

ANALISA KLASTERISASI PELANGGARAN LALU LINTAS MENGUNAKAN ALGORITMA K-PROTOTYPE (STUDI KASUS: PENGADILAN NEGERI KOTABUMI)

Nadia Apriliza¹⁾, Ryan Puji Cahyono²⁾
¹Sistem Informasi, ²Teknik Komputer
ryanpujic@gmail.com

Abstrak

Terjadinya pelanggaran lalu lintas merupakan salah satu bentuk problematika yang sering menimbulkan permasalahan di jalan raya, seperti kecelakaan dan kemacetan. Salah satu penyebab tingginya kasus pelanggaran lalu lintas adalah kurangnya pengetahuan dan kesadaran pengguna kendaraan dalam mematuhi peraturan lalu lintas. Penelitian ini dilakukan bertujuan untuk mengelompokkan dan menganalisis data pelanggaran lalu lintas di Pengadilan Negeri Kotabumi pada tahun 2020 dengan menggunakan algoritma K-Prototype guna mempermudah dalam mengetahui jenis pelanggaran yang kerap kali dilakukan oleh pengemudi. Atribut yang digunakan sebanyak 4 atribut yang terdiri dari jenis pelanggaran, jenis kendaraan, denda, dan pasal. Terdapat sebanyak 3 cluster yang terdiri dari C1 berjumlah 2202 dengan jumlah pelanggaran paling tinggi yaitu pasal 281, C2 berjumlah 861 dengan jumlah pelanggaran paling tinggi yaitu pasal 291(2), dan C3 berjumlah 586 dengan pelanggaran paling tinggi pasal 307. Dari hasil pengujian k optimal menggunakan elbow analysis didapatkan bahwa jumlah cluster yang optimal yaitu sebanyak 3 cluster.

Kata Kunci: *Clustering, Pelanggaran Lalu Lintas, K-Prototype, Elbow*

PENDAHULUAN

Transportasi berperan dalam mendorong, menunjang dan mendukung semua aspek kehidupan baik dalam pembangunan ekonomi, sosial budaya, politik dan pertahanan keamanan (Rusliyawati et al., 2021b). Peningkatan jumlah penduduk disertai dengan peningkatan aktivitas penduduk menyebabkan meningkatnya kebutuhan akan kendaraan. Seiring dengan pertumbuhan penduduk yang terus meningkat (Lalenoh et al., 2015) - (Ruslaini et al., 2021), berdampak dengan meningkatnya jumlah kendaraan di Indonesia. Jumlah penduduk yang terus meningkat berakibat pada meningkatnya jumlah pergerakan atau mobilitas masyarakat dalam rangka pemenuhan kebutuhan hidupnya (Kartawijaya, 2018). Peningkatan jumlah kendaraan terjadi seiring bertambahnya jumlah penduduk karena kendaraan dibutuhkan untuk memfasilitasi pergerakan masyarakat. Menurut (Sutanto et al.,

2014), seiring dengan pertumbuhan penduduk dan perkembangan ekonomi (Susanto & Puspaningrum, 2019), banyak negara-negara berkembang yang mengalami peningkatan signifikan dalam jumlah motor dan mobil (Pintoko & L., 2018). Banyaknya data pelanggaran lalu lintas pada Pengadilan Negeri Kotabumi maka data tersebut akan diolah dengan menggunakan data mining. Penerapan data mining pada data pelanggaran lalu lintas yaitu untuk mempermudah pengelompokan data-data pelanggaran lalu lintas dan menentukan pola cluster dengan menggunakan algoritma K-Prototype yang nantinya dapat dikelompokkan dan dianalisis (P. R. P. Adi & Indrawati, 2019). Data mining juga merupakan metode yang digunakan dalam pengolahan data berskala besar oleh karena itu data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi (Bakri, 2017). Berdasarkan hasil simulasi yang telah dilakukan maka dapat diambil kesimpulan bahwa secara umum algoritma K-Prototype dapat mempertahankan efisiensi algoritma K-means dalam menangani data berukuran besar tetapi menghilangkan keterbatasan penerapan hanya pada data numerik namun dapat diterapkan pada data kategorikal (Dewi et al., 2021) - (Nabila, Rahman Isnain, et al., 2021) d. Sehingga K-Prototype memberikan hasil clustering yang lebih baik karena dapat memberikan ciri atau karakteristik yang lebih mirip dalam cluster yang terbentuk (Herlinda et al., 2021b).

KAJIAN PUSTAKA

Data Mining

Adapun pengertian Data Mining menurut beberapa ahli, Data mining adalah suatu proses untuk mendapatkan informasi dengan melakukan pencarian pola dan relasi-relasi yang tersembunyi didalam suatu tumpukan data yang banyak (Anestiviya et al., 2021) - (Aldino et al., 2021). Data mining bisa digunakan untuk membantu pengambilan keputusan dimasa mendatang (Rani, 2016). Data mining adalah bidang ilmu yang digunakan untuk menangani masalah dalam pengambilan informasi dari database yang besar dengan menggabungkan teknik dari statistik (Wiguna et al., 2019), pembelajaran mesin, visualisasi data, pengenalan pola, dan database. Data mining juga merupakan metode yang digunakan dalam pengolahan data berskala besar oleh karena itu data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri (Aldino et al., 2021), bidang keuangan (Rusliyawati et al., 2021a), cuaca (Aji & Dewi, 2017), ilmu dan teknologi (Ahdan et al., 2021). Data Mining adalah proses dari menganalisa data dari prespektif yang berbeda dan menyimpulkannya ke dalam informasi yang berguna. Data mining membahas

penggalian atau pengumpulan informasi yang berguna dari kumpulan data (Herlinda et al., 2021a). Informasi yang biasanya dikumpulkan adalah pola-pola tersembunyi pada data, berhubungan antar elemen-element data ataupun pembuatan model untuk keperluan peramalan. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan (Rahmanto et al., 2021), menggali (Vitianingsih, 2016), atau menambang pengetahuan dari data atau informasi yang kita miliki (Alita, 2021).

