

## **Data Mining Menggunakan Metode K-Means Klaster untuk Mengelompokkan Pemegang Polis Asuransi Kendaraan Bermotor di Indonesia**

<sup>1</sup>Supiyah, <sup>2</sup>Aceng Komarudin Mutaqin, <sup>3</sup>Teti Sofia Yanti

<sup>1,2,3</sup>Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung,  
Jl. Tamansari No. 1 Bandung 40116

e-mail : <sup>1</sup>supiyah93@yahoo.com, <sup>2</sup>aceng.k.mutaqin@gmail.com, <sup>3</sup>buitet@yahoo.com

**Abstrak.** *Data mining* adalah suatu proses menemukan pola yang menarik dari data berukuran besar. Salah satu teknik dalam *data mining* adalah metode *K-means* klaster. Metode ini digunakan untuk mengelompokkan suatu kumpulan data ke dalam *k* buah klaster sedemikian sehingga titik-titik data dalam suatu klaster lebih mirip satu sama lain daripada titik-titik data dalam klaster yang berbeda. Dalam skripsi ini metode *K-means* klaster akan digunakan untuk mengelompokkan pemegang polis asuransi kendaraan bermotor di Indonesia berdasarkan variabel-variabel kode pertanggungan, kode kendaraan, kode penggunaan, kode wilayah, usia kendaraan, dan harga pertanggungan. Dengan ditetapkannya jumlah klaster sebanyak 5, maka klaster 1 mempunyai anggota sebanyak 826 pemegang polis, klaster 2 mempunyai anggota 136 pemegang polis, klaster 3 mempunyai anggota 522 pemegang polis, klaster 4 mempunyai anggota 442 pemegang polis, klaster 5 mempunyai anggota 2.151 pemegang polis. Klaster 1 sampai dengan klaster 5 didominasi oleh: pemegang polis yang mengambil produk asuransi dengan pertanggungan *comprehensive*, pemegang polis yang menggunakan merk Toyota, pemegang polis yang menggunakan kendaraan pribadi, dan pemegang polis yang kendaraannya berada di wilayah Jabodetabek.

**Kata Kunci:** *Data Mining*, *K-Means Klaster*, *Jarak Euclidean*, *Asuransi Kendaraan Bermotor*.

### **A. Pendahuluan**

*Data mining* (penambangan data) atau lebih tepat disebut sebagai *knowledge mining from data* (penambangan pengetahuan dari data) adalah proses menemukan pola yang menarik dari data berukuran besar (Han dkk., 2012). *Data mining* biasanya melibatkan pembersihan data, pengintegrasian data, pemilihan data, transformasi data, penemuan pola, evaluasi pola dan presentasi pengetahuan atau informasi (Han dkk., 2012). Pembersihan data dilakukan untuk menghilangkan gangguan dan data yang tidak konsisten. Pengintegrasian data dilakukan untuk menggabungkan data dari berbagai sumber. Pemilihan data dilakukan untuk memilih data yang relevan untuk dianalisis. Transformasi data ke dalam bentuk yang tepat dilakukan dalam *data mining* misalnya melalui operasi agregat. Penemuan dan evaluasi pola dilakukan untuk mengidentifikasi pola-pola yang menarik. Presentasi pengetahuan atau informasi dilakukan untuk menjelaskan visualisasi dan teknik-teknik yang digunakan dalam *data mining* kepada pengguna

*Data mining* dapat dilakukan pada setiap jenis data seperti data *database*, data *warehouse*, data transaksional, data deret waktu, data spasial dan spasiotemporal, data teks, data multimedia, data grafik, data jaringan, dan data web (Han dkk., 2012). *Data mining* telah sukses diaplikasi di berbagai bidang seperti dalam bidang intelijen bisnis, pencarian web, bioinformatika, informatika kesehatan, keuangan, asuransi dan perpustakaan digital (Han dkk., 2012).

