

ANALISIS CLUSTERING DENGAN K-MEANS UNTUK PENGELOMPOKAN PENJUALAN ELEKTRONIK PADA BATAM IT MART

Andrianto¹,
Erlin Elisa²

¹Mahasiswa Program Studi Sistem Informasi, Universitas Putera Batam

²Dosen Program Studi Sistem Informasi, Universitas Putera Batam

email: pb191510002@upbatam.ac.id

ABSTRACT

This study aims to help Batam IT Mart in the process of grouping sales, and to find out data on the types of goods that are in demand and not in demand, with that sales data obtained by the store can be processed into useful information in the future. The method used in this research is the K-means Clustering which is done by grouping data into several groups and then carrying out a calculation process to get results in the form of groups of data on the types of goods that are very hot-selling, best-selling, and not-selling. The data used in this research is electronic sales data at Batam IT Mart from March 2019 to November 2022. The research has resulting 3 clusters such as cluster 0 that contained 19 items, cluster 1 that contained 2 items and cluster 2 that contained 4 items. So the conclusion drawn is the analysis and grouping using k-means clustering method can help in doing the segmentation and grouping of the sales data. From this research, it is obtained renewal in the form of a cluster that can assist in analyzing and providing segments for goods sold at Batam IT Mart.

Keywords: Cluster, Data Mining, Elektronik, K-Means Algorithm

PENDAHULUAN

Pada zaman berkembang pesatnya teknologi, peralatan elektronik sangat dibutuhkan masyarakat. Meski bukan merupakan kebutuhan primer sehari-hari, tetapi peralatan elektronik sudah menunjang keperluan primer dalam kegiatan sehari-hari. Dengan tingginya kebutuhan akan peralatan elektronik, penjualan pada bidang ini pun berkembang dengan sangat baik. Adapun salah satu jenis penjualan yang pesat yaitu penjualan kabel data, di mana dalam keseharian, masyarakat memerlukan

kabel data untuk berbagai kebutuhan khususnya untuk mengisi daya dari telepon genggam. Dari data yang didapatkan dari Kementerian Perindustrian Republik Indonesia, diketahui bahwa sebesar 60% dari perkiraan produk impor yang ada di Indonesia dikuasai oleh sektor elektronik. Indonesia juga mengimpor bahan baku yang dominannya didatangkan dari Tiongkok, Taiwan, Singapura, Korea Selatan, dan Vietnam. Dari tingginya nilai impor yang dilakukan pada sektor tersebut dikarenakan terdapat

peningkatan dalam investasi pada bidang elektronika di Indonesia. Dengan perkembangan penjualan yang baik, data yang didapatkan oleh pihak penjual tentu banyak. Adapun pihak penjual yang dimaksud pada penelitian ini ialah Batam IT Mart. Dengan banyaknya data yang didapat, pihak Batam IT Mart tentu memiliki penyimpanan akan data yang diterima, tetapi data tersebut tidak pernah dilakukan pengolahan lebih lanjut dengan metode algoritma yang mendukung dan hanya disimpan yang bila mana media tempat penyimpanan data telah penuh, data tersebut akan dihapus. Dengan masalah demikian, pada penelitian ini akan dilakukan analisis terhadap data dengan mengaplikasikan metode data mining dengan teknik *clustering*, adapun metode yang peneliti pilih ialah metode *k-means*. Teknik ini dilakukan dengan mengelompokkan data dengan sifat yang menyamai lalu dilakukan pengujian terhadap kelompok yang sudah dibuat. Dari penelitian oleh Dinata, Safwandi, Hasdyna, & Azizah yang dilakukan pada tahun 2020, digunakan metode serupa dan didapatkan bahwa hasil pengujian dengan algoritma *k-means clustering* ini dapat membantu dalam proses pemilihan sepeda motor oleh pengguna sesuai dengan kebutuhannya. Dengan hasil pengujian yang baik tersebut, peneliti bermaksud menggunakan metode yang sama untuk melakukan pengujian pada lokasi penelitian, yakni Batam IT Mart, dengan harapan bahwasanya peneliti dapat membantu permasalahan yang dihadapi oleh lokasi penelitian, seperti mengolah data penjualan yang didapat agar dapat memberikan segmentasi dari penjualan yang terjadi, sehingga data yang didapat tidak hanya disimpan saja

dan tidak dilakukan pengolahan terhadap data.

