



Department of Digital Business

**Journal of Artificial Intelligence and Digital Business (RIGGS)**

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 5 No. 1 (2026) pp: 4759-4777

P-ISSN: 2963-9298, e-ISSN: 2963-914X

---

## Electric Vehicle Market Segmentation Based on Features and Specifications Using the K-Means Algorithm

Farhan Rivanto, Toni Arifin

Department Sistem Informasi Universitas Adhirajasa Reswara Sanjaya, Jawa Barat, Indonesia

[f2mas2000@gmail.com](mailto:f2mas2000@gmail.com)

### Abstrak

Kendaraan bermotor berbahan bakar fosil diketahui merupakan salah satu penyumbang utama emisi gas rumah kaca yang mempercepat perubahan iklim global. Sebagai alternatif yang lebih ramah lingkungan, mobil listrik hadir sebagai solusi karena tidak menghasilkan emisi gas buang secara langsung saat digunakan. Perkembangan teknologi kendaraan listrik mendorong meningkatnya minat konsumen terhadap berbagai tipe dan spesifikasi yang tersedia di pasar. Demikian, diperlukan segmentasi pasar untuk memahami karakteristik mobil listrik berdasarkan kebutuhan pengguna. Penelitian ini menerapkan algoritma K-Means Clustering untuk mengelompokkan mobil listrik berdasarkan fitur dan spesifikasi teknisnya. Dataset yang digunakan berasal dari platform Kaggle dengan judul *Cheapest Electric Cars 2023* yang dipublikasikan oleh koustubhk. Dataset tersebut terdiri dari 307 data dengan beberapa variabel utama, yaitu Name, Acceleration, TopSpeed, Range, Efficiency, FastChargeSpeed, Drive, dan Number of Seats. Variabel-variabel ini merepresentasikan performa, efisiensi, dan kapasitas kendaraan yang relevan dalam pengambilan keputusan konsumen. Hasil analisis menunjukkan bahwa pasar mobil listrik dapat dikelompokkan menjadi tiga kluster utama, yaitu kluster C0 (mobil keluarga) sebanyak 137 item, kluster C1 (mobil perkotaan) sebanyak 109 item, dan kluster C2 (mobil sport) sebanyak 61 item. Evaluasi kualitas clustering menggunakan indeks Davies-Bouldin menghasilkan nilai sebesar 0,641. Nilai tersebut menunjukkan hasil segmentasi yang diperoleh tergolong cukup baik karena semakin kecil nilai Davies-Bouldin Index, maka semakin optimal pemisahan antar kluster yang terbentuk..

*Kata kunci: Algoritma K-Means Clustering, Dataset, Mobil Listrik*

### 1. Latar Belakang

Kendaraan bermotor saat ini mayoritas menggunakan bahan bakar fosil yang tidak terbarukan. Semakin banyak kendaraan bermotor yang beroperasi, semakin tinggi juga konsumsi bahan bakar fosil. Penggunaan bahan bakar fosil oleh kendaraan bermotor menyebabkan peningkatan emisi gas rumah kaca yang berkontribusi pada perubahan iklim global (Triyatna, 2021). Menurut World Wide Fund, sektor transportasi pada tahun 2009 bertanggung jawab atas sekitar seperempat dari total emisi gas rumah kaca di atmosfer bumi. Data dari Kementerian ESDM menunjukkan bahwa konsumsi energi di sektor transportasi Indonesia meningkat dari 29% pada tahun 2007 menjadi 47% pada tahun 2017. Sektor transportasi juga menghasilkan emisi sebanyak 1,28 juta ton dengan tingkat peningkatan rata-rata sebesar 6,7% per tahun, yang melebihi peningkatan konsumsi bahan bakar sebesar 1,5 kali lipat (Ihsaf, 2022).

Bahan bakar fosil mengandung zat organik dan terbentuk melalui proses alami yang berlangsung selama jutaan tahun, di mana materi-materi tersebut terurai dan kemudian terkompresi untuk membentuk batu bara, minyak bumi, dan gas alam. Karena proses ini memakan waktu yang sangat lama dan tidak dapat diperbaharui, persediaan bahan bakar fosil secara perlahan akan habis (Sabrina, 2022). Penggunaan bahan bakar fosil memiliki dampak negatif yang signifikan seperti pemanasan global berdampak pada perubahan iklim yang ekstrem (Novita, 2022).

Seiring berjalannya waktu kendaraan bermotor berbahan bakar fosil ini, mulai tergantikan oleh mobil listrik produsen-produsen mobil mulai mengembangkan mobil ramah lingkungan ini untuk mengurangi konsumsi bahan bakar minyak, ada beberapa alasan utama kenapa mobil listrik lebih baik dari pada kendaraan bermotor berbahan bakar minyak fosil, yang pertama mobil listrik ini bahan bakar utama listrik pasti yang di keluarkan oleh mobil listrik tidak meninggalkan polusi (Krisna, 2020).

Peningkatan kesadaran akan manfaat lingkungan dan ekonomi dari mobil listrik, salah satu tantangan utama adalah memahami lebih dalam preferensi dan kebutuhan konsumen dalam memilih mobil listrik. Fokus utamanya adalah memperoleh pemahaman yang lebih baik tentang fitur dan spesifikasi yang dianggap penting oleh konsumen, serta melakukan segmentasi pasar untuk memahami kelompok konsumen yang berbeda dan preferensi mereka. Dengan demikian, segmentasi mobil listrik berdasarkan fitur dan spesifikasi menjadi krusial untuk memberikan wawasan yang lebih mendalam tentang preferensi konsumen, serta memfasilitasi pengembangan strategi pemasaran yang lebih efektif dan pengambilan keputusan yang lebih baik dalam pengembangan produk dan layanan mobil listrik. (Bhat, 2024).

Berdasarkan perkembangan teknologi yang sangat pesat saat ini, pemanfaatan teknologi informasi dapat kita temukan dalam berbagai bidang, salah satunya bidang pemasaran. Dibutuhkan suatu teknik untuk mengelola data mobil listrik. Metode pengelolaan data tersebut sering disebut dengan data mining adalah proses ekstraksi pola-pola menarik atau pengetahuan yang berharga dari kumpulan besar data dengan menggunakan berbagai teknik, termasuk machine learning, statistik, dan sistem database (Napizahni, 2023). Tujuan utama dari data mining adalah untuk menemukan informasi yang tersembunyi, yang tidak terdeteksi secara manual, dan untuk mengubah data mentah menjadi informasi yang berguna. Dalam konteks pasar mobil listrik, data mining dapat digunakan untuk menganalisis pola pembelian, preferensi konsumen, faktor-faktor yang memengaruhi keputusan pembelian, dan tren pasar yang dapat membantu dalam pengembangan dan pemasaran mobil listrik (Setiawan, 2021).

Algoritma yang digunakan pada teknik segmentasi ini adalah Algoritma Hierarchical yang didasarkan pada penelitian terdahulu yang dilakukan oleh (Kubiczek & Hadasik, 2021). Mengenai Penerapan Algoritma Hierarchical untuk pengelompokan mobil listrik di Polandia. Berdasarkan penelitiannya menunjukkan bahwa algoritma Hierarchical yang kurang efisien untuk data yang besar, sensitif terhadap noise dan outlier, dan membutuhkan penyesuaian parameter (Cahya, 2019). Selain itu ada juga penelitian terdahulu menggunakan segmentasi algoritma Fuzzy C-Means untuk mengelompokkan berbagai aspek kendaraan elektrik. Berdasarkan penelitiannya menunjukkan algoritma Fuzzy C-Means yang rentan terhadap noise (Nazari, 2023). Algoritma K-Means merupakan metode yang diciptakan untuk mengatasi kelemahan algoritma Hierarchical dan algoritma Fuzzy C-Means karena algoritma k-means dapat melakukan pengelompokan mobil listrik dengan baik, mampu mengelola dataset dengan jumlah data yang besar dengan efektif, dan hasil yang mudah diinterpresentasikan (Sutiono, 2023).

Berdasarkan uraian fenomena ini, penelitian ini bertujuan pada segmentasi pasar mobil listrik berdasarkan fitur dan spesifikasi menggunakan algoritma K-Means. Data mobil listrik yang digunakan akan dikelompokkan melalui teknik clustering untuk menghasilkan informasi mengenai karakteristik setiap segmen pasar. Proses ini bertujuan dalam memberikan *view* mengenai kelompok mobil listrik yang sesuai dengan kebutuhan konsumen tertentu. Kemudian, penelitian ini juga bertujuan untuk mengevaluasi hasil segmentasi guna memastikan kualitas dan akurasi pengelompokan yang dihasilkan. Dengan pendekatan berbasis data mining, penelitian ini dapat memberikan kontribusi dalam pengembangan strategi pemasaran dan inovasi produk mobil listrik. Pada akhirnya, hasil penelitian ini diharapkan menjadi referensi akademik maupun praktis dalam memahami dinamika segmentasi pasar kendaraan listrik.

## 2. Tinjauan Pustaka

Data mining merupakan proses ekstraksi informasi yang relevan dari kumpulan data berukuran besar dengan memanfaatkan teknik statistik, matematis, dan kecerdasan buatan (Serbi, 2020). Proses ini juga dikenal sebagai Knowledge Discovery in Databases (KDD), knowledge extraction, data atau pattern analysis, business intelligence, data archaeology, maupun data dredging (Harahap & Sulindawaty, 2020). Data mining memerlukan volume data yang signifikan agar mampu menghasilkan pola. Namun demikian, hasil yang diperoleh tidak selalu sempurna karena adanya kemungkinan noise atau ketidakakuratan data (Nayoan, 2022).

