

Penerapan Algoritma K-Means Clustering Dalam Pengelompokan Data Penjualan Pada Pabrik Mobil Toyota Indonesia

Novemli Firdaus ¹⁾*, Sarjon Defit ²⁾, Gunadi Widi Nurcahyo ³⁾

Universitas Putra Indonesia "YPTK" PADANG ¹⁾²⁾³⁾

novemlifirdaus05@gmail.com ¹⁾*, sarjon_defit@upiypk.ac.id ²⁾, gunadiwn@yptk.ac.id ³⁾

Abstrak

Data Mining merupakan upaya untuk mengeksplorasi data yang telah dipilih dengan tujuan menemukan wawasan dan pengetahuan yang bermanfaat. Metode Clustering merupakan pendekatan non-hirarki yang digunakan untuk memisahkan objek-objek ke dalam satu atau lebih Cluster berdasarkan karakteristik data. Saat ini Algoritma K-Means Clustering banyak digunakan pada perusahaan - perusahaan berskala besar pada tahun 2022 perusahaan mobil Toyota telah menjual 300 ribu lebih unit mobil dari semua type, Hasil dari klasterisasi berakhir sampai iterasi ke 11 karena pada iterasi 11 mendapatkan hasil yang sama dengan iterasi sebelumnya yaitu iterasi 10, Dalam pengelompokan penjualan mobil, terdapat tiga kelompok, yaitu C0 (Kurang Diminati), C1 (Diminati), dan C2 (Sangat Diminati). dapat di lihat pada Tabel 4.20. Dari kelompok mobil yang ada dapat dilihat bagaimana tingkat minat konsumen atau pembeli terhadap produk mobil toyota sehingga perusahaan dapat menyesuaikan produk mana yang ingin dijadikan prioritas untuk dipasarkan, untuk mobil-mobil yang kurang diminati perusahaan bisa melakukan evaluasi baik dari segi kualitas produk, pemasaran, iklan, dan harga yang kompetitif sehingga dapat menimbulkan daya tarik bagi pembeli serta mampu bersaing dengan kompetitor-kompetitor produk sejenis.

Kata kunci: Data Mining, Algoritma K-Means Clustering, RapidMiner, Penjualan Mobil Toyota

Abstract

[Application of the K-Means Clustering Algorithm in Grouping Sales Data at the Indonesian Toyota Car Manufacturer]. Data Mining is an attempt to explore selected data with the aim of finding useful insights and knowledge. The Clustering method is a non-hierarchical approach used to separate objects into one or more clusters based on data characteristics. Currently, the K-Means Clustering algorithm is widely used in large-scale companies in 2022 the Toyota car company has sold more than 300 thousand units of cars of all types, The results of clustering end until the 11th iteration because in iteration 11 it gets the same results as the previous iteration, namely iteration 10, In the grouping of car sales, there are three groups, namely C0 (Less Desirable), C1 (Desirable), and C2 (Highly Desirable). can be seen in Table 4.20. From the existing car group, it can be seen how the level of consumer interest or buyers in front of Toyota car products so that the company can adjust which products want to be a priority to be marketed, for cars that are less desirable the company can evaluate both in terms of product quality, marketing, advertising, and competitive prices so that it can cause attraction for buyers and be able to compete with competitors of similar products.

Keywords- Data Mining, Algoritma K-Means Clustering, RapidMiner, Toyota Car Sales

1. PENDAHULUAN

Salah satu metode dalam Data Mining adalah Knowledge Discovery In Database, digunakan untuk mencari informasi dari data yang sebelumnya tidak diketahui serta menemukan pola tersembunyi yang terdapat di dalam data[1]. Dalam proses Knowledge Discovery In Database, terdapat serangkaian langkah yang dimulai dari pemilihan data hingga tahap interpretasi dan evaluasi[2] Proses Knowledge Discovery In Database tidak hanya melibatkan interpretasi tetapi juga penilaian yang relevan[3]. Data

