

Pemetaan Wilayah Potensial Terhadap Penjualan Sepeda Motor Honda Menggunakan *K-Means Clustering*

Zulrahmadi¹✉

¹Universitas Islam Indragiri

zulrahmadi@gmail.com

Abstract

Indragiri Hilir regency consists of land and water which are divided into 20 districts, 39 sub-districts and 197 villages. Looking at the geographical condition of Indragiri Hilir Regency, motorcycle sales companies need to know the areas that have potential for motorcycle sales. Grouping potential areas is important in increasing sales profit for the company. This study aims to help PT. Capella Dinamik Nusantara in making the decision to increase sales to be more significant, promotion and marketing techniques were more targeted towards Honda motorcycle sales in the mapped areas. The data used in this study are Honda motorcycle sales data from 2017 to 2019. Data processing in this study uses the K-Means Clustering method with 3 clusters, namely Cluster 0 (C0) Less Potential, Cluster 1 (C1) Enough Potential, Cluster 2 (C2) Has the potential to sell Honda motorcycles. The result of the grouping process with 2 iterations states that for Cluster 0 there are 5 regions, for Cluster 1 there are 3 regions, and for Cluster 2 there are 2 regions.

Keywords: Data Mining, K-Means, Clustering, Sales, Motorbikes.

Abstrak

Wilayah kabupaten Indragiri Hilir terdiri dari daratan dan perairan yang terbagi menjadi 20 Kecamatan, 39 kelurahan dan 197 desa. Melihat kondisi geografis Kabupaten Indragiri Hilir, perusahaan penjualan sepeda motor perlu mengetahui wilayah yang potensial terhadap penjualan sepeda motor. Pengelompokan wilayah potensial merupakan hal penting dalam peningkatan profit penjualan bagi perusahaan. Penelitian ini bertujuan untuk membantu PT. Capella Dinamik Nusantara dalam pengambilan keputusan untuk meningkatkan penjualan menjadi lebih signifikan, teknik promosi dan marketing lebih tepat sasaran terhadap penjualan sepeda motor honda di wilayah-wilayah yang telah dipetakan. Data yang digunakan dalam penelitian ini adalah data penjualan sepeda motor honda tahun 2017 sampai dengan 2019. Pengolahan data dalam penelitian ini menggunakan metode *K-Means Clustering* dengan jumlah 3 Cluster yaitu Cluster 0 (C0) Kurang Berpotensi, Cluster 1 (C1) Cukup Berpotensi, Cluster 2 (C2) Sangat Berpotensi terhadap penjualan sepeda motor honda. Hasil dari proses pengelompokan dengan 2 kali iterasi menyatakan bahwa untuk Cluster 0 berjumlah 5 wilayah, untuk Cluster 1 berjumlah 3 wilayah, dan untuk Cluster 2 berjumlah 2 wilayah.

Kata kunci: *Data Mining*, *K-Means*, *Clustering*, Penjualan, Sepeda Motor.

© 2020 INFEB

1. Pendahuluan

PT. Capella Dinamik Nusantara (CDN) adalah perusahaan yang bergerak di bidang outlet penjualan sepeda motor honda yang mendirikan cabang wilayah pemasaran di kabupaten Indragiri Hilir sejak tahun 2011 yang berpusat di kota tembilahan.

Semenjak berdirinya CDN di kota tembilahan, semua data penjualan yang ada tersimpan didalam database kantor cabang dan tersusun dengan baik. Dalam proses penjualannya selama ini, PT. Capella Dinamik Nusantara mengalami kendala dalam pengelolaan produk yaitu manajemen stok sepeda motor. Pemilihan data penjualan sepeda motor honda sebagai objek penelitian untuk dianalisis dikarenakan dalam database penjualan tersebut tersimpan potensi-potensi atau fakta yang telah didokumentasikan untuk digali dan diproses menjadi sebuah pengetahuan baru bagi perusahaan ketika mengambil keputusan tentang permasalahan

penjualan, baik dari manajemen ketersediaan, maupun manajemen promosi atau marketing. Pengelompokan data penjualan dibagi berdasarkan wilayah desa-desa yang menjadi sasaran penjualan sepeda motor. Desa tersebut dikelompokkan menjadi wilayah desa potensial, cukup potensial, serta kurang potensial terhadap penjualan sepeda motor honda dengan menggunakan metode *Clustering*, data-data penjualan diperlukan sebagai acuan untuk melakukan pengelompokan. Data yang sudah dikelompokkan mampu mempermudah perusahaan dalam mengetahui wilayah desa yang potensial, cukup potensial, dan kurang potensial terhadap penjualan sepeda motor honda.