Tujuan utama data mining meliputi pendekatan *explanatory*, *confirmatory*, dan *exploratory*, yang masing-masing digunakan untuk menjelaskan kondisi tertentu, menguji hipotesis, dan menemukan pola baru dalam data (Setiawan, 2021). Secara umum, data mining terbagi menjadi *descriptive data mining* dan *predictive data mining* (Gede *et al.*, 2022). *Descriptive data mining* merupakan identifikasi pola atau karakteristik data, sedangkan *predictive data mining* bertujuan melakukan peramalan terhadap data di masa mendatang. Kemudian, menurut Sano (2019), data mining mencakup empat kelompok utama, yaitu *association*, *prediction*, *clustering*, dan *sequential relationships* yang masing-masing memiliki fungsi analitis berbeda.

Tahapan dalam proses *Knowledge Discovery in Databases* (KDD) terdiri dari lima langkah utama, yaitu:

1. Data Selection
2. Preprocessing/Cleaning
3. Transformation
4. Data Mining
5. Interpretation/Evaluation

Clustering merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik (Abdi, 2023). Tujuan utama clustering adalah memaksimalkan kemiripan dalam satu kelompok (*intra-cluster similarity*) dan meminimalkan kemiripan antar kelompok (*inter-cluster similarity*).

Salah satu algoritma clustering yang paling populer adalah K-Means. Algoritma ini pertama kali diperkenalkan oleh MacQueen pada tahun 1967 dan dikenal karena kesederhanaan serta efisiensinya dalam mengelompokkan data (elfanmauludi, 2023). Prinsip dasar K-Means adalah menentukan sejumlah K centroid awal secara acak, kemudian mengelompokkan data ke dalam cluster berdasarkan jarak terdekat terhadap centroid tersebut. Perhitungan jarak menggunakan {Euclidean Distance} dirumuskan sebagai berikut:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Persamaan (1) menyatakan bahwa  $d(p, q)$  adalah jarak antara dua titik data  $p$  dan  $q$ ,  $p_i$  dan  $q_i$  adalah nilai atribut ke- $i$ , serta  $n$  adalah jumlah atribut yang digunakan (Fahmi, 2021). Setelah proses penghitungan jarak, centroid diperbarui dengan menghitung rata-rata seluruh anggota cluster:

$$C_k = \frac{1}{|S_k|} \sum_{x_i \in S_k} x_i \quad (2)$$

di mana  $C_k$  adalah centroid cluster ke- $k$  dan  $S_k$  adalah himpunan anggota cluster ke- $k$ . Untuk mengevaluasi kualitas hasil clustering digunakan Silhouette Coefficient (Galela, 2023). Nilai Silhouette berada pada rentang  $-1$  hingga  $1$ . Rata-rata jarak intra-cluster dihitung sebagai:

$$a(i) = \frac{1}{|A|-1} \sum_{\substack{j \in A \\ j \neq i}} d(i, j) \quad (3)$$

Rata-rata jarak terhadap cluster lain:

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (4)$$

Nilai minimum jarak ke cluster lain:

$$b(i) = \min_{C \neq A} d(i, C) \quad (5)$$

Nilai Silhouette untuk setiap data:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (6)$$

Persamaan (3)-(6) menunjukkan  $a(i)$  adalah rata-rata jarak data ke- $i$  dalam cluster yang sama,  $b(i)$  adalah jarak minimum terhadap cluster lain, dan  $s(i)$  adalah nilai evaluasi akhir (Hidayati *et al.*, 2021). Nilai  $s(i) = 1$  menunjukkan data terkluster dengan baik,  $s(i) = 0$  berada di antara dua cluster, dan  $s(i) = -1$  menunjukkan pengelompokan kurang tepat.

Dalam implementasinya, penelitian ini menggunakan perangkat lunak RapidMiner yang sebelumnya dikenal sebagai YALE dan dikembangkan pada tahun 2001 (Fahmi, 2022). RapidMiner menyediakan lebih dari 500 operator untuk data preprocessing, modelling, dan evaluasi. Platform ini ditulis dalam bahasa Java dan mendukung pemodelan berbasis operator tree dengan representasi XML internal.

Penelitian terdahulu menunjukkan bahwa algoritma K-Means banyak digunakan dalam berbagai bidang. Kubiczek dan Hadasik (2021) menggunakan hierarchical clustering untuk segmentasi pasar mobil listrik di Polandia dan menghasilkan empat kelompok yaitu Premium, Kota, Kecil, dan Sport. Butsianto dan Wulan (2020) menggunakan K-Means dengan validasi DBI sebesar 0,341. Tamba dan Kesuma (2019) menerapkan K-Means pada penjualan sparepart Toyota. Pulungan *et al.* (2019) mengelompokkan kendaraan berdasarkan kegunaan. Kristian, Ginting, dan Ramadani (2022) menghasilkan tiga cluster kendaraan berdasarkan jenisnya.

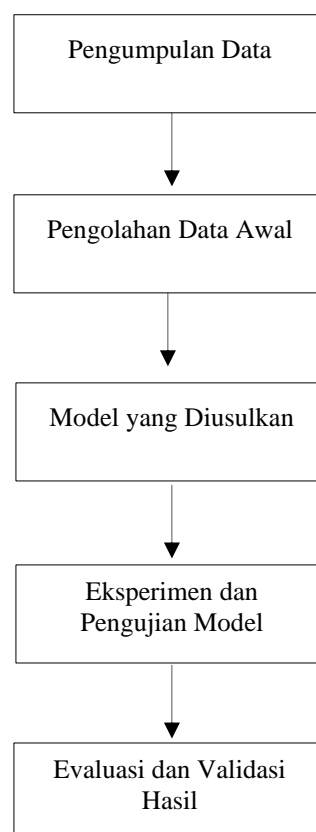
Penelitian lain seperti Sunggara (2019), Firdaus *et al.* (2023), Yusuf *et al.* (2023), Tetlageni dan Solichin (2023), Sman *et al.* (2020), serta Febriansyah dan Arifin (2023) juga menunjukkan efektivitas K-Means dengan evaluasi menggunakan DBI, SSE, dan Silhouette Coefficient.

Berdasarkan tinjauan studi ini, algoritma K-Means telah banyak digunakan karena kemampuannya dalam mengelompokkan data berdasarkan kemiripan karakteristik. Namun, segmentasi pasar mobil listrik berdasarkan fitur dan spesifikasi masih terbatas, sehingga penelitian ini berupaya mengisi celah tersebut dengan menerapkan algoritma K-Means serta evaluasi menggunakan Silhouette Coefficient.

### 3. Metode Penelitian

#### 3.1. Tahapan Penelitian

Dalam melakukan penelitian ini, metode penelitian yang sering digunakan meliputi penelitian tindakan, eksperimen, studi kasus, dan survei. Pada penelitian ini digunakan metode penelitian eksperimen. Penelitian eksperimen merupakan metode yang digunakan untuk mengetahui pengaruh suatu perlakuan tertentu terhadap variabel lain dalam kondisi yang terkendalikan (Qotrun, 2019). Berdasarkan metode tersebut, tahapan penelitian yang dilakukan meliputi beberapa langkah utama sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Desain Penelitian

Pertama, tahap pengumpulan data dilakukan dengan mengambil data publik yang tersedia secara daring. Dataset diperoleh dari situs penyedia data publik, yaitu Kaggle (<https://www.kaggle.com/>). Kedua, tahap pengolahan data awal dilakukan dengan mentransformasikan data ke dalam format yang sesuai untuk proses pemodelan,

termasuk pembagian data menjadi data latih (*training data*) dan data uji (*testing data*). Ketiga, tahap penentuan model yang diusulkan dilakukan dengan memilih metode data mining yang sesuai dengan karakteristik dataset. Keempat, tahap eksperimen dan pengujian model dilakukan untuk mengetahui hasil pengelompokan serta rule yang dapat digunakan dalam pengambilan keputusan. Kelima, tahap evaluasi dan hasil dilakukan dengan menilai performa model berdasarkan parameter evaluasi yang telah ditentukan.

### 3.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah dataset *Cheapest Electric Cars 2023* yang merupakan data sekunder atau data publik yang tidak diperoleh secara langsung dari penelitian lapangan, melainkan dari sumber lain. Dataset ini dipublikasikan pada tahun 2023 oleh koustubhk melalui platform Kaggle. Dataset tersebut terdiri dari 309 *records*, memiliki 7 atribut dan 1 label atau atribut target. Label pada dataset ini terdiri dari tiga cluster, yaitu cluster 0 (mobil sport), cluster 1 (mobil perkotaan), dan cluster 2 (mobil keluarga/mobil dengan jarak tempuh jauh). Distribusi awal data terdiri dari 61 *records* untuk cluster 0, 108 *records* untuk cluster 1, dan 140 *records* untuk cluster 2. Deskripsi atribut dataset dapat dilihat pada Tabel 1 berikut.

Tabel 1. Deskripsi Atribut Dataset Cheapest Electric Cars 2023

No	Attribute	Satuan	Deskripsi
1	Acceleration	sec	Kemampuan mobil bergerak dari posisi diam hingga mencapai kecepatan tertentu
2	TopSpeed	km/h	Kecepatan maksimal mobil
3	Range	km	Jarak tempuh kendaraan
4	Efficiency	wh/km	Tingkat efisiensi penggunaan daya listrik
5	FastChargeSpeed	km/h	Kecepatan pengisian daya mobil listrik
6	Drive	-	Sistem penggerak: 0 = All Wheel Drive, 1 = Front Wheel Drive, 2 = Rear Wheel Drive
7	NumberofSeats	-	Jumlah kapasitas tempat duduk
8	Cluster	-	0 = Mobil sport, 1 = Mobil perkotaan, 2 = Mobil keluarga/range jauh

### 3.3. Pengolahan Data Awal

Pada tahap ini dilakukan eksplorasi data (*data exploration*) terhadap dataset *Cheapest Electric Cars 2023*. Proses ini bertujuan untuk memastikan validitas atribut dan kelas dalam dataset sehingga dapat menjamin kesesuaiannya sebagai objek penelitian. Dataset yang digunakan terdiri dari 309 *records* dan siap untuk diproses menggunakan model yang telah ditentukan.

