

Pemanfaatan Fitur Analisis Data Menggunakan K-Means Cluster Dalam Point of Sales (POS)

Supangat
Program Studi Teknik Informatika
Universitas 17 Agustus 1945 Surabaya
supangat@untag-sby.ac.id

Anis R. Amna
Program Studi Teknik Informatika
Universitas 17 Agustus 1945 Surabaya
anis.r.amna@gmail.com

Abstrak – Pemanfaatan *big data* untuk meningkatkan performa usaha banyak menjadi pembahasan penelitian akhir-akhir ini. Ketersediaan data yang mampu diakses secara cepat untuk mereproduksi informasi baru yang penting dalam pengambilan keputusan menjadi salah satu faktor kunci yang menentukan keberhasilan organisasi. *K-Means cluster* sebagai salah satu algoritma *data mining* dengan kemampuan pengelompokan data merupakan salah satu *tools* yang penting untuk melakukan hal ini. Penelitian ini akan membahas implementasi algoritma *K-Means* untuk menghasilkan informasi baru berupa klasifikasi kelompok produk berdasarkan data transaksi penjualan di masa lalu. Algoritma ini selanjutnya akan menjadi fitur pada Sistem Informasi *Point of Sales* (POS) yang dikembangkan. Dengan adanya fitur baru pada sistem informasi POS, diharapkan pemilik usaha dapat merencanakan jumlah dan waktu pembelian produk dengan lebih baik, mengurangi jumlah persediaan barang di gudang, dan memberikan keleluasaan bagi pemilik usaha untuk menganalisis permintaan dengan mudah.

Kata Kunci: *Knowledge Management, K-Means Clustering, Rancang Bangun Sistem Informasi, Sistem Informasi Eksekutif, Use Case Driven Model.*

I. PENDAHULUAN

Pemanfaatan data dan informasi untuk meningkatkan kinerja perusahaan telah ramai menjadi topik pembahasan di bidang *knowledge management* (KM). Hal ini terjadi karena KM dianggap memainkan peran penting untuk meningkatkan nilai kolaborasi bisnis di dalam perusahaan. Pemanfaatan teknologi berbasis data seperti *robotics*, kecerdasan buatan, serta digitalisasi memungkinkan pelanggan untuk memperoleh pelayanan personal serta mengoptimalkan proses bisnis perusahaan pada saat bersamaan [1].

Dalam pembahasan lain, pemanfaatan KM dalam bisnis dipercaya mempengaruhi proses pengambilan keputusan secara rasional, yang pada akhirnya mampu meningkatkan performa organisasi. Hal ini disebabkan karena KM memungkinkan adanya kolaborasi antar berbagai disiplin ilmu yang didukung oleh pemanfaatan teknologi informasi dalam berbagai bentuk serta kemampuan perusahaan untuk

terus belajar mengoptimalkan proses bisnis yang ada sehingga melahirkan pengetahuan baru yang lebih baik [2].

Kemampuan perusahaan dalam memahami kebutuhan pelanggan merupakan salah satu faktor penting yang mempengaruhi kinerja organisasi. Tanpa pemahaman yang menyeluruh tentang keinginan pelanggan dan kondisi perusahaan, pemilik usaha akan mengambil keputusan menggunakan intuisinya. Oleh karena itu, fitur *data mining* dalam sistem informasi manajemen diperlukan untuk mengungkap informasi yang tersembunyi dalam data transaksi yang telah ada yang pada akhirnya memungkinkan pemilik usaha untuk membuat keputusan berdasarkan data dan mentransformasikannya dalam bentuk kebijakan bisnis yang bermanfaat untuk proses bisnisnya [3].

K-Means Clustering merupakan salah satu teknik *data mining* menggunakan strategi *unsupervised learning*. Strategi ini memiliki kemampuan untuk menemukan dan mengelompokkan data secara otomatis berdasarkan karakteristik tertentu yang tersembunyi di dalam data [4]. Meskipun memiliki potensi yang sangat besar, namun penggunaan *K-Means* sebagai salah satu fitur POS yang berfungsi memberikan rekomendasi pembelian produk di masa yang akan datang masih terbatas. Oleh karena itu, penelitian ini fokus pada implementasi fitur analisis data menggunakan *K-Means* yang akan memberikan rekomendasi pengambilan keputusan pembelian produk bagi pemilik usaha.

