

# PENERAPAN K-MEANS DALAM PENGELOMPOKAN DATA POLIS ASURANSI KENDARAAN BERMOTOR DI PT JASARAHARJA PUTERA

Winda Putri Utami<sup>1</sup>, Koko Handoko<sup>2</sup>

<sup>1</sup>Program Studi Teknik Informatika, Universitas Putera Batam

<sup>2</sup>Program Studi Teknik Informatika, Universitas Putera Batam

email: pb160210074@upbatam.ac.id

## ABSTRACT

*Motor Vehicle Insurance Policy is one of the insurance products at PT Jasaraharja Putera, being an insurance company certainly has a lot of data that causes that will only take place and make a large pile of records for that can be used to obtain information as product development and can be used as a target to get company income. The concept of the method used in this study is the K-Means Clustering method, this method is the most common and simple, which has the ability to group large enough data to be more efficient which aims to create information or add and find information from the data held. In this study the data is used to group Motor Vehicle Insurance Policy data from 2017 to 2019 apply the RapidMiner application. The test results obtained in the last iteration for the first cluster are 15 data, the second cluster is 6 data, the third cluster is 1 data, and the fourth cluster is 3 data. Based on the results of research on sample data, it can be seen that public interest in insurance is needed to develop insurance product marketing in order to increase the company's tagit and increase company revenue.*

*Keywords: Data Mining, Insurance Policy, K-means Clustering, Motor Vehicle,*

## PENDAHULUAN

Kebutuhan jasa asuransi saat ini makin dirasakan bagi penggunanya. Asuransi sendiri merupakan suatu pengendalian resiko dimana satu pihak mengalihkan resiko yang mungkin akan terjadi di masa mendatang kepada pihak lain, yang dalam hal ini adalah perusahaan asuransi. Asuransi menjadi penting perannya karena disetiap aspek perlu adanya perlindungan untuk mengurangi resiko yang dihadapi. Asuransi berasal dari Bahasa Inggris yaitu "insurance" yang artinya pertanggungan, sehingga dapat dikatakan asuransi adalah suatu perjanjian antara pihak tertanggung (nasabah) dengan pihak penanggung (perusahaan asuransi) dimana

penanggung bersedia untuk menggantikan kerugian yang mungkin dialami oleh nasabah dimasa mendatang.

Agar mendapat jaminan asuransi atas resiko yang mungkin terjadi, pihak tertanggung (nasabah) harus membayar premi kepada perusahaan asuransi dalam jangka waktu tertentu. Setelah itu baru pihak asuransi memiliki kewajiban dan memberikan jaminan sepenuhnya kepada nasabah. Asuransi kendaraan bermotor adalah pertanggungan terhadap kerugian atau kerusakan dan menggantikan ganti rugi pada kendaraan bermotor atau kepentingan yang dipertanggungkan. Pada prinsipnya, jaminan yang di tanggung terhadap kerusakan kendaraan itu sendiri dan

tanggung jawab hukum terhadap pihak lain yang dirugikan pada saat menggunakan kendaraan tersebut.

*Data Mining* adalah proses ekstraksi suatu data yang akan menjadi informasi atau pengetahuan atau pola dari data yang jumlahnya besar. Yang juga bertujuan untuk menemukan atau menambah pengetahuan dari data atau informasi yang kita miliki. Metode *k-Means clustering* adalah metode yang paling sederhana dan umum,. Mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan efisien (Alfina, Santosa, & Barakbah, 2012). Metode ini sendiri diterapkan pada data pemegang polis asuransi kendaraan bermotor, dimana variable-variabelnya yang dilibatkan adalah harga kendaraan, rate kendaraan, harga pertanggungan atau premi, tahun kendaraan.

