

KLUSTERISASI OPTIMAL DENGAN ELBOW METHOD UNTUK PENGELOMPOKAN DATA KECELAKAAN LALU LINTAS DI KOTA SEMARANG

by Arif Jananto

Submission date: 05-Jan-2023 01:20PM (UTC+0700)

Submission ID: 1988753121

File name: GELOMPOKAN_DATA_KECELAKAAN_LALU_LINTAS_DI_KOTA_SEMARANG_Vada.pdf (541.5K)

Word count: 3290

Character count: 19287

KLUSTERISASI OPTIMAL DENGAN *ELBOW METHOD* UNTUK PENGELOMPOKAN DATA KECELAKAAN LALU LINTAS DI KOTA SEMARANG

Vada Annisa Ekasetya¹, Arief Jananto²

^{1,2}Sistem Informasi, Fakultas Teknologi Informasi, Universitas Stikubank
e-mail : ¹vada.annisaekasetya@gmail.com dan ²ajananto09@edu.unisbank.ac.id

ABSTRAK

Kota Semarang merupakan kota metropolitan terbesar ke-5 se-Indonesia dengan angka kecelakaan lalu lintas yang dapat dikatakan tinggi. Dalam laporan tahunan Satlantas Kota Semarang dari tahun 2017, 2018 berurutan telah terjadi sebanyak 936 kasus dan 1.008 kasus. Kemudian pada tahun 2019 dengan hasil rekap sampai dengan bulan Agustus terjadi sebanyak 909 kasus.

Untuk mengetahui pola kecelakaan dalam mengolah data Laka Lintas dapat menggunakan sebuah metode penambangan data yang disebut data mining. Menggunakan teknik clustering dengan algoritma *K-Means* untuk menghasilkan pola dalam ekstraksi informasi, yang dikombinasikan dengan teknik optimisasi kluster yaitu metode *Elbow*.

Penelitian ini menggunakan proses uji coba manual dan aplikasi *RStudio*. Perhitungan manual menggunakan uji coba jumlah $K=2$ sampai $K=10$ dengan penerapan metode *Elbow* menggunakan perhitungan nilai *SSE* (*Sum of Square Error*) pada tiap kluster. Pada grafik selisih *SSE* tiap kluster menunjukkan patahan siku pada jumlah $K=3$. Perhitungan komputasi dengan aplikasi *RStudio* menghasilkan grafik selisih *SSE* dengan patahan siku pada jumlah $K=3$. Kemudian dengan menggunakan jumlah $K=3$ diimplementasikan kedalam Algoritma *K-Means* pada *RStudio*. Berdasarkan analisis yang dilakukan dengan data Laka Lintas periode bulan Oktober 2018 sampai bulan Juli 2019 menghasilkan tiga segmentasi data.

Kata Kunci : Kecelakaan lalu Lintas, Data Mining, Clustering, Segmentasi, Metode *Elbow*.

1. PENDAHULUAN

Kota Semarang merupakan kota metropolitan terbesar ke-5 se-Indonesia dengan angka kecelakaan lalu lintas yang dapat dikatakan tinggi. Jumlah penduduk Kota Semarang saat ini telah mencapai 1.815.729 jiwa [1]. Dalam laporan tahunan Satlantas Kota Semarang dari tahun 2017 terjadi sebanyak 936 kasus dengan jumlah 211 korban MD, 2 korban LB dan 968 korban LR. Kemudian tahun 2018 terjadi sebanyak 1.008 kasus, sekitar 189 korban meninggal dunia, 5 korban luka berat, dan 1.060 korban luka ringan. Kemudian pada tahun 2019 dengan hasil rekap sampai dengan bulan Agustus terjadi sebanyak 909 kasus dengan 134 korban MD, 1 korban LB, dan 1.034 korban LR

Untuk mengetahui pola kecelakaan dalam mengolah data Laka Lintas dapat menggunakan sebuah metode penambangan data yang disebut data mining. Menggunakan teknik clustering dengan algoritma *K-Means* untuk menghasilkan pola dalam ekstraksi informasi, yang dikombinasikan dengan teknik optimisasi kluster yaitu metode *Elbow*. Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [2]. Clustering adalah proses pengelompokan dari sebuah set objek – objek fisik maupun abstrak yang dimasukkan kedalam kelas – kelas dengan objek – objek yang serupa [3]. Kemudian Algoritma *K-Means* merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok [4].

