

## PENERAPAN DATA MINING UNTUK MENENTUKAN PENJUALAN SPAREPART TOYOTA DENGAN METODE K-MEANS CLUSTERING

Saut Parsaoran Tamba\*, Felix Toknady Kesuma, Feryanto  
Program Studi Sistem Infomasi Fakultas dan Ilmu Komputer Universitas Prima Indonesia  
E-mail : \*saut\_nabasa@yahoo.co.id

**ABSTRAK** - Semakin berkembangnya persaingan dalam dunia bisnis khususnya dalam industri penjualan sparepart mobil dan jasa service menuntut para pengembang untuk menemukan suatu pola yang dapat meningkatkan penjualan dan pemasaran barang di perusahaan, salah satunya adalah dengan pemanfaatan data transaksi. CV Terang Jaya merupakan perusahaan yang bergerak dalam bidang Otomotif yang melayani pembelian, penjualan sparepart mobil serta memberikan service untuk berbagai merek mobil. Namun demikian kurang dalam peninjauan produk-produk apa saja yang dibutuhkan konsumen dan penyimpanan data-data yang kurang efektif. Untuk mengatasi permasalahan tersebut analisis yang digunakan yaitu penerapan *Clustering* dengan menggunakan Algoritma *K-Means*. *Clustering* merupakan salah satu teknik dari salah satu fungsionalitas data mining, algoritma *clustering* merupakan algoritma pengelompokan sejumlah data menjadi kelompok-kelompok data tertentu (*cluster*). Sehingga dengan adanya pengelompokan data ini pihak perusahaan dapat mengetahui barang paling laris, laris dan tidak laris. Sehingga barang yang ada digudang tidak menumpuk. Dari penititan ini output yang dihasilkan yaitu, barang paling laris sebanyak 15, barang yang laris sebanyak 45 dan kurang laris sebanyak 13. Dengan adanya pengolahan data yang dilakukan diharapkan dapat memberikan solusi kepada pihak perusahaan agar dapat mengetahui mana barang yang paling laris, laris dan mana barang yang tidak laris.

Kata kunci: Data Mining, K-Means, *Clustering*, *Sparepart*

### 1. PENDAHULUAN

Pada era sekarang kebutuhan masyarakat terhadap alat transportasi semakin meningkat, untuk itu masyarakat memiliki kendaraan mobil pribadi guna dalam menjalani aktivitas sehari-hari. Setiap pengguna mobil pastinya ingin kendaraanya selalau dapat berjalan dengan baik. oleh karena itu perusahaan – perusahaan yang menjual suku cadang mobil beserta dengan pelayananan atau perawatan mobil harus meningkatkan produktifitas dan efisien kerja dalam usaha.

Semakin berkembangnya persaingan dalam dunia bisnis khususnya dalam industri penjualan sparepart mobil dan jasa service menuntut para pengembang untuk menemukan suatu pola yang dapat meningkatkan penjualan dan pemasaran barang di perusahaan, salah satunya adalah dengan pemanfaatan data transaksi. Penggunaan sistem informasi dalam persaingan yang ketat dalam suatu perusahaan dengan perusahaan yang lain merupakan salah satu masalah yang datang dari luar perusahaan.

CV Terang Jaya merupakan perusahaan yang bergerak dalam bidang Otomotif yang melayani pembelian, penjualan sparepart mobil serta memberikan service untuk berbagai merek mobil. Namun demikian CV Terang Jaya kurang dalam peninjauan produk yang dijual, produk-produk apa saja yang dibutuhkan konsumen dan penyimpanan data-data kurang efektif. Dengan demikian perlu adanya suatu sistem yang dapat mendukung perusahaan dalam mengambil keputusan secara cepat dan juga tepat. Dalam hal ini analisi yang digunakan untuk mengatasi permasalahan diatas

dengan menggunakan penerapan *Clustering* dengan menggunakan Algoritma K-Means.

