



**ANALISIS CLUSTERISASI PENGUNJUNG MALL BERDASARKAN
USIA DAN PENDAPATAN MENGGUNAKAN ALGORITMA DBSCAN**

*Clusterization Analysis Of Mall Visitors Based On Age And Income Using The
DBSCAN Algorithm*

Medy Wisnu Prihatmono

STMIK Profesional Makassar

***Correspondence Author: medy_wisnu_prihatmono@stmikprofesional.ac.id**

Abstract

Data mining can be used to extract valuable insights from irregular and high-volume mall customer data. More effective shopping strategies can be created by thoroughly understanding the demographics of mall customers to attract new customers, increase retention of existing customers, and create a more engaging shopping experience. The purpose of this study was to look at visitor data that has implemented strategies and to increase visitor volume. The research sample consisted of 200 mall visitor respondents who participated in the Ramadan quiz program during the period of 1 to 30 days of the month of Ramadan. Based on the data density analysis, five Cluster groups were formed with the following distribution: Noise (Cluster -1): 130 data (65% of the total sample), representing data points that are not associated with significant density, Cluster 0: 10 data (5% of the total sample), Cluster 1: 14 data (7% of the total sample), Cluster 2: 11 data (5.5% of the total sample), Cluster 3: 35 data (17.5% of the total sample).

Keywords: *Mall Visitor Analysis, Clustering, DBSCAN*

Abstrak

Data mining dapat digunakan untuk mengekstrak informasi berharga dari data tidak teratur dan bervolume tinggi pelanggan mall. Strategi belanja yang lebih efektif dapat dibuat dengan memahami secara menyeluruh demografi pelanggan mall untuk menarik pelanggan baru, meningkatkan retensi pelanggan yang sudah ada, dan menciptakan pengalaman belanja yang lebih menarik. Tujuan penelitian ini adalah untuk melihat data pengunjung yang telah menerapkan strategi dan untuk meningkatkan volume pengunjung. Sampel penelitian terdiri dari 200 responden pengunjung mall yang berpartisipasi dalam program kuis Ramadan selama periode 1 hingga 30 hari bulan Ramadan. Berdasarkan analisis kepadatan data, terbentuk lima kelompok Cluster dengan distribusi sebagai berikut: Noise (Cluster -1): 130 data (65% total sampel), merepresentasikan titik data yang tidak terasosiasi dengan kepadatan signifikan, Cluster 0: 10 data (5% total sampel), Cluster 1: 14 data (7% total sampel), Cluster 2: 11 data (5.5% total sampel), Cluster 3: 35 data (17.5% total sampel).

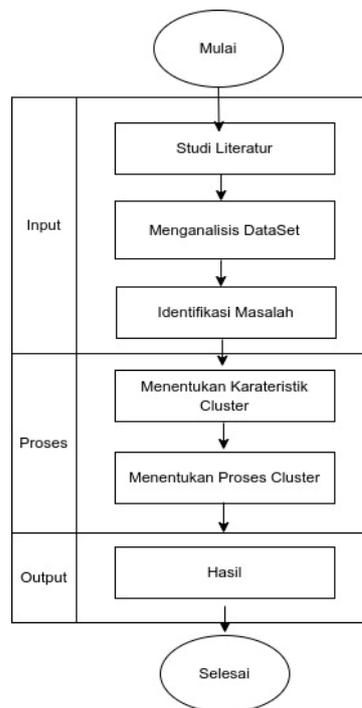
Kata Kunci: Analisis Pengunjung Mall, Clustering, DBSCAN

