JURNAL INFORMATIKA, Vol.11 No.1 April 2024, Halaman 1-8
ISSN: 2355-6579 | E-ISSN**:** 2528-2247

Implementasi Algoritma K-Means Clustering Untuk Pemilihan Dan Rekomendasi Mobil Bekas

Ihsanul Ependi[1*], Rachmat Adi Purnama[2]

[1,2] Universitas Bina Sarana Informatika
Jl. Margonda No.8, Pondok Cina, Kota Depok, Jawa Barat, Indonesia

E-mail korespondensi: 15200211@bsi.ac.id

Informasi Artikel	Diterima: 00-00-2023	Direvisi: 00-00-2023	Disetujui: 00-00-2023

Abstrak

Mobil bekas menjadi pilihan yang diminati masyarakat karena harganya yang terjangkau dan spesifikasinya yang beragam. Penelitian ini bertujuan untuk mengelompokkan mobil bekas berdasarkan atribut harga, jarak tempuh, dan tahun produksi menggunakan algoritma K-Means Clustering. Metode Elbow digunakan untuk menentukan jumlah klaster yang optimal. Penelitian ini bertujuan untuk menganalisis pengaruh harga, jarak tempuh, dan tahun produksi mobil terhadap keputusan rekomendasi mobil bekas. Selain itu, penelitian ini juga menganalisis perbedaan jarak antara centroid pada setiap atribut untuk memberikan rekomendasi yang relevan kepada konsumen. Hasil pengujian menunjukkan bahwa mobil bekas dapat dikelompokkan menjadi tiga klaster. Klaster 1 mencakup mobil dengan harga rendah, jarak tempuh tinggi, dan tahun produksi lama, sebanyak 45 dari 159 data. Klaster 2 berisi mobil dengan harga menengah, jarak tempuh sedang, dan tahun produksi menengah, sebanyak 68 data. Klaster 3 adalah mobil dengan harga tinggi, jarak tempuh rendah, dan tahun produksi baru, sebanyak 46 data. Penelitian ini diharapkan dapat membantu konsumen dalam memilih mobil bekas yang sesuai kebutuhan serta memberikan wawasan kepada pelaku bisnis untuk memahami segmentasi pasar mobil bekas.

Kata Kunci: K-Means Clustering, Mobil Bekas, Harga, Jarak Tempuh, Tahun produksi

Abstract

Used cars are a popular choice among people because of their affordable prices and various specifications. This research aims to group used cars based on the attributes of price, mileage and year of production using the K-Means Clustering algorithm. The Elbow method is used to determine the optimal number of clusters. This research aims to analyze the influence of price, mileage, and year of car production on used car recommendation decisions. Apart from that, this research also analyzes the differences in distance between the centroids for each attribute to provide relevant recommendations to consumers. The test results show that used cars can be grouped into three clusters. Cluster 1 includes cars with low prices, high mileage and old production years, 45 of the 159 data. Cluster 2 contains cars with medium prices, medium mileage and medium production years, totaling 68 data. Cluster 3 is cars with high prices, low mileage, and new production years, totaling 46 data. It is hoped that this research can help consumers choose used cars that suit their needs and provide insight to business people to understand used car market segmentation.

Keywords*: K-Means Clustering, Used Cars, Price, Mileage, Year of Production*

1. Pendahuluan

Kendaraan roda 4 (mobil) merupakan alat transportasi yang dianggap mampu mempermudah hidup manusia. Sejak ditemukannya alat transportasi tersebut, gerak gerik hidup manusia berubah menjadi lebih mudah dan dinamis. Semakin berkembangnya zaman semakin banyak pula pilihan mobil yang ditawarkan oleh produsen mobil ke konsumen.

Berdasarkan data ASEAN Automotive Federation (AAF), penjualan mobil di Asia Tenggara mencapai 2.188.497unit sepanjang periode Januari hingga Agustus 2023. AAF menganalisis kinerja penjualan kendaraan bermotor roda empat di tujuh negara Asia renggara, yaitu Indonesia, Malaysia, Myanmar, Filipina, Singapura, Thailand, dan Vietnam. Tabel diatas menunjukkan bahwa Indonesia memiliki manufaktur mobil terbesar pertama di Asia Tenggara dan di wilayah ASEAN dengan penjualan 675.287 unit selama periode Januari hingga Agustus 2023(Panny Dwi Lestari, Nursiah

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Fitri, Suriyadi, Dewi Comala Sari, Safaruddin, 2024).