*Clustering* merupakan salah satu teknik dari salah satu fungsionalitas data mining, algoritma *clustering* merupakan algoritma pengelompokan sejumlah data menjadi kelompok-kelompok data tertentu (*cluster*). Sedangkan Data mining, sering juga disebut sebagai knowledge discovery in database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Sehingga dengan adanya pengelompokan data ini pihak perusahaan dapat mengetahui barang paling laris, laris dan tidak laris, sehingga barang yang ada digudang tidak menumpuk. Dengan adanya pengolahan data yang dilakukan diharapkan akan dapat memberikan solusi nyata kepada pihak perusahaan agar dapat mengetahui mana barang yang paling laris, laris dan mana barang yang tidak laris.

### 2. METODE PENELITIAN

Penelitian ini menggunakan teknik klasifikasi dengan *K-Means* untuk mengelompokkan data kategorikal hingga menghasilkan klaster yang lebih stabil. Algoritma *K-means Clustering* adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode *K-Means* berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik

yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain.

Tahapan melakukan *clustering* atau pengelompokan dengan metode *K-Means* adalah sebagai berikut :

1. Pilih jumlah *cluster*  $k$ .
2. Inisialisasi  $k$  pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberiduberi nilai awal dengan angka-angka random.
3. Tempatkan semua data/ objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2} \quad (1)$$

dimana:

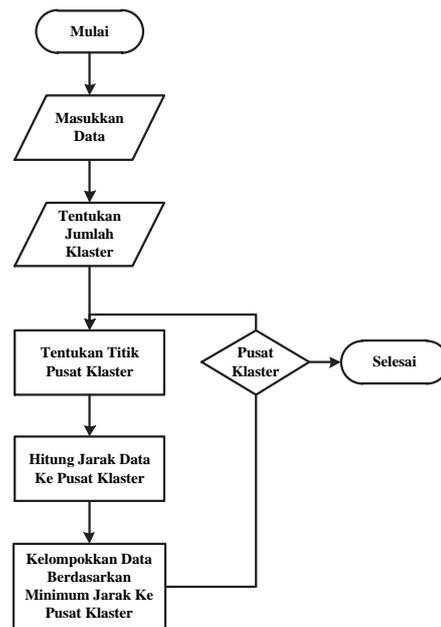
$D(i, j)$  = Jarak data ke  $i$  ke pusat *cluster*  $j$

$X_{ki}$  = Data ke  $i$  pada atribut data ke  $k$

$X_{kj}$  = Titik pusat ke  $j$  pada atribut ke  $k$

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi

Proses algoritma k-means dapat seperti dijabarkan dalam flowchart berikut



Gambar 1. Flowchart Algoritma K-Means

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Analisa Data

Data penelitian ini diapat dari CV. Terang Jaya, dimana data yang digunakan data penjualan sebanyak 50 data. Output yang diharapkan adalah menghasilkan 3 *cluster* yaitu (C1) data penjualan paling laris, (C2) data penjualan yang laris, dan (C3) data penjualan kurang laris. Variabel atau atribut digunakan dalam pengelompokan data penjualan ini terdiri dari **Stok Awal, dan Terjual**, seperti yang terlihat pada Tabel 1.

Tabel 1. Sampel Data

No	Nama Produk	Tipe Kendaraan	Stok Awal	Terjual
1	Bumper Depan	Fortuner 2010	29	16
2	Bumper Depan	Innova 2010	65	51
3	Bumper Depan	Avanza 2012	165	155
4	Bumper Depan	Rush 2012	36	14
5	Bumper Belakang	Innova 2012	36	25
6	Bumper Belakang	Fortuner 2012	27	10
7	Bumper Belakang	Avanza 2010	21	12
8	Kap Mesin	Avanza 2012	38	23
9	Kaca Depan	Avanza 2010	48	38
10	Kaca	Fortuner,	28	16

