

Analisis Clustering Kasus Bunuh Diri di Jawa Tengah dengan Menerapkan Algoritma *K-Means*

Widya Kurniawan, Faisal Reza Pradhana, & Khusna Amalia Zen*

Universitas Darussalam Gontor, Jl. Raya Siman, Ponorogo, Jawa Timur, Indonesia.

Email: rektorat@unida.gontor.ac.id, khusnazen172@gmail.com

Corresponding author: khusnazen172@gmail.com

Abstrak

Bunuh diri merupakan suatu tindakan yang bertujuan untuk mengakhiri hidup dengan sengaja. Di Indonesia fenomena ini masih banyak ditemukan dan dipengaruhi oleh beberapa faktor, diantaranya faktor sosial, faktor ekonomi, faktor psikologi, dan faktor lingkungan. Penelitian ini bertujuan untuk mengidentifikasi pola kasus bunuh diri dengan menggunakan teknik clustering, dengan data yang diambil dari Kepolisian Resor Semarang dan Boyolali. Tiga variabel utama yang akan dianalisis, yaitu: rentang usia, metode bunuh diri, dan Lokasi kejadian. Pendekatan CRISP-DM diterapkan untuk pengolahan data. Clustering dilakukan menggunakan algoritma *K-Means* untuk mengelompokkan data-data yang relevan berdasarkan variabel tersebut. Hasil Silhouette score sebesar 84% menunjukkan kualitas pemisahan antar cluster cukup baik. Visualisasi dengan metode Principal Component Analysis (PCA) digunakan untuk membantu memetakan cluster agar lebih komprehensif, yang mana kelompok paling rawan melakukan tindakan bunuh diri adalah pelaku dengan rentang usia produktif, dengan menggunakan metode gantung diri, dan melakukannya di rumah. Studi ini diharapkan dapat memberikan wawasan lebih lanjut terkait fenomena bunuh diri di Indonesia.

Kata Kunci: Klasterisasi, *K-Means*, Bunuh Diri, CRISP-DM

Abstract

Suicide is a deliberate act intended to end one's life. In Indonesia, this phenomenon remains prevalent and is influenced by various factors, such as psychological conditions, economic pressures, social issues, and environmental factors. This study aims to identify patterns of suicide cases using Clustering techniques, with data sourced from the Semarang and Boyolali Police Departments. The three main variables analyzed are age range, suicide method, and location of the incident. The CRISP-DM approach is applied for data processing, and the *K-Means* algorithm is used to group relevant data based on these variables. A Silhouette score of 84% indicates a good separation between clusters. Visualization with Principal Component Analysis (PCA) is used to map the clusters more comprehensively. The most vulnerable group to commit suicide is individuals in the productive age range, who tend to use hanging as the method and do so in private homes. This study is expected further insights into the suicide phenomenon in Indonesia.

Keywords: Clustering, *K-Means*, Suicide, CRISP-DM

1. PENDAHULUAN

Undang-undang Republik Indonesia No.36 Tahun 2009 berisi tentang Kesehatan, pemerintah mendefinisikan sehat sebagai kondisi yang mencakup kesejahteraan fisik, mental, spiritual, dan sosial yang memungkinkan seseorang untuk hidup produktif, baik secara sosial maupun ekonomi. Hal ini dijelaskan oleh *World Health Organization (WHO)* dari tahun 1980 sampai

tahun 1998 yang menyatakan bahwa: “*Health is a state of physical, mental, and social wellbeing and not merely the absence of disease of infirmity*”. Dari beberapa aspek kehidupan tersebut, salah satu isu sosial yang dapat menarik perhatian masyarakat Indonesia yaitu kesehatan mental yang mana memiliki definisi kondisi yang mencakup emosional, psikologis, dan sosial individu. Kesehatan mental dikatakan baik apabila kondisi

seseorang berada dalam keadaan sejahtera yang mampu memahami makna serta tujuan hidupnya. Sebaliknya, seseorang yang memiliki cara berpikir dan suasana hati yang sedang tidak baik menunjukkan bahwa kesehatan mental orang tersebut sedang terganggu atau tidak baik. Kesehatan mental yang buruk sering kali menjadi faktor risiko utama yang dapat menyebabkan seseorang rentan terhadap pemikiran atau perilaku bunuh diri [1].

