ANALISIS KLASTERISASI PENGGUNA SIM DI KABUPATEN MANOKWARI MENGGUNAKAN ALGORITMA K-MEANS

Yansi Ratte, Christian Dwi Suhendra, Lorna Yertas Baisa
Teknik Informatika, Universita Papua
Jalan Gunung Salju Amban Manokwari, Indonesia
c.suhendra@unipa.ac.id

ABSTRAK

Surat Izin Mengemudi (SIM) merupakan kartu identitas resmi yang harus dimiliki oleh setiap pengendara sepeda motor. Faktanya di Kabupaten Manokwari banyak masyarakat yang tidak memenuhi peraturan tersebut. Akibatnya, kasus pelanggaran lalu lintas semakin tinggi, sehingga untuk mengetahui pola pengguna SIM di perlukan klasterisasi dengan algoritma K-Means menggunakan data peserta uji SIM dengan atribut Umur, Jenis Pekerjaan dan Golongan SIM. Jumlah data sebanyak 1.155. Data tersebut di analisis menggunakan aplikasi *Jupyter notebook*, dengan menggunakan metode Elbow method di peroleh K=4 dengan nilai error SSE sebesar 48960.67. Nilai K tersebut akan di proses menggunakan algoritma K-Means berdasarkan Umur dan Jenis pekerjaan. Hasilnya *Cluster*_0 didominasi oleh pelajar/mahasiswa berusia 20 tahun, sementara *Cluster*_1 mayoritas berusia 25 tahun dan bekerja sebagai wiraswasta. *Cluster*_2 memiliki rata-rata usia 44 tahun dengan dominasi pekerja ASN, sedangkan *Cluster*_3 berusia 39 tahun dengan mayoritas bekerja sebagai wiraswasta. Hasil ini dapat digunakan oleh pihak kepolisian untuk menyusun program edukasi dan kebijakan lalu lintas yang lebih efektif guna mengurangi angka pelanggaran lalu lintas di Kabupaten Manokwari

Kata kunci : *Pelanggaran Lalu Lintas, SIM (Surat Izin Mengemudi), Elbow method, Klasterisasi, Algoritma K-Means, Aplikasi Jupyter Notebook.*

1. PENDAHULUAN

Transportasi darat merupakan bagian penting dalam kehidupan masyarakat modern, baik untuk kepentingan individu maupun kelompok. Seiring dengan meningkatnya jumlah penduduk dan bertambahnya kendaraan bermotor, arus lalu lintas menjadi semakin padat, terutama di kota-kota besar [1].

Kepolisian Negara Republik Indonesia (Polri) memiliki peran yang sangat penting dalam menjaga keamanan, ketertiban masyarakat, menegakkan hukum, melindungi, mengayomi, serta memberikan pelayanan publik kepada masyarakat. Salah satu bentuk pelayanan yang paling langsung dirasakan oleh masyarakat adalah pembuatan Surat Izin Mengemudi (SIM), yang merupakan pelayanan administratif dasar yang sangat penting. Aturan mengenai pembuatan SIM telah diatur secara jelas dalam Peraturan Kepala Kepolisian Republik Indonesia Nomor 9 Tahun 2012. Pada Pasal 52 Ayat 2 disebutkan bahwa kewenangan penerbitan SIM berada di tangan kepala kepolisian di masing-masing daerah, tetapi kewenangan tersebut didelegasikan kepada kepala satuan lalu lintas (Kasat Lantas) di wilayah tersebut. Artinya, penerbitan SIM memiliki dasar hukum yang jelas dan tidak dilakukan secara sembarangan[2].

Meskipun aturan tentang pembuatan SIM sudah jelas, masih banyak terjadi pelanggaran lalu lintas, seperti pengendara yang tidak memiliki SIM. Selain itu, banyak anak-anak yang belum cukup umur sudah mengendarai kendaraan bermotor di jalan raya. Salah satu alasan utama rendahnya pengurusan SIM adalah kurangnya pengetahuan masyarakat tentang cara mengurus SIM. Di samping itu, anggapan bahwa SIM tidak terlalu penting dan minimnya sosialisasi membuat sebagian orang tetap memilih untuk berkendara tanpa SIM. Situasi ini meningkatkan risiko kecelakaan dan pelanggaran lalu lintas. Kurangnya pengawasan juga menyebabkan banyak pengendara lolos dari tilang, sehingga kepatuhan hukum semakin menurun. Bahkan, ada masyarakat yang memilih jalur ilegal, seperti membuat SIM tembak, yang merugikan negara dan melemahkan aturan lalu lintas.