	Depan	Hilux 2012		
11	Kaca Depan	Rush 2010-2016	35	23
12	Kaca Bagasi	Avanza 2012	22	10
13	Kaca Bagasi	Innova 2012	20	13
14	Kaca Bagasi	Fortuner 2012	14	7
15	Lampu Besar RH	Innova 2012-2014	62	56
16	Lampu Besar RH	Avanza 2012	70	59
17	Lampu Besar RH	G Fortuner 2012	25	11
18	Lampu Besar RH	Rush 2012	12	4
19	Lampu Besar LH	Innova 2012-2014	28	16
20	Lampu Besar LH	Avanza 2012	90	74
21	Lampu Besar LH	G Fortuner 2012	15	7
22	Lampu Besar LH	Rush 2012	15	7
23	Lampu Stop RH	Innova 2012	60	48
24	Lampu Stop RH	Avanza 2012	70	59
25	Lampu Stop RH	Fortuner 2012	16	11
26	Lampu Stop RH	RUSH 2012	12	4
27	Lampu Stop LH	Innova 2012	65	47
28	Lampu Stop LH	Fortuner 2012	20	12
29	Lampu Stop LH	Rush 2012	21	10
30	Foglamp RH	Innova 2012	25	6
31	Foglamp LH	Avanza 2012	40	18
32	Grilld BUMPER	Innova 2012	50	27
33	Radiator	Innova Bensin 2012	15	3
34	Radiator	Avanza 2009	50	43
35	Spion RH	Rush 2012	35	12
36	Spion RH	Avanza 2010	40	19
37	Spion RH	Innova 2005-2012	50	33
38	Spion LH	Rush 2012	18	4

39	Spion LH	Avanza 2010	29	15
40	Spion LH	Innova 2005-2012	35	20
41	Tabung Hawa	Innova 2012	15	1
42	Liner FR RH	Innova 2010	30	21
43	Liner FR RH	Avanza 2010	29	11
44	Liner FR RH	Fortuner 2012	20	7
45	Liner FR LH	Innova 2010	30	26
46	Liner FR LH	Avanza 2010	30	21
47	Liner FR LH	Fortuner 2012	17	4
48	Shock FR RH	Rush 2007	20	5
49	Shock FR RH	Avanza 2009	25	9
50	Shock FR LH	Rush 2007	17	5

### 3.2 Penerapan K-Means Clustering

- Langkah pertama menentukan jumlah *cluster* yang ingin dibentuk. Dalam penelitian ini *cluster* yang akan dibentuk adalah sebanyak 3 *cluster*. Kemudian tentukan titik pusat awal dari setiap *cluster*, dimana pusat awal ditentukan secara acak/random sebagai berikut.

Table 2. Titik Pusat Awal Cluster

Cluster	Stok Awal	Terjual
cluster1	2	65
Cluster2	5	6
Cluster3	11	19

- Tempatkan semua data/ objek ke *cluster* terdekat dan Hitung jarak dari data ke-1 terhadap pusat *cluster*

$$\begin{aligned} \text{Cluster 1} &= \sqrt{(65 - 29)^2 + (51 - 16)^2} \\ &= 1.261 \end{aligned}$$

$$\begin{aligned} \text{Cluster 2} &= \sqrt{(16 - 29)^2 + (11 - 16)^2} \\ &= 38 \end{aligned}$$

$$\begin{aligned} \text{Cluster 3} &= \sqrt{(35 - 29)^2 + (20 - 16)^2} \\ &= 22 \end{aligned}$$

Jarak dari data ke-2 terhadap pusat *cluster*

$$\begin{aligned} \text{Cluster 1} &= \sqrt{(65 - 65)^2 + 51} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Cluster 2} &= \sqrt{(16 - 65)^2 + (11 - 51)^2} \\ &= 1.649 \end{aligned}$$

$$\begin{aligned} \text{Cluster 3} &= \sqrt{(35 - 65)^2 + 51} \\ &= 991 \end{aligned}$$

Berdasarkan hasil perhitungan jarak dari data ke-1 terhadap pusat *cluster* di atas dapat disimpulkan bahwa jarak data pertama yang paling dekat adalah dengan *cluster* 3. Hasil perhitungan selengkapnya untuk 50 data produk dapat di lihat pada tabel 3.