PENDAHULUAN

Dalam membuat dan memasarkan produk, perusahaan harus memperhatikan kebutuhan konsumen, yang dapat dilihat dari perilaku pembelian. Perilaku konsumen adalah studi tentang bagaimana orang, kelompok, atau organisasi memilih, membeli, atau menggunakan barang, jasa, ide, atau pengalaman untuk memenuhi keinginan dan kebutuhan yang tersebar luas. Perusahaan harus mengetahui bagaimana perilaku konsumen untuk memenuhi kebutuhannya [1]. Seberapa sering pelanggan berbelanja menentukan perilaku

mereka. Jika seseorang memiliki kebiasaan berbelanja yang kuat, mereka akan membuat keputusan untuk membeli sesuatu, bahkan jika mereka tidak benar-benar berniat untuk membelinya. Kebiasaan berbelanja pelanggan dapat memengaruhi keputusan pembelian mereka. Perusahaan dapat menggunakan kebiasaan berbelanja pelanggan ini untuk mempelajari apa yang dibutuhkan pelanggan [2]. Data Mining telah menjadi populer dan umum digunakan oleh berbagai pihak sebagai metode untuk mendapatkan informasi berharga. Pada saat ini, Data Mining menjadi sangat penting dalam berbagai industri bisnis, termasuk industri retail. Data mining adalah proses yang digunakan untuk menemukan pola penting dalam data yang sangat besar dan tidak terstruktur. Proses ini memungkinkan penggalian dan pemahaman informasi yang sebelumnya tidak diketahui untuk memulai tindakan lanjut yang tepat [3]. Pusat perbelanjaan modern, juga disebut mall, adalah tempat yang memiliki satu atau lebih department store besar yang menarik toko-toko kecil dan restoran. Toko-tokonya menghadap ke koridor utama atau area pejalan kaki, yang menjadi pusat dari pusat perbelanjaan modern. Koridor ini tidak hanya berfungsi sebagai jalur lalu lintas tetapi juga sebagai ruang komunal yang memudahkan interaksi antara pengunjung dan penjual [4].

Pada penelitian ini ingin diketahui suatu hasil simulasi pengetahuan yang dimana hubungan antara usia dan pendapatan terhadap skor belanja dalam proses kegiatan transaksi yang dilakukan di mall nipah. Adapun alur penelitian dengan menggunakan algoritma DBSCAN untuk simulasi pengelompokan berdasarkan usia dan pendapatan dapat dilihat ilustrasi pada Gambar 1, dibawah ini :



Gambar 1 Alur Penelitian clustering

METODE

Kumpulan data yang disimpan dalam memori disebut "dataset". Dataset dapat terdiri dari beberapa tabel yang saling berhubungan (hubungan), yang

memiliki semua atribut, fitur, dan fitur dari basis data biasa [5],[6]. Tabel yang disimpan dalam kumpulan data dapat menggunakan kunci asing dengan integritas referensial. Dataset terdiri dari kumpulan tabel dan data dan memiliki struktur yang mirip dengan data dalam database [7]. Ada dua data jenis dataset, yaitu: private database atau Kumpulan data pribadi adalah kumpulan data yang berasal dari komunitas atau organisasi yang menjadi subjek penelitian. Agensi, rumah sakit, pabrik, perusahaan jasa, dll, dan public dataset atau kumpulan data publik adalah kumpulan data yang tersedia melalui repositori publik yang disetujui oleh para peneliti. Data yang tersedia dari UCI (University of California, Irvine) dan sumber merupakan contohnya.

Tahap pengumpulan data dalam penelitian ini mengimplementasikan algoritma DBSCAN Clustering untuk menganalisis pola kunjungan pengunjung Mall Nipah. Sumber data utama diperoleh melalui aplikasi dashboard administrasi (data member) yang terintegrasi pada laman <https://nipah.id/kuisramadan>. Sampel penelitian terdiri dari 200 responden pengunjung mall yang berpartisipasi dalam program kuis Ramadan selama periode 1 hingga 30 hari bulan Ramadan. Seluruh data dikurasi secara sistematis dengan memperhatikan temporalitas aktivitas partisipan selama satu siklus penuh ibadah puasa.". Berikut ini merupakan dataset yang digunakan untuk melakukan clustering dengan menggunakan DBSCAN, dapat dilihat pada gambar 2, dibawah ini :