Clustering

Metode clustering secara umum dapat dibagi menjadi dua yaitu hierarchical clustering dan partitional clustering (Sulistiyawati & Supriyanto, 2021)–(Nabila, Isnain, et al., 2021), berikut ini merupakan metode clustering:

a. Hierarchical Clustering

Hierarchical clustering, data dikelompokkan melalui suatu bagan berupa hirarki (Wantoro & Priandika, n.d.), dimana terdapat penggabungan dua grup yang terdekat disetiap iterasinya ataupun pembagian dari seluruh set data kedalam cluster (Lestari & Savitri Puspaningrum, 2021).

b. Partitional Clustering.

Partitional clustering merupakan pengelompokan data ke dalam beberapa cluster tanpa adanya struktur hirarki antara satu dengan yang lainnya (Ali et al., 2020). Pada metode partitional clustering setiap cluster memiliki titik pusat cluster (centroid) dan secara umum metode ini memiliki fungsi tujuan yaitu meminimumkan jarak (dissimilarity) dari seluruh data ke pusat cluster masing-masing (Olsson et al., 2022).

Algoritma K-Prototype

Algoritma K-Prototype merupakan algoritma dengan pendekatan gabungan dari algoritma k-means dan algoritma k-modes (Fikri et al., 2017), yaitu data campuran penggabungan antara data numerik dan data kategorikal. Menurut (Ahluwalia, 2020), menganalisis algoritma clustering K-Prototype untuk data bertipe data campuran seperti data numerikal dan kategorikal. Metode berbasis K-Means memiliki efisiensi set data yang besar dan memiliki nilai numerik yang terbatas untuk dievaluasi (Setiawan, 2019). K-Prototype mempunyai keunggulan karena algoritma yang tidak terlalu kompleks dan mampu menangani data yang besar serta lebih baik dibandingkan dengan algoritma yang berbasis hirarki (Sari, 2021). Algoritma K-Prototype merupakan salah satu metode clustering

berbasis partitioning (R. P. Adi et al., 2020). Menurut (Kurniawan et al., 2019), telah mengusulkan algoritma pengelompokan K-Prototype untuk dataset yang tidak lengkap dengan atribut bertipe data campuran numerikal dan kategorikal. Algoritma K Prototype tradisional sangat ahli dalam mengelompokkan data dengan atribut numerikal dan kategorikal campuran, sedangkan data yang lengkap terbatas (Nabila, Rahman Isnain, et al., 2021).

METODE

Metode pengumpulan data adalah tahapan yang digunakan peneliti untuk mendapatkan data sebagai bahan dalam penyusunan proposal dengan menggunakan beberapa metode yaitu sebagai berikut:

Studi Pustaka dan Literatur

Studi Pustaka pada penelitian ini bersumber dari buku, jurnal, dan segala kepustakaan lain yang dianggap penting untuk mendukung dari penelitian ini (Suweda, 2009). Identifikasi masalah merupakan proses untuk mendefinisikan masalah dan membuat definisi tersebut dapat diuji dan diukur (Rudi Cahyono & Nurmahaludin, 2017). Pada proses ini dilakukan analisis masalah yang ada yaitu menemukan dan mengidentifikasi masalah. Setelah ditemukan, permasalahan dapat diangkat menjadi sebuah penelitian. Permasalahan pada penelitian ini adalah mengelompokkan kasus pelanggaran lalu lintas dengan metode clustering dengan menerapkan algoritma K-Prototype berdasarkan dari beberapa atribut (Rahmawati & Nani, 2021).

Pengumpulan data

Sumber data pada penelitian ini adalah data pelanggaran lalu lintas yang didapat dari website resmi pada Pengadilan Negeri Kotabumi tahun 2020 pada kasus perkara lalu lintas dengan jumlah data 3662 dan atribut yang digunakan sebanyak 4 atribut.

Studi Pustaka (*Library Research*)

Metode pengumpulan data dilakukan peneliti dengan cara mempelajari laporan-laporan ilmiah atau sumber bacaan dan buku-buku yang berkaitan atau berhubungan dengan judul penelitian (Damayanti, 2021).

Dokumentasi (*Documentation*)

Metode pengumpulan data yang dilakukan peneliti dengan cara mengumpulkan beberapa dokumen dan data yang diperlukan dalam penelitian. Hasil dokumentasi yang dilakukan disekolah sudah terlampir(Fakhrurozi et al., 2021).

HASIL DAN PEMBAHASAN

Seleksi Data

Merupakan tahapan awal, yaitu memilih dan menyeleksi data yang akan digunakan. Dimana hanya memilih atribut-atribut yang memenuhi tujuan penelitian. Atribut yang akan digunakan adalah jenis pelanggaran, jenis kendaraan, denda, dan bukti, data yang akan di clustering sebanyak 3662 data.