Penjabaran dari latar belakang membentuk suatu rumusan masalah, bagaimana karakteristik data korban kecelakaan lalu lintas berdasarkan kapan terjadinya (hari, waktu), kemudian karakteristik korbannya (umur, jenis kelamin dan pekerjaan) dan kendaraan yang terlibat dengan menerapkan metode *Elbow* dan algoritma *K-Means*.

2. TINJAUAN PUSTAKA

2.1 Pustaka yang Terkait dengan Penelitian

Penelitian pertama dijelaskan bahwa telah terjadi disrupsi dalam sistem pembelajaran perguruan tinggi yang banyak yang telah beralih menggunakan teknologi e-learning. Dari pengujian akhir mendapatkan hasil sebanyak 5 cluster dan dengan menggunakan metode elbow mendapatkan hasil yang lebih baik tingkat kemiripannya [6]. Penelitian kedua dijelaskan bahwa dalam perindustrian Indonesia, UKM tidak memiliki pemetaan pelanggan atau *Customer mapping*. Perhitungan hasil *SSE* pada 500 cluster

dari pengunjung peminat batik, menunjukan bahwa lekukan tajam menyiku terdapat pada $K=3$ [7]. Penelitian ketiga dijelaskan bahwa dalam proses peningkatan mutu pendidikan salah satu upaya pemerintah adalah Ujian Nasional. Penulis menggunakan 4 dataset yang berbeda menggunakan metode *elbow* dari grafik SSE. Hasil *cluster* terbaik dataset 1 dengan jumlah data 34 adalah 3 *cluster*, dataset 2 dengan jumlah data 68 adalah 4 *cluster*, dataset 3 dengan jumlah data 102 adalah 3 *cluster*, dan dataset 4 dengan jumlah data 137 adalah 3 *cluster* [8]. Penelitian keempat dijelaskan bahwa WSN atau *Wireless Sensor Network* dalam penerapannya memiliki ratusan ribu node sensor kecil yang digunakan untuk mendeteksi parameter lingkungan atau fisiologis, seperti temperatur, tekanan dll. saling terkoneksi secara *wireless*. Menggunakan algoritma K-means dengan optimasi metode *elbow* dalam menemukan 'k' terbaik. Hasil dari penelitian adalah jumlah *cluster* terbaik menekuk tajam berada di $K=3$ menggunakan pemrosesan *software* Matlab [9]. Penelitian kelima dijelaskan bahwa dalam penelitiannya menggunakan algoritma K-means dengan optimasi metode *elbow*, mengambil data selama 3 tahun terakhir untuk mahasiswa aktif pada Sekolah Tinggi Agama Hindu Negeri Tampung Penyang Palangkaraya. Uji coba 1 dengan 54 data mahasiswa menunjukkan $K=3$, uji coba 2 dengan 139 data mahasiswa menunjukkan $K=3$, uji coba 3 dengan 261 data mahasiswa menunjukkan $K=3$, dan uji coba 4 dengan 885 data mahasiswa menunjukkan $K=3$ [10].

2.2 Metodologi CRISP-DM

2.2.1 Pengertian Data Mining

Data Mining adalah proses penemuan *interesting knowledge* dari data yang berjumlah sangat besar. Sumber data yang besar pada data mining dapat berasal dari *database*, data *warehouse*, data transaksi, dan tipe data tingkat lanjut lainnya. Data Mining berfungsi untuk menentukan jenis pola atau pengetahuan [2].

2.2.2 Tahapan CRISP-DM

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah metodologi data mining yang komprehensif dan merupakan model proses yang menyediakan bagi pemula maupun pakar data mining sebuah konstruksi lengkap untuk melakukan proyek data mining dengan model referensi pada Gambar 1[5].



