

# ANALISA SEGMENTASI KONSUMEN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Sulistiyowati<sup>1)</sup>, Basma Eno Ketherin<sup>2)</sup>, Amalia Anjani Arifiyanti<sup>3)</sup>, Anwar Sodik<sup>4)</sup>

Jurusan Sistem Informasi, Institut Teknologi Adhi Tama Surabaya

e-mail : sulis\_tyowati@itats.ac.id

## ABSTRACT

*Consumer segmentation is a strategy carried out by dividing consumers into different consumer groups that have different characteristics, behaviors, or needs. Consumer segmentation is very important as supporting data for marketing supervisor to determine effective and efficient marketing strategies in product sales promotion activities. Segmenting consumer can be done by clustering or grouping data based on the consumer characteristics. In this research, clustering process can be done by grouping consumer data in three attributes i.e. motor type, purchasing type and job. The data used at Honda motorcycle dealers. The consumer grouping process was carried out using K-Means Algorithm, by divided data into k-clusters. Every cluster was determined by the proximity between data and mean point. The results of application comparison by SPSS software obtained that the difference among cluster was averagely 7% due to random selection of initial centroid.*

**Keywords:** *segmentation; consumer characteristics; k-means clustering.*

## ABSTRAK

Segmentasi konsumen adalah sebuah strategi yang dilakukan dengan membagi konsumen menjadi kelompok-kelompok konsumen yang berbeda yang memiliki karakteristik, perilaku, atau kebutuhan yang berbeda. Segmentasi konsumen ini penting karena dapat digunakan sebagai data pendukung untuk menentukan strategi pemasaran yang efektif dan efisien bagi *supervisor marketing* dalam kegiatan promosi penjualan produk. Untuk menentukan segmentasi konsumen, dilakukan dengan *clustering* yaitu dengan cara mengelompokkan data berdasarkan karakteristik konsumen. Sehingga pada penelitian ini, proses *clustering* dilakukan dengan cara mengelompokkan data konsumen berdasarkan tiga atribut yaitu, jenis motor, jenis pembelian dan pekerjaan. Data yang digunakan adalah data penjualan sepeda motor pada dealer Honda. Proses pengelompokan konsumen dilakukan dengan menggunakan Algoritma K-Means. Algoritma ini bekerja dengan cara membagi data dalam *k cluster*. Setiap *cluster* ditentukan atas kedekatan jarak tiap-tiap data dengan titik tengahnya (*mean point*). Dari hasil perbandingan aplikasi dengan menggunakan *software* SPSS diperoleh hasil selisih dari masing-masing *cluster* dengan jumlah rata-rata 7%. Hal ini disebabkan karena centroid awal yang dipilih secara acak.

**Kata kunci :** segmentasi; karakteristik konsumen; k-means *clustering*

## PENDAHULUAN

Saat ini pemanfaatan teknologi komputer sebagai alat bantu dalam mendukung kegiatan dalam bidang usaha sangat memudahkan manusia dalam mendapatkan informasi secara cepat, tepat, akurat sehingga tujuan dari suatu pekerjaan tersebut dapat tercapai secara efektif dan efisien. Pemanfaatan teknologi komputer juga sangat diperlukan untuk mendukung kegiatan promosi yaitu menentukan segmentasi konsumen berdasarkan data pembelian sepeda motor Honda sebelumnya. Pada proses pengelompokan segmentasi konsumen di dealer motor Honda, masih dilakukan secara manual, sehingga menyebabkan kesulitan dalam pengelompokkan data dan terjadinya kesalahan dalam pengelompokan yang dilakukan. Melalui penelitian ini, dilakukan sebuah analisa terhadap karakteristik konsumen di dealer Honda agar dapat menentukan segmentasi konsumen dengan menerapkan metode K-means. Hasil segmentasi konsumen ini dapat dimanfaatkan divisi *marketing* untuk memilih strategi promosi yang sesuai

dengan karakteristik calon konsumen, sehingga diharapkan promosi tersebut dapat berjalan lebih efektif dan efisien.