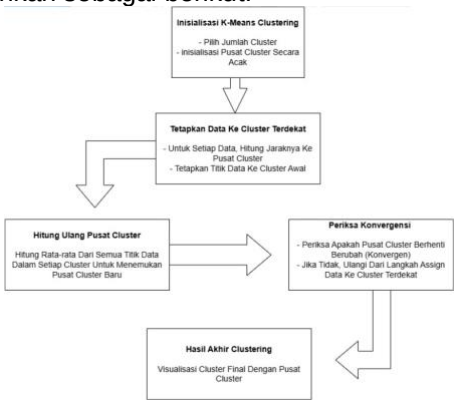
Dengan meningkatnya jumlah produksi mobil baru dan munculnya berbagai model berdampak terhadap melimpahnya jumlah mobil bekas atau yang disebut dengan mobil second, yang dapat menciptakan peluang bisnis bagi pembeli dan penjual. Adapun alasan lain mengapa konsunen maupun pelaku bisnis lebih memilih mobil bekas dikarenakan pajak yang lebih terjangkau dan proses peralihan kepemilikan yang tidak sulit, sehingga menjadikan bisnis ini menjadi lebih menjanjikan. Mobil bekas memiliki harga jual dan sfesifikasi yang beragam, oleh karena itu peneliti tertarik untuk membantu konsumen dalam menentukan pilihan berdasarkan harga, jarak tempuh dan tahun dengan metode k-means clustering.

Dalam analisis data, metode clustering membagi data menjadi beberapa kelompok atau cluster berdasarkan kesamaan karakteristik. Algoritma clustering K-Means dapat menentukan jumlah dari cluster (k) dengan secara acak, dengan nilai yang mewakili pusat pada cluster yang sementara.

Dengan menggunakan atribut harga, jarak tempuh dan tahun pada dataset, metode K-Means dapat digunakan untuk harga, jarak tempuh dan tahun mobil yang diinginkan. Dalam hal merekomendasikan sesuatu dan mengkelompokan data, metode ini dapat menghasilkan nilai yang akurat. “Dalam penyelesaiannya, algoritma K-Means akan menghasilkan titik centroid yang dijadikan tujuan dari algoritma K-Means”(Indraputra & Fitriana, 2020).

2. Metode Penelitian

Langkah tahapan penelitian yang akan dilakukan antara lain melakukan analisa dan mencari pola data yang dapat di jadikan cluster mobil bekas. Hal ini dapat memudahkan penelitian ini dan dapat dilaksanakan secara sistematis dan memungkinkan tercapainya tujuan yang diinginkan sebagai berikut:



Gambar 2.1
Proses dan Langkah Penelitian

inisialisasi K-Means clusterig merupakan langkah pertama didalam proses clustering menggunakan algoritma K-Means. Tahapan yang pertama yaitu menentukan jumlah cluster (K) sebelum melakukan clustering, penelitian perlu menentukan jumlah cluster (K) yang diinginkan. Pemilihan jumlah cluster bisa berdasarkan pengetahuan domain, analisis data sebelumnya, atau memakai metode evaluasi seperti Elbow Method atau Silhoute score.

Tahapan yang kedua yaitu menginisialisasi pusat cluster secara acak. Setelah menentukan jumlah cluster (K), langkah berikutnya yaitu yaitu menginisialisasi pusat cluster awal. Ini biasanya dilakukan secara acak dan melibatkan langkah-langkah seperti memilih (K) titik data acak sebagai pusat awal yaitu pilih secara acak (K) titik dari dataset sebagai pusat cluster, titik-titik ini akan menjadi centroid awal dari cluster yang akan di bentuk. Tahapan ini yaitu menentukan jumlah cluster yang diinginkan.

Langkah assign data ke cluster terdekat adalah salah satu tahapan penting dalam algoritma K-Means clustering. Setelah inisialisasi pusat cluster, proses ini melibatkan penempatan setiap titik data ke cluster yang paling dekat berdasarkan jarak euclidean. Untuk setiap titik data dalam dataset, kita perlu menghitung jarak ke setiap pusat cluster yang telah diinisialisasi atau diperbarui. Perhitungan jarak yang paling umum digunakan dalam K-Means adalah jarak euclidean.

Setelah itu menetapkan titik data ke cluster terdekat setelah menghitung jarak dari setiap titik data ke semua pusat cluster, kita kemudian menetapkan titik data tersebut ke cluster dengan pusat yang memiliki jarak minimum. Setelah semua titik data telah ditetapkan ke cluster terdekat, kita memiliki pembagian awal atau pembaruan dari data ke dalam cluster yang berbeda. Proses ini kemudian akan digunakan untuk menghitung ulang pusat cluster pada langkah berikutnya.