Mining adalah proses pengeksploasian data untuk menggali informasi yang memiliki nilai penting[4]. Data Mining dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan field[5] Salah satu metode data mining yang dapat diimplementasikan dalam data penjualan yaitu teknik clustering[6] Clustering merupakan proses mengelompokkan item data ke dalam kelompok yang lebih kecil dengan tujuan agar setiap kelompok memiliki tingkat kemiripan yang signifikan[7]. Clustering adalah metode analisis data yang dapat digunakan memecahkan masalah dalam

pengelompokan data.[8] Salah satu metode clustering yang sangat terkenal adalah algoritma k-means, karena k-means memiliki algoritma yang sederhana dan efisien[9] K-Means Clustering merupakan salah satu algoritma tanpa pengawasan yang paling banyak digunakan untuk mengelompokkan data dalam sejumlah klaster tertentu[10] Algoritma K-Means Clustering saat ini banyak digunakan pada perusahaan-perusahaan berskala besar yang memiliki banyak sekali data, salah satunya pada perusahaan teknologi transportasi. Tidak jarang suatu pabrik transportasi mengeluarkan banyak sekali model kendaraan, setidaknya pada tahun 2022 perusahaan mobil Toyota telah menjual 300 ribu lebih unit mobil dari semua type. Dari banyaknya total penjualan mobil ini tentu saja pengelompokan data penjualan akan sulit kalau hanya dilakukan dengan perhitungan secara manual dikarenakan penjualan produk toyota tersebar di seluruh provinsi di Indonesia. Bagi konsumen dengan banyaknya jenis kendaraan yang ada dipasaran konsumen perlu riset dalam menentukan jenis kendaraan yang sesuai dengan kebutuhan, seperti riset perbandingan harga, jenis mobil, kapasitas penumpang, dan lain-lain. menggunakan Algoritma K-Means Clustering ini diharapkan dapat membantu pengambilan keputusan yang tepat dalam menangan[7]

Secara umum, tujuan dari clusterisasi data penjualan mobil adalah menggunakan Algoritma K-Means Clustering untuk menentukan produk potensial[2] Analisa data tingkat penjualan produk mobil menggunakan Algoritma K-Means Clustering. Manfaat yang diperoleh dari penelitian ini adalah dengan metode Data Mining Algoritma K-Means Clustering peneliti dapat mengelompokkan data penjualan ke dalam sebuah Cluster. Dengan menggunakan metode K-Means Clustering dapat merekondisikan dan menentukan jenis mobil yang sesuai dengan yang dibutuhkan. Dengan konsep Data Mining menggunakan algoritma K-Means Clustering dapat menganalisa serta menentukan produk potensial Penelitian-penelitian tentang sistem pengatur lampu lalu lintas pintar telah dilakukan. Beberapa di antaranya Penelitian terdahulu menggunakan metode K-Means Clustering untuk menentukan penjualan sparepart Toyota dengan metode K-Means Clustering, data penjualan yang digunakan sebanyak 50 data. Output yang diharapkan adalah menghasilkan 3 Clustering yaitu (C1) data penjualan paling laris, (C2) data penjualan yang laris, dan (C3) data penjualan kurang laris (Kesuma dan Tamba, 2020). Penelitian selanjutnya menggunakan metode K-Means Clustering untuk prediksi penjualan mobil dengan data riset berdasarkan pada website resmi GaiKindo sebanyak 900 data. Cluster yang terbentuk setelah dilakukan proses K-Means Clustering terbagi menjadi tiga Clustering yaitu Cluster 0 jumlah anggota 235 dengan presentase 26% dikategorikan Laris, Cluster 1 jumlah anggota 604 dengan presentase 67% dikategorikan Kurang Laris, dan Cluster 2 jumlah anggota 61 dengan presentase 7% dikategorikan Paling Laris, dari proses

Clustering di atas dapat diperoleh validasi DBI (Davies Bouldin Index) dengan nilai 0,341[11]

2. BAHAN DAN METODE

a. Identifikasi Masalah.

Identifikasi masalah adalah langkah pertama dalam melakukan penelitian, yang memiliki tujuan untuk mengkaji dengan seksama objek penelitian yang akan diteliti.

b. Analisa masalah

Melakukan analisis masalah memiliki tujuan untuk memahami dengan lebih baik isu-isu yang muncul dalam kerangka penelitian. Dengan menggunakan Algoritma K-Means Clustering, diharapkan analisis masalah mampu memberikan solusi yang konstruktif dalam menyelesaikan tantangan penelitian tersebut.

