# CLUSTERING PELANGGARAN LALU LINTAS PADA KENDARAAN BERMOTOR MENGGUNKAN ALGORITMA K-PROTOTYPE (STUDI KASUS: PENGADILAN NEGERI CIREBON)

**Nilta Saniyah, Nana Suarna, Willy Prihartono**
Teknik Informatika, Stmik Ikmi Cirebon
Jl. Perjuangan No.10B Kota Cirebon, Indonesia
*niltasaniyah8@gmail.com*

**ABSTRAK**

Fenomena sering terjadi pelanggaran lalu lintas yaitu salah satu dari problematika yang masih banyak mengakibatkan permasalan di jalan raya, seperti kecelakaan lalu lintas dan kemacetan. Disebabkan dari tingginya kasus pelanggaran yaitu minimnya pengetahuan serta kesadaran pengendara kendaraan khususnya kendaraan bermotor bahkan ada beberapa pengguna juga abai dalam mematuhi peraturan lalu lintas dijalan raya. Pada daerah sektor lalu lintas di Cirebon masih cukup tinggi. Penelitian ini dilakukan bertujuan untuk pengelompokkan dan analisis data pelanggaran lalu lintas di Pengadilan Negeri Cirebon di tahun 2022. Metode penelitian ini menerapkan algoritma *K-Prototype* untuk memudahkan dalam mengetahui jenis pelanggaran yang sering kali dilakukan oleh pengendara. Algoritma *K-prototype* merupakan salah satu metode analisis cluster pada data berukuran besar dengan bertipe campuran yang terdiri data numerik dan kategorikal. penelitian ini akan menggunakan sebanyak 3 atribut yang terdiri dari jenis pelanggaran, denda dan pasal. Setelah data dikumpulkan dan dilakukan analisis kemudian dilakukan pra-pemrosesan data dengan melakukan melakukan pembersihan data yang tidak valid data yang tidak diperlukan. Selanjutnya data diproses dengan metode clustering menggunakan algoritma *K-Prototype* untuk pengelompokkan jenis pelanggaran dengan atribut yang digunakan tahap berikutnya dilakukan analisis dan interpretasi terhadap hasil dari klasterisasi yang dilakukan sehingga dapat mengetahui pola yang terjadi. Hasil dari penelitian ini terdapat 2 cluster yang terdiri dari C1 yaitu pelanggaran paling sering dilakukan pada pasal 291 dan C2 yaitu pelanggaran yang jarang dilakukan pada pasal 280 dari pengujian yang optimal menggunakan *elbow analysis* didapatkan bahwa jumlah cluster yang optimal sebanyak 2 cluster.

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

## 1. PENDAHULUAN

Transportasi adalah salah satu unsur penting dalam sistem kehidupan masyarakat dan dapat menentukan efektifitas di suatu daerah. Peningkatan jumlah penduduk memiliki pengaruh besar terhadap lajunya transportasi. Transportasi dan mobilitas perpindahan penduduk menjadi hal yang tidak dapat dipisahkan. Perpindahan penduduk dan kegiatan ekonomi yang dilakukan sangat bergantung pada sistem transportasi yang digunakan. Transportasi yaitu sarana penghubung atau yang menghubungkan antara daerah produksi dan pasar, atau seringkali menjembatani produsen dengan konsumen. Transportasi atau pengangkutan dapat didefinisikan sebagai suatu proses pergerakan atau perpindahan orang atau barang dari suatu tempat ketempat lainnya dengan menggunakan suatu teknik atau cara tertentu untuk maksud dan tujuan tertentu.

Peningkatan jumlah penduduk disertai dengan peningkatan aktivitas penduduk menyebabkan meningkatnya kebutuhan akan kendaraan. Oleh karena itu terjadinya banyak pelanggaran lalu lintas menjadi salah satu bentuk problematika yang masih sering menimbulkan permasalahan dijalan raya, seperti kecelakaan dijalan raya dan kemacetan. Penyebab tingginya kasus pelanggaran lalu lintas adalah kurangnya pengetahuan dan kesadaran bahkan banyak juga pengguna kendaraan yang abai dalam mematuhi peraturan lalu lintas. Faktor -faktor yang sering menyebabkan pelanggaran lalu lintas yaitu manusia, kendaraan dan jalan raya. Masih tingginya kasus pelanggaran lalu lintas di Indonesia salah satunya di kota Kabupaten Cirebon memiliki kasus pelanggaran yang masih dikatakan cukup tinggi. Oleh karena itu dapat dilihat pada Pengadilan Negeri Cirebon yang memiliki kasus pelanggaran yang tinggi. Dapat dilihat dari *website* resmi milik Pengadilan Negeri Cirebon 2021 pada tiap bulannya. Pada bulan Januari tidak terdapat pelanggara, bulan Februari 157, bulan Maret 308, bulan April 711, bulan Mei 346, bulan Juni 426, bulan Juli 385, bulan Agustus 139, bulan September 1021, bulan Oktober 702, bulan November 575 bulan Desember 294 dan jumlah data pelanggaran pada tahun 2021 sebanyak 4.434 data ini terbilang cukup tinggi. Dari data tersebut dapat disimpulkan bahwa jumlah pelanggar lalu lintas masih relatif tinggi. Dimana data-data pelanggaran tesebut belum dikelompokkan, sehingga belum diketahui jenis pelanggaran apa saja yang kerap kali dilanggar oleh para pengendara.