## TINJAUAN PUSTAKA

### Data

Data adalah sesuatu yang nyata diketahui atau dianggap yang dipakai untuk keperluan suatu analisa, diskusi, presentasi ilmiah atau tes statistik [1]. Data dapat dibagi menjadi tiga jenis yakni data cross section, data berkala (*time series*), dan data panel. Data *cross section* data yang sering disebut juga data silang adalah data yang menunjukkan suatu titik tertentu. Contoh dari data cross section adalah data penjualan sepeda motor pada bulan Desember 2016, data pelanggan Juni 2005, data keuangan pada bulan Agustus 2009, dan sebagainya. Berbeda halnya dengan data runtun waktu atau data *time series* yang merupakan data yang terdiri dari satu obyek akan tetapi meliputi beberapa periode waktu tertentu yaitu misalnya harian, mingguan, bulanan tahunan dan sebagainya. Sedangkan data panel adalah data yang mengkombinasikan antara data *cross section* dan *data time series*, dengan kata lain data panel (*pooled data*) akan mempunyai beberapa objek dan beberapa periode waktu.

### Segmentasi

Menurut Philip Kotler, yang dimaksud dengan segmentasi pasar adalah suatu usaha untuk meningkatkan ketepatan pemasaran perusahaan. Segmen pasar terdiri dari kelompok besar yang dapat diidentifikasi dalam sebuah pasar dengan keinginan, daya beli, lokal geografis, perilaku pembelian dan kebiasaan pembelian yang serupa [2].

### Regresi

Analisis regresi adalah hubungan yang didapat dan dinyatakan dalam bentuk persamaan matematik yang menyatakan hubungan fungsional antar variabel-variabel [3]. Regresi dibagi menjadi dua, yaitu [3] :

1. Analisis regresi linier sederhana digunakan untuk mendapatkan hubungan matematis dalam bentuk suatu persamaan antara variabel tak bebas dengan variabel bebas tunggal.
2. Analisis regresi linier berganda ini digunakan untuk memprediksi berubahnya nilai variabel tertentu bila variabel lain berubah.

### Clustering

*Clustering* merupakan teknik data mining yang berfungsi untuk mengelompokkan data berdasarkan kemiripan data pada suatu kelompok dan meminimalkan kemiripan pada kelompok lain [4]. *Clustering* merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan record yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam *cluster* lain[5]. Ada dua metode *clustering* yang kita kenal, yaitu *hierarchical clustering* dan *partitioning*.

### K-Means Clustering

*K-means clustering* merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil [6].

Langkah-langkah melakukan *clustering* dengan metode *K-Means* adalah sebagai berikut [7] :

1. Pilih jumlah *cluster*  $k$ .
2. Inisialisasi  $k$  pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara acak. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka *random*.
3. Alokasikan 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 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 :  
$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

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* (nilai tengah) 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.

## METODE

### Alur Proses Cluster Menggunakan K-Means Clustering

Proses *cluster* menggunakan metode *K-Means clustering* memiliki beberapa tahapan, yaitu :

#### 1. Pengumpulan Data

Pengumpulan data dilakukan dengan cara *observasi* langsung ke dealer motor Honda dan juga melakukan wawancara dengan *stakeholder* yang terkait. Penelitian ini mengambil data penjualan selama empat tahun terakhir yaitu pada tahun 2013-2016. Data yang diuji adalah data jenis motor, data jenis pembelian dan data pekerjaan.

#### 2. Praproses Data

Tahap praproses data merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain, yang dilakukan antara lain :

##### 1. Analisis Atribut

Sebelum data tersebut diolah, terlebih dahulu dilakukan pemilihan atribut yaitu memilih data –data apa saja yang digunakan dalam proses pengelompokan data. Atribut-atribut data tersebut dipilih berdasarkan kebutuhan dari perusahaan. Atribut yang dipilih tersebut, kemudian dianalisis dengan menggunakan aplikasi SPSS. Analisis atribut dilakukan untuk mengetahui keeratan hubungan antara tiga variabel (data jenis motor, jenis pembelian, dan pekerjaan).

