

## SISTEM PENDUKUNG KEPUTUSAN PEMASARAN PRODUK MENGUNAKAN *DATA MINING* DENGAN *K-MEANS CLUSTERING*

Arief Samuel Gunawan<sup>1)</sup>, Evasaria Magdalena Sipayung<sup>2)</sup> Alvin<sup>3)</sup>

<sup>1)</sup>Departemen Sistem Informasi, Institut Teknologi Harapan Bangsa

Jl. Dipati Ukur 80-84, Bandung, 40132

Telp: (022) 2506604, Fax : (022) 2507901

E-mail: arief@ithb.ac.id<sup>1)</sup>, evasaria@ithb.ac.id<sup>2)</sup>

---

### Abstrak

*Sistem informasi yang dapat membantu mengolah pengetahuan (knowledge) dari data penjualan berperan penting untuk mendukung pengambilan keputusan pemasaran di sebuah perusahaan. Studi kasus ini dilakukan di PT. XYZ yang merupakan sebuah perusahaan yang menjual minuman partai besar seperti sirup dan kopi. Target penjualan seringkali tidak tercapai serta terdapat banyak produk yang memiliki potensi penjualan tinggi tetapi hasilnya rendah. Solusi yang diusulkan adalah sistem informasi yang dapat mengolah laporan penjualan dengan menggunakan teknik data mining dengan metode k-means clustering. Proses ini dilakukan terhadap atribut harga satuan maksimum dan rata-rata pembelian per-transaksi dari masing-masing customer. Sistem ini menyajikan pengetahuan mengenai rincian transaksi dan kontribusi dari masing-masing customer, brand, dan informasi hasil clustering. Selain itu terdapat customer breakdown untuk melihat rincian kontribusi brand terhadap masing-masing customer. Sistem Informasi ini juga dapat membantu dalam memberikan usulan terhadap pemasaran produk baru dengan melihat kecocokan dari tiap cluster yang ada dengan atribut harga.*

**Kata kunci:** Data mining, clustering, K-Means Clustering

### Abstract

*An information system which can help to process knowledge from sales data in a company has an important role to support decision making for marketing purposes. This case study was conducted in PT.XYZ, a distribution company selling large-batch drinks like syrup and coffee. The sales target is often underachieved and there are many products with high sales potential but turned out to have low sales volume. The proposed solution is an information system which can process the sales report using data mining and k-means clustering method. This process is conducted to maximum per unit price and average purchase per transaction attributes, from each customer. This system provides knowledge about transaction details and contribution from each customer, brands as well as the clustering result. Besides that, a customer breakdown is also provided to show the contribution of each brands to each customers. This information system can also help to give suggestions of new product marketing by matching the suitability of each clusters with the price attribute.*

**Keywords:** Data mining, clustering, K-Means Clustering

## 1. PENDAHULUAN

PT. XYZ merupakan sebuah perusahaan penyedia peralatan rumah tangga yang juga berfokus untuk menjual minuman partai besar seperti sirup dan kopi, serta beberapa alat yang berhubungan dengan minuman seperti mesin *coffee control*. PT. XYZ saat ini mengoperasikan sebuah aplikasi yang berfungsi untuk mencatat setiap transaksi penjualan dan memberikan data laporan penjualan. Namun, hingga saat ini belum ada pemanfaatan data laporan penjualan yang dilakukan. Sehingga, banyaknya data yang tersedia di laporan penjualan tidak memberikan pengetahuan yang dapat membantu proses pemasaran dan penjualan.

Dengan mengadopsi teknik *direct selling*, perusahaan cenderung menawarkan produk dengan cara menelepon serta mengirimkan brosur satu persatu ke lokasi *customer*. Tercatat hanya 7,5% terjadinya transaksi penjualan yang terjadi dari penawaran produk yang dilakukan oleh tim *marketing* setiap tiga bulannya. Hal ini juga dibuktikan dengan tidak tercapainya target penjualan dari tim *marketing*. Selain itu juga penjualan produk kurang merata, terdapat dominasi produk yang menyebabkan produk yang berpotensi kurang menghasilkan kuantitas penjualan yang tinggi. Berdasarkan hasil wawancara, permasalahan ini dinilai karena kurangnya pengetahuan yang dimiliki oleh tim *marketing* dan para pengambil keputusan mengenai kemampuan dan perilaku dari masing-masing *customer* yang ada, serta sasaran *customer* dalam proses penjualan.