KAJIAN TEORI

2.1 *Knowledge Discovery in Database (KDD)*

Pada tahun 1989, nama *Knowledge Discovery in Database* pertama kali dikemukakan untuk menandakan keterampilan yang bisa didapatkan dari penambangan data di dalam suatu basis data yang bertujuan untuk menambang data-data penting dari dalam basis data tersebut (Shirazi, Baziyad, & Karimi, 2019).

Terdapat 5 tahapan dari KDD (Alam, Resmi, & Masripah, 2022).

1. *Data Selection*
Pada tahap ini dilakukan pemilihan data dari keseluruhan data yang akan dianalisa lebih lanjut.
2. *Pre-processing*
Pada tahap ini, data akan memasuki proses pembersihan untuk menyaring data-data yang akan digunakan dalam proses perhitungan.
3. *Transformation*
Transformasi dilakukan untuk mengubah format data ke dalam format yang dapat diolah dalam proses *data mining*.
4. *Data Mining*
Tahap ini merupakan tahapan data dilakukan proses perhitungan sesuai dengan algoritma yang telah ditentukan.
5. *Evaluation*
Pada tahapan ini, data hasil pengolahan dievaluasi dengan menggunakan aplikasi pengolah untuk memastikan validnya hasil yang didapat.

2.2 Data Mining

Data Mining merupakan suatu prosedur yang menggunakan data dalam skala yang besar untuk menggali informasi berharga yang terdapat dalam kumpulan data yang nantinya dapat digunakan dalam membantu proses untuk mengambil keputusan yang bersifat krusial (Mahalisa & Arminarahmah, 2022).

2.3 Clustering

Clustering adalah suatu prosedur yang biasa digunakan untuk membagi data ke dalam beberapa kluster dengan sifat yang saling menyamai dan berbeda dengan kluster lainnya (Pradana, Kusumawardani, & Permanasari, 2020).

2.4 K-Means

K-Means bekerja dengan cara melakukan penghimpunan data dengan mencari jarak terdekat dari *centroid* tiap kluster secara berulang dari beberapa kluster yang ada. Perhitungan titik *centroid* dapat dilakukan dengan menggunakan rumus *Euclidean Distance* (Sitinjak, Pangestu, & Sari, 2022).

$$d(x, y) = \sqrt{\sum_{i=0}^n (xi - yi)^2}$$

Rumus 1. Rumus Euclidean Distance

Untuk mencari jarak antara titik objek berupa x dan titik *centroid* berupa y , maka akan dilakukan perhitungan atribut berjumlah i dari hasil kuadrat dari hasil pengurangan titik x dan titik y .

2.5 RapidMiner

Aplikasi *RapidMiner* adalah aplikasi *open-source* yang dirancang memiliki *framework* tersendiri dalam proses penganalisaan data dan dapat dibaurkan ke aplikasi lainnya untuk melakukan kegiatan *data mining* (Kovács & Ghous, 2020).

2.6 Penjualan

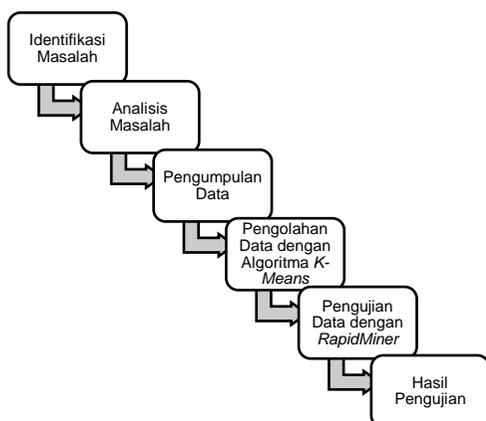
Penjualan adalah kegiatan pertukaran barang antara pihak penjual dan pihak pembeli yang bertujuan untuk mencukupi keperluan dalam keseharian. Kegiatan penjualan dianggap sah apabila terjadi kesepakatan antar kedua belah pihak (Robani, Hadi, Nurdiawan, Dwilestari, & Suarna, 2021).