Salah satu teknik *data mining* yang dapat digunakan adalah analisis klaster (Guo, 2001). Salah satu metode analisis klaster adalah metode *K-means* klaster. Metode *K-means* klaster membutuhkan jumlah klaster yang ditentukan terlebih dahulu oleh pengguna. Metode ini dapat digunakan untuk menempatkan data baru untuk dikelompokkan ke dalam klaster terdekat. Metode *K-means* klaster telah diterapkan

oleh Guo (2001) pada data pemegang polis asuransi kendaraan bermotor, dimana variabel-variabel yang dilibatkannya adalah usia pemegang polis, usia kendaraan, jenis kendaraan, jenis kelamin pemegang polis, tingkat cakupan asuransi, pendidikan, lokasi tempat tinggal, cuaca, skor kredit pemegang polis, dan nomor identifikasi pemegang polis. Dalam skripsi ini akan diaplikasikan *data mining* menggunakan metode *K-means* klaster pada data pemegang polis asuransi kendaraan bermotor di Indonesia berdasarkan variabel kode pertanggung, kode kendaraan, kode penggunaan, kode wilayah, usia kendaraan, dan harga pertanggung.

## B. Tinjauan Pustaka

### 1. Pengertian Asuransi

Asuransi adalah istilah yang digunakan untuk merujuk pada tindakan, sistem, atau bisnis dimana perlindungan finansial (atau ganti rugi secara finansial).

### 2. Asuransi Kendaraan Bermotor

Asuransi Kendaraan Bermotor adalah jenis asuransi khusus kendaraan bermotor, dimana risiko yang mungkin terjadi pada kendaraan dialihkan kepada perusahaan asuransi.

**Tabel 2.1** Kategori Kendaraan Bermotor Menurut Kementerian Keuangan Republik Indonesia Pada Tahun 2011

Jenis kendaraan	Kategori	Uang Pertanggung
Jenis Kendaraan Non Bus dan Non Truk	Kategori 1	0 s.d Rp150.000.000
	Kategori 2	Rp150.000.001 s.d. Rp300.000.000
	Kategori 3	Rp300.000.001 s.d. Rp500.000.000
	Kategori 4	Rp500.000.001 s.d. Rp800.000.000
	Kategori 5	Lebih dari Rp800.000.000
Jenis Kendaraan Bus dan Truk	Kategori 6	Truk, semua uang pertanggung
	Kategori 7	Bus, semua uang pertanggung
Jenis Kendaraan Roda 2 (dua)	Kategori 8	Semua uang pertanggung

### 3. Proses Data Mining

*Data mining* menggabungkan teknik dari mesin pembelajaran, pengenalan pola, statistik, teori *database* dan visualisasi untuk mengekstrak konsep, keterkaitan konsep dan pola yang menarik secara otomatis dari *database* yang berukuran besar. Suatu proses *data mining* secara umum melibatkan 4 tahapan berikut ini.

- Tahap 1: Pemilihan jenis data yang akan digunakan.
- Tahap 2: Pengolahan data sebelum diproses seperti pembersihan data, transformasi data, penanganan data outlier dan data hilang.
- Tahap 3: Eksplorasi data dan pembentukan model.
- Tahap 4: Interpretasi dan evaluasi pola yang ditemukan.

#### 4. Teknik Data Mining

*Data mining* yaitu proses eksplorasi dan pembentukan sebuah model untuk mewakili kumpulan data. Berbagai teknik *data mining* tersedia untuk proses eksplorasi dan pembentukan sebuah model untuk mewakili kumpulan data, diantaranya adalah *neural network*, algoritma genetika, dan teknik kluster.

*Neural network* merupakan teknik yang paling banyak digunakan dalam *data mining*. Algoritma genetika adalah metode optimasi kombinatorik yang didasarkan pada proses evolusi biologi. Teknik kluster digunakan untuk mensegmentasi data ke dalam kelompok-kelompok, masing-masing kelompok memiliki sifat-sifat yang umum dan menarik.

#### 5. Data Mining dalam Industri Asuransi

*Data mining* dapat membantu perusahaan asuransi membuat keputusan bisnis penting dan menerapkan pengetahuan yang baru ditemukan ke dalam praktek bisnis seperti pengembangan produk, pemasaran, analisis distribusi klaim, manajemen pertanggungjawaban aset dan analisis solvabilitas.