Beberapa penelitian yang pernah dilakukan sebelumnya menggunakan *algoritma k-means clustering* diantaranya adalah Putra & Wadisman (2018) menggunakan *algoritma k-means clustering* dalam pemilihan pelanggan potensial dengan data dengan

pusat centroid terbesar yang termasuk ke dalam pelanggan yang paling potensial [1]. Penelitian Waworuntu & Amin (2018) algoritma *k-means clustering* dalam pemetaan calon penerima jamkesda dengan hasil penelitian diperoleh hasil pada C0 cluster tidak mampu sebanyak 334 anggota, C1 cluster mampu sebanyak 106 anggota [2]. Penelitian Silalahi (2018) algoritma *k-means* dalam *clustering* terhadap penjualan produk di PT Batamas Niaga Jaya diperoleh hasil pada C0 kelompok tidak laris dengan jumlah 657 produk, C1 kelompok yang laris dengan jumlah 70 produk [3]. Penelitian Rosmini dkk (2018) algoritma *k-means clustering* dalam pemetaan kelompok mahasiswa melalui data aktivitas kuliah diperoleh hasil pada cluster 1 kelulusan tepat waktu berjumlah 10 mahasiswa, cluster 2 kelulusan tidak tepat waktu berjumlah 10 mahasiswa [4].

Penelitian Parlina dkk (2018) algoritma *k-means clustering* dalam pengelompokan pegawai yang layak mengikuti assessment dengan hasil diperoleh C1 cluster lolos dengan pusat cluster 8;66;13, C2 cluster hampir lolos dengan pusat cluster 10;71;14, dan C3 cluster tidak lolos dengan pusat cluster 7;60;12 [5]. Penelitian Fatmawati & Windarto (2018) algoritma *k-means clustering* dalam pengelompokan daerah terjangkit demam berdarah dengue diperoleh hasil C1 cluster tingkatan tinggi dengan jumlah 4 provinsi, C2 cluster tingkatan sedang dengan jumlah 13 provinsi, dan C3 cluster tingkatan rendah dengan jumlah 17 provinsi [6]. Penelitian Nur dkk (2017) algoritma *k-means* dalam *clustering* jurusan pada Sekolah Menengah Kejuruan tahun ajaran 2014/2015 diperoleh hasil C1 cluster tidak lulus, C2 cluster rekayasa perangkat lunak, C3 cluster teknik komputer jaringan [7]. Penelitian Febianto & Palasara (2019) algoritma *k-means clustering* dalam analisa data informasi kemiskinan di Jawa Barat diperoleh hasil 5 cluster dengan cluster pertama yang menjadi prioritas dalam peningkatan kesejahteraan penduduknya [8].

Penelitian Aprianti & Permadi (2018) algoritma *k-means* dalam *clustering* data kecelakaan lalu lintas di jalan raya diperoleh hasil bahwa kecelakaan sering terjadi pada kendaraan roda 2, korbannya adalah pengemudi di hari kerja pada waktu pagi hari [9]. Penelitian Gustientiedina dkk (2019) algoritma *k-means* dalam *clustering* data obat-obatan pada RSUD Pekanbaru diperoleh hasil 3 cluster [10]. Penelitian Alfiandi dkk (2018) algoritma *k-means clustering* dalam pemetaan permukiman kumuh di kota Bengkulu diperoleh hasil cluster 1 kawasan kumuh berat, cluster 2 kawasan kumuh sedang, cluster 3 kawasan kumuh ringan, cluster 4 kawasan tidak kumuh [11]. Penelitian Triyansyah & Fitrihanah (2018) algoritma *k-means clustering* dalam menentukan strategi marketing diperoleh hasil 2 cluster untuk informasi reseller potensial, dan 3 cluster untuk informasi tipe sepatu yang laris. Isi dari pendahuluan adalah jawaban atas pertanyaan [12].

Penelitian Maulida (2018) algoritma *k-means* dalam pengelompokan kunjungan wisatawan ke objek wisata unggulan di Prov. DKI Jakarta dengan hasil C1 = Taman Impian Jaya Ancol, C2 = Taman Mini Indonesia Indah dan Kebon Binatang Ragunan, dan C3 = Monumen Nasional, Museum Nasional, Museum Satria Mandala, Museum Sejarah Jakarta dan Pelabuhan Sunda Kelapa [13]. Penelitian Widodo & Wahyuni (2017) algoritma *k-means clustering* untuk mengetahui bidang skripsi mahasiswa multimedia diperoleh cluster 1 bidang video dan animasi, cluster 2 bidang aplikasi pengembangan perangkat lunak [14]. Penelitian Praja dkk (2019) algoritma *k-means clustering* dalam mengelompokkan data penumpang dan kapal angkutan laut di Indonesia dengan hasil 5 cluster, cluster 1 sangat rendah, cluster 2 rendah, cluster 3 sedang, cluster 4 tinggi, dan cluster 5 sangat tinggi [15]. Penelitian Efendi dkk (2019) algoritma *k-means clustering* dalam pengelompokan dan pemetaan derajat kesehatan kota Bengkulu dengan hasil cluster tertinggi berjumlah 2 puskesmas, cluster menengah berjumlah 15 puskesmas, dan cluster terendah berjumlah 3 puskesmas [16].