Gambar 1. Model Referensi fase-fase dari CRISP-DM

1. Business Understanding Phase (Fase Pemahaman Bisnis) berfokus pada pemahaman dari tujuan proyek dari prespektif bisnis.
2. Data Understanding Phase (Fase Pemahaman Data) melibatkan empat langkah diantaranya : Pengumpulan data utama bertujuan untuk beradaptasi dalam meningkatkan pemahaman terhadap data; Mendeskripsikan data; Eksplorasi data; dan verifikasi kualitas dari data.
3. Data Preparation Phase (Fase Persiapan Data) mencakup semua kegiatan untuk membangun set data final atau data yang akan dimasukan kedalam tools pada fase modeling dari inisiasi data mentah.
4. Modeling Phase (Fase Pemodelan) berbagai macam teknik untuk pemodelan dapat dipilih dan diterapkan dengan parameter yang disesuaikan sampai dengan hasil paling optimal.
5. Evaluation Phase (Fase Evaluasi) merupakan fase untuk mereview dan mengevaluasi hasil model dan konstruksi pada proses – proses sebelumnya untuk menjelaskan kembali pencapaian dari tujuan bisnis.
6. Deployment Phase (Fase Penyebaran) merupakan fase untuk mengatur dan mempresentasikan hasil dari pencarian pengetahuan pada proses data mining sebelumnya yang mudah untuk dipahami.

2.3 Metode Clustering

Clustering adalah proses pengelompokan dari sebuah set objek – objek fisik maupun abstrak yang dimasukkan kedalam kelas – kelas dengan objek – objek yang serupa yang terbagi menjadi 4 kategori, yaitu : *partitioning method, hierarchical method, density-based method, dan grid-based method* [2].

2.4 Metode Optimasi Kluster

2.4.1 Metode Elbow

Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* yang akan membentuk siku pada suatu titik.[10].

2.4.2 Sum of Square Error (SSE)

Sum of Square Error (SSE) sering digunakan dalam acuan penelitian yang berhubungan dengan optimalisasi kluster. Akan ada kluster yang mengalami penurunan paling ekstrim yang dilanjutkan dengan nilai K yang akan turun secara perlahan – lahan sampai nilai K tersebut stabil dengan persamaan pada rumus 1[10].

$$SSE = \sum_{i=1}^k \sum_{x_i \in C_i} D(x_i, C_i)^2 \quad (1)$$

Dimana :

- k = Jumlah kluster
- $x_i \in C_i$ = Nilai keanggotaan titik data x_i ke pusat kelompok C_i
- C_i = Pusat kluster ke- i
- $D(x_i, C_i)$ = Jarak dari titik x_i ke kelompok C_i yang diikuti

2.5 Algoritma K-Means

Algoritma *K-Means* merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok [4].

Berikut adalah langkah pengelompokan menggunakan Algoritma *K-Means* yang akan diterapkan pada penelitian ini :

1. Menentukan jumlah kluster
2. Menentukan titik pusat (*centroid*) pada tiap kluster yang diuji coba
3. Menghitung jarak obyek ke *centroid*
Perhitungan jarak obyek ke *centroid* pada tiap kluster pada penelitian ini menggunakan rumus *Euclidian Distance* dengan persamaan pada rumus 2

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Dimana :

- x_i = obyek x ke- i
- y_i = obyek y ke- i
- n = banyaknya obyek

4. Mengelompokan obyek ke dalam masing – masing kluster berdasarkan jarak terdekat terhadap titik pusat (*centroid*)
5. Menentukan dan menghitung *centroid* baru menggunakan hasil nilai rata – rata dari setiap anggota pada masing – masing kluster dapat menggunakan rumus 3

$$C_k = \frac{1}{n_k} \sum d_i \quad (3)$$

Dimana :

- n_k : Jumlah data dalam kluster k
- d_i : Jumlah dari nilai jarak dalam masing – masing kluster

Kembali ulangi mulai dari langkah ke-3, apabila perhitungan rata – rata tiap anggota belum stabil atau belum memiliki nilai yang sama persis dengan rata – rata pada iterasi sebelumnya.