Tabel 1. Seleksi Data

NAMA	PLAT	JENIS PELANGGARAN (PASAL)	JENIS KENDARAAN	DENDA (Rp)	BUKTI
AM [REDACTED]	BE 86 [REDACTED]	289 UULLAJ Jo 106(6) ULLAJ	TRUK	124000	STNK
FE [REDACTED] WA [REDACTED]	BE [REDACTED]	289 UULLAJ Jo 106(6) ULLAJ	MOBIL PENUMPANG PRIBADI	74000	STNK
SALW [REDACTED] O [REDACTED]	-	293(2) ULLAJ JO 106 ULLAJ	SEPEDA MOTOR	39000	STNK
NA [REDACTED]	-	291 (2) UULLAJ JO 106 ULLAJ	SEPEDA MOTOR	39000	STNK
.....
ME [REDACTED] SAI [REDACTED]	BE 43 [REDACTED]	293(2)UULLAJ JO 106 ULLAJ	SEPEDA MOTOR	49000	SIM C

Tabel 2. Deskripsi Data

Nama Atribut	Tipe Data
Jenis Pelanggaran (Pasal)	Kategorikal
Jenis Kendaraan	Kategorikal
Denda (Rp)	Numerikal
Bukti	Kategorikal

Preprocessing Data

Preprocessing dilakukan untuk membuang data yang tidak layak untuk memasuki proses data mining, seperti data noise dan missing value. Cara untuk memperbaiki noise dapat dilakukan dengan beberapa cara yaitu dapat dihapus, diisi secara manual, diabaikan, diisi dengan rata-rata parameter, atau dengan rata - rata kelas. Pada variabel non-numerik atau kategorikal, dapat mengganti missing values dengan nilai modus atau nilai yang paling banyak muncul dalam kolom tersebut. Gambar dibawah ini merupakan contoh dari data processing.

Tabel 3 Pembersihan Data

NAMA	PLAT	JENIS PELANGGARAN (PASAL)	JENIS KENDARAAN	DENDA (Rp)	BUKTI
AN [REDACTED]	BE 86 [REDACTED]	289 UULLAJ Jo 106(6) ULLAJ	TRUK	124000	STNK
FE [REDACTED] W [REDACTED]	BE [REDACTED]	289 UULLAJ Jo 106(6) ULLAJ	MOBIL PENUMPANG PRIBADI	74000	STNK
SA [REDACTED] NT [REDACTED]	-	293(2) ULLAJ JO 106 ULLAJ	SEPEDA MOTOR	39000	STNK
NAN [REDACTED]	-	291 (2) UULLAJ JO 106 UULLAJ	SEPEDA MOTOR	39000	STNK
.....
MEIRI [REDACTED] SA [REDACTED] A [REDACTED]	BE 432 [REDACTED]	293(2)UULLAJ JO 106 UULLAJ	SEPEDA MOTOR	49000	SIM C

Berdasarkan Tabel diatas, atribut yang akan digunakan yaitu Jenis Pelanggaran (Pasal), Jenis Kendaraan, Denda (Rp), dan Bukti. Sedangkan atribut Nama dan Plat tidak digunakan.

JENIS PELANGGARAN (PASAL)	JENIS PELANGGARAN (PASAL)
289 UULLAJ Jo 106(6) ULLAJ	Setiap pengemudi atau penumpang yang duduk disamping pengemudi mobil tidak menggunakan sabuk keselamatan (Pasal 289)
289 UULLAJ Jo 106(6) ULLAJ	Setiap pengemudi atau penumpang yang duduk disamping pengemudi mobil tidak menggunakan sabuk keselamatan (Pasal 289)
293(2) ULLAJ JO 106 ULLAJ	Pengemudi Sepeda Motor di Jalan tanpa menyalakan lampu utama pada siang hari (Pasal 293(2))
291 (2) UULLAJ JO 106 UULLAJ	Penumpang tidak menggunakan helm standar nasional (Pasal 291(2))
.....
293(2)UULLAJ JO 106 UULLAJ	Pengemudi Sepeda Motor di Jalan tanpa menyalakan lampu utama pada siang hari (Pasal 293(2))

Pada Tabel Jenis Pelanggaran (Pasal) ditambahkan keterangan atau penjelasan terkait pasal yang dilanggar. Pada tahun 2020, pasal yang dilanggar sebanyak 24 pasal mengenai pelanggaran lalu lintas.

Transformasi Data

Pada tahap ini dilakukan perubahan data yang memiliki tipe data yang awalnya tidak bisa diolah secara matematis menjadi data yang dapat diolah. Tujuan dari transformasi data untuk menghindari data yang rusak dan tidak valid, bentuk data dibagi menjadi beberapa golongan dengan skala tertentu, dengan tujuan agar variasi data pada atribut-atribut tertentu menjadi lebih sedikit. Berikut merupakan Tabel inialisasi data.

Tabel 5. Inisialisasi Atribut Jenis Pelanggaran (pasal)