2. Metodologi Penelitian

Objek penelitian yang dilakukan ini menggunakan data yang bersumber dari penjualan sepeda motor honda tahun 2017 sampai dengan 2019 pada PT. Capella Dinamik Nusantara Tembilahan.

Knowledge Discovery in Database (KDD) merupakan proses *non-trivial* yang berguna untuk mencari dan mengidentifikasi pola di dalam data, dimana pola yang ditemukan tersebut bersifat benar, agar dapat bermanfaat dan dapat dipahami. KDD juga dapat diartikan sebagai inti dari informasi, yang bersifat tidak dapat dikenal dari kumpulan data. Proses KDD menyertakan dari proses sistem *data mining*. *Data Mining* merupakan proses pengelompokan data yang diolah dari data dengan jumlah besar, pemanfaatan sebuah data menjadi suatu yang penting dalam segala bidang, baik dari bidang akademik, bisnis maupun edisi.

Clustering merupakan teknik pengelompokan *record* pada basis data berdasarkan kriteria tertentu. Hasil *clustering* diberikan kepada pengguna akhir untuk memberikan gambaran tentang apa yang terjadi pada *database*. Konsep dasar dari *clustering* adalah mengelompokkan sejumlah objek ke dalam cluster dimana cluster yang baik adalah cluster yang memiliki tingkat kesamaan yang tinggi antar objek di dalam suatu cluster dan tingkat ketidaksamaan yang tinggi dengan objek cluster yang lainnya. *Clustering* adalah teknik data mining yang digunakan untuk menganalisis data dalam memecahkan permasalahan pengelompokan data atau mempartisi dari dataset ke dalam subset. Teknik cluster mempunyai dua metode dalam pengelompokannya yaitu *hierarchical clustering* dan *non-hierarchical clustering*. *Non hierarchical clustering*

pada teknik ini dimulai dengan menentukan jumlah *cluster* yang diinginkan (dua *cluster*, tiga *cluster*, empat *cluster* atau lebih), setelah menentukan jumlah *cluster* yang diinginkan maka proses *cluster* dimulai tanpa mengikuti proses hirarki, metode ini juga disebut sebagai metode *k-means clustering* [17]. Metode *k-means clustering* merupakan salah satu metode data *clustering non hierarchical* yang berusaha mempartisi data-data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik berbeda, dikelompokkan ke dalam kelompok lainnya. Kemiripan anggota terhadap *cluster* diukur dengan kedekatan objek terhadap nilai mean pada *cluster* atau dapat disebut sebagai centroid *cluster* atau pusat massa [18].

Langkah-langkah dalam penerapan algoritma *k-means clustering* yaitu dengan menentukan jumlah *cluster* *k* yang dibuat, selanjutnya menentukan nilai dari pusat/centroid awal yang ditentukan secara acak atau penentuan sendiri, setelah mendapatkan nilai titik pusat *cluster* langkah selanjutnya adalah menghitung jarak setiap data ke masing-masing *centroid* dengan menggunakan rumus korelasi antar dua objek, seperti yang terdapat pada rumus *Euclidean Distance* berikut ini.

$$D_{fg} = \sqrt{(w_f - w_g)^2 + (x_f - x_g)^2 + \dots + (n_f - n_g)^2} \quad (1)$$

Dimana :

D_{fg} = Jarak antara f dan g

w_f = Koordinat w objek

w_g = Koordinat w pusat

x_f = Koordinat x objek

x_g = Koordinat x pusat

Setelah mendapatkan hasil dari perhitungan jarak setiap data, langkah selanjutnya adalah menggabungkan setiap data yang dilihat berdasarkan jarak terdekat antara data dan *cluster*. Langkah selanjutnya menentukan titik pusat *centroid* baru dengan menghitung nilai rata-rata berdasarkan data pada *cluster* yang serupa, untuk menentukan pusat *centroid* baru untuk menghitung jarak antara data dan *centroid* dapat menggunakan rumus berikut ini.

$$C_k = \frac{a_1 + a_2 + a_3 + \dots + a_n}{\sum a} \quad (2)$$

Dimana :

C_k = data *cluster* ke *k*

a_1 = nilai *record* data ke 1

a_2 = nilai *record* data ke 2

$\sum a$ = jumlah *record* data

Algoritma *k-means* merupakan algoritma yang dilakukan secara berulang-ulang hingga menghasilkan *cluster* sesuai kelompok pada data-data yang sama, dan yang berbeda di tempatkan pada kelompok yang berbeda.