3 METODE PENELITIAN

3.1 Fase Pemahaman Bisnis

Penelitian ini bertujuan untuk dapat menerapkan algoritma K-Means yang dikombinasikan dengan metode *elbow* menggunakan data kecelakaan lalu lintas yang dalam penerapannya dapat diperoleh jumlah *cluster* optimal dalam pengelompokan data kecelakaan lalu lintas untuk mendapatkan profil dari tiap *cluster*.

3.2 Fase Pemahaman Data

Sumber data yang digunakan adalah data kecelakaan lalu lintas yang telah terjadi di Kota Semarang dari bulan Oktober 2018 sampai dengan Juli 2019 yang dirubah menjadi format horizontal.

3.3 Fase Persiapan Data

1. Pemilihan Data

Tahap akan dilakukannya seleksi atribut yang akan digunakan pada fase selanjutnya. Terdapat perubahan terhadap penyusunan atribut dataset. Karena pada bagian seluruh atribut korban yaitu atribut kendaraan yang terlibat, usia, jenis kelamin, dan pekerjaan masih memiliki dua sampai dengan 3 korban dalam sekali kejadian. Dengan hasil seleksi seperti pada Tabel 1.

Tabel 1. Hasil Seleksi Atribut

N	Atribut	N	Atribut
1	Hari Kejadian	5	Kendaraan yang Terlibat (korban)
2	Waktu Kejadian	6	Usia (korban)
3	Jenis Kecelakaan	7	Jenis Kelamin (korban)
4	Kendaraan yang Terlibat (pelaku)	8	Pekerjaan (korban)

2. Pembersihan Data

Akhir proses pembersihan semua *record* data lengkap, diperoleh 913 *record* data laka lantas dari jumlah awal yaitu 1572 *record* data.

3. Transformasi Data

Dalam transformasi data terdapat 2 tahap utama, yaitu proses penyederhanaan *record* dan proses konversi *record*. Pada proses penyederhanaan *record* terdapat beberapa atribut yang terpilih, diantaranya

a. Atribut Jenis Kecelakaan

Dari banyaknya jenis dari kecelakaan, ditransformasi menjadi tunggal dan tidak tunggal.

b. Atribut Kendaraan yang Terlibat

Isi atribut ini sangat bermacam, karena diikuti dengan merk kendaraan, maka disederhanakan menjadi hanya garis besarnya saja, seperti SPM, KBM.

c. Atribut Pekerjaan

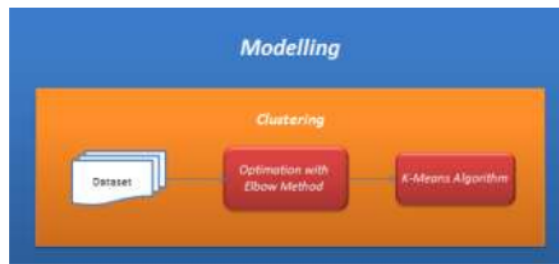
Pada pekerjaan, terdapat pengelompokan kembali untuk kategori PNS dan Swasta.

Selanjutnya adalah proses konversi *record* yang telah melalui proses penyederhanaan, yang hasilnya dapat dilihat pada Tabel 2.

Tabel 2. Potongan dataset akhir

hari	waktu	jenis kecelakaan	kendaraan yang terlibat	usia (korban)	jenis kelamin (korban)	pekerjaan (korban)
1	2	2	3	7	3	6
1	4	2	3	7	3	6
3	4	2	3	5	1	2
2	4	2	3	7	3	6
3	3	1	1	2	1	3
7	2	2	3	3	1	2
5	2	2	4	3	1	2
5	2	1	1	3	1	4
5	2	2	4	7	3	6