#### 6. Pengklasteran (Clustering)

Pengklasteran (*clustering*) adalah salah satu pekerjaan yang paling berguna dalam proses *data mining* untuk menemukan kluster (kelompok) dan mengidentifikasi distribusi dan pola yang menarik dalam data. Metode pengklasteran melakukan analisis kluster berdasarkan jarak *Euclidean* dihitung dari satu atau lebih variabel kuantitatif dan pusat kluster awal yang dihasilkan dan diperbarui oleh algoritma. Salah satu algoritma dalam analisis kluster adalah *K-means* kluster. Asuransi isalkan bahwa objek atau individu atau data yang dikumpulkan ada sebanyak  $n$  titik pengamatan. Masing-masing objek diwakili oleh vektor pengamatan  $\mathbf{x}'_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ ,  $i = 1, 2, \dots, n$ , dari  $p$  buah variabel. Tujuan dari algoritma *K-means* kluster adalah untuk menentukan partisi alami dari data menjadi  $k$  buah kluster. Algoritma *K-means* kluster sebagai berikut:

1. Menentukan banyaknya kluster ( $k$ ).
2. Memilih  $k$  pusat kluster awal (*centroid*).
3. Menetapkan titik-titik data yang jaraknya paling dekat dengan pusat kluster awal  $j$  sebagai anggota kluster  $j$ ,  $j = 1, 2, \dots, k$ . Ukuran kedekatannya didasarkan pada ukuran jarak *Euclidean*. Misalkan pusat kluster  $j$  direpresentasikan oleh vektor  $\mathbf{c}'_j = (c_{j1}, c_{j2}, \dots, c_{jp})$ ,  $j = 1, 2, \dots, k$ , maka jarak *Euclidean* antara titik data  $i$  ke pusat kluster  $j$  adalah  $d_{ij} = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2 + \dots + (x_{ip} - c_{jp})^2}$ ,  $i = 1, 2, \dots, n$ , dan  $j = 1, 2, \dots, k$ .
4. Menghitung rata-rata dari titik-titik data dalam setiap kluster, kemudian memindahkan  $k$  pusat kluster awal ke rata-rata klasternya.
5. Menetapkan kembali titik-titik data yang jaraknya paling dekat dengan pusat kluster  $j$  sebagai anggota kluster  $j$ ,  $j = 1, 2, \dots, k$ .
6. Menggunakan rata-rata dari titik-titik data dalam setiap kluster sebagai pusat kluster baru.

Prosedur ini diulang sampai tidak ada perubahan lebih lanjut dalam pengklasteran.

## C. Bahan dan Metode

### 1. Bahan

Asuransi ntuk mengaplikasikan *data mining* menggunakan *K-means* klaster, akan digunakan data sekunder pemegang polis asuransi kendaraan bermotor di Indonesia hasil pencatatan yang diperoleh dari Kementerian Keuangan Republik Indonesia pada tahun 2009-2010.. Variabel-variabel yang akan digunakan untuk mengaplikasikan *data mining* menggunakan *K-means* klaster disajikan dalam Tabel 3.1.

**Tabel 3.1** Variabel untuk Profil Risiko Pemegang Polis

Nama Varibel	Skala Pengukuran
Kode Pertanggungan	Nominal
Kode Kendaraan	Nominal
Kode Penggunaan	Nominal
Kode Wilayah	Nominal
Usia Kendaraan	Interval
Harga Pertanggungan	Rasio

Sumber : Kementerian Keuangan Republik Indonesia 2011

Kode pertanggungan adalah kode yang digunakan untuk mengidentifikasi jenis pertanggungan. Kode kendaraan adalah kode merk kendaraan yang digunakan Kode penggunaan adalah kode standar yang digunakan untuk mengidentifikasi kendaraan yang digunakan untuk pribadi, kantor atau umum, kode untuk setiap jenis penggunaan. Kode wilayah adalah kode atas dasar alamat yang tercantum dalam Surat Tanda Nomor Kendaraan (STNK). Usia kendaraan (dalam tahun) dihitung sejak kendaraan dikeluarkan oleh pabrikan sampai kendaraan diasuransikan. Diasumsikan bahwa kendaraan dikeluarkan di awal tahun, dan 1 tahun dihitung 365 hari. Sedangkan harga pertanggungan adalah harga kendaraan saat baru atau taksiran harga kendaraan apabila dibeli pada saat pertanggungan dimulai dengan kondisi yang sama. Tabel 3.2 menyajikan contoh data yang memuat variabel-variabel untuk profil risiko pemegang polis.