c. Menentukan Tujuan

Setelah mengevaluasi permasalahan dan merumuskan permasalahan dengan tepat, langkah berikutnya adalah menetapkan tujuan yang ingin dicapai dalam penelitian. Tujuan utama penelitian ini adalah mengatasi permasalahan yang ada dengan sukses melalui implementasi Algoritma K-Means Clustering untuk mengelompokkan data penjualan mobil Toyota.

d. Melakukan Studi Literatur

Studi literatur adalah langkah penting untuk memperoleh pemahaman mendalam tentang metode dan juga sebagai panduan referensi yang mendukung seluruh proses penelitian dengan memberikan landasan yang kuat. Dalam konteks penelitian ini, referensi utamanya berasal dari jurnal-jurnal ilmiah sebelumnya yang mendiskusikan topik Data Mining dan implementasi Algoritma K-Means Clustering.

e. Mengumpulkan Data

Salah satu metode pengumpulan data yang digunakan oleh peneliti adalah Studi Pustaka, di mana penelitian ini mencakup pengumpulan bahan dan pemahaman terhadap jurnal-jurnal yang berhubungan dengan isu penelitian. Data penelitian yang saya gunakan diperoleh dari website Gaikindo, yaitu Gabungan Industri Kendaraan Bermotor Indonesia. Website ini menjadi sumber utama informasi karena menyediakan berbagai data yang relevan dengan penelitian ini. Dengan mengandalkan data dari website Gaikindo, saya dapat memastikan bahwa penelitian saya didukung oleh sumber informasi yang terpercaya dan terkini

dalam menganalisis permasalahan yang menjadi fokus penelitian.

f. Mengolah Data

Pada proses pengolahan data, data diolah dengan menggunakan Algoritma K-Means Clustering [12] dengan tahapan sebagai berikut:

- 1) Menentukan jumlah Cluster.
- 2) Menentukan sentroid awal secara acak.
- 3) Mengumpulkan data kedalam cluster terdekat menggunakan rumus berikut:

$$De = \sqrt{(Ri - Ti)^2 + (Rj - Tj)^2} \quad (1)$$

- 4) Melakukan perhitungan ulang pusat centroid baru dengan perhitungan rata-rata data yang ada dalam Cluster yang telah terbentuk menggunakan rumus berikut:

$$CI = (R1 + R2 + R3 + \dots + Rn) / (\Sigma R) \quad (2)$$

- 5) Hitungan ulang setiap objek dengan menggunakan pusat Cluster (centroid) baru.

g. Proses Pengujian

Beberapa tahapan untuk melakukan pengujian sistem antara lain:

- 1) Dalam tahap pengujian manual, Microsoft Excel digunakan untuk menerapkan Algoritma K-Means Clustering.
- 2) Dalam tahap pengujian selanjutnya, Software RapidMiner digunakan untuk menerapkan Algoritma K-Means Clustering.
- 3) Melakukan perbandingan hasil pengujian manual dengan aplikasi untuk validasi hasil.

h. Analisa Hasil

Analisa hasil adalah proses evaluasi dan pemahaman terhadap data dan temuan yang diperoleh dari pengujian atau percobaan yang telah dilakukan, dengan tujuan untuk mengidentifikasi pola, tren, atau informasi penting yang dapat digunakan untuk membuat keputusan atau kesimpulan dalam konteks penelitian atau studi yang sedang dilakukan

3. HASIL DAN PEMBAHASAN

a. Analisis Data

Tabel 1 Data yang di peroleh sebanyak 80 tipe mobil Toyota dengan total penjualan sebanyak 186.247 unit dari rentang semester 1 (Januari-Juni) dan semester ke2 (Juli-Desember) 2022. Data tersebut diolah menggunakan metode K-Means Clustering. Data penjualan dapat di lihat pada

