

Implementasi K-Means Clustering untuk Analisis Non-Numerik Dataset Spare Part Mobil

Implementation of K-Means Clustering for Non-Numerical Analysis of Car Spare Parts Dataset

Mailia Putri Utami¹, Gita Mustika Rahma², Finna Suroso³
^{1,2,3}Sistem Informasi Industri Otomotif, Politeknik STMI Jakarta

E-mail: ¹Mailiap2206@gmail.com, ²g.m.rahmah@gmail.com, ³finna.suroso@gmail.com

Abstrak

Industri otomotif merupakan salah satu sector ekonomik terbesar di dunia, dengan berbagai rantai pasok yang kompleks. Kompleksitas data otomotif yang beragai sering kali banyak mengandung atribut data non-numerik, seperti nama, jenis spare part, merk, dan atribut kualitatif lainnya. Analisis non-numerik dapat memberikan wawasan berharga dengan pola dan hubungan antar suku cadang. Tujuan dari penelitian ini yaitu untuk mengembangkan metode k-means clustering pada dataset spare part yang mengandung atribut non-numerik. K-means clustering adalah Teknik yang umumnya digunakan untuk analisis data numerik dan memerlukan modifikasi atau keterlibatan metode lain dalam mengatasi data non-numerik. Adapun proses yang dilibatkan yaitu proses normalisasi dengan menggunakan metode binning. Implementasi K-means Clustering pada dataset non-numerik memiliki manfaat potensial. Pertama, itu memungkinkan identifikasi kelompok dari suku cadang yang memiliki karakteristik serupa, yang dapat digunakan untuk mengelompokkan produk serupa. Kedua yaitu untuk membantu dalam pengelolaan ketersediaan dengan lebih efisien, menghindari kelebihan persediaan, dan memenuhi permintaan pelanggan dengan lebih baik. Penelitian ini menghadapi tantangan dalam menentukan metrik kesamaan yang tepat untuk data non-numerik dan dalam menentukan jumlah cluster yang optimal. Namun, metodologi yang cermat dan eksperimen yang berkelanjutan, mampu mengembangkan pendekatan yang dapat digunakan dalam pengelompokan suku cadang mobil non-numerik. Hasil dari penelitian ini diperoleh sebaran data dengan menggunakan cluster $K=2$ dengan nilai Silhouette sebagai nilai dari sebaran data yaitu 0,925.

Kata kunci: non-numerik, *K-means clustering*, *Sparepart*.

Abstract

The automotive industry is one of the largest economic sectors in the world, with various complex supply chains. The complexity of various automotive data often contains many non-numerical data attributes, such as name, type of spare part, brand, and other qualitative attributes. Non-numerical analysis can provide valuable insight into patterns and relationships between parts. The aim of this research is to develop a k-means clustering method on spare part datasets that contain non-numerical attributes. K-means clustering is a technique that is generally used for numerical data analysis and requires modification or involvement of other methods in dealing with non-numerical data. The process involved is the normalization process using the binning method. Implementing K-means Clustering on non-numerical datasets has potential benefits. First, it allows the identification of groups of parts that have similar characteristics, which can be used to group similar products. Second, to help manage availability more efficiently, avoid excess inventory, and meet customer demand better. This research faces challenges in determining appropriate similarity metrics for non-numerical data and in determining the optimal number of clusters. However, careful methodology and ongoing experimentation have been able to develop approaches that can be used for non-numerical grouping of auto parts. The

results of this research obtained data distribution using cluster $K=2$ with a Silhouette value as the value of the data distribution, namely 0.925.

Keywords: Non-numeric, K-means clustering, Spare part

1. PENDAHULUAN

Industri otomotif adalah salah satu industry yang sangat besar dan kompleks, dengan sejumlah besar produsen yang penyedia suku cadang mobil. Dalam industry ini, manajemen rantai pasokan dan pengelolaan investaris suku cadang adalah komponen penting untuk memastikan efisiensi operasional dan kepuasan pelanggan. Dalam hal ini, analisis data dapat membantu perusahaan otomotif dalam pengambilan keputusan yang lebih baik terkait pengelolaan suku cadang mobil [1].

K-means clustering adalah metode clustering yang telah terbukti efektif dalam analisis data numerik. Namun, implementasi K-means clustering pada dataset suku cadang mobil yang mengandung atribut non-numerik memerlukan adaptasi dan pengembangan teknik yang sesuai. Penelitian ini bertujuan untuk menjawab tantangan tersebut dengan mengembangkan pendekatan yang efektif untuk mengelompokkan suku cadang mobil berdasarkan atribut non-numerik mereka [2].

Penelitian terdahulu yang relevan dengan penelitian ini yaitu melakukan penelitian dengan memanfaatkan data mining untuk penjualan suku cadang sepeda motor di PT. Motorindo Perkasa Raya. Pada penelitiannya, data suku cadang terdiri dari beberapa jenis suku cadang yang mudah untuk dikelompokkan sehingga dapat secara cepat menentukan dan menemukan nilai barang dari tiap transaksi jual beli. Data mining sendiri dalam penelitian tersebut dijadikan serangkaian proses untuk menggali nilai tambah berupa informasi yang tidak diketahui secara manual dari suatu basis data [3].

