

## KLASTERISASI PENYEWAAN KENDARAAN MENGGUNAKAN METODE K-MEANS PADA PT. MARDIKA DAYA TRIBUANA

Muhamad Ridho Tetlageni<sup>1\*</sup>, Achmad Solichin<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur  
Email: <sup>1</sup>1911502159@student.budiluhur.ac.id, <sup>2</sup>achmad.solichin@budiluhur.ac.id

(\*: Corresponding Author)

(Naskah masuk: 2 Agustus 2023, diterima untuk diterbitkan: 25 Agustus 2023)

### Abstrak

Pada zaman sekarang kebutuhan perusahaan terhadap moda transportasi semakin meningkat, untuk itu perusahaan menginginkan kendaraan mobil untuk menjalankan kegiatan sehari-hari seperti kebutuhan operasional perusahaan dan lain-lain. Setiap perusahaan dalam memilih kendaraan mobil yang disewa pasti mencari kendaraan yang nyaman untuk dipakai. PT. Mardika Daya Tribuana merupakan salah satu perusahaan yang berfokus pada bidang *penyewaan*. Masalah yang terjadi pada PT. Mardika Daya Tribuana yaitu data penyewaan yang sangat banyak tidak bisa diolah, sehingga pihak perusahaan belum menemukan suatu metode untuk mengolah data agar menemukan pola dari data yang ada. Maka pada penelitian ini penulis bertujuan untuk mengelompokkan kendaraan-kendaraan yang dimana hasil dari pengelompokan dapat digunakan oleh pihak PT. Mardika Daya Tribuana untuk mengambil keputusan yang tepat sesuai kebutuhan yang diinginkan. Tentunya untuk mengolah data penyewaan pada PT. Mardika Daya tribuana diperlukan suatu metode yang dapat menghasilkan suatu keputusan yang tepat. Pada penelitian ini penulis menerapkan analisa data *Mining* dengan Teknik *Clustering* menggunakan algoritma *K-Means*. Metode *K-Means Clustering* digunakan dan diolah berdasarkan kebutuhan untuk mendapatkan informasi pada setiap kelompoknya. Hasil kluster pada data penyewaan sebanyak 280 baris data dan dibagi menjadi 3 (tiga) kluster dimana pada setiap kluster menghasilkan, 11 data dengan level sedang rendah dari segi peminatnya pada kluster 1, 113 data dengan level sedang dari segi peminatnya pada kluster 2, dan 156 data dengan level tinggi dari segi peminat pada kluster 3. Berdasarkan pengujian menggunakan *Sum of Square Error* (SSE) kluster yang direkomendasikan yaitu 3 kluster dengan nilai selisih 10.44. Sehingga kluster tersebut optimal untuk dipakai.

**Kata kunci:** *data mining, clustering, k-means, penyewaan, SSE*

## VEHICLE RENTAL CLUSTERING USING THE K-MEANS METHOD AT PT. MARDIKA DAYA TRIBUANA

### Abstract

*In this day and age the company's need for transportation modes is increasing, for that the company wants a car vehicle to carry out daily activities such as the company's operational needs and others. Every company in choosing a rented car vehicle must be looking for a vehicle that is comfortable to use. PT. Mardika Daya Tribuana is one of the companies that focuses on the field of rental. Problems that occur at PT. Mardika Daya Tribuana is that a lot of rental data cannot be processed, so the company has not found a method to process the data to find patterns from existing data. So, in this study the author aims to group vehicles that are mandated as a result of grouping can be used by PT. Mardika Daya Tribuana to make the right decision according to the desired needs. Of course, to process rental data at PT. Mardika Daya tribuana needed a method that could produce the right decision. In this study, the author applies data mining analysis with clustering techniques using the K-Means algorithm. The K-Means Clustering method is used and processed based on the need to obtain information in each group. The cluster results in rental data as many as 280 rows of data and divided into 3 (three) clusters in each cluster produce, 11 data with medium low levels in terms of interest in cluster 1, 113 data with medium levels in terms of enthusiasts in cluster 2, and 156 data with high levels in terms of enthusiasts in cluster 3. Based on testing using the recommended cluster Sum of Square Error (SSE), there are 3 clusters with a difference value of 10.44. So that the cluster is optimal for use.*

**Keywords:** *data mining, clustering, k-means, leasing, SSE*

## 1. PENDAHULUAN

Pada zaman sekarang kebutuhan perusahaan terhadap Moda transportasi semakin meningkat, untuk itu perusahaan menginginkan kendaraan mobil untuk menjalankan kegiatan sehari-hari seperti kebutuhan operasional perusahaan dan lain-lain. Setiap perusahaan pasti menginginkan mobil yang nyaman untuk digunakan dalam beraktivitas, dan saat memilih mobil pun mereka pasti akan mencari mobil yang nyaman untuk digunakan. Berbagai merek mobil banyak dijual dan disewakan, dengan berbagai jenis tipe mobil. Mobil sendiri memiliki banyak merek di antaranya Mitsubishi, Toyota, Suzuki, Honda, dan lain sebagainya.