II. LANDASAN TEORI

A. *Knowledge Management* dan *Data Mining*

Knowledge Management dapat diartikan sebagai paradigma, *framework*, konsep, fungsi, maupun nilai yang berfungsi mendukung individu maupun organisasi untuk mengelola nilai yang dimiliki [5]. Dalam terminologi yang lain, KM juga dapat dipahami sebagai proses identifikasi, pengumpulan, dan pengkodean pengetahuan eksplisit yang telah ada baik dalam bentuk prosedur, dokumen, maupun kebijakan yang sebelumnya tidak terekam secara implisit yang dapat bermanfaat bagi orang lain [6]. Sehingga secara umum KM dapat dipahami sebagai serangkaian metode, *framework*, maupun konsep untuk mengumpulkan pengetahuan yang berasal dari individu atau kelompok dan

mengelolanya untuk menghasilkan manfaat bagi banyak orang.

Perkembangan teknologi internet khususnya dalam bidang *database*, *data warehouse*, *machine learning*, dan manajemen berbasis pengetahuan telah banyak mempengaruhi cara organisasi untuk memenangkan kompetisi. Pengetahuan yang tersembunyi di dalam aktivitas operasional perusahaan telah banyak disadari sebagai faktor kunci untuk mendukung berbagai keputusan organisasi [3]. Sebagai sebuah teori, DM didefinisikan sebagai proses yang menggunakan teknik-teknik statistika, matematika, kecerdasan buatan dan *machine learning* untuk mengekstrak dan mengidentifikasi informasi yang bermanfaat dan mengubahnya menjadi pengetahuan yang bersumber dari *database* [7]. Pemanfaatan *data mining* (DM) dalam proses penemuan pengetahuan yang tersimpan di dalam *database* mampu menyediakan berbagai metodologi pengambilan keputusan, penyelesaian masalah, analisis dan perencanaan, diagnosis, deteksi, *integrasi*, sampai pada penyediaan tindakan preventif, pembelajaran, maupun motivasi.

B. K-Means Clustering

K-Means Clustering merupakan salah satu algoritma pengklasteran yang mudah digunakan dan memiliki efisiensi komputasi yang tinggi. Meskipun demikian, algoritma ini memiliki sensitifitas yang cukup tinggi pada *centroid* [8]. Pada *clustering* berbasis *centroid*, sejumlah titik awal data tertentu diinisiasi sebagai *centroid* awal. Selanjutnya, pengelompokan data dilakukan secara iteratif melalui penghitungan jarak masing-masing objek observasi terhadap titik *centroid* terdekat hingga mencapai kriteria tertentu [9].

Secara umum metode *K-Means* dilakukan mengikuti algoritma berikut:

1. Menentukan jumlah kelas *cluster*
2. Menentukan titik *centroid* awal
3. Menghitung jarak antara data dengan *centroid* menggunakan rumus jarak Euclidean sebagai berikut:

$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \dots\dots\dots (1)$$

4. Menghitung rata-rata data dari setiap kelas menggunakan rumus sebagai berikut:

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \dots\dots\dots (2)$$

5. Untuk seluruh kelas, jika perbedaan antara nilai *means* dan *centroid* > nilai *error*, ganti nilai *centroid* dengan rata-rata nilai kelas dan ulangi langkah 3.

C. Sistem Informasi Eksekutif

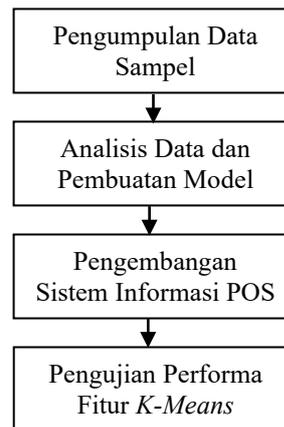
EIS merupakan sistem yang didesain untuk memenuhi kebutuhan eksekutif, merupakan gabungan dari sistem informasi *enterprise* dan teknologi web yang menyandarkan sebagian besar prosesnya pada manajemen data [10]. Selain berfungsi memberikan rekomendasi yang lebih baik bagi pengambil keputusan, EIS memiliki beberapa kelebihan sebagai berikut [11]:

1. Akses yang lebih cepat pada *database* dan informasi
2. Informasi yang lebih relevan dan terbaru
3. Meningkatkan kemampuan untuk mengidentifikasi riwayat tren di masa lalu
4. Meningkatkan efektifitas dan efisiensi di kalangan eksekutif.