Untuk pengujian model, dataset dibagi menjadi dua bagian menggunakan metode *split validation*, yaitu 90% data training dan 10% data testing. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk menguji kinerja model yang telah dibangun. Distribusi data training dan testing ditunjukkan pada Tabel 2 berikut.

Tabel 2 Sebaran Dataset Cheapest Electric Cars 2023

No	Kelas	Jumlah Record	Data Training	Data Testing
1	Cluster 0 (Mobil Keluarga)	137	123	13
2	Cluster 1 (Mobil Perkotaan)	109	98	10
3	Cluster 2 (Mobil Sport)	61	54	6
	Jumlah	307	275	29

### 3.4. Model yang Diusulkan

Dataset *Cheapest Electric Cars 2023* merupakan data yang telah diproses dan siap untuk dilakukan analisis data mining. Penelitian ini menggunakan metode clustering K-Means sebagai model yang diusulkan. Model tersebut dipilih karena kemampuannya dalam mengelompokkan data berdasarkan kemiripan karakteristik.

Alur model yang diusulkan ditunjukkan pada Gambar 2.



Gambar 2. Model yang Diusulkan

Tahapan model K-Means yang diterapkan dalam penelitian ini meliputi: memasukkan dataset, menentukan nilai K sebagai jumlah cluster yang akan dibentuk, menentukan centroid awal secara acak, menghitung jarak setiap data terhadap centroid menggunakan rumus Euclidean Distance, mengelompokkan data berdasarkan jarak terdekat, menghitung ulang nilai centroid sebagai rata-rata anggota cluster, dan melakukan iterasi hingga tidak terjadi perubahan pada centroid atau keanggotaan cluster.

### 3.5. Eksperimen dan Pengujian Model

Pada tahap ini dilakukan eksperimen dengan menerapkan metode K-Means pada dataset menggunakan perangkat lunak RapidMiner Studio versi 9.10. Proses eksperimen bertujuan untuk menghasilkan segmentasi pasar mobil listrik berdasarkan atribut yang tersedia dalam dataset. Hasil pengujian digunakan untuk melihat performa clustering serta kualitas pengelompokan yang terbentuk.

### 3.6. Evaluasi dan Validasi

Tahap evaluasi dilakukan menggunakan RapidMiner Studio versi 9.10 untuk menilai kualitas hasil clustering. Evaluasi dalam penelitian ini menggunakan metode Davies Bouldin Index (DBI). Nilai DBI digunakan sebagai indikator kualitas cluster, di mana semakin kecil nilai DBI maka semakin baik hasil pengelompokan yang diperoleh. Evaluasi ini bertujuan untuk memastikan bahwa segmentasi pasar mobil listrik yang dihasilkan memiliki kualitas cluster yang optimal dan representatif terhadap karakteristik data.

## 4. Hasil dan Diskusi

### 4.1. Perhitungan Algoritma K-Means

Pada Penelitian ini penulis mengambil *data public* dari web [www.kaggle.com](http://www.kaggle.com). Data yang digunakan merupakan data *Cheapest Electric Cars 2023* dengan jumlah sebanyak 307 data. Pada proses perhitungan menggunakan algoritma *K-Means* ada beberapa tahapan, yaitu:

Tahap pertama dalam algoritma *K-Means* menentukan jumlah *cluster*, pada penelitian ini terdapat 3 *cluster*. Penentuan *centroid* awal dilakukan secara *random* dengan atribut *acceleration*, *topspeed*, *range*, *efficiency*, *fast charge speed*, *drive*, *number of seats*. berikut tabel di bawah ini merupakan *dataset* yang telah melalui tahap transformasi.

Tabel 3. Dataset Hasil Transformasi

No	Name	Drive	Acceleration	TopSpeed	Range	Efficiency	FastChargeSpeed	NumberOfSeats
ML1	Lucid Air Dream Edition P	0	2	270	645	183	820	5
ML2	Porsche Taycan Turbo S	0	2	260	400	209	980	4
ML3	Audi e-tron GT RS	0	3	250	405	210	1000	4
ML4	Renault Zoe ZE50 R110	1	11	135	315	165	230	5
ML5	Audi Q4 e-tron 35	2	9	160	285	182	360	5
ML6	Tesla Model Y Performance	0	3	250	415	181	640	5
ML7	Porsche Taycan 4 Cross Turismo	0	5	220	425	197	1050	4
ML8	Renault Zoe ZE50 R135	1	9	140	310	168	230	5
ML9	Porsche Taycan Turbo	0	3	260	415	202	1020	4
ML10	Smart EQ fortwo coupe	2	11	130	100	167	230	2
ML11	Smart EQ fortwo cabrio	2	11	130	95	176	370	2
ML12	Honda e Advance	2	8	145	170	168	180	4
ML13	Porsche Taycan 4S	0	4	250	395	180	970	4
ML14	Porsche Taycan 4S Plus	0	4	250	460	182	1130	4
ML15	Mercedes EQV 300 Long	1	12	160	305	295	310	7
ML16	Lexus UX 300e	1	7	160	235	191	150	5
ML17	BMW i4 eDrive40	2	5	190	470	172	730	5
ML18	Renault Twingo Electric	2	12	135	130	164	840	4
ML19	Fiat 500e Cabrio 42 kWh	1	9	150	230	162	380	4
ML20	Skoda Enyaq iV 60	2	8	160	325	178	440	5
ML21	Skoda Enyaq iV 80	2	8	160	420	183	530	5
ML22	Skoda Enyaq iV 80x	0	6	160	400	193	500	5
ML23	Skoda Enyaq iV RS	0	6	180	390	197	540	5
ML24	Fiat 500e Hatchback 42 kWh	1	9	150	235	159	390	4
ML25	Audi Q4 Sportback e-tron 35	2	9	160	295	176	370	5
ML26	Nissan Ariya 63kWh	1	7	160	320	197	430	5
ML27	Nissan Ariya 87kWh	1	7	160	435	200	520	5
ML28	Nissan Ariya e-4ORCE 87kWh - 225 kW	0	5	200	420	207	500	5
ML29	Nissan Ariya e-4ORCE 87kWh - 290 kW	0	5	200	420	207	500	5
ML30	Volkswagen ID.4 Pro Performance	2	8	160	410	188	520	5
ML31	Mercedes EQV 300 Extra-Long	1	12	160	305	295	310	7
ML32	Lucid Air Grand Touring	0	3	270	665	168	840	5
ML33	Lucid Air Touring	0	3	250	550	160	960	5
ML34	Lucid Air Pure	2	4	200	560	157	980	5
ML35	Fiat 500e Hatchback 24 kWh	1	9	135	135	158	230	4
ML36	Fiat 500e 3+1 42 kWh	1	9	150	235	159	390	4
ML37	Mercedes EQC 400 4MATIC	0	5	180	370	216	440	5
ML38	Citroen e-SpaceTourer M 50 kWh	1	12	130	185	250	290	7
ML39	Citroen e-SpaceTourer XL 50 kWh	1	12	130	180	257	290	7
ML40	Citroen e-SpaceTourer M 75 kWh	1	13	130	265	257	290	7
ML41	Citroen e-SpaceTourer XL 75 kWh	1	13	130	260	262	280	7
ML42	Opel Zafira-e Life M 50 kWh	1	12	130	185	250	290	7
ML43	Opel Zafira-e Life L 50 kWh	1	12	130	180	257	290	7
ML44	Opel Zafira-e Life M 75 kWh	1	13	130	265	257	290	7
ML45	Opel Zafira-e Life L 75 kWh	1	13	130	260	262	280	7
ML46	Peugeot e-Traveller Standard 50 kWh	1	13	130	185	250	290	7

ML47	Peugeot e-Traveller Long 50 kWh	1	13	130	180	257	290	7
ML48	Peugeot e-Traveller Standard 75 kWh	1	14	130	265	257	290	7
ML49	Peugeot e-Traveller Long 75 kWh	1	14	130	260	262	280	7
ML50	Porsche Taycan	2	5	230	410	173	1010	4
ML51	Porsche Taycan Plus	2	5	230	475	176	1170	4
ML52	Tesla Model S Dual Motor	0	3	250	570	167	790	5
ML53	Tesla Model S Plaid	0	2	282	550	173	770	5
ML54	Tesla Model X Dual Motor	0	3	250	470	202	650	7
ML55	Tesla Model X Plaid	0	2	262	455	209	630	7
ML56	Mini Cooper SE	1	7	150	180	161	260	4
ML57	JAC iEV7s	1	12	132	225	173	160	5
ML58	Volkswagen ID.4 Pure Performance	2	9	160	285	182	360	5
ML59	Audi e-tron GT quattro	0	4	245	420	202	1030	4
ML60	Hyundai Kona Electric 39 kWh	1	9	155	250	157	220	5
ML61	Hyundai Kona Electric 64 kWh	1	7	167	395	162	370	5
ML62	Porsche Taycan 4S Cross Turismo	0	4	240	425	197	1050	4
ML63	Porsche Taycan Turbo Cross Turismo	0	3	250	400	209	980	4
ML64	Porsche Taycan Turbo S Cross Turismo	0	2	250	385	217	950	4
ML65	Kia EV6 GT	0	3	260	370	200	970	5
ML66	BMW iX xDrive40	0	6	200	360	197	480	5
ML67	BMW iX xDrive50	0	4	200	505	208	680	5
ML68	MG MG5 Electric Long Range	1	7	185	335	171	330	5
ML69	MG Marvel R Performance	0	4	200	330	197	390	5
...	...	...	...	...	...	...	...	...
ML183	Mercedes EQS 450 4MATIC	0	5	210	615	175	920	5
ML184	Mercedes EQS 500 4MATIC	0	4	210	605	178	900	5
ML185	BMW iX1 xDrive30	0	5	180	355	182	510	5
ML186	Volkswagen ID.4 Pro 4MOTION	0	6	180	405	190	510	5
...	...	...	...	...	...	...	...	...
ML301.	Maxus MIFA 9	1	9	180	345	261	290	7
ML302.	BMW i7 M70 xDrive	0	3	250	490	208	730	5
ML303.	Volkswagen ID.7 Pro	2	8	180	470	164	700	5
ML304.	Volkswagen ID.7 Pro S	2	7	180	520	165	800	5
ML305.	Polestar 4 Long Range Single Motor	2	7	180	510	184	760	5
ML306.	Polestar 4 Long Range Dual Motor	0	3	200	475	198	710	5
ML307.	Maserati Grecale Folgore	0	4	220	405	235	440	5

Tahapan selanjutnya, penulis membentuk 3 cluster berdasarkan *acceleration*, *topspeed*, *range*, *efficiency*, *fast charge speed*, *drive*, *number of seats*. Selanjutnya adalah menentukan titik pusat dari setiap *cluster* (*centroid*) yang dipilih secara acak dari data yang telah di transformasi dari data 12, 55 dan 183.