Maka dari itu tim *marketing* dinilai perlu memiliki pengetahuan mengenai kemampuan dan perilaku dari tiap-tiap *customer*, serta keterkaitan *customer* dan produk yang ada agar penawaran produk lebih tepat sasaran.

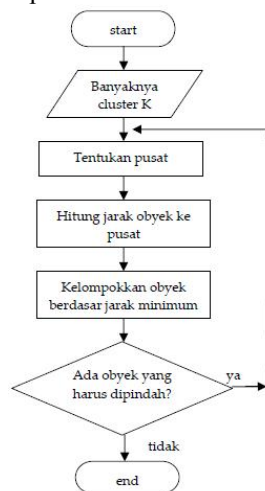
Pengetahuan yang diharapkan ini dapat dicapai dengan pemanfaatan data laporan penjualan yang dimiliki dari PT. XYZ. Informasi setiap transaksi penjualan perlu diolah dan digali sehingga akhirnya menjadi pengetahuan. Dengan adanya pengetahuan tersebut, diharapkan PT. XYZ dapat menciptakan strategi penjualan yang lebih efektif dalam menentukan sasaran penawaran produk yang dapat disesuaikan dengan kemampuan dari masing-masing *customer* yang terdapat pada PT. XYZ

## 2. LANDASAN TEORI

### 2.1 K-Means Clustering

*Cluster* adalah suatu kumpulan dari entitas yang hampir sama. Pengertian lain *cluster* adalah kumpulan dari objek yang mirip dengan objek lainnya dan berada pada kelompok yang sama. Sedangkan proses untuk mengelompokkan data baik itu bersifat fisik atau abstrak kedalam suatu kelompok atau kelas yang memiliki kesamaan sifat disebut *clustering*. Metode *K-Means* merupakan salah satu metode analisis *cluster* non-hierarki yang digolongkan sebagai metode pengelompokan yang bersifat *unsupervised* (tanpa arahan) karena data yang dianalisis tidak mempunyai label kelas, yang berarti dalam proses pengelompokannya tidak mempunyai anggota *cluster* yang pasti. Melainkan objek yang sudah masuk ke dalam *cluster* tertentu masih bisa berpindah ke *cluster* yang lain [1]. Tujuan utama *clustering* adalah untuk menemukan atau mencari pola yang bermanfaat atau berguna pada suatu *database*, kemudian merangkumnya dan membuat lebih mudah untuk dipahami [2].

*Flowchart K-Means Clustering* digambarkan pada Gambar 1.



Gambar 1. Flowchart K-Means

Algoritma *K-Means clustering* memiliki langkah-langkah berikut [3].

1. Menentukan banyaknya/jumlah *cluster* k
2. Menentukan nilai pusat (*centroid*)

Dalam menentukan nilai *centroid* untuk awal iterasi, nilai awal *centroid* dilakukan secara acak. Sedangkan jika menentukan nilai *centroid* yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut

$$\bar{v}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj}$$

dimana :

- $v_{ij}$  adalah *centroid* atau rata-rata *cluster* ke-I untuk variable ke-j
  - $N_i$  adalah jumlah data yang menjadi anggota *cluster* ke-i
  - $i, k$  adalah indeks dari *cluster*
  - $j$  adalah indeks dari variabel
  - $x_{kj}$  adalah nilai data ke-k yang ada di dalam *cluster* tersebut untuk variable ke-j
3. Menghitung jarak antara titik *centroid* dengan titik tiap objek

Untuk menghitung jarak tersebut dapat menggunakan *Euclidean Distance*, yaitu

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Dimana :