##### 2. Analisis Jumlah Cluster

Setelah menganalisa jumlah atribut, proses yang dilakukan selanjutnya adalah menganalisa jumlah *cluster*. Dalam proses analisis jumlah *cluster* digunakan metode Elbow untuk mencari nilai  $k$  terbaik. Jumlah *cluster* yang akan diuji adalah  $k=2$  sampai dengan  $k=8$ . Dari hasil proses perhitungan *Sum of Square Error* terhadap 7603 data konsumen maka hasil yang paling mengalami penurunan yang paling besar adalah pada  $K=3$

3. Pengelompokan Data

Agar data penjualan dapat diolah dengan menggunakan *k-means clustering*, maka data penjualan yang berjenis nominal seperti jenis motor, jenis pembelian dan pekerjaan harus dilakukan proses inialisasi data terlebih dahulu. Sebelum dilakukan proses inialisasi, data tersebut harus dikelompokkan terlebih dahulu. Proses pengelompokan data ini diurutkan berdasarkan urutan abjadnya:

- a. **Jenis Motor** : Beat, Beat cbs iss, Beat cw, Beat pop cbs, Beat pop cbs iss, Beat pop cw, Blade, CB150R, CBR150, Megapro, Revo, Scoopy, Sonic, Spacy, Supra, Supra GTR, Vario 110, Vario 125 cbs, Vario 125 cbs is, Vario 150, Verza.
- b. **Jenis Pembelian** : Kredit dan Tunai
- c. **Pekerjaan** : Buruh, Dokter, Guru/Dosen, Ibu Rumah Tangga, Lain-Lain, mahasiswa/Pelajar, Ojek, Pegawai Negeri, Pegawai Swasta, Pengacara, Petani/Nelayan, TNI/POLRI, Wartawan, Wiraswasta/Pedagang.

4. Inialisasi Data

Pada tahapan ini, data-data penjualan tersebut diberi insial dengan angka 1,2,3 dan seterusnya. Hasil dari inialisasi data-data penjualan tersebut dapat dilihat pada Tabel 1, Tabel 2 dan Tabel 3

Tabel 1. Inialisasi Data Jenis Motor

Jenis Motor	Inisial	Jenis Motor	Inisial
beat cbs	1	Revo	11
beat cbs iss	2	Scoopy	12
beat cw	3	Sonic	13
beat pop cbs	4	Spacy	14
beat pop cbs iss	5	Supra	15
beat pop cw	6	supra gtr	16
Blade	7	vario 110	17
cb150r	8	vario 125 cbs	18
cbr150r	9	vario 125 cbs iss	19
mega pro	10	vario 150	20
		Versa	21

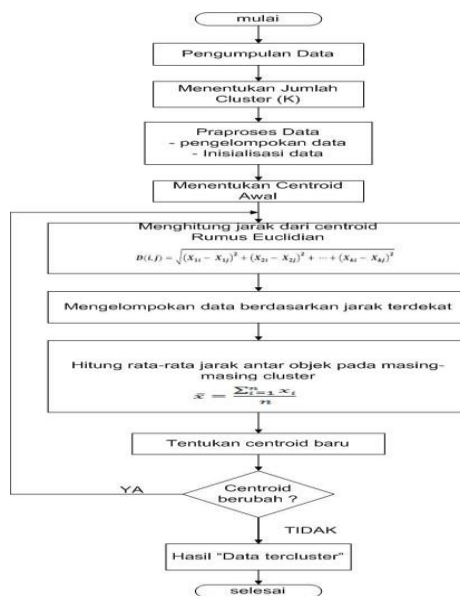
Tabel 2. Inialisasi Data Jenis Motor

Jenis Motor	Inisial
kredit	1
tunai	2

Tabel 3. Inialisasi Data Jenis Pekerjaan

Jenis Pekerjaan	Inialisasi	Jenis Pekerjaan	Inialisasi
Buruh	1	Pegawai Negeri	8
Dokter	2	Pegawai Swasta	9
Guru/Dosen	3	Pengacara	10
Ibu Rumah Tangga	4	Petani/Nelayan	11
Lain-Lain	5	TNI/POLRI	12
Mahasiswa/Pelajar	6	Wartawan	13
Ojek	7	Wiraswasta/Pedagang	14