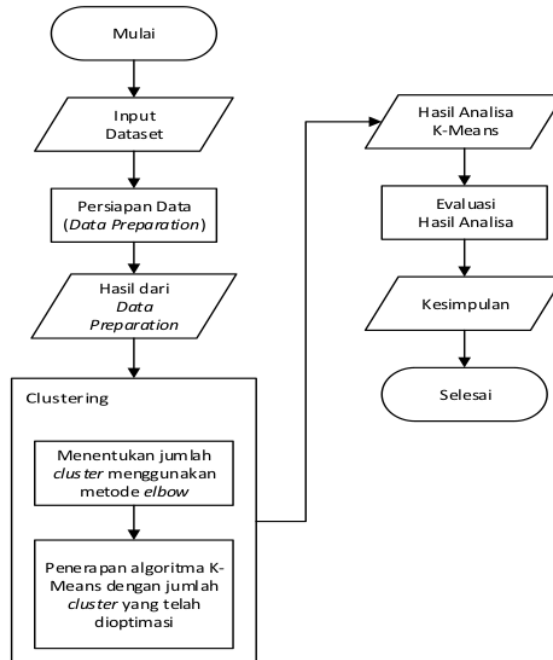
3.4 Fase Pemodelan



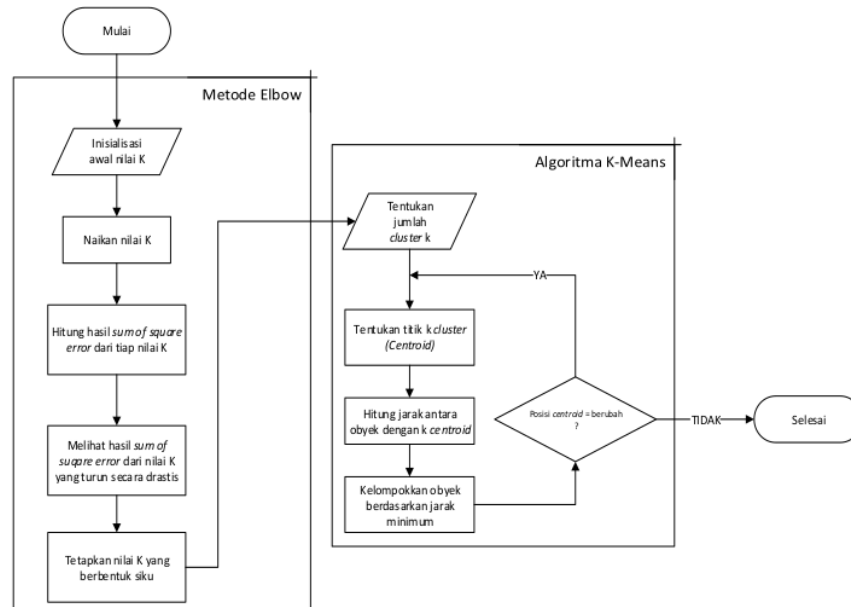
Gambar 2. Model Penelitian

Teknik data mining dengan metode *clustering* menggunakan algoritma *K-Means* dikombinasikan dengan metode *Elbow* sebagai metode optimasi jumlah kluster. Menggunakan *tools* RStudio, yang akhirnya akan menghasilkan hasil pengelompokan atau *clustering* dengan jumlah kluster yang telah dioptimasi dalam mengelompokkan data kecelakaan lalu lintas yang diharapkan mendapatkan profil dari tiap kluster.

Flowchart proses pemodelan (skema pada Gambar 2) dengan algoritma yang digunakan dapat dilihat pada Gambar 3 yang dalam prosesnya terdapat penerapan metode *Elbow* dan algoritma *K-Means* pada Gambar 4.



Gambar 3. *Flowchart* penelitian



Gambar 4. Flowchart metode Elbow dan algoritma K-Means

3.5 Fase Evaluasi

Pada fase ini akan didapatkan hasil analisa *clustering* data menggunakan algoritma *K-Means* dengan metode *Elbow* untuk optimasi kluster. Hasil analisa kemudian dapat dijabarkan yang selanjutnya dapat diambil kesimpulan dengan memperlihatkan profil dari tiap kluster.