PT Jasaraharja Putera merupakan perusahaan asuransi yang bergerak dalam bidang penjualan asuransi kerugian. Perusahaan mengakui pendapatan premi adalah salah satu target agar pendapatan perusahaan lebih meningkat. Maka penulis melakukan pengujian dan melakukan pengelompokan dengan data variabel yang dikhususkan untuk data polis kendaraan bermotor. Tujuan pelaksanaan penelitian ini untuk mendapatkan hasil pengelompokan sebagai landasan dasar untuk mengambil keputusan dan dimanfaatkan sebagai acuan mengembangkan produk dan pemasaran, agar perusahaan dapat mencapai target pendapatan, dan memenuhi pangsa pasar.

## KAJIAN TEORI

### 2.1 KDD (*Knowledge Discovery in Database*)

Konsep *Data Mining* sering disebut sebagai *Knowledge Discovery in Database* (KDD) bertugas untuk mengekstrak dan mengenali pola atau model yang menarik dari data yang terdapat pada basis data untuk mencari pengetahuan yang terdapat dalam basis data yang cukup besar (Tan, Vlandari, 2004). Berhubungan dengan teknik

integrasi dan penemuan ilmiah, interpretasi dan

visualisasi dari pola-pola sejumlah data. Serangkaian proses data tersebut memiliki tahap sebagai berikut :

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*).
2. Integrasi data (penggabungan data dari beberapa sumber).
3. Transformasi data ( data diubah menjadi bentuk yang sesuai untuk di-*mining*).
4. Aplikasi teknik *Data Mining* (proses ekstraksi pola dari data yang ada).
5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan).
6. Presentasi pengetahuan (dengan teknik isualisasi)

### 2.2 *Data Mining*

*Data Mining* adalah proses ekstraksi suatu data (sebelumnya tidak diketahui, bersifat implisit, dan tidak dianggap tidak berguna) menjadi informasi atau pengetahuan atau pola dari data yang jumlahnya besar (Witten, 2011).

*Data Mining* adalah proses ekstraksi atau mengidentifikasi informasi dari database yang jumlahnya besar. *Data Mining* sebenarnya untuk membentuk informasi atau pola yang baru yang berguna.

### 2.3 *K-Means Clustering*

Metode *K-Means Clustering* adalah metode yang sederhana dan umum. Dikarenakan metode ini mempunyai kemampuan mengelompokkan data yang cukup besar dengan waktu yang efisien dan cukup cepat karena menggunakan komputer. Hasil *cluster* yang terbentuk dari metode ini tergantung pada inisiasi pusat awal *cluster* nilai yang diberikan (Handoko, 2016).

*K-Means* merupakan metode algoritma untuk menganalisa data dimana prosesnya menetapkan nilai-nilai yang akan di *cluster* secara *random*, dan nantinya akan dikelompokkan menjadi

beberapa kelompok dimana pada setiap kelompoknya mempunyai karakteristik yang mirip, atau sama dengan lainnya tapi dengan kelompok lainnya memiliki karakteristik yang berbeda (Novianto & Goeirmanto, 2019).

#### 2.4 RapidMiner

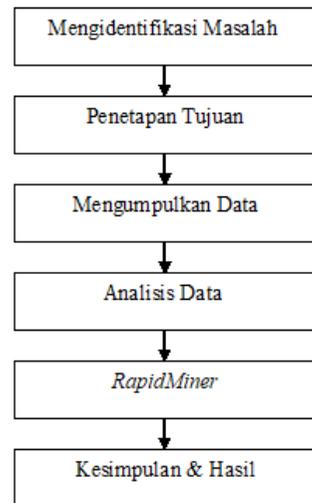
*RapidMiner* adalah suatu aplikasi yang dapat meningkatkan keakuratan analisis. *RapidMiner* juga merupakan perangkat lunak yang memudahkan penggunaannya untuk menggunakan perangkat lunak ini. Dalam *RapidMiner* ini juga tidak dibutuhkan kemampuan membuat koding yang khususnya untuk penggunaan *data mining*, karena fasilitasnya sudah tersedia. Metode yang disediakan oleh *RapidMiner* mulai dari *classification*, *clustering*, *association*, dll.