**Tabel 3.2** Data Pemegang Polis Asuransi Kendaraan Bermotor di Indonesia

No Pemegang Polis	Kode Pertanggungan	Kode Kendaraan	Kode Penggunaan	Kode Wilayah	Usia Kendaraan (Dalam Tahun)	Harga Pertanggungan (Rp)
1	1	13	1	1	9,69	10000000
2	1	13	1	1	8,16	13100000
3	1	13	1	1	7,78	11000000
4	1	13	1	1	7,64	21000000
...	...	...	...	...	...	...
4077	1	3	2	11	7,87	19000000

Sumber: Kementerian Keuangan Republik Indonesia 2011

## 2. Metode

Akan diaplikasikan *data mining* menggunakan *K-means* kluster untuk mengelompokkan pemegang polis asuransi kendaraan bermotor di Indonesia. Tahapan dalam menggunakan metode *K-means* kluster adalah sebagai berikut:

1. Data dalam variabel kode pertanggung, kode kendaraan, kode penggunaan dan kode wilayah perlu dikuantifikasikan karena datanya bersifat kualitatif. Proses inisialisasi (Ramadhani, 2013) akan digunakan untuk mengkuantifikasikannya. Langkah-langkah proses inisialisasinya adalah sebagai berikut:
  - a. Buat tabel distribusi frekuensi untuk setiap variabel (kode pertanggung, kode kendaraan, kode penggunaan dan kode wilayah).
  - b. Untuk variabel tertentu, kategori (kode) yang memiliki frekuensi tertinggi pertama diberi nilai kuantitatif 1. Kategori (kode) yang memiliki frekuensi tertinggi kedua diberi nilai kuantitatif 2. Begitu seterusnya.
2. Menghitung nilai *Z-score* untuk setiap pengamatan dalam variabel kode pertanggung, kode kendaraan, kode kendaraan, kode wilayah, usia kendaraan, dan harga pertanggung. Hal ini dilakukan karena beberapa variabel satuannya tidak sama. Langkah-langkah menghitung nilai *Z-score* adalah sebagai berikut:
  - a. Menghitung nilai rata-rata ( $\bar{x}_j, j = 1, 2, \dots, 6$ ) dan simpangan baku ( $S_j, j = 1, 2, \dots, 6$ ) untuk setiap variabel.
  - b. Menghitung nilai *Z-score* untuk setiap pengamatan dalam setiap variabel dengan menggunakan persamaan berikut:
  - c.  $Z_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}; \dots$  (3.1) untuk  $i = 1, 2, \dots, 6; j = 1, 2, \dots, 4.077$ . Nilai *Z-score* inilah yang akan diolah menggunakan metode *K-means* kluster.
3. Tentukan jumlah kluster ( $k$ ).
4. Tentukan pusat kluster (*centroid*). Pusat kluster (*centroid*) adalah  $k$  buah pengamatan pertama. Misalkan  $k$  buah *centroid* tersebut direpresentasikan oleh direpresentasikan oleh vektor  $c'_j = (c_{j1}, c_{j2}, \dots, c_{jp}), j = 1, 2, \dots, k$ . Pada paket program SPSS, *centroid* telah ditentukan secara acak dan secara otomatis sehingga tidak perlu menentukan *centroid* sendiri.
5. Hitung jarak *Euclidean* antara tiap titik data dengan setiap *centroid* dengan menggunakan rumus:
 
$$d_{ij} = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2 + \dots + (x_{ip} - c_{jp})^2}, i = 1, 2, \dots, n, \text{ dan } j = 1, 2, \dots, k.$$
6. Masukkan tiap titik data ke dalam suatu kluster berdasarkan jarak *Euclidean* terdekat dengan pusat kluster (*centroid*) yang berpadanan.
7. Bentuk kluster awal yang berisikan titik-titik data yang di dalamnya belum tetap menjadi anggota kluster tersebut karena mungkin masih mengalami pemindahan titik data antar kluster.
8. Hitung kembali pusat kluster (*centroid*) yang baru terbentuk di tiap-tiap kluster dengan merata-ratakan nilai tiap variabel yang masuk menjadi anggota kluster awal.
9. Ulangi perhitungan jarak antara setiap titik data dengan *centroid* yang baru dengan menggunakan jarak *Euclidean*.
10. Masukkan tiap titik data pada suatu kluster berdasarkan jarak terdekat

dengan pusat klaster (*centroid*) yang baru.

11. Bentuk klaster baru yang berisikan titik-titik data yang di dalamnya belum tetap menjadi anggota klaster tersebut karena mungkin masih mengalami pemindahan titik data antar klaster.
12. Lakukan pengecekan apakah klaster yang baru terbentuk sudah tidak ada lagi pemindahan titik data antar klaster. Jika ternyata masih ada pemindahan titik data antar klaster maka kembali ke langkah 6 sampai tidak ada lagi pemindahan titik data antar klaster.
13. Lakukan interpretasi dan profilisasi klaster. Interpretasi dilakukan untuk mencari karakteristik tiap klaster yang khas.