Tabel 1. Data penjualan Mobil Toyota Tahun 2022

| Data | Semester 1 | Semester 2 |
|------|------------|------------|
|------|------------|------------|

| | | |
|-----|--------|-------|
| 1 | 200 | 650 |
| 2 | 86 | 110 |
| 3 | 314 | 573 |
| 4 | 183 | 101 |
| 5 | 622 | 5232 |
| 6 | 6962 | 7249 |
| 7 | 1316 | 3154 |
| 8 | 9695 | 10686 |
| 9 | 6185 | 8417 |
| 10 | 546 | 542 |
| 11 | 4154 | 2485 |
| 12 | 2 | 992 |
| 13 | 6997 | 9427 |
| 14 | 6678 | 1282 |
| 15 | 416 | 418 |
| 16 | 732 | 1814 |
| 17 | 157 | 253 |
| 18 | 2303 | 2787 |
| 19 | 2906 | 3636 |
| 20 | 1133 | 238 |
| ... | | ... |
| 70 | 85 | 65 |
| 71 | 4.642 | 5.315 |
| 72 | 4.248 | 6.638 |
| 73 | 0 | 1567 |
| 74 | 0 | 1248 |
| 75 | 0 | 15873 |
| 76 | 0 | 7545 |
| 77 | 490 | 1 |
| 78 | 711 | 2 |
| 79 | 11.524 | 866 |
| 80 | 3740 | 15 |

(Sumber: website GAIKINDO Indonesia Automobile Industry Data Tahun 2022)

Kemudian dari data yang ada akan dilakukan pengolahan data penjualan mobil dengan menggunakan Algoritma K-Means Clustering sampai didapatkan hasil klaster untuk setiap mobil. Tahapan algoritma K-Means Clustering sebagai berikut:

- 1) Menentukan jumlah cluster
- 2) Menentukan titik centroid secara acak
- 3) Menghitung jarak terdekat dengan centroid
- 4) Mengelompokkan data ke masing-masing cluster
- 5) Jika cluster berubah maka proses dilanjutkan ke iterasi berikutnya, jika tidak ada perubahan maka proses dihentikan.

b. Hasil Perhitungan Manual

- 1) Centroid awal
Tabel 2 Nilai centroid awal ditentukan secara acak dari dataset yang tersedia.

Tabel 2. Centroid Awal

| Centroid Awal | X | Y |
|---------------|-------|-------|
| C0 | 416 | 418 |
| C1 | 1.066 | 1.946 |
| C2 | 1.612 | 3.243 |

2) Menghitung Jarak

Tabel 3 Menghitung Jarak Terdekat Data dengan Centroid Untuk melakukan perhitungan jarak setiap data terhadap pusat cluster dapat dilakukan dengan menggunakan rumus persamaan (1)

Pusat cluster 0

$$\begin{aligned}
 C01 &= \sqrt{((200-416)^2+(650-418)^2)} \\
 &= \sqrt{((-216)^2+(232)^2)} \\
 &= \sqrt{(46.656+53.824)} \\
 &= 317
 \end{aligned}$$

$$\begin{aligned}
 C02 &= \sqrt{((84-416)^2+(110-418)^2)} \\
 &= \sqrt{((-330)^2+(-308)^2)} \\
 &= \sqrt{(108.900+94.864)} \\
 &= 451
 \end{aligned}$$

Pusat cluster 1

$$\begin{aligned}
 C11 &= \sqrt{((200-1.066)^2+(650-1.946)^2)} \\
 &= \sqrt{((-866)^2+(-1.296)^2)} \\
 &= \sqrt{(749.956+1.679.616)} \\
 &= 1.559
 \end{aligned}$$

$$\begin{aligned}
 C12 &= \sqrt{((84-1.066)^2+(110-1.946)^2)} \\
 &= \sqrt{((-980)^2+(-1.836)^2)} \\
 &= \sqrt{(960.400+3.370.896)} \\
 &= 2.081
 \end{aligned}$$

Pusat cluster 2

$$\begin{aligned}
 C21 &= \sqrt{((200-1.612)^2+(650-3.243)^2)} \\
 &= \sqrt{((-1.412)^2+(-2.593)^2)} \\
 &= \sqrt{(1.993.744+6.723.649)} \\
 &= 2.953
 \end{aligned}$$

$$\begin{aligned}
 C22 &= \sqrt{((84-1.612)^2+(110-3.243)^2)} \\
 &= \sqrt{((-1.526)^2+(-3.133)^2)} \\
 &= \sqrt{(2.328.676+9.815.689)} \\
 &= 3.485
 \end{aligned}$$