CustomerID	Sex	Usia	Gaji (juta)	Skor Belanja (1-100)	kluster	
0	1	Pria	19	15	39	0
1	2	Pria	21	15	81	0
2	3	Wanita	20	16	6	0
3	4	Wanita	23	16	77	0
4	5	Wanita	31	17	40	-1
...
195	196	Wanita	35	120	79	-1
196	197	Wanita	45	126	28	-1
197	198	Pria	32	126	74	-1
198	199	Pria	32	137	18	-1
199	200	Pria	30	137	83	-1

200 rows x 6 columns

Gambar 2. Jumlah Dataset Untuk Proses Clustering

Kelompok yang umum dalam kumpulan data dapat diidentifikasi melalui teknik clustering. Unit biasanya lebih mirip satu sama lain di setiap grup. Segmentasi berbasis cluster telah menjadi sangat populer dalam penelitian data, terutama dalam bidang pemasaran, sejak tahun 1970-an. Meskipun pengelompokan tidak ada dalam metode analisis data terstruktur, itu sangat fleksibel dan bergantung pada jenis data atau sampel yang digunakan [8],[9]. Metode DBSCAN termasuk dalam algoritma unsupervised learning dan merupakan salah satu metode pengelompokan data [10]. Dalam metode DBSCAN, data dikelompokkan menggunakan ide kepadatan data. Metode DBSCAN mengelompokkan data dengan menggunakan parameter epsilon (jarak antar titik) dan MinPts (minimal titik untuk membentuk Cluster). Langkah pertama metode DBSCAN adalah mengambil titik acak untuk digunakan sebagai centroid (pusat Cluster), kemudian menggunakan parameter epsilon untuk menghitung jarak antara titik terdekat dengan jarak geometris [11]. Namun, keunggulan metode ini dibandingkan dengan pendekatan lain, seperti K-Means, adalah kemampuan untuk menemukan data outlier [12],[13]. Data yang dapat diakses oleh epsilon dan memenuhi syarat

minimal point untuk Cluster dimasukkan ke dalam Cluster. Data yang tidak dapat diakses atau tingkat kepadatan datanya tidak memenuhi persyaratan MinPts dianggap sebagai data outlier. Berikut adalah definisi metode DBSCAN [14]:

- a) Epsilon-Neighborhood dari satu titik

$$N_{Eps}(p) = \{q \in D \mid \text{dist}(p, q) < eps\}$$

(1.1)

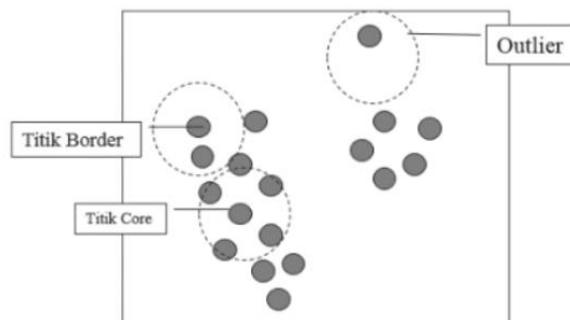
Nilai epsilon menunjukkan bahwa data yang masuk ke dalam Cluster harus memiliki minimal satu data lain yang lebih dekat ke data tersebut.

- b) Directly Density-Reachable

$$p \in N_{Rps}(Q)$$

(1.2)

Jika border point memiliki epsilon neighborhood dari q (core point), border point masih dianggap sebagai anggota Cluster. Adapun ilustrasinya dapat dilihat pada gambar 3. Berikut ini :



Gambar 3 Pengelompokan DBSCAN

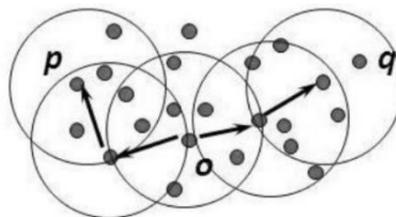
- c) Supaya q point menjadi core point, point tersebut harus memiliki MinPts (minimal point untuk membentuk Cluster). Adapun kondisinya ditunjukkan melalui persamaan berikut :

$$||N_{Eps}(Q)|| \geq MinPts$$

(1.3)