Tabel 3. Hasil Perhitungan Setiap Data Pada Iterasi 1

Jarak Ke			Jarak Terdekat Ke Cluster		
C1	C2	C3	C1	C2	C3
1.261	38	22	0	0	1
-	1.649	991	1	0	0
10.916	20.885	18.355	1	0	0
1.398	29	37	0	1	0
705	216	26	0	0	1
1.719	12	108	0	1	0
1.565	6	78	0	1	0
811	166	12	0	0	1
186	761	337	1	0	0
1.262	37	23	0	0	1
814	163	9	0	0	1
1.724	7	113	0	1	0
1.489	8	64	0	1	0
1.987	18	190	0	1	0
28	2.071	1.323	1	0	0
69	2.358	1.556	1	0	0
1.640	9	91	0	1	0
2.262	53	279	0	1	0
1.262	37	23	0	0	1
554	4.043	2.971	1	0	0
1.986	17	189	0	1	0
1.986	17	189	0	1	0
14	1.413	809	1	0	0
69	2.358	1.556	1	0	0
1.649	-	100	0	1	0
2.262	53	279	0	1	0
16	1.345	759	1	0	0
1.566	5	79	0	1	0
1.725	6	114	0	1	0
2.065	34	206	0	1	0
1.114	73	9	0	0	1

591	290	64	0	0	1
2.354	65	309	0	1	0
79	1.058	544	1	0	0
1.551	20	64	0	1	0
1.049	88	6	0	0	1
339	518	184	0	0	1
2.256	51	273	0	1	0
1.332	29	31	0	1	0
991	100	-	0	0	1
2.550	101	381	0	1	0
935	114	6	0	0	1
1.636	13	87	0	1	0
1.981	20	184	0	1	0
660	239	41	0	0	1
935	114	6	0	0	1
2.257	50	274	0	1	0
2.161	40	240	0	1	0
1.804	13	131	0	1	0
2.164	37	243	0	1	0

- Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut.

Pembangkitan ulang *centroid* baru dihitung dengan rumus sebagai berikut

$$C = \frac{\sum m}{n} \quad (2)$$

Keterangan :

C = *centroid* data

m = anggota data yang termasuk kedalam *centroid* tertentu

n = jumlah data yang menjadi anggota *centroid* tertentu.

Berdasarkan Tabel 3, titik pusat *cluster* baru untuk parameter Stok Awal memiliki 10 anggota.:

$$C1 = (0 + 10.916 + 186 + 28 + 69 + 554 + 14 + 69 + 16 + 79) / 10 = 74.50$$

$$\begin{aligned} \text{Cluster Kedua (C2) memiliki 26 anggota} \\ C2 = (1.398 + 1.719 + 1.565 + 1.724 + 1.489 + 1.987 + 1.64 + 2.262 + 1.986 + 1.986 + 1.649 + 2.262 + 1.566 + 1.725 + 2.065 + 2.354 + 1.551 + 2.256 + 1.332 + 2.55 + 1.636 + 1.981 + \end{aligned}$$

$$(2.257 + 2.161 + 1.804 + 2.164) / 26 = 20.81$$

ClusterKetiga (C3) memiliki 14 anggota  
 $C3 = (1.261 + 705 + 811 + 1.262 + 814 + 1.262 + 1.114 + 591 + 1.049 + 339 + 991 + 935 + 660 + 935) / 14 = 35.64$

Ulangi perhitungan *cluster* baru untuk parameter Terjual, sehingga hasil perhitungan diperoleh titik pusat baru dari setiap *cluster* dibawah ini.

Table 4. Titik Pusat Baru Cluster

Cluster	Stok Awal	Terjual
cluster1	74.50	63
Cluster2	20.81	8.23
Cluster3	35.64	21.71

- Setelah didapatkan titik pusat yang baru dari setiap *cluster*, lakukan kembali dari langkah ketiga hingga titik pusat dari setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain.