### METODE PENELITIAN

#### 3.1 Desain Penelitian

Algoritma *K-Means* masuk kedalam penerapan *data mining*. Tujuan dari *clustering* adalah untuk pengelompokan data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan, dimana data harus masuk dalam *cluster* tertentu dan memungkinkan setiap data untuk masuk dalam *cluster* tertentu melalui tahapan sebuah proses, yang tahapan berikutnya berpindah ke *cluster* yang lainnya. *K-Means* akan memisahkan data ke *K* daerah bagian terkenal dan mampu melakukan klasifikasi data yang besar dengan sangat cepat.

Berikut ini merupakan gambar yang digunakan untuk menggambarkan langkah-langkah dalam algoritma *K-Means* :



Gambar 1. Desain Penelitian  
(Sumber: Data Penelitian 2020)

Berikut penjelasan tahapan-tahapan pada desain penelitian diatas, yaitu:

1. Mengidentifikasi Masalah  
Menjelaskan permasalahan untuk menyelesaikan masalah yang ada pada penelitian, seperti pada penelitian ini peneliti lebih fokus meneliti tentang polis asuransi kendaraan bermotor.
2. Penetapan Tujuan  
Berdasarkan permasalahan yang ada, penetapan tujuan ini bertujuan untuk menentukan tujuan yang berisi hal apa saja yang ingin dicapai, dan untuk sebagai apa penelitian ini dilakukan.
3. Mengumpulkan Data  
Data yang dikumpulkan pada penelitian ini dilakukan dengan cara melakukan wawancara, observasi dan dokumentasi langsung di PT Jasaraharja Putera supaya mendapatkan data yang valid kebenarannya.
4. Analisis Data  
Analisis data yang dilakukan pada penelitian ini mengambil data asuransi kendaraan bermotor dari tahun 2017 sampai dengan 2019, data dilakukan analisis perhitungan manual menggunakan Excel.

5. *RapidMiner*  
Data perhitungan manual yang dihitung menggunakan Excel dan sudah dilakukan pemilihan data yang dijadikan sampel pada penelitian dilakukan perhitungan kembali menggunakan *RapidMiner* untuk mengevaluasi dan mengetahui keakuratan hasil pada hasil akhir.
6. Kesimpulan dan Hasil  
Tahapan penarikan kesimpulan dan hasil pada penelitian ini adalah proses akhir pada judul penelitian yang diangkat oleh peneliti.

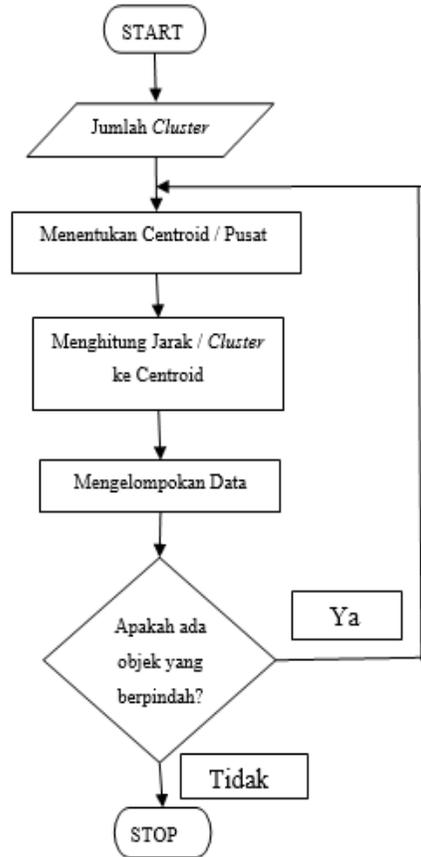
### 3.2 Metode Perancangan Sistem

Metode perancangan sistem adalah tahapan yang dilakukan dalam proses perancangan untuk mempermudah dalam mengembangkan ide, dan mempermudah menganalisis sebuah penelitian.