- $D_e$  adalah *Euclidean Distance*
- $i$  adalah banyaknya objek,
- $(x, y)$  merupakan koordinat objek dan

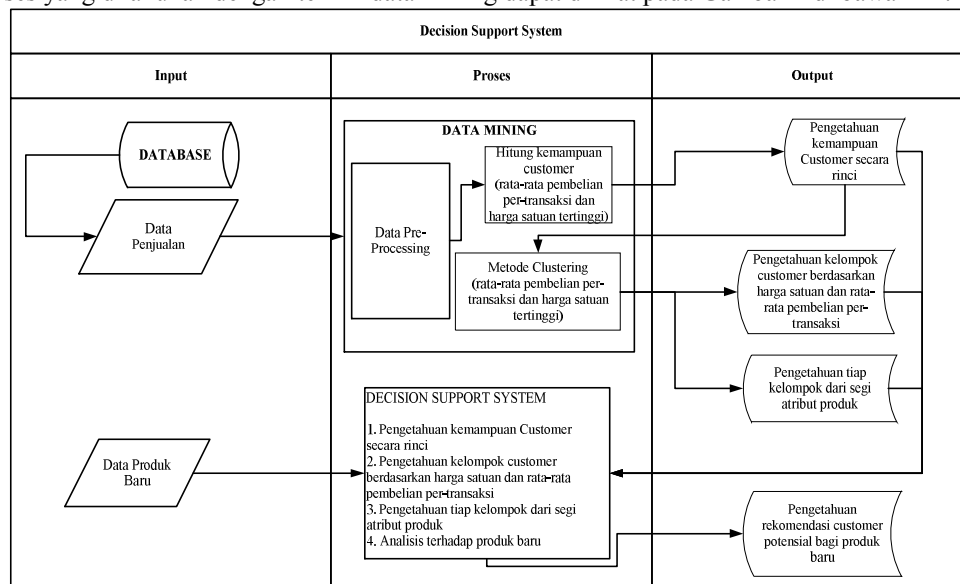
- (s,t) merupakan koordinat *centroid*.
- 4. Pengelompokan objek  
Untuk menentukan anggota *cluster* adalah dengan memperhitungkan jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada distance matriks adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke *cluster* dan nilai 0 untuk data yang dialokasikan ke *cluster* yang lain.
- 5. Kembali ke tahap 2, lakukan perulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke *cluster* lain [6].

### 3. ANALISIS DAN PERANCANGAN SISTEM

#### 3.1 Analisis Kebutuhan Sistem

Sistem pendukung pengambilan keputusan ini pada akhirnya ditujukan untuk memberikan pengetahuan yang dapat membantu tim *marketing* dan para *decision maker* untuk mengetahui kemampuan dan perilaku dari tiap *customer* dari PT. XYZ. Selain itu sistem ini juga dapat membantu dalam memberikan referensi *customer* untuk proses penawaran produk baru. Maka dari itu diperlukan pengelompokan *customer* sesuai dengan homogenitas yang paling mirip, sehingga dapat dilihat karakteristik *customer* berdasarkan kemampuannya dari masing-masing kelompok yang ada. Proses perancangan ini memuat hasil dari pengolahan data laporan penjualan yang dimiliki (laporan penjualan tahun 2012) dengan menggunakan teknik *data mining*. Metode dari *data mining* yang digunakan adalah *clustering*. Proses *clustering* yang digunakan adalah K-Means *clustering* dengan menggunakan data penjualan.

Proses-proses yang dilakukan dengan teknik data mining dapat dilihat pada Gambar 2 di bawah ini.