III. METODOLOGI PENELITIAN

Metodologi yang digunakan pada penelitian ini menggunakan pendekatan *research and development* (R&D). Metode ini dipilih dengan mempertimbangkan produk yang dihasilkan dan kontribusi yang dapat diberikan kepada masyarakat [12]. Pada proses ini dilakukan tahap-tahap sebagai berikut:

1. Pengumpulan Data Sampel
2. Analisis Data dan Pembuatan Model *K-Means*
3. Pengembangan Sistem Informasi POS
4. Pengujian Efektivitas Fitur Analisis Data Menggunakan *K-Means* sebagai fitur POS



Gambar 1. Metodologi Penelitian

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data Sampel

Data sampel diperoleh dari 27.000 data transaksi penjualan *spare part* kendaraan bermotor di sebuah toko di Surabaya selama 1 tahun. Pemilihan data sampel ini dikhususkan pada produk kendaraan roda 2 dengan pertimbangan besarnya jumlah pembelian dan efisiensi persediaan barang di gudang.

Tabel 1. Data Transaksi Penjualan 76 Jenis Produk Sparepart Selama 1 Tahun Terakhir

ID_Brg	Harga	Jml Jual	ID_Brg	Harga	Jml Jual
1	88000	1	39	30250	9
2	88000	2	40	90000	16
3	93500	3	41	120000	10
4	88000	229	42	82500	17
5	102000	62	43	95000	18
6	102000	36	44	160000	7
7	93500	20	45	90000	71
8	120000	213	46	104500	351
9	110000	1179	47	69000	9
10	50000	339	48	104500	6
11	27500	336	49	129500	229
12	27500	3323	50	72600	190
13	35100	80	51	32000	28
14	25000	1573	52	71500	428
15	65000	1013	53	77000	366
16	65000	83	54	33000	230
17	92000	81	55	12000	237
18	39000	452	56	23500	374
19	63500	543	57	27000	113
20	110000	507	58	30000	39
21	150000	148	59	34000	191
22	175000	2	60	25000	450
23	60000	12	61	26000	277
24	120000	12	62	27500	393
25	30000	3	63	26000	40
26	40000	4	64	28350	303
27	45000	107	65	29000	40
28	33000	1184	66	30500	40
29	28000	429	67	27500	403
30	33000	589	68	31500	17
31	63500	131	69	29000	814
32	102000	14	70	31500	845
33	125000	16	71	8000	325
34	150000	4	72	7500	78
35	65000	10	73	8000	102
36	60000	5	74	31500	70
37	60000	960	75	10000	151
38	31000	6335	76	34000	406

B. Analisis Data dan Pembuatan Model

Berdasarkan data awal dilakukan proses *clustering* menggunakan pendekatan *K-Means*. Pada proses *K-Means* ditentukan jumlah *cluster* yang akan dihasilkan sebanyak 3 *cluster*, yaitu *cluster* penjualan rendah, penjualan sedang, dan penjualan tinggi. Penentuan *centroid* awal ditentukan secara *random* sesuai Tabel 2 berikut.

Tabel 2. Centroid Awal

Cluster	P1	P2
1	7500	78
2	102000	14
3	175000	2

Selanjutnya proses penghitungan *means* dan *update* nilai *centroid* terus dilakukan hingga tidak ada perubahan keanggotaan pada masing-masing *cluster*. Setelah iterasi ke 22, diperoleh hasil *cluster* untuk masing-masing data seperti pada Tabel 3 berikut.

Tabel 3. Hasil Akhir Analisis Cluster Menggunakan K-Means

id_brg	cluster	id_brg	cluster	id_brg	cluster
1	2	27	1	53	2
2	2	28	1	54	1
3	2	29	1	55	1
4	2	30	1	56	1
5	2	31	1	57	1
6	2	32	2	58	1
7	2	33	2	59	1
8	2	34	3	60	1
9	2	35	1	61	1
10	1	36	1	62	1
11	1	37	1	63	1
12	1	38	1	64	1
13	1	39	1	65	1
14	1	40	2	66	1
15	1	41	2	67	1
16	1	42	2	68	1
17	2	43	2	69	1
18	1	44	3	70	1
19	1	45	2	71	1
20	2	46	2	72	1
21	3	47	2	73	1
22	3	48	2	74	1
23	1	49	3	75	1
24	2	50	2	76	1
25	1	51	1		
26	1	52	2		

Sehingga, berdasarkan Tabel 3 dapat disimpulkan bahwa 60,5% data merupakan produk dengan kategori tingkat penjualan tinggi, 32,9% merupakan produk dengan tingkat penjualan sedang, dan 6,6% produk merupakan produk dengan tingkat penjualan rendah (Tabel 4).