Tujuan utama dari penelitian ini adalah untuk mengembangkan sebuah solusi berbasis teknologi informasi yang dapat memberikan pemahaman yang lebih mendalam mengenai tingginya kasus pelanggaran lalu lintas di Cirebon. Dengan menerapkan algoritma *K-Prototype* pada data

pelanggaran lalu lintas, penelitian ini bertujuan untuk menghasilkan hasil clustering yang akurat dan informatif. Hasil analisis data ini akan memberikan informasi yang berguna bagi Pengadilan Negeri Cirebon dalam menekan tingginya pelanggaran lalu lintas. Adapun *novelty* atau kontribusinya adalah untuk mengoptimalisasi peraturan lalu lintas dan menekan pelanggaran pada kendaraan bermotor.

Selain itu temuan penelitian ini dampaknya akan merambah ke teknologi untuk analisis pengolahan data pelanggaran. Pemanfaatan teknologi informasi dalam mengevaluasi dan menganalisis data pelanggaran dapat untuk mengurangi angka pelanggaran di Cirebon. Dengan demikian, penelitian ini memiliki potensi untuk memberikan kontribusi nyata terhadap masalah pelanggaran yang tinggi.

## 2. TINJAUAN PUSTAKA

### 2.1. Pelanggaran Lalu Lintas

Pada undang- undang No. 22 tahun 2009 pasal 265 tentang pemeriksaan kendaraan bermotor yang dijalankan petugas kepolisian atau penyidik pegawai negeri sipil dibidang lalu lintas dan angkutan jalan meliputi: surat izin mengemudi, surat tanda nomor kendaraan bermotor, surat tanda coba kendaraan bermotor, fisik kendaraan bermotor, tanda bukti lulus uji dan daya angkut kendaraan. Dalam hal ini jika pengendara tidak sesuai peraturan maka pada pasal 267 ayat 1 "setiap pelanggaran dibidang lalu lintas dan angkutan jalan yang dicek menurut pemeriksaaan cepat dapat dikenakan pidana denda sesuai dengan penetapan pengadilan" [1].

### 2.2. Algoritma K-Prototype

Algoritma ini ialah salah satu metode pengelompokkan yang berbasis partisi. Algoritma *K-prototype* adalah ekspansi di antara K-*Means* dan *K-Modes* untuk mengatasi clustering pada tipe data campuran numerik dan kategorikal. *K-means* cluster analysis merupakan metode analisis cluster non hirarki yang dapat digunakan untuk mempartisi objek kedalam kelompok-kelompok berdasarkan pendekatan karakteristik, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam cluster lain [2].

### 2.3. Lalu Lintas

Lalu lintas ialah undang-undang No. 22 tahun 2009 yang mendefinisikan sebagai laju kendaraan dan orang diruang lalu lintas jalanan, sedangkan yang dimaksud dengan ruang lalu lintas jalan yaitu prasarana yang diperuntukan bagi Gerakan pindah kendaraan, orang dan barang berupa jalan dan fasilitas mendukung. Menurut Djajoesman yang dimaksud secara harfia lalu lintas artian sebagai gerak(bolak balik) manusia atau barang dari satu tempat ke tempat yang lain dengan menggunakan sarana jalan umum [3].

### 2.4. Python

Python merupakan Bahasa pemograman interpreatif multiguna dengan perancangan yang terfokus pada skala keterbacaan kode. *Python* diakui sebagai Bahasa yang mengkombinasikan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas dan kelengkapan dengan fungsionalitas pustaka standar yang besar komprehensif. *Pytho*n bisa dibilang Bahasa pemograman dengan tujuan umum yang dikembangkan secara khusus untuk membuat *source code* mudah dibaca. Python juga memiliki *library* yang lengkap sehingga memungkinkan *programmer* untuk membuat aplikasi yang mutakhir dengan menggunakan *source code* yang tampak sederhana [4].