Analisis data peserta uji SIM berperan dalam memahami pola-pola partisipasi masyarakat berdasarkan karakteristik tertentu. Melalui pengelompokan data, pihak berwenang dapat mengidentifikasi kelompok masyarakat dengan tingkat partisipasi tinggi maupun rendah dalam pengurusan SIM. Informasi ini dapat dimanfaatkan untuk merancang strategi yang lebih efektif guna meningkatkan kesadaran dan partisipasi dalam pengurusan SIM, khususnya pada kelompok yang masih kurang aktif.

Penelitian sebelumnya telah membahas penerapan metode k-nearest neighbors (KNN) dalam klasifikasi golongan SIM di Kabupaten Manokwari [3].

Penelitian tersebut menggunakan atribut Golongan SIM, Umur, dan Jenis Permohonan untuk mengelompokkan data. Sementara itu, penelitian yang di lakukan pada saat ini menggunakan metode algoritma K-Means dengan atribut Umur, Jenis pekerjaan, dan Golongan SIM. Penggunaan algoritma K-Means bertujuan untuk mengelompokkan data peserta uji SIM berdasarkan karakteristik yang serupa, seperti atribut Usia dan Jenis pekerjaan. Melalui pengelompokan data ini, kelompok yang kurang aktif

dalam mengurus SIM dapat dikenali, sehingga strategi yang tepat dapat disusun untuk meningkatkan kesadaran masyarakat, seperti melalui sosialisasi langsung, penyebaran informasi tentang pentingnya memiliki SIM, serta mempermudah proses pengurusannya agar lebih banyak masyarakat yang ingin mengurus SIM.

2. TINJAUAN PUSTAKA

2.1. Surat Izin Mengemudi (SIM)

Surat Izin Mengemudi (SIM) adalah dokumen yang berfungsi sebagai identitas pengendara kendaraan bermotor, yang juga digunakan sebagai bukti data forensik oleh kepolisian. SIM diberikan kepada seseorang yang telah berhasil melewati ujian pengetahuan, keterampilan, dan kemampuan mengemudi sesuai dengan aturan yang ditetapkan dalam Undang-Undang Lalu Lintas dan Angkutan Jalan [4].

Proses pemberian SIM ini bertujuan untuk memastikan bahwa pengendara memiliki kompetensi yang diperlukan dalam mengoperasikan kendaraan dengan aman, serta memahami peraturan lalu lintas yang berlaku. Dengan demikian, SIM tidak hanya berfungsi sebagai identitas, tetapi juga sebagai indikator bahwa pengendara telah lulus ujian dan siap untuk mengemudi di jalan raya dengan mematuhi semua ketentuan yang ada [5].

2.2. Data Mining

Data mining atau biasa di sebut sebagai maching learning atau kecerdasan buatan adalah proses mengidentifikasi fakta dari banyaknya data yang besar untuk menemukan fakta informasi yang di inginkan dengan menggunakan metode dan Teknik yang beragam terkait dengan penelitian. Selain itu data mining di gunakan untuk mengidentifikasi informasi yang terkait dari berbagai database [6].

Salah satu contoh penerapan data mining adalah menggunakan algoritma K-Means *clustering*. Dalam hal ini penerapan data mining dengan algoritma K-Means *clustering* dapat menjadi solusi yang tepat untuk mengidentifikasi pola karakteristik peserta uji SIM di Kabupaten Manokwari.

2.3. Pengoptimalan K-Means

Menentukan jumlah *cluster* optimal menggunakan metode Elbow. Metode Elbow adalah teknik untuk menentukan jumlah *cluster* yang optimal dalam analisis data. Metode ini bekerja dengan membandingkan hasil dari berbagai jumlah *cluster* dan melihat titik di mana penurunan nilai kesalahan SSE mulai berkurang secara signifikan. Titik ini disebut sebagai Elbow karena bentuk grafiknya menyerupai siku. Dari sini, jumlah *cluster* yang paling sesuai dapat dipilih untuk mengelompokkan data dengan baik [7].