Gambar 2. Flowchart Sistem

Berdasarkan 6 atribut yang ada pada data laporan penjualan, maka perlu dilakukan analisis atribut-atribut yang terkait dan dapat dimanfaatkan untuk proses *data mining* berikutnya. Menurut analisis faktor yang dilakukan, ada dua atribut yang dapat dimanfaatkan untuk diolah menggunakan teknik *data mining*. Melalui atribut-atribut ini dapat dilihat kemampuan setiap *customer* dalam melakukan pembelian serta melihat kecenderungan produk yang sesuai dari masing-masing *customer*. Atribut-atribut tersebut adalah sebagai berikut:

1. HRGSAT (Harga Satuan): Atribut harga merupakan merupakan salah satu bagian yang sangat penting dalam pemasaran suatu produk karena harga adalah suatu nilai tukar dari produk dan barang manapun [7]. Dalam penelitian ini, atribut harga satuan produk untuk melihat kemampuan *customer* dalam membeli sebuah produk berdasarkan harga tertinggi yang pernah dibeli dalam satu transaksi.
2. NETTO (Total Pembayaran): Atribut *netto* dapat dimanfaatkan untuk melihat pembelian bersih yang dilakukan dari tiap *customer* dalam 1 kali transaksi. NETTO didapatkan dari hasil perkalian perhitungan = (HRGSAT x QTY) – Discount + PPN.

#### 3.2 Proses dan Hasil Clustering

Proses *clustering* akan dilakukan dua kali terhadap dua variabel yang berbeda. Proses yang pertama dilakukan terhadap rata-rata pembelian per-transaksi masing-masing *customer*. Rata-rata pembelian per-transaksi dari masing-masing *customer* diperoleh dari penjumlahan nominal keseluruhan transaksi (*SUM of NETTO*) dibagi dengan jumlah transaksi yang dilakukan oleh masing-masing *customer*. Hasil *clustering* dari variabel ini bertujuan untuk melihat kemampuan *customer* dalam melakukan pembelian dalam satu kali transaksi. Hasil ini juga

ditujukan untuk memberikan *marketing* data sebagai bahan untuk mengevaluasi perilaku customer dalam melakukan pembelian. Proses *clustering* yang kedua dilakukan terhadap variabel kemampuan dari masing-masing *customer* dalam melakukan pembelian berdasarkan harga satuan produk tertinggi (*MAX of Unit Price*). Hal ini bertujuan untuk melihat kemampuan *customer* dalam membeli sebuah produk berdasarkan harga tertinggi yang mampu dibeli. Hasil dari *clustering* ini juga akan digunakan untuk proses pemberian usulan produk baru. Pengelompokan dibagi menjadi 3 yaitu: kelompok di level *high*, *average*, dan *low* untuk memetakan *customer*. Sehingga pada proses *clustering* ini dapat ditentukan bahwa nilai *k* yang digunakan adalah 3.

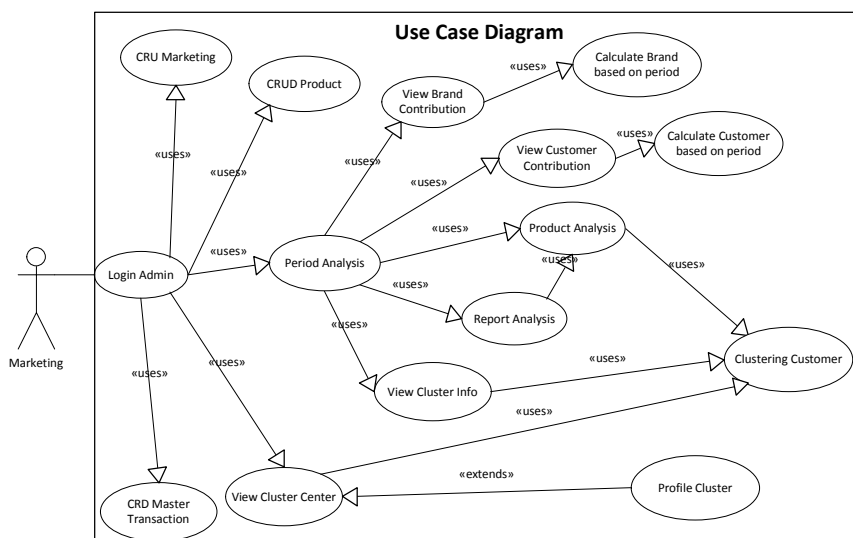
Hasil dari profil *clustering* tersebut dijelaskan pada Tabel 1.