### 3.4 Hasil

Berdasarkan hasil pengelompokan dari seluruh data menggunakan metode k-means *clustering*, didapatkan hasil akhir pengelompokan hingga iterasi ke-5, dimana titik pusat tidak lagi berubah dan tidak ada data yang berpindah antar *cluster*. Hasil *cluster* yang terbentuk dari setiap data tersebut dapat dilihat pada tabel 4.

Tabel 5. Hasil Akhir Pengelompokan

Nama Cluster	Data Barang
Cluster 1	9, 10, 11, 12, 13, 14, 18, 27, 28, 29, 30, 31, 32, 34, 35, 36, 45, 46, 50
Cluster 2	1, 2, 3, 4, 5, 6, 7, 8, 16, 21, 22, 23, 33, 40, 47, 51, 55, 56, 57, 58, 59, 60
Cluster 3	15, 17, 19, 20, 24, 25, 26, 37, 38, 39, 41, 42, 43, 44, 48, 49, 52, 53, 54

Berdasarkan hasil dari *cluster* yang terbentuk pada tabel 4 dapat dilihat bahwa hasil pengelompokan dari *cluster* diatas seperti terlihat pada dibawah ini.

Nama Cluster  
 Barang paling laris (C1) = 8 barang  
 Barang yang laris (C2) = 26 barang  
 Barang kurang laris (C3) = 16 barang

## 4. PENUTUP

### 4.1 Kesimpulan

- Dengan adanya pengelompokan data ini, pihak perusahaan dapat mengetahui barang paling laris, laris dan tidak laris. Sehingga barang yang ada digudang tidak menumpuk.
- Dari penelitian ini output yang dihasilkan yaitu, barang paling laris sebanyak 8, barang yang laris sebanyak 26 dan kurang laris sebanyak 16.
- Dengan adanya pengolahan data yang dilakukan diharapkan dapat memberikan solusi kepada pihak perusahaan agar dapat mengetahui mana barang yang paling laris, laris dan mana barang yang tidak laris.

### 4.2 Saran

- Untuk penelitian lebih lanjut dapat menambah parameter atau variabel lain untuk proses *clustering*.
- Untuk penelitian berikutnya data yang digunakan lebih banyak lagi agar menghasilkan data yang akurat.
- Pada penelitian selanjutnya dapat dilanjutkan dengan metode lain untuk pengelompokan atau klasifikasi data.

## DAFTAR PUSTAKA

- Zainul ArasZ, Sarjono, "Analisis Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan Bedah Rumah Menggunakan Metode Clustering K-Means Vol 1 No. 2, pp 159-170, 2016.
- Fina Nasari, Surya Darma, "Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru", 2015.
- Aldi Nurzahputra, Much Aziz Muslim, Miranita Khusniati, "Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa", Vol. 16, No. 1, pp17-24, 2017.
- Totok Suprawoto, "Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran" Vol. 1, No. 1, pp 12-16, 2016.
- Khairul Umami, "Analisa Data Mining Dalam Penjualan Sparepart Mobil Dengan Menggunakan Metode Algoritma Apriori" Vol.8 No.3 pp 155-164, 2015.
- Eva Rianti, "Data Mining Dalam Menentukan Penjualan Kacamata Pada Optik Zal Laris Dan Tidak Laris Menggunakan Metode Clustering" Vol. 4, No. 2, pp 267-283, 2017.
- EJohan Oscar Ong, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University" Vol. 12, No. 1, 2013.

- [8] Gunawan Abdillah, Firman Ananda Putra, Faiza Renaldi, "Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means", 2016.
- [9] Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C, "Application of K-Means Clustering Algorithm for Prediction of Students' Academic Performance", Vol. 7, No. 1, pp 292-295, 2010.
- [10] Unnati R. Raval, Chaita Jani, "Implementing & Improvisation of K-means Clustering Algorithm", Vol. 5, Issue. 5, pp 191-203, 2016.