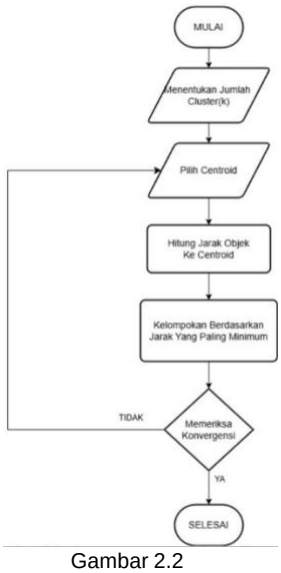
Tujuan menghitung ulang pusat cluster adalah untuk memperbarui posisi pusat (centroid) dari setiap cluster berdasarkan rata-rata posisi dari titik-titik data yang ditetapkan ke cluster tersebut. Dengan menghitung ulang pusat cluster, kita dapat mencari posisi yang lebih representatif yang terbentuk pada iterasi saat ini. Hitung rata-rata dari setiap cluster untuk setiap cluster, hitung rata-rata dari semua titik data yang berada dalam cluster tersebut. Rata-rata ini akan menjadi pusat cluster (centroid) yang baru. Setelah semua dihitung maka perbarui pusat cluster dengan ganti pusat cluster lama dengan pusat cluster baru yang telah dihitung untuk setiap cluster.

Langkah periksa konvergensi algiritna K-Means clustering menentukan kapan algoritma harus berhenti melakukan iterasi. Algoritma K-

Means terus berulang hingga mencapai keadaan stabil (konvergensi) dimana pusat cluster (centroid) tidak berubah lagi secara signifikan. Konvergensi dalam konteks K-Means clustering bahwa posisi pusat cluster tidak lagi berubah atau perubahan yang terjadi sangat kecil dan tidak mempengaruhi hasil clustering secara signifikan. Algoritma berhenti ketika mencapai konvergensi karna pembagian cluster telah dianggap optimal. Algoritma memeriksa perubahan posisi centroid antara dua iterasi berturut-turut. Jika pergerakan centroid kurang dari ambang batas tertentu, maka dianggap telah konvergen.

Langkah terakhir adalah hasil akhir clustering dalam proses K-Means clustering merujuk pada kondisi setelah algoritma K-Means mencapai konvergensi dan berhenti. Pada tahap ini, kita telah memiliki pembagian data ke dalam cluster yang berbeda berdasarkan kedekatan mereka dengan pusat cluster yang dihitung ulang secara iteratif. Setiap titik data dalam dataset telah ditetapkan ke salah satu dari K cluster berdasarkan jarak ke pusat cluster terdekat. Visualisasi hasil clustering adalah cara yang efektif untuk memahami bagaimana data dikelompokkan.

Dalam metode pengolahan cluster peneliti akan membuat langkah atau urutan melalui flowchart. Diagram flowchart algoritma sistem dibawah ini menunjukan penerapan metodologi dan algoritma pada clustering yang akan dibangun untuk menentukan kelompok mobil bekas. Gambaran ini dinyatakan sebagai simbol. Dengan demikian setiap simbol menggambarkan proses tertentu. Sedangkan antara proses digambarkan dengan garis penghubung. Dengan menggunakan flowchart, akan memudahkan peneliti untuk melakukan pengecekan bagian yang terlupa dalam menganalisis masalah.



Gambar 2.2
Flowchart Clustering K-Means

Langkah pertama dalam pengolahan data adalah pemilihan data untuk diproses dengan algoritma K-Means Clustering. Ini melibatkan inisialisasi K-Means Clustering. Dalam inisialisasi K-Means Clustering ini penelitian akan menentukan jumlah cluster (k), setelah penelitian menentukan jumlah cluster selanjutnya adalah menginisialisasi pusat cluster secara acak, memilih (k) titik data acak sebagai pusat awal dan titik-titik ini akan menjadi centroid awal dari cluster yang akan di bentuk.

Setelah pusat cluster terinisialisasi, data yang tersedia akan ditugaskan ke dalam cluster yang paling dekat berdasarkan perhitungan jarak Euclidean. Berikut adalah rumus untuk menentukan jarak.

$$d = \sqrt{[(x2 - x1)2 + (y2 - y1)2 + (z2-z1)2]}$$

Setelah menghitung jarak dari setiap titik data ke semua pusat cluster kita kemudian menetapkan titik data tersebut ke cluster dengan pusat yang memiliki jarak minimum. Jarak minimum itulah yang akan dijadikan cluster. Ketika semua sudah di tentukan jarak mana yang paling minimum dari centroid 1, centroid 2 dan centroid 3 maka perhitungan iterasi 1 sudah selesai.