**Tabel 1. Hasil Profiling Cluster**

<p><b>Cluster 1:</b> Cluster ini untuk harga satuan dan rata-rata pembelian <i>low</i>. Harga satuan untuk cluster ini adalah antara Rp 95.000 – Rp 750.000 dan rata-rata pembelian selama setahun kurang dari Rp 3.956.793.</p>
<p><b>Cluster 2:</b> Cluster ini untuk harga satuan dan rata-rata pembelian <i>average</i>. Harga satuan untuk cluster ini adalah antara Rp 750.000 – Rp 30.700.000 dan rata-rata pembelian selama setahun antara Rp dari Rp 3.956.793 – 25.097.2287.</p>
<p><b>Cluster 3:</b> Cluster ini untuk harga satuan dan rata-rata pembelian <i>high</i>. Harga satuan untuk cluster ini adalah di atas Rp 30.700.000 dan rata-rata pembelian selama setahun lebih besar dari Rp 125.548.411.</p>

### 3.3 Perancangan Sistem

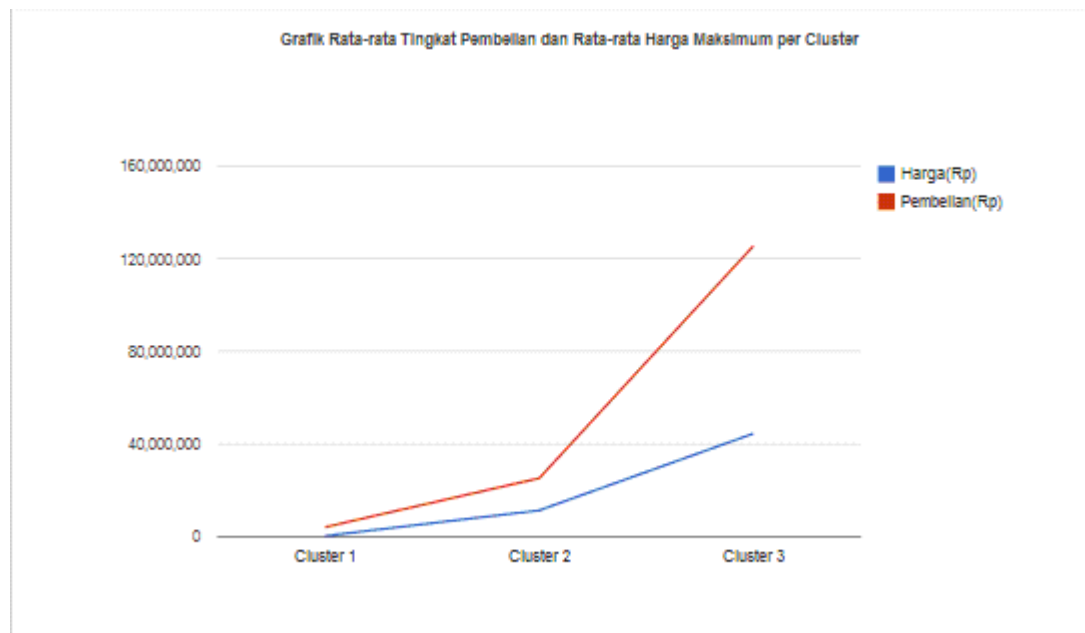
*Use case diagram* menggambarkan fungsionalitas yang diharapkan dari sebuah sistem dari sudut pandang *user*, siapa saja aktor yang terlibat dalam sistem serta kegiatan-kegiatan yang dapat dilakukan oleh aktor tersebut. *Use case diagram* sistem pendukung pengambilan keputusan untuk pemasaran produk yang dikembangkan pada Gambar 4.