Kelebihan yang terdapat pada algoritma *k-means clustering* adalah sebagai berikut [19]:

- Sangat mudah dalam pengimplementasian dan proses menjalankan
- Mudah untuk beradaptasi
- Sangat umum digunakan

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini adalah data total penjualan sepeda motor honda tahun 2017 sampai dengan tahun 2019 yang diperoleh dari PT. Capella Dinamik Nusantara Tembilahan. Adapun sampel data penjualan sepeda motor honda yang menjadi objek penelitian seperti pada Tabel 1 di bawah ini:

Tabel 1. Sampel Data Penjualan Sepeda Motor Honda

No	Nama Desa	Kecamatan	Kabupaten	Penjualan Tahun 2017 (unit)	Penjualan Tahun 2018 (unit)	Penjualan Tahun 2019 (unit)
1	Kuala Enok	Tanah Merah	Indragiri Hilir	4	10	6
2	Kuala Gaung	Gaung Anak Serka	Indragiri Hilir	7	8	12
3	Nusantara Jaya	Keritang	Indragiri Hilir	8	10	6
4	Teluk Pinang	Gaung Anak Serka	Indragiri Hilir	41	28	44
5	Tembilahan Hilir	Tembilahan	Indragiri Hilir	115	123	130
6	Lahang Hulu	Gaung	Indragiri Hilir	7	8	10
7	Tembilahan Barat	Tembilahan Hulu	Indragiri Hilir	50	76	41
8	Metro	Reteh	Indragiri Hilir	3	2	1
9	Pulau Palas	Tempuling	Indragiri Hilir	23	43	36
10	Tembilahan Kota	Tembilahan	Indragiri Hilir	103	123	135

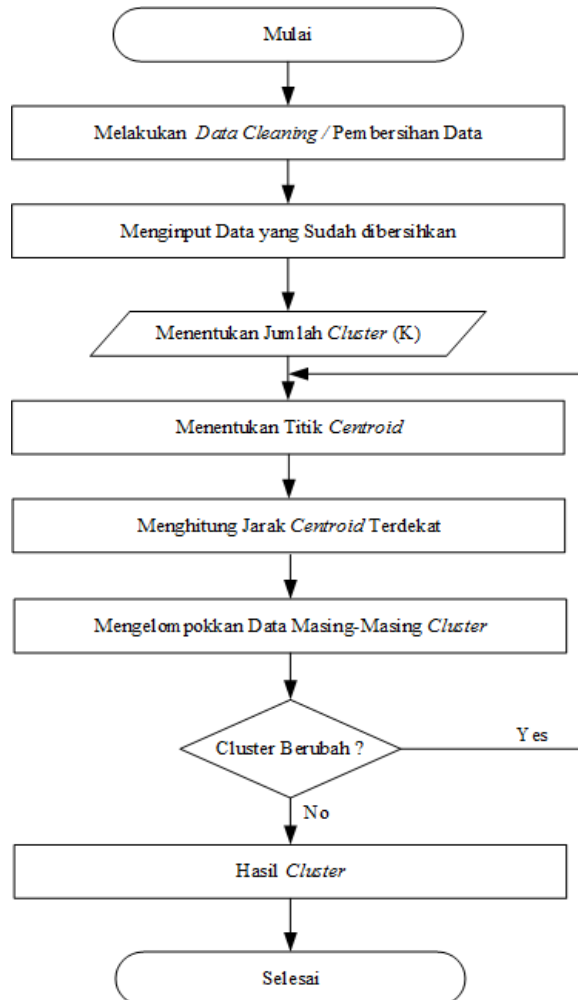
Penelitian ini menggunakan teknik analisa dengan metode *k-means clustering* untuk memetakan wilayah potensial terhadap penjualan sepeda motor honda berdasarkan total penjualan tahun 2017, tahun 2018, dan tahun 2019 dengan menjadi 3 *cluster*, yaitu *cluster* C0 untuk wilayah kurang berpotensi, *cluster* C1 untuk wilayah cukup berpotensi, dan *cluster* C2 untuk

wilayah sangat berpotensi. Berikut ini merupakan tahapan algoritma *k-means*, yaitu:

1. Melakukan *Data Cleaning* / pembersihan data
2. Menginput data yang sudah dibersihkan
3. Menentukan jumlah *cluster* (K)
4. Menentukan titik *centroid* secara random