Peneliti terdahulu lainnya dilakukan dengan penerapan metode K-means, namun pada penelitian tersebut dataset yang digunakan sesuai dengan kebutuhan dari prosedur penelitian dengan mengimplementasikan metode *K-means* [4]. Peneliti lainnya dengan implementasi yang sama yaitu dengan mengadopsi proses *K-means clustering* yang diproses dengan membandingkan dengan metode *Ant Colony Optimization*. Dari *Ant Colony Optimization* belum menjadi metode yang cukup optimal dikarenakan center dari perkelompokan hanya berdasarkan satu atribut saja [5]. Peneliti lainnya dengan mengadopsi anomali data berupa data non-numerik, data tersebut diperuntukan untuk deteksi dan digunakannya metode data mining [6].

Sebagai *research gap* yaitu terletak pada pemanfaatan algoritma K-means clustering sebagai metode untuk menganalisis kebutuhan dalam menentukan kebutuhan dalam menganalisa sparepart. Kebaruan penelitian ini yaitu terletak pada implementasi proses pengolahan data dengan menggunakan pemanfaatan implementasi metode binning. Metode dari proses binning sendiri merupakan cara untuk mengubah data non-numerik menjadi data numerik, dikarenakan algoritma K-means clustering hanya dapat berproses dengan menggunakan data numerik. Sehingga dapat meningkatkan proses kerja dari metode yang diterapkan pada penelitian ini [6].

Implementasi K-means Clustering pada dataset suku cadang mobil non-numerik akan memberikan manfaat yang signifikan, termasuk [4]:

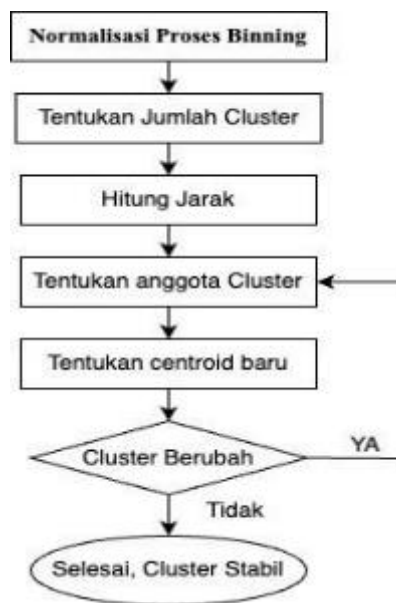
- 1) Identifikasi segmentasi produk: dengan melakukan clustering perusahaan otomotif dapat mengidentifikasi kelompok suku cadang yang memiliki karakteristik serupa. Ini dapat membantu dalam merancang strategi manajemen rantai pasok yang lebih efektif.
- 2) Pengelolaan stok: dengan memahami kelompok suku cadang yang ada, perusahaan

dapat mengoptimalkan persediaan dan menghindari kelebihan persediaan yang tidak perlu.

- 3) Kepuasan pelanggan: dengan memahami preferensi pelanggan, perusahaan dapat lebih efektif memenuhi kebutuhan pelanggan, meningkatkan kepuasan, dan membangun loyalitas pelanggan.
- 4) Peningkatan keuntungan: analisis clustering yang efektif dapat membantu perusahaan mengidentifikasi peluang penghematan biaya dan peningkatan keuntungan.

2. METODE PENELITIAN

Pada tahapan ini diuraikan berdasarkan beberapa tahapan, yang pertama yaitu Tahap Pengambilan Data, Tahap Pengolahan Data, Tahap Pengujian atau Eksperimen. Dan dijabarkan berdasarkan alur dari proses sistem yang dilakukan, tergambar pada Gambar 1.



Gambar 1. Prosedur Penelitian

Sebelum penelitian masuk pada tahap normalisasi, peneliti membutuhkan dataset yang diperlukan. Data yang diperoleh berdasarkan hasil observasi dan wawancara oleh pihak bengkel mobil. Data tersebut berdasarkan periode waktu transaksi penjualan spare part dalam kurung waktu dari bulan September 2021 s/d September 2023. Data tersebut digunakan untuk dijadikan dataset untuk proses clustering, digunakan untuk menjadi pondasi awal pada penelitian ini, yang mana sesuai dengan prosedur penelitian data tersebut diperlukan normalisasi data karena data yang diperoleh tidak semuanya merupakan data numerik [7].

Proses normalisasi yang digunakan yaitu dengan proses binning, yaitu proses untuk mentransformasikan atau mengubah nilai dari data non-numerik menjadi data numerik atau data yang dapat biner atau data yang dapat dikalkulasikan [8]. Nilai dari data yang telah dilakukan proses binning digunakan untuk mencari jarak terdekat antar tiap data pada proses k-means clustering, yang mana mencari nilai dari jarak terdekat antar tiap data dari centroid / pusat cluster, atau digunakan juga pada proses melakukan pencarian cluster center dengan cluster center baru pada tiap iterasinya [9].

2.1 Prosedur Pemikiran

Algoritma *K-Means* merupakan algoritma yang dapat diterapkan pada data yang direpresentasikan dalam ruang dimensi. Tujuannya untuk mengelompokan set-data ruang dimensi, yaitu $X = \{x_i | i = 1, \dots, N\}$, dimana $x_i \in R^d$ yang menyatakan data ke- i , sebagai “titik data”. *K-means* mempartisi X ke dalam K *Cluster*, algoritma *K-Means* mengelompokan semua titik data dalam X sehingga setiap titik x_i hanya jatuh dalam satu dari K partisi. Titik di cluster, dilakukan dengan cara memberikan setiap titik sebuah ID *cluster*. Titik dengan ID cluster yang sama berarti berada dalam satu cluster yang sama, sedangkan titik dengan ID cluster yang berbeda pada cluster yang berbeda. Vektor keanggotaan cluster m dengan panjang N , m_i , bernilai ID cluster titik x_i [10].