- d) Density-Connected

Jika dua titik perbatasan termasuk dalam suatu Cluster, tetapi mereka tidak membagi core-point tertentu, titik-titik tersebut tidak dianggap dapat mencapai density satu sama lain. Adapun gambarnya dapat dilihat pada gambar 4 berikut :



Gambar 4 Pengelompokan DBSCAN

- e) Cluster

Apabila satu titik termasuk dalam Cluster C dan point q adalah densitas yang dapat dicapai dari point p berdasarkan epsilon dan MinPts yang ditentukan, maka titik q merupakan bagian dari Cluster [15].

- f) Noise adalah data yang tidak dapat dimasukkan ke dalam Cluster mana pun atau tidak memenuhi syarat untuk dimasukkan ke dalam Cluster sesuai dengan parameter epsilon dan MinPts.

PEMBAHASAN

Tahap uji coba awal dilaksanakan dengan menggunakan data kuesioner yang dikumpulkan selama periode satu bulan, data ulasan terdiri dari 200 record, kemudian selanjutnya langkah awal menggunakan metode DBSCAN terlebih dahulu menentukan Eps (ϵ) dan MinPts. Yang dimaksud dengan Eps (ϵ) adalah jarak maksimum dua titik yang dapat ditempuh meskipun masih berada dalam cluster yang sama pada penelitian ini Eps (ϵ) sebesar $\epsilon = 0.3$, sedangkan MinPts adalah sampel jumlah titik paling sedikit yang diperlukan untuk membentuk sebuah Cluster, pada penelitian ini MinPts yang digunakan sebesar MinPts = 10, untuk lebih jelasnya penggunaan Eps (ϵ) dan MinPts dalam bahasa pemrograman python dapat dilihat pada gambar 5 dibawah ini :

```
[36] x = np.array(x)
     XX = StandardScaler().fit_transform(x)
     dbscan = DBSCAN (eps=0.3,min_samples=10)
     dbscan.fit(XX)
```

Gambar 5. Penentuan Nilai Eps (ϵ) dan MinPts

Dalam DBSCAN, penentuan Cluster sepenuhnya bergantung pada ϵ dan MinPts, yang menentukan kepadatan lokal. Cluster yang valid terdiri dari titik inti yang terhubung secara kepadatan, sedangkan titik yang tidak memenuhi kriteria diklasifikasikan sebagai noise. Untuk memastikan kualitas pengelompokan, analisis distribusi Cluster dan metrik seperti Silhouette Score digunakan, pada penelitian jumlah cluster yang diperoleh sebanyak lima cluster mulai dari cluster -1, cluster 0, cluster 1, cluster 2 dan cluster 3, untuk lebih jelasnya bisa dilihat pada gambar 6 dibawah ini :

```
[0 0 0 0 -1 0 -1 0 -1 -1 -1 -1 -1 0 -1 0 -1 0 -1 -1 -1 0 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
 1 -1 1 2 2 1 -1 -1 1 2 -1 -1 -1 -1 -1 2 3 2 2 2 -1 1 1 1
 3 3 3 3 -1 3 3 3 -1 3 1 3 3 3 -1 3 -1 3 -1 3 -1 3 3 3
 3 3 -1 3 3 3 -1 3 -1 3 -1 3 3 3 3 3 -1 3 -1 3 -1 -1 -1 3
-1 3 -1 -1 -1 -1 -1 3 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -1 -1 -1 -1 -1 -1]
```