Tabel 4. Centroid Awal

Centroid	No	Nama	Drive	Acceleration	TopSpeed	Range	Efficiency	FastCharge	NumberofSeat
Cluster0	ML55	Tesla Model X Plaid	0	2	262	455	209	630	7
Cluster1	ML12	Honda e Advance	2	8	145	170	168	180	4
Cluster2	ML183	Mercedes EQS 450 4MATIC	0	5	210	615	175	920	5

Setelah menentukan centroid awal, maka tahap selanjutnya yaitu menghitung jarak setiap data yang ada ke centroid terdekat untuk menentukan cluster menggunakan rumus *Euclidean distance*. Dari perhitungan data ke-1 dengan *cluster* ke-2 maka dihasilkan jarak 120.8

Hasil perhitungan tersebut yaitu jarak data ke-1 dengan *cluster* ke-0 sebesar 269.9, jarak data ke-1 dengan *cluster* ke-1 sebesar 806.9, jarak data ke-1 dengan *cluster* ke-2 sebesar 120.8. Perhitungan tersebut dilakukan pada semua data yang mana terdapat 307 dataset mobil listrik. Setelah dilakukan perhitungan pada semua data tersebut, maka proses selanjutnya adalah memilah cluster mana dari tiap data dimana nilai terkecil dari tiap

perhitungan cluster akan menjadi penentu cluster yang akan diikuti dari data tersebut. Berikut merupakan tabel 5. hasil perhitungan jarak pada iterasi ke-1.

Tabel 5. Hasil Perhitungan dengan Menggunakan Rumus Euclidean pada Iterasi ke-1

Data Ke-i	Jarak ke Centroid			Jarak Terdekat	Cluster diikuti
	C0	C1	C2		
ML1	270,08	806,92	120,72	120,72	2
ML2	354,31	841,34	231,28	231,28	2
ML3	373,57	860,49	230,93	230,93	2
ML4	444,69	153,77	756,22	153,77	1
ML5	336,14	214,59	651,97	214,59	1
ML6	51,31	531,84	346,47	51,31	0
ML7	423,35	910,17	231,48	231,48	2
ML8	444,57	148,75	757,69	148,75	1
ML9	392,13	883,20	230,72	230,72	2
ML10	552,56	87,40	864,78	87,40	1
ML11	464,57	205,00	761,15	205,00	1
ML12	546,94	0,00	865,98	0,00	1
ML13	346,70	828,20	229,19	229,19	2
ML14	500,91	998,92	264,15	264,15	2
ML15	377,89	226,95	696,53	226,95	1
...	...	...	...	...	...
ML295	250,03	326,04	555,25	250,03	0
ML296	236,03	324,86	541,95	236,03	0
ML297	120,59	575,72	292,82	120,59	0
ML298	253,90	750,89	170,92	170,92	2
ML299	235,22	729,02	187,72	187,72	2
ML300	242,95	331,35	539,11	242,95	0
ML301	370,38	229,37	691,46	229,37	1
ML302	106,65	646,18	233,28	106,65	0
ML303	117,97	601,37	265,44	117,97	0
ML304	204,49	712,84	156,31	156,31	2
ML305	165,25	673,41	193,94	165,25	0
ML306	103,78	614,72	253,64	103,78	0
ML307	202,60	364,64	527,45	202,60	0

Selanjutnya kelompokkan data sesuai dengan jarak *cluster* terdekat. Dari data yang sudah di kelompokkan akan didapat *centroid* baru dari hasil rata-rata setiap *cluster*. Berdasarkan perhitungan terdapat 152 data yang tergabung dalam *cluster*0, yaitu sebagai berikut table 6 di bawah ini:

Tabel 6 Anggota *cluster* 0 pada Iterasi ke-1

No	Name	Drive	Acceleration	Top Speed	Range	Efficiency	Fast Charge Speed	Number of Seats
ML6	Tesla Model Y Performance	0	3	250	415	181	640	5
ML17	BMW i4 eDrive40	2	5	190	470	172	730	5
ML18	Renault Twingo Electric	2	12	135	130	164	840	4
ML20	Skoda Enyaq iV 60	2	8	160	325	178	440	5
ML21	Skoda Enyaq iV 80	2	8	160	420	183	530	5
ML22	Skoda Enyaq iV 80x	0	6	160	400	193	500	5
ML23	Skoda Enyaq iV RS	0	6	180	390	197	540	5
ML26	Nissan Ariya 63kWh	1	7	160	320	197	430	5
ML27	Nissan Ariya 87kWh	1	7	160	435	200	520	5
ML28	Nissan Ariya e-4ORCE 87kWh - 225 kW	0	5	200	420	207	500	5
...	...	...	...	...	...	...	...	...
ML53	Tesla Model S Plaid	0	2	282	550	173	770	5
ML54	Tesla Model X Dual Motor	0	3	250	470	202	650	7
ML55	Tesla Model X Plaid	0	2	262	455	209	630	7
ML61	Hyundai Kona Electric 64 kWh	1	7	167	395	162	370	5
ML66	BMW iX xDrive40	0	6	200	360	197	480	5
Jumlah	152							
Rata-rata		0,70	5,89	188,38	407,93	191,50	565,00	5,16

Selanjutnya kelompokan data sesuai dengan jarak *cluster* terdekat. Dari data yang sudah di kelompokan akan didapat *centroid* baru dari hasil rata-rata setiap *cluster*. Berdasarkan perhitungan terdapat 103 data yang tergabung dalam *cluster1*, yaitu sebagai berikut table 7 di bawah ini:

Tabel 7. Anggota *cluster* 1 pada Iterasi ke-1

No	Name	Drive	Acceleration	Top Speed	Range	Efficiency	Fast Charge Speed	Number of Seats
ML4	Renault Zoe ZE50 R110	1	11	135	315	165	230	5
ML5	Audi Q4 e-tron 35	2	9	160	285	182	360	5
ML8	Renault Zoe ZE50 R135	1	9	140	310	168	230	5
ML10	Smart EQ fortwo coupe	2	11	130	100	167	230	2
ML11	Smart EQ fortwo cabrio	2	11	130	95	176	370	2
ML12	Honda e Advance	2	8	145	170	168	180	4
ML15	Mercedes EQV 300 Long	1	12	160	305	295	310	7
ML16	Lexus UX 300e	1	7	160	235	191	150	5
ML19	Fiat 500e Cabrio 42 kWh	1	9	150	230	162	380	4
...	...	...	...	...	...	...	...	...
ML140	Toyota PROACE Shuttle M 75 kWh	1	13	130	260	262	280	7
ML141	Toyota PROACE Shuttle L 75 kWh	1	13	130	260	262	280	7
ML142	Mercedes eVito Tourer Long 90 kWh	1	12	160	310	290	310	7
ML143	Mercedes eVito Tourer Extra-Long 90 kWh	1	12	160	305	295	310	7
ML144	Mercedes eVito Tourer Long 60 kWh	1	12	160	210	286	310	7
Jumlah	103							
Rata -rata		1,05	10,42	145,29	242,18	216,43	293,20	5,79

Selanjutnya kelompokan data sesuai dengan jarak *cluster* terdekat. Dari data yang sudah di kelompokan akan didapat *centroid* baru dari hasil rata-rata setiap *cluster*. Berdasarkan perhitungan terdapat 52 data yang tergabung dalam *cluster2*, yaitu sebagai berikut table 8. di bawah ini:

Tabel 8. Anggota *cluster* 2 pada Iterasi ke-1

No	Name	Drive	Acceleration	Top Speed	Range	Efficiency	Fast Charge Speed	Number of Seats
ML1	Lucid Air Dream Edition P	0	2	270	645	183	820	5
ML2	Porsche Taycan Turbo S	0	2	260	400	209	980	4
ML3	Audi e-tron GT RS	0	3	250	405	210	1000	4
ML7	Porsche Taycan 4 Cross Turismo	0	5	220	425	197	1050	4
ML9	Porsche Taycan Turbo	0	3	260	415	202	1020	4
ML13	Porsche Taycan 4S	0	4	250	395	180	970	4
ML14	Porsche Taycan 4S Plus	0	4	250	460	182	1130	4
...	...	...	...	...	...	...	...	...
ML32	Lucid Air Grand Touring	0	3	270	665	168	840	5
ML33	Lucid Air Touring	0	3	250	550	160	960	5
ML34	Lucid Air Pure	2	4	200	560	157	980	5
ML50	Porsche Taycan	2	5	230	410	173	1010	4
ML51	Porsche Taycan Plus	2	5	230	475	176	1170	4
Jumlah	52							
Rata - rata		0,58	4,29	228,75	463,17	189,27	974,81	4,67

Hasil rata-rata di dapatkan dari 3 tabel setiap *cluster* diatas adalah *centroid* baru yang akan digunakan untuk menghitung jarak terdekat pada iterasi ke-2. Berikut tabel 9. merupakan rata rata dari setiap *cluster* yang terdapat pada iterasi ke-1.