Setiap cluster dari *centroid cluster* diwakili oleh titik tunggal dalam R^d . Set representative cluster dinyatakan $C = \{c_j | j = 1, \dots, K\}$. Sejumlah K *representative cluster* tersebut disebut dengan *centroid cluster*. Set data dalam X dikelompokan berdasarkan konsep kemiripan/kedekatan [11]. Pada saat data sudah dihitung ketidakmiripan terhadap *centroid*, maka selanjutnya dipilih ketidakmiripan yang paling kecil sebagai *cluster* yang akan diikuti sebagai relokasi pada *cluster* yang disebut juga dengan iterasi. Relokasi data dalam cluster yang diikuti dapat dinyatakan dengan nilai keanggotaan α yang bernilai 1 atau 0. 1 jika data tersebut sebagai anggota *cluster* dan 0 tidak menjadi anggota cluster dan nilai α sebuah data pada semua cluster, hanya akan bernilai 1, sedangkan yang lainnya 0 [12].

$$\alpha_{ij} = \begin{cases} 1 & \text{arg arg } \{d(x_{ic_j})\} \\ 0 & \text{ } \end{cases} \quad (1)$$

Dimana $d(x_{ic_j})$ merupakan ketidakmiripan (jarak) dari data ke- i ke cluster c_j . Berikut ini formula perumusan untuk mencari rata-rata setiap fitur dari semua data dalam semua cluster [13].

$$C_j = \frac{1}{N} \sum_{i=1}^N x_j \quad (2)$$

Berikut ini adalah alur program dasar dari algoritma *K-Means* [5]:

1. Inisialisasi, dimana pertama kali yaitu menentukan nilai K sebagai jumlah cluster yang diinginkan dan metrik ketidakmiripan (jarak). Tetapkan ambang batas sebagai perubah fungsi objektif dan ambang batas perubah posisi pusat (*centroid*).
2. Pilih K data dari set data C sebagai pusat (*centroid*).
3. Alokasi semua data ke pusat (*centroid*) terdekat dengan metrik jarak yang sudah ditetapkan.
4. Hitung Kembali nilai pusat (*centroid*) C berdasarkan data yang mengikuti cluster dari masing-masing.
5. Mengulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu perubahan fungsi objektif sudah dibawah ambang batas, tidak ada data yang berpindah cluster, dan perubahan posisi *centroid* sudah di bawah ambang batas yang ditetapkan.

Pemilihan K sebagai titik data sebagai pusat (*centroid*) awal juga memengaruhi hasil clustering. Sifat ini menjadi salah satu ciri khas atau karakteristik alami *K-Means* yang dapat mengakibatkan hasil *cluster* yang dilakukan memiliki nilai atau hasil yang berbeda. Kondisi ini dikenal juga dengan solusi yang local optima, artinya algoritma *K-means* sangat sensitid terhadap lokasi awal *centroid*. Penyelesaian masalah local optima dapat diselesaikan dengan menjalankan algoritma beberapa kali dengan inisial *centroid* yang berbeda kemudian memilih hasil yang terbaik [14].

3. HASIL DAN PEMBAHASAN

Pembahasan hasil yang dicapai dapat diuraikan Tahap Pengolahan Data, Tahap Pengujian atau Eksperimen, Analisis Awal, dan Penjabaran terkait yang telah dilakukan dalam penelitian.

3.1 Pengolahan Data

Pada proses pengolahan data, yang pertama kali dilakukan dengan melakukan Teknik observasi untuk memperoleh data secara periode, proses memperoleh data dengan adanya pendekatan secara observasi dan wawancara kepada pihak bengkel. Dimana sebagai peneliti wajib mengetahui keapsahan dari data yang diperoleh, adapun proses memperoleh data yaitu dengan kurang waktu 1 bulan penuh pada proses kerja magang yang dilakukan anggota peneliti.

Adapun data transaksi penjualan spare part tersebut didapat dalam kurung waktu dari bulan September 2021 s/d September 2023. Data tersebut digunakan untuk dijadikan dataset untuk proses clustering. Adapun berikut ini data yang diperoleh.

