Damayanti, et al., 2021). Ini juga membantu organisasi dalam, meningkatkan efisiensi pemasaran mereka, menentukan peluang pasar baru, membuat merek yang lebih baik strategi, mengidentifikasi retensi pelanggan(Febrian et al., 2021);(Rahman Isnain et al., 2021);(Permatasari & Anggarini, 2020).

Clustering

Clustering adalah proses berulang untuk menemukan informasi dari sejumlah besar data mentah yang tidak terstruktur. Clustering adalah jenis eksplorasi data yang digunakan dalam banyak aplikasi seperti pembelajaran mesin, klasifikasi, dan pengenalan pola.[4]

Elbow Methods

Elbow method adalah sebuah metode optimalisasi yang paling sering digunakan dalam mengidentifikasi jumlah cluster optimal dalam suatu dataset (Masud dkk., 2018). Pengertian lain dari Elbow method yaitu metode tertua untuk mengidentifikasi jumlah cluster optimal pada dataset yang dianalisis, yang mana ide dasar dari metode tersebut yaitu menentukan $k = 2$ sebagai jumlah cluster optimal awal k . Kemudian, jumlah cluster optimal k tersebut akan terus bertambah dari langkah pertama ke langkah yang paling maksimum untuk menentukan perkiraan jumlah cluster optimal potensial dan akhirnya akan dibedakan jumlah cluster optimal k sesuai dengan levelnya (Shi dkk., 2021).

K -Means

Algoritma pengelompokan membuat grup sedemikian rupa sehingga grup tersebut berisi objek yang serupa berdasarkan beberapa karakteristik. Kesamaan ditentukan oleh seberapa dekat objek dalam ruang. Menurut, algoritma K-means merupakan salah satu

algoritma berbasis centroid yang paling populer. Misalkan kumpulan data D berisi n objek di luar angkasa. Metode distribusi membagi objek D menjadi k kelompok, C_1, \dots, C_k , itu $C_i \subset D$ dan $C_i \cap C_j = \emptyset$ ($1 \leq i, j \leq k$). Teknik distribusi berbasis centroid menggunakan centroid C_i untuk mewakili suatu kelompok. Secara konseptual, pusat suatu kelompok adalah pusatnya. Selisih antara benda $p \in C_i$ dan benda c_i yang mewakili grup diukur dengan $\text{dist}(p, c_i)$, dimana $\text{dist}(x, y)$ adalah jarak Euclidean antara dua titik x dan y . [5]

Algoritma: Algoritma k-means untuk partisi, di mana pusat setiap kluster diwakili oleh nilai rata-rata objek dalam kluster tersebut. Input: k : jumlah kluster, D : set data yang berisi n objek. Output: Seperangkat k kluster. Metode: (1) secara sembarang memilih k objek dari D sebagai pusat kluster awal; (2) ulangi (3) (re)memasukkan setiap objek ke dalam kluster yang paling mirip dengannya, berdasarkan nilai rata-rata objek dalam kluster; (4) perbarui nilai rata-rata kluster, yaitu, hitung nilai rata-rata objek untuk setiap kluster; (5) hingga tidak ada perubahan.

METODE

Clustering

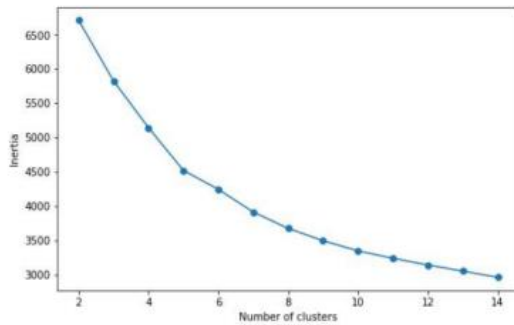
Clustering adalah salah satu metode paling umum yang digunakan dalam mengeksplorasi data untuk memperoleh pemahaman yang jelas tentang struktur data. Ini dapat dicirikan sebagai tugas menemukan sub tipe dan subkelompok dalam dataset lengkap. Data yang serupa dikelompokkan dalam banyak subkelompok. Sebuah kelompok merujuk pada kumpulan titik data yang teragregasi karena adanya kesamaan tertentu. Clustering digunakan dalam analisis tren belanja untuk membagi pelanggan berdasarkan perilaku dan transaksi mereka.

K -means Clustering Algorithm

Pengelompokan K-Means adalah algoritma machine learning yang paling umum dan sederhana. Algoritma ini mengikuti pendekatan iteratif yang berusaha membagi dataset menjadi berbagai subkelompok " k " yang telah ditentukan dan saling tidak tumpang tindih, di mana setiap titik data hanya termasuk ke dalam satu subkelompok berdasarkan karakteristik serupa mereka.[6]

Elbow Method

Elbow Method adalah alat yang digunakan untuk menganalisis kluster yang terbentuk dari dataset kita dan membantu menginterpretasikan jumlah optimal kluster yang sesuai dalam dataset. Dari metode ini, ditemukan bahwa jumlah optimal kluster.