PT. Mardika Daya Tribuana merupakan salah satu perusahaan yang berfokus pada bidang jasa penyewaan, dan penjualan mobil bekas. Masalah yang terjadi pada PT. Mardika Daya Tribuana yaitu data penyewaan yang sangat banyak tidak bisa diolah, sehingga pihak perusahaan belum menemukan suatu metode untuk mengolah data agar menemukan pola dari data yang ada. Kemudian data dari penyewaan semestinya dapat dijadikan pengetahuan baru dan dapat dimanfaatkan untuk bisa mengambil keputusan yang tepat sesuai kebutuhan bagi PT. Mardika Daya Tribuana ke depannya.

Dengan adanya kegiatan penyewaan pada PT. Mardika Daya Tribuana, data penyewaan semakin lama semakin bertambah banyak dan membengkak. Data tersebut disimpan di dalam *excel* dan tidak dimanfaatkan, maka penulis mengambil keputusan untuk mengolah data penyewaan pada PT. Mardika Daya Tribuana. Tentunya untuk mengolah data penyewaan pada PT. Mardika Daya Tribuana diperlukan suatu metode yang dapat menghasilkan suatu keputusan yang tepat. Pada penelitian ini penulis akan menerapkan analisa data *Mining* dengan Teknik *Clustering* menggunakan algoritma *K-Means*.

Data Mining merupakan suatu proses mencari pola-pola yang khas dari data yang tersebar dalam jumlah besar, dengan tujuan mengidentifikasi hubungan antara data yang berbeda dan membentuk pengetahuan atau informasi baru. Data tersebut biasanya diarsipkan dalam bentuk digital pada berbagai sistem komputer, seperti *database*, data warehouse, dan media penyimpanan lainnya[1].

Penelitian ini berpatokan pada beberapa penelitian sebelumnya yang terkait dengan *K-Means Clustering* untuk berbagai kebutuhan antara lain untuk menganalisis data kemiskinan di Jawa Barat [2], klasterisasi persebaran Virus Corona di DKI Jakarta[3], klasterisasi kasus Covid-19 di Provinsi Lampung[4], pengelompokan keluarga miskin Desa Bina Baru[5], pengelompokan kondisi gizi balita pada posyandu[6], dan lain-lain.

Penelitian yang pernah dilakukan oleh [7] menggunakan algoritma *K-Means Clustering* untuk memprediksi penjualan mobil. Penelitian ini bertujuan untuk mengelompokkan data penjualan ke

dalam sebuah *Cluster* dengan metode data *mining* Algoritma *K-Means Clustering*. Hasil dari penelitian memakai tiga *Cluster* atau  $K=3$  yang didapatkan terbagi menjadi 3 *Cluster* dimana terbagi menjadi tiga kategori yaitu, kategori Laris untuk *Cluster* 0, kategori Kurang Laris untuk *Cluster* 1, dan kategori Paling Laris untuk *Cluster* 2. Masing-masing *Cluster* terbentuk sehingga menghasilkan nilai dan persentase yaitu, *Cluster* 0 = 235;26%, *Cluster* 1 = 604;67%, *Cluster* 2 = 61;7%. Proses *Clustering* diatas divalidasi oleh Davies Bouldin Index (DBI) dengan nilai 0,341.

Keterkaitan antara penelitian di atas dengan penelitian yang dilakukan sama-sama memakai metode *K-Means Clustering* tetapi peneliti menambahkan metode Normalisasi min-max untuk penyebaran data yang merata dan menggunakan metode *Elbow* untuk penentuan jumlah klaster terbaik berdasarkan teknik SSE.

Klasterisasi merupakan salah satu teknik yang digunakan dalam data mining. Tujuannya adalah mengelompokkan sejumlah data menjadi kelompok-kelompok tertentu (klaster). Dalam proses klasterisasi, tahap yang sangat krusial adalah menentukan atau mendeskripsikan nilai kuantitatif dari tingkat kemiripan atau ketidakmiripan data, yang juga dikenal sebagai ukuran kedekatan (*proximity measure*). Oleh karena itu, perbandingan beberapa metode yang umum digunakan menjadi sangat penting. Beberapa metode tersebut mencakup perhitungan jarak Euclidean, Manhattan, dan Minkowski[8].

Algoritma *K-Means* adalah salah satu algoritma *Clustering* yang sering digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristiknya. Proses ini menghasilkan kelompok data yang disebut sebagai *Cluster* atau klaster. Kelebihan dari algoritma ini adalah kemudahan dalam implementasi serta kompleksitas waktu dan ruang yang relatif rendah. Hal ini membuatnya menjadi efisien dalam komputasinya dan menghasilkan hasil yang memuaskan jika klasternya kompak, berbentuk hiper-spherical, dan mampu memisahkan fitur-fitur ruangnya dengan baik[9].

Klasterisasi *K-Means* sangat bergantung pada proses perhitungan jarak. Dengan demikian standarisasi data atau nilai atribut menjadi salah satu faktor penting yang mempengaruhi kualitas hasil klasterisasi. Salah satu metode untuk menstandarisasi data adalah dengan proses normalisasi. Normalisasi Min-Max adalah sebuah metode dalam proses data yang menggunakan nilai maksimum dan minimum dari suatu atribut untuk mengubah data ke dalam rentang baru secara linier, di mana nilai minimum akan diubah menjadi 0 dan nilai maksimum akan diubah menjadi 1. Tujuan dari normalisasi ini adalah untuk menciptakan perbandingan antara data yang seimbang, sehingga data-data tersebut memiliki skala yang serupa dalam rentang 0 hingga 1[10].