Nama Sparepart	Banyaknya barang yang dipesan (barang)	Tersedia	Harga	Kode Sparepart
SERVICE CHILLER ANGES 15 HP	1	10	1	100.000.201.1001.4
SERVICE BLOWER 3 PHASE 380V	1	10	1	100.000.819.0000038
SKUN Y 1,5MM	1	10	1	100.000.2005000038
SKUN KABEL 3 MM	1	10	1	100.000.21050001
SELANG PU 12MM X 8MM CLEAR	1	10	1	100.000.2
RODA TROLLY 6" (HIDUP)	1	10	1	100.000.2
STOP KONTAK + SAKLAR	1	10	1	100.000.4
STOP KONTAK + SAKLAR	1	10	1	100.000.4
KABEL SERABUT 2 X 0.5 MM	1	10	1	100.000.2103001.5
OVERLOAD FUSI TR - 13A 7 - 11A	1	10	1	100.000.21030041
BEARING 6307 SKF	1	10	1	100.000.38
BESI BEHEL 6 MM	1	10	1	100.000.21030032
MEKANIKSEAL POMPA 25HP	1	10	1	100.000.8
PROXIMITY AUTONICS PRW12-4CP 12-24 VDC	1	10	1	100.000.2009001.1
WD 40	1	10	1	100.000.1
T-SLOT ENDMILL ATT06D-05 D06X5MM	1	10	1	100.000.21030043
PROXIMITY AUTONICS PRW12-4CP 12-24 VDC	1	10	1	100.000.9001.1
PROXIMITY AUTONICS PRW12-4CP 12-24 VDC	1	10	1	100.000.2009001.1
HEATER 30X30, 220V 200W	1	10	1	100.000.20100030
HEATER 35X35, 220V 150W	1	10	1	100.000.77
HEATER 40X40, 280V 250W	1	10	1	100.000.20100036
KAWAT BENDIPAT	1	10	1	100.000.200300035
KAWAT LAS ARGON STAINLESS STEEL, TECH 318 DIA 2 MM	1	10	1	100.000.200900383
MATA GERINDA WD BLACK 4 INCH	1	10	1	100.000.20070037
ELBOW GALVANIS 1 1/4 INCH	1	10	1	100.000.200300391
ELBOW SOCK GALVANIS 1 1/2 INCH	1	10	1	100.000.20070037
VLOK SOCK GALVANIS 1 1/2 X 1 INCH	1	10	1	100.000.20070036
VLOK SOCK GALVANIS 1 1/4 X 1 INCH	2	10	1	100.000.200300222
VLOK SOCK GALVANIS 1 1/2" KE 1 1/4"	2	10	1	100.000.21030034
NEAPLE FITTING 12X3/8	2	10	1	100.000.5
NEAPLE FITTING 12X1/4	2	10	1	100.000.5
NEAPLE FITTING 12X3/8	2	10	1	100.000.5
BEARING 6307 SKF	2	10	1	100.000.38
DOP GALVANIS DRAT DALAM 1 1/4"	2	10	1	100.000.21030030

Gambar 2. Dataset transaksi sparepart 2021

Nama Sparepart	Banyaknya barang yang dibeli	Ketersediaan barang	Trankasi	Harga	Kode Sparepart
WAN BELT XPA-1432	1	10	1	85000	1
SELANG PU 12MM X 8MM CLEAR	1	10	1	50000	2
FUSE KERAMIK 32A	1	10	1	210000	10
BATU CERAMIK XERPEC #220 BROWN 1X6X100 A-C-100RM	3	10	1	145000	7
REMOTE HOIST WIRELESS TELECONTROL F24-6D.6 TOMBOL - E/S	2	10	1	112000	21080013
SEAL UHS 80X90 NOK	2	10	1	100000	19
SEAL DHS NOK 80 X 90	1	10	1	200000	21120030
HEATER 140X120,220V,750W	4	10	1	132000	31
SKUN KABEL 25 MM, BAUT 8 MM	2	10	1	100000	20110367
HEAT DRYER BOSCH GHG 18-60	3	10	1	130000	21120003
KUNCI INGGRIS 12" TEKIRO	1	10	1	120000	4
TANG KOMBINASI 7"	3	10	1	200000	1
TANG POTONG (KABEL)	2	10	1	100000	200300002
TANG SNAPRING TEKIRO 7 INCH	4	10	1	120000	20110024
KUNCI L SET MILI TEKIRO 1,5MM - 10 MM	1	10	1	125000	819000007
CENTER DRILL 8 X 3 MM	3	10	1	100000	20110019
SHRINK TUBE KABEL 6MM X 100MM	2	10	1	150000	21070003
HIGH PRESSURE BUSHING BU-PT-20AX15A 3/4 X 1/2 12MPa	3	10	1	110000	21110016
NEAPLE FITTING 12X1/4	1	10	1	135000	3
SELONGSONG KABEL HEATER 8MM	3	10	1	125000	8
HEATER 220X80, 220V 1800W	2	10	1	112000	88
HEATER 35X100 220 V 400 W	2	10	1	100000	75
HEATER 210X80 220V	1	10	1	100000	21120029
HEATER 60X100 220V	1	10	1	100000	21120028
HEATER 60X100, 380V 840W	2	10	1	100000	20100047
HEATER 120X55 380V 500W	3	10	1	210000	22
TWINFLEX TOZEN 3 INCH JIS10K KARET FLEXIBEL	3	10	1	100000	21090013
PROXIMITY AUTONICS PRW12-4DP 12-24 VDC	1	10	1	120000	20080011
PROXIMITY AUTONICS PRW12-4DP 12-24 VDC	3	10	1	210000	20080011
CORONG PVC 4"	4	10	1	100000	200300001
BALL VALVE PVC 3/4 INCH	2	10	1	150000	22010004
KLEM KABEL SUPER MAS 6 MM	2	10	1	120000	21020043