Table 4. Presentase Tingkat Penjualan Produk berdasarkan Cluster

Cluster	Jumlah Data	Persen
1	46	60,5%
2	25	32,9%
3	5	6,6%

C. Pengembangan Sistem Informasi POS

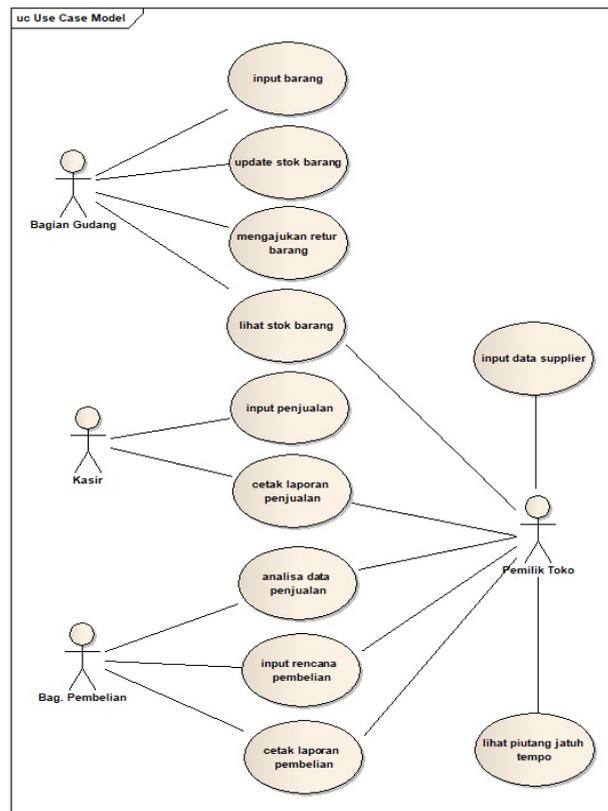
Untuk menentukan produk-produk yang masuk dalam kategori tingkat penjualan tinggi, sedang, dan rendah, dibuatlah alat bantu berupa fitur POS yang dapat digunakan pemilik usaha untuk melakukan pembelian ulang produk.

Sistem informasi POS dengan fitur analisis *K-Means* ini dikembangkan menggunakan pendekatan pengembangan *waterfall* dengan perancangan sistem menggunakan pendekatan *usecase driven object*.

Kemampuan fungsional yang diperlukan pengguna terhadap sistem POS tersebut dapat didefinisikan sebagai berikut:

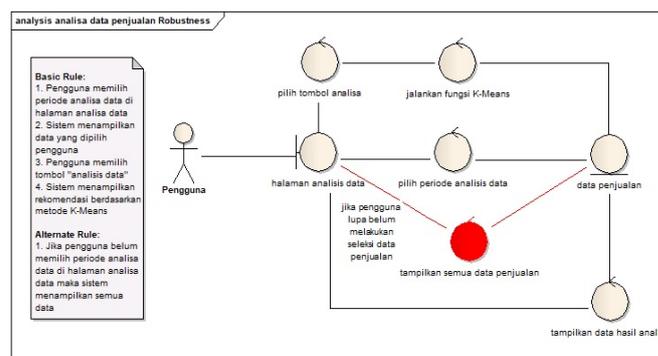
1. Bagian gudang dapat menginputkan barang masuk
2. Bagian gudang dapat mengupdate stok barang
3. Bagian gudang dapat mengajukan retur
4. Kasir dapat melakukan input data penjualan
5. Kasir dapat mencetak laporan penjualan
6. Bagian pembelian dapat melakukan analisis data
7. Bagian pembelian dapat mengajukan rencana pembelian
8. Bagian pembelian dapat mencetak laporan analisis data
9. Pemilik toko dapat melakukan analisis data
10. Pemilik toko dapat melihat hasil analisis data
11. Pemilik toko dapat mencetak laporan penjualan
12. Pemilik toko dapat mencetak laporan pembelian
13. Pemilik toko dapat mencetak hasil analisis data
14. Pemilik toko dapat menginput data *supplier*
15. Pemilik toko dapat melihat piutang jatuh tempo

