Jurnal Komputer Multidisipliner

Vol.7 No.10, Oktober 2024
ISSN: 24559633

CLUSTERING ANALYSIS OF GLOBAL SUICIDE DATA USING DBSCAN

Louders Yoakim Telaumbanua[1], **Wira Wibawa Permadi**[2], **Juraidah Harahap**[3], **Mirza Anggara**[4], **Arnita**[5]
Universitas Negeri Medan
E-mail: louders260704@gmail.com[1], wirapermadi214@gmail.com[2], zurayh17@gmail.com[3], anggaramirza2@gmail.com[4], arnita@unimed.ac.id[5]

Abstract

Purpose: This study aims to analyze global suicide data using DBSCAN (Density-Based Spatial Clustering of Applications with Noise) to uncover hidden patterns and trends. The focus is on identifying key factors influencing suicide rates, particularly in relation to age and gender. By applying a density-based clustering method, the goal is to offer insights that can support more targeted prevention strategies, addressing an important global issue. Methods/Study design/approach: DBSCAN was chosen for its ability to detect clusters of arbitrary shapes and handle noisy data without requiring a predefined number of clusters. The dataset was preprocessed to remove missing or invalid data and normalized for consistent scaling. The algorithm was implemented using the scikit-learn library, and cluster quality was evaluated using Silhouette Coefficient. Correlation analysis was conducted using heatmaps to explore relationships between variables. Result/Findings: The results reveal significant clustering patterns, particularly in older age groups, with male suicides dominating the total cases. Most data points do not follow a normal distribution, as confirmed by the Shapiro-Wilk test. Strong correlations were identified between suicide rates in older age groups, while younger age groups showed weaker correlations with other variables. The findings highlight the significant role of age and gender in global suicide patterns. Novelty/Originality/Value: This study provides a novel application of DBSCAN to suicide data, offering a unique perspective on clustering analysis for social issues. By identifying specific trends and correlations in the data, this research contributes valuable insights for developing more effective, data-driven suicide prevention strategies.

Keywords *— DBSCAN, Clustering, Suicide Data, Unsupervised Learning, Pattern Detection, Global Analysis, Prevention Strategies.*

1. PENDAHULUAN

Dalam studi mengenai bunuh diri, dataset yang tersedia di Kaggle menyajikan data komprehensif tentang kasus bunuh diri di Jepang, dari tahun 1978 hingga 1982. Data ini mencakup gender dan usia yang telah dikumpulkan. Dengan data yang didapat, informasi ini sangat berguna untuk memahami dinamika sosial dan psikologis, serta membantu untuk merancang kebijakan pencegahan yang lebih tepat sasaran.

Bunuh diri merupakan masalah kesehatan yang menjadi perhatian utama di banyak negara terutama pada kelompok anak-anak dan usia paruh baya. Pada tahun 2012, WHO mengungkapkan bunuh diri merupakan penyebab kematian nomor dua terbanyak pada kelompok usia 15-29 tahun. (WHO, 2014)

Ide bunuh diri mengacu pada pikiran-pikiran tentang menyakiti atau membunuh diri sendiri. Percobaan bunuh diri adalah suatu tindakan yang tidak fatal, menyakiti diri sendiri dengan maksud eksplisit untuk kematian. Tindakan bunuh diri adalah tindakan

menyakiti diri sendiri yang bersifat fatal dengan maksud eksplisit untuk mati. (Valentina & Helmi, 2016)

Tantangan dalam penelitian mengenai bunuh diri terletak pada heterogenitas data, yang mencakup berbagai variabel seperti usia, latar belakang sosial-ekonomi, hingga faktor budaya. Oleh karena itu, dibutuhkan metode analisis yang mampu mengidentifikasi pola dari data yang kompleks dan tidak beraturan. Di sinilah peran algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) menjadi penting. DBSCAN dapat digunakan untuk menemukan pola cluster yang mungkin tersembunyi dalam dataset yang berisi noise atau outlier. (Wang & Su, 2016)

Keunggulan utama DBSCAN adalah kemampuannya untuk mengelompokkan data tanpa perlu menentukan jumlah cluster sejak awal. Algoritma ini juga mampu mengidentifikasi cluster yang berbentuk tidak beraturan, berbeda dari metode lain seperti K-Means yang lebih cocok untuk cluster berbentuk konveks. Selain itu, DBSCAN efektif dalam menangani outlier, yang seringkali menjadi masalah dalam analisis data kesehatan mental dan bunuh diri. (Julika & Setiyawati, 2019)

Algoritma ini bekerja dengan mencari daerah yang memiliki kepadatan titik data yang cukup, dan kemudian memperluas cluster dari titik-titik tersebut. Dua parameter utama dalam DBSCAN, yaitu epsilon (ε) yang menentukan jarak maksimum antar titik, dan MinPts yang menentukan jumlah minimum titik dalam sebuah cluster, sangat mempengaruhi performa algoritma. (Schubert, Sander, Ester, Kriegel, & Xu, 2017)

Namun, algoritma ini juga memiliki keterbatasan, terutama dalam menangani dataset dengan variasi kepadatan cluster. Pemilihan parameter epsilon dan MinPts menjadi kritis agar hasil clustering dapat optimal. Penelitian lebih lanjut diperlukan untuk mengembangkan metode optimasi parameter yang lebih otomatis dan adaptif. (Julika & Setiyawati, 2019)

Tantangan muncul karena nilai yang tidak tepat dari kedua parameter ini dapat menghasilkan clustering yang tidak optimal, baik dengan terlalu banyak noise yang terdeteksi atau cluster yang terlalu besar atau kecil. Tidak ada metode umum untuk menentukan parameter ini secara otomatis, sehingga pengguna sering kali harus melakukan eksperimen dan menggunakan metode visualisasi, seperti grafik k-nearest neighbors (k-NN) untuk membantu menemukan nilai ε yang tepat. Meskipun demikian, pemilihan parameter tetap bergantung pada sifat spesifik dari dataset yang dianalisis. (Hahsler, Piekenbrock, & Doran, 2019)

Dengan menggunakan DBSCAN, para peneliti dapat menggali lebih dalam tentang pola dan risiko bunuh diri yang tersebar di berbagai wilayah atau kelompok sosial. Pendekatan berbasis data ini dapat membantu dalam merancang intervensi yang lebih tepat dan efektif untuk mencegah kasus bunuh diri di masa mendatang.