Setelah semua data penjualan diinisialisasi ke dalam bentuk angka, maka data-data tersebut telah dapat dikelompokkan dengan menggunakan algoritma *K-means Clustering*. Alur proses *cluster* menggunakan metode *K-Means clustering* dapat dilihat pada Gambar 1.



Gambar 1. Alur Proses Cluster Menggunakan K-Means Clustering

**HASIL DAN PEMBAHASAN**

Berdasarkan proses *cluster* yang telah dilakukan pada aplikasi *clustering* yang menggunakan algoritma *k-means*, diperoleh tiga cluster yang terdiri dari beberapa anggota yang memiliki karakteristik hampir sama. Anggota *cluster* tersebut dapat dijelaskan pada Tabel 4 dan Tabel 5

Tabel 4. Anggota Cluster 1

	Cluster 1		Cluster 2	
<b>Jenis Motor</b>	Revo	241	Beat cbs	305
	Scoopy	147	Beat cbs iss	86
	Sonic	21	Beat cw	2801
	Spacy	20	Beat pop cbs	24
	Supra	568	Beat pop cbs iss	18
	Supra GTR	8	Beat pop cw	53
	Vario 110	263	Blade	69
	Vario 125 cbs	1456	CB150R	258
	Vario 125 cbs iss	621		
	Vario 150	432		
<b>Jenis Pembelian</b>	Verza	121		
	Kredit	3317	Kredit	3284
<b>Pekerjaan</b>	Tunai	581	Tunai	330
	Guru/Dosen	44	Guru/Dosen	40
	Ibu Rumah Tangga	701	Ibu Rumah Tangga	636
	Mahasiswa/Pelajar	491	Mahasiswa/Pelajar	562
	Pegawai Negeri	87	Pegawai Negeri	70

Cluster 1			Cluster 2	
	Pegawai Swasta	2006	Pegawai Swasta	1811
	Petani/Nelayan	8	Petani/Nelayan	10
	TNI/POLRI	31	TNI/POLRI	21
			Wiraswata/Pedagang	464
<b>Jumlah anggota cluster 3898 orang</b>			<b>Jumlah anggota cluster 3614 orang</b>	

Tabel 5. Anggota Cluster 3

Cluster 3		
<b>Jenis Motor</b>	Beat cbs	31
	Beat cbs iss	3
<b>Jenis Pembelian</b>	Kredit	6
	Tunai	28
<b>Pekerjaan</b>	Pegawai Negeri	12
	Pegawai Swasta	22
Jumlah anggota cluster 34 orang		

Dari data diatas diperoleh informasi dari bahwa persebaran dari 7546 konsumen sebanyak 3898 atau sebesar 51% masuk di cluster 1, sedangkan sebanyak 3614 atau sebesar 47% masuk dalam anggota cluster 2, dan untuk yang masuk dalam anggota cluster 3 adalah sebanyak 34 atau sekitar 2%. Persebaran hasil cluster dapat dilihat pada Tabel 6. Dari sejumlah 3898 konsumen yang masuk dalam anggota cluster 1 paling banyak membeli type motor Vario 125 cbs dan hampir sebagian besar konsumen yang berada dalam cluster 1 melakukan pembelian sepeda motor secara kredit. Sama halnya pada anggota cluster 2 sebagian besar konsumen melakukan pembelian secara kredit yakni sebanyak 3284 orang dan type motor yang paling banyak dibeli oleh 2801 konsumen adalah type beat cw, selain itu konsumen yang masuk dalam cluster 2 yang melakukan pembelian motor paling banyak bekerja sebagai pegawai swasta sebanyak 1811 orang. Namun jauh berbeda dengan hasil cluster 3 yakni sebanyak 34 orang diantaranya 31 orang membeli sepeda motor type beat cbs dan sebagian besar konsumen tersebut melakukan pembelian secara tunai.

