

PENERAPAN DATA MINING MENGGUNAKAN METODE K-MEANS CLUSTERING PADA PENJUALAN HANDPHONE DAN ELEKTRONIK (STUDI KASUS : CV REY GASENDRA)

Tria Setyani¹⁾, Setiawansyah²⁾, A. Ferico Octaviansyah³⁾, Rio Andika⁴⁾

^{1,3,4}Teknologi Informasi, Univeritas Teknokrat Indonesia

²Informatika, Universitas Teknokrat Indonesia

^{1,2,3,4}Jl. H. ZA. Pagar Alam No. 90-11 Kedaton, Bandarlampung

Email: ²setiawansyah@teknokrat.ac.id, ³fericopasaribu@teknokrat.ac.id, ⁴Rio_Andika@teknokrat.ac.id

Abstrak

Abstrak CV Rey Gasendra merupakan salah satu toko handphone dan elektronik yang bergerak dalam penjualan berbagai macam merk handphone, mesin cuci, kulkas, tv, kipas angin dan barang elektronik lainnya. Pengelompokan data penjualan pada CV Rey Gasendra masih dilakukan dengan cara manual di excel. Cara pengelompokan tersebut membutuhkan waktu dan memungkinkan data hilang. Dalam melakukan pengelompokan data diperlukan histori data penjualan, apabila data penjualan besar maka di perlukan teknik data mining, salah satu teknik data mining yaitu dengan menggunakan metode K-means Clustering. Clustering merupakan salah satu metode data mining yang bersifat unsupervised dan K-Means adalah salah satu metode clustering non-hirarki yang berusaha membagi data yang ada ke dalam bentuk satu atau lebih kelompok. Metode K-Means clustering dapat diterapkan untuk mengelompokkan suatu data penjualan berdasarkan jenis barang, tipe pelangan, atau jumlah barang. Data yang digunakan adalah data penjualan pada bulan Desember-Februari sebanyak 2027 data. Hasil pengujian dilakukan dengan menggunakan aplikasi RapidMiner dimana hasilnya terdapat 2 cluster yaitu cluster 0 berjumlah 102 data traksaksi dan cluster 1 berjumlah 1925 data transaksi.

Kata Kunci: Data mining, K-Means, Clustering, CV Rey Gasendra, Penjualan

1. Pendahuluan

Dengan pesatnya perkembangan teknologi informasi saat ini, kebutuhan akan informasi yang akurat semakin penting dalam kehidupan sehari-hari. Informasi kini menjadi elemen krusial dalam perkembangan masyarakat baik saat ini maupun di masa depan. Namun, seringkali kebutuhan informasi yang tinggi tidak diimbangi dengan penyajian informasi yang memadai; informasi sering kali harus digali kembali dari kumpulan data yang sangat besar. CV Rey Gasendra adalah toko yang menjual berbagai merk handphone, mesin cuci, kulkas, televisi, kipas angin, dan barang elektronik lainnya. Di CV Rey Gasendra, terdapat masalah yang berulang terkait penjualan, terutama dalam hal pengelompokan data penjualan. Untuk pengelompokan data yang efektif, diperlukan histori data penjualan. Ketika jumlah data

penjualan sangat besar, teknik data mining diperlukan. Penelitian ini menggunakan metode K-Means clustering untuk mengelompokkan data berdasarkan barang yang paling laku dan kurang laku. K-Means adalah salah satu metode dalam data mining yang melakukan pengelompokan data (*clustering*) [1].

Penggunaan data mining dapat membantu dalam menganalisis data transaksi untuk mengidentifikasi pola penjualan barang di CV Rey Gasendra. Data Mining adalah teknik yang berkaitan dengan penggalian informasi dari data yang biasanya berupa pola tersembunyi dan hubungan antara elemen data, serta model yang diperlukan [2]. Dalam penelitian ini, analisis data mining dilakukan dengan teknik clustering menggunakan metode K-Means. Teknik clustering adalah metode dalam data mining yang memungkinkan pengguna untuk mengelompokkan sejumlah data atau objek ke dalam *cluster* (kelompok). Pada akhir proses, setiap *cluster* akan berisi data yang mirip dan berbeda dari objek di cluster lainnya [3]. Metode K-Means adalah algoritma pengelompokan iteratif yang membagi set data ke dalam sejumlah cluster yang telah ditetapkan sebelumnya. Algoritma K-Means dikenal karena kesederhanaannya dalam implementasi, kecepatan, kemudahan adaptasi, dan umum digunakan dalam praktik [4].