JENIS PELANGGARAN (PASAL)	Inisialisasi
Tidak dipasang Tanda Nomor Kendaraan (Pasal 280)	A
Tidak memiliki SIM (Pasal 281)	B
Setiap Pengguna Jalan Tidak mematu perintah yang diberikan petugas Polri s ebagaimana dimaksud dalam pasal 104 ayat (3), yaitu dalam keadaan tertentu untuk ketertiban dan kelancaran lalu lintas wajib untuk : Berhenti, jalan terus, mempercepat, memperlambat, dan / atau mengalihkan arus kendaraan (Pasal 282)	C
Pengemudi Sepeda Motor tidak memenuhi persyaratan teknis dan laik jalan (Pasal 285(1))	D
Pengemudi Mobil tidak memenuhi persyaratan teknis dan laik jalan meliputi : kaca spion, klakson, ampu utama, lampu rem, lampu penunjuk arah, atau alat pemantul cahaya, alat pengukur kecepatan, knalpot dan kedalaman alur ban(Pasal 285(2))	E
Kesesuaian kinerja roda dan kondisi Ban; Kesesuaian daya mesin penggerak terhadap berat kendaraan(Pasal 286)	F
Melanggar rambu lalu lintas Pasal (Pasal 287(1))	G
Melanggar aturan perintah atau melanggar yang dinyatakan dengan Alat Pemberi Isyarat Lalu Lintas (Pasal 287(2))	H
Tidak memberi prioritas jalan bagi Kendaraan Bermotor memiliki hak utama yang menggunakan alat peringatan dengan bunyi dan sinar dan/atau yang di kawal oleh Petugas Polri (Pasal 287(4))	I
Tidak Memiliki STNK (288(1))	J
Pengendara Sepeda Motor yang memiliki SIM namun tidak dapat menunjukkannya saat razia (Pasal 288(2))	K
Kendaraan tidak dilengkapi dengan surat keterangan uji berkala dan tanda lulus uji berkala (Pasal 288(3))	L
Setiap pengemudi atau penumpang yang duduk disamping pengemudi mobil tidak menggunakan sabuk keselamatan (Pasal 289)	M
Pengemudi dan Penumpang tidak mengenakan sabuk keselamatan dan helm (Pasal 290)	N
Tidak mengenakan helm standar nasional (Pasal 201)	O
Setiap pengendara atau penumpang sepeda motor tidak menggunakan helm standar nasional (Pasal 291(1))	P
Penumpang tidak menggunakan helm standar nasional (Pasal 291(2))	Q
Muatan Tanpa Kereta samping mengangkut penumpang lebih dari 1 orang (Pasal 292)	R

JENIS PELANGGARAN (PASAL)	Inisialisasi
Pengemudi Sepeda Motor di Jalan tanpa menyalakan lampu utama pada siang hari (Pasal 293(2))	S
Berpindah lajur atau bergerak ke samping Tidak memberikan isyarat saat akan berpindah lajur atau bergerak kesamping (Pasal 295(2))	T
Pengemudi mobil barang untuk mengangkut orang (Pasal 303)	U
Kendaraan angkutan orang dengan tujuan tertentu, tapi menaikkan atau menurunkan Penumpang lain di sepanjang perjalanan atau menggunakan Kendaraan angkutan tidak sesuai dengan angkutan untuk keperluan lain (Pasal 304)	V
Tata Cara Pemuatan Barang Tidak mematu ketentuan mengenai tata cara pemuatan, daya angkut (Pasal 307)	W
(a) tidak memiliki izin menyelenggarakan angkutan orang dalam trayek; (b) tidak memiliki izin menyelenggarakan angkutan orang tidak dalam trayek; (c) tidak memiliki izin menyelenggarakan angkutan barang khusus dan alat berat (Pasal 308)	X

Gambar 1 Activity Diagram Registrasi

Normalisasi Data

Normalisasi yaitu proses penskalaan nilai pada atribut data sehingga data tersebut dapat pada range tertentu, untuk menghasilkan keseimbangan nilai atribut antar data saat sebelum dan sesudah proses data mining. Pada tahap ini menggunakan normalisasi min-max.

Tabel 4.

Denda (Rp)	Normalisasi
39000	0
49000	0.06250
54000	0.09375
59000	0.12500
69000	0.18750
74000	0.21875
99000	0.37500
100000	0.38125
124000	0.53125
149000	0.68750
174000	0.84375
199000	1

Berikut ini merupakan perhitungan dari proses normalisasi pada dataa denda :

$$Normalisasi = \frac{(Nilai Atribut - nilai min atribut)}{(Nilai max atribut - nilai min atribut)} \quad (1)$$

Diketahui :

Nilai Atribut Denda : 120000

Nilai Max Atribut : 199000

Nilai Min Atribut : 39000

Maka:

$$Normalisasi = \frac{(124000 - 39000)}{(199000 - 39000)} = 0,53125$$

(12)

Data Mining (K-Prototype)

Beberapa tahapan yang dilakukan yaitu inialisasi awal. Pada proses inialisasi ini dilakukan pemilihan k prototype secara acak melalui dataset. Setelah inialisasi awal tahap selanjutnya yaitu menghitung jarak seluruh data pada dataset terhadap inialisasi awal, lalu alokasikan data tersebut kedalam x cluster terdekat. Tahapan selanjutnya adalah mengitung titik pusat cluster yang baru setelah semua objek pada cluster telah dialokasikan lalu lakukan realokasikan semua objek, jika titik pusat cluster tidak berubah maka proses akan berhenti, proses ini akan terus dilakukan sampai tidak adalagi perubahan prototype atau sampai kriteria stopping terpenuhi.

Implentasi Exploratory Data Analysis

Exploratory Data Analysis adalah untuk dapat memahami komponen dan isi dalam penyusunan data. Berikut merupakan Exploratory Data Analysis yang digunakan dalam penelitian ini.