Setelah setiap data terklasifikasi, pusat setiap cluster dihitung ulang dengan mencari rata-rata posisi data dalam cluster tersebut. Rata-rata ini akan menjadi pusat cluster (centroid) yang baru. Berikut adalah rumus rata-rata dari clustering K-Means.

$$\bar{x} = \frac{1}{n} \sum_{i=1} n\, xi$$

Setelah menentukan perhitungan rata-rata maka hasil dari perhihtungan tersebut adalah hasil centroid baru dari centroid 1, centroid 2 dan centroid 3.

Memeriksa perhitungan datanya konvergen atau tidak, jika tidak ulangi dari langkah assign data ke cluster terdekat seperti iterasi 1 tadi, jika konvergen maka untuk centroid berikutnya tidak akan ada lagi perubahan data. Dalam penelitian ini untuk centroid nya berhenti berubah di iterasi 2 setelah menghitung rata-rata setiap cluster.

Dalam penelitian clustering mobil bekas ini penelitian menentukan hasil-hasil cluster, langkah hasil akhir clustering dalam proses K-Means clustering adalah tahap dimana kita menganalisis dan memahami output akhir setelah algoritma mencapai konvergensi. Pembagian data ke cluster setiap titik data dalam dataset telah ditetapkan ke salah satu K cluster berdasarkan kedekatan mereka dengan pusat cluster yang dihitung. Ini berati bahwa kita memiliki label

cluster untuk setiap titik data yang menunjukan cluster mana titik tersebut dimiliki.

3. Hasil dan Pembahasan

Dalam menjalankan penelitian ini, penulis mengumpulkan dan mempersiapkan data untuk dijalankan di model clustering. Data yang penulis muat atau upload adalah data yang telah penulis kodekan pada proses sebelumnya.

1. Hasil Penelitian

Yang pertama yaitu peneliti akan menentukan centroid secara acak, berikut tabel pusat awal cluster secara acak. Setelah itu peneliti akan menentukan jumlah cluster, pada penelitian ini jumlah cluster yang digunakan ialah 3 cluster.

Tabel 3.1
Menentukan Centroid Awal Secara Acak

Centroid Awal	Model	Harga	Jarak Tempuh	Tahun
Data 17	Toyota Yaris S	147000000	110000	2015
Data 10	Toyota Innova	200000000	12000	2012
Data 3	Toyota Innova V	270000000	500000	2017

Jarak euclidean merupakan jarak suatu benda ke pusat yang harus dihitung pada saat pengujian menggunakan pemodelan yang telah ditetapkan. Berikut adalah perhitungan eulidean distance.

```
Data point 1:
  Jarak ke centroid 1: 275300012.3619
  Jarak ke centroid 2: 90000001.3347
  Jarak ke centroid 3: 80154387.0678
Data point 2:
  Jarak ke centroid 1: 385300010.5113
  Jarak ke centroid 2: 200000000.1600
  Jarak ke centroid 3: 30410531.0707
Data point 3:
  Jarak ke centroid 1: 355333648.9836
  Jarak ke centroid 2: 170073161.1513
  Jarak ke centroid 3: 0.0000
Data point 4:
  Jarak ke centroid 1: 121300030.4233
  Jarak ke centroid 2: 64000001.1417
  Jarak ke centroid 3: 234052899.3418
```

Gambar 3.1
Menghitung Jarak ke Titik Centroid

Setelah menghitung jarak antara data ke masing-masing centroid penelitian akan menentukan jarak mana yang paling minimum dari 20 baris atas dari masing-masing centroid dan menentukan cluster.

	Distance_to_Nearest_Centroid	Cluster
0	8.015439e+07	3
1	3.041053e+07	3
2	0.000000e+00	3
3	6.400000e+07	2
4	7.400002e+03	2
5	8.000000e+07	2
6	8.300002e+07	2
7	2.700035e+07	2
8	2.000000e+07	2
9	0.000000e+00	2
10	7.500000e+07	2
11	1.370000e+05	2
12	7.000033e+07	2
13	9.100039e+07	2
14	9.030004e+07	1
15	0.000000e+00	1
16	9.030000e+07	1
17	4.300014e+07	2
18	7.200004e+07	2
19	7.100003e+07	2

Gambar 3.2
Data 20 Baris Atas

Langkah berikutnya setelah menghitung jarak terdekat penelitian perlu menentukan cluster dari jarak yang paling minimum. Berikut adalah tabel 10 baris atas untuk menentukan cluster dari jarak yang paling minimum.