3.6 Fase Penyebaran

Pada tahap ini hasil analisa yang telah ditata dan disimpulkan dalam bentuk laporan dapat diberikan atau direkomendasikan ke pihak urmin Satlantas Polres Semarang Barat sebagai bahan acuan dalam pengelompokan profil data kecelakaan lalu lintas.

4 HASIL DAN PEMBAHASAN

4.1 Perhitungan Manual Algoritma Data Mining

4.1.1 Algoritma K-Means

Banyak iterasi perhitungan manual yang dilakukan peneliti dari 2 kluster sampai dengan 10 kluster dapat dilihat pada Tabel 3.

Tabel 3. Jumlah iterasi seluruh kluster

Kluster	Banyak Iterasi
1 Kluster	1 iterasi
2 Kluster	5 iterasi
3 Kluster	11 iterasi
4 Kluster	9 iterasi
5 Kluster	13 iterasi
6 Kluster	12 iterasi
7 Kluster	12 iterasi
8 Kluster	13 iterasi
9 Kluster	14 iterasi
10 Kluster	13 iterasi

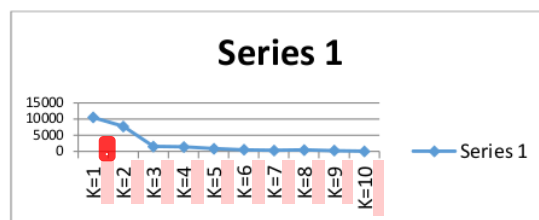
4.1.2 SSE (Sum of Square Error)

Menghitung selisih dari masing – masing nilai SSE kluster dengan hasil dapat dilihat pada Tabel 4.

Tabel 4. Hasil selisih SSE

Kluster	Nilai SSE	Selisih
K=1	10476,21516	0
K=2	9309,938937	1166,276219
K=3	7723,987667	1585,95127
K=4	6325,53373	1398,453937
K=5	5442,941461	882,592269
K=6	4923,029412	519,912049
K=7	4591,289984	331,739428
K=8	4116,789919	474,500065
K=9	3805,89493	310,894989
K=10	3751,107829	54,787101

Jelas bahwa pada K=3 mengalami penurunan yang paling besar yang kemudian akan membentuk lekukan siku yang dapat dilihat pada Gambar 6.



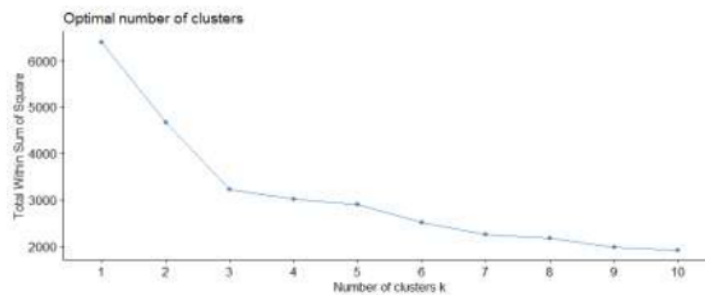
Gambar 6. Grafik selisih SSE

4.2 Implementasi Algoritma K-Means pada RStudio

4.2.1 Mencari K Optimal Kluster

```
fviz_nbclust(datafix, kmeans, method = "wss")
```

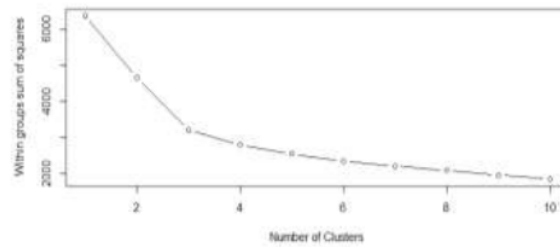
Source code diatas berfungsi untuk mencari optimal kluster menggunakan dataset 'datafix' dan proses K-Means dengan metode *Elbow* yang menggunakan nilai total wss (*within sum square*). Dengan hasil yang dapat dilihat pada Gambar 7.