Pada penelitian ini, diterapkan metode klusterisasi K-Means dengan metode perhitungan jarak euclidean dan penerapan normalisasi min-max. Selain itu, metode *Elbow* juga digunakan untuk menentukan jumlah kluster yang paling optimal. Proses metode ini melihat presentasi hasil perbandingan dari jumlah total K dan akan memberikan sebuah lekukan pada grafik dan akan disebut siku[11]. Penelitian ini memiliki kontribusi praktis bagi PT Mardika Daya Tribuana dalam membantu keputusan dan strategi pemasaran produk dan jasanya.

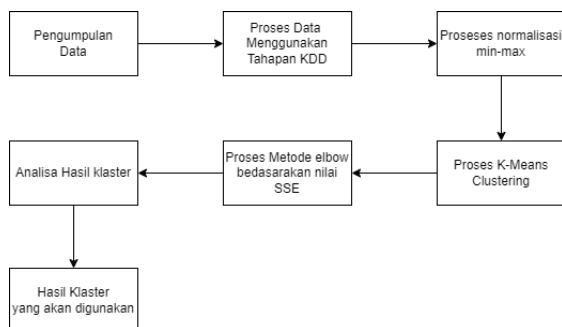
## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Data penelitian yang digunakan dalam penelitian ini adalah hasil informasi atau data murni tanpa adanya manipulasi sehingga data dalam penelitian ini bersumber dari data penyewaan PT. Mardika Daya Tribuana. Jenis data yang diberikan berupa *file worksheet* dengan ekstensi xls, jumlah keseluruhan data pada bulan 4 Januari 2021 – 29 maret 2023 adalah sebanyak 299 data. Dengan atribut sebanyak 13 yaitu Tahun, Bulan, Nama Perusahaan, Merk, Tipe Kendaraan, Nomor Polisi, Tahun, Masa Sewa(bulan), Mulai Sewa, Akhir Sewa, Harga Sewa/Bulan, Marketing, Keterangan.

### 2.2 Penerapan Metode

Pada penelitian ini penerapan metode untuk mengolah data yang didapatkan dari PT. Mardika Daya Tribuana dengan menerapkan *data mining* menggunakan metode K-Means *Clustering*. Terdapat beberapa tahapan yang dibuat dan bisa dilihat pada Gambar 1 yang merupakan proses bagaimana data penyewaan yang ada pada PT. Mardika Daya Tribuana diolah dan kemudian mengelompokkan kendaraan mana saja yang di minati oleh konsumen dari segi penyewaannya.



Gambar 1 Tahapan Penelitian

Berikut penjelasan tiap tahapan penerapan metode yang dilakukan pada penelitian ini

#### 1) Pengumpulan Data

Pada tahap ini peneliti melakukan pengumpulan data yang akan diolah untuk menjadi kebutuhan penelitian. Proses pengumpulan data terbagi menjadi 2 tahap yaitu Studi Literatur dan Wawancara.

#### 2) Proses Data Menggunakan Tahapan KDD

Data diolah manual menggunakan tahapan KDD yaitu, Data Selection, Preprocessing/Cleaning, Transformation, Data mining.

#### 3) Proses Normalisasi Min-Max

Pada tahap ini setelah *dataset*, jumlah *Cluster*, dan titik *centroid* yang telah melalui proses yang sudah ditetapkan, maka proses normalisasi diperlukan agar sebaran data yang dihasilkan merata. Dengan menggunakan rumus:

$$x^1 = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Dimana:

$x^1$  = Data atribut yang akan dinormalisasikan

$x_{min}$  = Nilai terkecil atribut tersebut

$x_{max}$  = Nilai tertinggi atribut tersebut

#### 4) Proses K-Means Clustering

Pada tahap ini setelah *dataset* dan titik *centroid* sudah di normalisasi, penentuan jumlah kluster dilakukan agar dapat mengelompokkan dari data yang ada. Berikut tahapan algoritma K-Means[12].

- Tentukan  $k$  sebagai jumlah *Cluster* yang akan dibentuk
- Tentukan  $k$  *Centroid* (titik pusat *Cluster*) awal secara acak. Adapun rumus yang digunakan untuk menghitung *centroid Cluster* ke-1 berikutnya sebagai berikut:

$$v = \frac{\sum_{i=1}^n x_i}{n}; i = 1, 2, 3, \dots, n \quad (2)$$

Dimana:

$V$  = *centroid* pada *Cluster*

$x_i$  = objek ke- $i$

$n$  = banyaknya objek / jumlah objek yang menjadi anggota *Cluster*

- Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *Cluster*. Untuk menghitung jarak antara objek dengan *centroid* dapat menggunakan *Euclidean Distance*.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1, 2, 3, \dots, n \quad (3)$$

Dimana:

$x_i$  = objek  $x$  ke- $i$

$y_i$  = data  $y$  ke- $i$

$n$  = banyaknya objek

- Alokasikan masing-masing objek ke dalam *centroid* yang paling dekat.

- e. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (2).
- f. Ulangi langkah ke 3 jika posisi *centroid* baru tidak sama.