Berikut adalah tahapan dalam metode Elbow :

a. Menentukan jumlah K yang akan di uji dari 1-10
b. Menerapkan K-Means untuk setiap nilai K yang sudah di tentukan
c. Menghitung nilai SSE (Sum of Squared Errors)

$$SSE = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(x_i - C_j\right)^2 \qquad (1)$$

Yang dimana:
x_i = titik data dalam *cluster*
C_j = pusat *cluster* terdekat
k = jumlah *cluster*
n = jumlah total data

d. Membuat grafik Elbow
e. Menentukan titik Elbow

2.4. Sum of Square Error (SSE)

SSE adalah cara untuk menghitung seberapa jauh data asli berbeda dari hasil yang diperkirakan oleh model. Dalam K-Means, SSE digunakan untuk menentukan jumlah *cluster* yang paling baik. Caranya, setiap titik data dibandingkan dengan pusat klasternya, lalu selisihnya dikuadratkan dan dijumlahkan. Jika SSE kecil, berarti *cluster* yang dibuat lebih akurat karena datanya lebih dekat ke pusat *cluster* [8].

Rumus SSE telah dijelaskan pada bagian 2.3. Pengoptimalan K-Means, sebagaimana tercantum dalam rumus (1) . Rumus ini digunakan untuk menghitung total kesalahan kuadrat antara setiap titik data dengan pusat *cluster* terdekatnya dalam algoritma K-Means.

2.5. Algoritma K-Means *clustering*

K-Means merupakan algoritma *clustering* yang di gunakan untuk mengelompokkan suata data berdasarkan variable yang sama dengan menetapkan titik pusat *cluster* (centroid) dan mengelompokkan data yang berdekatan dengan centroid. Dari semua data tersebut akan di masukkan ke dalam *cluster* dan setiap centroid yang berisi anggota [9].

K-Means *clustering* merupakan metode pengelompokkan data dalam algoritma berdasarkan kemiripan karakteristiknya [10].

Berikut adalah urutan proses algoritma K-Means:

a. Pemilihan titik awal *cluster*
b. Pengukuran jarak setiap data ke masing-masing *centroid* . Menggunakan rumus Euclidean distance sebagai berikut :

$$d(x,y) = \sqrt{\sum(x_i - y_i)^2} \qquad (2)$$

Yang dimana :
$d(x,y)$ = Jarak anatara 2 titik data x dan data y
x_i = Nilai x dari anggota ke-i dalam data
y_i = Nilai y dari anggota ke-i dalam data

c. Mengelompokkan data *cluster* dengan centroid
d. Menyesuaikan letak centroid dengan menggunakan persamaan berikut :

$$C_j = \frac{1}{n_i}\sum_{i=1}^{n_j} x_1 \qquad (3)$$

Yang dimana:
C_j = Centroid baru untuk *cluster* ke-j
n_j = Jumlah data dalam *cluster* ke-j
x_i = data ke-i dalam *cluster*

e. Kembali ke proses dari b – d hingga hasilnya stabil

3. METODE PENELITIAN

3.1. Metode Penelitian

Dalam penelitian ini, peneliti menggunakan pendekatan kuantitatif karena fokus utama penelitian adalah menganalisis data numerik untuk menemukan pola tertentu melalui algoritma K-Means. Pendekatan ini dianggap paling tepat karena mampu memberikan hasil yang objektif dan terukur dalam proses klasterisasi data peserta uji SIM di Kabupaten Manokwari. Analisis dilakukan dengan menggunakan atribut Umur, Jenis pekerjaan, dan Golongan SIM, dengan tujuan untuk mengidentifikasi pola atau karakteristik umum dalam kelompok data yang sama. Penelitian ini menggunakan aplikasi Jupyter Notebook sebagai alat yang digunakan untuk mengelola data. Aplikasi Jupyter notebook yaitu singkatan dari kata Julia (Ju), Python (Py), dan R. Sedangkan aplikasi Jupyter notebook merupakan aplikasi online gratis untuk membuat dan berbagi berupa kode, hasil perhitungan, visualisasi hasil analisis, dan teks dalam membuat narasi komputasi [11].

Jupyter Notebook dipilih karena dapat menganalisis data dengan mudah dan juga dapat di gunakan secara gratis. berikut adalah alur kerja penelitian.