#### D. Hasil dan Pembahasan

##### 1. Transformasi Data

Asuransi atau yang digunakan untuk proses *data mining* menggunakan metode *K-means* klaster adalah data profil risiko pemegang polis asuransi kendaraan bermotor di perusahaan asuransi umum XYZ untuk kendaraan jenis non bus dan non truk untuk kategori 1 s/d kategori 5. Dalam data tersebut terdapat variabel-variabel yang skala pengukurannya nominal (kualitatif). Variabel-variabel tersebut adalah kode pertanggungan, kode kendaraan, kode penggunaan, kode wilayah, sedangkan usia kendaraan, dan harga pertanggungan mempunyai skala pengukuran rasio. Nilai pengamatan dari variabel-variabel tersebut perlu dikuantifikasikan melalui proses inialisasi (Ramadhani, 2013) agar bisa digunakan metode *K-means* klaster.

##### 2. Menghitung Z-score

Enam variabel yang akan dianalisis menggunakan *K-means* klaster memiliki satuan yang berbeda, oleh karena itu perlu distandarisasi dengan menghitung nilai *Z-score* untuk setiap pengamatan dari variabel-variabel tersebut. Tabel 4.1 menyajikan rata-rata dan simpangan baku tersebut.

**Tabel 4.1** Hasil Perhitungan Rata-rata dan Simpangan Baku

Variabel	Rata-rata	Simpangan Baku
Kode Pertanggungan	1,1094	0,3122
Kode Kendaraan	3,4677	3,4294
Kode Penggunaan	1,1781	0,4727
Kode Wilayah	2,0839	2,2201
Usia Kendaraan	4,7549	3,7027
Harga Pertanggungan	214.299.438	223.177.745

Nilai *Z-score* untuk setiap pengamatan dalam variabel kode pertanggungan dihitung sebagai berikut. Untuk pengamatan pertama nilai *Z-score* nya adalah:

$$Z_{11} = \frac{x_{11} - \bar{x}_{.1}}{S_{.1}} = \frac{1 - 1,1094}{0,3122} = -0,3504.$$

**Tabel 4.2** Nilai *Z-score*

No Pemegang Polis	Kode Pertanggung	Kode Kendaraan	Kode Penggunaan	Kode Wilayah	Usia Kendaraan (Dalam Tahun)	Harga Pertanggung (Rp)
1	-0,3504	2,7796	-0,3767	-0,4882	1,3328	-0,5122
2	-0,3504	2,7796	-0,3767	-0,4882	0,9195	-0,3732
3	-0,3504	2,7796	-0,3767	-0,4882	0,8169	-0,4673
4	-0,3504	2,7796	-0,3767	-0,4882	0,7791	-0,0193
5	-0,3504	2,7796	-0,3767	-0,4882	-0,1581	0,0641
...	...	...	...	...	...	...
4077	-0,3504	-0,1364	1,7388	4,0162	0,8412	-0,1089

Sumber: Hasil Perhitungan *Software* SPSS

### 3. Pembentukan Kluster

Dalam analisis *K-means* kluster, langkah pertama yang harus dilakukan adalah menentukan jumlah kluster yang akan digunakan. Dalam skripsi ini jumlah kluster yang akan digunakan adalah 5. Nilai ini didasarkan pada banyaknya kategori jenis kendaraan. Langkah selanjutnya adalah menentukan pusat masing-masing kluster (*centroid*) awal. Dengan bantuan *Software* SPSS 21 pusat kluster awal disajikan dalam Tabel 4.3.