Algoritma K-Means Clustering pernah diterapkan untuk pengelompokan data penjualan aksesoris [5], pengelompokan data pembelian tinta [6], penjualan sparepart Toyota [1], dan analisis pada penyakit menular manusia [7]. Hasil dari beberapa penelitian terdahulu dengan menerapkan metode K-Means Clustering menunjukkan bahwa metode tersebut mampu memberikan keakurasian dalam mengelompokkan data sesuai kebutuhan.

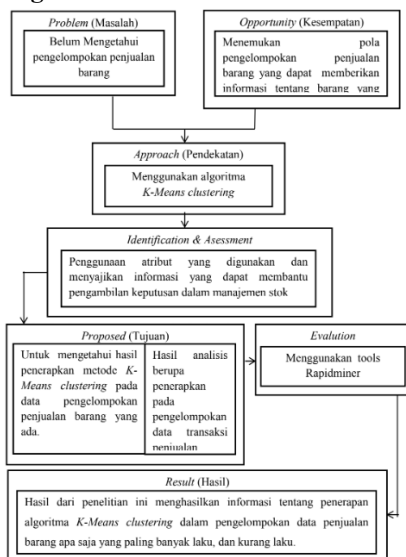
Dalam uraian tersebut, penulis membahas penerapan pengolahan data penjualan, khususnya mengenai masalah clustering, dengan judul 'Penerapan Data Mining Menggunakan Metode K-Means Clustering pada Handphone dan Elektronik di CV Rey Gasendra.'

2. Metode Penelitian

Pada penelitian ini metode penelitian yang digunakan adalah kuantitatif. Kuantitatif merupakan penelitian yang

datanya berupa angka dan dianalisis menggunakan teknik statistik.

2.1 Kerangka Penelitian



Gambar 1. Kerangka Penelitian

a) Masalah

Langkah awal dalam penelitian ini adalah mengidentifikasi masalah penelitian, yaitu belum diketahui pola pengelompokan penjualan barang di CV Rey Gasendra.

b) Kesempatan

Peluang yang ditemukan adalah menentukan pola pengelompokan penjualan barang untuk mengidentifikasi barang-barang yang paling banyak terjual dengan menggunakan algoritma K-Means clustering.

c) Pendekatan

Penelitian ini menggunakan pendekatan dengan menerapkan algoritma K-Means clustering untuk menemukan pola pengelompokan penjualan barang.

d) Identifikasi & Penilaian

Tahap identifikasi dalam penelitian ini melibatkan atribut yang akan digunakan, sehingga informasi yang dihasilkan dapat sesuai dengan tujuan, yaitu menyiapkan aturan untuk pengelompokan data transaksi penjualan barang.

e) Usulan

Usulan penelitian ini adalah untuk mengetahui hasil penerapan metode K-Means clustering pada data pengelompokan penjualan barang.

f) Evaluasi

Dalam penelitian ini, evaluasi dilakukan menggunakan tools RapidMiner.

g) Hasil

Penelitian ini diharapkan menghasilkan informasi tentang penerapan algoritma K-Means clustering dalam pengelompokan data penjualan barang, untuk mengetahui

barang apa saja yang paling laku dan yang kurang laku.

2.2 Prosedur Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data primer dan sekunder. Data sekunder merupakan data yang didapat secara langsung dari tempat studi kasus. Data primer yang digunakan dalam penelitian ini adalah data transaksi penjualan dari bulan Desember 2020 – Februari 2021. Sedangkan data sekunder yaitu melakukan kajian pustaka tentang penerapan algoritma K-Means clustering dari jurnal, buku dan artikel.