Gambar 1. Alur kerja penelitian

3.2. Sumber Data

Data yang di gunakan dalam penelitian ini adalah data yang di dapatkan dari SATPAS POLRES MANOKWARI dengan jumlah 1.155 entri data yang terdiri dari 10 atribut yaitu Nama, Jenis Kelamin, Tempat Lahir, Tanggal Lahir, Pekerjaan, Alamat, Kota, Provinsi, Jenis Permohonan dan Golongan SIM. Data ini di dapatkan pada tahun 2023

3.3. Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini menggunakan metode pengumpulan data sekunder, yaitu data yang telah dikumpulkan dan digunakan dalam penelitian sebelumnya sebagai tugas dari salah satu mata kuliah. Data ini diperoleh dari hasil kerja praktik mahasiswa di SATPAS Polresta Manokwari pada tahun 2023 dan digunakan kembali untuk dianalisis lebih lanjut dalam penelitian ini

3.4. Teknik Analisis

Setelah melewati tahap pre-processing, data dianalisis menggunakan metode K-Means *clustering*. Metode ini mengelompokkan data peserta uji SIM ke dalam beberapa *cluster* berdasarkan kesamaan atribut, seperti Umur, Jenis pekerjaan, dan Golongan SIM. Untuk menentukan jumlah *cluster* yang optimal, dilakukan 10 kali percobaan menggunakan metode Elbow, dengan mengacu pada nilai SSE. SSE digunakan karena dapat mengukur seberapa dekat data dalam satu *cluster* dengan pusatnya. Semakin kecil nilai SSE, semakin baik hasil pengelompokan. Dibandingkan dengan error lainnya, SSE lebih cocok untuk K-Means karena dapat menunjukkan titik optimal jumlah *cluster* melalui metode Elbow.

4. HASIL DAN PEMBAHASAN

4.1. Pre- processing

Pre-processing data yaitu suatu fase yang di mana di anggap penting karena sebelum melakukan analisis data perlu untuk membersihkan data yang akan di gunakan, biasanya pada data mentahan sering kali datanya tidak terstruktur, kurang lengkap sehingga perlu untuk di bersihkan, selain itu, data juga harus ditransformasi agar data tersebut dapat di gunakan secara efektif dalam model analisis. Berikut adalah tahap pembersihan data peserta uji SIM di Kabupaten Manokwari :

4.1.1. Data Selection

Pada tahap ini proses bertujuan untuk memilih data yang akan di gunakan. Berikut adalah data yang akan di gunakan dalam penelitian :

Tabel 1. Data yang digunakan

No	Tanggal Lahir	Pekerjaan	Gol. SIM
1	4/8/1982	WIRASWASTA	C
2	4/9/1982	WIRASWASTA	A
3	7/9/1996	HONORER	C
4	5/10/1979	PENGEMUDI	BI Umum
5	4/20/1967	ASN	A
6	4/3/2000	PEG.BUMN	C
7	4/11/2001	PEG.BUMN	C
8	4/27/1987	HONORER	C
9	8/6/1976	WIRASWASTA	C
10	12/3/2004	PELAJAR/MHS	C
...	...	...	...
1146	3/7/1980	WIRASWASTA	C
1147	31/05/1992	WIRASWASTA	C
1148	20/05/1992	WIRASWASTA	C
1149	20/05/1992	WIRASWASTA	A
1150	26/05/1966	KARYAWAN SWASTA	C
1151	26/05/1966	KARYAWAN SWASTA	A
1152	18/01/2001	PELAJAR/MHS	C
1153	14/12/1997	PELAJAR/MHS	A
1154	25/01/1992	KARYAWAN SWASTA	C
1155	16/08/2000	PELAJAR/MHS	C

Pada tabel 1 diatas dapat dilihat dataset yang akan dipakai, dataset yang di gunakan terdiri dari tiga

atribut yaitu Tanggal Lahir, Pekerjaan dan atribut tambahan yaitu Gol. SIM**.**

4.1.2. Transformasi

Pada tahap ini, data akan dikonversi menggunakan label encoding untuk mengubah kategori menjadi angka numerik. Hal ini dilakukan agar data dapat diolah dengan baik menggunakan metode K-Means. Selain itu, kolom Tanggal Lahir juga dikonversi ke dalam format datetime, Hasil perhitungan ini disimpan dalam atribut baru bernama Umur. Berikut adalah tabel hasil label encoding untuk atribut Jenis pekerjaan, dan Golongan SIM dari data peserta uji SIM.