2.7 Elektronik

Elektronik adalah alat yang dapat melakukan pekerjaannya dengan keakurasian yang tinggi dengan sendirinya sehingga dapat meringankan pekerjaan manual (Nugraha & Hasan, 2019).

METODE PENELITIAN

Penelitian yang dilakukan menggunakan metode penelitian kuantitatif. Gambaran akan metode penelitian yang dilakukan oleh penulis dapat dilihat pada gambar di bawah.



(Sumber: Data Penelitian, 2022)

Gambar 1. Desain Penelitian

HASIL DAN PEMBAHASAN

Hasil dari perhitungan dengan menggunakan algoritma *k-means clustering* didapatkan berupa segmentasi penjualan pada Batam IT Mart yang dikluster ke dalam 3 kluster. Kluster tersebut telah ditetapkan sebelumnya oleh peneliti sebelum dilakukan perhitungan dengan algoritma *k-means clustering*.

Kluster yang terbentuk adalah kluster 0 yang berisi barang yang terjual tidak laris, kluster 1 yang berisi barang

yang terjual laris, dan kluster 2 yang berisi barang yang terjual sangat laris. Dari hasil perhitungan manual didapatkan bahwa kluster 0 terdiri atas 17 segmen, kluster 1 yang terdiri atas 4 segmen dan kluster 2 yang terdiri atas 4 segmen.

Tahap pertama penelitian dilakukan dengan melakukan pra-proses data dengan melakukan penarikan sampel penelitian yang akan dilakukan perhitungan.

Tahapan setelah melakukan pra-proses data adalah melakukan pembersihan dari data yang telah didapat pada tahap pra-proses data. Pembersihan dilakukan dengan melakukan penyaringan terhadap data yang tidak digunakan sehingga menyisakan data yang nantinya digunakan dalam proses perhitungan.

Setelah data dibersihkan, data selanjutnya dilakukan seleksi. Hasil dari seleksi data adalah data yang akan dilakukan proses perhitungan dengan menggunakan algoritma *k-means clustering*.

Sebelum masuk ke dalam tahap *data mining*, data akan melalui tahap transformasi. Pada penelitian ini, data yang didapat sudah dalam bentuk angka sehingga data tidak mengalami perubahan. Hasil dari tahap-tahap sebelumnya dapat dilihat pada tabel di bawah.

Tabel 1. Data Penelitian

No	Nama Barang	2019	2020	2021	2022
1	LIGHT BULB	12	4	6	28
2	5050 5V USB LIGHT	20	17	15	0
3	LIGHT CLIP/4 PCS	12	20	6	1

4	3AAA BATTERY	84	14	11	2
5	3.5 M/ 6.5F	18	16	18	10
6	3.5 M/M CABLE	144	23	38	49
7	Cat6e Cable	83	13	170	90
8	DC 5.5 CABLE	4	32	12	10
9	HDMI F/F	27	25	13	5
10	HDMI TO MINI + MICRO HDMI	7	33	4	5
11	IPHONE CABLE	202	57	47	58
12	LED CABLE TYPE C	14	7	0	9
13	METAL CABLE TYPE-C	19	32	0	0
14	ROUND CABLE 20CM	8	18	3	8
15	ROUND CABLE 30CM	101	9	11	7
16	USB 2.0 M/F	27	35	22	42
17	UNIVERSAL BATTERY CHARGER	66	13	13	28
18	USB 3.4A CHARGER MJ-A05	15	62	0	11
19	16GB USB	30	0	8	2
20	32GB USB	32	53	2	7
21	CARD READER	52	132	90	8
22	HEADSET	149	0	84	62
23	OTG TYPE C 16 GB	6	10	12	3
24	SANDISK ULTRA MICROSDXC 64 GB	20	9	20	23
25	ON/OFF ADAPTER	18	93	13	14