2.3 Metode Pengumpulan Data

a) Observasi Langsung

Pengumpulan data melalui observasi dilakukan dengan mengamati secara langsung tempat penelitian dan melihat proses transaksi yang terjadi di sana.

b) Wawancara

Wawancara digunakan sebagai instrumen untuk memperoleh data dan informasi secara lisan dari responden.

c) Dokumentasi

Dokumentasi merupakan tahap penting dalam pengumpulan data. Dokumentasi dilakukan untuk memperoleh data yang valid dan sesuai dengan kebutuhan penelitian, yaitu dengan mendokumentasikan data transaksi penjualan dari Desember 2020 hingga Februari 2021.

3. Hasil dan Pembahasan

Data yang dikumpulkan terdiri dari data transaksi penjualan dan data stok persediaan dari CV Rey Gasendra dalam format file Excel. Penelitian ini menggunakan algoritma K-Means Clustering untuk menghasilkan prediksi barang yang paling laku dan kurang laku di CV Rey Gasendra. Setelah melakukan proses data mining dengan bantuan aplikasi RapidMiner, penelitian ini berhasil memprediksi penjualan barang dari data transaksi periode Desember 2020 hingga Februari 2021. Hasilnya menunjukkan bahwa ada 102 jenis barang yang paling laku dan 1925 jenis barang yang kurang laku.

3.1 Data Selection

Data yang digunakan dalam penelitian ini adalah data penjualan barang, khususnya handphone dan elektronik, dari CV Rey Gasendra selama periode Desember 2020 hingga Februari 2021. Agar data dapat diolah menggunakan *software* RapidMiner, dilakukan preprocessing secara manual. Atribut yang tidak relevan, seperti tanggal, faktur, kode, jumlah beli, harga, dan diskon, akan dihapus. Data transaksi awal yang digunakan pada penelitian ini dapat dilihat pada gambar 2.

Tanggal	Kode	Faktur	Nama	Jumlah	Sat.	Diskon
02/01/2021	401B70 072727 82	I/SL- 009534	TV COOCOA 40S6G	1	PCS	0
02/01/2021	508243 727212 313	I/SL- 009535	TV CHANGHONG 24	1	PCS	0
	674874 847647 4		TV CHANGHONG 24	1	PCS	0
	1032		BRACKET TV STANDAR	1	PCS	0
	19-32		BRACKET TV STANDAR	1	PCS	0
02/01/2021	8,68061 E+14	I/SL- 009536	VIVO Y12S 3/32	1	PCS	0
02/01/2021	8,60418 E+14	I/SL- 009537	NOTE 9 PRO 6/64	1	PCS	0
02/01/2021	8,63802 E+14	I/SL- 009538	NOTE 9 4/64	1	PCS	0
02/01/2021	3,64924 E+11	I/SL- 009539	MESIN CUCI POLYTRON PMW 9366	1	PCS	0
02/01/2021	8,37645 E+14	I/SL- 009540	TV SAMSUNG LED 32	1	PCS	0
02/01/2021	3,57281 E+14	I/SL- 009541	INFINIX SMART 5 2/32	1	PCS	0
02/01/2021	LAZ86 699889 99	I/SL- 009542	SETRIKA COSMOC C15- 438	1	PCS	0

Gambar 2. Data transaksi awal

Pada penelitian ini data yang diperlukan hanya data transaksi penjualan yang terjadi pada CV Rey Garendra, maka dilakukan penyaringan data pada data tersebut. Penyaringan dilakukan dengan menghapus data transaksi yang bukan transaksi penjualan. Berikut contoh data transaksi yang siap digunakan dapat dilihat pada Gambar 3.