3) Hasil nilai cluster iterasi 1

Tabel 3. Hasil nilai cluster iterasi 1

| Data | C0 | C1 | C2 | K |
|------|--------|--------|--------|---|
| 1 | 317 | 1.559 | 2.953 | 0 |
| 2 | 451 | 2.081 | 3.485 | 0 |
| 3 | 186 | 1.565 | 2.969 | 0 |
| 4 | 393 | 2.045 | 3.452 | 0 |
| 5 | 4.818 | 3.316 | 2.222 | 2 |
| 6 | 9.461 | 7.930 | 6.684 | 2 |
| 7 | 2.880 | 1.234 | 309 | 2 |
| 7 | 2.880 | 1.234 | 309 | 2 |
| 8 | 13.839 | 12.282 | 10.988 | 2 |
| 9 | 9.862 | 8.251 | 6.905 | 2 |
| 10 | 180 | 1.497 | 2.904 | 0 |
| 11 | 4.271 | 3.135 | 2.653 | 2 |
| 12 | 708 | 1.429 | 2.768 | 0 |
| 13 | 11.157 | 9.547 | 8.200 | 2 |
| 14 | 6.321 | 5.651 | 5.432 | 2 |
| 15 | 0 | 1.661 | 3.068 | 0 |
| 16 | 1.431 | 359 | 1.678 | 1 |

| | | | | |
|----|--------|--------|--------|---|
| 17 | 307 | 1.922 | 3.325 | 0 |
| 18 | 3.029 | 1.496 | 828 | 2 |
| 19 | 4.069 | 2.498 | 1.352 | 2 |
| 20 | 739 | 1.709 | 3.043 | 0 |
| 21 | 1.535 | 724 | 1.815 | 1 |
| 22 | 369 | 2.022 | 3.428 | 0 |
| 23 | 2.562 | 954 | 657 | 2 |
| 24 | 2.276 | 618 | 796 | 1 |
| 25 | 1.661 | 0 | 1.407 | 1 |
| 26 | 14.453 | 12.917 | 11.645 | 2 |
| 27 | 11.323 | 9.809 | 8.571 | 2 |
| 28 | 435 | 2.094 | 3.501 | 0 |
| 29 | 1.631 | 1.134 | 2.109 | 1 |
| 30 | 243 | 1.900 | 3.307 | 0 |
| 31 | 3.560 | 1.899 | 493 | 2 |
| 32 | 2.514 | 991 | 884 | 2 |
| 33 | 418 | 2.053 | 3.457 | 0 |
| 34 | 4.137 | 2.679 | 1.757 | 2 |
| 35 | 5.364 | 4.407 | 3.980 | 2 |
| 36 | 279 | 1.927 | 3.334 | 0 |
| 37 | 979 | 838 | 2.192 | 1 |
| 38 | 383 | 2.034 | 3.440 | 0 |
| 39 | 160 | 1.810 | 3.217 | 0 |
| 40 | 4.770 | 3.887 | 3.605 | 2 |
| 41 | 96 | 1.614 | 3.021 | 0 |
| 42 | 3.190 | 2.372 | 2.472 | 1 |
| 43 | 843 | 1.330 | 2.641 | 0 |
| 44 | 1.966 | 1.136 | 1.848 | 1 |
| 45 | 416 | 1.860 | 3.249 | 0 |
| 46 | 420 | 1.910 | 3.302 | 0 |
| 47 | 432 | 1.769 | 3.152 | 0 |
| 48 | 888 | 1.300 | 2.601 | 0 |
| 49 | 1.192 | 587 | 1.937 | 1 |
| 50 | 3.068 | 1.407 | 0 | 2 |
| 51 | 368 | 1.982 | 3.385 | 0 |
| 52 | 514 | 2.136 | 3.538 | 0 |
| 53 | 215 | 1.632 | 3.034 | 0 |
| 54 | 293 | 1.939 | 3.345 | 0 |
| 55 | 215 | 1.875 | 3.282 | 0 |
| 56 | 495 | 2.150 | 3.556 | 0 |
| 57 | 1.027 | 1.964 | 3.249 | 0 |
| 58 | 385 | 1.286 | 2.692 | 0 |
| 59 | 503 | 2.143 | 3.547 | 0 |
| 60 | 499 | 2.124 | 3.527 | 0 |
| 61 | 501 | 2.099 | 3.499 | 0 |
| 62 | 595 | 1.159 | 2.549 | 0 |
| 63 | 645 | 1.082 | 2.475 | 0 |
| 64 | 565 | 1.101 | 2.507 | 0 |
| 65 | 1.353 | 460 | 1.776 | 1 |
| 66 | 426 | 2.020 | 3.420 | 0 |
| 67 | 6.035 | 4.450 | 3.169 | 2 |
| 68 | 654 | 1.030 | 2.432 | 0 |
| 69 | 360 | 2.017 | 3.424 | 0 |
| 70 | 484 | 2.121 | 3.526 | 0 |
| 71 | 6.468 | 4.913 | 3.671 | 2 |
| 72 | 7.306 | 5.669 | 4.298 | 2 |
| 73 | 1.222 | 1.131 | 2.325 | 1 |
| 74 | 928 | 1.274 | 2.565 | 0 |