**5) Proses Metode Elbow**

Pada tahap ini fungsi dari SSE adalah menentukan jumlah K atau klaster yang terbaik. Peneliti memberikan batasan jumlah klaster minimal 2 dan jumlah maksimal klaster 6. Berikut persamaan SSE disajikan[3]

$$SE = \sum_{k=1}^k \sum_{x_i \in S_k} ||x_i - c_k|| \quad (4)$$

Dimana:

- K = jumlah klaster
- $x_i$  = data k i
- $c_k$  = *centroid* klaster

**6) Analisa Hasil Klaster**

Setelah didapatkan hasil SSE, tahap selanjutnya melakukan analisis klaster yaitu mencari klaster yang memiliki nilai selisih SSE yang terbesar maka klaster tersebutlah yang paling optimal untuk digunakan.

**2.3 Rancangan Layar**

Rancangan layar adalah rancangan tampilan atau kata lain bisa digunakan sebagai antarmuka antara pengguna (*user*) dan aplikasi (sistem), atau biasa disebut sebagai *user interface* (UI).

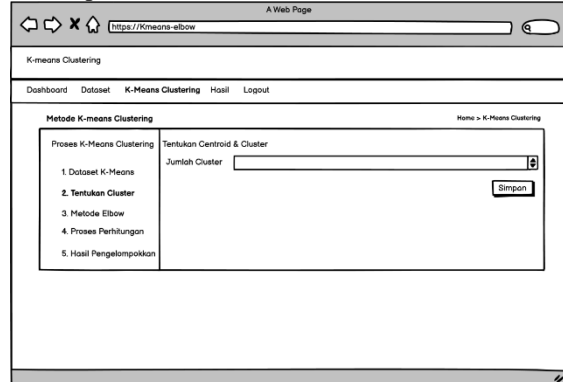
Rancangan layar menu *dataset* ini akan tampil setelah admin memilih menu *dataset*, admin bisa menambahkan atau *upload* data berupa *excel* melalui *button Choosen File* selain itu, admin juga dapat menambahkan data secara manual, mengubah dan menghapus data sesuai kebutuhan. Tampilan bisa dilihat pada Gambar 2

Pilihan	Nomor Polisi	Tahun Kendaraan	Merak	Tipe Mobil	Jenis Kendaraan	Status Kendaraan	Nilai Rata-Rata
<input type="checkbox"/>	B 21696 SJD	2020	Toyota	Fortuner VRZ 2.4 A/T Diesel	SUV	Baru	1250000
<input type="checkbox"/>	B 9513 SCF	2017	Daihatsu	Granmax BlindVan 1.3 M/T	Minibus	Baru	2454545
<input type="checkbox"/>	B 9515 SCF	2017	Daihatsu	Granmax BlindVan 1.3 M/T	Minibus	Baru	2454545
<input type="checkbox"/>	B 9587 SCF	2017	Daihatsu	Granmax BlindVan 1.3 M/T	Minibus	Baru	2454545
<input type="checkbox"/>	B 9396 SCF	2017	Daihatsu	Granmax BlindVan 1.3 M/T	Minibus	Baru	2454545
<input type="checkbox"/>	B 9575 SCF	2017	Daihatsu	Granmax BlindVan 1.3 M/T	Minibus	Baru	2454545
<input type="checkbox"/>	B 9233 SCG	2018	Mitsubishi	FE 71 Long 110 PS	Truck	Baru	4818182
<input type="checkbox"/>	B 9514 SCF	2016	Daihatsu	Granmax BlindVan 1.3 M/T	Minibus	Baru	2500000
<input type="checkbox"/>	B 2702 S01	2016	Toyota	Avanza 1.3 E M/T	MPV	Baru	3000000

Gambar 2 Rancangan Layar Dataset

Menu Tentukan *Cluster* merupakan sub menu dari menu *K-Means Clustering*, dimana pada saat admin menekan tombol Tentukan *Cluster* akan menampilkan halaman penentuan *Cluster*. Admin

dapat menentukan *Cluster* sesuai kebutuhan setelah menentukan *Cluster* maka admin akan menekan *button* simpan untuk proses *K-Means*. Tampilan bisa dilihat pada Gambar 3



Gambar 3 Rancangan Layar (Tentukan Cluster)

Menu Proses Perhitungan merupakan sub menu dari menu *K-Means Clustering*, dimana pada saat admin menekan tombol Proses Perhitungan akan menampilkan hasil perhitungan setiap perulangan dan perhitungan jarak menggunakan *euclidian distance*. Tampilan bisa dilihat pada Gambar 4

Centroid 1	Nilai C1-1	Nilai C1-2	Nilai C1-3	Nilai C1-4	Nilai C1-5	Nilai C1-6
Centroid 2	Nilai C2-1	Nilai C2-2	Nilai C2-3	Nilai C2-4	Nilai C2-5	Nilai C2-6

Perulangan X	Hasil C1	Hasil C2
B 21696 SJDToyotaFortuner VRZ 2.4 A/T Diesel2020	Hasil C1	Hasil C2
B 9513 SCFDaihatsuGranmax BlindVan 1.3 M/T2017	Hasil C1	Hasil C2
B 9515 SCFDaihatsuGranmax BlindVan 1.3 M/T2017	Hasil C1	Hasil C2
B 9587 SCFDaihatsuGranmax BlindVan 1.3 M/T2017	Hasil C1	Hasil C2
B 9396 SCFDaihatsuGranmax BlindVan 1.3 M/T2017	Hasil C1	Hasil C2