### 2.5. Clustering

Clustering ialah proses untuk pengorganisasian kelompok data didalam berbagai kelompok-kelompok sedemikian rupa maka objek-objek yang sama akan menjadi satu *cluster* sedangkan objek-objek yang tidak sama menjadi anggota cluster yang lain. Disetiap cluster berisikan data yang sama. Ukuran kemiripan biasanya dihitung dengan jarak. Jarak pada satu cluster dibuat sedekat mungkin dan jarak antar cluster diusahakan untuk sejauh mungkin. Jadi pada satu cluster harus serupa dan dengan cluster yang lain [5].

## 3. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode pendekatan kuantitatif yang merupakan jenis penelitian yang menghasilkan penemuan-penemuan yang dicapai(diperoleh) dengan prosedur-prosedur statistic atau cara lain dari kuantitatif [6], oleh karena itu ada beberapa tahapan dalam penelitian ini dengan melalui tahapan awal yaitu studi Pustaka dan literatur, pengumpulan data, evaluasi clustering, kesimpulan.

### 3.1. Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini yaitu melalui memanfatkan jurnal, buku atau prosiding dan artikel yang relevan sebagai media untuk bahan referensi pada penelitian ini melalui penelusuran literatur atau mengumpulkan data untuk mendapatkan data lain yang berhubungan dengan penelitian ini, oleh sebab itu peneliti mengambil data sekunder dimana data tersebut didapatkan dari sumber yang relevan dan resmi yaitu pada website resmi pengadilan negeri Cirebon berupa data perkara pelanggaran lalu lintas di Cirebon pada tahun 2022. Berikut data pelanggaran lalu lintas di pengadilan negeri Cirebon.

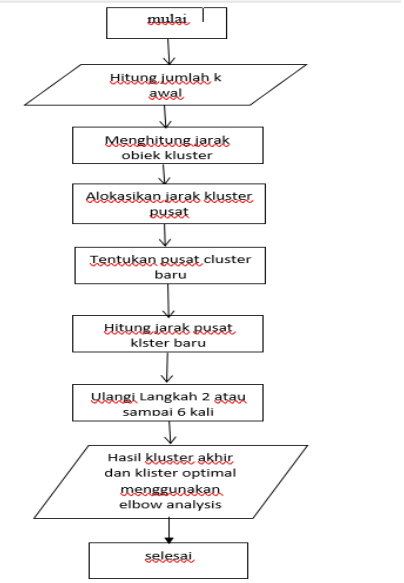
### 3.1. Teknik Analisis Data

Teknik analisis data dari penelitian ini menerapkan metode KDD (*Knowledge Discovery and in Database)*. KDD ialah *terminology computer science* yang mengacu pada proses pencarian informasi yang berkaitan dengan ekstraksi dan perhitungan data-data yang ditelaah dari kumpulan

data digital dalam jumlah besar, seperti pada database atau dataset[5].

a. Selection data
   Diawal tahapan ini terdapat tahap *selection* adalah pemilihan dan penyeleksian data yang akan digunakan. dimana penelitian ini hanya memilih atribut yang memenuhi tujuan dari penelitian ini. Atribut yang akan yaitu jenis pelanggaran, denda dan bukti. Selanjutnya data yang akan di *clustering* sejumlah 2325 Data.
b. Pre- processing
   Tahap *processing* ini dilakukan untuk membersihkan data yang tidak bisa untuk memasuki proses data mining, seperti data dan *missing value*. Cara untuk memperbaiki noise dapat melakukannya dengan cara yaitu penghapusan data, dengan diisi secara manual, dibiarkan, dengan diisi dengan rata-rata parameter, atau dengan rata-rata kelas.
c. Transformation
   Tahap ini yaitu melakukan perubahan data yang memiliki tipe data yang di awalnya bisa diolah secara matematis agar menjadi data yang dapat proses. Dengan tujuan untuk mencegah data yang rusak dan ketidakvalidan data, bentuk data dibagi menjadi beberapa kelompok dengan skala tertentu, dengan tujuan agar variasi data pada atribut tertentu menjadi sedikit.
d. Data mining
   Data mining merupakan suatu proses pengambilan atau penelusuran data yang belum diketahui sebelumnya, tetapi bisa dipahami dan bermanfaat dari *database* yang besar serta digunakan untuk mengambil suatu keputusan yang krusial [7].Pada tahap ini dilakukan clustering data pelanggaran lalu lintas menggunakan algoritma *K-Prototype. K-Protoype* mempunyai keunggulan karena algoritma yang tidak terlalu kompleks dan mampu menangani data yang besar serta lebih baik dibandingkan dengan algoritma yang berbasis hierarki. Oleh karena itu algoritma ini merupakan salah satu metode pengelompokkan yang berbasis partisi. Algoritma *K-prototype* adalah ekspansi antara *K-Means* dan *K-Modes* untuk mengatasi clustering dengan tipe data campuran numerik dan kategorikal. *K-means cluster analysis* merupakan metode analisis cluster non hirarki yang dapat digunakan untuk mempartisi objek kedalam kelompok-kelompok berdasarkan pendekatan karakteristik, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam cluster lain. Jika perubahan centeroid pada *K-Means cluster analysis* menggunakan rata-rata, maka *K-modes cluster analysis* pada perubahan centeroidnya menggunaka modus. K-Modes cluster analysis juga sama tahapan analisisnya dengan *K-means*. Perubahan yang mendasar terdapat pada pengukuran kesamaan (*similarity measure*) antara objek dengan *centeroid (prototype*)-nya.