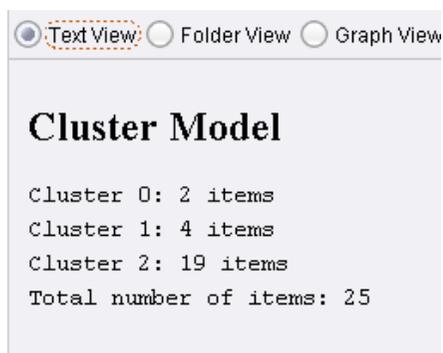
Setelah tahap transformasi data, kemudian akan ditentukan titik *centroid*. Titik *centroid* ditentukan dengan mengambil data dengan tingkat penjualan terendah (C0), menengah (C1), dan tertinggi (C2), sehingga didapatkan titik *centroid* pada data nomor 12 (C0), 18 (C1), dan 11 (C2). Dengan terbaginya ketiga kluster tersebut, didapatkan kluster 0 dengan anggota sebanyak 19 segmen, kluster 1 dengan anggota sebanyak 2

segmen, dan kluster 2 dengan anggota sebanyak 4 segmen. Perhitungan dilakukan dengan rumus *Euclidean Distance* dan dihitung secara berulang hingga didapatkan hasil yang sama. Pada penelitian ini, perhitungan dilakukan sebanyak 6 kali untuk mendapatkan hasil perhitungan yang sama, nilai titik *centroid* yang telah sama dapat dilihat pada tabel di bawah.

Tabel 2. Titik *Centroid* iterasi keenam

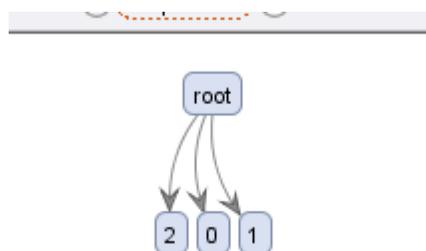
c0	27.47	21.53	9.26	10.58
c1	35	112.5	51.5	11
c2	144.5	23.25	84.75	64.75

Untuk memastikan valid atau tidaknya hasil perhitungan yang dihitung secara manual, digunakan perangkat lunak pembantu berupa *RapidMiner* untuk melakukan perhitungan. Hasil yang didapat dari perangkat lunak *RapidMiner* dapat dilihat di bawah.



Gambar 2. Text View

Terdapat perbedaan urutan pada kluster dari perhitungan manual dengan aplikasi disebabkan karena pada perhitungan manual, titik *centroid* yang diambil tidak sesuai pada urutan data. Pada aplikasi, kluster 0 merupakan kluster barang yang terjual laku, kluster 1 merupakan kluster barang yang terjual sangat laku, dan kluster 2 merupakan kluster barang yang terjual tidak laku.



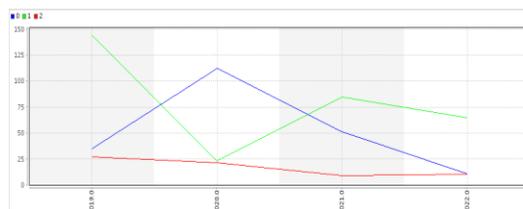
Gambar 3 Graph View

Pada *graph view*, ditampilkan jumlah kluster hasil perhitungan sebanyak 3 buah kluster.

Attribute	cluster_0	cluster_1	cluster_2
2019.0	35	144.500	27.474
2020.0	112.500	23.250	21.526
2021.0	51.500	84.750	9.263
2022.0	11	64.750	10.579

Gambar 4. Centroid Table

Pada *Centroid Table*, ditampilkan angka yang menjadi titik *centroid* yang sudah sama dari perhitungan.