5. Menghitung jarak *centroid* terdekat
 - a. Proses iterasi 1
 - b. Proses iterasi 2
 - c. Proses iterasi ... n
6. Mengelompokkan data masing-masing *cluster*
7. Mengecek perubahan data
8. Hasil *cluster*

Agar lebih jelas dan terarah, penerapan algoritma *k-means* pada penelitian ini digambarkan dalam bentuk *flowchart* seperti pada Gambar 1 berikut:



Gambar 1. *Flowchart* Proses *k-means*

3.1 Melakukan Pembersihan Data

Tahapan ini merupakan proses pengoreksian, pengurangan data sebelumnya atau membuang sebagian data yang ganda, menghilangkan data yang kotor, serta memverifikasi data yang dianggap tidak konsisten dan memverifikasi kesalahan yang ada pada data. Tujuannya yaitu untuk memudahkan dalam proses perhitungan algoritma *k-means*.

Hasil dari pembersihan data dapat dilihat pada Tabel 2 berikut ini:

Tabel 2. Sampel Data Cleaning Penjualan Sepeda Motor Honda

No	Nama Desa	Penjualan Tahun 2017 (unit)	Penjualan Tahun 2018 (unit)	Penjualan Tahun 2019 (unit)
1	Kuala Enok	4	10	6
2	Kuala Gaung	7	8	12
3	Nusantara Jaya	8	10	6
4	Teluk Pinang	41	28	44
5	Tembilahan Hilir	115	123	130
6	Lahang Hulu	7	8	10
7	Tembilahan Barat	50	76	41
8	Metro	3	2	1
9	Pulau Palas	23	43	36
10	Tembilahan Kota	103	123	135

3.2 Menginput Data yang Sudah dibersihkan

Proses penginputan data dilakukan ke dalam sistem setelah melakukan pembersihan data (*data cleaning*). Data yang digunakan dalam proses penginputan pada penelitian ini adalah data *training* yang sudah dilakukan pembersihan dengan jumlah sampel sebanyak 10 data.

3.3 Menentukan Jumlah Cluster K

Pada penelitian ini jumlah *cluster* ditentukan menjadi 3, dimana nantinya ke-3 *cluster* ini menentukan hasil penelitian mengenai wilayah yang masuk dalam kategori kurang berpotensi, cukup berpotensi, dan sangat berpotensi terhadap penjualan sepeda motor honda di Kabupaten Indragiri Hilir.

3.4 Menentukan Titik Centroid

Penentuan *centroid* awal pada penelitian ini dilakukan dengan cara acak atau random. Peneliti mengambil mengambil angka terendah, angka menengah, dan angka tertinggi mengenai total penjualan sepeda motor honda selama 3 tahun yang ada di Kabupaten Indragiri Hilir, dengan data *centroid* sebagai berikut:

Tabel 3. Sampel *Centroid* Acak

<i>Centroid</i>	Nama Desa	Penjualan Tahun 2017 (unit)	Penjualan Tahun 2018 (unit)	Penjualan Tahun 2019 (unit)
C0	Kuala Enok	4	10	6
C1	Teluk Pinang	41	28	44
C2	Tembilahan Kota	103	123	135

3.5 Menghitung Jarak Centroid Terdekat

Rumus yang digunakan untuk menghitung jarak setiap data kemasing-masing *centroid* adalah menggunakan rumus korelasi antara tiga objek (*Euclidean distance*), sehingga dengan rumus ini ditemukan jarak yang paling terdekat dari setiap data dengan masing-masing mendapatkan kecocokan data *centroid*. Proses perhitungannya seperti berikut ini:

1. Proses iterasi 1
 - a. Menghitung *cluster* 0 (C0)

Proses perhitungan diambil dari data desa nomor urut 1 dan nilai *centroid* C0, yaitu:

$$D_1 = \sqrt{(4-4)^2 + (10-10)^2 + (6-6)^2}$$

$$D_1 = 0$$

b. Menghitung cluster 1 (C1)

Proses perhitungan diambil dari data desa nomor urut 1 dan nilai centroid C1, yaitu:

$$D_1 = \sqrt{(4-41)^2 + (10-28)^2 + (6-44)^2}$$

$$D_1 = 56$$

c. Menghitung cluster 2 (C2)

Proses perhitungan diambil dari data desa nomor urut 1 dan nilai centroid C2, yaitu:

$$D_1 = \sqrt{(4-103)^2 + (10-123)^2 + (6-135)^2}$$

$$D_1 = 198$$

3.6 Mengelompokan data masing-masing cluster

Setelah proses perhitungan dilakukan pada tahapan menghitung jarak centroid terdekat sebelumnya, selanjutnya data dikelompokan sesuai dengan cluster masing-masing data, seperti yang terlihat pada Tabel 4 dibawah ini.