Berikut alur dari algoritma *K-Prototype.*




Proses algoritma *K-prototype* diuraikan sebagai berikut:

1. Menentukan $k$ awal prototype yaitu $Z_1, Z_2 \ldots Z_k$ sebagai pusat cluster di masing-masing *cluster*.
2. Menghitung jarak objek pada dataset terhadap pusat cluster awal dengan menggunakan persamaan.

$$d(X_j, Z_i = \left( \sum_{l=1}^{m_r} (X_{jl}^r - Z_{il}^r)^2 + Y_i \sum_{l=l+1}^{m_c} \delta(X_{jl}^c, Z_{il}^c) \right)^{\frac{1}{2}}$$

3. Mengalokasikan setiap objek kedalam cluster yang memiliki jarak terdekat dengan pusat cluster awal.
4. Menentukan pusat cluster baru untuk masing-masing cluster yang telah dialokasikan lalu lakukan realokasi pada semua objek jika titik pusat terjadi perubahan maka proses akan berhenti. Proses in ikan terus dilakukan sampai tidak ada lagi perubahan prototype atau sampai kriteria stopping terpenuhi.

*Dissimilarity*: $\delta(x_j, y_j)$

$$\begin{cases} (x_j = y_j)0 \\ (x_j \neq y_j)1 \end{cases}$$

*Euclidean distance* (x,y) = $\sum_{j=1}^{p} (x_j - y_j)^2$

*Hamming distance*

$$d(x,y) = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|$$

5. Menghitung jarak objek pada dataset terhadap pusat cluster baru, dengan menggunakan persamaan seperti langkah 2.
6. Mengalokasikan setiap objek ke dalam cluster yang memiliki jarak terdekat dengan pusat cluster baru.
7. Mengulangi langkah 2 sampai 6 hingga tidak ada lagi perpindahan objek atau anggota masing-masing cluster tetap.

e. Interpretation dan Evaluation

Pada tahap *interpretation* yaitu proses untuk menganalisis pengklusteran menggunakan algoritma K-Prototype dengan menggunakan metode *Elbow* analisis. Metode *Elbow* merupakan 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, jika nilai cluster pertama dan nilai cluster kedua memeberikan sudut dalam grafik dan nilainya mengalami penurunan besar maka jumlah cluster tersebut tepat. Yang menggunakan perbandingannya dengan menghitung *Sum of Square Error (SEE)* nilai dari masing-masing cluster[8]. Untuk mendapatkan perbandingannya adalah dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai cluster. Karena semakin besar jumlah cluster k maka nilai SSE akan semakin kecil. untuk mengukur titik data dalam suatu cluster dengan pusatnya menggunakan metrik SSE (*Sum Of Squared Erros*) berikut ini:

1. Inisialisasi awal nilai *k.*
2. Menaikan nilai k sampai jumlah cluster yang ditentukan.
3. Menghitung nilai SSE dari setiap k.

$$SSE = \sum_{k=1}^{k}\sum_{x_i \in s_k}\|X_i - C_k\|_2^2$$

k : jumlah *cluster* yang digunakan pada algoritma *K-prototype*
$X_i$ : nilai atribut dari data ke-*i*
$C_k$ : jumlah *cluster i* pada cluster ke-*k*

4. Melakukan perhitungan SSE sampai k yang ditentukan.
5. menganalisis hasil SSE yang nilai k-nya turun secara signifikan.
6. Menetapkan nilai cluster dilihat dari hasil SSE yang nilai k-nya turun secara signifikan.