- 1) import data didukung dengan menggunakan library pandas, numpy, matplotlib. Pandas dan numpy digunakan untuk mengolah data numeric, pandas juga dapat digunakan untuk membaca file, membuat tabel, dan lain sebagainya, sedangkan numpy digunakan untuk operasi matrik dan vektor. Sedangkan matplotlib digunakan untuk untuk visualisasi data.

```
!pip install kmodes
import pandas as pd
import numpy as np
from kmodes.kprototypes import KPrototypes
from tqdm.auto import tqdm
import matplotlib.pyplot as plt
```

Gambar 1 *Exploratory data analysis*

Pada Gambar diatas dapat dilihat bahwa pada proses ini perlu menginstal K- Modes dikarenakan algoritma K-Prototype belum dapat menjalankan perintah sendiri. Selanjutnya yang digunakan yaitu read data dan head, read data digunakan untuk memasukan data dengan format CSV atau dengan format lainnya seperti XLSX. Sedangkan library head digunakan untuk menampilkan data, dan untuk dapat melihat 5 baris pertama dari dataframe

```
df=pd.read_csv('drive/MyDrive/DataOlahh.csv')
df.head()
```

Gambar 2. *Library Read Data dan Head*

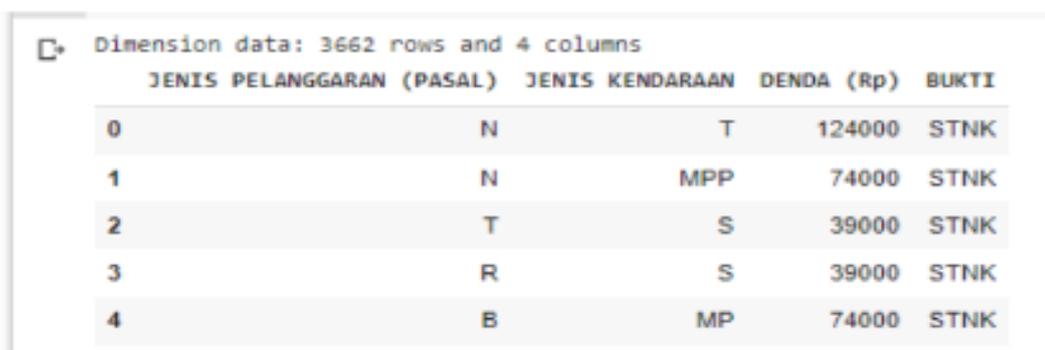
Implementasi Preprocessing Data

Tahapan ini merupakan proses untuk menyeleksi data yang akan digunakan pada proses pengolahan untuk memastikan data yang digunakan tidak terdapat kesalahan data. Proses preprocessing pertama menghapus atribut yang tidak digunakan yaitu nama dan plat dengan perintah `df.drop()`.

```
# Data pre-processing
df.drop(['NAMA', 'PLAT'], 1, inplace=True)
```

Gambar 3. Drop Data

Setelah data dihapus, data ditampilkan kembali untuk melihat apakah atribut telah terhapus

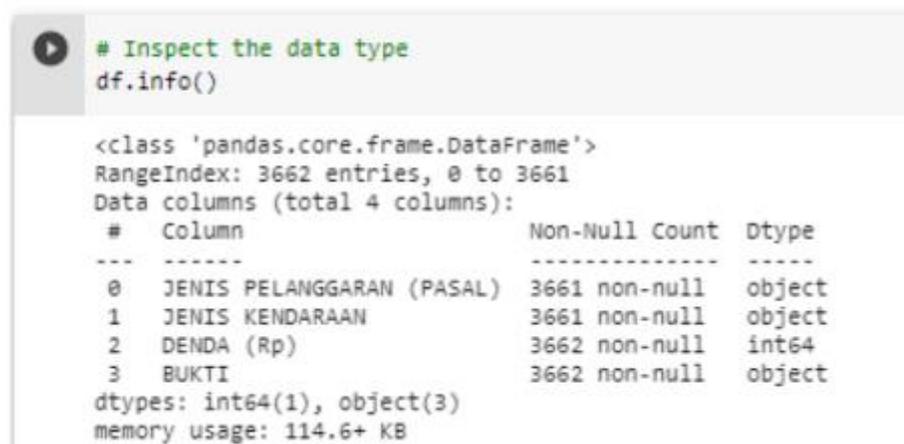


```
Dimension data: 3662 rows and 4 columns
```

	JENIS PELANGGARAN (PASAL)	JENIS KENDARAAN	DENDA (Rp)	BUKTI
0	N	T	124000	STNK
1	N	MPP	74000	STNK
2	T	S	39000	STNK
3	R	S	39000	STNK
4	B	MP	74000	STNK

Gambar 4. Data Setelah Dihapus

`print()` digunakan untuk mencetak output program di layar. Selanjutnya, yaitu menggunakan perintah `df.info()` untuk mengetahui informasi mengenai detail dataframe, seperti data tipe, nama kolom, dan untuk memastikan tipe data setiap kolom dipetakan dengan benar. Pada penelitian ini terdapat 4 kolom atribut dan data tipenya yaitu object untuk data kategorikal dan `int64` untuk data bertipe numerikal.



```
# Inspect the data type
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3662 entries, 0 to 3661
Data columns (total 4 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   JENIS PELANGGARAN (PASAL)            3661 non-null   object
1   JENIS KENDARAAN                      3661 non-null   object
2   DENDA (Rp)                          3662 non-null   int64
3   BUKTI                                3662 non-null   object
dtypes: int64(1), object(3)
memory usage: 114.6+ KB
```

Gambar 5. Informasi Dataframe

Untuk mengetahui jumlah variabel data kategorikal menggunakan perintah `df.select_dtypes('object').nunique()` dengan menyeleksi tipe data yang berbentuk objek. Pada jenis pelanggaran terdapat sebanyak 24 variabel, jenis kendaraan terdapat 4 variabel, dan bukti terdapat 6 variabel.

```
# Inspect the categorical variables
df.select_dtypes('object').nunique()

JENIS PELANGGARAN (PASAL)    24
JENIS KENDARAAN              4
BUKTI                        6
dtype: int64
```