Proses perancangan pada penelitian ini sebagai berikut :

1. Melakukan pengumpulan data, pengumpulan data pada penelitian ini diambil dengan variabel polis asuransi kendaraan.
2. Setelah dilakukan pengumpulan data, menentukan mana saja data yang akan dilakukan pengclusteran nantinya.
3. Menentukan titik pusat secara acak sesuai kategori
4. Menghitung jarak masing masing variabel ke titik pusatnya sebagai data pengujian.
5. Melakukan pengelompokan data berdasarkan jarak terdekat antara korelasi beberapa objek sampai pengolahan data menghasilkan *output*.
6. Jika belum didapatkan data yang berpindah, maka pengujian dilakukan lagi dari mengacak titik pusatnya sesuai kategori.
7. Setelah *output* sudah didapat dari pengolahan data, maka penyelesaian yang dilakukan dari awal hingga adanya *output* berarti data tersebut sudah

memperoleh hasil dari pengolahan data.



Gambar 2. Desain Penelitian (Sumber: Data Penelitian 2020)

### 3.3 Teknik Pengumpulan Data

Pada penelitian ini Teknik pengumpulan data yang digunakan sebagai berikut :

1. Wawancara  
Teknik pengumpulan data untuk mendapatkan informasi langsung dari narasumbernya dilakukan di PT Jasaraharja Putera. Wawancara pada
2. Observasi  
Studi lapangan dilakukan peneliti langsung terhadap lokasi penelitian. Bertujuan untuk mendapatkan data asuransi kendaraan bermotor yang ada di PT Jasaraharja Putera dan melakukan pengamatan secara langsung.

3. Dokumentasi  
Teknik pengumpulan data menggunakan dokumentasi bertujuan untuk memperoleh data dari tempat peneliti berupa foto serta laporan kegiatan yang ada di tempat penelitian untuk dapat mempelajari masalah yang berhubungan pada objek yang akan diteliti.

sudah disetujui oleh pemegang polis.

#### 4.1 Analisis Data

Tahapan-tahapan yang akan dilakukan untuk menganalisa penelitian ini sebagai berikut:

1. Seleksi Data

Data yang dikumpulkan berasal dari data polis asuransi kendaraan bermotor dari tahun 2017 sampai dengan tahun 2019 di PT Jasaraharja Putera.

2. Pembersihan Data

Sampel yang akan diambil dan akan diteliti sebanyak 25 data. Sebelum data diolah akan ada tahap proses pembersihan data untuk menghilangkan variabel yang tidak digunakan pada penelitian ini. Data yang dihapus seperti data variabel nomor polis asuransi karena atribut ini hanya untuk melihat data pada polis asuransinya saja.

3. Transformasi Data

Untuk pengolahan data yang lebih mudah, maka dilakukan transformasi data dari data mentah menjadi sampel data yang ingin diambil.

4. *Data Mining*

Data pada penelitian ini akan diinput menggunakan sistem, sistem yang digunakan pada proses *data mining* ini dengan *RapidMiner*.

5. Evaluasi

Tahapan akhir ini sebagai informasi yang menentukan apakah tahapan yang dilakukan dan hasil yang didapatkan benar atau tidak, seperti pada penelitian ini dilihat dari perhitungan yang dilakukan secara manual sama atau tidak dengan perhitungan menggunakan aplikasi *RapidMiner*.

### HASIL DAN PEMBAHASAN

Pada penelitian ini sampel atau hasil data yang digunakan mengambil data polis asuransi kendaraan bermotor pada PT Jasaraharja Putera, sampel data yang diambil nantinya akan dianalisa menggunakan metode *k-means clustering* untuk mendapatkan nilai yang akurat dan relevan. Ada 4 atribut yang akan diambil pada penelitian ini antara lain harga kendaraan, premi kendaraan, tahun, dan rate kendaraan. Berikut variabel yang ada dalam penelitian ini meliputi :