| | | | | |
|----|--------|--------|--------|---|
| 75 | 15.461 | 13.968 | 12.732 | 2 |
| 76 | 7.139 | 5.700 | 4.594 | 2 |
| 77 | 424 | 2.028 | 3.431 | 0 |
| 78 | 510 | 1.976 | 3.364 | 0 |
| 79 | 11.117 | 10.514 | 10.193 | 2 |
| 80 | 3.348 | 3.298 | 3.866 | 1 |

Dalam hasil iterasi pertama ini, dapat dilihat bahwa data telah tercluster menjadi tiga kelompok utama, yaitu C0, C1, dan C2, yang menggambarkan kesamaan antar data di dalam masing-masing cluster tersebut.

- a) Anggota cluster 0 (C0) terdiri dari 43 data mobil dengan nomor : 1, 2, 3, 4, 10, 12, 15, 17, 20, 22, 28, 30, 33, 36, 38, 39, 41, 43, 45, 46, 47, 48, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 66, 68, 69, 70, 74, 77 dan 78
- b) Anggota cluster 1 (C1) terdiri dari 12 data mobil dengan nomor : 16, 21, 24, 25, 29, 37, 42, 44, 49, 65, 73 dan 80
- c) Anggota cluster 0 (C2) terdiri dari 25 data mobil dengan nomor : 5, 6, 7, 8, 9, 11, 13, 14, 18, 19, 23, 26, 27, 31, 32, 34, 35, 40, 50, 67, 71, 72, 75, 76 dan 79

setelah perhitungan iterasi pertama selesai dilanjutkan dengan iterasi berikutnya sampai dataset pada cluster tidak berubah lagi, dengan langkah-langkah yang sama pada pencarian iterasi sebelumnya. Pada penelitian ini total iterasi pada perhitungan manual berhenti sampai pada perhitungan iterasi ke 11 karna sudah mendapatkan hasil yang sama dengan cluster itersi 10 dimana pada C0 terdapat 54 item dataset C1 terdapat 19 dataset dan C2 terdapat 7 dataset

c. Hasil Pengujian RapidMiner

1) Hasil cluster

| id | TYPE MOBIL | cluster | Semester 1 | Semester 2 |
|----|------------------|-----------|------------|------------|
| 1 | All New Coro... | cluster_0 | 200 | 650 |
| 2 | All New Coro... | cluster_0 | 86 | 110 |
| 3 | All New Cam... | cluster_0 | 314 | 573 |
| 4 | All New Cam... | cluster_0 | 183 | 101 |
| 5 | Avanza 1.3 A/... | cluster_1 | 622 | 5232 |
| 6 | All New Avan... | cluster_2 | 6962 | 7249 |
| 7 | All New Avan... | cluster_1 | 1316 | 3154 |
| 8 | All New Avan... | cluster_2 | 9695 | 10686 |
| 9 | All New Avan... | cluster_2 | 6185 | 8417 |
| 10 | All New Avan... | cluster_0 | 546 | 542 |
| 11 | All New Velo... | cluster_1 | 4154 | 2485 |

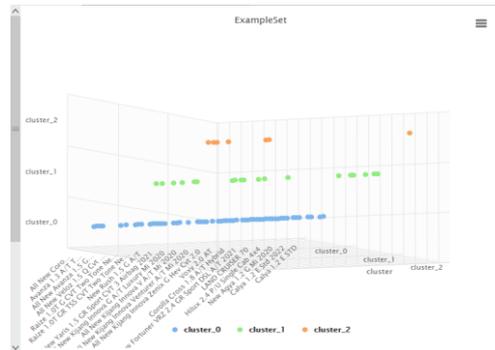
Gambar 1. Contoh Penulisan Nama Gambar

- a) Gambar 1 Setelah proses dijalankan maka pada ExampleSet Clustering data yang diinputkan dikelompokkan menjadi 3 cluster, cluster_0 (C0), cluster_1 (C1) dan cluster_2

(C2) seperti Gambar 5.5.