Gambar 4 Rancangan Layar (Proses Perhitungan)

Menu Hasil Pengelompokan merupakan sub menu dari menu *K-Means Clustering*, dimana pada saat admin menekan tombol Hasil Pengelompokan akan menampilkan hasil setiap *Cluster* dari proses perhitungan pada menu sebelumnya. Tampilan bisa dilihat pada Gambar 5

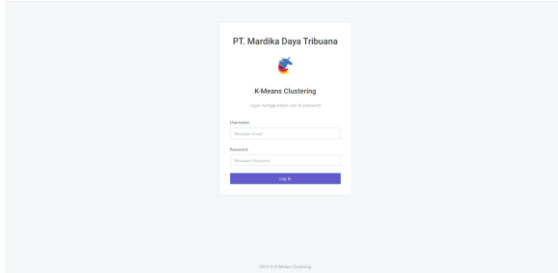
Object	Cluster
B 21696 SJDToyotaFortuner VRZ 2.4 A/T Diesel2020	1
B 9513 SCFDaihatsuGranmax BlindVan 1.3 M/T2017	2
B 9515 SCFDaihatsuGranmax BlindVan 1.3 M/T2017	2
B 9587 SCFDaihatsuGranmax BlindVan 1.3 M/T2017	2
B 9396 SCFDaihatsuGranmax BlindVan 1.3 M/T2017	2

Gambar 5 Rancangan Layar (Hasil Pengelompokan)

### 3. HASIL DAN PEMBAHASAN

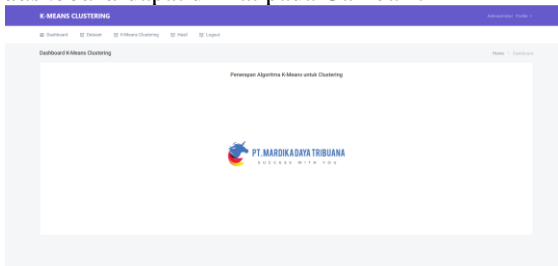
#### 3.1 Tampilan Layar

Berdasarkan rancangan layar dan aplikasi yang sudah dijelaskan pada bagian sebelumnya, dikembangkan aplikasi. Gambar 6 adalah Tampilan layar *login*, halaman yang pertama kali akan muncul kepada admin. Untuk menuju ke halaman selanjutnya, admin diminta untuk memasukkan *username* dan *password* yang benar untuk masuk ke halaman *dashboard*.



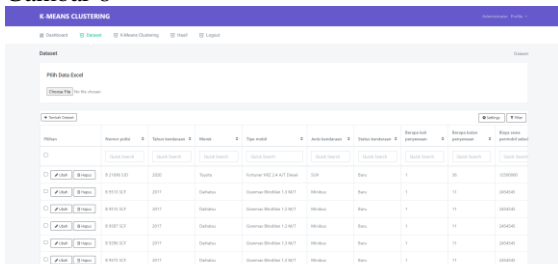
Gambar 6 Tampilan Layar Login

Tampilan layar *dashboard* ini akan tampil setelah admin berhasil *login*, pada halaman *dashboard* akan ditampilkan logo dari perusahaan PT. Mardika Daya Tribuana, dan juga terdapat 5 (Lima) navbar yang tersedia yaitu *Dashboard*, *dataset*, *K-Means Clustering*, *Hasil* dan *Logout*. Tampilan layar *dashboard* dapat dilihat pada Gambar 7



Gambar 7 Tampilan Layar Dashboard

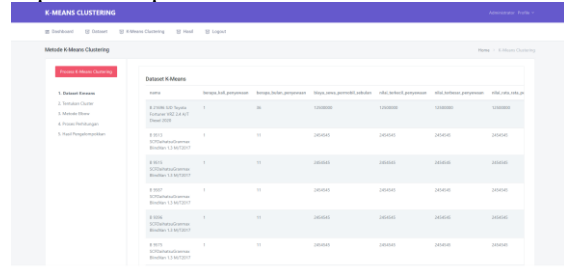
Tampilan layar dataset akan tampil ketika admin menekan tombol menu dataset, pada halaman dataset akan ditampilkan button untuk mengunggah file excel dan setelah diunggah akan tampil dataset, pada halaman ini juga admin bisa menambahkan data secara manual, menghapus dan mengubah data sesuai kebutuhan. Tampilan layar dataset dapat dilihat pada Gambar 8



Gambar 8 Tampilan Layar Dataset

Tampilan layar K-Means Clustering adalah yang menjadi inti dari program ini karna di dalam menu ini terdapat sub menu-sub menu yang menjadi proses

dari K-Means dan metode *Elbow*, ketika admin menekan tombol *dataset* K-Means maka akan muncul *dataset* yang sudah di unggah dan *dataset* yang sudah di Normalisasikan. Tampilan layar *dataset* K-Means dapat dilihat pada Gambar 9



Gambar 9 Tampilan Layar Dataset K-Means

#### 3.2 Implementasi Metode

Setelah data terkumpul, tahapan yang digunakan yaitu data di proses menggunakan *Knowledge Discovery in Database (KDD)*. Proses data akan dijabarkan sebagai berikut di bawah ini:

##### 1) Data Selection

Data penyewaan yang diperoleh dari PT. Mardika Daya Tribuana memiliki 13 atribut dan 299 baris data. Berikut data penyewaan disajikan dalam sampel data pada Tabel 2 sebelum melalui proses data selection.

