

## Penerapan Teknik Data Mining untuk *Clustering* Armada pada PT. Siaga Transport Indonesia Menggunakan Metode *k-Means*

Nuzul Hikmah <sup>1)</sup>, Dyah Ariyanti <sup>2)</sup>, Melista Sugesti <sup>3)</sup>  
*Universitas Panca Marga Probolinggo* <sup>1)</sup>  
*n.hikmah1807@upm.ac.id*

**Abstrak** – PT. Siaga Transport Indonesia adalah perusahaan yang bergerak dibidang ekspedisi. Armada menjadi aset penting untuk perusahaan yang bergerak di bidang ini. Beratnya muatan yang diangkut, jauhnya jarak yang ditempuh, dan usia armada menjadi faktor penyebab kerusakan armada. Hal ini berpengaruh terhadap potensi armada. Untuk itu diperlukan *analysis clustering* menggunakan metode *k-means* untuk mengetahui armada yang potensial dan yang kurang potensial. *Clustering* merupakan teknik pengelompokan objek yang didasarkan pada kemiripan antar objek dalam satu *cluster*. *K-means* merupakan metode *clustering* yang mempunyai kemampuan mengelompokkan data dalam jumlah cukup besar dengan waktu komputasi yang relatif cepat dan efisien. Dalam penelitian ini sistem akan mengelompokkan 33 armada dengan menggunakan variabel biaya operasional, tahun kendaraan, produktivitas armada dan kilometer armada menjadi 3 kelompok dengan kategori *cluster* armada berpotensi rendah, sedang dan tinggi. Dari hasil akhir sistem diketahui *cluster* 1 atau armada dengan potensi rendah ada 7 armada, *cluster* 2 atau armada dengan potensi sedang ada 14 armada, *cluster* 3 atau armada dengan potensi tinggi ada 12 armada. Pengujian dilakukan dengan metode *silhouette coefficient* dengan rentang variasi nilai yang dihasilkan yaitu dibawah nol sampai dengan 0,9.

**Kata kunci:** *Data Mining, Clustering, K-means*

### 1. PENDAHULUAN

#### Latar Belakang

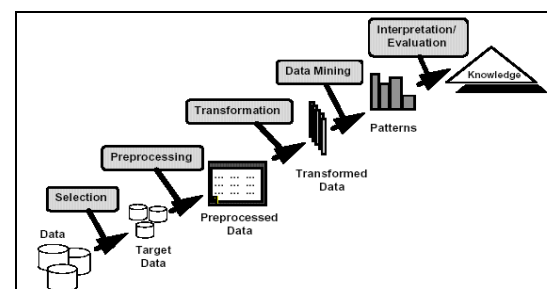
PT. Siaga Transport Indonesia merupakan perusahaan yang bergerak dibidang ekspedisi. Tujuan umum didirikannya sebuah perusahaan adalah untuk menghasilkan barang, jasa yang menjadi kebutuhan konsumen. Armada menjadi aset penting untuk perusahaan yang bergerak di bidang ini sehingga harus diperhatikan kondisi setiap armada. Beratnya muatan yang diangkut dan jauhnya jarak yang ditempuh oleh setiap armada, dapat membuat armada mengalami kerusakan. Usia armada juga menjadi faktor penyebab kerusakan armada. Hal ini sangat berpengaruh terhadap kinerja armada karena produktivitas armada akan menurun, sehingga keuntungan yang didapat oleh perusahaan akan mengalami penurunan pula. Untuk mengatasi hal tersebut, pimpinan perusahaan melakukan regenerasi terhadap kendaraan dengan menjual armada yang telah mengalami penurunan potensi. Dalam menentukan armada yang akan dijual, perlu dilakukan identifikasi berdasarkan beberapa faktor yang mempengaruhi kondisi kendaraan. Faktor ini meliputi tahun kendaraan yang berkaitan dengan usia kendaraan, produktivitas kendaraan dalam satu tahun terakhir, biaya operasional dan kilometer kendaraan. Salah satu cara untuk mempermudah dalam pengelompokan armada

adalah dengan menggunakan teknik data mining yaitu dengan kluster analisis (*analysis cluster*).

### 2. DASAR TEORI

#### **Knowledge discovery in databases (KDD)**

*Knowledge discovery in databases* (KDD) adalah keseluruhan proses *non-trivial* untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. Proses KDD itu ada 5 tahapan yang dilakukan secara terurut, yaitu:



Gambar 1. Tahapan KDD

#### a. Data selection

Pemilihan (seleksi) data dari sekumpulan data operasional.