TRANSAKSI	NAMA
1	TV COOCOA 40S6G
2	TV CHANGHONG 24
	TV CHANGHONG 24
	BRACKET TV STANDAR
3	VIVO Y12S 3/32
4	NOTE 9 PRO 6/64

Gambar 3. Data yang sudah diseleksi

3.2 K-Means Clustering

Dalam proses *Knowledge Discovery in Database* (KDD), seleksi data adalah tahap di mana himpunan data dikumpulkan dan data target pada sampel data diciptakan, sehingga penemuan berikutnya dapat dilakukan. Hasil seleksi disimpan dalam berkas yang terpisah dari basis data operasional. Setelah memperoleh data dari CV Rey Gasendra, dilakukan proses seleksi data untuk memudahkan perhitungan dengan algoritma *K-Means clustering*. Data sampel yang akan digunakan pada penelitian ini dapat dilihat pada Gambar 4.

KODE Barang	NAMA	Stok Awal	Terjual
KB001	MESIN CUCI LG T2108 8KG	10	6
KB002	MI NOTE 10 8/256	2	0
KB003	MESIN CUCI SAMSUNG 2 TABUNG 7,5 KG	4	1
KB004	KULKAS POLYTRON 18BN	10	3
KB005	REALME 7PRO 8/128	4	1
KB006	BLENDER PHILIPS	1	1
KB007	REALME C17 6/256	33	21
KB008	VIVO Y30I	41	24
KB009	VIVO Y20S 8/128	18	12
KB010	KULKAS SHARP 2 PINTU 197	6	0
KB011	TV LG 43N7000 UHD	2	1
KB012	TV LED LG 24	18	1
KB013	BOX FREEZER 200 LITER	3	1
KB014	TV SAMSUNG 50 TU6900	3	0
KB015	INFINIX SMART 5 2/32	63	10
KB016	BRACKET TV STANDAR	254	111
KB017	VIVO Y20 3/64	34	19
KB018	KEYBOARD MINI	30	18
KB019	SETRIKA PHILIPS	10	2
KB020	AC AQUA 1PK	3	0
KB021	REDMI 9A 2/32	95	26
KB022	MESIN CUCI SHARP 68MW	30	12
KB023	KULKAS LG 2 PINTU 185	4	0
KB024	BOX FREEZER SHARP 300L	2	1
KB025	MESIN CUCI SHARP 1 TABUNG 806	6	0
KB026	AC PANASONIC 1,5 PK	1	1
KB027	TV TCL 32IN	1	0
KB028	SAMSUNG A20S 4/64	2	2
KB029	TV LG 50 SMART UN70	1	1
KB030	KULKAS FREZER LG 304 SL	1	0
KB031	ADVAN TAB	10	3
KB032	TV SAMSUNG LED 24	39	14
KB033	V20 SE 8/128	4	1

Gambar 4. Data Sampel

3.3 Menentukan Titik Pusat Centroid Secara Random

Pada tahap ini dilakukan proses utama yaitu pengelompokan data penjualan barang yang diakses dari database, yaitu metode clustering algoritma K-Means. Dari banyak data penjualan yang diperoleh, sampel percobaan dilakukan dengan menggunakan parameter-parameter berikut :

- Jumlah cluster : 2
- Jumlah data : 8
- Jumlah atribut : 2

Setelah menentukan cluster sebanyak 2 cluster, selanjutnya ditentukan titik pusat secara random. C0 diambil dari nomor 1, C1. Tampilan penentuan centroid secara random dapat dilihat pada Gambar 5.

KODE Barang	SA	ST
KB001	10	6
KB002	2	0
KB003	4	1
KB004	10	3
KB005	4	1
KB006	1	1
KB007	33	21
KB008	41	24
KB009	18	12
KB010	6	0
KB011	2	1
KB012	18	1
KB013	3	1
KB014	3	0
KB015	63	10
KB016	254	111
KB017	34	19
KB018	30	18

Gambar 5. Penentuan Centroid secara Random

3.4 Menghitung Jarak Terdekat Centroid

Menghitung Jarak dari Centroid Menghitung jarak antara titik centroid dengan titik tiap objek dengan menggunakan Euclidian Distance. Adapun penghitungan centroid awal secara manual. Perhitungannya adalah sebagai berikut:

$$D(i,f) = \sqrt{(X_{i1} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Setelah dilakukan perhitungan berdasarkan rumus diatas, maka hasil perhitungan selengkapnya sebagai berikut:

Centroid	SA	ST
C0	10	6
C1	10	3

Gambar 6. Hasil Centroid Iterasi-1

Kode Barang	SA	ST	JC0	JC1	C0	C1
KB001	10	6	0	3	0	
KB002	2	0	10	8,544		1
KB003	4	1	7,81025	6,32456		1
KB004	10	3	3	0		1
KB005	4	1	7,81025	6,32456		1
KB006	1	1	10,2956	9,21954		1
KB007	33	21	27,4591	29,2062	0	
KB008	41	24	35,8469	21		1

Gambar 7. Perhitungan Jarak dan Pengelompokan Data Iterasi Ke-1

Iterasi selanjutnya melakukan perhitungan lagi dengan menggunakan titik centroid yang baru iterasi ke-2. Dari centroid baru iterasi ke-2, dilakukan perhitungan kembali, sehingga didapatkan hasil Tabel Perhitungan Jarak dan Pengelompokan Data Iterasi Ke-2 dimana C1 memiliki 2 data, C2 memiliki 6 data.

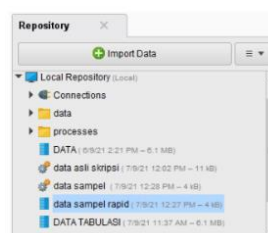
Dari centroid baru iterasi ke-3, dilakukan perhitungan kembali, sehingga didapatkan hasil Tabel Perhitungan Jarak dan Pengelompokan Data Iterasi Ke-3 dimana C0 memiliki 2 data, dan C1 memiliki 6 data. Karena pada iterasi ke-3 posisi cluster tidak berubah atau sama dengan posisi cluster pada iterasi ke-2 pada tabel 4.11, maka proses perhitungan iterasi tidak dilanjutkan.

3.5 Implementasi RapidMiner

RapidMiner adalah software data mining yang digunakan untuk mengolah dataset dan menemukan pola data sesuai dengan tujuan pengolahan. Tidak semua algoritma yang tersedia cocok untuk setiap dataset, sehingga diperlukan penyesuaian pola data agar sesuai dengan tujuan pengolahan.

3.6 Import Data Sampel ke RapidMiner

Pada tahap ini dilakukan persiapan data yang akan diolah, yaitu data sampel penjualan yang didapatkan dari bulan Desember 2020-Februari 2021. Kemudian data format tabular tersebut di import ke dalam local repository yang ada pada rapidminer. Seperti pada gambar 8.



Gambar 8. Import Data Sampel ke Repository

3.7 Desain Operator Data Sampel

Pada tahap ini, proses yang dilakukan melibatkan drag and drop tabel data ke dalam proses. Selanjutnya, tambahkan operator Select Attributes untuk menghubungkan dengan data yang ada. Kemudian, tambahkan operator K-Means dengan Minimum Support 2 dan isi Max Optimization Steps dengan 10. Setelah semua operator ditambahkan, hubungkan seluruh operator tersebut sesuai dengan gambar 9.



Gambar 9. Desain Operator Data Sampel

Setelah proses running, akan tampil *Example Set (Read Excel)*, seperti yang terlihat pada Gambar 10.

Gambar 10. Tabel Cluster

Gambar 10 menampilkan hasil data penjualan dalam bentuk text view, dapat dilihat bahwa cluster 0 berjumlah 8 items dan cluster 1 berjumlah 0 items, dari total data sampel penjualan 8 items.

Cluster Model

```
Cluster 0: 8 items
Cluster 1: 0 items
Total number of items: 8
```

Gambar 10. Tabel Hasil Model Cluster

3.8 Hasil Clustering

1. Perhitungan Manual menggunakan Data Sampel

Berdasarkan hasil pengelompokan dari seluruh data menggunakan metode K-Means clustering, didapatkan hasil akhir pengelompokan hingga iterasi ke-3, dimana titik pusat tidak lagi berubah dan tidak ada data yang 48 berpindah antar cluster. Hasil dari perhitungan manual didapatkan C0 dengan 2 data dan C1 dengan 6 data. Hasil cluster perhitungan manual dapat dilihat pada gambar 11.