Tabel 9. *Centroid* Baru Hasil Iterasi ke-1

Centroid	Drive	Acceleration	TopSpeed	Range	Efficiency	FastChargeSpeed	NumberofSeats
Cluster 0	0,58	4,29	228,75	463,17	189,27	974,81	4,67
Cluster 1	1,05	10,42	145,29	242,18	216,43	293,20	5,79
Cluster 2	0,58	4,29	228,75	463,17	189,27	974,81	4,67

Selanjutnya lakukan langkah-langkah diatas sampai nilai *centroid* tidak mengalami perubahan. Berikut tabel 10. hasil nilai *centroid* dari iterasi ke-3.

Tabel 10. *Centroid* Baru Hasil Iterasi ke-3

Centroid	Drive	Acceleration	TopSpeed	Range	Efficiency	FastChargeSpeed	NumberofSeats
Cluster 0	0,69	5,87	187,62	410,64	191,06	562,14	5,18
Cluster 1	1,04	10,19	147,33	248,03	215,79	299,17	5,74
Cluster 2	0,62	4,45	227,02	457,07	189,19	956,21	4,69

Pada perhitungan yang telah dilakukan. Percobaan perhitungan berhenti pada iterasi ke-4. Berikut merupakan tabel 11. nilai *centroid* hasil iterasi ke-4 yang sudah tidak mengalami perubahan sehingga perhitungan berhenti pada iterasi tersebut.

Tabel 11. Jarak Terdekat Hasil Iterasi ke – 4

Data Ke-i	Jarak ke Centroid			Jarak Terdekat	Cluster diikuti
	C0	C1	C2		
ML1	358,16	667,11	236,14	236,14	2
ML2	424,61	706,70	72,87	72,87	2
ML3	442,73	725,55	74,79	74,79	2
ML4	350,63	109,56	746,09	109,56	1
ML5	239,80	79,82	624,21	79,82	1
ML6	100,41	394,77	319,93	100,41	0
ML7	489,18	775,06	99,68	99,68	2
ML8	351,08	104,71	746,36	104,71	1
ML9	463,70	748,61	84,23	84,23	2
ML10	459,08	171,45	815,37	171,45	1
ML11	374,35	174,16	695,97	174,16	1
ML12	454,20	150,30	831,92	150,30	1
ML13	413,05	695,32	68,24	68,24	2
ML14	573,48	864,25	175,48	175,48	2
ML15	293,84	99,01	675,62	99,01	1
...	...	...	...	...	...
ML295	156,28	196,09	553,04	156,28	0
ML296	132,84	189,20	522,25	132,84	0
ML297	130,81	436,15	284,96	130,81	0
ML298	312,98	609,79	102,24	102,24	2
ML299	292,29	587,65	120,93	120,93	2
ML300	139,99	200,61	505,49	139,99	0
ML301	288,68	112,26	681,02	112,26	1
ML302	196,62	504,79	230,52	196,62	0
ML303	152,72	462,26	262,05	152,72	0
ML304	263,20	573,11	176,54	176,54	2
ML305	221,65	532,05	208,67	208,67	2
ML306	161,91	472,69	248,49	161,91	0
ML307	133,92	223,97	520,89	133,92	0

Selanjutnya kelompokkan data sesuai dengan jarak *cluster* terdekat. Dari data yang sudah di kelompokkan akan didapat *centroid* baru dari hasil rata-rata setiap *cluster*. Berdasarkan perhitungan terdapat 140 data yang tergabung dalam *cluster0*, yaitu sebagai berikut table 12 di bawah ini

Tabel 12. Anggota Cluster ke-0 pada iterasi k-4

No	Name	Drive	Acceleration	Top Speed	Range	Efficiency	Fast Charge Speed	Number of Seats
ML6	Tesla Model Y Performance	0	3	250	415	181	640	5
ML17	BMW i4 eDrive40	2	5	190	470	172	730	5
ML20	Skoda Enyaq iV 60	2	8	160	325	178	440	5
ML21	Skoda Enyaq iV 60	2	8	160	325	178	440	5
ML22	Skoda Enyaq iV 80	2	8	160	420	183	530	5
ML23	Skoda Enyaq iV 80x	0	6	160	400	193	500	5
ML27	Nissan Ariya 87kWh	1	7	160	435	200	520	5
ML28	Nissan Ariya e-4ORCE 87kWh - 225 kW	0	5	200	420	207	500	5
ML29	Nissan Ariya e-4ORCE 87kWh - 290 kW	0	5	200	420	207	500	5
ML30	Volkswagen ID.4 Pro Performance	2	8	160	410	188	520	5
...	...	...	...	...	...	...	...	...
ML300	Peugeot e-208	1	8	150	310	155	480	5
ML302	BMW i7 M70 xDrive	0	3	250	490	208	730	5
ML303	Volkswagen ID.7 Pro	2	8	180	470	164	700	5
ML306	Polestar 4 Long Range Dual Motor	0	3	200	475	198	710	5
ML307	Maserati Grecale Folgore	0	4	220	405	235	440	5
Jumlah	137							
Rata-rata		0,69	5,87	187,62	410,64	191,06	562,14	5,18

Selanjutnya kelompokan data sesuai dengan jarak *cluster* terdekat. Dari data yang sudah di kelompokan akan didapat *centroid* baru dari hasil rata-rata setiap *cluster*. Berdasarkan perhitungan terdapat 109 data yang tergabung dalam *cluster1*, yaitu sebagai berikut table 13. di bawah ini

Tabel 13. Anggota Cluster ke-1 pada iterasi k-4

No	Name	Drive	Acceleration	Top Speed	Range	Efficiency	Fast Charge Speed	Number of Seats
ML40	Citroen e-SpaceTourer M 75 kWh	1	13	130	265	257	290	7
ML41	Citroen e-SpaceTourer XL 75 kWh	1	13	130	260	262	280	7
ML42	Opel Zafira-e Life M 50 kWh	1	12	130	185	250	290	7
ML43	Opel Zafira-e Life L 50 kWh	1	12	130	180	257	290	7
ML44	Opel Zafira-e Life M 75 kWh	1	13	130	265	257	290	7
ML45	Opel Zafira-e Life L 75 kWh	1	13	130	260	262	280	7
ML46	Peugeot e-Traveller Standard 50 kWh	1	13	130	185	250	290	7
ML47	Peugeot e-Traveller Long 50 kWh	1	13	130	180	257	290	7
ML48	Peugeot e-Traveller Standard 75 kWh	1	14	130	265	257	290	7
ML49	Peugeot e-Traveller Long 75 kWh	1	14	130	260	262	280	7
...	...	...	...	...	...	...	...	...
ML279	ORA Funky Cat 63 kWh	1	8	160	340	174	240	5
ML280	ORA Funky Cat 63 kWh	1	8	160	340	174	240	5
ML292	Mini Cooper SE Convertible	1	8	150	170	170	240	4
ML294	Hyundai Kona Electric 48 kWh	1	10	155	300	161	290	5
ML301	Maxus MIFA 9	1	9	180	345	261	290	7
Jumlah	109							
Rata-rata		1,04	10,19	147,33	248,03	215,79	299,17	5,74

Selanjutnya kelompokan data sesuai dengan jarak *cluster* terdekat. Dari data yang sudah di kelompokan akan didapat *centroid* baru dari hasil rata-rata setiap *cluster*. Berdasarkan perhitungan terdapat 58 data yang tergabung dalam *cluster2*, yaitu sebagai berikut table 14. di bawah ini

Tabel 14. Anggota *Cluster* ke-2 pada iterasi k-4

No	Name	Drive	Acceleration	Top Speed	Range	Efficiency	Fast Charge Speed	Number of Seats
ML1	Lucid Air Dream Edition P	0	2	270	645	183	820	5
ML2	Porsche Taycan Turbo S	0	2	260	400	209	980	4
ML3	Audi e-tron GT RS	0	3	250	405	210	1000	4
ML7	Porsche Taycan 4 Cross Turismo	0	5	220	425	197	1050	4
ML9	Porsche Taycan Turbo	0	3	260	415	202	1020	4
ML13	Porsche Taycan 4S	0	4	250	395	180	970	4
ML14	Porsche Taycan 4S Plus	0	4	250	460	182	1130	4
ML18	Renault Twingo Electric	2	12	135	130	164	840	4
ML32	Lucid Air Grand Touring	0	3	270	665	168	840	5
ML33	Lucid Air Touring	0	3	250	550	160	960	5
...	...	...	...	...	...	...	...	...
ML291	XPENG G9 AWD Long Range	0	3	200	440	214	920	5
ML298	Kia EV9 99.8 kWh RWD	2	9	180	460	217	870	7
ML299	Kia EV9 99.8 kWh AWD	0	5	180	450	222	850	7
ML304	Volkswagen ID.7 Pro S	2	7	180	520	165	800	5
ML305	Polestar 4 Long Range Single Motor	2	7	180	510	184	760	5
Jumlah	61							
Rata-rata		0,62	4,45	227,02	457,07	189,19	956,21	4,69

Hasil dari tabel 15. di bawah ini rata-rata yang didapatkan dari 3 tabel setiap *cluster* diatas adalah centroid yang sudah tidak mengalami perubahan.