**Gambar 4. Use Case Diagram**

## 4. IMPLEMENTASI

Pada bagian ini, memperlihatkan tampilan dari aplikasi yang telah dikembangkan. Ketiga *cluster customer* berdasarkan harga satuan dan rata-rata pembelian diperlihatkan pada Gambar 5.



Gambar 5. Cluster Customer

Untuk mengetahui siapa saja *customer* yang berada di *cluster* harga dan rata-rata pembelian diperlihatkan pada Gambar 6.

Customer	Area	Harga Maksimum	Pembelian per-Transaksi	Cluster-Harga	Cluster-Pembelian	Kategori
CST001	Utara	Rp. 200.000	Rp. 141.365	Cluster 1	Cluster 1	Low-Low
CST002	Timur	Rp. 115.000	Rp. 115.000	Cluster 1	Cluster 1	Low-Low
CST003	Utara	Rp. 21.500.000	Rp. 9.557.500	Cluster 2	Cluster 3	Average-High
CST004	Utara	Rp. 200.000	Rp. 163.571	Cluster 1	Cluster 1	Low-Low
CST005	Utara	Rp. 260.000	Rp. 191.100	Cluster 1	Cluster 1	Low-Low
CST006	Barat	Rp. 139.000	Rp. 152.888	Cluster 1	Cluster 1	Low-Low
CST007	Utara	Rp. 27.200.000	Rp. 10.165.000	Cluster 2	Cluster 3	Average-High
CST008	Utara	Rp. 231.500	Rp. 183.588	Cluster 1	Cluster 1	Low-Low
CST009	Utara	Rp. 9.500.000	Rp. 373.550	Cluster 2	Cluster 2	Average-Average
CST010	Timur	Rp. 375.000	Rp. 321.473	Cluster 1	Cluster 1	Low-Low

Showing 1 to 10 of 226 entries Previous Next

Gambar 6. Detail Customer per Cluster

Sistem ini dapat memberikan referensi *customer* untuk produk baru yang akan dijual dengan melihat kesesuaian dari tiap *cluster* yang ada dengan atribut harga.

Product Analysis

Product List

NamaProduct17

BrandLAZ

JenisKE

HargaRp. 47.500.000

Cluster3

[View Customer](#)

Gambar 7. Product Analysis

## 5. KESIMPULAN

Berdasarkan penelitian yang dilakukan, maka kesimpulan yang dapat diambil adalah sebagai berikut:

1. Sistem informasi dengan menggunakan metode *K-Means clustering* dapat digunakan untuk pengelompokan terhadap dua atribut, yaitu harga satuan dan total pembayaran (netto) dan dengan *customer* dikelompokkan menjadi tiga kelompok dengan kategori kemampuan *low*, *average*, dan *high*. *Cluster low* dengan harga satuan antara Rp 95.000-Rp 750.000 dan rata-rata pembelian selama setahun kurang dari Rp 3.956.793, *cluster medium* dengan harga satuan antara Rp 750.000-Rp 30.700.000 dan rata-rata pembelian selama setahun antara dari Rp 3.956.793-25.097.2287, dan *cluster high* dengan harga satuan di atas Rp 30.700.000 dan rata-rata pembelian selama setahun di atas Rp 3.956.793.
2. Sistem informasi yang dikembangkan yang dapat mendukung aktivitas tim *marketing* dan para *decision maker* dalam proses pemasaran dilakukan dengan memuat pengetahuan mengenai *customer* untuk melihat kemampuan dari masing-masing *customer* dalam melakukan pembelian.

## 6. DAFTAR RUJUKAN

- [1] McLeod, Raymon (1998). *Management Information Systems. 7th Edition*, New Jersey : Prentice Hall, Inc.
- [2] Soraya, Yani (2011). Perbandingan Kinerja Metode *Single Linkage*, Metode *Complete Linkage* Dan Metode *K-means* Dalam Analisis *Cluster*. Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
- [3] Wakhidah, Nur (2009). *Clustering Menggunakan K-Means Algorithm*. Jurnal TRANSFORMATIKA Vol.7/ No.1/ Juli 2010.