1. Harga Kendaraan, sebuah angka untuk dijadikan sebagai tolak ukur yang nantinya dikalikan dengan rate asuransi.
2. Tahun Kendaraan, yang menunjukkan usia kendaraan yang akan diasuransikan juga untuk menentukan rate mana yang cocok untuk kendaraan yang akan diasuransikan nantinya.
3. Rate Kendaraan, rate yang sudah ditetapkan langsung oleh OJK (Otoritas Jasa Keuangan), rate asuransi ini berbeda satu sama lain hal ini dikarenakan dilihat berdasarkan tahun kendaraan, harga kendaraan, jenis kendaraan, yang nantinya akan mempengaruhi besarnya premi yang harus dibayarkan oleh pemegang polis.
4. Premi, tarif nominal pembayaran yang akan dibayarkan dan

**Tabel 1.** Data Rekapitulasi Polis Asuransi Kendaraan Bermotor

No.	Harga Kendaraan	Premi	Tahun	Rate
1	439,000,000	5,787,500	2015	1.2
2	126,692,000	3,546,022	2015	2.67
3	85,000,000	4,035,500	2010	4.58
4	205,000,000	5,573,500	2010	2.62
5	150,000,000	4,180,000	2016	2.67
6	230,000,000	5,831,000	2017	2.42
7	65,000,000	1,989,500	2011	2.78
8	205,000,000	5,573,500	2010	2.62
9	220,000,000	5,006,000	2018	2.18
10	100,000,000	4,350,000	2012	4.2
11	110,000,000	4,357,000	2014	3.82
12	130,000,000	3,805,000	2013	2.8
13	186,000,000	4,967,400	2013	2.54
14	60,000,000	1,632,000	2014	2.42
15	140,000,000	3,838,000	2014	2.67
16	65,000,000	3,363,000	2008	4.97
17	420,000,000	5,686,000	2019	1.2
18	90,000,000	2,373,000	2016	2.42
19	1,500,000,000	16,700,000	2019	1.05
20	130,000,000	2,208,000	2002	1.41
21	120,000,000	2,007,000	2003	1.36
22	120,000,000	2,007,000	2003	1.36
23	330,000,000	4,275,000	2013	1.2
24	45,000,000	1,536,000	2009	3.03
25	250,000,000	5,675,000	2014	2.18

(Sumber: Data Penelitian 2020)

#### 4.1 Analisa Proses Algoritma

Penelitian yang akan digunakan pada penelitian ini menggunakan algoritma *k-means clustering* untuk mengelompokan data yang tersedia. Proses yang digunakan sebagai berikut:

1. Menentukan Jumlah *Cluster*

Pada tahapan ini menentukan pengelompokan banyaknya jumlah pada *cluster* yang ingin dibuat dan akan diproses melalui pertimbangan yang diusulkan untuk menentukan banyak *cluster* yang ingin dibentuk. Pada penelitian ini jumlah *cluster*

yang ditentukan pada pengelompokan data Polis Asuransi Kendaraan Bermotor.

*Cluster* yang akan dibentuk untuk digunakan pada penelitian ini sebanyak 4 kelompok  $k=4$ , dan dimana digunakan sebanyak 4 atribut.

2. Menentukan Titik Pusat *Cluster*

Untuk melakukan *centroid* awal dapat dilakukan secara acak dari data yang telah tersedia sebanyak jumlah *cluster*, jumlah *centroid* awal sebanyak 4 *centroid*. Nilai untuk C1 diambil

dari baris data ke-20, nilai C2 diambil dari baris data ke-9, nilai C3 diambil dari baris data ke-1, nilai C4 diambil dari baris data ke-25. Berikut nilai *centroid* awal pada penelitian :

C1 = (130.000.000, 2.208.000, 2002, 1.41)

C2 = (220.000.000, 5.006.000, 2018, 2.18)

C3 = (439.000.000, 5.787.500, 2015, 1.20)

C4 = (250.000.000, 5.675.000, 2014, 2.18)