## 4. HASIL DAN PEMBAHASAN

Hasil penelitian ini menjelasan tentang penerapan algoritma *K-Prototype* berdasarkan data pelanggaran lalu lintas yang diperoleh dari pengadilan negeri Cirebon. Proses pengelompokkan yang dilakukan menggunakan *Python* pada *Google Collab*. Hasil penelitian ini bertujuan untuk menyajikan fakta-fakta yang ditemukan dari data tersebut, memberikan informasi tentang temuan yang sebenarnya

### Data Selection

Read data serta head, read data dapat diterapkan untuk memasukkan data dengan bentuk *CSV* atau dengan bentuk yang lain. Sedangkan *library head* diterapkan untuk menunjukkan data dari *CSV* yang sudah diimport, dan library head juga dapat melihat 5 baris pertama dari data frame . berikut tampilan dari hasil read data ditunjukkan pada gambar 1.

```
#read data
df=pd.read_csv('/content/drive/MyDrive/tilang.csv')
df
```

| | Nomor Perkara | Plat | Nama Pelanggar | Bukti | Pasal | Denda |
|---|---|---|---|---|---|---|
| 0 | 4652/Pid.LL/2022/PN Cbn | E4008BD | Aldy | STNK | 291 | 49000 |
| 1 | 4583/Pid.LL/2022/PN Cbn | E4985BR | Salbi Agung D U | STNK | 291 | 49000 |
| 2 | 4628/Pid.LL/2022/PN Cbn | E5467BM | Rizky M | STNK | 291 | 49000 |
| 3 | 4629/Pid.LL/2022/PN Cbn | E6758BR | Mumu Mumhaimin | STNK | 291 | 49000 |
| 4 | 4586/Pid.LL/2022/PN Cbn | E4098BS | Moh. Nadip Naufal | STNK | 291 | 49000 |

Gambar 1. Read Data

Selanjutnya setelah read data, langkah berikutnya adalah melakukan seleksi data yang akan digunakan pada proses pengolahan bertujuan untuk memastikan data yang digunakan tidak terdapat kesalahan data selanjutnya proses menghilangkan atribut yang tidak digunakan dalam pengolahan data yaitu nomor perkara, plat dan nama pelanggar dengan perintah *df.drop()* yaitu perintah untuk manghapus baris atau kolom tertentu dengan menentukan sumbu kolom ataupun sumbu baris yang ditunjukan pada gambar 2

```
#menghapus data yang tidak dipakai
df.drop(['Nomor Perkara', 'Plat', 'Nama Pelanggar'], 1, inplace=True)
df
```

```
<ipython-input-6-bb4278811122>:2: FutureWarning: In a future version o
```

| | Bukti | Pasal | Denda |
|---|---|---|---|
| 0 | STNK | 291 | 49000 |
| 1 | STNK | 291 | 49000 |
| 2 | STNK | 291 | 49000 |
| 3 | STNK | 291 | 49000 |
| 4 | STNK | 291 | 49000 |

Gambar 2. Drop Data

### 4.1. Data Preprocessing

validasi data missing value tersebut bertujuan untuk dapat menampilkan apakah masih terdapat missing value pada data. Pada gambar 3 yang menghasilkan tidak adanya perubahan data dan masih tetap sama berjumlah 2325 data tahun 2022, dikarenakan data yang digunakan merupakan data lengkap tidak ada data kosong, data ganda ataupun *missing.*

```
#cek missing value
df.isna().sum()
```

```
Bukti    0
Pasal    0
Denda    0
dtype: int64
```

Gambar 3. Missing Value

### 4.2. Data transformasi

Standarisasi atau normalisasi adalah salah satu teknik persiapan data yang paling sering digunakan untuk membantu mengubah nilai kolom numerik pada dataset untuk menggunakan skala umum sehingga atribut berada dalam rentang yang lebih kecil, seperti -1 hingga 1 ataupun 0 hingga 1. Normalisasi pada penelitian ini menggunakan fungsi *minMax_normalize* atau disebut dengan *linear normalization* yang digunakan untuk mempertahankan bentuk data distribusi dan nilai pasti dari data minimum dan maksimum. Setelah di standarisasi atau dinormalisasi data pada atribut pada denda akan berubah sesuai dengan normalisasi yang dilakukan. Kemudian tampilan hasil dari normalisasi data dapat dilakukan dengan perintah *df()* atau yang dimaksud tampilan dari Dataframe Proses standarisasi dari data numerikal dapat dilihat pada gambar 4.

```
df["Denda"]=(df["Denda"]-df["Denda"].min())/(df["Denda"].max()-df["Denda"].min())
df
```

| | Bukti | Pasal | Denda |
|---|---|---|---|
| 0 | STNK | 291 | 0.000000 |
| 1 | STNK | 291 | 0.000000 |
| 2 | STNK | 291 | 0.000000 |
| 3 | STNK | 291 | 0.000000 |
| 4 | STNK | 291 | 0.000000 |