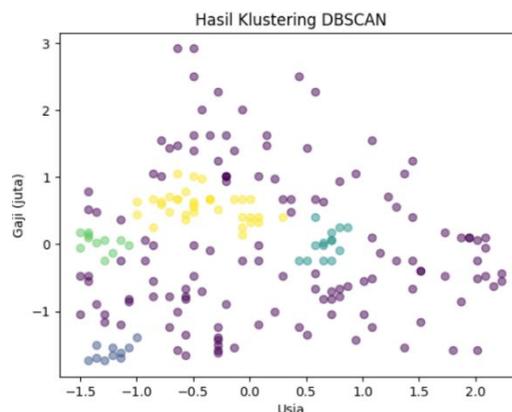
Gambar 6. Jumlah cluster yang diperoleh sebanyak lima (5)

Kelompok cluster yang berisikan data usia dan gaji yang masuk dalam setiap cluster memiliki nilai yang beragam mulai cluster -1 berisikan data sebanyak 130 data, jumlah untuk cluster 0 berisikan data sebanyak 10 data, jumlah untuk cluster 1 berisikan data sebanyak 14 data, jumlah untuk cluster 2 berisikan data sebanyak 11 data dan yang terakhir cluster 3 berisikan data sebanyak 35 data, untuk lebih jelasnya bisa dilihat penjelasan Tabel 1, tentang jumlah dan persentase yang diperoleh dari setiap cluster:

Tabel 1. Jumlah Persentasi data tiap cluster

No	Cluster	Jumlah	Persentase
1	Cluster -1	130	65 %
2	Cluster 0	10	5%
3	Cluster 1	14	7 %
4	Cluster 2	11	5.5%
5	Cluster 3	35	17.5%

Dengan menggunakan data yang berjumlah 200 data dengan menggunakan algoritma DBSCAN didapatkan sebaran data dengan menggunakan library matplotlib yang dimana sebaran data yang berwarna ungu tersebar merupakan cluster -1 dan ini merupakan noise yang tidak membentuk density, density yang artinya metode kerapatan yang digunakan untuk mengelompokkan data yang berdekatan, sedangkan sebaran dari warna kuning adalah merupakan bagian dari cluster 3, untuk sebaran dari warna hijau muda adalah merupakan bagian dari cluster 2, sedangkan sebaran dari warna hijau tua adalah merupakan bagian dari cluster 1 dan sebaran dari warna biru tua adalah merupakan bagian dari cluster 0, untuk lebih jelasnya bisa dilihat pada gambar 7, dibawah berikut ini :



Gambar 7. Sebaran Clustering Menggunakan DBSCAN

KESIMPULAN

Penelitian ini mengimplementasikan algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) untuk menganalisis sebaran data pengunjung mall berdasarkan dua variabel kategorikal: usia, dan pendapatan, Proses Clusterisasi dilakukan terhadap 200 sampel data responden yang dikumpulkan selama periode satu bulan. Berdasarkan analisis kepadatan data, terbentuk lima kelompok Cluster dengan distribusi sebagai berikut: Noise (Cluster -1): 130 data (65% total sampel), merepresentasikan titik data yang tidak terasosiasi dengan kepadatan signifikan, Cluster 0: 10 data (5% total sampel), Cluster 1: 14 data (7% total sampel), Cluster 2: 11 data (5.5% total sampel), Cluster 3: 35 data (17.5% total sampel).

UCAPAN TERIMA KASIH

Ucapan terima kasih penulis haturkan kepada STMIK Profesional Makassar yang telah memberi dukungan selama penelitian.