Gambar 3. Dataset transaksi sparepart 2022

Nama Sparepart	Banyaknya barang yang dibeli	Ketersediaan barang	Trankasi	Harga	Kode Sparepart
AIR WATER OIL FILTER SEPARATOR MOLLAR ARF2000	1	10	1	100000	22110038
ACB SCHNEIDER NS1000N 3P 1000A 50kV MOTORIZED	1	10	1	125000	23010006
ORING VITON 8,8 X 1,9	1	10	1	115000	22040012
CHECK VALVE SCREW 1/4" 3.0MM	3	10	1	125000	22120017
HEAD SCREW MIRROR DIA 30 + CHUCKRING	2	10	1	100000	22120012
BAUT JOINT TRAY MB X 20 + MUR MB	2	10	1	150000	23010019
PLUG DOP PVC 1/2 INCH	1	10	1	110000	21050017
SELANG PU 12MM X 8MM CLEAR	4	10	1	100000	2
SEAL UHS 60 NOK	2	10	1	175000	15
SEAL DHS 60 NOK	3	10	1	100000	16
BAUT 5X10	1	10	1	200000	116
BESI SIKU 50 X 50 CM	3	10	1	150000	1219000038
HAND TAP MS X O,8 YAMAWA SPIRAL	2	10	1	100000	5
BAUT + MUR 38MM X 150MM	4	10	1	115000	22120001
BAUT MUR M24	1	10	1	100000	13
BAUT 6X10	3	10	1	175000	130
HAND TAP SPIRAL M6X1 YAMAWA	2	10	1	100000	17
NEAPLE HIDROLIK POLDS 1/2X3/8	3	10	1	115000	9
STEKER LEGRAND MALE FORT 32 A	1	10	1	100000	20070014
PISAJ GERINDA BETON 4 INCH	3	10	1	110000	22120008
BAUT 10X50	2	10	1	120000	16
BESI AS VCN 95 X 96 MM	2	10	1	100000	21080001
FLANGE PVC 2"	1	10	1	145000	21100042
BALL VALVE KUNINGAN 3/8 INCH	1	10	1	125000	22010034
DYNABOLT 12X50	2	10	1	110000	3
NEAPLE HIDROLIK POLDS 3/8X3/8	3	10	1	100000	11
SELANG GAS(2 MTR)	3	10	1	150000	10
BAUT 8X10	1	10	1	125000	146
HAND TAP SPIRAL M8X1,25 YAMAWA	3	10	1	135000	18
MATA BOR 4,5 mm	4	10	1	100000	6
MATA BOR 5 mm	2	10	1	115000	7
MATA BOR 6 mm	2	10	1	110000	9
BAUT + MUR 10X100	2	10	1	150000	23130000

Gambar 4. Dataset transaksi sparepart 2023

3.2 Proses Menghitung Jarak

Proses menghitung jarak digunakan untuk menentukan jarak terdekat antara sebuah data dengan suatu *cluster center*. Proses ini dilakukan dengan menggunakan data hasil proses *binning*. Data-data tersebut dihitung jaraknya terhadap setiap *cluster* yang ada dengan cara membandingkan nilai dari masing- masing atribut pada masing-masing data. Misalkan terdapat sebuah data mempunyai 5 atribut yaitu atribut banyak barang yang dibeli, ketersediaan barang,

transaksi, harga dan kode barang. Dengan nilai (0,0100,1000,001,010) dan sebuah *cluster center* dengan atribut (1,0010,0100,001,010). Maka jarak antar keduanya adalah :

$$\text{Jarak} = ((0 \text{ XOR } 1) + (0100 \text{ XOR } 0010))/4 + 1/4 (1000 \text{ XOR } 0100) + 1/3 (001 \text{ XOR } 001) + 1/3 (010 \text{ XOR } 010))$$

$$\text{Jarak mean} = (1+2/4+2/4+ 0+0)/5 = 2/5 = 0,4$$

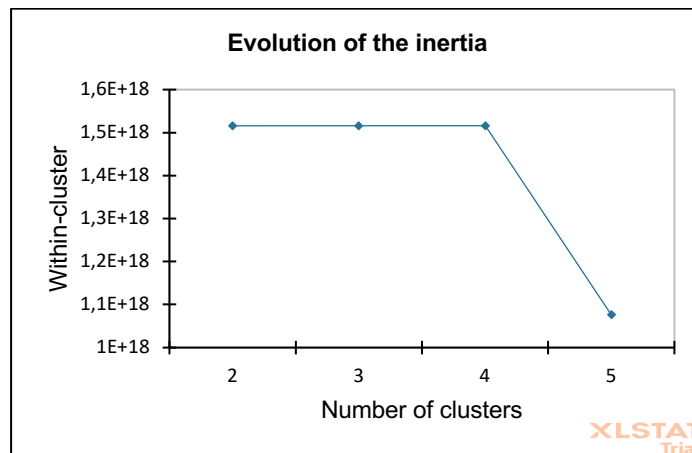
Keterangan: *Hamming Distance* dapat pula diimplementasikan sebagai operasi *bitwise XOR* antara dua objek, karena hasil dari operasi antar dua data yang berbeda yang akan bernilai 1, dan antar dua data yang sama akan bernilai 0.

3.3 Proses Clustering

Inti dari proses *clustering* ada di sini. Apabila telah diketahui jarak-jarak antara setiap data dengan setiap *cluster center*, maka data- data selanjutnya dapat dikelompok- kelompokkan berdasarkan jaraknya terhadap suatu *cluster center*. Suatu data akan dijadikan anggota suatu *cluster* yang mempunyai jarak paling minimum dengan data tersebut. Kemudian, setelah terbentuk *cluster-cluster*, *cluster center* akan dicari lagi dengan mencari “nilai tengah” dari *cluster* yang bersangkutan. Nilai tengah ini bisa berupa *mean*, *median*, atau *modus* dari atribut-atribut data yang menjadi anggota *cluster* tersebut. Jika *cluster center* yang baru telah didapatkan, maka akan dibandingkan dengan *cluster* yang lama. Jika berbeda (berubah), proses iterasi berjalan kembali. Namun jika ternyata *cluster center* bernilai sama, maka proses iterasi selesai dan *cluster* dinyatakan sebagai *cluster* yang stabil.

3.4 Hasil Analisis

Setelah dilakukannya implementasi dari metode K-means clustering maka dapat diamati hasilnya dalam membentuk *cluster-cluster*. Dataset yang digunakan yaitu dari rentang waktu September 2021 s/d sepeptember 2023.