Suatu tindakan yang memiliki tujuan untuk mengakhiri hidup dengan sengaja disebut tindakan bunuh diri [2]. Di Negara Indonesia telah terjadi peningkatan kematian yang disebabkan oleh bunuh diri. Seperti data yang ditunjukkan oleh Asosiasi Pencegahan Bunuh Diri di Indonesia bahwasannya jumlah kasus bunuh diri resmi pada tahun 2020 sebanyak 670 kasus. Sedangkan, pada data yang telah dipaparkan oleh databoks terdapat 971 kasus bunuh diri di Indonesia yang terjadi pada Januari hingga Oktober tahun 2023. Data ini membuktikan bahwasannya kasus bunuh diri di Indonesia mengalami peningkatan yaitu sekitar 45%. Data di Indonesia menunjukkan bahwa bunuh diri merupakan salah satu penyebab kematian tertinggi di komunitas usia produktif, yaitu sekitar usia 15-29 tahun [3]. Para korban pasti memiliki berbagai macam alasan untuk melakukan hal. Namun, faktanya mereka melakukan hal ini dikarenakan kematian adalah jalan keluar satu-satunya yang mereka pikirkan [4]. Dalam kasus ini, bukan hanya kelompok usia produktif saja yang melakukannya, korban para lansia pun dapat kita temukan, bahkan dengan anak usia dini juga sudah ada yang melakukan hal tersebut. Bunuh diri tidak hanya menyebabkan penderitaan bagi individu yang terlibat, namun hal ini juga dapat berdampak luas pada keluarga, teman-teman, dan masyarakat di sekitarnya. Teknologi yang kita ketahui sudah berkembang pesat di zaman ini, memungkinkan untuk mengelompokkan kasus ini berdasarkan beberapa atribut. Salah satu hal yang muncul karena berkembangnya teknologi yang terkait dengan Analisa data, yaitu data mining.

Data mining merupakan proses menganalisis data dalam jumlah yang cukup besar guna menemukan pola, tren, serta informasi berguna yang tidak langsung terlihat.

Tujuan dari data mining untuk mendapatkan wawasan yang dapat mendukung guna mengambil keputusan dengan lebih baik di beberapa bidang. Teknik data mining memiliki 5 metode, yaitu asosiasi, klasifikasi, klusterisasi, estimasi, dan *forecasting*. Dalam penelitian ini penulis mengidentifikasi dengan memanfaatkan metode *Clustering* pada objek penelitian, yaitu kasus bunuh diri. *Clustering* adalah proses pembagi data di suatu himpunan yang bertujuan untuk mengelompokkan data-data yang ada berdasarkan kemiripannya. *Clustering* termasuk dalam kategori *Unsupervised Learning* yang berguna untuk mempartisi data tanpa label ke dalam suatu kelompok [5].

Terdapat berbagai macam algoritma yang bisa digunakan dalam metode *clustering*, salah diantaranya yaitu algoritma *K-Means*. Metode ini memiliki proses yang dijalankan secara berulang-ulang. Pada penelitian kali ini, algoritma *K-Means* akan dilakukan untuk pengelompokkan data kasus bunuh diri di beberapa wilayah di Jawa Tengah dengan sampel data sebanyak 103 kasus dan akan dikelompokkan menjadi 3 cluster berdasarkan usia, metode bunuh diri, dan Lokasi kejadian terjadinya kasus bunuh diri di beberapa wilayah Jawa Tengah.

Dalam hal melakukan klusterisasi, algoritma *K-Means* banyak digunakan oleh peneliti sebelumnya. Salah satunya seperti penelitian yang dilakukan oleh Putra May Chandra Abirianto yang berjudul “Penerapan Metode *K-Means Clustering* untuk Pengelompokkan Pasien Penyakit Liver”. Penelitian ini membahas mengenai Pengelompokkan pasien penyakit liver yang telah diuji menggunakan metode *K-Means Clustering*, dengan hasil akhir menunjukkan tingkat akurasi sebesar 86% antara hasil perhitungan sistem dan perhitungan manual. Temuan ini mengindikasikan bahwa metode *K-Means Clustering* dapat diterapkan pada sistem dan berfungsi secara optimal [6].

Data yang didapat oleh peneliti bersumber dari data yang dimiliki oleh Polres (Kepolisian Resor) Kabupaten Semarang dan Boyolali. Data penelitian diambil dari Polres kedua kabupaten tersebut karena sumber ini merupakan institusi resmi dan terpercaya yang menyediakan data yang akurat, konsisten, dan terstruktur mengenai kasus bunuh diri.

Berdasarkan permasalahan yang telah dipaparkan, penelitian ini bertujuan untuk mengelompokkan kasus-kasus bunuh diri di beberapa wilayah di Jawa Tengah berdasarkan usia, metode bunuh diri, dan Lokasi kejadian dengan menggunakan algoritma *K-Means Clustering*.

2. DASAR TEORI

2.1 Bunuh Diri

Bunuh diri merupakan tindakan agresif merusak diri sendiri dan dapat mengakhiri kehidupan seseorang. Bunuh diri dapat ditafsirkan dengan merenggut nyawanya sendiri dengan menganggap bahwa dengan bunuh diri maka semua masalah terselesaikan. Hal ini dapat terjadi akibat perilaku spontan yang timbul akibat seseorang tidak dapat mengontrol emosi, berputus asa, dan menanggapi bahwa hidupnya tidak dapat dilanjutkan dan tidak berharga lagi [7]. Perilaku bunuh diri pada seseorang disebabkan karena stress tinggi dan kegagalan mekanisme koping dalam mengatasi masalah [8]. Hari demi hari semakin tinggi angka kematian yang disebabkan oleh bunuh diri.