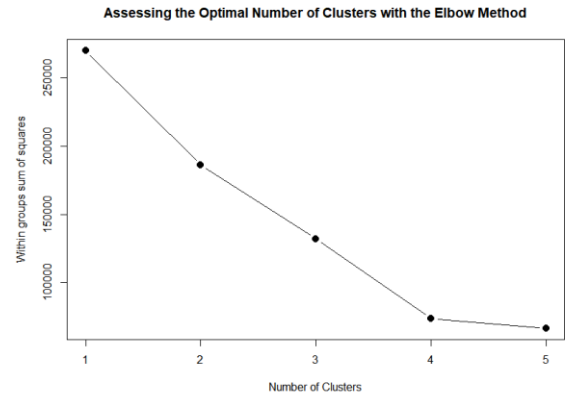


PEMBAHASAN

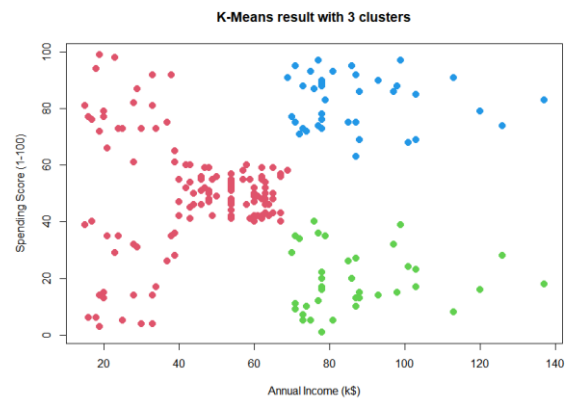
Dalam menganalisis suatu data, kita memerlukan untuk mengetahui kondisi awal data yang akan digunakan. Kondisi awal suatu data dapat diidentifikasi melalui analisis statistik, yang memberikan gambaran deskriptif tentang data tersebut. Proses ini bertujuan untuk memahami karakteristik variabel perilaku tren belanja customer dengan mengevaluasi nilai rata-rata, nilai minimum, dan nilai maksimum.

Pada project ini data yang digunakan adalah data Customer mall yang berjumlah 200 baris dan 5 kolom, selanjutnya data customer mall akan dibagi menjadi tiga cluster berdasarkan rata-rata dari masing-masing fitur yang digunakan, yaitu data Annual Income(k\$) dan juga Spending Score. Nantinya cluster 1 menunjukkan pelanggan dengan pendapatan tinggi, tetapi skor pengeluaran rendah. Lalu, cluster 2 menunjukkan pelanggan muda dengan pendapatan tinggi dan skor pengeluaran tinggi, yang terakhir cluster 3 menunjukkan pelanggan dengan pendapatan yang lebih rendah dibandingkan dengan klaster lainnya, tetapi dengan skor pengeluaran yang lebih tinggi.

Langkah pertama yang perlu dilakukan pada metode clustering yaitu menentukan jumlah cluster optimal k yang akan digunakan dalam metode clustering selanjutnya. Pada artikel ini akan digunakan Elbow method untuk menentukan jumlah cluster optimalnya. Dengan melakukan pengolahan data menggunakan bantuan software Rstudio diketahui bahwa jumlah cluster optimal untuk data Tren Belanja Customer Mall yaitu 3 cluster. Hasil pengolahan data dengan Elbow method dapat dilihat seperti pada Gambar berikut.



Langkah selanjutnya yaitu proses clustering menggunakan metode K-Means Clustering. Berdasarkan hasil optimasi menggunakan Elbow Method sebelumnya, diketahui bahwa jumlah cluster optimal dalam pengelompokan menggunakan K-Means clustering adalah 3 cluster. Maka hasil pengelompokan menggunakan algoritma K-means clustering dengan jumlah cluster optimal yaitu 3 cluster, sebelum melakukan perhitungan menggunakan K-Means Clustering dilakukan pemilihan fitur terlebih dahulu dari data yang akan digunakan. Fitur yang dipilih untuk perhitungan K-Means Clustering disini yaitu Annual Income(k\$) dan Spending Score dapat direpresentasikan dalam Gambar sebagai berikut.



Gambar diatas merupakan representasi dari K-Means Clustering 3 cluster

Selanjutnya, yaitu membagi data ke dalam 3 cluster berdasarkan rata-rata setiap fiturnya. Dengan melihat nilai rata-rata pada setiap cluster yang dihasilkan oleh 'kmeans_result\$centers', bisa memberikan interpretasi awal terhadap karakteristik masing-masing cluster seperti pada gambar berikut ini.