Tabel 2. Hasil konversi tanggal lahir

No	Tanggal lahir	Umur
1	4/8/1982	42
2	4/9/1982	42
3	7/9/1996	28
4	5/10/1979	45
5	4/20/1967	57
6	4/3/2000	24
7	4/11/2001	23
8	4/27/1987	37
9	8/6/1976	48
10	12/3/2004	20
...	...	...
1146	3/7/1980	44
1147	31/05/1992	32
1148	20/05/1992	32
1149	20/05/1992	32
1150	26/05/1966	58
1151	26/05/1966	58
1152	18/01/2001	24
1153	14/12/1997	27
1154	25/01/1992	33
1155	16/08/2000	24

Pada Tabel 2, kolom Tanggal Lahir diubah ke dalam format tipe datetime agar bisa diproses dengan benar. Setelah itu, umur dihitung dengan cara mengambil tahun saat ini lalu dikurangi dengan tahun lahir. Hasil perhitungan ini disimpan dalam atribut baru bernama Umur, yang berisi angka umur dari setiap orang berdasarkan tahun lahir mereka

Tabel 3. Hasil label encoding atribut pekerjaan

Pekerjaan	Jenis pekerjaan
APOTEKER	0
ASN	1
BELUM BEKERJA	2
DOKTER	3
DOSEN	4
GURU	5
HONORER	6
IBU RUMAH TANGGA	7
KARYAWAN SWASTA	8
NELAYAN	9
PEG. SWASTA	10
PEG.BUMN	11
PELAJAR/MHS	12
PENDETA	13
PENGEMUDI	14
PENSIUNAN ASN	15
PETANI	16
POLRI	17
SWASTA	18
TEKNISI	19
TNI AD	20
TNI AL	21
TNI AU	22
WIRASWASTA	23

Pada tabel 3, nama atribut Pekerjaan diubah menjadi Jenis Pekerjaan. Setelah itu, setiap jenis pekerjaan dikonversi ke dalam bentuk angka menggunakan Label Encoding. Karena terdapat 24 jenis pekerjaan, maka masing-masing jenis diberikan nilai numerik dari 0 hingga 23. Proses ini dilakukan dengan LabelEncoder, yang mengubah data teks menjadi angka agar lebih mudah digunakan dalam analisis dan pemodelan machine learning.

Tabel 4. Hasil label encoding atribut Golongan SIM

Gol. SIM	Golongan SIM
A	0
A Umum	1
BI Umum	2
BII Umum	3
C	4

Pada Tabel 4, atribut Gol. SIM diubah menjadi Golongan SIM untuk memperjelas informasi. Selanjutnya, dilakukan Label Encoding untuk mengonversi 5 jenis Golongan SIM ke dalam bentuk angka numerik. Setiap kategori Golongan SIM diberikan nilai dari 0 hingga 4 sesuai dengan jumlah kategori yang ada. Proses ini bertujuan untuk mempermudah analisis data

Proses *clustering* telah melalui beberapa tahap pre-processing, yaitu data selection, yang mencakup pemilihan data yang akan digunakan, serta transformasi, di mana Tanggal Lahir dikonversi ke dalam format datetime untuk menghitung Umur. Selain itu, atribut Pekerjaan diubah menjadi Jenis Pekerjaan dan dikonversi menjadi angka menggunakan Label Encoding dengan rentang 0-23, sementara atribut Gol. SIM diubah menjadi Golongan SIM dan dikonversi dengan Label Encoding dalam rentang 0-4. Setelah tahap pre-processing selesai, data siap diterapkan pada algoritma K-Means.

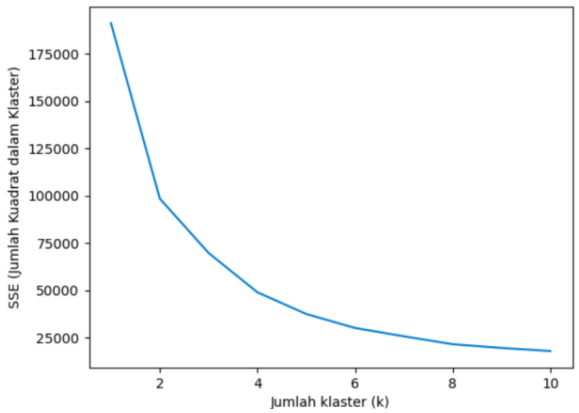
Tabel 5. Data yang telah melewati tahap pre-processing

No	Umur	Jenis Pekerjaan	Golongan SIM
1	43	23	4
2	43	23	0
3	29	6	4
4	46	14	2
5	58	1	0
6	25	11	4
7	24	11	4
8	38	6	4

No	Umur	Jenis Pekerjaan	Golongan SIM
9	49	23	4
10	21	12	4
...	...	...	...
1146	45	23	4
1147	33	23	4
1148	33	23	4
1149	33	23	0
1150	59	8	4
1151	59	8	0
1152	24	12	4
1153	28	12	0
1154	33	8	4
1155	24	12	4