DAFTAR PUSTAKA

- [1] T. M. Dista and F. F. Abdulloh, “Clustering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization,” *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 3, p. 1339, Jul. 2022, doi: 10.30865/mib.v6i3.4172.
- [2] N. R. Noperahila, “Analisis Pola Perilaku Konsumen dan Kebiasaan Belanja menggunakan Metode K-Means Clustering,” 2024.
- [3] G. Schuh *et al.*, “Data Mining Definitions and Applications for the Management of Production Complexity,” *Procedia CIRP*, vol. 81, pp. 874–879, 2019, doi: 10.1016/j.procir.2019.03.217.
- [4] N. Jayanti, C. G. Selan, and M. W. Prihatmono, “Analisis Data Pengunjung Mall Nipah Mendukung Strategi Digital Marketing Menggunakan K-Means Clustering,” *J. Penelit. Inov.*, vol. 2, no. 1, pp. 209–220, May 2022, doi: 10.54082/jupin.65.
- [5] M. Honda, “Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS,” *J. Inform. dan Teknol.* vol. 2, no. 2, pp. 109–118, 2019.
- [6] M. Faisal and T. K. A. Rahman, “Optimally Enhancement Rural Development Support Using Hybrid Multy Object Optimization (MOO) and Clustering Methodologies: A Case South Sulawesi - Indonesia,” *Int. J. Sustain. Dev. Plan.*, vol. 18, no. 6, pp. 1659–1669, Jun. 2023, doi: 10.18280/ijstdp.180602.
- [7] Bima Yudha Putra, “Klasterisasi Pengunjung Mall Menggunakan Algoritma K-Means Berdasarkan Pendapatan Dan Pengeluaran,” *JITET (Jurnal Inform. dan Tek. Elektro Ter. Vol.11 No.3 S1, pISSN2303-0577 eISSN2830-7062*, 2023.
- [8] O. Dogan, J.-L. Bayo-Monton, C. Fernandez-Llatas, and B. Oztaysi, “Analyzing of Gender Behaviors from Paths Using Process Mining: A Shopping Mall Application,” *Sensors*, vol. 19, no. 3, p. 557, Jan. 2019, doi: 10.3390/s19030557.
- [9] M. Faisal *et al.*, “A Hybrid MOO, MCGDM, and Sentiment Analysis Methodologies for Enhancing Regional Expansion Planning: A Case Study Luwu - Indonesia,” *Int. J. Math. Eng. Manag. Sci.*, vol. 10, no. 1, pp. 163–188, Feb. 2025, doi: 10.33889/IJMEMS.2025.10.1.010.
- [10] J. Zhang, “Advancements of Outlier Detection: A Survey,” *ICST Trans. Scalable Inf. Syst.*, vol. 13, no. 1, p. e2, Feb. 2013, doi: 10.4108/trans.sis.2013.01-03.e2.
- [11] M. Ramadhani, “Implementation of Data Mining Analysis to Determine the Tuna Fishing Zone Using DBSCAN Algorithm,” *Int. J. Mach. Learn. Comput.*, vol. 9, no. 5, pp. 706–711, Oct. 2019, doi: 10.18178/ijmlc.2019.9.5.862.
- [12] T. H. F. khan, N. N. Alleema, N. Yadav, S. Mishra, and A. Shahi, “Text Document Clustering using K-Means and Dbscan by using Machine Learning,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 6327–6330, Oct. 2019, doi: 10.35940/ijeat.A2040.109119.
- [13] M. Faisal and T. K. A. Rahman, “Determining rural development priorities using a hybrid clustering approach: a case study of South Sulawesi, Indonesia,” *Int. J. Adv. Technol. Eng. Explor.*, vol. 10, no. 103, pp. 696–719,



- Jun. 2023, doi: 10.19101/IJATEE.2023.10101215.
- [14] D. Armiady, “Analisis Metode DBSCAN (Density-Based Spatial Clustering of Application with Noise) dalam Mendeteksi Data Outlier,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 6, p. 2158, Dec. 2022, doi: 10.30865/jurikom.v9i6.5080.
- [15] M. Faisal, T. K. A. Rahman, I. Mulyadi, K. Aryasa, Irmawati, and M. Thamrin, “A Novelty Decision-Making Based on Hybrid Indexing, Clustering, and Classification Methodologies: An Application to Map the Relevant Experts Against the Rural Problem,” *Decis. Mak. Appl. Manag. Eng.*, vol. 7, no. 2, pp. 132–171, Feb. 2024, doi: 10.31181/dmame7220241023.