Tabel 15. Centroid Baru Hasil Iterasi ke-4

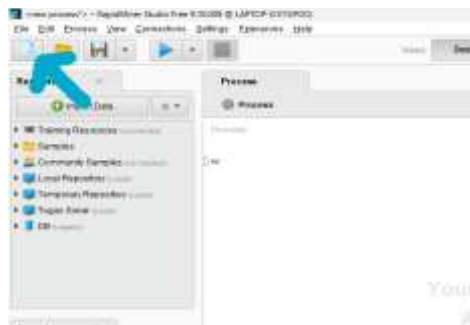
Centroid	Drive	Acceleration	TopSpeed	Range	Efficiency	FastChargeSpeed	NumberofSeats
Cluster 0	0,69	5,87	187,62	410,64	191,06	562,14	5,18
Cluster 1	1,04	10,19	147,33	248,03	215,79	299,17	5,74
Cluster 2	0,62	4,45	227,02	457,07	189,19	956,21	4,69

Dari perhitungan iterasi ke-4 diperoleh hasil bahwa *cluster* C1: mobil perkotaan dengan hasil drive= 1,04, acceleration= 10,19, top speed= 147,33, range= 248,03, efficiency= 215,79, fast charge speed= 299,17, number of seats= 5,74. *Cluster* C2: mobil sport menjadi hasil di tengah tengah antar *cluster* 1 dan *cluster* 0 dengan hasil drive= 0,62, acceleration= 0,62, top speed= 227,02, range= 457,07, efficiency =189,19, fast charge speed= 956,21, number of seats= 4,69. *Cluster* C0: mobil keluarga menjadi hasil terendah dengan hasil drive= 0,69, acceleration= 5,87, top speed= 187,62, range= 410,64, efficiency= 191,06, fast charge speed= 562,14, number of seats= 5,18.

#### 4.2. Perhitungan Algoritma K-Means Menggunakan Rapidminer

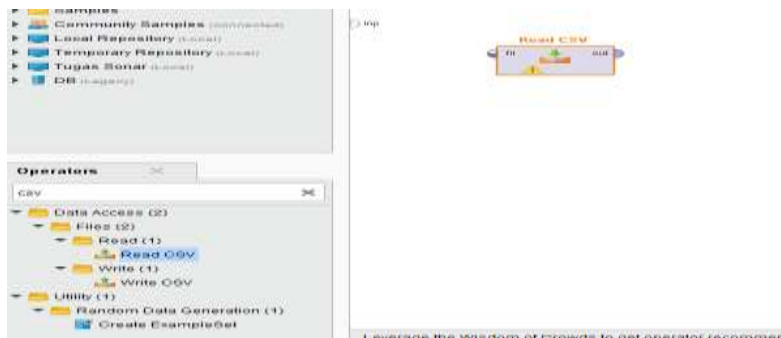
Pada tahap ini pemodelan dilakukan dengan menggunakan Rapidminer versi 9.10 dengan langkah sebagai berikut:

1. Buka rapidminer 9.10 lalu klik logo kertas “new process”



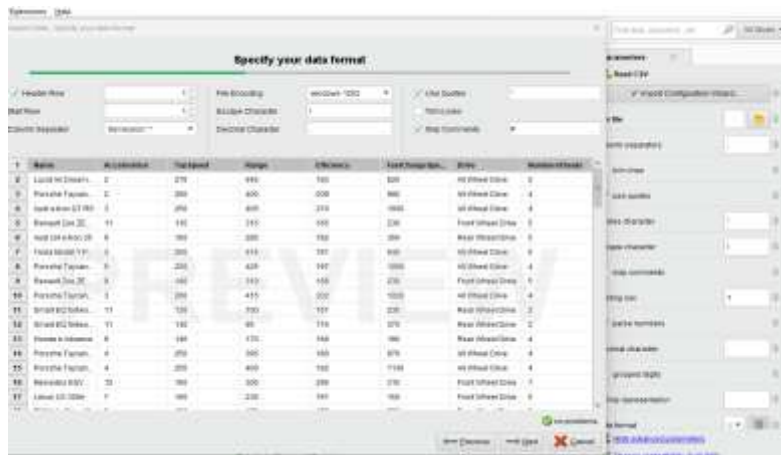
Gambar 3. Klik Logo Kertas

2. Pada panel *operators* ketik “read csv” lalu klik 2 kali atau drag ke panel *process*



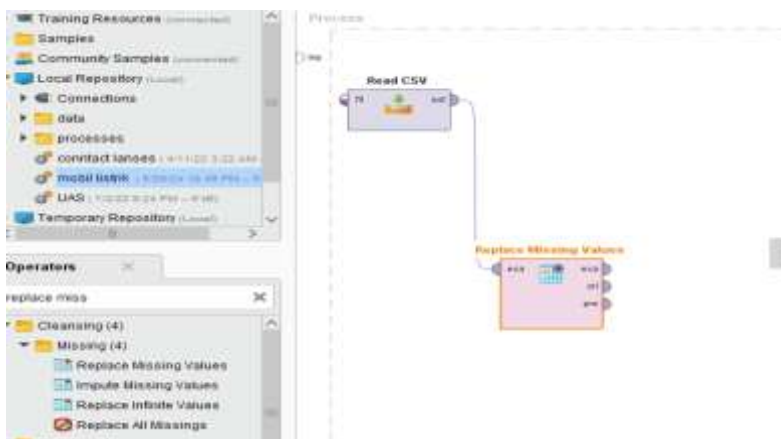
Gambar 4. Pilih *Operators read csv*

3. *Import Data csv* pada panel *parameters*, dengan klik *read csv* file kemudian pilih file *excel* yang akan diuji



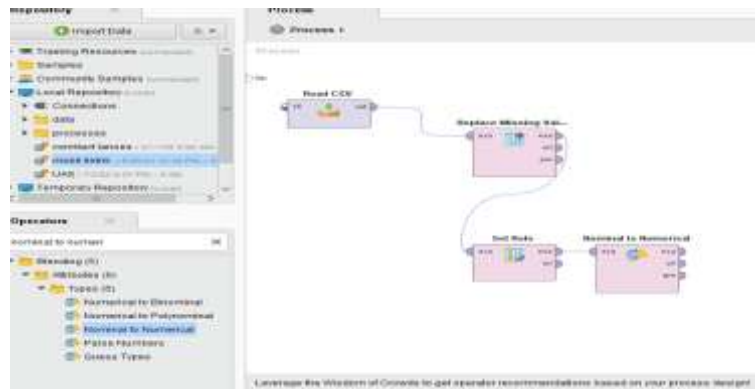
Gambar 5. Isi Dataset

4. Selanjutnya kita cari di *operators* “*Replace Missing Values*” fungsinya itu untuk mengganti variabel-variabel yang hilang di salah satu atribut.



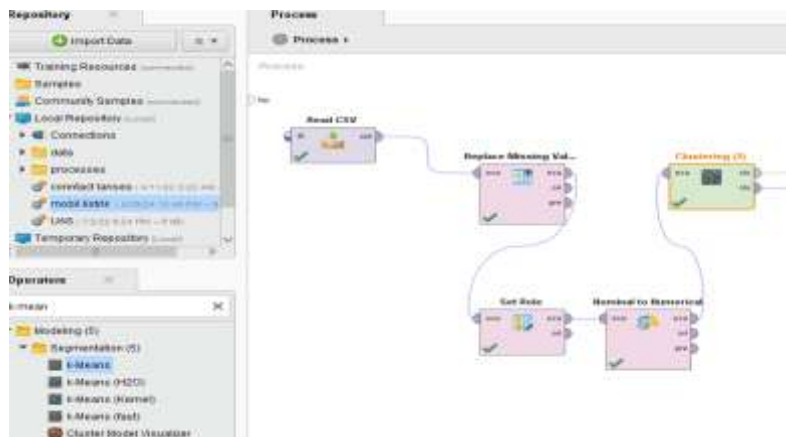
Gambar 6. Pilih *Replace Missing Values*

5. Salah satu atribut yang bernama *drive* ini bervariasi *nominal*, sehingga harus diubah menjadi *numeric*. Selanjutnya cari *operators* “*Set Role*” dan “*Nominal to Numerical*” fungsinya yaitu untuk mengubah data dari atribut *drive* dari variabel *nominal* ke variabel *numeric*.



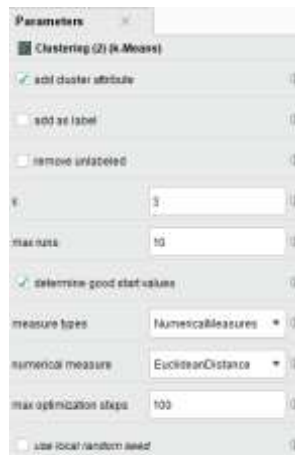
Gambar 7. Pilih *Set Role* dan *Nominal to Numerical*

6. Pada panel *operators* ketik “*K-Means*” lalu klik dua kali atau drag ke panel *process* .



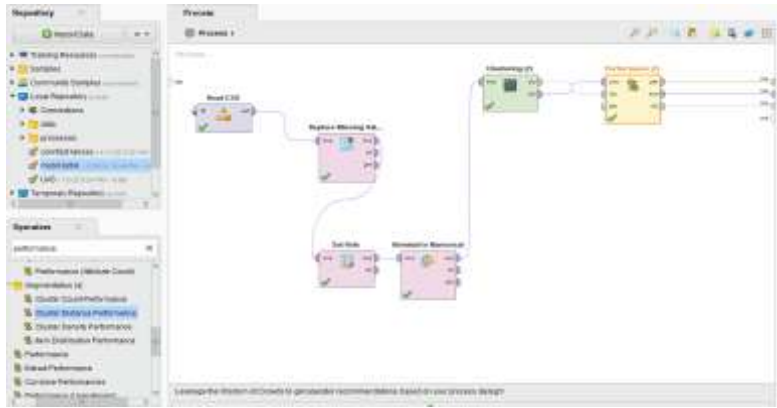
Gambar 8. Pilih Metode *K-Means*

7. Tentukan jumlah *cluster* yang akan digunakan pada panel *parameters*



Gambar 9. Tentukan Jumlah *Cluster*

8. Pada panel *operators* ketik “*Cluster Distance performance*” yang berfungsi untuk melihat performa dataset yang diolah oleh metode *k-means*. lalu klik dua kali atau drag ke panel *process* dan hubungkan semua *connector*.