- b) Data 1, 2, 3, 4 dikelompokkan ke dalam cluster_0, data 5, 7, 11 dikelompokkan ke dalam cluster_1 dan data 6, 8, 9 dikelompokkan ke dalam cluster_2 dan seterusnya sampai data ke 80.

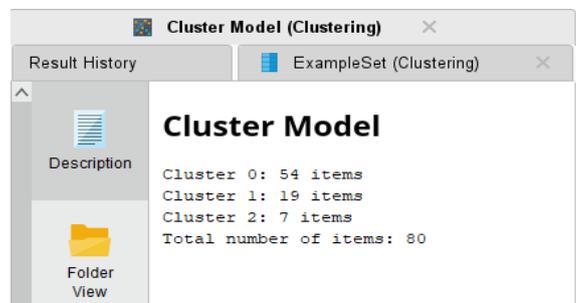
2) Hasil Hasil Visualizations Cluster



Gambar 2. Hasil Visualizations Cluster

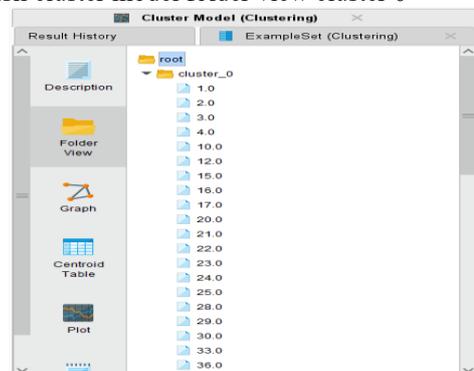
Gambar 2 Pada toolbar Visualizations terlihat bahwa hasil cluster 2 (C2) Sangat Diminati, cluster 1 (C1) Diminati, cluster 0 (C0) Kurang Diminati, terlihat pada Gambar 2.

- 3) Gambar 3 Pada bagian cluster model cluster 0 ada 54 items, cluster 1 ada 19 items dan cluster 2 ada 7 items dari total 80 data.



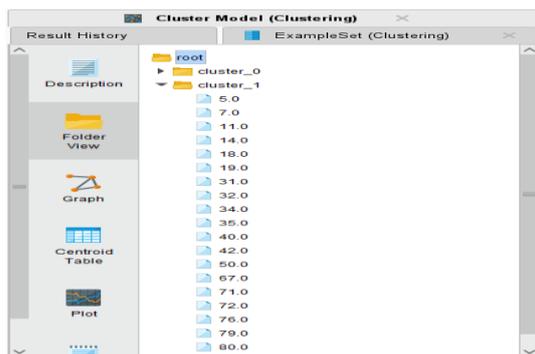
Gambar 3. Hasil Cluster Model Description

- 4) Hasil cluster model folder view cluster 0



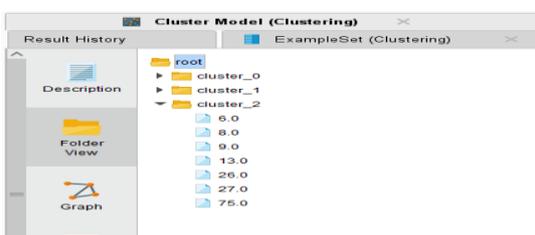
Gambar 4. Hasil Cluster Model Folder View Cluster 0

- 5) Hasil cluster model folder view cluster 1



Gambar 5. Hasil Cluster Model Folder View Cluster 1

6) Hasil cluster model folder view cluster



Gambar 6. Hasil Cluster Model Folder View Cluster 2

Gambar 4, Gambar 5 dan Gambar 6 Setelah dilakukan pengujian dan perbandingan hasil terhadap 80 data sampel pada pengujian Microsoft Excel 2019 Pada Bab IV dan dengan menggunakan software RapidMiner Studio versi 9.10 didapatkan hasil cluster dan anggota cluster yang sama.