Gambar 5. Evolusi inerti

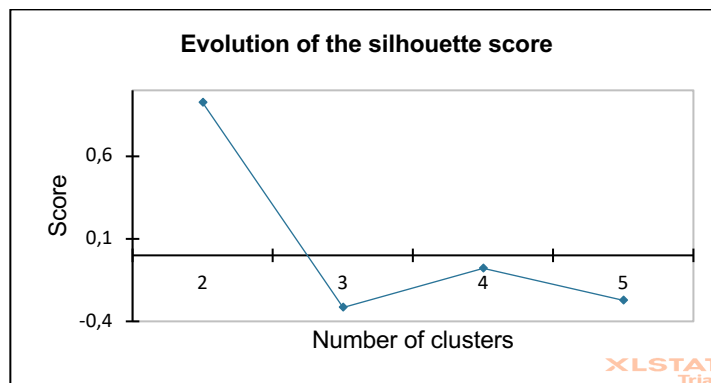
Nilai dari sebaran berdasarkan hasil dari hasil evolusi inersia pada penelitian ini untuk mengukur seberapa baik kumpulan data yang dikelompokkan oleh metode k-means. Hal ini dihitung dengan mengukur jarak antara setiap titik data dan pusat massanya, mengkuadratkan jarak dan menjumlahkan kuadrat-kuadrat dalam satu cluster. Hasil yang baik yaitu dari model yang memiliki nilai inersia yang rendah dan jumlah cluster yang sedikit. Namun, hal ini merupakan trade-off karena seiring K bertambahnya, inersia berkurang. Berdasarkan hasil pengamatan dengan data yang digunakan dari gambar diatas, Untuk mencari K dataset yang optimal, gunakan *metode Elbow* [15]; temukan titik di mana penurunan inersia mulai melambat. K=5 adalah "siku" dari grafik di atas. Oleh karena itu, digunakan uji coba clustering

dengan menggunakan 5 cluster, namun disamping itu diperlukannya uji coba lebih untuk mendapatkan penggunaan jumlah cluster yang tepat dengan data yang digunakan.

Uji coba clustering dengan menggunakan 2 hingga 5 cluster menghasilkan nilai untuk evaluasi dari nilai Silhouette yaitu:

Tabel 1. Nilai Silhouette

Cluster	2	3	4	5
Silhouette scores	0,925	-0,316	-0,079	-0,273



Gambar 6. Nilai Silhouette

Nilai dari Silhouette berguna untuk melihat kualitas dan kekuatan dari cluster yang terbentuk, atau seberapa baik/buruknya suatu obyek ditempatkan dalam suatu cluster. Pada tabel di atas dan gambar di atas menggambarkan jumlah cluster dan nilai Silhouette yang terbentuk pada cluster yang digunakan. Ketika nilai maksimum Silhouette dengan menggunakan k=2 menghasilkan nilai 0,925. Berdasarkan pengamatan ketika jumlah cluster semakin besar maka nilai silhouette yang dihasilkan cenderung lebih kecil disbanding jumlah cluster sebelumnya. Dalam pengamatan pada penelitian ini nilai Silhouette cenderung tidak konsisten adanya penurunan dari cluster 3 dan Kembali meningkat cluster 4 dan kembali menurun pada cluster 5.

Sebaran nama sparepart berdasarkan *object by cluster* untuk analisa ketersediaan penjualan *spare part* berdasarkan penyebaran dengan menggunakan (k=2) cluster yang paling ideal berdasarkan nilai silhouette. Berikut merupakan hasil sebaran cluster yang tertera pada Tabel 2.

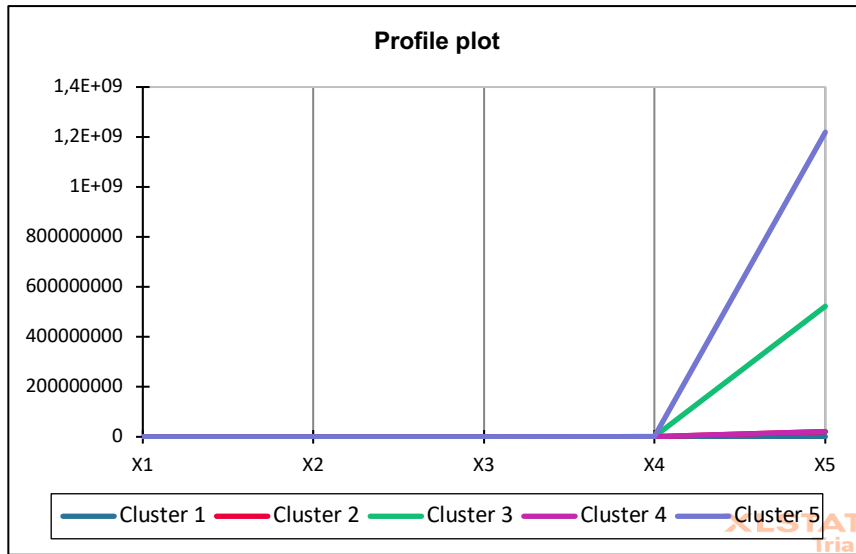
Tabel 2. Penyebaran Data Cluster 2

NAMA SPAREPART
REMOTE HOIST WIRELESS TELECONTROL F24-6D 6 TOMBOL + E/S
SEAL DHS NOK 80 X 90
SKUN KABEL 25 MM, BAUT 8 MM
HEAT DRYER BOSCH GHG 18-60
TANG SNAPRING TEKIRO 7 INCH
CENTER DRILL 8 X 3 MM
SHRINK TUBE KABEL 6MM X 100MM