Tabel 6. Persebaran Hasil Clustering

<b>Cluster 1</b>	3898	51%
<b>Cluster 2</b>	3614	47%
<b>Cluster 3</b>	34	2%

### Pengujian Perbandingan

Dengan data dan metode yang sama pengujian dilakukan dengan membandingkan hasil output program dengan hasil output program SPSS. Hasil pengujian clustering dapat dilihat pada Tabel 7.

Tabel 7. Hasil Pengujian Clustering menggunakan SPSS

<b>Cluster 1</b>	3321	43%
<b>Cluster 2</b>	3491	45%
<b>Cluster 3</b>	791	12%

Berdasarkan hasil pengujian Clustering dengan aplikasi dan software SPSS, maka dapat dilihat hasil selisih seperti pada Tabel 9.

<b>Cluster 1</b>	51%	43%	8%	Tabel 9. Hasil Perbandingan Pengujian
<b>Cluster 2</b>	47%	45%	2%	
<b>Cluster 3</b>	2%	12%	10%	
	Rata-Rata Selisih		7%	

## KESIMPULAN

Berdasarkan hasil pembahasan dan pengujian, dapat diambil kesimpulan sebagai berikut :

1. Proses cluster yang dilakukan dengan menggunakan K-Means menghasilkan sebuah informasi gambaran karakteristik konsumen yang telah terkluster atau terkelompok. Proses cluster pada data penjualan di PT Calista Alba menghasilkan tiga cluster/kelompok konsumen berdasarkan tiga variabel yaitu jenis motor, jenis pembelian dan pekerjaan.
2. Hasil dari pengelompokan data konsumen tersebut dapat digunakan sebagai data pendukung untuk divisi marketing dalam menentukan strategi pemasaran.
3. Pengujian perbandingan menggunakan software SPSS menghasilkan jumlah cluster yang sama dengan penelitian yang dilakukan. Namun pada prosentase persebaran jumlah anggota cluster masih terdapat selisih, rata-rata selisih perbandingan kedua aplikasi tersebut adalah sebesar 7%.

## DAFTAR PUSTAKA

- [1] Oscar, Johan Ong.(2013). *Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University*. Bekasi : Jurnal Ilmiah Teknik Industri, Vol.12, No.1
- [2] Rahayu, Ekawati. (2008). *Manajemen Pemasaran Buku Daros*. Kudus : STAIN
- [3] Pratomo, Suwarsito,dkk (2015). *Analisis Regresi Dan Korelasi Antara Pengunjung Dan Pembeli Terhadap Nominal Pembelian Di Indomaret Kedungmundu Semarang Dengan Metodr Kuadrat Terkecil*. Semarang : Jurnal Teknik Informatika
- [4] Budiman, Irwan. (2012). *Data Clustering Menggunakan Metodologi Crisp-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma*. Semarang :Tesis Sistem Informasi
- [5] Ardhyanti, Joanna & Kusumawati, Yupie. (2014). *Data Mining Dengan Metode Clustering Untuk Pengolahan Informasi Persediaan Obat Pada Puskesmas Pandanaran Semarang*. Semarang : Jurnal Sistem Informasi
- [6] Augusta, Y. (2007). *K-Means – Penerapan, Permasalahan dan Metode Terkait*. Jurnal Sistem dan Informatika, Vol. 3: 47-6
- [7] Ong, Johan Oscar. (2013). *Implementasi Algoritma K Means Clustering Untuk Menentukan Strategi Marketing President University*. Jurnal Ilmiah Teknik Industri, Vol.12, No.1, ISSN 1412-6869

*Halaman ini sengaja dikosongkan*