4. KESIMPULAN

Berdasarkan dari uraian pada bab-bab sebelumnya dapat ditarik kesimpulan sebagai berikut:

1. Dengan Data Mining didapatkan pengetahuan baru (knowledge) bagaimana cara pengelompokan data dengan algoritma K-means Clustering. Clustering pengelompokan data penjualan mobil Toyota Indonesia berdasarkan pada total penjualan selama tahun 2022. Dengan clusterisasi dapat dilihat bagaimana minat konsumen terhadap tipe-tipe mobil Toyota Indonesia, maka dapat disimpulkan apa saja tipe-tipe mobil yang (Diminati), (Kurang Diminati) dan (Sangat Diminati) oleh konsumen maka perlu dilakukan penyesuaian atau upgrade baik dari segi kualitas maupun harga.
2. Dari segi penjualan mobil Toyota pada tahun 2022 jumlah data yang digunakan adalah 80 tipe mobil dengan total penjualan keseluruhan sebanyak 310.569 yang telah dikelompokkan menjadi 3 cluster. Dimana cluster 0 (C0) berjumlah 54 data mobil dengan rata-rata penjualan 1.105 selama tahun 2022 maka cluster

0 (C0) Kurang Diminati, cluster 1 (C1) berjumlah 19 data mobil dengan rata-rata penjualan 6.910 selama tahun 2022 maka cluster 1(C1) Termasuk tipe-tipe mobil yang Diminati, cluster 2 (C2) berjumlah 7 data mobil dengan rata-rata penjualan 17.087 selama tahun 2022 maka cluster 2(C2) Termasuk tipe-tipe mobil yang Sangat Diminati

5. DAFTAR PUSTAKA

- [1] M. R. Nugroho, I. E. Hendrawan, and P. P. Purwantoro, "Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI," *Nuansa Inform.*, vol. 16, no. 1, pp. 125–133, 2022, doi: 10.25134/nuansa.v16i1.5294.
- [2] A. Permadi and Y. A. Wijaya, "Pengelompokan Dataset Bus Menggunakan Algoritma," vol. 8, no. 1, pp. 19–33, 2023.
- [3] W. Romadhona, B. Indarmawan Nugroho, and A. Alim Murtopo, "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means," *J. Minfo Polgan*, vol. 11, no. 2, pp. 100–104, 2022, doi: 10.33395/jmp.v11i2.11797.
- [4] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [5] Noviyanto and P. Ekasari, "Algoritma K-Means Untuk Klasterisasi Jabatan Fungsional Dosen Pada Perguruan Tinggi Swasta Di Lingkungan LLDikti Wilayah III," *Paradig. J.*, vol. 24, no. 1, pp. 103–107, 2022.
- [6] I. S. Mangku Negara, P. Purwono, and I. A. Ashari, "Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 6, no. 3, p. 153, 2022, doi: 10.31328/jointecs.v6i3.2693.
- [7] T. Informasi, M. Nurrohman, and P. Sukmasetya, "Jurnal Sistem Klasterisasi Volume Sampah Organik di Kota Magelang menggunakan K-Means Clustering System of Organic Waste Volume in Magelang City using K-Means," vol. 5, pp. 146–153, 2023.
- [8] T. Hidayat, "Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means Clustering," *J. Sistim Inf. dan Teknol.*, vol. 4, pp. 19–24, 2022, doi: 10.37034/jsisfotek.v4i1.115.
- [9] M. C. Untoro, L. Anggraini, M. Andini, H. Retnosari, and M. A. Nasrulloh, "Penerapan metode k-means clustering data COVID-19 di Provinsi Jakarta," *Teknologi*, vol. 11, no. 2, pp. 59–68, 2021, doi:

- 10.26594/teknologi.v1i1i2.2323.
- [10] S. A. Putri and M. Mustakim, "Dimensional Data Unsupervised Learning Using an Analytic Hierarchy Process in Determining Attributes in the Classification Algorithm," *Build. Informatics, Technol. Sci.*, vol. 4, no. 1, pp. 235–240, 2022, doi: 10.47065/bits.v4i1.1752.
- [11] S. Butsianto and N. T. Mayangwulan, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 3, pp. 187–201, 2020, doi: 10.32672/jnkti.v3i3.2428.
- [12] N. Mirantika, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat," *Nuansa Inform.*, vol. 15, no. 2, pp. 92–98, 2021, doi: 10.25134/nuansa.v15i2.4321.