Gambar 7. Grafik SSE pada aplikasi Rstudio

```
wssplot <- function(data, nc=10, seed=1234){
  wss <- (nrow(data)-1)*sum(apply(data,2,var))
  for (i in 2:nc){
    set.seed(seed)
    wss[i] <- sum(kmeans(data, centers=i)$withinss)}
  plot(1:nc, wss, type="b", xlab="Number of Clusters",
       ylab="Within groups sum of squares")
  wss
}
```


Source code diatas menunjukkan bagaimana fungsi dari perhitungan wws / metode Elbow yang masuk kedalam variabel 'wssplot', selanjutnya dapat dipanggil dengan diikuti dataset yang ingin diproses.



Gambar 8. Grafik Hasil Perhitungan WWS

4.2.2 Eksekusi K-Means

```
final <- kmeans(datafix, 3, nstart = 25)
```

Source code yang berfungsi untuk memproses dataset 'datafix' dengan parameter 3 kluster dan random set / nstart sebanyak 25 yang artinya R akan melakukan proses klusterisasi sebanyak 25 kali secara acak dan akan menghasilkan kluster dengan variasi terendah. Seluruh hasil perhitungan akan dimasukkan kedalam variabel baru yang bernama 'final'.

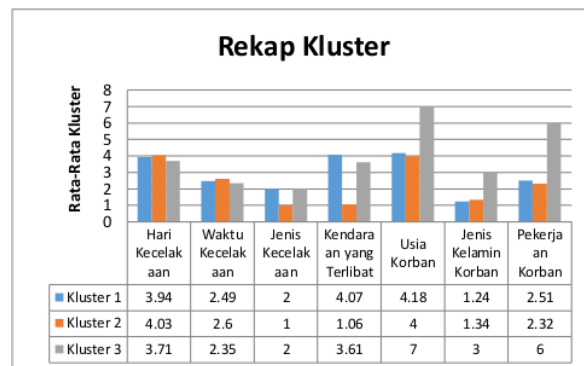
```
lakalantas %>%
  mutate(Cluster = final$cluster) %>%
  group_by(Cluster) %>%
  summarise_all("mean")
```

Source code diatas berfungsi untuk menampilkan profilisasi kluster yang diringkas berdasarkan rata – rata dari masing – masing atribut per kluster. Dengan hasil yang dapat dilihat pada Gambar 8.

```
> lakalantas %>%
+   mutate(Cluster = final$cluster) %>%
+   group_by(Cluster) %>%
+   summarise_all("mean")
# A tibble: 3 x 8
  Cluster hari waktu `jenis kecelakaan` `kendaraan yang terlibat` usia `jenis kelamin` pekerjaan
  <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1     1  3.94  2.49  2.00  4.07  4.18  1.24  2.51
2     2  4.03  2.60  1.06  1.06  4.00  1.34  2.32
3     3  3.71  2.35  2.00  3.61  7.00  3.00  6.00
```

Gambar 9. Profilisasi kluster berdasarkan rata-rata

4.3 Analisa Rekap Kluster



Gambar 10. Diagram dan tabel rekap kluster

Dengan acuan pada gambar 10, maka dapat dijabarkan makna angka yang dicocokkan dengan konversi data yang menghasilkan pernyataan-pernyataan sebagai berikut :

1. Kluster 1

Laka lintas rata - rata ke hari Kamis pada pukul 06:15 WIB sampai 12:00 WIB dengan jenis kecelakaan tidak tunggal antara SPM dengan KBM dengan korban rata – rata di usia 31 sampai 40 tahun berjenis kelamin laki – laki dengan pekerjaan sebagai pegawai swasta.

2. Kluster 2

Laka lintas rata – rata terjadi pada hari Kamis, rawan pada pukul 06:15 WIB sampai 12:00 WIB. terjadi dengan jenis kecelakaan tunggal, rata - rata SPM (Tunggal), berada di usia 31 sampai 40 tahun berjenis kelamin laki – laki, dengan pekerjaan pegawai swasta.