Tabel 4 Sampel Data Perhitungan Iterasi 1

No	Nama Desa	Penjualan Tahun 2017 (unit)	Penjualan Tahun 2018 (unit)	Penjualan Tahun 2019 (unit)	C0	C1	C2	Jarak Terdekat	C0	C1	C2
1	Kuala Enok	4	10	6	0	56	198	0	1	0	0
2	Kuala Gaung	7	8	12	7	51	194	7	1	0	0
3	Nusantara Jaya	8	10	6	4	53	196	4	1	0	0
4	Teluk Pinang	41	28	44	56	0	145	0	0	1	0
5	Tembilahan Hilir	115	123	130	201	148	13	13	0	0	1
6	Lahang Hulu	7	8	10	5	52	195	5	1	0	0
7	Tembilahan Barat	50	76	41	88	49	118	49	0	1	0
8	Metro	3	2	1	9	63	206	9	1	0	0
9	Pulau Palas	23	43	36	48	25	150	25	0	1	0
10	Tembilahan Kota	103	123	135	198	145	0	0	0	0	1
Jumlah									5	3	2

Tabel 3 diatas merupakan hasil klasterisasi masing-masing data sampel pada setiap desa dengan menggunakan rumus *euclidean distance*. Adapun hasil untuk cluster 0 berjumlah 5 desa, untuk cluster 1 berjumlah 3 desa, dan untuk cluster 2 berjumlah 2 desa.

3.7 Menentukan Titik Centroid

Tahapan selanjutnya adalah melakukan iterasi kedua dengan menghitung rata-rata hasil data sampel yang telah dikelompokan pada iterasi pertama untuk mencari centroid baru, dengan perhitungan sebagai berikut:

1. Cluster 0 (C0) berjumlah 5 data

$$C0(X1) = (4+7+8+7+3) / 5$$

$$C0(X1) = 6$$

$$C0(X2) = (10+8+10+8+2) / 5$$

$$C0(X2) = 8$$

$$C0(X3) = (6+12+6+10+1) / 5$$

$$C0(X3) = 7$$

2. Cluster 1 (C1) berjumlah 3 data

$$C1(X1) = (41+50+23) / 3$$

$$C1(X1) = 38$$

$$C1(X2) = (28+76+43) / 3$$

$$C1(X2) = 49$$

$$C1(X3) = (44+41+36) / 3$$

$$C1(X1) = 40$$

3. Cluster 2 (C2) berjumlah 2 data

$$C2(X1) = (115+103) / 2$$

$$C2(X1) = 109$$

$$C2(X2) = (123+123) / 2$$

$$C2(X2) = 123$$

$$C2(X3) = (130+135) / 2$$

$$C2(X3) = 133$$

Setelah proses perhitungan dilakukan, maka didapatkan data centroid baru yang akan digunakan untuk proses iterasi kedua. Data centroid baru dapat dilihat pada Tabel 5 berikut ini:

Tabel 5. Centroid Baru

Centroid	Penjualan Tahun 2017 (unit)	Penjualan Tahun 2018 (unit)	Penjualan Tahun 2019 (unit)
C0	6	8	7
C1	38	49	40
C2	109	123	133

3.8 Menghitung Jarak Centroid Terdekat

Rumus yang digunakan untuk menghitung jarak setiap data kemasing-masing centroid adalah menggunakan rumus korelasi antara tiga objek (Euclidean distance), sehingga dengan rumus ini ditemukan jarak yang paling terdekat dari setiap data dengan masing-masing mendapatkan kecocokan data centroid. Data yang

digunakan untuk *centroid* baru diperoleh dari hasil perhitungan sebelumnya. Proses perhitungan jarak masing-masing data ketitik pusat *cluster* seperti berikut ini:

2. Proses iterasi 2

a. Menghitung *cluster* 0 (C0)

Proses perhitungan diambil dari data desa nomor urut 1 dan nilai *centroid* baru C0, yaitu:

$$D_1 = \sqrt{(4-6)^2 + (10-8)^2 + (6-7)^2}$$

$$D_1 = 3$$

b. Menghitung *cluster* 1 (C1)

Proses perhitungan diambil dari data desa nomor urut 1 dan nilai *centroid* baru C1, yaitu:

$$D_1 = \sqrt{(4-38)^2 + (10-49)^2 + (6-40)^2}$$

$$D_1 = 62$$

c. Menghitung *cluster* 2 (C2)

Proses perhitungan diambil dari data desa nomor urut 1 dan nilai *centroid* baru C2, yaitu:

$$D_1 = \sqrt{(4-109)^2 + (10-123)^2 + (6-133)^2}$$

$$D_1 = 200$$

3.9 Mengelompokkan data masing-masing *cluster*

Setelah proses perhitungan dilakukan pada tahapan menghitung jarak *centroid* terdekat sebelumnya, menggunakan nilai *centroid* baru selanjutnya data dikelompokkan sesuai dengan *cluster* masing-masing, seperti yang dapat dilihat pada Tabel 6 di bawah ini.