Gambar 6. Informasi Jumlah Variable Data Kategorikal

```
# Summary statistics of numerical variable
for i in df.select_dtypes('object').columns:
    print(df[i].value_counts(), '\n')
```

Gambar 7. Informasi Data Atribut

Hasil Clustering

Setelah proses preprocessing dilakukan selanjutnya proses pengolahan data atau training data. Pada Gambar 4.16 memperlihatkan perintah yang digunakan pada proses pengolahan data, seperti yang telah ditentukan pada penelitian ini menggunakan 3 cluster.

```
#training
model = KPrototypes(n_clusters=3, verbose=3)
member = model.fit_predict(df, categorical=[0, 1, 3])
```

Gambar 8. Pengolahan data

Setelah training data, proses selanjutnya yaitu array. Array merupakan tipe data terstruktur dalam pemrograman, array memungkinkan untuk menyimpan data maupun referensi objek dalam jumlah terindeks dan banyak.

```
df['member']=cluster_dict
df
```

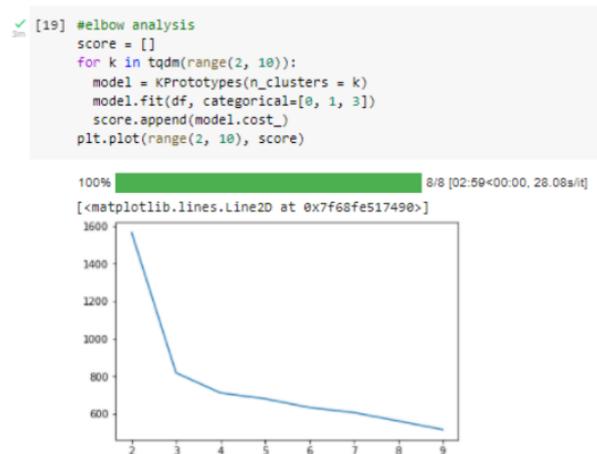
	JENIS PELANGGARAN (PASAL)	JENIS KENDARAAN	DENDA (Rp)	BUKTI	member
0	M	T	0.53125	STNK	2
1	M	MPP	0.21875	STNK	0
2	S	S	0.00000	STNK	0
3	Q	S	0.00000	STNK	0
4	B	MP	0.21875	STNK	0
...
3657	J	T	0.68750	SIM BI UMUM	2
3658	B	MPP	0.37500	STNK	0
3659	K	S	0.06250	STNK	0
3660	K	S	0.06250	STNK	0
3661	S	S	0.06250	SIM C	1

3662 rows x 5 columns

Gambar 9. Hasil Clustering

Evaluasi Clustering

Pada penelitian ini untuk melakukan evaluasi clustering menggunakan elbow analysis dalam menentukan jumlah cluster optimal yang digunakan. Metode Elbow adalah metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik.



Gambar 10. Evaluasi Clustering

Pemilihan jumlah kluster dilakukan secara bertahap dimulai dari jumlah kluster sebanyak 2 hingga 10 cluster. Dari hasil pengolahan data diatas didapatkan hasil cluster optimal menggunakan elbow analysis yaitu sebanyak 3 cluster yang terdiri dari C1 berjumlah 2202 dengan jumlah pelanggaran paling tinggi yaitu pasal 281, C2 berjumlah 861 dengan jumlah pelanggaran paling tinggi yaitu pasal 291(2), dan C3 berjumlah 586 dengan pelanggaran paling tinggi pasal 307.

SIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan penulis, maka dapat disimpulkan bahwa penelitian yang berjudul “Arsitektur Enterprise Sebagai Pendukung Keputusan Penerimaan Siswa Baru Menggunakan Metode SAW” sebagai berikut:

- Menghasilkan website sistem informasi penerimaan siswa baru dengan menggunakan metode SAW dalam perhitungan yang dilakukan, website ini dapat membantu mempermudah pekerjaan panitia penerimaan siswa baru melakukan seleksi nilai dan dalam pembuatan laporan. Tidak hanya itu, bahkan calon siswa yang ingin mendaftar pada SMAN 1 Bumi Agung tidak perlu lagi untuk datang ke sekolah dan membawa

- persyaratan saat melakukan pendaftaran, karena pendaftaran dapat dilakukan dimana saja asalkan terhubung dengan internet.
- Website penerimaan siswa baru dengan menggunakan metode SAW dibangun dengan menggunakan metode pengembang sistem Extreme Programming yang memiliki 4 tahapan yaitu perencanaan, desain, pengkodean dan pengujian, kemudian sistem digambarkan menggunakan UML seperti usecase, activity dan class diagram. Berdasarkan hasil pengujian yang dilakukan dengan menggunakan pengujian black box diperoleh hasil 100% yang artinya dapat disimpulkan sistem tersebut sudah efektif, dan pada pengujian usability diperoleh hasil 93% yang dapat disimpulkan responden sangat setuju.