3. Kluster 3

Laka lintas rata – rata terjadi pada hari Kamis, rawan pada pukul 06:15 WIB sampai 12:00 WIB dengan jenis kecelakaan tidak tunggal, antara SPM dengan SPM, korban rata - rata tidak diketahui, jenis kelaminnya tidak diketahui dan pekerjaan yang tidak diketahui.

5. KESIMPULAN

Berdasarkan uraian di atas maka dapat diambil kesimpulan sebagai berikut :

1. Dalam penerapan algoritma *K-Means* dengan hasil optimasi menggunakan metode *Elbow*, menghasilkan 3 segmentasi kecelakaan.
2. Rata – rata Laka Lintas terjadi pada hari Kamis pada pukul 06:00 sampai 12:00 WIB. Kemudian rata – rata jenis kecelakaan yang terjadi adalah kecelakaan tidak tunggal. Jenis kendaraan yang terlibat antara sepeda motor dengan sepeda motor (SPM – SPM). Melibatkan korban berusia 31 sampai 40 tahun, yang rata – rata berjenis kelamin laki – laki dengan memiliki profesi sebagai pegawai swasta.

DAFTAR PUSTAKA

- [1] Dinas Kependudukan Dan Pencatatan Sipil Semarang. (2018), Jumlah Penduduk Kota Semarang Periode Desember 2018 [Internet]. Semarang : Dispendukcapil. Tersedia dalam: <<http://dispendukcapil.semarangkota.go.id/statistik/jumlah-penduduk-kota-semarang/2018-12-02>> [diakses 8 Agustus 2019 pukul 17.17 WIB]
- [2] Retnosari, P. and Jananto, A. (2013), *Implementasi Data Mining Untuk Menemukan Hubungan Antara Kota Kelahiran Mahasiswa Dengan Tingkat Kelulusan Mahasiswa Pada Fakultas Teknologi Informasi Unisbank*, Jurnal Dinamika Informatika, vol. 5(2), pp. 112–121.
- [3] Han, J., Kamber, M., & Pei, J. (2012), *Data Mining : Concepts and Techniques (Third Edition)*, Waltham, MA: Morgan Kaufmann Publishers.
- [4] Prasetyo, Eko. (2012). *DATA MINING - Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta : ANDI.
- [5] Shearer, Colin. (2000). *The CRISP-DM Model: The New Blueprint for Data Mining*. JOURNAL of Data Warehousing, vol. 5(4), p. 13-22.
- [6] Bahri, S. (2019). *Optimasi Cluster K-Means dengan Modifikasi Metode Elbow untuk Menganalisis Disrupsi Pendidikan Tinggi*, Tesis, Program Studi Teknik Informatika FASILKOM USU, Sumatera Utara
- [7] Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). *Integration k-means clustering method and elbow method for identification of the best customer profile cluster*. IOP Conference Series: Materials Science and Engineering, vol. 336(1), p. 012017
- [8] Alatubir, D. A. V. (2017). *Penerapan Algoritma K-Means Untuk Pengelompokan Sekolah Menengah Atas Di Provinsi Daerah Istimewa Yogyakarta Berdasarkan Nilai Daya Serap Ujian Nasional Bahasa Indonesia*, Skripsi, Program Studi Teknik Informatika FST Universitas Sanata Dharma, Yogyakarta.
- [9] Bholowalia, P., & Kumar, A. (2014). *EBK-means: A clustering technique based on elbow method and k-means in WSN*. International Journal of Computer Applications, vol. 105(9), p. 17-24.
- [10] Merliana, N. P. E., & Santoso, A. J. (2015). *Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-means Clustering*, Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers Unisbank (SENDI-U). ISBN: 978-979-3649-81-8

KLUSTERISASI OPTIMAL DENGAN ELBOW METHOD UNTUK PENGELOMPOKAN DATA KECELAKAAN LALU LINTAS DI KOTA SEMARANG

ORIGINALITY REPORT

13%

SIMILARITY INDEX

13%

INTERNET SOURCES

7%

PUBLICATIONS

6%

STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

2%

★ semilkom.apps.cs.ipb.ac.id

Internet Source

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On