#### b. *Pre-processing*

Sebelum proses data mining dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi

fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

*c. Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining.

*d. Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

*e. Interpretation/evaluation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

**Clustering**

*Clustering* adalah proses mengelompokkan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas-nya. Penerapan metode *clustering* menghasilkan pengetahuan berupa penentuan beberapa *cluster* catatan data yang memiliki kemiripan atribut.

**K-Means**

Algoritma *K-Means* merupakan sebuah metode sederhana, cepat dan efisien untuk pengelompokkan data dalam suatu angka spesifik dari *cluster*, yaitu *k*. Algoritma *K-Means* dimulai dengan pemilihan secara acak *K*, *K* disini merupakan banyaknya cluster yang ingin dibentuk. Kemudian tetapkan nilai-nilai *K* secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan centroid. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil).

**Algoritma K-Means**

1. Menentukan jumlah cluster
2. Menentukan nilai centroid dalam menentukan nilai centroid untuk awal iterasi, nilai awal centroid dilakukan secara acak.
3. Menghitung jarak antara titik centroid dengan titik tiap objek Untuk menghitung jarak tersebut dapat menggunakan Euclidean Distance, yaitu

$$D_e = \sqrt{(x_i - s_j)^2 + (y_i - t_j)^2}$$

4. Pengelompokan objek untuk menentukan anggota cluster adalah dengan memperhitungkan jarak minimum objek.

5. Kembali ke tahap 2, lakukan perulangan hingga nilai centroid yang dihasilkan tetap dan anggota cluster tidak berpindah kecluster lain.

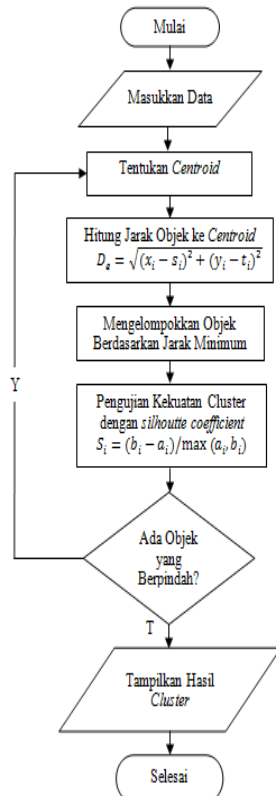
**Metode Pengujian Silhouette Coefficient**

Metode ini merupakan metode validasi *cluster* yang menggabungkan metode *cohesion* dan *separation*. Penggunaan metode ini untuk menilai kualitas serta kekuatan *cluster*. *silhouette coefficient* dari tiap objek dalam suatu *cluster* adalah suatu ukuran yang menunjukkan seberapa ketat data dikelompokkan dalam *cluster* tersebut. Dalam hal ini ditinjau pula seberapa baik suatu objek ditempatkan dalam suatu *cluster* tertentu. Berikut ini merupakan ukuran nilai *silhouette* menurut Kaufman dan Rousseeuw:

Tabel 1. Ukuran Nilai Silhouette

Nilai <i>silhouette coefficient (SC)</i>	Keterangan
$0.7 < SC \leq 1$	<i>strong structure</i>
$0.5 < SC \leq 0.7$	<i>medium structure</i>
$0.25 < SC \leq 0.5$	<i>weak structure</i>
$SC \leq 0.25$	<i>no structure</i>

### 3. METODE PENELITIAN



Gambar 2. Diagram Alir Sistem

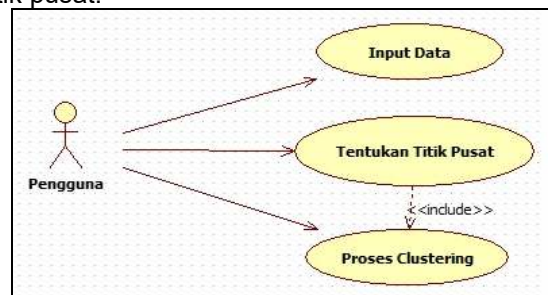
Tahap pertama adalah *input* data, data yang digunakan adalah data training yang telah melalui tahap *pre-processing* sehingga menjadi data dengan format excel yang dapat diproses dalam sistem clustering data. Tahap selanjutnya adalah me-*nentukan* titik *centroid* awal dimana titik *centroid* awal ditentukan dari hasil perhitungan rata-rata data sampel. Selanjutnya adalah me-*nghitung* jarak antara titik *centroid* dengan titik tiap objek menggunakan persamaan *euclidean distance*. Tahap selanjutnya adalah me-*ngelompokkan* objek berdasarkan jarak ter-*kecil* objek ke titik *centroid*. Semakin kecil jarak objek ke *centroid*, semakin baik objek tersebut berada di kelas *centroid*-nya. Langkah terakhir adalah pengujian, dimana pada tahapan ini pengujian dilakukan menggunakan metode *silhouette coefficient*. jika hasil perhitungan baru berbeda dengan hasil perhitungan sebelumnya, maka ulangi untuk menentukan *centroid* baru dengan persamaan sebagai berikut:

$$\text{Pusat cluster baru} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\text{jumlah objek pada cluster}}$$

Namun apabila hasil perhitungan baru sama dengan hasil perhitungan sebelumnya, maka selesailah perhitungannya dan tampilkan hasil *cluster* terakhir.

### Use Case Diagram

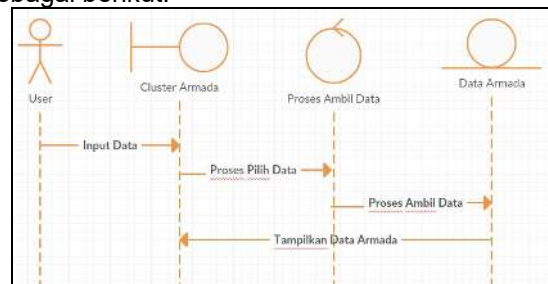
Pengguna sistem ini sebanyak satu orang. Fungsi yang dapat dijalankan oleh pengguna pada sistem ini adalah fungsi memasukkan data. Data yang dimaksud disini adalah data training pada 33 kendaraan. Fungsi berikutnya adalah menentukan titik pusat. Titik pusat ini didapat dari hasil rata-rata dari masing-masing variabel pada data sampel. Kemudian fungsi proses *Clustering* dimana pada fungsi ini dilakukan perhitungan jarak objek dengan titik *centroid* menggunakan metode *Euclidean Distance* dan pengujian dengan metode *silhouette coefficient* untuk menilai kualitas serta kekuatan *cluster*. Fungsi proses cluster sangat bergantung pada fungsi penentuan titik pusat.



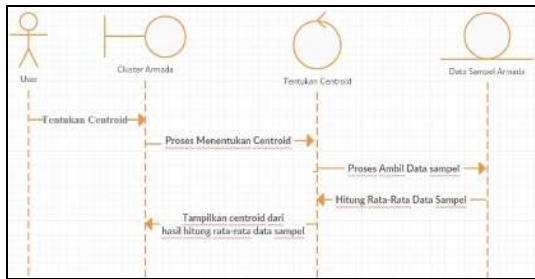
Gambar 3. Use Case Diagram

### Sequence Diagram

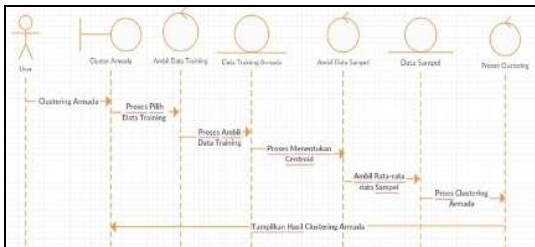
*Sequence Diagram* menggambarkan urutan interaksi antar objek ketika satu *use case* dilakukan, sehingga *sequence diagram* sebaiknya sebanyak *use case* nya. Ada 3 tahap *Sequence diagram* untuk sistem ini, yaitu proses pertama adalah *input* data, selanjutnya proses menentukan titik *centroid* yang dilakukan oleh user dan terakhir adalah proses *clustering*. Masing-masing *Sequence Diagram* dapat dilihat sebagai berikut:



Gambar 4. Sequence Diagram input data



Gambar 5. Sequence Diagram centroid



Gambar 6. Sequence Diagram proses Cluster

**4. HASIL DAN PEMBAHASAN**

**Pre-Processing Data**

Sebelum melakukan data mining perlu dilakukan *pre-processing* terhadap data sampel dan data uji yang masing-masing berjumlah kurang lebih 3500 data dalam satu tahun. Tujuan dari proses *pre-processing* data ini adalah untuk memastikan data yang akan diolah di data mining adalah data yang baik dan siap diproses pada data mining. Data sampel diambil dari Januari 2016 samapi dengan Desember 2016. Sedangkan data uji diambil dari bulan Januari 2017 sampai dengan Desember 2017.