	Age	Annual Income (k\$)	Spending score (1-100)
1	40.39474	87.00000	18.63158
2	32.69231	86.53846	82.12821
3	40.32520	44.15447	49.82927

Dari gambar diatas menunjukkan bahwasannya cluster 1 rata – rata usia customer yaitu 40 tahun dengan pendapatan tahunan sekitar \$87.00, dan skor pengeluaran 18.63 menunjukkan pelanggan dengan pendapatan tinggi, tetapi skor pengeluaran rendah. Lalu, cluster 2 rata -rata usia customer 32 tahun dengan pendapatan tahunan sekitar \$86.53, dan skor pengeluaran 82.12 menunjukkan pelanggan muda dengan pendapatan tinggi dan skor pengeluaran tinggi, yang terakhir cluster 3 rata -rata usia 40 tahun dengan pendapatan tahunan sekitar \$44.15, dan skor pengeluaran 49.82 menunjukkan pelanggan dengan pendapatan yang lebih rendah dibandingkan dengan klaster lainnya, tetapi dengan skor pengeluaran yang lebih tinggi.

```
> table(kmeans_result$cluster)
```

```
1 2 3
38 39 123
```

Dengan table(kmeans_result\$cluster) diketahui masing -masing frekuensi dari setiap cluster. Data diatas menunjukkan cluster 1 memiliki jumlah customer 38, cluster 2 memiliki jumlah customer 39, dan juga cluster 3 memiliki jumlah customer 123

KESIMPULAN

Artikel ini menggunakan algoritma K -Means Clustering yang bertujuan untuk melihat bagaimana perilaku belanja customer berdasarkan data yang digunakan yang berisi fitur -fitur seperti CustomerID, Gender, Age, Annual Income (k\$), Spending Score (1-100).

Pada artikel ini digunakan pula Elbow Methods untuk mencari jumlah cluster optimal. Berdasarkan Elbow Method diketahui bahwa jumlah cluster optimal yaitu berjumlah 3 cluster. Setelah dilakukan Elbow Method untuk mencari jumlah cluster optimal selanjutnya yaitu melakukan perhitungan rata -rata menggunakan algoritma K -Means clustering terhadap fitur – fitur yang sudah dipilih seperti Age, Annual Income(k\$), dan Spending Score(1-100) untuk melakukan segmentasi terhadap perilaku belanja customer.

Setelah melakukan perhitungan dengan pemodelan algoritma K-Means clustering didapatkan rata -rata untuk setiap fitur pada masing -masing cluster.

Cluster	Age	Annual Income(k\$)	Spending Score(1-100)
Cluster 1	40.39	\$87.00	18.634
Cluster 2	32.69	\$86.53	82.128
Cluster 3	40.32	\$44.15	49.829

Dari tabel diatas menunjukkan bahwasannya berdasarkan perhitungan segmentasi customer menggunakan K -Means Clustering didapatkan cluster 1 dengan rata – rata usia customer yaitu 40 tahun dengan pendapatan tahunan sekitar \$87.00, dan skor pengeluaran 18.63 menunjukkan pelanggan dengan pendapatan tinggi, tetapi skor pengeluaran rendah. Lalu, cluster 2 rata -rata usia customer 32 tahun dengan pendapatan tahunan sekitar \$86.53, dan skor pengeluaran 82.12 menunjukkan pelanggan muda dengan pendapatan tinggi dan skor pengeluaran tinggi, yang terakhir cluster 3 rata -rata usia 40 tahun dengan pendapatan tahunan sekitar \$44.15, dan skor pengeluaran 49.82 menunjukkan pelanggan dengan pendapatan yang lebih rendah dibandingkan dengan klaster lainnya, tetapi dengan skor pengeluaran yang lebih tinggi.

DAFTAR PUSTAKA

- [1] B. Turkmen, “Customer Segmentation With Machine Learning for Online Retail Industry,” *The European Journal of Social and Behavioural Sciences*, vol. 31, no. 2, pp. 111–136, Apr. 2022, doi: 10.15405/EJSBS.316.
- [2] C. Qian, M. Yang, P. Li, and S. Li, “Application of Customer Segmentation for Electronic Toll Collection: A Case Study,” *J Adv Transp*, vol. 2018, 2018, doi: 10.1155/2018/3635107.
- [3] M. Anggara, H. Sujaini, and H. Nasution, “Pemilihan Distance Measure

Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness,” *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. 4, no. 1, pp. 186–191, Jan. 2016, Accessed: Dec. 14, 2023. [Online]. Available: <https://jurnal.untan.ac.id/index.php/justin/article/view/13119>

- [4] “IEEE Xplore Full-Text PDF:” Accessed: Dec. 14, 2023. [Online]. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9072123>
- [5] H. Firdaus, “Tahun 2022 50 ANALISA CLUSTER MENGGUNAKAN K-MEANS DAN FUZZY C-MEANS DALAM PENGELOMPOKAN PROVINSI MENURUT DATA INTESITAS BENCANA ALAM DI INDONESIA TAHUN 2017-2021”.
- [6] “View of ANALISIS SEGMENTATION CUSTOMER MALL MENGGUNAKAN CLUSTERING ALGORITMA K-MEANS.” Accessed: Dec. 14, 2023. [Online]. Available: <http://ilmudata.org/index.php/ilmudata/article/view/266/255>