Seseorang yang melakukan bunuh diri cenderung sadar dalam melakukan tindakan ini yang didasarkan pada keinginan diri sendiri tanpa paksaan dari manapun. Banyak terjadi pada remaja karena masa ini termasuk masa peralihan pencarian jati diri yang mana dapat mengalami banyak perubahan baik biologis, psikologis, ataupun sosial. Berdasarkan hasil penelitian oleh Kezia Yunitasari Kusuma [9] bahwasannya penyebab bunuh diri yang terjadi di Toraja yaitu karena faktor disharmonis keluarga, faktor kepribadian dan isolasi sosial.

2.2 Clustering

Salah satu teknik dalam Data Mining adalah *Clustering* yang digunakan untuk menganalisis data dengan tujuan mengelompokkan objek-objek dalam suatu dataset ke dalam grup atau klaster berdasarkan kesamaan karakteristik atau fitur tertentu tanpa label yang telah ditentukan sebelumnya [10]. Objek-objek dalam satu klaster memiliki tingkat kemiripan yang cukup tinggi, sementara objek-objek yang berada pada klaster berbeda memiliki perbedaan yang signifikan. Hal ini berbeda dengan klasifikasi yang memerlukan data yang telah diberi label untuk melatih model.

Tujuan *clustering* yaitu untuk mengidentifikasi pola atau stuktur dalam data [11]. Dengan mengelompokkan data menjadi klaster yang berbeda, maka karakteristik dan hubungan yang ada dalam data tersebut dapat lebih mudah dipahami. Metode *clustering* memiliki beberapa algoritma yang bisa digunakan untuk mengelompokkan data, yaitu *K-Means Clustering*, DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), GMM (*Gaussian Mixture Models*), dan lain-lain.

Penggunaan *clustering* dapat diterapkan diberbagai hal, seperti segmentasi pasar dalam bisnis, pengelompokkan dokumen dalam teks mining, teks anomali dalam keamanan, pengelompokkan pelanggan dalam CRM, dan analisis gen dalam cabang ilmu biologi.

2.3 K-Means

Algoritma *K-Means* adalah salah satu algoritma dalam analisis *clustering* yang digunakan untuk mengelompokkan data ke dalam beberapa kategori yang berbeda secara otomatis. *K-Means* adalah salah satu algoritma *clustering* yang paling umum digunakan dalam analisis data. Algoritma ini tergolong *unsupervised learning*, dimana data tidak memiliki label atau kategori yang telah ditentukan sebelumnya [12].

K-Means adalah algoritma klasterisasi yang bertujuan untuk membagi N data ke dalam K klaster, Dimana setiap data termasuk ke dalam klaster dengan mean (rata-rata) terdekat. Algoritma ini mencari posisi optimal dari K centroid (titik tengah klaster) sehingga jumlah jarak kuadrat antar tiap data dan *centroid* klasternya bisa minimal [13]. Langkah-langkah mengolah data dengan Algoritma *K-Means Clustering* adalah sebagai berikut:

- Menentukan jumlah cluster K tergantung pada kebutuhan sistem
- Menentukan nilai titik tengah (centroid) secara acak sejumlah cluster
- Alokasikan masing-masing data ke pusat cluster terdekat, dengan rumus:

$$Be = \sqrt{((O_i - T_i)^2 + (O_j - T_j)^2)} \quad (1)$$

Dimana:

Be = Jarak data ke pusat cluster;

O = Data record;

T = Data centroid;

d. Hitung ulang pusat cluster (centroid) baru, dengan rumus:

$$KI = (O1+O2+O3+...+On)/(\Sigma O) \quad (2)$$

Dimana:

KI = Centroid baru;

O1 = Nilai data record ke-1;

O_n = Nilai data record ke-n;

ΣO = Jumlah data record;

e. Tempatkan kembali data ke pusat cluster yang baru terbentuk. Jika pusat cluster masih berubah, maka proses diulangi dari langkah c, namun jika pusat cluster tidak berubah, maka pencarian dapat dihentikan.

Algoritma ini memiliki sejumlah keunggulan, yaitu lebih sederhana untuk dipahami dan diterapkan, efisien dalam menangani data dalam jumlah besar, dapat memproses data dengan tipe data numerik, serta hasilnya stabil dan konvergen. Namun, disisi lain algoritma ini juga memiliki kelemahan, diantaranya jumlah kluster (K) harus ditentukan di awal, sensitif terhadap inialisasi, memerlukan data numerik, serta tidak dapat menangani data dengan bentuk yang kompleks. Algoritma *K-Means* juga cocok untuk berbagai aplikasi, seperti analisis data, visualisasi data, dan tugas lain yang membutuhkan pengelompokan data.