REFERENSI

- Adi, P. R. P., & Indrawati, K. R. (2019). Perbedaan keterikatan kerja berdasarkan generasi kerja karyawan pada perusahaan berkonsep thk ditinjau dari etos kerja. *Jurnal Psikologi Udayana*, 6(1), 46–57.
<https://ojs.unud.ac.id/index.php/psikologi/article/view/47148/28328>
- Adi, R. P., Koswara, Y., Tashika, J., Devi, Y., & Saifudin, A. (2020). Pengujian Black Box pada Aplikasi Pertokoan Minimarket Menggunakan Metode Equivalence Partitioning. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 3(2), 100.
<https://doi.org/10.32493/jtsi.v3i2.4695>
- Ahdan, S., Sucipto, A., Priandika, A. T., & ... (2021). Peningkatan Kemampuan Guru SMK Kridawisata Di Masa Pandemi Covid-19 Melalui Pengelolaan Sistem Pembelajaran Daring. *Jurnal ABDINUS ...*, 5(2), 390–401.
<http://ojs.unpkediri.ac.id/index.php/PPM/article/view/15591>
- Ahluwalia, L. (2020). EMPOWERMENT LEADERSHIP AND PERFORMANCE: ANTECEDENTS. *Angewandte Chemie International Edition*, 6(11), 951–952., 7(1), 283.
http://www.nostarch.com/javascriptforkids%0Ahttp://www.investopedia.com/terms/i/in_speceie.asp%0Ahttp://dspace.ucuenca.edu.ec/bitstream/123456789/35612/1/Trabajo de Titulacion.pdf%0Ahttps://educacion.gob.ec/wp-content/uploads/downloads/2019/01/GUIA-METODOL
- Aji, G. F. S., & Dewi, N. (2017). Prosiding Seminar Nasional: Membongkar Sastra, Menggugat Rezim Kepastian. In *Prosiding Seminar Nasional: Membongkar Sastra, Menggugat Rezim*

· *Kepastian.*

- Aldino, A. A., Darwis, D., Prastowo, A. T., & Sujana, C. (2021). Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency. *Journal of Physics: Conference Series*, 1751(1). <https://doi.org/10.1088/1742-6596/1751/1/012038>
- Ali, Z. M., Hassoon, N. H., Ahmed, W. S., & Abed, H. N. (2020). The Application of Data Mining for Predicting Academic Performance Using K-means Clustering and Naïve Bayes Classification. *International Journal of Psychosocial Rehabilitation*, 24(03), 2143–2151. <https://doi.org/10.37200/ijpr/v24i3/pr200962>
- Alita, D. (2021). Multiclass SVM Algorithm for Sarcasm Text in Twitter. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(1), 118–128. <https://doi.org/10.35957/jatisi.v8i1.646>
- Anestiviya, V., Ferico, A., & Pasaribu, O. (2021). Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : Sman 1 Natar). *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(1), 80–85. <http://jim.teknokrat.ac.id/index.php/JTISI>
- Bakri, M. (2017). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Vol, 11*, 1–4.
- Damayanti. (2021). Digitalisasi Sistem Peminjaman Buku Pada Smk Negeri 2 Kalianda Lampung Selatan. *Journal of Social ...*, 2(2), 128–138. <https://ejurnal.teknokrat.ac.id/index.php/JSSTCS/article/view/1368>
- Dewi, P. S., Sastradipraja, C. K., & Gustian, D. (2021). Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Algoritma Naïve Bayes Classifier. *Jurnal Teknologi Dan Informasi*, 11(1), 66–80. <https://doi.org/10.34010/jati.v11i1.3593>
- Fakhrurozi, J., Pasha, D., Jupriyadi, J., & Anggrenia, I. (2021). Pemertahanan Sastra Lisan Lampung Berbasis Digital Di Kabupaten Pesawaran. *Journal of Social Sciences and Technology for Community Service (JSSTCS)*, 2(1), 27. <https://doi.org/10.33365/jsstcs.v2i1.1068>
- Fikri, C. M., Agustin, F. E. M., & Mintarsih, F. (2017). Pengelompokan kualitas kerja pegawai menggunakan algoritma K-Means++ dan Cop-Kmeans untuk merencanakan program

- pemeliharaan kesehatan pegawai di pt. Pln p2b jb depok. *Pseudocode*, 4(1), 9–17.
- Herlinda, V., Darwis, D., & Dartono, D. (2021a). Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 94–99.
- Herlinda, V., Darwis, D., & Dartono, D. (2021b). ANALISIS CLUSTERING UNTUK RECREDESIALING FASILITAS KESEHATAN MENGGUNAKAN METODE FUZZY C-MEANS. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 94–99.
- Kartawijaya, K. (2018). Pengaruh Lingkungan Keluarga dan Motivasi terhadap Prestasi Belajar Matematika (Eksperiment pada SMP Negeri di Kabupaten Bekasi). *Alfarisi: Jurnal Pendidikan MIPA*, 1(2), 211–217.
<https://journal.lppmunindra.ac.id/index.php/alfarisi/article/view/5668>
<https://journal.lppmunindra.ac.id/index.php/alfarisi/article/download/5668/2939>
- Kurniawan, D. E., Ahmad, I., Ridho, M. R., Hidayat, F., & Anggra Js, A. (2019). Analysis of performance comparison between Software-Based iSCSI SAN and Hardware-Based iSCSI SAN. *Journal of Physics: Conference Series*, 1351(1).
<https://doi.org/10.1088/1742-6596/1351/1/012009>
- Lalenoh, R. H., Sendow, T. K., & Jansen, F. (2015). Analisa Kapasitas Ruas Jalan Sam Ratulangi Dengan Metode Mkji 1997 Dan Pkji 2014. *Jurnal Sipil Statik*, 3(11), 737–746.
- Lestari, G., & Savitri Puspaningrum, A. (2021). Sistem Pendukung Keputusan Pemberian Tunjangan Karyawan Menggunakan Metode Analytical Hierarchy Process (Ahp) Studi Kasus: Pt Mutiara Ferindo Internusa. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(3), 38–48. <http://jim.teknokrat.ac.id/index.php/JTISI>
- Nabila, Z., Isnain, A. R., & Permata, P. (2021). Mining Data Analysis for Clustering of Covid-19 Case in Lampung Province Using K-Means Algorithm. *The 1st International Conference on Advanced Information Technology and Communication (IC-AITC)*.
- Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTISI>
- Olsson, T., Ericsson, M., & Wingkvist, A. (2022). To automatically map source code entities to architectural modules with Naive Bayes. *Journal of Systems and Software*, 183, 111095.