Tabel 1. Sampel Data

Atribut	Data
No.	1
Tahun	2021
Bulan	Januari
Nama Perusahaan	PT Dharma Harapan Raya
Merk	Toyota
Tipe Kendaraan	Fortuner VRZ 2.4 A/T Diesel
Nomor Polisi	B 21696 SJD
Tahun	2020
Masa Sewa (Bulan)	36
Mulai Sewa	04/01/2021
Akhir Sewa	03/01/2024
Harga Sewa/Bulan	Rp.12,500,000
Marketing	Group
Keterangan	New Car

Dari 13 atribut data dan 299 baris data, penulis hanya menggunakan 6 atribut data dan 280 baris data dari data penyewaan pada PT. Mardika Daya Tribuana, penulis juga melakukan penambahan atribut sebanyak 6. Atribut yang diseleksi dan diambil dari data asli yaitu atribut: “Nomor Polisi”, “Tahun”, “Merk”, “Tipe Mobil”, “Jenis Kendaraan”, “Masa Sewa (Bulan)”. Sedangkan penulis melakukan penambahan atribut sebanyak 6, yaitu atribut: “Status Kendaraan”, “Berapa Kali Penyewaan”, “Berapa Bulan Penyewaan”, “Biaya Sewa Mobil”, “Nilai Terkecil Penyewaan”, “Nilai Terbesar Penyewaan”, “Nilai Rata-Rata Penyewaan”. Tujuan dari seleksi atribut, penambahan atribut, dan seleksi jumlah data dilakukan untuk kebutuhan pengolahan data menggunakan metode *K-Means Clustering*. Berikut

sampel data disajikan setelah melalui proses seleksi pada Tabel 2

Tabel 2. Sampel Data Setelah Seleksi

Atribut	Data
No.	1
nomor_polisi	B 21696 SJD
tahun_kendaraan	2020
merek	Toyota
tipe_mobil	Fortuner VRZ 2.4 A/T Diesel
jenis_kendaraan	SUV
status_kendaraan	Baru
berapa_kali_penyewaan	1
berapa_bulan_penyewaan	36
biaya_sewa_permobil_sebulan	12500000
nilai_terkecil_penyewaan	12500000
nilai_terbesar_penyewaan	12500000
nilai_rata_rata_penyewaan	12500000

## 2) Preprocessing/Cleaning

Pada tahap ini data penyewaan PT. Mardika Daya Tribuana ada beberapa atribut yang tidak digunakan pada penelitian ini, atribut-atribut yang digunakan adalah atribut yang diperlukan untuk melakukan proses K-Means Clustering. Atribut yang tidak digunakan yaitu: Tahun, bulan, nama perusahaan, mulai sewa, akhir sewa, marketing. Atribut-atribut yang digunakan dan penambahan atribut bisa dilihat pada Tabel 2 Penulis juga menemukan data yang rancu pada nomor polisi yaitu: “B 2229 SOK”, “B 2379 SOW”, dan “B 2380 SOW” dan ada beberapa nomor polisi lainnya, yaitu data mengalami duplikat. Penulis juga menemukan merek dan tipe kendaraan “Suzuki”, “Ambulance” 2 buah tidak ada nomer polisinya, maka penulis mengambil langkah, data yang duplikat dan tidak ada nomor polisinya dihapus dengan tujuan agar data bisa diolah pada data mining. Total data yang tadinya ada 299 baris setelah melakukan tahapan cleaning ini berkurang menjadi 280 baris.

## 3) Transformation

Pada tahap ini atribut dan data penyewaan PT. Mardika Daya Tribuana akan disesuaikan dengan kebutuhan proses data *mining* menggunakan *K-Means Clustering*. Contohnya seperti pada atribut **Nomor Polisi** diubah menjadi **nomor\_polisi**, atribut **Tahun** diubah menjadi **tahun\_kendaraan**, atribut **Tipe Kendaraan** diubah menjadi **tipe\_mobil**, atribut **Keterangan** diubah menjadi **status\_kendaraan**, atribut **Masa sewa(bulan)** diubah menjadi **berapa\_bulan\_penyewaan**, atribut **Harga Sewa/Bulan** diubah menjadi **biaya\_sewa\_permobil\_sebulan**, dan penambahan atribut **Jenis kendaraan** diubah menjadi **jenis\_kendaraan**. Format pada isi atribut adapun juga yang diubah seperti pada isi atribut **status\_kendaraan** yang awalnya “**New Car**” menjadi “**baru**”, isi atribut **biaya\_sewa\_permobil\_sebulan**