Tabel 6 Sampel Data Perhitungan Iterasi 2

No	Nama Desa	Penjualan Tahun 2017 (unit)	Penjualan Tahun 2018 (unit)	Penjualan Tahun 2019 (unit)	C0	C1	C2	Jarak Terdekat	C0	C1	C2
1	Kuala Enok	4	10	6	3	62	200	3	1	0	0
2	Kuala Gaung	7	8	12	5	59	196	5	1	0	0
3	Nusantara Jaya	8	10	6	3	60	198	3	1	0	0
4	Teluk Pinang	41	28	44	55	22	147	22	0	1	0
5	Tembilahan Hilir	115	123	130	201	140	7	7	0	0	1
6	Lahang Hulu	7	8	10	3	60	197	3	1	0	0
7	Tembilahan Barat	50	76	41	88	30	119	30	0	1	0
8	Metro	3	2	1	9	70	208	9	1	0	0
9	Pulau Palas	23	43	36	49	17	152	17	0	1	0
10	Tembilahan Kota	103	123	135	198	137	6	6	0	0	1
Jumlah									5	3	2

Tabel diatas merupakan hasil dari pengeompokan pada iterasi kedua, dimana dapat disimpulkan bahwa data pada iterasi kedua tidak didapati perubahan untuk setiap *cluster* atau hasil iterasi kedua sama dengan hasil iterasi pertama. Dengan demikian proses iterasi hanya dilakukan sampai iterasi kedua dengan hasil *cluster* 0 berjumlah 5 desa, *cluster* 1 berjumlah 3 desa, dan *cluster* 2 berjumlah 2 desa. Hasil untuk setiap *cluster* dapat dilihat pada masing-masing tabel di bawah ini.

Tabel 7. *Cluster* 0 Kategori Kurang Berpotensi

No	Nama Desa	Penjualan Tahun 2017 (unit)	Penjualan Tahun 2018 (unit)	Penjualan Tahun 2019 (unit)
1	Kuala Enok	4	10	6
2	Kuala Gaung	7	8	12
3	Nusantara Jaya	8	10	6
4	Lahang Hulu	7	8	10
5	Metro	3	2	1
Jumlah		6	8	7

Tabel 7 diatas menunjukkan hasil *cluster* untuk kategori wilayah kurang berpotensi terhadap penjualan sepeda motor honda di Kabupaten Indragiri Hilir dengan jumlah 5 desa.

Tabel 8. *Cluster* 1 Kategori Cukup Berpotensi

No	Nama Desa	Penjualan Tahun 2017 (unit)	Penjualan Tahun 2018 (unit)	Penjualan Tahun 2019 (unit)
1	Teluk Pinang	41	28	44
2	Tembilahan Barat	50	76	41
3	Pulau Palas	23	43	36
Jumlah		38	49	40

Tabel 8 diatas menunjukkan hasil *cluster* untuk kategori wilayah cukup berpotensi terhadap penjualan sepeda motor honda di Kabupaten Indragiri Hilir dengan jumlah 3 desa.

Tabel 9. *Cluster* 2 Kategori Sangat Berpotensi

No	Nama Desa	Penjualan Tahun 2017 (unit)	Penjualan Tahun 2018 (unit)	Penjualan Tahun 2019 (unit)
1	Tembilahan Hilir	115	123	130
2	Tembilahan Kota	103	123	135
Jumlah		109	123	133

Tabel 9 diatas menunjukkan hasil *cluster* untuk kategori wilayah sangat berpotensi terhadap penjualan sepeda motor honda di Kabupaten Indragiri Hilir dengan jumlah 2 desa.

4. Kesimpulan

Algoritma *K-Means Clustering* dapat digunakan sebagai metode untuk mengelompokkan wilayah potensial terhadap penjualan sepeda motor honda. Dari 10 data penjualan sepeda motor yang diolah didapatkan hasil untuk kategori kurang berpotensi (C0) sebanyak 5 wilayah, kategori cukup berpotensi (C1) sebanyak 3 wilayah, dan kategori sangat berpotensi (C2) sebanyak 2 wilayah.