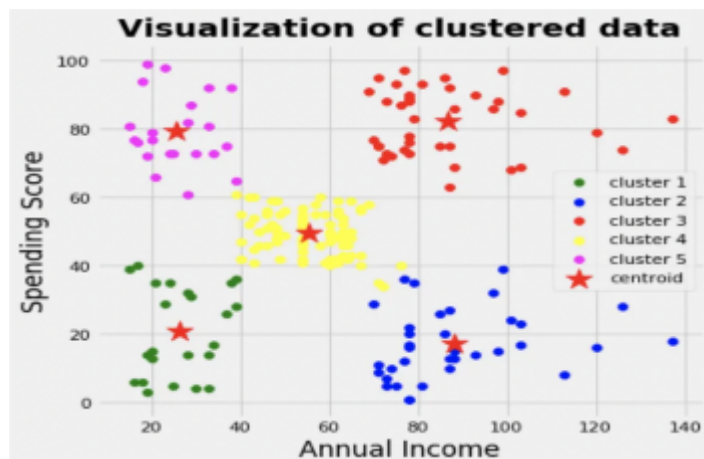
HIGH PRESSURE BUSHING BU-PT-20AX15A 3/4 X 1/2 12MPa
HEATER 210X80 220V
HEATER 60X100 220V
HEATER 60X100, 380V 840W
TWINFLEX TOZEN 3 INCH JIS10K KARET FLEXIBEL
PROXIMITY AUTONICS PRW12-4DP 12-24 VDC
PROXIMITY AUTONICS PRW12-4DP 12-24 VDC
BALL VALVE PVC 3/4 INCH
KLEM KABEL SUPER MAS 8 MM
PAKU BETON 3 CM
HEATER 30X30, 220V 200W
CONTACTOR SCHNEIDER LC1D40A 220V
CONTACTOR SCHNEIDER LC1D32M7 220V
GRC BOARD 3MM 120 X 240
PIPA PVC 3/4 INCH
PAKU GRC 3 CM
PISAU BUBUT INSERT DCMT 11
FUSE KACA 15A
SCREW INJECTION D120MM MESIN MITSUBISHI 1300T
SOK DRAT DALAM PVC 1 INCH
BESI AS VCN 85x1200
REMOTE HOIST WIRELESS TELECONTROL F24-6D 6 TOMBOL + E/S
SEAL DHS NOK 80 X 90
SKUN KABEL 25 MM, BAUT 8 MM
HEAT DRYER BOSCH GHG 18-60
TANG SNAPRING TEKIRO 7 INCH
CENTER DRILL 8 X 3 MM
SHRINK TUBE KABEL 6MM X 100MM
HIGH PRESSURE BUSHING BU-PT-20AX15A 3/4 X 1/2 12MPa
HEATER 210X80 220V
HEATER 60X100 220V
HEATER 60X100, 380V 840W
TWINFLEX TOZEN 3 INCH JIS10K KARET FLEXIBEL
PROXIMITY AUTONICS PRW12-4DP 12-24 VDC
PROXIMITY AUTONICS PRW12-4DP 12-24 VDC
BALL VALVE PVC 3/4 INCH
KLEM KABEL SUPER MAS 8 MM
PAKU BETON 3 CM

HEATER 30X30, 220V 200W
CONTACTOR SCHNEIDER LC1D40A 220V
CONTACTOR SCHNEIDER LC1D32M7 220V
GRC BOARD 3MM 120 X 240
PIPA PVC 3/4 INCH
PAKU GRC 3 CM
PISAU BUBUT INSERT DCMT 11
FUSE KACA 15A
SCREW INJECTION D120MM MESIN MITSUBISHI 1300T
SOK DRAT DALAM PVC 1 INCH
BESI AS VCN 85x1200
AIR WATER OIL FILTER SEPARATOR MOLLAR ARF2000
ACB SCHNEIDER NS1000N 3P 1000A 50Ka MOTORIZED
ORING VITON 8,8 X 1,9
CHECK VALVE SCREW 1/4" 3.0MM
HEAD SCREW MIXING DIA 90 + CHUCKRING
BAUT JOINT TRAY M8 X 20 + MUR M8
PLUG DOP PVC 1/2 INCH
BAUT + MUR 33MM X 150MM
STEKER LEGRAND MALE FORT 32 A
PISAU GERINDA BETON 4 INCH
BESI AS VCN 95 X 96 MM
FLANGE PVC 2"
BALL VALVE KUNINGAN 3/8 INCH
BAUT M 30 X 100MM
BOX PANEL 60 X 40 + BRACKET
TEMBAGA DIAMETER 50 x 200 MM
HEATER 40X40, 220V 250W
HEATER 40X40, 220V 250W
HEATER 40X40, 380V 250W
KAWAT LAS ARGON TECH 492 TG 1,6 MM
VLOK SOCK GALVANIS 2 X 1 1/4 INCH
TEE GALVANIS 1 1/4 INCH
DOUBLE DRAT GALVANIS 1 1/4 INCH
ELBOW GALVANIS 2 INCH
DOUBLE DRAT GALVANIS 2 INCH
CLAMP U OMEGA 1 INCH

Dari data yang digunakan berikut nama *sparepart* berdasarkan sebaran data dengan menggunakan $k=2$, yang mana memiliki nilai Silhouette.