Nama Cluster	Nama Barang
C0 (Barang Paling Laku)	REALME C17 6/256, VIVO Y30i
C1 (Barang Kurang Laku)	MESIN CUCI LG T2108 8KG, MI NOTE 10 8/256, MESIN CUCI SAMSUNG 2 TABUNG 7,5 KG, KULKAS POLYTRON 18BN, REALME 7PRO 8/128, BLENDER PHILIPS

Gambar 11. Perhitungan Manual menggunakan data sampel

2. Perhitungan menggunakan Tool Rapidminer dengan Data Sampel

Berdasarkan hasil pengelompokan dari seluruh data menggunakan metode K-Means clustering, didapatkan hasil akhir pengelompokan dari perhitungan menggunakan tools Rapidminer didapatkan C0 dengan 8 data dan C1 dengan 0 data. Hasil perhitungan menggunakan tools dapat dilihat pada Gambar 7.

Nama Cluster	Nama Barang
C0 (Barang Paling Laku)	REALME C17 6/256, VIVO Y30i, MESIN CUCI LG T2108 8KG, MI NOTE 10 8/256, MESIN CUCI SAMSUNG 2 TABUNG 7,5 KG, KULKAS POLYTRON 18BN, REALME 7PRO 8/128, BLENDER PHILIPS
C1 (Barang Kurang Laku)	-

Gambar 7. Hasil Perhitungan menggunakan Rapid Miner

4. Penutup

Berdasarkan analisis yang dilakukan dalam penelitian dengan menggunakan Algoritma K-Means Clustering, dapat disimpulkan bahwa metode ini berhasil diterapkan untuk mengelompokkan data penjualan. Hasil penelitian ini memudahkan CV Rey Gasendra dalam menentukan data penjualan. Analisis data penjualan untuk periode Desember 2020 hingga Februari 2021 dengan perhitungan manual menggunakan 8 data sampel menunjukkan bahwa terdapat 2 data dalam cluster 0 (C0) dan 6 data dalam cluster 1 (C1), dengan proses iterasi manual sebanyak 3 kali. Sementara itu, menggunakan alat RapidMiner 9.2 dengan 8 data sampel, seluruh data masuk dalam cluster 0 (C0). Dengan alat RapidMiner yang menggunakan 2027 data, hasil pengelompokan menunjukkan 102 data dalam cluster 0 (C0) yang berlabel barang paling laku, dan 1925 data dalam cluster 1 (C1) yang berlabel barang kurang laku.

Daftar Pustaka

[1] Saut PT, Felix TK & Feryanto. (2019). “Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering”. Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA), 2(2), 67-72.

[2] Adinugroho, S. & Sari, Y., 2018. Implementasi Data Mining Menggunakan WEKA. Malang: UB Press.

[3] Ahmad LR & Hafiz BF. (2018). “Clustering Application for UKT Determination Using Pillar K-Means Clustering Algorithm and Flask Web Framework”. Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM), 1(2), 53-59.

[4] Iin P, Agus PW, Anjar W & M. Ridwan L. (2018). “Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Assessment Center Untuk Clustering Program SDP”. CESS (Journal of Computer Engineering, System and Science), 3(1), 87-93.

[5] Mardalius, Mardalius. (2018). “Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means.” Jurteksi 4(2): 123–32.

[6] Susliansyah, Susliansyah, Heny Sumarno, Hendro Priyono, and Noer Hikmah. (2019). “Pengelompokkan Data Pembelian Tinta Dengan Menggunakan Metode K-Means.” J-SAKTI (Jurnal Sains Komputer dan Informatika) 3(2): 381.

[7] A. Bastian, H. Sujadi, & G. Febrianto. (2018). “Penerapan Algoritma KMeans Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka),” J. Sist. Inf., vol. 14, no. 1, pp. 26–32.