3. Menghitung Jarak Setiap Data  
Jarak yang ada disetiap data akan dihitung menggunakan rumus *Euclidean Distance*. Setelah data dilakukan perhitungan kemudian didapatkan perbandingan hasil akhir.
4. Menghitung data masing-masing *centroid* ke yang paling dekat dan dilihat antara jarak dengan *centroid* pada setiap clusternya.
5. Menentukan *centroid* yang baru dengan cara melihat rata-rata pada *centroid* yang sama untuk melakukan iterasi selanjutnya. Berikut hasil *cluster* pada iterasi pertama  
Cluster 1 = 2, 3, 5, 7, 10, 11, 12, 14, 15, 16, 18, 20, 21,22, 24  
Cluster 2 = 4, 6, 8, 9, 13  
Cluster 3 = 1, 17, 19  
Cluster 4 = 23, 25

6. Setelah hasil *cluster* pada iterasi telah diketahui, maka berikut adalah data *centroid* baru untuk iterasi pertama  
C1 = (102.446.133,33 , 3.015.134,83 , 1.876,40 , 2,88)  
C2 = (209.200.000 , 5.390.280 , 2.013,60 , 2,48)  
C3 = (786.333.333,33 , 9.391.166,67 , 2.017,67 , 1,15)  
C4 = (290.000.000 , 4.975.000 , 2.013,50 , 1,69)

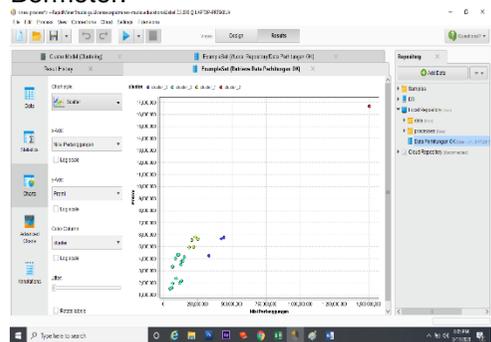
7. Menghitung ulang iterasi hingga nilai pada *centroid* tidak berubah. Jika nilai iterasi berikutnya sama dengan yang dihasilkan pada iterasi sebelumnya maka proses iterasi

selesai, lain halnya apabila iterasi berikutnya nilainya masih berbeda maka iterasi belum selesai dan ulangi iterasinya.

8. Iterasi selesai dilakukan pada iterasi ketiga dengan hasil iterasi yaitu :  
Cluster 1 = 2, 3, 5, 7, 10, 11, 12, 14, 15, 16, 18, 20, 21,22, 24  
Cluster 2 = 4, 6, 8, 9, 13, 25  
Cluster 3 = 19  
Cluster 4 = 1, 17, 23
9. Dan berikut hasil data *centroid* baru untuk iterasi terakhir pada pusat cluster yaitu :  
C1 = (102.446.133,33 , 3.015.134,83, 2.010,67 , 2,88)  
C2 = (216.000.000 , 5.437.733,33 , 2.013,67 , 337,73)  
C3 = (1.500.000.000 , 16.700.000 , 2.019 , 1,05)  
C4 = (396.333.33 , 5.249.500 , 2.015,67 , 1,20)

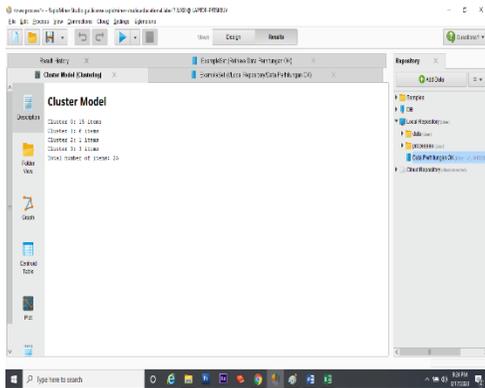
#### 4.2 Proses Pengujian

Pada penelitian ini pengujian yang dilakukan menggunakan aplikasi *RapidMiner*, untuk meyakinkan apakah analisis perhitungan manual akurat atau tidak. Maka untuk mengecek kebenarannya dilakukan pengujian *k-means clustering* menggunakan *RapidMiner*. Berikut grafik penyebaran data Polis Asuransi Kendaraan Bermotor.