Gambar 5. Centroid Plot View

Grafik yang ditampilkan pada *Centroid Plot View* adalah titik lokasi nilai yang ditampilkan pada *centroid table*.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -1841.762
Avg. within centroid distance_cluster_0: -2160.500
Avg. within centroid distance_cluster_1: -5177.812
Avg. within centroid distance_cluster_2: -1105.884
Davies Bouldin: -0.742
```

Gambar 6. Performance Vector

Pada bagian *Performance Vector* ditampilkan nilai rata-rata antar *centroid*. Selain nilai rata-rata, ditampilkan juga nilai optimum dari perhitungan sebesar 0,742.

SIMPULAN

Setelah penelitian dilakukan, penulis dapat ditarik kesimpulan berupa:

1. Pengelompokan dan analisa pada penjualan dengan metode *data mining* dengan menggunakan algoritma *k-means clustering* dapat digunakan untuk memberikan segmentasi pada barang-barang yang dijual pada Batam IT Mart. Hasil yang didapat dari pengujian berupa kluster barang yang terjual tidak laris, laris dan sangat laris.
2. Kluster yang didapat dari perhitungan dari data yang terbentuk menggunakan algoritma *k-means clustering* berupa kluster 0 yang berisi barang yang terjual tidak laris sebanyak 19 data, kluster 1 yang berisi barang yang terjual laris sebanyak 2 data, dan kluster 2 berisi barang yang terjual sangat laris sebanyak 4 data.

DAFTAR PUSTAKA

Alam, S., Resmi, M. G., & Masriyah, N. (2022). Classification of Covid-19

vaccine data screening with Naive Bayes algorithm using Knowledge Discovery in database method. *Journal of Computer Networks, Architecture and High Performance Computing*, 4(2), 177–185. <https://doi.org/10.47709/cnahpc.v4i2.1584>

Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis K-Means Clustering pada Data Sepeda Motor. *INFORMAL: Informatics Journal*, 5(1), 10. <https://doi.org/10.19184/isj.v5i1.17071>

Kovács, L., & Ghous, H. (2020). Efficiency comparison of Python and RapidMiner. *Multidiszciplináris Tudományok*, 10(3), 212–220. <https://doi.org/10.35925/j.multi.2020.3.26>

Mahalisa, G., & Arminarahmah, N. (2022). Diabetes Classification Analysis Using the Euclidean Distance Method Based on the K-Nearest Neighbors Algorithm. *Jurnal Teknologi Komputer dan Sistem Informasi*, 5(3), 178–182.

Nugraha, A. R., & Hasan, A. (2019). Kendali Perangkat Elektronik Menggunakan Aplikasi Berbasis Web Menggunakan Arduino. *Jumantaka*, 03(1), 11–21. Diambil dari <http://jurnal.stmik-dci.ac.id/index.php/jumantaka/article/view/364>

Pradana, C., Kusumawardani, S. S., & Permanasari, A. E. (2020). Comparison Clustering Performance Based on Moodle Log Mining. *Pradana, C Kusumawardani, S S Permanasari, A E*, 722(1), 1–11.

<https://doi.org/10.1088/1757-899X/722/1/012012>

Robani, A. M., Hadi, S., Nurdiawan, O., Dwilestari, G., & Suarna, N. (2021). Sistem Informasi Penjualan Motor Bekas Berbasis Android Untuk Meningkatkan Penjualan di Mokascirebon. *Com. JURIKOM (Jurnal Riset Komputer)*, 8(6), 205–212.

<https://doi.org/10.30865/jurikom.v8i6.3629>

Shirazi, S., Baziyad, H., & Karimi, H. (2019). An Application-Based Review of Recent Advances of Data Mining in Healthcare. *J Biostat Epidemiol.*, 5(4), 268–278.

Sitinjak, D. K., Pangestu, B. A., & Sari, B. N. (2022). Clustering Tenaga Kesehatan Berdasarkan Kecamatan di Kabupaten Karawang Menggunakan Algoritma K-Means. *Journal of Applied Informatics and Computing*, 6(1), 47–54.
<https://doi.org/10.30871/jaic.v6i1.3855>



Biodata

Penulis kedua, Erlin Elisa, S.Kom., M.Kom., merupakan Dosen Prodi Sistem Informasi Universitas Putera Batam. Penulis banyak berkecimpung di bidang Sistem Informasi



Biodata

Penulis pertama, Andrianto, merupakan mahasiswa Prodi Sistem Informasi Universitas Putera Batam.