**Tabel 4.3** Pusat Kluster Awal

	Kluster				
	1	2	3	4	5
Z-score (Kode Pertanggung)	-0,3504	-0,3504	-0,3504	2,8529	2,8529
Z-score (Kode Kendaraan)	5,9871	1,3216	1,3216	-0,7196	0,1552
Z-score (Kode Penggunaan)	-0,3767	-0,3767	-0,3767	-0,3767	3,8544
Z-score (Kode Wilayah)	-0,0378	-0,4882	7,6197	0,4127	-0,0378
Z-score (Usia Kendaraan)	1,1086	-0,9251	-0,9251	6,6208	-0,5064
Z-score (Harga Pertanggung)	-0,7362	13,2437	0,5588	-0,8706	-0,6018

Sumber: Hasil Perhitungan *Software* SPSS

Langkah selanjutnya adalah menghitung jarak *Euclidean* untuk setiap titik data ke pusat kluster yang ada di Tabel 4.7. Untuk titik data pertama atau pemegang polis 1, jarak *Euclidean* nya terhadap pusat kluster 1 sampai dengan pusat kluster 5 masing-masing adalah

$$d_{1,1} = \sqrt{(-0,3504 - (-0,3504))^2 + (2,7796 - 5,9871)^2 + (-0,3767 - (-0,3767))^2 + (-0,4882 - (-0,3504))^2 + (1,3328 - 1,1086)^2 + (-0,5122 - (-0,7362))^2}$$

$$d_{1,1} = 3,2545$$

Terlihat bahwa titik data pertama atau pemegang polis 1 lebih dekat ke pusat kluster 1 dibandingkan ke pusat kluster lainnya. Dengan demikian titik data pertama atau pemegang polis 1 masuk menjadi anggota kluster 1.

Tabel 4.4 menyajikan contoh jarak *Euclidean* pemegang polis dengan pusat kluster 1 sampai kluster 5 dan keanggotaan klasternya.

**Tabel 4.4** Jarak dan Keanggotaan Klaster

Pemegang Polis	Jarak dengan Pusat Klaster					Anggota Klaster
	1	2	3	4	5	
1	3,2545	14,0160	8,6086	7,1701	6,2165	1
2	3,2648	13,8185	8,4932	7,4880	6,1107	1
3	3,2633	13,8980	8,4824	7,5608	6,0847	1
4	3,3338	13,4513	8,4322	7,6268	6,1030	1
5	3,5423	13,4094	8,2970	8,3619	5,9720	1
6	3,7852	12,6290	8,2713	8,5082	6,0906	1
7	3,8329	12,4763	8,2800	8,5022	6,1290	1
8	3,9766	12,0828	8,3210	8,4766	6,2502	1
9	4,2266	11,6418	8,3684	8,6209	6,3961	1
10	6,4113	8,7561	9,2227	10,008	7,9268	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
4073	7,9381	14,4611	4,3161	8,2777	6,7544	3
4074	7,9441	14,4342	4,3050	8,3036	6,7538	3
4075	8,2754	13,3828	4,1511	9,2185	6,9032	3
4076	8,2298	13,6267	4,1435	9,1721	6,8543	3
4077	7,6729	14,4327	4,8117	7,8770	5,7720	3

Langkah selanjutnya adalah menghitung rata-rata setiap variabel *Z-score* untuk setiap klaster yang telah terbentuk di atas. Hasil perhitungan rata-ratanya disajikan dalam Tabel 4.5.

**Tabel 4.5** Rata-rata Variabel *Z-score* untuk Setiap Klaster

Klaster	Kode Pertanggung	Kode Kendaraan	Kode Penggunaan	Kode Wilayah	Usia Kendaraan (Dalam Tahun)	Harga Pertanggung (Rp)
1	-0,2929	1,7418	-0,3629	-0,1607	0,0855	0,1399
2	0,1662	0,8795	-0,3767	-0,4737	-0,5133	5,1357
3	-0,3125	-0,0984	-0,2015	3,1899	-0,0490	-0,1546
4	0,9462	-0,0913	-0,3641	0,1258	2,4573	-0,6126
5	0,0219	-0,3427	0,1062	-0,1444	-0,1424	-0,0374

Rata-rata setiap variabel *Z-score* untuk setiap klaster di atas menjadi pusat klaster baru. Setelah diperoleh pusat klaster baru, langkah selanjutnya adalah menghitung kembali jarak *Euclidean* untuk setiap titik data ke pusat klaster baru. terlihat bahwa masih ada pemindahan anggota klaster, sehingga perlu dihitung pusat klaster baru untuk klaster 1 sampai klaster 5 dan penempatan pemegang polis ke klaster-klaster tersebut. Penghitungan pusat klaster baru dan penempatan pemegang polis ke klaster-klaster dilakukan terus menerus sampai tidak ada lagi pemindahan anggota klaster. Hasil akhir banyaknya anggota klaster, klaster terdekat dan jarak dengan klaster terdekat dalam analisis *K-means* klaster disajikan dalam Tabel 4.6.