- <https://doi.org/10.1016/j.jss.2021.111095>
- Pintoko, B. M., & L., K. M. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier. *E-Proceeding of Engineering*, 5(3), 8121–8130.
- Rahmanto, Y., Alfian, J., Damayanti, D., & Borman, R. I. (2021). Penerapan Algoritma Sequential Search pada Aplikasi Kamus Bahasa Ilmiah Tumbuhan. *Jurnal Buana Informatika*, 12(1), 21. <https://doi.org/10.24002/jbi.v12i1.4367>
- Rahmawati, D., & Nani, D. A. (2021). PENGARUH PROFITABILITAS, UKURAN PERUSAHAAN, DAN TINGKAT HUTANG TERHADAP TAX AVOIDANCE. *Jurnal Akuntansi Dan Keuangan*, 26(1), 1–11. <https://doi.org/10.23960/jak.v26i1.246>
- Rani, L. N. (2016). Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *INOVTEK Polbeng - Seri Informatika*, 1(2), 126. <https://doi.org/10.35314/isi.v1i2.131>
- Rudi Cahyono, G., & Nurmahaludin, N. (2017). Rancang Bangun Sistem Monitoring Populasi Hama Tanaman Padi Berbasis Web dan Gateway. *Poros Teknik*, 8(2), 55. <https://doi.org/10.31961/porosteknik.v8i2.388>
- Ruslaini, R., Abizar, A., Ramadhani, N., & Ahmad, I. (2021). PENINGKATAN MANAJEMEN DAN TEKNOLOGI PEMASARAN PADA UMKM OJESA (OJEK SAHABAT WANITA) DALAM MENGATASI LESS CONTACT EKONOMI MASA COVID-19. *Martabe: Jurnal Pengabdian Kepada Masyarakat*, 4(1), 139–144.
- Rusliyawati, Putri, T. M., & Darwis, D. (2021a). Penerapan Metode Garis Lurus dalam Sistem Informasi Akuntansi Perhitungan Penyusutan Aktiva Tetap pada PO Puspa Jaya. *Jurnal Ilmiah Sistem Informasi Akuntansi (JIMASIA)*, 1(2), 1–13.
- Rusliyawati, R., Putri, T. M., & Darwis, D. (2021b). Penerapan Metode Garis Lurus Dalam Sistem Informasi Akuntansi Perhitungan Penyusutan Aktiva Tetap Pada Po Puspa Jaya. *Jurnal Ilmiah Sistem Informasi Akuntansi*, 1(1), 1–13. <http://jim.teknokrat.ac.id/index.php/jimasia/article/view/864>
- Sari, N. (2021). Implementation of the AHP-SAW method in the decision support system for selecting the best tourism village. *Jurnal Teknik Informatika CIT Medicom*. <https://www.medikom.iocspublisher.org/index.php/JTI/article/view/51>
-

- Setiawan, S. (2019). Analisis Cluster Menggunakan Algoritma K-Means Untuk Mengetahui Kemampuan Pegawai Dibidang It Pada Cv. Roxed Ltd. *Pelita Informatika: Informasi Dan Informatika*, 7(3), 341–347.
- Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25.
<https://doi.org/10.33365/jtk.v15i2.1162>
- Susanto, E. R., & Puspaningrum, A. S. (2019). *Rancang Bangun Rekomendasi Penerima Bantuan Sosial Berdasarkan Data Kesejahteraan Rakyat*. 15(1), 1–12.
- Sutanto, F., Samsurizal, E., & Budi, G. S. (2014). Analisa Perhitungan Struktur Bangunan Gedung Head Office Dan Showroom Yamaha Pontianak. *Jurnal Mahasiswa Teknik Sipil Universitas Tanjungpura*, 3(2), 1–9.
- Suweda, I. (2009). Pentingnya Pengembangan Zona Selamat Sekolah Demi Keselamatan Bersama Di Jalan Raya (Suatu Tinjauan Pustaka). *Jurnal Ilmiah Teknik Sipil*, 13(1), 1–12.
- Vitianingsih, A. V. (2016). Game Edukasi Sebagai Media Pembelajaran Pendidikan Anak Usia Dini. *Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 1(1), 1–9.
- Wantoro, A., & Priandika, A. T. (n.d.). *DETERMINATION OF TARGET VALUE AND VALUE CONVERSION OF SCALE IN MATCHING PROFILE (PM) WITH COMBINATION METHOD ANALITYCAL HIERARCHY PROCESS (AHP) AS METHOD DEVELOPMENT IN SYSTEM DECISION SUPPORT*.
- Wiguna, P. D. A., Swastika, I. P. A., & Satwika, I. P. (2019). Rancang Bangun Aplikasi Point of Sales Distro Management System dengan Menggunakan Framework React Native. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 4(3), 149–159.
<https://doi.org/10.25077/teknosi.v4i3.2018.149-159>