yang awalnya “**Rp. 12.500.000**” menjadi “**12500000**”, pada penambahan atribut, ada isi data yang diubah mengikuti rujukan pada data aslinya seperti isi pada atribut **berapa\_kali\_penyewaan** isinya berupa numerik dimana itu diambil dari berapa kali penyewaan permobil dari kurun waktu 2021-2023 sesuai data penyewaan pada PT. Mardika Daya Tribuana, isi pada atribut **nilai\_terkecil\_penyewaan** isinya berupa numerik dimana itu diambil dari harga sewa terendah per mobil sesuai data penyewaan pada PT. Mardika Daya Tribuana, isi pada atribut **nilai\_terbesar\_penyewaan** isinya berupa numerik dimana itu diambil dari harga sewa terbesar per mobil sesuai data penyewaan pada PT. Mardika Daya Tribuana, isi pada atribut **nilai\_rata\_rata\_penyewaan** isinya berupa numerik dimana itu diambil dari harga sewa terendah dan terbesar dari per mobil kemudian di rata-ratakan sesuai data penyewaan pada PT. Mardika Daya Tribuana. Tujuan dari transformasi ini adalah untuk kebutuhan dalam proses *K-Means Clustering* yang diimplementasikan pada sistem aplikasi. Berikut sampel data disajikan sebelum di transformasi terdapat pada Tabel 3 dan sampel data disajikan sesudah di transformasi terdapat pada Tabel 4

Tabel 3. Sampel Data Sebelum Transformasi

Atribut	Data
Nomor Polisi	B 21696 SJD
Tahun	2020
Merek	Toyota
Tipe Kendaraan	Fortuner VRZ 2.4 A/T Diesel
Jenis Kendaraan	SUV
Keterangan	New Car
Berapa Kali Penyewaan	1
Masa Sewa (bulan)	36
Harga Sewa Mobil/Bulan	Rp. 125000000
Nilai Terkecil Penyewaan	Rp.125000000
Nilai_Terb Besar Penyewaan	Rp. 125000000
Nilai Rata Rata Penyewaan	Rp. 125000000

Tabel 4. Sampel Data Sesudah Transformasi

Atribut	Data
<b>nomor_polisi</b>	B 21696 SJD
<b>tahun_kendaraan</b>	2020
<b>Merek</b>	Toyota
<b>tipe_mobil</b>	Fortuner VRZ 2.4 A/T Diesel
<b>jenis_kendaraan</b>	SUV
<b>status_kendaraan</b>	Baru
<b>berapa_kali_penyewaan</b>	1
<b>Berapa_bulan_penyewaan</b>	36
<b>biaya_sewa_permobil_sebulan</b>	125000000
<b>nilai_terkecil_penyewaan</b>	125000000
<b>nilai_terbesar_enyewaan</b>	125000000
<b>nilai_rata_rata_penyewaan</b>	125000000

## 4) Data Mining

Setelah melewati proses *selection*, *preprocessing*, *transformation*, kemudian data diolah untuk dicari informasinya dalam data menggunakan

algoritma K-Means *Clustering*. Data yang sudah diolah disimpan pada *database* dan diproses dengan sistem yang dibangun menggunakan K-Means *Clustering*

### 3.3 Pengujian Kualitas Kluster

Pengujian kualitas kluster diterapkan menggunakan data penyewaan pada PT. Mardika Daya Tribuana dengan total data sebanyak 280 baris data, penulis menguji kualitas kluster dengan *range* 2 kluster hingga 6 kluster menggunakan metode *Elbow*, dengan melihat jumlah nilai SSE yang tertinggi, dimana kluster yang mempunyai nilai SSE tertinggi merupakan kluster yang optimal untuk dipakai pada penelitian ini.

### 3.4 Pengujian Jumlah Kluster Terbaik

Pengujian untuk menentukan jumlah atau nilai kluster terbaik menggunakan metode *Elbow* dengan *range* jumlah kluster 2 hingga 6 kluster. Dimana hasil dengan nilai SSE tertinggi akan menjadi acuan pemakaian kluster untuk penelitian ini. Berikut Tabel 2 disajikan untuk hasil perhitungan menggunakan metode *Elbow*.

Tabel 5. Hasil SSE

Jumlah kluster	Nilai SSE	Nilai Selisih
2	23.700015185454	23.700015185454
3	13.255548218698	10.444466966756
4	17.874803680496	4.6192554617973
5	11.055794912194	6.819008768302
6	7.3935736791951	3.6622212329984

Berdasarkan Tabel 1 nilai selisih tertinggi terdapat pada  $k=2$  yakni 23.700015185454. oleh karena itu nilai ini menunjukkan bahwa *Cluster* optimal terdapat pada  $k=2$ . Namun demikian, pembagian *Cluster* tersebut tentunya harus melalui pertimbangan lainnya. jika pembagian *Cluster* terlalu sedikit maka tidak maksimal. Oleh karena itu penulis merekomendasikan menggunakan  $k=3$  dimana pada pengujian nilai selisih  $k=3$  yakni 10.444466966756.  $k=3$  pada pengujian memiliki nilai terbesar kedua setelah  $k=2$ .

### 3.5 Hasil Klasterisasi

Berdasarkan Tabel 1 pengelompokan data menjadi 3 (tiga) kluster juga dapat dipilih karena memiliki nilai SSE terbesar kedua setelah 2 (dua) kluster. Pemakaian 3 (tiga) kluster, penyajian data menjadi lebih efektif dan informatif dibandingkan dengan 2 (dua) kluster karena terlalu sedikit pembagiannya.