**Tabel 4.6** Banyaknya Anggota Kluster

Kluster	Banyaknya Anggota	Kluster Terdekat	Jarak dengan Kluster Terdekat
1	826 pemegang polis	5	1,954
2	136 pemegang polis	5	4,232
3	522 pemegang polis	5	2,641
4	442 pemegang polis	1	3,749
5	2.151 pemegang polis	1	1,954

Berdasarkan Tabel 4.13 terlihat bahwa kluster yang jumlah anggotanya paling banyak adalah kluster 5. Sedangkan kluster yang jumlah anggotanya paling sedikit adalah kluster 2. Sementara itu, kluster 1, kluster 2, dan kluster 3 lebih dekat ke kluster 5 dibandingkan dengan kluster yang lain. Sedangkan kluster 4 dan kluster 5 lebih dekat ke kluster 1 dibandingkan dengan kluster yang lain.

#### E. Kesimpulan

Berdasarkan hasil penerapan *data mining* menggunakan metode *K-means* kluster pada data polis asuransi kendaraan bermotor di Indonesia berdasarkan variabel kode pertanggungan, kode kendaraan, kode penggunaan, kode wilayah, usia kendaraan dan harga pertanggungan dapat disimpulkan bahwa:

1. Dengan ditetapkan jumlah kluster sebanyak 5, maka kluster 1 mempunyai anggota sebanyak 826 pemegang polis, kluster 2 mempunyai anggota 136 pemegang polis, kluster 3 mempunyai anggota 522 pemegang polis, kluster 4 mempunyai anggota 442 pemegang polis, kluster 5 mempunyai anggota 2.151 pemegang polis.
2. Kluster 1, kluster 2, dan kluster 3 lebih dekat ke kluster 5 dibandingkan dengan kluster yang lain. Sedangkan kluster 4 dan kluster 5 lebih dekat ke kluster 1 dibandingkan dengan kluster yang lain.
3. Kluster 1 sampai dengan kluster 5 didominasi oleh: pemegang polis yang mengambil produk asuransi dengan pertanggungan *comprehensive*, pemegang polis yang menggunakan merk Toyota, pemegang polis yang menggunakan kendaraan pribadi, dan pemegang polis yang kendaraannya berada di wilayah Jabodetabek.
4. Kode kendaraan dalam kluster 1 dan kluster 5 lebih beragam dibandingkan dengan kode kendaraan dalam kluster 2, 3, dan 4. Kode penggunaan dalam kluster 1, 3, 4 dan 5 lebih beragam dibandingkan dengan kode penggunaan dalam kluster 2. Kode wilayah dalam kluster 1 lebih beragam dibandingkan dengan kode wilayah dalam kluster 2, 3, 4 dan 5.
5. Rata-rata usia kendaraan untuk kluster 1, 2, 4 dan 5 lebih tua dibandingkan dengan rata-rata usia seluruh kendaraan dan rata-rata usia kendaraan dalam kluster 3. Rata-rata harga pertanggungan untuk kluster 1 dan 5 lebih tinggi dibandingkan dengan rata-rata harga pertanggungan seluruh kendaraan dan rata-rata harga pertanggungan untuk kluster 2, 3 dan 4.

#### Daftar Pustaka

- Guha, S., Rastogi, R., Shim, K. (1998). "CURE: An Efficient Clustering Algorithm for Large Databases." Proceedings of the ACM SIGMOD Conference.
- Guo, L. (2001). *Applying Data Mining Techniques in Property/Casualty Insurance*.

University of Central Florida.

Han, J., Kamber, M., Pei, J. (2012). *Data Mining: Concepts and Techniques*. Elsevier, USA.

Kementerian Keuangan Republik Indonesia. (2011). Perubahan Atas Peraturan Menteri Keuangan Nomor 74/pmk.010J2007 Tentang Penyelenggaraan Pertanggungjawaban Asuransi Pada Lini Usaha Asuransi Kendaraan Bermotor. Kementerian Keuangan Republik Indonesia Badan Pengawas Modal dan Lembaga Keuangan.

Putri, R.N.. (2012). Pemodelan Regresi Hurdle Untuk Data Asuransi Kendaraan Bermotor di Indonesia. *Skripsi Program Studi Statistika, Universitas Islam Bandung*, Bandung.

Ramadhani. (2013). Data Mining Menggunakan Algoritma K-means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro. *Jurnal Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro*.