Tabel 3.2
Menentukan Cluster

Centroid 1	Centroid 2	Centroid 3	Jarak terpendek	Menetukan
275300012.4	90000001.33	80154387.07	80154387.07	C3
385300010.5	200000000.2	30410531.07	30410531.07	C3
355333649	170073161.2	0	0	C3
121300030.4	64000001.14	23405289.3	64000001.14	C2
185300022.1	7400.001689	17007294.3	7400.001689	C2
26530012	80000002.04	90137119.29	80000002.04	C2
102300008.6	83000018.89	25304806.78	83000018.89	C2
158300005.1	27000352.66	19705969 2.7	27000352.66	C2
165300024.1	20000001.94	19006523 1.8	20000001.94	C2
185300025.9	0	17007316 1.2	0	C2
110300032.8	75000001.13	24505056.3	75000001.13	C2
185300004.1	137000.0001	17006919 8.3	137000.0001	C2
115300059.4	70000330.18	24004745 6.8	70000330.18	C2
943001533.23	91000394.64	26104267 5.4	91000394.64	C2
903000377.23	95000001.35	26504663 8.9	90300037.23	C1
0	185300025.9	35533364 9	0	C1
903000000.01	95000049.52	26504513 1.8	90300000.01	C1
142300000.4	43000135.63	21305589 5	43000135.63	C2
113300002.6	72000037.72	24204989 2.3	72000037.72	C2
113300005.3	72000027.95	24205010 0.8	72000027.95	C2

Untuk proses selanjutnya yaitu pengujian pada perhituungan ulang pusat cluster dengan pemodelan yang sudah di tetapkan sebelumnya, maka pada tahapan ini penelitian akan menghitung rata-rata untuk membuat centroid baru. Proses perhitungan jarak untuk centroid yang baru ini sama seperti perhitungan jarak sebelumnya hanya saja kita akan menghitung dengan centroid yang baru.

```
Data point 1:
  Jarak ke centroid 1: 215100006.9894
  Jarak ke centroid 2: 128357159.0097
  Jarak ke centroid 3: 63354953.5421
Data point 2:
  Jarak ke centroid 1: 325100005.9758
  Jarak ke centroid 2: 238357153.6992
  Jarak ke centroid 3: 46696270.5462
Data point 3:
  Jarak ke centroid 1: 295140972.1564
  Jarak ke centroid 2: 208414943.0711
  Jarak ke centroid 3: 16993633.6974
Data point 4:
  Jarak ke centroid 1: 61100027.7611
  Jarak ke centroid 2: 25642946.7849
  Jarak ke centroid 3: 217339660.6782
Data point 5:
  Jarak ke centroid 1: 125100015.8298
  Jarak ke centroid 2: 38357211.3608
  Jarak ke centroid 3: 153342352.3089
```

Gambar 3.3
Menghitung Jarak Titik Data ke Centroid Baru

Setelah menghitung jarak antara data ke masing-masing centroid yang baru penelitian akan tampilkan data untuk menentukan jarak mana yang paling minimum dari 20 baris atas dari masing-masing centroid yang baru pada iterasi ke 2.

	Distance_to_Nearest_Centroid	Cluster
0	6.335495e+07	3
1	4.669627e+07	3
2	1.699363e+07	3
3	2.564295e+07	2
4	3.835721e+07	2
5	7.335195e+07	3
6	4.210000e+07	1
7	1.135729e+07	2
8	1.835728e+07	2
9	3.835723e+07	2
10	3.664292e+07	2
11	3.835719e+07	2
12	3.164315e+07	2
13	3.410057e+07	1
14	3.010005e+07	1
15	6.020001e+07	1
16	3.010001e+07	1
17	4.642942e+06	2
18	3.364286e+07	2
19	3.264286e+07	2

Gambar 3.4
Data 20 Baris Atas Centroid Baru

Pada tahap iterasi ke-2, proses klasterisasi menunjukkan bahwa algoritma masih belum mencapai kondisi konvergen, yang ditandai oleh perubahan signifikan pada posisi cluster dibandingkan dengan hasil iterasi sebelumnya. kasus ini mengindentifikasikan bahwa data dalam ruang belum terganti secara optimal oleh pusat cluster yang telah dihitung.

Oleh karena itu, dibutuhkan langkah perhitungan ulang terhadap posisi pusat cluster berdasarkan rata-rata pada data di dalam setiap cluster yang terbentuk pada iterasi saat ini atau pada iterasi 3. Proses ini bertujuan untuk memperbaiki posisi pusat cluster sehingga mendapatkan hasil yang lebih mendekati pola distribusi data yang sesungguhnya. Perhitungan ulang ini akan dilakukan sampai perubahan posisi pusat cluster menjadi sangat kecil atau berada dalam ambang toleransi yang telah ditentukan, sehingga algoritma dapat mencapai kondisi konvergen dengan hasil klasterisasi yang stabil.

```
Rata-rata dari nilai harga:
90140000.0

Rata-rata dari nilai jarak_tempuh:
119000.0

Rata-rata dari nilai tahun:
2014.2
```