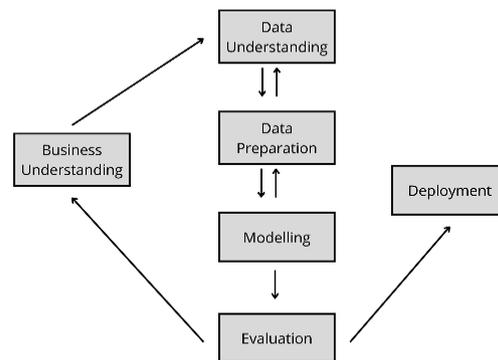
2.4 Data Mining

Data Mining adalah metode yang digunakan untuk mengolah data guna menemukan pola tersembunyi dalam kumpulan data yang diolah. Dengan data mining informasi baru atau pengetahuan baru akan dihasilkan dari data lama, dan informasi ini dapat menjadi dasar dalam pengambilan Keputusan di masa depan[6].

Dengan menggunakan metode Data mining pengguna dapat mengolah data dalam skala besar. Sehingga perannya sangat penting dalam berbagai bidang kehidupan, seperti industry, keuangan cuaca, ilmu, dan teknologi. Data mining juga sering disebut dengan istilah *pattern recognition* (pengenalan pola). Untuk mendapatkan pengetahuan yang berguna dengan mencari beragam informasi dalam dataset, kita dapat memanfaatkan bantuan data mining yang berguna untuk memenuhi kebutuhan pengguna dalam waktu yang lebih cepat dan lebih akurat [14].

3. METODOLOGI

Dalam pengimplementasian algoritma *K-Means Clustering*, pada penelitian ini metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) akan digunakan. Definisi dari CRISP-DM yaitu kerangka kerja standar yang digunakan dalam proses data mining untuk memandu alur kerja proyek analisis data. Menurut Larose, data mining memiliki enam fase yang mana ditunjukkan pada Gambar 1 berikut.



Gambar 1. Metode CRISP-DM

3.1 Business Understanding

Tahap ini fokus untuk mengidentifikasi masalah yang akan dipecahkan, menetapkan tujuan yang jelas, dan merumuskan rencana proyek.

3.2 Data Understanding

Pada tahapan ini akan menyertakan proses untuk mengumpulkan data dan pemahaman tentang data yang akan dipakai dalam penelitian kali ini. Pada tahapan ini, penulis mengumpulkan data kasus bunuh diri yang bersumber dari Kepolisian Resor di Semarang dan Boyolali. Pada penelitian ini ada 3 variabel yang dikelompokkan yaitu variabel usia, metode bunuh diri, dan Lokasi kejadian.

Selain itu, penulis juga melihat beberapa data yang dipaparkan oleh katadoks.com yang mana menunjukkan bahwasannya terdapat kenaikan jumlah individu yang melakukan bunuh diri tahun 2022 sampai tahun 2023, ditunjukkan oleh jumlah korban pada tahun 2022 sebanyak 670 kasus, dan terdapat sekitar 971 kasus pada tahun 2023 dengan hanya rentang waktu 10 bulan. Hal ini membuktikan bahwa kenaikan kasus bunuh diri di Indonesia dapat mencapai 45%.

3.3 Data Preparation

Pada tahapan ini melibatkan proses transformasi serta pembersihan data agar dapat digunakan dalam proses modelling. Tahapan ini sangat penting untuk sebuah penelitian guna memastikan data yang digunakan siap untuk dianalisis dengan algoritma *K-Means*. Berikut hal-hal yang perlu diperhatikan dalam data preparation.

- Pengumpulan data, yang dilakukan peneliti guna mendapatkan data untuk dianalisis berikut. Pada kesempatan ini, sumber data diambil dari Kepolisian Resor Kabupaten Semarang dan Boyolali. Peneliti dalam mengumpulkan datanya menggunakan metode wawancara secara langsung dengan petugas Kepolisian Resor di kedua kabupaten tersebut.
- Pembersihan data (Data Cleaning), yang terdiri dari pemeriksaan kualitas data (mengecek integritas dan keakuratan data untuk menghindari analisis yang salah), Penanganan Missing Value (Mengisi atau menghapus data yang hilang untuk mencegah bias dalam analisis). Pada tahap ini peneliti mengecek keberadaan nilai yang hilang (missing value) di dalam dataset. Namun berdasarkan hasil pengecekan, tidak ditemukan adanya kekosongan data dalam dataset.
- Transformasi data, hal ini dilakukan untuk mengubah data mentah menjadi format untuk analisis. Transformasi data terdiri dari Normalisasi dan Standarisasi (menyelaraskan skala data untuk memastikan variabel-variabel yang berbeda memiliki pengaruh yang setara dalam analisis), Encoding Kategorikal data (mengubah data kategorikal menjadi bentuk numerik yang bisa diproses oleh algoritma machine learning).