Gambar 1. Grafik Penyebaran Data (Sumber: Data Penelitian 2020)

Serta berikut merupakan hasil akhir yang diperoleh pada clustering yang dilakukan pada pengujian data pengelompokan Polis



**Gambar 2.** Hasil Akhir Penelitian  
(Sumber: Data Penelitian 2020)

Pada hasil pengelompokan dengan menggunakan metode *k-means clustering* terdapat 4 *cluster* dan membentuk perolehan pada *cluster* pertama 15 data, *cluster* kedua 6 data, *cluster* ketiga 1 data, dan *cluster* keempat 3 data, perhitungan yang dilakukan menggunakan aplikasi *RapidMiner* sudah sesuai dengan perhitungan secara manual.

## SIMPULAN

Data yang digunakan merupakan data Polis Asuransi Kendaraan Bermotor dari tahun 2017 sampai dengan tahun 2019. Data di dapat dari PT Jasaraharja Putera yang merupakan Perusahaan Asurasi, data dibagi dalam masing-masing variabel seperti data harga kendaraan, tahun kendaraan, premi kendaraan, rate kendaraan dan akan dilakukan proses clustering menggunakan metode *k-means clustering* dengan aplikasi *RapidMiner*. Untuk hasil clustering yang lebih akurat, perhitungan dilakukan dengan cara manual yang dilakukan di Excel menggunakan rumus  $d(x,y) = ||x-y|| = \sqrt{\sum_{i=1}^n \{x_i - y_i\}^2}$ ;  $i = 1,2,3,\dots,n$ . Sampel data yang digunakan sebanyak 25 sampel data yang akan dilakukan perhitungan manual dan menggunakan aplikasi *RapidMiner*. Hasil Clustering Pertama mendapatkan 15 data, hasil clustering kedua mendapatkan 6 data,

hasil *clustering* ketiga mendapatkan 1 data, dan clustering keempat mendapatkan 3 data. Perhitungan yang dilakukan secara manual dan menggunakan aplikasi *RapidMiner* dengan metode *k-means clustering* hasilnya sama dan sinkron berarti dapat dikatakan bahwa perhitungan yang dilakukan hasilnya akurat. Penerapan data mining untuk pengelompokan data Polis Asuransi Kendaraan Bermotor yang telah didapat dan hasil perhitungannya digunakan untuk mengelompokkan data-data agar pihak perusahaan bisa mengembangkan dan memperoleh informasi dari data yang sudah dikelompokkan untuk membantu pengembangan pemasaran terhadap hasil pengelompokan.

## DAFTAR PUSTAKA

- Alfina, T., Santosa, B., & Barakbah, R. (2012). Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data ( Studi kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS ), 1.
- Handoko, K. (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 ), 02(03), 31–40.
- Novianto, R., & Goeirmanto, L. (2019). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi, 6(1), 85–95.
- Tan, Vlandari, R. T. (2004). Dok baru 2019-07-26 22.49.31\_1.pdf. Yogyakarta: Penerbit Gava Media. Retrieved from www.gavamedia.net
- Witten, I. H. F. (2011). Dok baru 2019-07-26 23.06.55\_1.pdf. Jakarta: PT Elex Media Komputindo.

	<p><b>Biodata</b> <b>Winda Putri</b> <b>Utami,</b> <b>merupakan</b> <b>mahasiswa Prodi</b> <b>Teknik</b> <b>Informatika</b> <b>Universitas</b> <b>Putera Batam.</b></p>
	<p><b>Biodata</b> <b>Koko Handoko,</b> <b>merupakan</b> <b>Dosen Prodi</b> <b>Teknik</b> <b>Informatika</b> <b>Universitas</b> <b>Putera Batam.</b></p>