Gambar 3.5
Rata-rata Centroid 1 Iterasi 2

Setelah kita melakukan perhitungan untuk menentukan jarak setiap titik data ke masing-masing centroid yang baru, proses perhitungan jarak untuk centroid yang baru ini sama seperti perhitungan jarak sebelumnya hanya saja kita akan menghitung dengan centroid yang baru,

```
Data point 1:
  Jarak ke centroid 1: 199860020.9458
  Jarak ke centroid 2: 130272738.9195
  Jarak ke centroid 3: 45017133.3440
Data point 2:
  Jarak ke centroid 1: 309860015.8152
  Jarak ke centroid 2: 240272735.4237
  Jarak ke centroid 3: 65012006.4036
Data point 3:
  Jarak ke centroid 1: 279902561.1905
  Jarak ke centroid 2: 210330218.4737
  Jarak ke centroid 3: 35198259.1507
```

Gambar 3.6
Menghitung Jarak Titik Data ke Centroid Baru

Setelah menghitung jarak antara data ke masing-masing centroid yang baru penelitian akan tampilkan kodingan data untuk menentukan jarak mana yang paling minimum dari 20 baris atas dari masing-masing centroid yang baru pada iterasi ke 2.

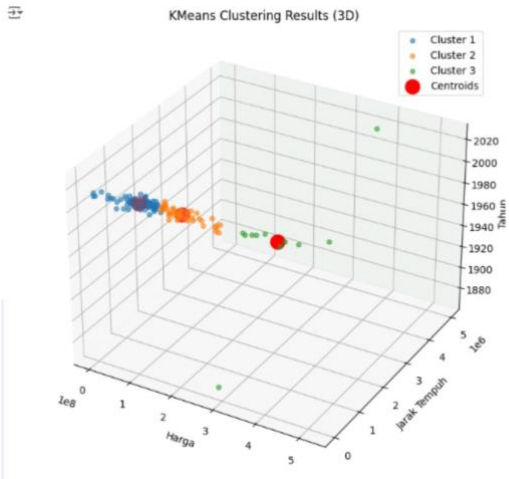
	Distance_to_Nearest_Centroid	Cluster
0	4.501713e+07	3
1	6.501201e+07	3
2	3.519826e+07	3
3	2.372734e+07	2
4	4.027278e+07	2
5	5.501396e+07	3
6	2.686005e+07	1
7	1.327290e+07	2
8	2.027282e+07	2
9	4.027279e+07	2
10	3.472732e+07	2
11	4.027278e+07	2
12	2.972762e+07	2
13	1.886069e+07	1
14	1.486028e+07	1
15	7.544000e+07	1
16	1.486000e+07	1
17	2.727529e+06	2
18	3.172727e+07	2
19	3.072727e+07	2

Gambar 3.7
Data 20 Baris Atas Centroid Baru

Setelah semua data dikelompokkan ke dalam cluster, langkah berikutnya ialah menghitung ulang pusat cluster yang baru menggunakan perhitungan rata-rata nilai data dalam setiap cluster. Jika centroid baru yang dihasilkan konvergen dengan centroid sebelumnya, maka iterasi akan dihentikan. Namun, jika belum maka lanjutkan iterasi selanjutnya sampai konvergensi. Penghentian iterasi terjadi dikarenakan centroid baru telah sama dengan centroid lama yang menunjukkan bahwa data yang berada dalam cluster sudah stabil, sehingga tidak perlu lagi menghitung ulang jarak data pada centroid.

Pada iterasi ketiga dari proses K-Means clustering dengan menggunakan tiga variabel telah mencapai konvergensi sepenuhnya. Pada iterasi ini, centroid dari setiap cluster diperbarui berdasarkan rata-rata posisi data dalam cluster yang diperoleh pada iterasi kedua. Setelah pembaruan centroid, data dalam setiap cluster diperiksa kembali untuk memastikan apakah ada data yang perlu dipindahkan ke cluster lain. Tetapi, dalam iterasi ketiga ini, tidak ada lagi data yang berpindah cluster. Berarti ini menunjukkan bahwa algritma telah mencapai konvergensi, dikarenakan ttidak ada lagi perubahan dalam pembagian data terhadap cluster yang terjadi. Oleh karena itu proses iterasi berhenti pada iterasi ke tiga, dikarenakan setiap data sudah berada dalam cluster dengan centroid terdekat telah sesuai.