**Proses Clustering K-Means**

1. Langkah awal metode *k-means* yaitu menentukan jumlah *cluster* yang diinisialkan dengan “K”. Dalam hal ini penulis menentukan ada 3 kategori *cluster*. Dengan ketentuan untuk *cluster* 1 adalah armada dengan potensi rendah, *cluster* 2 adalah armada dengan potensi sedang dan *cluster* 3 merupakan armada dengan potensi tinggi.
2. Selanjutnya langkah kedua dalam algoritma *k-means* adalah menentukan *centroid*. *Centroid* diambil dari hasil rata-rata masing-maing variabel pada data sampel. Dimana data sampel dibagi menjadi 3 golongan berdasarkan tahun kendaraan sebagai berikut:
  - a. Kendaraan yang telah berumur lebih dari atau sama dengan 9 tahun, maka kendaraan tersebut dikategorikan sebagai kendaraan yang kurang potensial.
  - b. Kendaraan yang telah berumur lebih dari atau sama dengan 5 tahun dan kurang dari 9 tahun, maka kendaraan tersebut

dikategorikan sebagai kendaraan dengan potensi sedang.

- c. Kendaraan yang telah berumur kurang dari atau sama dengan 4 tahun, maka kendaraan tersebut dikategorikan sebagai kendaraan dengan potensi tinggi.

Sehingga didapat hasil *centroid* seperti pada tabel berikut ini:

Tabel 2. Centroid Pertama

3. Selanjutnya menghitung jarak tiap objek ke titik *centroid* menggunakan persamaan *euclidean distance* sebagai berikut:

$$D(X_n, C_m) = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Hasil dari perhitungan jarak objek ke *centroid* untuk iterasi pertama adalah sebagai berikut:

Tabel 3. Perhitungan Euclidean Distance

No	No. Pol	Perhitungan Jarak Objek Ke Centroid		
		C1	C2	C3
1	8552	28052512	37930651	42967325
2	8003	12166621	22047001	27085814
3	8943	7963153	2021365	7012790
4	8445	3278503	13158434	18197996
5	8523	2420003	12300125	17339906
...	...	...	...	...

4. Tahap selanjutnya adalah mengelompokkan objek berdasarkan jarak minimum pada setiap iterasi untuk mengetahui objek menempati *cluster* ke berapa.

Tabel 4. Jarak Minimum

No	No. Pol	Jarak Minimum	Cluster
1	8552	28052512	1
2	8003	12166621	1
3	8943	2021365	2
4	8445	3278503	1
5	8523	2420003	1
...	...	...	...

Apabila hasil *cluster* pada iterasi yang baru tidak sama dengan hasil *cluster* pada iterasi sebelumnya maka, tentukan *centroid* baru dan hitung kembali jarak objek dengan *centroid* yang baru. Apabila hasil *cluster* pada iterasi yang baru sama dengan hasil *cluster* pada

iterasi sebelumnya maka, proses *clustering* berakhir dan lanjutkan untuk pengujian kualitas *cluster* tersebut.

5. Untuk menilai kualitas serta kekuatan *cluster* dilakukan pengujian dengan menggunakan metode *silhouette coefficient*.

### Pengujian Sistem

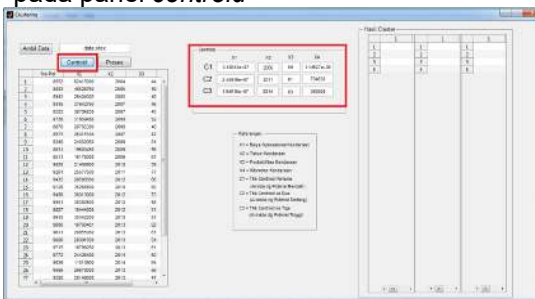
Langkah-langkah untuk menjalankan aplikasi pada tampilan Menu Utama adalah sebagai berikut:

1. Memilih data training yang telah melalui *pre-processing* data dan siap di lakukan proses *clustering*.



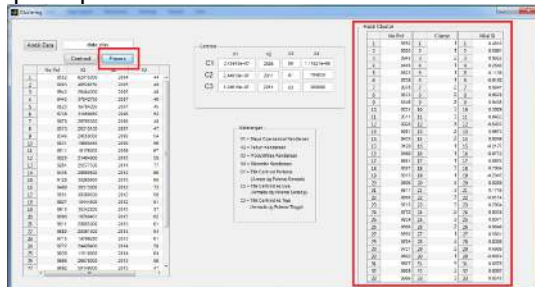
Gambar 7. Tampilan Proses Input Data

2. Mengambil centroid dari perhitungan rata-rata data sampel yang hasilnya akan ditampilkan pada panel *centroid*



Gambar 8. Tampilan Proses Centroid

3. Hasil dari proses *clustering* ini akan ditampilkan pada panel hasil *cluster*.



Gambar 9. Tampilan Proses Clustering

### 5. KESIMPULAN

Berdasarkan hasil penelitian yang penulis lakukan tentang *clustering* armada menggunakan metode

*k-means* dalam mengelompokkan armada disimpulkan bahwa:

1. Metode *k-means clustering* dapat diterapkan pada data *clustering* armada di PT.Siaga Transport Indonesia sehingga metode ini sangat membantu dalam pengambilan keputusan untuk menjual armada
2. Objek penelitian dikelompokkan kedalam 3 (tiga) kelompok dengan hasil pengelompokkan sebagai berikut:
  - a. *Cluster* 1 untuk kategori kendaraan dengan potensi rendah beranggotakan 7 armada.
  - b. *Cluster* 2 untuk kategori kendaraan dengan potensi sedang beranggotakan 14 armada.
  - c. *Cluster* 3 untuk kategori kendaraan dengan potensi tinggi beranggotakan 12 armada.
3. Dari hasil pengelompokkan armada menggunakan metode *k-means* pada iterasi terakhir, diketahui bahwa armada yang berada pada *cluster* 1 ada 7 data dengan 4 (empat) data yang *error* atau tidak sesuai. Sehingga nilai akurasi yang diperoleh pada pengujian ini sebesar 87,88%.

### 6. PUSTAKA

Andayani, Sri. 2007. *Pembentukan cluster dalam Knowledge Discovery in Database dengan Algoritma K-Means*.

Bahar, Apriadi. 2016. *Penentuan Strategi Penjualan Alat-Alat Tattoo Di Studio Sonyx Tattoo Menggunakan Metode K-Means Clustering*

Fahrudin, Alex. Bambang Eka Purnama dan Berliana Kusuma Riasti. 2011. *Pembangunan Sistem Informasi Layanan Haji Berbasis Web Pada Kelompok Bimbingan Ibadah Haji Ar Rohman Maburur Kudus*

Firdaus, Gita Rosyada. 2017. *Clustering Untuk Pemetaan Daerah Potensial Budidaya Perikanan Di Jawa Timur Menggunakan Algoritma Fuzzy C-Means*

Handoko, Koko. 2016. *Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus di Program Studi TKJ Akademi Komunitas Solok Selatan)*.

Hendini, Ade. 2016. *Pemodelan UML Sistem Informasi Monitoring Penjualan Dan Stok Barang (Studi Kasus: Distro Zhezha Pontianak)*

Hidayat, Rahman., Rochid Wasono dan Moh. Yamin Darsyah. 2017. *Pengelompokan Kabupaten Di Jawa Tengah Menggunakan Metode K-Means dan Fuzzy C-Means*.

Hikmawati, Nur Lailatul. 2017. *Pengenalan Ucapan Menggunakan Metode MFCC dan Algoritma Backpropagation*.

- J, W, Satzinger., Jackson R. B dan Burd S. D. 2011. *Systems Analysis and Design in a Changing World, Sixth ed.*
- L, Kaufman and P. J. Rousseuw. 1990. *Finding Groups in Data*. New York: John Wiley & Sons.
- Murti, Mikael Aditya Wahyu Krisna. 2017. *Penerapan Metode K-Means Clustering Untuk Mengelompokan Potensi Produksi Buah-Buahan di Provinsi Daerah Istimewa Yogyakarta*.
- Purba, Yugi Trianto. 2008. *Penerapan Data Mining Untuk Menemukan Pola Antara Nilai Ujian Saringan Masuk Terhadap Indeks Prestasi*.
- Saragi, Cytra Sari Rewati. 2017. *Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Parameter Nilai Magang Dan Job Description Pada Jurusan Teknik Informatika Politeknik Negeri Batam*.
- Siregar, Gellysa Urva dan Helmi Fauzi. 2015. *Pemodelan UML E-Marketing Minyak Goreng*
- Zahrotun, Lisna. 2015. *Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan Metode Clustering K-Means Dan Agglomerative Hierarchical Clustering (Ahc)*.