Gambar 7. Sebaran data sparepart dari cluster 1 hingga cluster 5



Gambar 8. Sebaran Data

Berdasarkan sebaran di atas dapat dilihat penyeabran data yang paling rapat yaitu dengan $k=1$ dan $k=2$. Karena data terlihat memiliki kemiripan dan penyebaran yang rapat.

4. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini yaitu metode k -means clustering secara keseluruhan dapat menganalisa penjualan untuk ketersediaan suku cadang kendaraan, yang mana pada proses clustering diperlukan data yang numerik, pada penelitian ini data yang diperoleh merupakan data non numerik, maka dari itu diperlukan proses normalisasi menggunakan proses binning. Setelah proses binning dataset dapat diproses menggunakan k -means clustering, dan didapatkan sebaran data yang paling ideal yaitu menggunakan cluster $k=2$ dengan nilai Silhouette 0,925, kesimpulan

tersebut diperoleh karena sebaran data pada cluster $k=2$ memiliki miripan dan sebaran data yang rapat. Hasil penelitian ini diharapkan akan memberikan panduan praktis bagi perusahaan otomotif dalam memanfaatkan analisis data untuk meningkatkan efisiensi dan efektivitas operasional mereka dalam mengelola suku cadang mobil. Identifikasi segmentasi produk yang telah ditetapkan pada cluster yang didapat pada penelitian ini memberikan kemudahan pengelolaan stok, dengan atribut nama sparepart yang cukup banyak. Sehingga diharapkan dapat mempermudah dalam meningkatkan keuntungan.

Saran dari hasil penelitian ini peneliti memberikan rekomendasi untuk peneliti selanjutnya diperkenankan untuk dilakukan uji coba lebih banyak cluster yang digunakan dengan menggunakan periode data yang lebih banyak, agar data yang didapat bisa lebih optimal. Dan diperlukan penggunaan tools analisa terbaik untuk menganalisa dengan clustering, seperti menggunakan python ataupun spss.

DAFTAR PUSTAKA

- [1] M. P. Utami, "Aplikasi C-Service Motor Dengan Algoritma Artificial Neural Network Terintegrasi Sistem Pakar," *Jurnal sistem dan teknologi informasi*, vol. 11, no. 3, pp. 496-503, 2023.
- [2] Y. Z. Q. T. W. H. Y. J. S. T. X. Yming li, "K-Means: A Robust and Stable K-means Variant," *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, pp. 3120-3124, 2021.
- [3] F. T. K. F. Saut Parsaoran Tamba, "Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering," *Jurnal Sistem Informasi Ilmu Komputer Prima*, vol. 2, pp. 67-72, 2019.
- [4] D. W. Z. M. ., B. X. Jinyan Liu, "Improved K-means clustering algorithm for screwlocking classification," *Advanced Information Technology, Electronic and Automation Control Conference*, vol. 4, pp. 1955-1958, 2019.
- [5] S. K. P. T Namratha Reddy, "Optimization of K-Means Algorithm: Ant Colony Optimization," *International Conference on Computing Methodologies and Communication*, no. IEEE, pp. 530-535, 2017.
- [6] N. G. a. E. T. Alexander Grusho, "Detection of anomalies in Non-numerical Data," *International Congress on ultra Modern Telecommunication and Control Systems and Workshops (ICUMT)*, vol. 8, no. IEEE Explore, pp. 273-276, 2018.
- [7] M. s. B. B. Fransisca Angelia, "Optimization on Purity K-Means Using Variant Distance Measure," *Intenational Conference on Mechanical, Electronics, Computer and Industrial Techology*, vol. 3, no. IEEE Xplore, pp. 143-147, 2020.
- [8] J. H. F. B. G. Shivani shah, "ProxiClust: Data Sparsification and Community Detection for Assembly-free Metagenomic Binning," *Intenational Conference on Advances in Social Networks Analysis and Mining*, vol. 1, no. IEEE Xplore, pp. 17-20, 2018.
- [9] L. W. W. F. L. F. W. G. Chen Xinxin, "Network Loss Allocation Method Based on Improved K-means Clustering And Correlation Analysis," *UTC*, no. IEEE Explore, pp. 37-41, 2020.
- [10] Y. Z. zhuo wang, "Anomaly Detection by using streaming K-Means and Batch K-Means," *International Conference on Big Data Analysis*, vol. 5, no. IEEE Xplore, pp. 11-17, 2020.
- [11] T. H. C. R. Rui Maximo Esteves, "Competitive K-means," *International Conference on Cloud Computing Technology and Science*, vol. 8, no. IEEE, pp. 17-24, 2018.
- [12] D. X. Z. N. Wenhaou, "Improved K-Means Text Clustering Algorithm Based on BERT and Density Peak," *Infomation Communication Techologies Conference*, vol. 2, no. IEEE Xplore, pp. 260-264, 2021.

- [13] M. S. Y. Kristina P. Sinaga, "Unxupervised K-Means Clustering Algorithm," no. IEEE Access, pp. 80716 - 80727, 2020.
- [14] S. H. H. ., E. W. ., M. Dhendra Marutho, "The Determination of Cluster Number at k-mean using Elbow Method and Purity Evaluation on Headline News," *International Seminar on Application for Technology of Information and Communication (iSemantic)*, no. IEEE, pp. 533-538, 2018.
- [15] H. N. B. V. a. S. W. L. Thanh Quoc Phan, "Passive Elbow Movement Assistant (PEMA): A portable exoskeleton to compensates angle-dependent tone profile of the elbow joint post-stroke," *International Conference on Rehabilitation Robotics (ICORR) Toronto*, vol. 16, no. IEEE Exploe, pp. 1209-1214, 2019.