3.4 Modelling

Tahapan ini akan melibatkan pemilihan dan penerapan teknik modelling yang tepat untuk data yang disiapkan. Pada tahapan inilah algoritma *K-Means* akan diterapkan untuk mengelompokkan data berdasarkan rentang usia, metode bunuh diri, dan Lokasi kejadian. Pada tahap ini juga akan ditentukan jumlah kluster optimal dan pengujian model. Dalam pemilihan model, algoritma *K-Means* dipilih karena keandalannya dalam mengelompokkan

data kedalam kluster yang homogen berdasarkan jarak *Euclidean* antar data.

3.5 Evaluation

Tahap ini melibatkan pemilihan dan penerapan teknik modelling yang tepat untuk data yang telah disiapkan. Pada tahap inilah algoritma *K-Means* akan diterapkan untuk mengelompokkan data berdasarkan rentang usia, metode bunuh diri, dan Lokasi kejadian. Pada tahap ini juga akan ditentukan jumlah kluster optimal dan pengujian model. Dalam pemilihan model, algoritma *K-Means* dipilih karena keandalannya dalam mengelompokkan data kedalam kluster yang homogen berdasarkan jarak *Euclidean* antar data.

3.6 Deployment

Setelah tahap evaluasi selesai dilakukan untuk menilai secara detail hasil dari sebuah model akan dilakukan implementasi dari keseluruhan model yang telah dibangun yang mana disebut tahap *deployment*. Pada penelitian kali ini implementasi yang dilakukan akan diterapkan dan dibuat menjadi sebuah laporan yang berisi hasil yang sesuai dengan tahapan CRISP-DM.

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan dan Persiapan Data

Pada penelitian ini, data yang akan digunakan untuk melakukan pemodelan atau implementasi Algoritma *K-Means* adalah data kasus bunuh diri di daerah Semarang dan Boyolali selama kurang lebih 5 tahun dari tahun 2020 hingga tahun 2024. Jumlah data sebanyak 103 data. Kemudian data tersebut akan dibentuk menjadi 3 kelompok berdasarkan atribut usia, metode bunuh diri, dan lokasi kejadian. Selanjutnya Algoritma *K-Means* akan digunakan untuk melakukan pengolahan data tersebut. Berikut Tabel 1 yang berisi data kasus bunuh diri di Semarang dan Boyolali:

Tabel 1. Data Kasus Bunuh Diri di Daerah Kabupaten Semarang dan Boyolali

No	Nama	Usia	Rentang Usia	Metode Bunuh Diri	Lokasi Kejadian
1	K1	38	Dewasa	Gantung Diri	Rumah
2	K2	51	Lansia	Gantung Diri	Rumah
3	K3	19	Remaja	Gantung Diri	Rumah
4	K4	50	Lansia	Gantung Diri	Kebun

5	K5	47	Lansia	Gantung Diri	Rumah
6	K6	47	Lansia	Gantung Diri	Rumah
7	K7	35	Dewasa	Gantung Diri	Rumah
8	K8	67	Lansia	Gantung Diri	Rumah
9	K9	55	Lansia	Gantung Diri	Hotel
10	K10	37	Dewasa	Minum Racun	Rumah

Dari tabel diatas, atribut yang memiliki tipe data kategorikal akan diubah menjadi tipe data numerik. Berikut beberapa atribut serta pengkodeannya yang tipe datanya perlu diubah.

Tabel 2. Atribut Rentang Usia

Anak-Anak	1
Remaja	2
Dewasa	3
Lansia	4

Tabel 3. Atribut Metode Bunuh Diri

Gantung Diri	1
Minum Racun	2
Lompat Dari Gedung	3
Menceburkan Diri	4
Lain-Lain	5

Tabel 4. Atribut Lokasi Kejadian

Rumah	1
Hotel	2
Kebun	3
Jembatan	4
Lain-lain	5

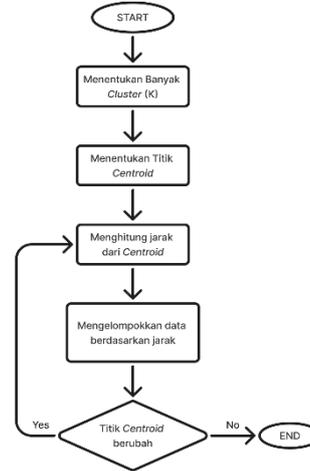
Dengan melakukan pengkodean dan pengkategorian ini, data atribut kasus bunuh diri siap untuk dikelompokkan dengan metode *Clustering* menggunakan Algoritma *K-mean*. Berikut sampel data yang siap diolah akan ditunjukkan pada tabel 5.

Tabel 5. Sampel Data Implementasi K-Means Clustering

Nama	Rentang Usia	Metode Bunuh Diri	Lokasi Kejadian
K1	3	1	1
K2	4	1	1
K3	2	1	1
K4	4	1	3
K5	4	1	1
K6	4	1	1
K7	3	1	1
K8	4	1	1
K9	4	1	2
K10	3	2	1

4.2 Data Understanding

Diagram alur di bawah ini menunjukkan tahapan perencanaan Algoritma *K-Means Clustering* dalam mengelompokkan data kasus bunuh diri di wilayah tertentu. Berikut Gambar 1 yang menampilkan secara keseluruhan tahapan Algoritma *K-Means Clustering*.