Daftar Rujukan

- [1] Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science*, 1(1), 72–77. DOI: <https://doi.org/10.31539/intecom.v1i1.141> .
- [2] Waworuntu, M. N. V., & Faisal Amin, M. (2018). Penerapan Metode K-Means untuk Pemetaan Calon Penerima Jamkesda. *KLIK-Kumpulan Jurnal Ilmu Komputer*, 5(2), 190. DOI: <https://doi.org/10.20527/klik.v5i2.157> .
- [3] Silalahi, M. (2018). Analisis Clustering Menggunakan Algoritma K-Means Terhadap Penjualan Produk Pada PT Batamas Niaga Jaya. *Computer Based Information System Journal*, 6(2). DOI: <https://doi.org/10.33884/cbis.v6i2.709> .
- [4] Rosmini, R., Fadlil, A., & Sunardi, S. (2018). Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah. *It Journal Research And Development*, 3(1), 22–31. DOI: [https://doi.org/10.25299/itjrd.2018.vol3\(1\).1773](https://doi.org/10.25299/itjrd.2018.vol3(1).1773) .
- [5] Parlina, I., Windarto, A. P., Wanto, A., & Lubis, M. R. (2018). Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center Untuk Clustering Program Sdp. *Computer Engineering, Science and System Journal*, 3(1), 87. DOI: <https://doi.org/10.24114/cess.v3i1.8192> .
- [6] Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi. *Computer Engineering, Science and System Journal*, 3(2), 173. DOI: <https://doi.org/10.24114/cess.v3i2.9661> .
- [7] Nur, F., Zarlis, M., & Nasution, B. B. (2017). Penerapan Algoritma K-Means Pada Siswa Baru Sekolah Menengah Kejuruan Untuk Clustering Jurusan. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, 1(2), 100–105. DOI: <https://doi.org/10.30743/infotekjar.v1i2.70> .
- [8] Febianto, N. I., & Palasara, N. (2019). Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 8(2), 130. DOI: <https://doi.org/10.32736/sisfokom.v8i2.653> .
- [9] Aprianti, W., & Permadi, J. (2018). K-Means Clustering untuk Data Kecelakaan Lalu Lintas Jalan Raya di Kecamatan Pelaihari. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(5), 613. DOI: <https://doi.org/10.25126/jtiik.2018551113> .
- [10] Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. DOI: <https://doi.org/10.25077/teknosi.v5i1.2019.17-24> .
- [11] Alfandi, D., Ernawati, & Purwandari, E. P. (2018). Implementasi K-Means Clustering Dan Pemetaan Pemukiman Kumuh Di Kota Bengkulu Berbasis Web. *Jurnal Rekursif*, 6(2).
- [12] Triyansyah, D., & Fitrihanah, D. (2018). Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing. *Jurnal Telekomunikasi Dan Komputer*, 8(3), 163. DOI: <https://doi.org/10.22441/incomtech.v8i3.4174> .
- [13] Maulida, L. (2018). Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 167. DOI: <https://doi.org/10.14421/jiska.2018.23-06> .
- [14] Widodo, W., & Wahyuni, D. (2017). Implementasi Algoritma K-Means Clustering Untuk Mengetahui Bidang Skripsi Mahasiswa Multimedia Pendidikan Teknik Informatika Dan Komputer Universitas Negeri Jakarta. *PINTER: Jurnal Pendidikan Teknik Informatika Dan Komputer*, 1(2), 157–166. DOI: <https://doi.org/10.21009/pinter.1.2.10> ,
- [15] Praja, B. S., Kusuma, P. D., & Setianingsih, C. (2019). Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Penumpang Dan Kapal Angkutan Laut Di Indonesia. *E-Proceeding of Engineering*, 6(1), 1442.
- [16] Efendi, R., Coastera, F. F., & Tanjung, F. (2019). Pengelompokan dan Pemetaan Derajat Kesehatan Kota Bengkulu dengan Metode K-means Clustering. *Jurnal Rekursif*, 7(1), 91–97.
- [17] Lase, Y., & Panggabean, E. (2019). Implementasi Metode K-Means Clustering Dalam Sistem Pemilihan Jurusan Di SMK Swasta Harapan Baru. *Jurnal Teknologi Dan Ilmu Komputer Prima (JUTIKOMP)*, 2(2), 43. DOI: <https://doi.org/10.34012/jutikomp.v2i2.723> .
- [18] Ali, A. (2019). Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(1), 186–195. DOI: <https://doi.org/10.30812/matrik.v19i1.529> .
- [19] Kelebihan Algoritma k-means clustering [online] (diperbaharui 21 Januari 2017) <https://id.wikipedia.org/wiki/K-means> [Diakses 11 Mei 2020].