4.2. Penentuan Jumlah *Cluster*

Jumlah *cluster* optimal ditentukan menggunakan Elbow Method, yang didasarkan pada perhitungan SSE. Grafik berikut menunjukkan hubungan antara jumlah K dan nilai SSE untuk K = 1 hingga K = 10. Titik di mana grafik mulai melandai, menyerupai siku (Elbow), menjadi acuan dalam menentukan jumlah *cluster* optimal. Berikut adalah grafik metode Elbow:



Gambar 2. Grafik Elbow method

Grafik di atas menunjukkan metode Elbow untuk menentukan jumlah *cluster* optimal. Sumbu X menunjukkan jumlah *cluster* (K), sedangkan sumbu Y menunjukkan nilai SSE. Terlihat bahwa nilai SSE menurun tajam dari K=1 hingga K=4, kemudian setelah K=4, penurunannya mulai melambat. Titik ini disebut Elbow atau siku, di mana setelahnya tambahan *cluster* tidak lagi memberikan pengurangan SSE yang signifikan. Oleh karena itu, K=4 dipilih sebagai jumlah *cluster* optimal karena memberikan keseimbangan antara akurasi pengelompokan dan kompleksitas model.

4.3. Penerapan Algoritma K-Means (data mining)

Setelah melakukan tahapan pre-processing dan penentuan jumlah *cluster* optimal, tahapan selanjutnya adalah penerapan algoritma K-Means *clustering* untuk mengelompokkan data peserta uji SIM berdasarkan Jenis pekerjaan, Golongan SIM dan Umur. Berikut adalah tahapan penerapan algoritma K-Means *clustering :*

4.3.1. Menentukan Jumlah *Cluster*

Berdasarkan hasil dari Gambar 1, grafik Elbow method di peroleh jumlah *cluster* optimal adalah K=4

a. Melakukan *clustering* dengan K-Means

Berdasarkan metode Elbow yang terdapat pada Gambar 2, menghasilkan nilai K=4, dengan demikian, proses klasterisasi menggunakan algoritma K-Means dilakukan, dan hasilnya ditampilkan sebagai berikut :

Tabel 6. Hasil klasterisasi algoritma K-Means

No	Jenis Pekerjaan	Golongan SIM	Umur	*Cluster*
1	23	4	43	3
2	23	0	43	3
3	6	4	29	0
4	14	2	46	3
5	1	0	58	2
6	11	4	25	0
7	11	4	24	0
8	6	4	38	2
9	23	4	49	3
10	12	4	21	0
...	...	...	...	...
1146	23	4	45	3
1147	23	4	33	1
1148	23	4	33	1
1149	23	0	33	1
1150	8	4	59	2
1151	8	0	59	2
1152	12	4	24	0
1153	12	0	28	0
1154	8	4	33	0
1155	12	4	24	0

b. Evaluasi menggunakan SSE

Tahap ini akan menampilkan nilai error dari penentuan jumlah *cluster* optimal yang dilakukan sebanyak 10 kali percobaan. Berikut adalah tabel nilai error

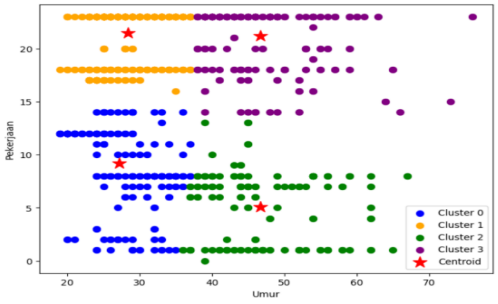
Tabel 7. Nilai SSE dan selisih error K-Means

K	SSE	Selisih dari K sebelumnya
1	191125.2	-
2	98488.1	92637.1
3	69725.72	28762.38
4	48960.67	20765.05
5	37501.83	11458.84
6	30112.38	7389.45
7	25721.08	4391.3
8	21534.32	4186.76
9	19554.8	1979.52
10	17938.76	1616.04

Dari tabel di atas, terlihat bahwa nilai SSE menurun tajam dari K=1 ke K=2, yaitu sebesar 92637.1, menunjukkan pemisahan awal sangat berdampak. Penurunan terus berlanjut namun semakin kecil, dari K=2 ke K=3 sebesar 28762.38, dan dari K=3 ke K=4 sebesar 20765.05. Setelah K=4, penurunan SSE tidak lagi signifikan. Karena itu, K=4 dipilih sebagai jumlah *cluster* optimal sesuai metode Elbow.

c. Visualisasi hasil *clustering*

Tahap ini menampilkan scatter plot hasil klasterisasi menggunakan k=4. Scatter plot ini menampilkan hasil pengelompokan berdasarkan atribut Umur dan Jenis pekerjaan. Berikut adalah visualisasi hasil *clustering* menggunakan metode K-Means.