Gambar 2. Diagram Implementasi Algoritma *K-Means Clustering*

Dari diagram diatas berikut langkah-langkah atau urutan dari kinerja metode *K-Means* yang dapat dijelaskan.

1. Penentuan banyaknya cluster (K)

Dalam menentukan banyak cluster, penulis memanfaatkan metode *silhouette score* yang mana dalam data ini mendapatkan skor sebanyak 84% dengan 3 jumlah cluster.

2. Penentuan centroid awal

Pada implementasi Algoritma *K-Means*, dalam memilih titik centroid awal dapat dilakukan dengan cara acak pada data sampel. Proses ini tidak mempertimbangkan atribut atau karakteristik tertentu dalam penentuan centroid awal.

Tabel 6. Centroid Awal Cluster

Centroid	Rentang Usia	Metode Bunuh Diri	Lokasi Kejadian
C1	4	1	3
C2	3	2	1
C3	3	1	1

Kemudian akan diproses pengelompokkan data kedalam kelompok-kelompok dengan menggunakan titik centroid yang sebelumnya telah ditentukan sebagai titik awal dalam proses tersebut.

3. Penghitungan jarak dari centroid

Persamaan *Euclidean Distance* akan

digunakan dalam perhitungan ini untuk mendapatkan seberapa dekat atau jauh suatu data dengan titik pusat. Berikut rumus *Euclidean Distance*.

$$dist(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- dist = jarak
- i = data ke-i
- y = centroid
- x = data
- n = banyak data

Perhitungan jarak akan dilihat dari perbedaan nilai atribut antar data dengan pusat *cluster* pada setiap atribut yang relevan. Dari perhitunagn tersebut akan didapatkan nilai matrik jarak sebagai berikut:

Jarak data-1 ke pusat *cluster*

$$dist(1,1) = \sqrt{(3 - 4)^2 + (1 - 1)^2 + (1 - 3)^2} = 2,236$$

$$dist(1,2) = \sqrt{(3 - 3)^2 + (1 - 2)^2 + (1 - 1)^2} = 1$$

$$dist(1,3) = \sqrt{(3 - 3)^2 + (1 - 1)^2 + (1 - 1)^2} = 0$$

Jarak data-2 ke pusat *cluster*

$$dist(2,1) = \sqrt{(4 - 4)^2 + (1 - 1)^2 + (1 - 3)^2} = 2$$

$$dist(2,2) = \sqrt{(4 - 3)^2 + (1 - 2)^2 + (1 - 1)^2} = 1,414$$

$$dist(2,3) = \sqrt{(4 - 3)^2 + (1 - 1)^2 + (1 - 1)^2} = 1$$

Kemudian, perhitungan akan dilanjutkan pada data ke-3 dan seterusnya. Hal ini akan menunjukkan matriks jarak yang memperlihatkan jarak antar setiap data dengan pusat awal *cluster*.

4. Pengelompokkan data berdasarkan jarak

Langkah berikutnya yaitu membandingkan jarak tiap data ke pusat *cluster* untuk mencari jarak terdekat antara data ke titik pusat. Jarak ini yang nantinya akan

mengindikasikan setiap data akan tergabung dalam satu kelompok *cluster* dengan jarak terdekat. Berikut tabel yang menunjukkan kelompok data setelah dibandingkan dengan titik pusat awal *cluster*.

Tabel 7. Pengelompokkan Jarak Terkecil Pusat Cluster Iterasi ke-1

Nama	C1	C2	C3	Cluster
K1	2,236	1	0	3
K2	2	1,414	1	3
K3	2,828	1,414	1	3
K4	0	2,449	2,236	1
K5	2	1,414	1	3
K6	2	1,414	1	3
K7	2,236	1	0	3
K8	2	1,414	1	3
K9	1	1,732	1,414	1
K10	3	1	2	2

5. Penentuan Pusat Cluster Baru

Setelah tahap penentuan sebelumnya, langkah berikutnya adalah menentukan pusat *cluster* baru berdasarkan rata-rata nilai data dari iterasi 1. Berikut perhitungan rata-rata pusat awal *cluster* baru.

Tabel 8. Pusat Cluster Baru

Centroid	Rentang Usia	Metode Bunuh Diri	Lokasi Kejadian
C1	3,366	1,439	4,02
C2	3,143	3,143	1,286
C3	3,491	1	1,036

Kemudian untuk menentukan iterasi berikutnya dilakukan dengan langkah yang sama seperti tahap sebelumnya sampai tidak ada atau terjadi sedikit perubahan data yang terjadi dalam suatu pengelompokkan *cluster*. Pada penelitian ini, proses iterasi dilakukan hingga mencapai iterasi ke-2, dikarenakan di iteasi 2 tidak terjadi banyak perubahan dari iterasi sebelumnya. Berikut hasil dari iterasi ke-2.