2. Hasil Pengujian



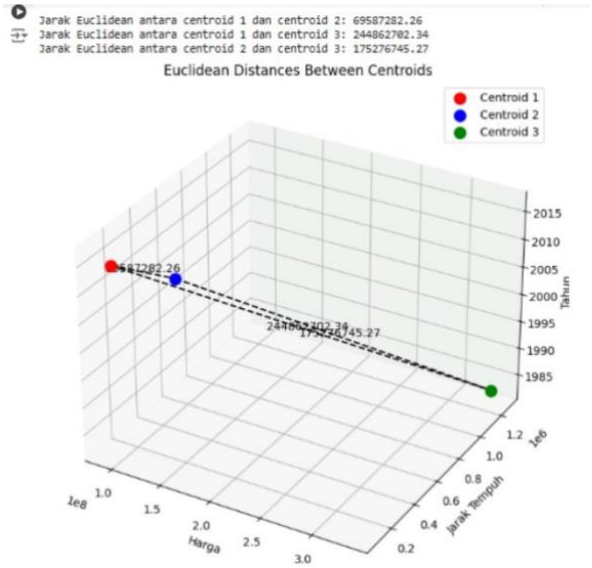
Gambar 3.8
Visualisasi Hasil K-Means Clustering Mobil Bekas

Dari 159 titik data yang digunakan didalam dataset mobil bekas dan pengelompokan tiga cluster dengan penggunaan variable harga, jarak_tempuh dan tahun yang ada selama pengujian dengan menggunakan google colab. Berikut tabel jumlah data yang terdapat dalam masing-masing cluster.

Tabel 3.3
Cluster Model

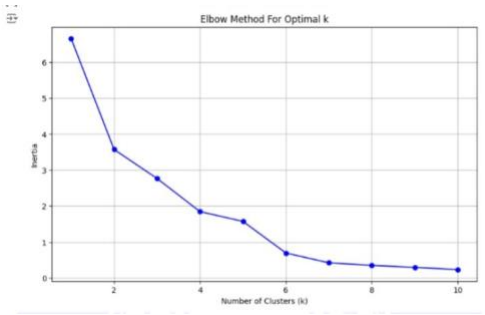
Cluster Model	
Cluster 1	100
Cluster 2	48
Cluster 3	11
Total Semua Data	159

Setelah hasil clustering mobil bekas penelitian akan menampilkan hasil dari perhitungan jarak antara masing-masing centroid untuk menggambarkan tingkat pemisahan antar masing-masing cluster, dan juga digunakan sebagai indikator bahwa setiap cluster memiliki posisi yang berbeda secara signifikan, sehingga pengelompokan dapat dinilai efektif dan akurat.



Gambar 3.9
Visualisasi Jarak Antar Centroid

Hasil dari clustering mobil bekas dengan 159 data yang memiliki 3 cluster ini sudah bisa di tentukan mana saja mobil yang dibagi berdasarkan harga, jarak_tempuh dan tahun. Berikut hasil dan penjelasan tentang masing-masing cluster. Demikian juga metode Elbow untuk menentukan hasil cluster mana yang signifikan, berikut tampilan visualisasi dari metode Elbow Clustering K-Means mobil bekas.



Gambar 3.10
Visualisasi Hasil Elbow Clustering K-Means Mobil Bekas

Berdasarkan hasil visualisasi diatas bahwa lekukan atau siku-siku bisa menentukan di titik yang ke-3, dikarenakan peneliti bertujuan untuk mengelompokan mobil bekas berdasarkan tiga kategori atau cluster. Yaitu murah , sedang dan mahal.

4. Kesimpulan

Dalam pengujian algoritma K-Means clustering menunjukkan bahwa penggunaan algoritma K-Means clustering efektif di dalam menganalisis dan mengelompokkan mobil bekas. Algoritma ini telah berhasil mengelompokkan data harga menjadi beberapa kategori yang jelas, seperti kelompok harga rendah, menengah, dan tinggi, memungkinkan identifikasi karakteristik spesifik dari setiap kelompok.

Hasil clustering ini memudahkan pemberian rekomendasi yang lebih tepat sasaran kepada pengguna, sehingga mereka dapat memilih mobil bekas yang sesuai dengan anggaran dan karateristik mereka. Selain itu, implementasi K-Means Clustering juga dapat mendukung sebagai pengambilan keputusan yang lebih berbasis analisis data. Secara keseluruhan, penggunaan algoritma ini dalam pemilihan dan rekomendasi mobil bekas terbukti dapat memberikan manfaat yang signifikan dalam analisis dan rekomendasi mobil bekas yang lebih efektif dan efisien.