Berdasarkan kebutuhan fungsional tersebut maka dapat disusun *use case* dengan empat *rule actor*, yaitu pemilik, bagian gudang, kasir, dan bagian pembelian seperti yang dapat dilihat pada Gambar 2. Selain fitur standar dari POS yang memiliki kemampuan dokumentasi data, sistem informasi yang dikembangkan juga dilengkapi dengan fitur analisis data yang mengimplementasikan metode *K-Means* sesuai dengan simulasi model manual yang telah dilakukan sebelumnya.

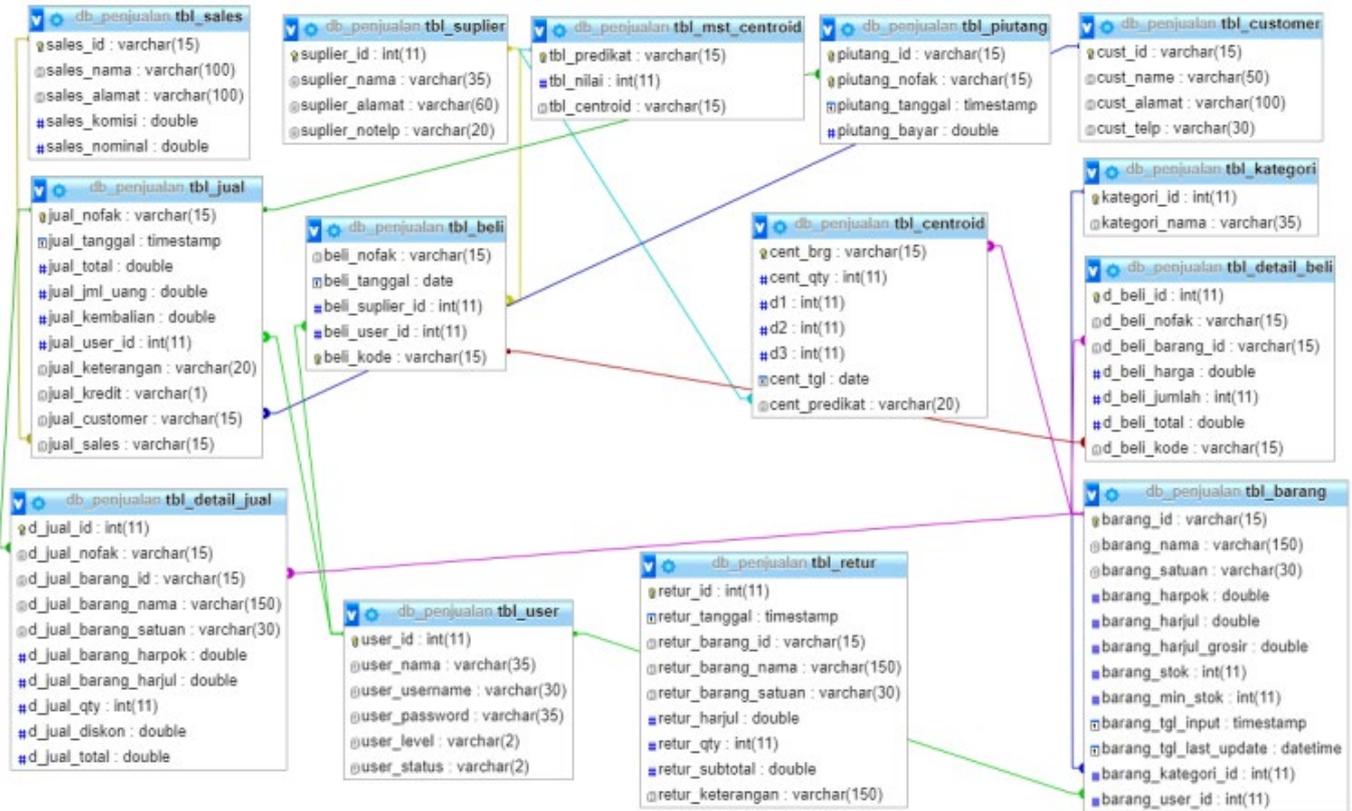


Gambar 2. Use Case Sistem Informasi POS dengan Fitur Analisis Data Menggunakan *K-Means*