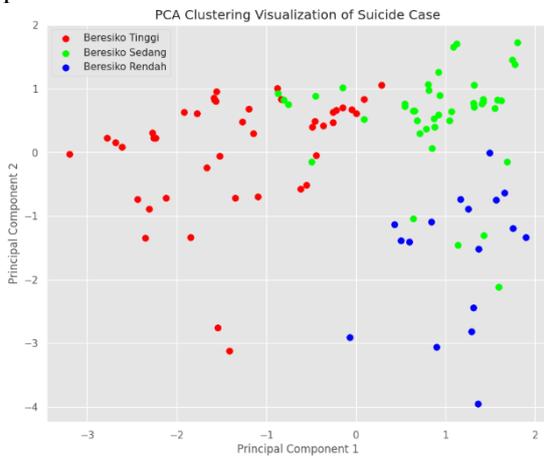
Tabel 9. Hasil Pengelompokkan Iterasi ke-2

Nama	C1	C2	C3	Cluster
K1	3,078	2,167	0,492	3
K2	3,121	2,326	0,510	3
K3	3,347	2,445	1,491	3
K4	1,282	2,875	2,029	1
K5	3,121	2,326	0,510	3
K6	3,121	2,326	0,510	3
K7	3,078	2,167	0,492	3
K8	3,121	2,326	0,510	3
K9	2,166	2,416	1,090	3
K10	3,423	0,350	2,060	2

Dari hasil pengelompokkan data berdasarkan iterasi 2 diatas menunjukkan

bahwa didapatkan sebanyak 61 data pada *cluster* 3, sedangkan *cluster* 2 memiliki anggota sebanyak 7 data, dan pada *cluster* 1 memiliki data sebanyak 35.

Setelah melakukan proses klasterisasi dengan metode *K-Means*, berikut akan ditampilkan visualisasi dengan menggunakan metode PCA (*Principal Component Analysis*), yang mana dengan metode ini multidimensi diubah menjadi dua dimensi, memungkinkan setiap kluster untuk divisualisasikan secara lebih sederhana dan jelas sehingga dapat memberikan gambaran lebih intuitif terkait distribusi data dalam setiap klaster dan perbedaan karakteristik di antara klaster.



Gambar 3. Visualisasi Clustering Kasus Bunuh Diri

Data pada gambar di atas terbagi menjadi tiga kelompok. Kelompok pertama ditunjukkan oleh titik-titik berwarna merah, kelompok kedua ditandai oleh titik-titik berwarna hijau, dan kelompok ketiga diwakili oleh titik-titik berwarna biru. Adapun hasil pengelompokan berdasarkan klasterisasi sebelumnya pada tiap *cluster* nya adalah sebagai berikut:

Tabel 10. Pengelompokan Pada Tiap Cluster

Atribut		C1	C2	C3
Rentang Usia	Lansia	18	3	36
	Dewasa	12	2	19
	Remaja	18	5	2
	Anak-anak	-	-	0
Metode Bunuh Diri	Gantung Diri	26	-	61
	Minum Racun	-	6	-
	Lompat Dari Gedung	5	-	-
	Menceburkan Diri	3	1	-
	Lain-lain	1	-	-
Lokasi Kejadian	Rumah	-	5	60
	Hotel	-	2	1
	Kebun	7	-	-

	Jembatan	1	-	-
	Lain-lain	27	-	-

Berdasarkan tabel pengelompokan pada setiap *cluster*, dapat disimpulkan bahwa pada Rentang Usia, kelompok C1 mayoritas adalah anak remaja dan lansia dengan berbagai metode bunuh diri dan terjadi di berbagai lokasi. Sedangkan pada kelompok C2 anak-anak remaja menempati jumlah korban terbanyak dengan metode yang sering digunakan yaitu meminum racun dan lokasi kejadian di tempat tinggal. Untuk kelompok C3 mayoritas korban adalah lansia dengan metode yang digunakan oleh keseluruhan korban dalam kelompok ini yaitu gantung diri, dan lokasi kejadian sebagian besar terjadi di rumah. Hal ini menunjukkan adanya perbedaan risiko di tiap *cluster*, dengan metode gantung diri dan kejadian di rumah menjadi pola umum di seluruh kelompok.