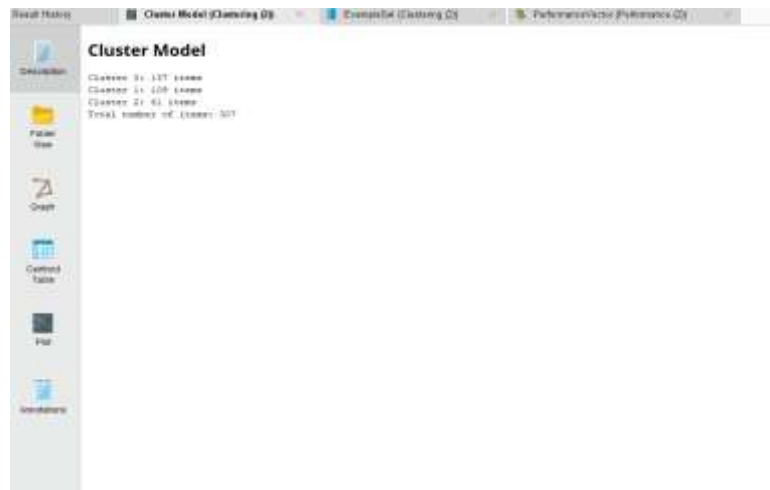
Gambar 10. Pilih *Cluster Distance performance* dan Hubungkan Connector

- Setelah proses *running* selesai maka akan tampil hasil *clustering* dan *cluster model* yang dihasilkan dari *dataset* dan metode yang dipilih.

Item No.	Item	Cluster	Size	Accuracy	Precision	Recall	Silhouette	Entropy	Homogeneity
1	Lexus LS 460	cluster_2	0	2	210	240	100	400	0
2	Porsche Tayc	cluster_2	0	2	250	400	200	800	4
3	Audi A8 D 5	cluster_2	0	2	250	400	210	1000	4
4	Mercedes C63	cluster_2	1	11	130	215	100	200	0
5	Audi Q7 4.0	cluster_2	2	0	190	205	100	300	0
6	Jeep Wrangler	cluster_2	0	2	250	415	100	540	0
7	Porsche Tayc	cluster_2	0	2	220	405	100	1050	4
8	Mercedes C63	cluster_2	1	0	140	210	100	200	0
9	Porsche Tayc	cluster_2	0	2	250	415	100	1000	4
10	Mercedes C63	cluster_2	2	11	130	200	100	200	0
11	Mercedes C63	cluster_2	2	11	130	90	100	210	0
12	Mercedes C63	cluster_2	0	0	140	110	100	100	4
13	Porsche Tayc	cluster_2	0	4	250	205	100	410	4
14	Porsche Tayc	cluster_2	0	4	250	400	100	1120	4
15	Mercedes C63	cluster_2	1	12	150	205	200	210	0
16	Lexus LS 460	cluster_2	1	1	150	200	100	100	0
17	Mercedes C63	cluster_2	2	2	140	410	100	100	0
18	Mercedes C63	cluster_2	0	12	130	130	100	840	4
19	Mercedes C63	cluster_2	1	0	120	200	100	300	4

Gambar 11. Tampilan Hasil *Cluster*

- Berdasarkan pemodelan 3 *cluster* maka diperoleh data yang tergabung ke *cluster 0* : mobil keluarga sebanyak 137 *items*, *cluster 1* : mobil per-kotaan sebanyak 109 *items* dan *cluster 2* : mobil sport sebanyak 61 *items*.



Gambar 12. Tampilan *Cluster Model*

11. Dari tiga *cluster* yang dibentuk diperoleh *centroid* akhir untuk masing-masing *cluster* seperti yang terdapat pada gambar 13.

Attribute	Cluster_0	Cluster_1	Cluster_2
Dimensi	6.761	1.207	6.294
Akselerasi	3.999	15.193	4.479
TopSpeed	166.944	147.328	224.957
Range	489.799	348.228	499.839
Efisiensi	161.897	276.199	166.391
FuelConsumption	166.028	296.174	164.866
BeratKendaraan	9.481	8.766	4.769

Gambar 13. Centroid Table

12. Validasi Hasil

Berdasarkan validasi menggunakan metrik *Davies Bouldin Index* pada aplikasi Rapidminer maka diperoleh dua *cluster* yang dinilai dapat mengelompokan data dengan baik. Performance Vector hasil evaluasi *cluster* yang dibentuk sebanyak dua *cluster* dengan nilai 0,558 yang dapat dilihat pada gambar 14.

Gambar 14. Performance Vector

Sebelumnya telah dilakukan pengujian dari  $k= 3$  sampai dengan  $k= 6$  maka didapat *cluster* yang paling optimal yaitu  $k= 3$  dengan nilai *Davies Bouldin* sebesar 0,641. *Davies Bouldin Index* dapat dikatakan baik jika nilai semakin kecil atau minimal maka hasil clustering yang didapat semakin baik (Sopyan et al., 2022).

Tabel 16. Hasil DBI pengujian

Cluster	Nilai Davies Bouldin
K= 3	0,641
K= 4	0,659
K= 5	0,781
K= 6	0,844

Hasil dari analisis ini menunjukkan bahwa mobil listrik dapat dibagi menjadi tiga *cluster* utama, yaitu *Cluster 0*: mobil keluarga, *Cluster 1*: mobil per-kotaan, dan *Cluster 2*: mobil sport. Setiap *cluster* memiliki karakteristik unik yang membedakan mereka satu sama lain. Berikut table 17 adalah kesimpulan dari penelitian ini berdasarkan hasil clustering dan visualisasi pada gambar.

Tabel 17. Kesimpulan Penelitian

Cluster	Kesimpulan
0 (Mobil Keluarga )	Jumlah item : 137 Karakteristik : Mobil listrik dalam cluster ini memiliki ciri- ciri yang sesuai dengan kebutuhan keluarga. Mobil-mobil ini mungkin memiliki ukuran yang lebih besar, kapasitas tempat duduk yang lebih banyak, dan fitur kenyamanan serta keamanan yang sesuai untuk perjalanan keluarga.
1 (Mobil Perkotaan)	Jumlah item : 109 Karakteristik : Mobil listrik dalam cluster ini dirancang untuk kebutuhan perkotaan. Mereka mungkin memiliki ukuran yang lebih kecil, lebih efisien dalam penggunaan energi, dan cocok untuk perjalanan jarak pendek serta kemudahan parkir di area perkotaan yang padat.
2 (Mobil Sport)	Jumlah item : 61 Karakteristik: Mobil listrik dalam cluster ini memiliki performa tinggi yang sesuai dengan kebutuhan pengendara yang menginginkan kecepatan dan akselerasi tinggi. Mobil-mobil ini mungkin memiliki desain yang sporty dan fitur-fitur canggih untuk meningkatkan pengalaman berkendara.

Penelitian ini berhasil mengelompokkan mobil listrik ke dalam tiga kategori utama berdasarkan karakteristik mereka menggunakan algoritma K-Means.

Hasil ini dapat memberikan panduan bagi produsen mobil dalam mengembangkan dan memasarkan mobil listrik yang sesuai dengan kebutuhan spesifik dari berbagai segmen konsumen. Selain itu, hasil ini juga dapat digunakan oleh pembuat kebijakan untuk merancang strategi yang mendukung adopsi mobil listrik dalam berbagai konteks penggunaan.

Perbedaan antara hasil dari Excel dan *RapidMiner* biasanya terkait dengan presisi perhitungan, otomatisasi, penanganan data, serta algoritma yang dioptimalkan di *RapidMiner*. Excel memberikan lebih banyak kontrol manual namun dengan risiko perhitungan yang tidak sepenuhnya akurat atau optimal, sementara *RapidMiner* secara otomatis menangani berbagai aspek penting dari proses analisis, sehingga menghasilkan perbedaan hasil.

Perhitungan optimal adalah tentang menemukan solusi terbaik berdasarkan metrik yang ditentukan, dan dalam kasus algoritma seperti *K-Means*, ini biasanya dicapai ketika perbedaan antara iterasi sudah minimal, dan hasilnya sudah memuaskan secara statistik.

## 5. Kesimpulan

Berdasarkan hasil analisis dan pembahasan tentang segmentasi mobil listrik maka dapat disimpulkan bahwa Penelitian ini berhasil mengelompokkan mobil listrik berdasarkan fitur dan spesifikasi menggunakan algoritma K-Means. Proses clustering menghasilkan tiga klaster utama yang mencerminkan segmentasi pasar mobil listrik, yaitu: Cluster 0: Mobil Keluarga yang terdiri dari 137 item. Mobil dalam klaster ini memiliki karakteristik yang cocok untuk penggunaan keluarga dengan kapasitas kursi yang lebih besar, jangkauan yang cukup jauh, dan efisiensi yang baik. Cluster 1: Mobil Perkotaan yang terdiri dari 109 item. Mobil dalam klaster ini memiliki karakteristik yang sesuai untuk penggunaan di perkotaan, dengan ukuran yang lebih kecil, akselerasi yang baik, dan efisiensi yang tinggi dalam jarak tempuh yang lebih pendek. Cluster 2: Mobil Sport yang terdiri dari 61 item. Mobil dalam klaster ini memiliki karakteristik performa tinggi, dengan kecepatan puncak yang tinggi, akselerasi cepat, dan desain yang lebih sporty. Hasil evaluasi kualitas klaster menggunakan Davies-Bouldin Index menunjukkan nilai sebesar 0,641. Nilai ini menunjukkan bahwa klaster yang dihasilkan memiliki tingkat kompak dan terpisah yang cukup baik, sehingga dapat disimpulkan bahwa algoritma K-Means mampu mengelompokkan mobil listrik dengan baik berdasarkan fitur dan spesifikasi yang dianalisis.