Gambar 3. Scatter plot hasil *clustering*

Hasil scatter plot pada Gambar 3 menunjukkan penyebaran data berdasarkan dua atribut, yaitu Umur dan Jenis Pekerjaan. Grafik ini terdiri dari 4 *cluster*, yang masing-masing ditampilkan dengan warna berbeda. *Cluster*_0 (biru) berisi lebih banyak individu dengan usia muda dan jumlah pekerjaan yang lebih sedikit, *cluster*_1 (orange) didominasi oleh individu dengan pekerjaan lebih tinggi dan usia menengah, *cluster*_2 (hijau tua) mencakup individu dengan pekerjaan menengah ke bawah dengan variasi usia yang cukup luas. Sementara itu, *cluster*_3 (ungu) terdiri dari individu dengan pekerjaan lebih tinggi dan rentang usia yang beragam. Selain itu, centroid dari setiap *cluster* ditandai dengan warna merah berbentuk bintang, yang menunjukkan titik tengah dari semua *cluster*.

d. Menampilkan data dalam setiap *cluster*

Pada tahap ini akan menampilakan tabel yang berisi jumlah data dalam ke-4 *cluster:*

Tabel 8. Jumlah hasil klasterisasi

Cluster	**Jumlah data**
*Cluster*_0	249
*Cluster*_1	507
*Cluster*_2	150
*Cluster*_3	249

Pada tabel 8 di atas, ditunjukkan hasil klasterisasi data peserta uji SIM dengan jumlah K=4 dari total 1.155 data. Hasil klasterisasi tersebut terdiri dari *cluster*_0 dengan 249 anggota, *cluster*_1 dengan 507 anggota, *cluster*_2 dengan 150 anggota, dan *cluster*_3 dengan 249 anggota.

4.4. Analisis dan Interprensi

Setelah proses klasterisasi menggunakan algoritma K-Means, data hasil pengelompokan disimpan ke dalam file Excel untuk dianalisis lebih lanjut. Dalam proses analisis ini, digunakan fungsi MODE di Excel untuk menemukan nilai yang paling sering muncul dalam kumpulan data. Dengan menggunakan fungsi ini, kita dapat mencari dan menampilkan nilai yang paling dominan dalam suatu rentang data. Fungsi MODE sangat berguna untuk mengidentifikasi pola yang sering muncul dalam dataset yang akan di analisisis

Analisis ini bertujuan untuk memahami karakteristik masing-masing *cluster* dengan menghitung nilai rata-rata dari atribut yang digunakan, yaitu Jenis pekerjaan sebagai sumbu y dan Umur sebagai sumbu x. Berikut adalah hasil analisis data dari hasil *cluster* peserta uji SIM.

Tabel 9. Hasil analisis

Cluster	**Rata-rata Umur**	**Rata-rata Jeni pekerjaan**
0	20	12
1	25	23
2	44	1
3	39	23

Berdasarkan hasil analisis pada Tabel 9, terdapat pola hubungan antara Umur dan Jenis pekerjaan di setiap *cluster*. *Cluster*_0 didominasi oleh individu dengan rata-rata usia 20 tahun, yang sebagian besar berstatus sebagai pelajar atau mahasiswa. Sementara itu, *cluster*_1 memiliki rata-rata usia 25 tahun, dengan mayoritas bekerja sebagai wiraswasta. Pada *cluster*_2, individu yang tergabung memiliki rata-rata usia 44 tahun dan sebagian besar bekerja sebagai ASN (Aparatur Sipil Negara). Sedangkan *cluster* 3 memiliki rata-rata usia 39 tahun dengan mayoritas pekerjaan yang sama seperti *cluster*_1, yaitu wiraswasta. Dari pola ini, dapat disimpulkan bahwa individu dengan usia lebih muda cenderung masih berstatus sebagai pelajar, sedangkan pada usia yang lebih dewasa, mereka mulai memasuki dunia kerja dengan profesi yang lebih stabil, seperti ASN dan wiraswasta.