Bagian pembelian dapat mengakses fitur analisis data untuk mengetahui tren penjualan barang pada periode tertentu. Selain itu, pemilik toko juga dapat memanfaatkannya untuk memonitor kondisi persediaan barang. Untuk memanfaatkan fitur ini, baik bagian pembelian maupun pemilik toko harus melakukan seleksi periode data yang akan dianalisis pada halaman analisis data. Selanjutnya, sistem akan menampilkan rekomendasi dari *K-Means* di halaman analisis data. Untuk mencetak rekomendasi hasil analisis data, pengguna data dapat memilih tombol cetak yang terdapat di halaman analisis data. Proses ini secara detil dapat dilihat pada *robustness* diagram pada gambar 3 berikut.



Gambar 3. Robustness Diagram Analisis Data Menggunakan *K-Means*



Gambar 4. Database Sistem Informasi POS Dengan Fitur Analisis Data Menggunakan *K-Means*

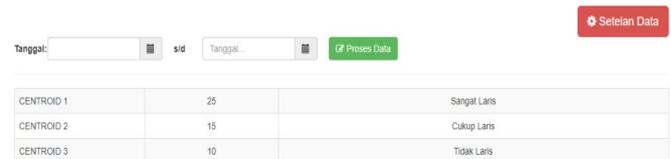
Sebagai media penyimpanan dari sistem informasi POS, maka dibuatlah *database* seperti terlihat pada Gambar 4 di atas. Setelah diimplementasikan, sistem informasi POS dengan fitur analisis data memiliki tampilan awal yang dapat dilihat pada Gambar 5. Selain memiliki menu yang terletak di atas aplikasi, sistem juga dilengkapi dengan *shortcut* berupa menu-menu yang paling sering digunakan dalam sistem POS, antara lain: penjualan eceran, *sales*, *supplier*, barang, pembelian, dan laporan.



Gambar 5. Tampilan Pada Halaman Awal Sistem Informasi POS Dengan Fitur Analisis Data *K-Means*

Pada Gambar 6, fitur analisis data terletak di atas sistem informasi dan terbatas hanya dapat diakses oleh pemilik toko dan bagian pembelian. Penentuan *centroid* terupdate otomatis setelah pengguna memilih periode tanggal awal dan akhir transaksi dan mengklik tombol proses, dan jika

pengguna tidak memilih data transaksi maka secara *default* sistem akan mengacak data *centroid* awal.



Gambar 6. Fitur Analisis Data Menggunakan *K-Means Cluster*

Setelah pengguna mengklik tombol proses data, maka sistem akan menampilkan hasil rekomendasi berupa data produk untuk masing-masing *cluster*. Selain itu, informasi juga dilengkapi dengan rata-rata jumlah penjualan pada periode tertentu (Gambar 7).

D. Pengujian Performa Fitur Analisis Data *K-Means*

Pengujian performa penggunaan metode *K-Means* pada studi kasus penjualan produk *sparepart* menunjukkan tingkat *error* sebesar 14,47%. Sehingga dapat disimpulkan bahwa metode *K-Means* memiliki tingkat akurasi sebesar 85,53%.

Sangat Laris				
ID Barang	Nama Barang	Jumlah Order	Distance	Predikat
BR000001	D-P221 Suzuki – Shinai Shogun125 Salmat18 Thunder125 Spin Skywave Skyblue	82	27 37 42	Sangat Laris
BR000006	D-M753 Yamaha – MioSport	43	18 28 33	Sangat Laris
BR000015	D-M756 Honda – Newbeat1 Newcopy11 NewVelo11	143	118 130 133	Sangat Laris
BR000018	D-M781 Honda – Newbeat125R1 PCK181	202	177 187 192	Sangat Laris
BR000018	D-2181 TOYOTA – Avanza 5th Agia Cella DAIHATSU – Kalia Terra Greaser Ludo Serox Cella Ayta Agia	220	185 200 210	Sangat Laris
BR000018	D-2188 TOYOTA – Innova Crossover TeraCam Innova Fortuner Velox Vaggleme12	41	16 26 31	Sangat Laris