Gambar 4. transformasi data

### 4.3. Data Mining

Data mining merupakan suatu proses menggali sekumpulan data dan mengubah data dalam bentuk informasi yang bermanfaat. Data mining juga memiliki beberapa teknik yang sering digunakan oleh peneliti, diamtaranya clustering, classification, association dan yang lainnya [9]. Setelah proses preprocessing untuk standarisi nilai numerikal dapat dilakukannya proses pengolahan data atau training data merupakan bagian dari kumpulan dataset yang disediakan untuk menjadi bahan pembelajaran model dan dapat menentukan *cluster* yang diinginkan sehingga mendapatkan informasi terbaru. Pada proses training dapat dapat menggunakan perintah *model=KPrototype(n_cluster=2,verbose=2_,max_iter=10)* yang bermaksud menentukan model *k-prototype* dengan jumlah 2 kluster dengan jumlah iterasi atau baisa disebut eksekusi berulang yang dilakukan sebanyak 10 setelah perintah iterasi yaitu perintah untuk pengelompokkan kluster menggunakan data array dan data kategorikal. Gambar 5 memperlihatkan perintah yang digunakan pada proses pengolahan data, seperti yang telah ditentukan pada penelitian ini menggunakan 2 kluster.

```
#training data
model=KPrototypes(n_clusters=2,verbose=2,max_iter=10)
klaster=model.fit_predict(my_array,categorical=[0])
```

```
Initialization method and algorithm are deterministic. Setting n_init to 1.
Init: initializing centroids
Init: initializing clusters
Starting iterations...
```

Gambar 5. training data

Setelah *training* data, proses selanjutnya yaitu *array*. *Array* yang merupakan tipe data terstruktur dalam pemrograman dan variabel yang menyimpan lebih dari 1 buah data yang memiliki sifat atau tipe data yang sama. Dapat dikatakan juga bahwa array merupakan kumpulan dari data tunggal yang dijadikan dalam 1 variabel. array juga dapat memungkinkan untuk menyimpan data maupun referensi objek dalam jumlah terindeks dan dalam jumlah banyak. Berikut tampilan dari hasil training data *array* ditunjukkan pada gambar 6.

```
my_array=df.values
my_array[:,1]=my_array[:,1].astype(float)
my_array[:,2]=my_array[:,2].astype(float)
my_array
```

```
array([['STNK ', 291.0, 0.0],
       ['STNK', 291.0, 0.0],
       ['STNK', 291.0, 0.0],
       ...,
       ['STNK', 280.0, 0.2857142857142857],
       ['STNK', 280.0, 0.2857142857142857],
       ['STNK', 280.0, 0.2857142857142857]], dtype=object)
```

Gambar 6. Hasil cluster

Kemudian proses selanjutnya yaitu menentukan centroid. Centroid merupakan titik pusat *cluster* jumlah centroid identic dengan jumlah cluster. Penentuan centroid awal ditentukan secara melalui titik pusat centroid baru yang dilakukan dengan proses berulang atau dapat dilakukan dengan perintah untuk melihat *Centroid* dapat menggunakan perintah *print(model.cluster_centroid_)*. Berikut tampilan dari *cluster centroid* terdapat pada gambar 7.

```
print(model.cluster_centroids_)
```

```
[['290.0797186400938' '0.16680623011220982' 'STNK']
 ['280.0' '0.28479113777982895' 'STNK']]
```

Gambar 7. Centroid

Selanjutnya data yang telah diolah dikelompokkan dengan menambah atribut baru yaitu atribut member untuk mempermudah dalam proses menganalisis data, dengan menggunakan printah *cluster_dict* atau kluster terbaru dan perintah *cluster_dict.append(c)* yaitu penambahan nilai array kedalam kluster yang sudah ditentukan ditunjukkan pada gambar 8.

```
#proses pengelompokkan data
cluster_dict=[]
for c in klaster:
    cluster_dict.append(c)
```

Gambar 8. pengelompokkan data

Pada proses akhir adalah ditampilkan data yang sudah diolah serta tampilan pengelompokkan kluster di data pelanggaran lalu lintas. Data yang diolah dengan *tools python* yaitu sebanyak 2325, 3 atribut serta menggunakan 2 cluster. Hasil kluster ditunjukkan pada gambar 9.

```
#hasil cluster
df['cluster']=cluster_dict
df()
```

| | Bukti | Pasal | Denda | cluster |
|---|---|---|---|---|
| **0** | STNK | 291 | 0.000000 | 0 |
| **1** | STNK | 291 | 0.000000 | 0 |
| **2** | STNK | 291 | 0.000000 | 0 |
| **3** | STNK | 291 | 0.000000 | 0 |
| **4** | STNK | 291 | 0.000000 | 0 |