5. KESIMPULAN DAN SARAN

Klasterisasi peserta uji SIM di Kabupaten Manokwari dengan algoritma K-Means menghasilkan 4 *cluster*, berdasarkan metode Elbow dengan nilai SSE sebesar 48960,67. *Cluster*_0 didominasi oleh peserta berusia sekitar 20 tahun, mayoritas pelajar atau mahasiswa. *Cluster*_1 terdiri dari peserta berusia sekitar 25 tahun yang mayoritas wiraswasta. *Cluster*_2 mencakup peserta berusia rata-rata 44 tahun, sebagian besar bekerja sebagai ASN, dan *Cluster*_3 memiliki rata-rata usia 39 tahun dengan mayoritas juga wiraswasta. Pola ini menunjukkan adanya hubungan antara Umur dan Jenis pekerjaan, yang dapat menjadi dasar bagi pihak kepolisian dalam menyusun program edukasi dan kebijakan lalu lintas yang lebih tepat sasaran. Meskipun *Cluster*_2 merupakan kelompok dengan jumlah peserta paling sedikit (150 peserta), hal ini kemungkinan disebabkan karena banyak dari mereka sudah memiliki SIM sebelumnya, sehingga tidak termasuk dalam data pengurusan saat ini. Namun demikian, sosialisasi tetap penting dilakukan, terutama di lingkungan instansi pemerintahan di Kabupaten Manokwari, untuk memastikan informasi terkait

perpanjangan SIM dan kesadaran berlalu lintas tetap terjaga. Untuk meningkatkan keakuratan hasil klasterisasi, disarankan agar penelitian selanjutnya menggunakan dataset yang lebih beragam serta menerapkan metode evaluasi seperti Silhouette Score atau Davies-Bouldin Index guna memastikan jumlah *cluster* yang diperoleh sudah optimal. Dengan pengembangan lebih lanjut, diharapkan hasil penelitian ini dapat memberikan manfaat bagi instansi terkait dalam merencanakan kebijakan lalu lintas yang lebih efektif.

DAFTAR PUSTAKA

[1] E. N. Ibie and O. Orinasanti, "Pelayanan Psikotes Sebagai Persyaratan Pembuatan Surat Izin Mengemudi (Sim) Di Wilayah Hukum Polisi Resort Kabupaten Gunung Mas," *J. Adm. Publik dan Pembang.*, vol. 5, no. 2, pp. 135–144, 2024.

[2] A. Maryati, T. Arbain, and M. R. Syafari, "Kualitas Pelayanan Pembuatan Surat Izin Mengemudi (Sim C) Di Kantor Satuan Lalu Lintas Kepolisian Resort Hulu Sungai Utara," *J. Adm. Publik dan Pembang.*, vol. 5, no. 1, p. 1, 2023.

[3] I. Luqman Hakim, A. Saputra, M. Florensia Sahetapi, Y. Ratte, and A. Ansi Ahoren, "Penerapan Metode K-Nearest Neighbors (Knn) Dalam Klasifikasi Golongan Sim Di Daerah Kabupaten Manokwari," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3693–3698, 2024.

[4] R. Dimas Sulistiyo and M. R. Shihab, "Transformasi Digital dalam Pelayanan Surat Izin Mengemudi (SIM): Studi Kasus Korlantas Polri," *Technomedia J.*, vol. 8, no. 2SP, pp. 189–204, 2023.

[5] E. Ebing and A. F. M. Karmiza, "Implementasi Sosialisasi Aplikasi Sinar Dalam Pembuatan Surat Izin Mengemudi (Sim) Di Polres Banyuasin," *Gov. Insight*, vol. 1, no. 2, 2024.

[6] A. Abdussalam, F. A. Nisa, A. Susanto, and ..., "Klasterisasi Perkara Pelanggaran Lalu Lintas Menggunakan Algoritma K-Means Dan Davies-Bouldin Index," *Sci. …*, vol. 5, no. Sens 5, pp. 366–375, 2020.

[7] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model *clustering* dengan Menggunakan Algoritma K-Means," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 8, no. 1, p. 56, 2023.

[8] A. Winarta and W. J. Kurniawan, "Optimasi *Cluster* K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python," *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 5, no. 1, pp. 113–119, 2021.

[9] T. Pelayanan, P. Di, and U. Prima, "Implementasi data mining *clustering* dalam mengukur kepuasan terhadap pelayanan perpustakaan di universitas prima indonesia," vol. 7, pp. 318–324, 2024.

[10] T. Hartini *et al.*, "IMPLEMENTASI ALGORITMA K-MEANS *clustering*," vol. 9, no. 1, pp. 76–83, 2025.

[11] R. R. Asyrofi and R. Asyrofi, "Implementasi Aplikasi Jupyter Notebook Sebagai Analisis Kreteria Plagiasi Dengan Teknik Simantik," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 8, no. 2, pp. 627–637, 2023.