Hasil penelitian dengan nilai  $k=3$  atau dengan 3 kluster menggunakan data penyewaan pada PT. Mardika Daya Tribuana sebanyak 280 baris data ditampilkan mulai dari kluster 1 hingga kluster 3, pengujian pada klasterisasi dengan 3 kluster, memiliki dengan jumlah anggota masing-masing kluster adalah 113 data pada kluster 1, 156 data pada

kluster 2, dan 11 data pada kluster 3. Tabel 4.12 merupakan penyajian hasil klasterisasi data penyewaan jika dilakukan dengan nilai kluster 3.

Tabel 6. Klasterisasi dan informasi data penyewaan dengan nilai  $k=3$

Kluster	Informasi
1(11)	Kelompok ini merupakan kendaraan dengan level rendah dari segi peminatnya
2(113)	Kelompok ini merupakan kendaraan dengan level sedang dari segi peminatnya
3(156)	Kelompok ini merupakan kendaraan dengan level tinggi dari segi peminatnya

## 4. KESIMPULAN

Pada penelitian ini metode K-Means *Clustering* dapat diterapkan untuk mengelompokkan kendaraan pada PT. Mardika Daya Tribuana, pada proses pengujian terhadap 280 data penyewaan, didapatkan hasil 11 data pada kluster 1, 113 data pada kluster 2, dan 156 data pada kluster 3. Adapun jumlah *Cluster* yang direkomendasikan berdasarkan perhitungan nilai SSE adalah 2 *Cluster*. Tetapi pada prakteknya, jumlah *Cluster* sendiri dapat ditentukan sendiri oleh pengguna sesuai kebutuhan dan berdasarkan pertimbangan-pertimbangan tertentu, oleh karena itu pada penelitian ini penulis menggunakan 3 (tiga) *Cluster* yang dimana nilai selisih SSE pada 3 *Cluster* adalah nilai tertinggi kedua setelah nilai selisih *Cluster* 2, maka dapat dikatakan 3 *Cluster* juga optimal untuk dipakai pada penelitian ini. Diharapkan pada penelitian selanjutnya, klasterisasi dapat dikembangkan dengan menerapkan berbagai metode klasterisasi lainnya seperti *Fuzzy K-Means*, *Fuzzy C-Means*, *K-Medoids*, dan sebagainya. Dalam penentuan titik *centroid* awal atau nilai *centroid* awal, diharapkan menggunakan metode yang tepat agar hasil pengelompokan jauh lebih maksimal.

## DAFTAR PUSTAKA

- [1] T. Humaira Mufidah, M. Risky, Z. Anshori, and D. Jenderal Penyelenggaraan Haji dan Umrah Kementerian Agama, "Ciledug Raya No.99, RT.10/RW.2, Petukangan Utara," *Jurnal Media Infotama*, vol. 19, no. 1, pp. 3–4, 2023.
- [2] N. I. Febianto and N. Palasara, "Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 8, no. 2, pp. 130–140, Aug. 2019, doi: 10.32736/sisfokom.v8i2.653.
- [3] A. Solichin and K. Khairunnisa, "Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means," *Fountain of Informatics Journal*, vol. 5, no. 2, p. 52, Oct. 2020, doi: 10.21111/fij.v5i2.4905.
- [4] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan

- Algoritma K-Means,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [5] F. Amelia, I. Iskandar, S. Kurnia Gusti, and E. Haerani, “KREA-TIF: JURNAL TEKNIK INFORMATIKA Clustering Keluarga Miskin Desa Bina Baru dengan Metode K-Medoids,” vol. 11, no. 1, pp. 1–13, 2023, doi: 10.32832/kreatif.v11i1.14104.
- [6] C. Adi Rahmat, H. Permatasari, E. Rasywir, and Y. Pratama, “Penerapan K-Means Untuk Clustering Kondisi Gizi Balita Pada Posyandu,” vol. 7, no. 1, pp. 207–213, 2023, doi: 10.30865/mib.v7i1.5142.
- [7] S. Butsianto and N. T. Mayangwulan, “Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 3, no. 3, 2020.
- [8] A. Aditya, I. Jovian, and B. N. Sari, “Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 1, p. 51, Jan. 2020, doi: 10.30865/mib.v4i1.1784.
- [9] C. Selvi, D. Sembiring, L. Hanum, and S. Parsaoran Tamba, “Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus FTIK UNPRI),” *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, 2022.
- [10] S. Dwididanti and D. A. Anggoro, “Analisis Perbandingan Algoritma Bisecting K-Means dan Fuzzy C-Means pada Data Pengguna Kartu Kredit,” *Emitor: Jurnal Teknik Elektro*, vol. 22, no. 2, pp. 110–117, Aug. 2022, doi: 10.23917/emitor.v22i2.15677.
- [11] I. Wahyudi, M. Burhanis Sulthan, and L. Suhartini, “Analisa Penentuan Cluster Terbaik Pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi Di Pamekasan,” 2021.
- [12] A. Winarta and W. J. Kurniawan, “Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python,” *Jurnal Teknik Informatika Kaputama (JTik)*, vol. 5, no. 1, 2021.