Dari kesimpulan diatas dapat disimpulkan bahwa harga mobil bekas memiliki hubungan yang erat dengan kualitas dan kondisi mobil. Harga mahal mencerminkan keadaan mobil dengan kualitas terbaik, ataupun menandakan sebagai kendaraan yang klasik yang memiliki tahun yang cukup tua tetapi memiliki nilai harga jual yang mahal, sedangkan harga menengah menawarkan keseimbangan antara kualitas mobil dan biaya. Sementara itu, harga murah menjadi pilihan utama bagi konsumen yang fokus pada keterjangkauan meskipun dengan konsekuensi kendaraan yang lebih tua dan potensi kebutuhan perawatan yang lebih besar. Oleh karena itu, rekomendasi mobil bekas yang sesuai sangat dipengaruhi oleh anggaran, kebutuhan, dan preferensi konsumen terhadap kualitas kendaraan.

Referensi

Adinda, P. R., & Komputer, T. (n.d.). MENSORTIR JENIS DAUN MINT MENGGUNAKAN EUCLIDEAN DISTANCE DAN K-MEANS CLUSTERING DENGAN EKSTRAKSI CIRI BENTUK DAN TEKSTUR. In Portaldata.org (Vol. 2, Issue 9).

Analisis Sistem Informasi Akuntansi Siklus Pendapatan Menggunakan DFD Dan Flowchart Pada Bisnis Porobico. (n.d.).

Arif, M., & Faisal, M. (2023). Penerapan Model Regresi Linear Untuk Estimasi Mobil Bekas Menggunakan Bahasa Python. Euler : Jurnal Ilmiah Matematika, Sains Dan Teknologi, 11(2), 182–191. https://doi.org/10.37905/euler.v11i2.20698

Cahyadi Iman, Nurhayati, & Fitri Iskandar. (2020). Aplikasi Pengaduan dan Maintenance Elektronik Berbasis Web dengan Metode First in First Out (FIFO) di Universitas Nasional. Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi). http://journal.lembagakita.org/index.php/jtik

Darmawan Sidik, A., Ansawarman, A., Kunci, K., Kendaraan Bermotor, J., Regresi, M., & Jalan, F. (2022). Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning. Formosa Journal of Multidisciplinary Research (FJMR), 1(3), 559–568. https://doi.org/10.55927

Farisi, A., & Fatah, Z. (2024). Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Menggunakan Metode K-Means Clustering. Volume 2, 205–212. https://doi.org/10.59435/gjmi.v2i11.1059

Fauziah, R., & Purnamasari, A. I. (2023). Implementasi Algoritma K-Means pada Kasus Kekerasan Anak dan Perempuan Berdasarkan Usia. Hello World Jurnal Ilmu Komputer, 2(1), 34–41. https://doi.org/10.56211/helloworld.v2i1.232

Gelar Guntara, R. (2023). Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7. Jurnal Teknologi Dan Sistem Informasi Bisnis, 5(1), 55–60. https://doi.org/10.47233/jteksis.v5i1.750

Halimatussya'diyah Purba, & Yahfizham. (2023). Konsep Dasar Pemahaman Algoritma Pemrograman. Jurnal Arjuna, Volume 1. https://doi.org/10.61132/arjuna.v1i6

Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. Jurnal Teknik Industri, Vol. 10.

Puspita, R. N., Kependudukan, D., Sipil, P., & Tangerang, K. (n.d.). PERBANDINGAN METODE CENTROID DAN WARD DALAM PENGELOMPOKKAN TINGKAT PENYELESAIAN PENDIDIKAN DI INDONESIA. 3(3), 2022. https://doi.org/10.46306/lb.v3i3

Rahayu, K., Fitria, V., Septhya, D., Rahmaddeni, R., & Efrizoni, L. (2023). Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning. MALCOM: Indonesian Journal of

Machine Learning and Computer Science, 3(2), 108–114. https://doi.org/10.57152/malcom.v3i2.780

Rahman, S., Sembiring, A., Siregar, D., Khair, H., Prahmana, G., Puspadini, R., & Zen, M. (n.d.). PYTHON : DASAR DAN PEMROGRAMAN BERORIENTASI OBJEK TAHTA MEDIA GROUP.

Rahmat, A., Syafiih, M., & Faid, M. (2023). IMPLEMENTASI KLASIFIKASI POTENSI PENYAKIT JANTUNG DENGAN MENGGUNAKAN METODE C4.5 BERBASIS WEBSITE (STUDI KASUS KAGGLE.COM). INFOTECH Journal, 9(2), 393–400. https://doi.org/10.31949/infotech.v9i2.6295

Syahfitri, N., Budianita, E., Nazir, A., & Afrianty, I. (2023). KLIK: Kajian Ilmiah Informatika dan Komputer Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid. Media Online, 4(3), 1668–1675. https://doi.org/10.30865/klik.v4i3.1525