Gambar 9. Hasil cluster

### 4.4. Interpretation dan Evaluation

Pada proses selanjutnya evaluasi dan analisis hasil cluster dengan menampilkan tabel pada setiap cluster seperti ditampilkan pada gambar 10 untuk cluster 1 itu diisi dengan pasal 291 yaitu pengendara kendaraan bermotor tidak menggunakan helm. Kemudian pada gambar 11 tampilan dari data cluster 2 yang diisi dengan pasal 280 yaitu pengemudi kendaraan bermotor tidak menggunakan tanda nomor kendaraan(plat).

```
df[df['cluster']== 0].head(5)
```

| | Bukti | Pasal | Denda | cluster |
|---|---|---|---|---|
| **0** | STNK | 291 | 0.0 | 0 |
| **1** | STNK | 291 | 0.0 | 0 |
| **2** | STNK | 291 | 0.0 | 0 |
| **3** | STNK | 291 | 0.0 | 0 |
| **4** | STNK | 291 | 0.0 | 0 |

Gambar 10. Kluster 1

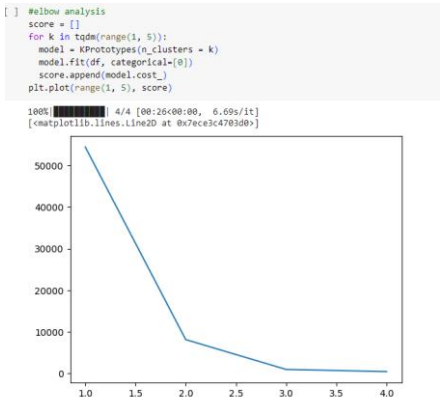
```
df[df['cluster']== 1].head(5))
```

| | Bukti | Pasal | Denda | cluster |
|---|---|---|---|---|
| **12** | STNK | 280 | 0.285714 | 1 |
| **29** | STNK | 280 | 0.285714 | 1 |
| **34** | SIM C | 280 | 0.285714 | 1 |
| **36** | STNK | 280 | 0.285714 | 1 |
| **37** | STNK | 280 | 0.285714 | 1 |

Gambar 11. Kluster 2

Selanjutnya implementasi metode *Elbow* dalam penentuan cluster optimal berdasarkan. Yang menggunakan perbandingannya dengan perhitungan *Sum of Square Error (SEE)* nilai dari setiap cluster. Proses ini dilakukan dengan menerapkan *Python* di *Google Collab*. Dan penelitian ini juga menjelaskan jarak antar titik pusat cluster dengan menerapkan *Euclidean Distance* yaitu jarak antar titik pusat cluster yang bertipe numerikal dan *Hamming Distance* yaitu jarak antar pusat titik yang bertipe kategorikal.

Metode *Elbow* yaitu metode yang diterapkan untuk menghasilkan informasi dalam mencakup jumlah cluster optimal dengan cara melihat rata-rata hasil perbandingan antara jumlah cluster yang akan terbentuk siku di satu titik, jika nilai *cluster* 1 dan nilai *cluster* ke 2 memberikan sudut pada grafik dan terjadinya penurunan besar pada nilainya maka jumlah cluster tersebut adalah tepat. Dapat menggunakan perbandingan dengan perhitungan *Sum of Square Error (SEE)* nilai pada setiap *cluster* [10]. Berikut proses *elbow* analisis ditampilkan pada Gambar 12.




Gambar 12. Metode elbow

Penentuan jumlah kluster dilakukan secara berurutan dimulai dari jumlah kluster sebanyak 1 hingga 5 cluster. Dari hasil pengolahan data menggunakan analisis elbow menghasilkan cluster optimal sebanyak 2 cluster yang terdiri dari C1 pasal yang penggaran paling tinggi yaitu pasal 291, C2 jarang dilakukan pelanggaran yaitu pasal 280.