Cukup Laris				
ID Barang	Nama Barang	Jumlah Order	Distance	Predikat
BR000004	D-Q219 Kawasaki – Kaze 182z KLX – HONDA – CBR250cc TRX	14	11 1 4	Cukup Laris
BR000007	D-M714 Yamaha – Mio	13	12 2 3	Cukup Laris
BR000005	D-5834 SUZUKI – Swift ADV Ertiga TOYOTA – AllNewAvanza 2013Avanza GrandNewAvanza Veloz2012 keajaiban GrandNewVelo Agria Ayta DAIHATSU – GranMia Ludo New5	18	7 3 8	Cukup Laris
Total Data			1380	
Total Terjual			48	
Rata Rata			0.03	

Tidak Laris				
ID Barang	Nama Barang	Jumlah Order	Distance	Predikat
BR000002	D-P222 Suzuki – Nex	51	14 4 1	Tidak Laris
BR000003	D-P223 Suzuki – NewShogun120FL	6	19 9 4	Tidak Laris
BR000005	D-M213 Yamaha – Vision Japhe10X ScorpioZ Japhe1Z1	1	24 14 9	Tidak Laris
BR000006	D-M754 Yamaha – Mio MioGT MioPro11 X-Ride ScoaGT	3	22 12 7	Tidak Laris
BR000009	D-M780 Yamaha – Mio104 MioPro	5	20 10 5	Tidak Laris

Gambar 7 Tampilan Rekomendasi K-Means Cluster Untuk Fitur Analisis Data

V. KESIMPULAN

Penambahan fitur analisis data menggunakan pendekatan K-Means clustering pada sistem informasi POS dapat memberikan nilai tambah bagi pemilik usaha untuk merencanakan pembelian dengan lebih baik. Selain itu, rekomendasi yang diberikan dalam bentuk fitur dalam sistem informasi POS dapat digunakan dengan mudah sehingga dapat meningkatkan efisiensi organisasi. Penggunaan fitur K-Means secara signifikan juga terbukti memiliki tingkat akurasi yang tinggi yaitu sebesar 85,53% sehingga dalam jangka panjang metode ini dapat terus digunakan sebagai salah satu fitur decision support system bagi perusahaan.

REFERENSI

[1] Paschek, D. & Draghici, A. (2018). Knowledge Management – The Foundation for a Successful Business Process Management. *Procedia - Soc. Behav.*

Sci., Vol. 238, pp. 182–191.

[2] Abubakar, A.M., Elrehail, H., Alatailat, M.A. & Elçi, A. (2017). Knowledge Management, Decision-Making Style and Organizational Performance. *J. Innov. Knowl.*, pp. 1–15.

[3] Shaw, M.J., Subramaniam, C., Woo, G. & Welge, M.E. (2001). Knowledge Management and Data Mining for Marketing. *Decis. Support Syst.*, Vol. 31, pp. 127–137.

[4] Thomas, M.C., Zhu, W. & Romagnoli, J.A. (2017). Data Mining and Clustering in Chemical Process Databases for Monitoring and Knowledge Discovery. *J. Process Control.*

[5] Tsai, H. (2013). Knowledge Management vs. Data Mining: Research Trend, Forecast and Citation Approach. *Expert Syst. Appl.*, Vol. 40, No. 8, pp. 3160–3173

[6] Cooper, P. (2016). Data, Information, Knowledge, and Wisdom. *Anaesth. Intensive Care Med.*, pp. 1–2.

[7] Turban, E., Aronson, J.E. & Liang, T.P. (2007). Decision Support Systems and Business Intelligence. *Decis. Support Bus. Intell. Syst. 7/E*, pp. 1–35.

[8] Majhi, S.K. & Biswal, S. (2018). Optimal Cluster Analysis Using Hybrid K-Means and Ant Lion Optimizer. *Karbala Int. J. Mod. Sci.*, Vol. 4, No. 4, pp. 347–360.

[9] Ben, S., Naouali, S. & Chtourou, Z. (2018). A Fast and Effective Partitioning Clustering Algorithm for Large Categorical Datasets Using a K-Means Based Approach. *Comput. Electr. Eng.*, Vol. 68, April, pp. 463–483.

[10] Koh, C.E. & Watson, H.J. (1998). Data Management in Executive Information Systems. *Inf. Manag.*, Vol. 33, pp. 301–312.

[11] Turban, E., Sharda, R. & Delen, D. (2011). *Decision Support and Business Intelligence Systems*. Pearson Education International.

[12] Yoshikawa, H. (2012). *Design Methodology for Research and Development Strategy*. Japan: Center for Research and Development Strategy, Japan Science and Technology Agency.