## Referensi

1. Sano, V. D. (2019, January 15). *Cara Kerja Data Mining – Seri Data Mining for Business Intelligence* . Binus. <https://binus.ac.id/malang/2019/01/cara-kerja-data-mining-seri-data-mining-for-business-intelligence-3/>
2. Krisna, A. (2020, March 3). *Agar Mobil Listrik Benar-benar Ramah Lingkungan*. Www.Kompas.Id. <https://www.kompas.id/baca/ilmu-pengetahuan-teknologi/2020/03/03/agar-mobil-listrik-benar-benar-ramah-lingkungan>
3. Nayoan, A. (2022, May 9). *Apa Itu Data Mining? Manfaat, Proses, dan Contohnya*. Niagahoster. <https://www.niagahoster.co.id/blog/data-mining-adalah/>

4. Fahlevi, A. (2021, September 30). *Proses Data Mining KDD*. Binus. <https://sis.binus.ac.id/2021/09/30/proses-data-mining-kdd/>
5. Sunggara, L. A., & Lengkap, N. (2019). *Artikel Skripsi Universitas Nusantara PGRI Kediri Yang bertanda tangan dibawah ini*. <http://kaggle.com>.
6. Bhat, F. A., Verma, M., & Verma, A. (2024). Who will buy electric vehicles? Segmenting the young Indian buyers using cluster analysis. *Case Studies on Transport Policy*, 15. <https://doi.org/10.1016/j.cstp.2024.101147>
7. Butsianto, S., & Mayangwulan, N. T. (2020). Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, 3(3).
8. Cahya. (2019, August 28). *Data Mining Algoritma Hierarchical Clustering*. <https://Extra.Cahyadsn.Com/>. <https://extra.cahyadsn.com/hierarchical.clustering>
9. Novita, C. (2022, July 19). *Apa Penyebab Pemanasan Global dan Bagaimana Dampaknya?* <https://Tirto.Id/>. <https://tirto.id/apa-penyebab-pemanasan-global-dan-bagaimana-dampaknya-gudm>
10. elfanmauludi. (2023, December 29). *Algoritma K-Means dan Implementasinya*. PENELITIAN.ID. <https://www.penelitian.id/2023/12/algoritma-k-means-dan-implementasinya.html>
11. envihsa.fkm.ui.ac.id. (2022, November 25). *Mobil Listrik: Persoalan atau Pemecahan Masalah?* <https://envihsa.fkm.ui.ac.id/2022/11/25/mobil-listrik-persoalan-atau-pemecahan-masalah/>
12. Fahmi. (2022, October 19). *RapidMiner, Definisi dan Fitur-Fitur, dan Tutorialnya*. Idmetafora. <https://idmetafora.com/news/read/2061/RapidMiner-Definisi-dan-Fitur-Fitur-dan-Tutorialnya.html>
13. Febriansyah, A. R., & Arifin, T. (2023). *Penerapan Algoritma K-Means Clustering Pada Agent Video Game Valorant*. 4(1).
14. Firdaus, N., Defit, S., & Widi Nurcahyo, G. (2023). Penerapan Algoritma K-Means Clustering Dalam Pengelompokan Data Penjualan Pada Pabrik Mobil Toyota Indonesia. *Fountain of Informatics Journal*, 8(2), 2548–5113. <https://doi.org/10.21111/fij.v8i2.10798>
15. Gede, A., Pradnyana, S., Kom, M., Kom, K., Agustini, S., & Si, M. S. (2022). *Konsep Dasar Data Mining*.
16. Harahap, P. N., & Sulindawaty, S. (2020). Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT.Arma Anugerah Abadi Cabang Sei Rampah). *MATICS*, 11(2), 46. <https://doi.org/10.18860/mat.v1i1i2.7821>
17. Hidayati, R., Zubair, A., Hidayat Pratama, A., Indana, L., Studi Sistem Informasi, P., & Teknologi Informasi, F. (2021). *Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering* (Vol. 20, Issue 2).
18. Abdi, H. (2023, May 15). *Clustering adalah Metode Pengelompokan Data, Berikut Penjelasannya*. Liputan6. <https://www.liputan6.com/hot/read/5287033/clustering-adalah-metode-pengelompokan-data-berikut-penjasannya?page=4>
19. Sabrina, I. Z. (2022, March 21). *Pemanfaatan Bahan Bakar Fosil dan Dampaknya bagi Lingkungan*. Solarindustri.Com. <https://solarindustri.com/blog/bahan-bakar-fosil/>
20. Kubiczek, J., & Hadasik, B. (2021). Segmentation of passenger electric cars market in Poland. *World Electric Vehicle Journal*, 12(1), 1–20. <https://doi.org/10.3390/wevj12010023>
21. Prastya, M. (2022, March 31). *Pengertian Apa Itu Mobil Listrik Dan Jenis-Jenisnya*. Carmudi.Co.Id. <https://www.carmudi.co.id/journal/pengertian-apa-itu-mobil-listrik-dan-jenis-jenisnya/>
22. Napizahni, M. (2023, June 6). *Data Mining: Pengertian, Metode, Fungsi, & Penerapannya*. [Www.Dewaweb.Com](http://Www.Dewaweb.Com).
23. Galela, M. R. (2023, November 21). *Koefisien Silhouette untuk Menentukan Jumlah Klaster Ideal*. Kemenkeu. <https://klc2.kemenkeu.go.id/kms/knowledge/koefisien-silhouette-untuk-menentukan-jumlah-klaster-ideal-d0e05a09/detail/>
24. Nazari, M., Hussain, A., & Musilek, P. (2023). Applications of Clustering Methods for Different Aspects of Electric Vehicles. In *Electronics (Switzerland)* (Vol. 12, Issue 4). MDPI. <https://doi.org/10.3390/electronics12040790>
25. Tomba, S. P., & Kesuma, F. T. (2019). PENERAPAN DATA MINING UNTUK MENENTUKAN PENJUALAN SPAREPART TOYOTA DENGAN METODE K-MEANS CLUSTERING. *Jurnal Sistem Informasi Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2).
26. Pojiah. (2022, August 31). *Mengenal Clustering: Sejarah, Pengertian, Jenis, dan Cara Kerjanya*. Idmetafora. <https://idmetafora.com/news/read/830/Mengenal-Clustering-Sejarah-Pengertian-Jenis-dan-Cara-Kerjanya.html>
27. Pulungan, W., Poningsih, P., & Satria, H. (2019). PENGELOMPOKAN PADA KENDARAAN BERMOTOR MENURUT KEGUNAANNYA MENGGUNAKAN METODE DATA MINING K-MEANS. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1). <https://doi.org/10.30865/komik.v3i1.1687>
28. Qotrun A. (2019, May 13). *Penelitian Eksperimen: Pengertian, Karakteristik, Subjek, Prosedur, Kelebihan dan Kekurangannya*. <https://Www.Gramedia.Com/>. <https://www.gramedia.com/literasi/desain-penelitian/>
29. Fahmi, R. N. (2021, February 7). *Algoritma K-Means Clustering*. Medium. <https://restinf21.medium.com/algoritma-k-means-clustering-70625a84eb5>
30. Maulid, R. (2023, January 7). *RapidMiner: Tools Data Science Andalan Data Experts*. Dqlab. <https://dqlab.id/rapidminer--tools-data-science-andalan-data-experts>
31. Tetlageni, M. R., & Solichin, A. (2023). *KLASTERISASI PENYEWAAN KENDARAAN MENGGUNAKAN METODE K-MEANS PADA PT. MARDIKA DAYA TRIBUANA VEHICLE RENTAL CLUSTERING USING THE K-MEANS METHOD AT PT. MARDIKA DAYA TRIBUANA* (Vol. 20, Issue 2).
32. Setiawan, R. (2021, October 30). *Apa itu Data Mining dan Bagaimana Metodenya?* [Www.Dicoding.Com](http://Www.Dicoding.Com).
33. Kristian, A., Serasi Ginting, B., & Ramadani, S. (2022). *Data Mining Pengujian Kendaraan Bermotor Berdasarkan Jenis Kendaraan menggunakan Metode K-Means Studi Kasus Dinas Perhubungan Kota Binjai A B S T R A K* Sejarah artikel. 2(1), 6–12.
34. Serbi, S. (2020, August 30). *Apa itu Data Mining: Metode dan Penerapannya*. IDCloudHost. <https://idcloudhost.com/blog/apa-itu-data-mining-definisi-fungsi-metode-dan-penerapannya/#comments>
35. Sman, D. I., Rahmawati, B., & Arifin, T. (2020). PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN SISWA LOLOS SNMPTN. *JURNAL RESPONSIF*, 2(2), 184–190. <http://ejurnal.ars.ac.id/index.php/jti>
36. Sopyan, Y., Lesmana, A. D., & Juliane, C. (2022). Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan. *Building of Informatics, Technology and Science (BITS)*, 4(3). <https://doi.org/10.47065/bits.v4i3.2697>
37. Triyatna, S. O. (2021, September 27). *Tingkat Ketergantungan Energi Fosil Masih Tinggi*. [Www.Kompas.Id/](http://Www.Kompas.Id/).
38. Sutiono, S. (2023, June 4). *Algoritma K-Means: Pengertian, Cara Kerja, Kelebihan dan Contoh*. DosenIT.Com. <https://dosenit.com/kuliah-it/algoritma-k-means>
39. Yusuf, D., Sestri, E., Razi, F., Informasi, T., & Dahlan Jakarta, A. (2023). Volume 6; Nomor 2. *Juli*, 307–313. <https://ojs.trigunadharna.ac.id/index.php/jsk/index>