## 5. KESIMPULAN DAN SARAN

Berdasarkan temuan dari analisis dan penelitian pada data pelanggaran lalu lintas pada Pengadilan Negeri Cirebon yang sudah dilakukan bisa disimpulkan yaitu pengelompokkan pelanggaran lalu lintas pada pengadilan negeri Cirebon menggunakan teknik algoritma *K-Prototype* yang dapat di implementasikan untuk pengelompokkan data pelanggaran lalu lintas pada Pengadilan Negeri Cirebon sangat efisien dikarenakan yang memiliki data campuran, Dengan dibuktikan pada hasil pengolahan data diatas menghasilkan cluster sebanyak 2 cluster yang terdiri dari C1 pelanggaran paling tinggi yaitu pasal 291, C2 pelanggaran paling tinggi yaitu pasal 280. Dibuktikan dengan pemilihan kluster terbaik pada metode *elbow* pada gambar 4.21 merupakan penentuan jumlah kluster dilakukan secara berurutan dimulai dari kluster sebanyak 1 hingga 5 kluster. Dari hasil pengolahan data digunakan analisis elbow dapat menghasilkan kluster optimal adalah 2 cluster. Kesimpulan penelitian yang telah dilakukan, peneliti menyadari bahwa terdapat kekurangan dalam penelitian yang dilakukan. Hal ini terkait dengan batasan waktu, keterbatasan pengetahuan, dan juga keterbatasan dalam ruang lingkup topik penelitian. Untuk mengatasi kekurangan ini dan meningkatkan kesempurnaan penelitian, sejumlah saran dapat diberikan adalah penelitian ini sebatas penerapan pada algoritma *K-Prototype* diharapkan pada penelitian selanjutnya dapat dikembangkan dan diimplementasikan kedalam Sistem Informasi agar lebih efisien secara waktu, serta ada penelitian ini penulis menggunakan 3 atribut yaitu 2 atribut numerikal dan 1 atribut kategorikal yang diiharapkan pada penelitian selanjutnya atribut yang digunakan lebih banyak dibandingkan dengan atribut yang penulis gunakan pada penelitian ini.

## DAFTAR PUSTAKA

[1] B. RI, "UU No.22 Tahun 2009 Peraturan Presiden Republik Indonesia," *Demogr. Res.*, pp. 4–7, 2009.

[2] R. Wijayati, D. Retno, and S. Saputro, "Clustering Data Campuran Numerik Dan Kategorik Menggunakan Algoritme K-Prototype," vol. 6, pp. 702–706, 2023.

[3] Sunaryo, M. Fakih, R. Syamsiar, and Kasmawati, "Peningkatan Kesadaran Hukum Masyarakat Terhadap Mewujudkan Terciptanya Tertib," *Sakai Sambayan*, vol. 4, no. 2, pp. 156–164, 2020, [Online]. Available: http://jss.lppm.unila.ac.id/index.php/ojs/article/view/186/153

[4] Muhammad Al Faruqi, "Pemrograman Phyton Pada Citra Digital," *Unikom*, pp. 12–26, 2021.

[5] T. Tendean and W. Purba, "Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means," *J. Sains dan Teknol.*, vol. 1, no. 2, pp. 5–11, 2020.

[6] R. B. Pratama, "Metodologi Penelitian," *Angew. Chemie Int. Ed. 6(11), 951–952.*, pp. 28–55, 2019.

[7] Library Binus, "Bab 2 landasan teori," *Apl. dan Anal. Lit. Fasilkom UI*, pp. 4–25, 2006.

[8] M. R. Koni, I. Djakaria, and N. I. Yahya, "Pengelompokkan Kabupaten/Kota Provinsi Jawa Timur Berdasarkan Indikator Kesejahteraan Rakyat Menggunakan Metode Elbow dan Algoritma K-Prototype," *Estimasi J. Stat. Its Appl.*, vol. 4, no. 1, pp. 10–19, 2023, doi: 10.20956/ejsa.vi.24837.

[9] E. Elisawati, D. Wahyuni, and A. Arianto, "Analisa Clustering Pada Data Pelanggaran Lalulintas Di Pengadilan Negeri Dumai Dengan Menggunakan Metode K-Means," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 4, no. 2, p. 1, 2019, doi: 10.14421/jiska.2019.42-01.

[10] D. A. I. C. Dewi and D. A. K. Pramita, "Comparative Analysis of Elbow and Silhouette Methods On K-Medoids Clustering Algorithm in Balinese Craft Production Grouping," *Matrix J.*, vol. 9, no. 3, pp. 102–109, 2019.

[11] D. H. Murti, N. Suciati, and D. J. Nanjaya, "Clustering Data Non-Numerik Dengan Pendekatan Algoritma K-Means Dan Hamming Distance Studi Kasus Biro Jodoh," *JUTI J. Ilm. Teknol. Inf.*, vol. 4, no. 1, p. 46, 2005, doi: 10.12962/j24068535.v4i1.a245.