5. SIMPULAN

Dengan menerapkan algoritma *K-Means* clustering pada penelitian ini dapat menghasilkan pengelompokan data kasus bunuh diri menjadi tiga kategori risiko, yakni cluster berisiko tinggi, sedang, dan rendah. Setiap cluster memiliki jumlah data yang berbeda, yaitu 61 anggota pada cluster berisiko tinggi, 7 anggota pada cluster berisiko sedang, dan 35 anggota data pada cluster berisiko rendah. Analisis mendalam terhadap karakteristik masing-masing cluster mengungkapkan bahwa cluster berisiko tinggi atau cluster 3, terdiri dari kelompok dengan jumlah kasus bunuh diri paling besar. Di cluster ini, lansia mendominasi jumlah pelaku bunuh diri, di mana metode yang paling banyak digunakan adalah gantung diri, dan kasus-kasus ini sering terjadi di area rumah. Hal ini menunjukkan bahwa lansia mungkin lebih rentan terhadap faktor-faktor tertentu yang memicu tindakan tersebut, terutama ketika berada dalam lingkungan yang akrab seperti rumah mereka sendiri.

Pengelompokan data bunuh diri dengan menggunakan algoritma *K-Means* yang menghasilkan tiga kategori risiko di beberapa wilayah di Jawa Tengah ini membawa wawasan baru tentang faktor-faktor utama yang berkaitan dengan tindakan bunuh diri, seperti usia yang rentan, lokasi kejadian, dan metode yang digunakan. Temuan ini

berpotensi menjadi dasar penting bagi perumusan program pencegahan dan edukasi yang lebih terarah bagi masyarakat Indonesia. Selain itu, penelitian ini memberikan kontribusi penting dalam memahami profil risiko pada setiap cluster sehingga strategi pencegahan dapat lebih tepat sasaran, baik melalui edukasi kepada kelompok usia rentan maupun upaya memperketat pengawasan di lokasi-lokasi yang sering digunakan untuk bunuh diri. Dengan demikian, hasil penelitian ini diharapkan dapat mendukung pendekatan yang lebih efektif dalam mengurangi angka bunuh diri di Indonesia sesuai dengan karakteristik risiko yang ada.

DAFTAR REFERENSI

- [1] D. Novita, "Kesehatan Mental - Penyebab, Gejala, Dan Cara Mengobati - Klikdokter." Accessed: May 18, 2024. [Online]. Available: <https://www.klikdokter.com/penyakit/masalah-mental/kesehatan-mental>
- [2] A. Zulaikha And N. Febriyana, "Bunuh Diri Pada Anak Dan Remaja," *Jps*, Vol. 7, No. 2, P. 62, Dec. 2018, Doi: 10.20473/Jps.V7i2.19466.
- [3] "Setiap Jam, Satu Orang Bunuh Diri." Accessed: May 18, 2024. [Online]. Available: <https://health.kompas.com/read/2016/09/08/140000323/setiap.jam.satu.orang.bunuh.diri>
- [4] H. Mallo And D. Ronda, "Analisis Faktor Penyebab Utama Kecenderungan Bunuh Diri Di Kalangan Remaja Yang Berusia 15-17 Tahun Di Makassar".
- [5] S. Suryadi, "Penerapan Metode Clustering K-Means Untuk Pengelompokan Kelulusan Mahasiswa Berbasis Kompetensi," *Informatika*, Vol. 6, No. 1, Pp. 52–72, Oct. 2019, Doi: 10.36987/Informatika.V6i1.738.
- [6] P. M. C. Abrianto, "Penerapan Metode K-Means Clustering Untuk Pengelompokan Pasien Penyakit Liver," Vol. 2, No. 2, 2018.
- [7] K. Kunci, "The Role Of Mental Health As A Moderator Of Suicidal Tendencies Abstract".
- [8] D. Febrianti And N. Husniawati, "Hubungan Tingkat Depresi Dan Faktor Resiko Ide Bunuh Diri Pada Remaja Smpn," *J. Ilm. Kesehat.*, Vol. 13, No. 1, Pp. 85–94, Mar. 2021, Doi: 10.37012/Jik.V13i1.422.
- [9] K. Y. Kusuma, "Faktor Penyebab Kecenderungan Bunuh Diri Dikalangan Remaja Toraja," Vol. 2, 2022.
- [10] N. Rahmawati, M. H. E. Pramesty, F. Karno, U. Aprilia, And A. Hapsery, "Klasterisasi Kesehatan Ibu Dan Anak Di Indonesia Dengan Menggunakan Algoritma K-Means".
- [11] Y. Elda, S. Defit, Y. Yunus, And R. Syaljumairi, "Klasterisasi Penempatan Siswa Yang Optimal Untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means," *Jidt*, Pp. 103–108, Sep. 2021, Doi: 10.37034/Jidt.V3i3.130.
- [12] S. Dasgupta, N. Frost, M. Moshkovitz, And C. Rashtchian, "Explainable K-Means And K-Medians Clustering".
- [13] M. Ahmed, R. Seraj, And S. M. S. Islam, "The K-Means Algorithm: A Comprehensive Survey And Performance Evaluation," *Electronics*, Vol. 9, No. 8, P. 1295, Aug. 2020, Doi: 10.3390/Electronics9081295.
- [14] I. Virgo, S. Defit, And Y. Yuhandri, "Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering," *Jsisfotek*, Pp. 23–28, Mar. 2020, Doi: 10.37034/Jsisfotek.V2i1.17.