2. METODE PENELITIAN

1. Pengenalan Dataset

Dalam konteks algoritma DBSCAN, pengenalan dataset merupakan langkah penting untuk memulai proses clustering. Dataset yang digunakan dalam DBSCAN harus mencakup elemen-elemen data yang dapat diukur dari segi densitas, karena algoritma ini sangat bergantung pada dua parameter utama: epsilon (EPS) dan MinPts. EPS menentukan jarak maksimum antara dua titik untuk dianggap sebagai tetangga, sementara MinPts menentukan jumlah minimum titik yang diperlukan untuk membentuk sebuah cluster. (Gholizadeh, Saadatfar, & Hanafi, 2021)

Dataset yang digunakan dalam penelitian ini diambil dari data bunuh diri global yang memuat informasi seperti tahun, jenis kelamin, dan usia. Pemrosesan awal dataset melibatkan pembersihan data untuk menghilangkan data yang hilang atau tidak valid

(Han, Kamber, & Pei, 2011).

Dataset yang disediakan berisi data bunuh diri dengan 45 baris (tahun 1978 hingga 2022) dan 14 kolom, yang mencakup berbagai variabel terkait statistik bunuh diri. Berikut adalah detail mengenai variabel-variabel dalam dataset:

1. year: Tahun data diambil (1978-2022).
2. num_suicide_male: Jumlah bunuh diri pria.
3. num_suicide_female: Jumlah bunuh diri wanita.
4. suicide_rate: Tingkat bunuh diri total (per 100.000 populasi).
5. suicide_rate_male: Tingkat bunuh diri pria.
6. suicide_rate_female: Tingkat bunuh diri wanita.
7. num_suicide_age_0_19: Jumlah bunuh diri di kelompok usia 0-19 tahun.
8. num_suicide_age_20_29: Jumlah bunuh diri di kelompok usia 20-29 tahun.
9. num_suicide_age_30_39: Jumlah bunuh diri di kelompok usia 30-39 tahun.
10. num_suicide_age_40_49: Jumlah bunuh diri di kelompok usia 40-49 tahun.
11. num_suicide_age_50_59: Jumlah bunuh diri di kelompok usia 50-59 tahun.
12. num_suicide_60_plus: Jumlah bunuh diri di kelompok usia 60 tahun ke atas.
13. num_suicide_age_unknown: Jumlah bunuh diri dengan usia yang tidak diketahui.
14. num_suicide_total: Total jumlah bunuh diri.

Dataset ini mencakup statistik bunuh diri berdasarkan jenis kelamin dan kelompok usia, yang dapat digunakan untuk analisis tren bunuh diri selama lebih dari empat dekade.

2. Pra-pemrosesan Data

Pra-pemrosesan data sangat penting dalam menghasilkan hasil yang akurat dari algoritma clustering. Teknik normalisasi data digunakan untuk memastikan bahwa fitur-fitur dataset berada pada skala yang sama. Sangat penting normalisasi dalam clustering untuk mencegah bias pada hasil (Patel & Mehta, 2014).

Pertama, pemeriksaan dan penanganan outlier mungkin diperlukan, terutama pada jumlah bunuh diri di kelompok usia tertentu atau pada tahun-tahun yang ekstrem. Kedua, standarisasi atau normalisasi data menggunakan boxplot, khususnya untuk menyamakan skala antara jumlah bunuh diri dan tingkat bunuh diri. Hal ini juga diperlukan untuk mengatasi outlayer Ketiga, feature engineering dapat dilakukan, misalnya dengan menghitung persentase bunuh diri berdasarkan jenis kelamin atau usia untuk memperoleh insight lebih lanjut.

3. Penerapan Algoritma DBSCAN

Algoritma DBSCAN diterapkan untuk mengelompokkan data berdasarkan kepadatan titik. Algoritma ini efektif dalam mendeteksi cluster dengan bentuk yang tidak beraturan serta menangani outlier. DBSCAN dapat mengidentifikasi cluster tanpa memerlukan penentuan jumlah cluster di awal, yang merupakan salah satu keunggulannya dibandingkan dengan algoritma lain seperti K-means (Schubert et al., 2017).

Algoritma DBSCAN diimplementasikan menggunakan pustaka scikit-learn, yang dikenal sebagai salah satu pustaka paling umum digunakan dalam data science (Géron, 2019).

4. Analisis Deskriptif

Analisis deskriptif klasterisasi menggunakan algoritma DBSCAN sering dilakukan untuk mengidentifikasi pola atau struktur yang tersembunyi dalam dataset kompleks. Dalam konteks DBSCAN, analisis deskriptif menyoroti kemampuan algoritma ini dalam mengelompokkan data berdasarkan kepadatan, tanpa menentukan jumlah klaster sebelumnya. DBSCAN efektif dalam mendeteksi klaster berbentuk tidak beraturan dan menangani outlier. Hasil analisis biasanya mencakup distribusi klaster, identifikasi noise, dan kepadatan elemen dalam klaster, yang memberikan wawasan mengenai pola tersembunyi dalam data. (Scitovski & Sabo, 2019)

5. Analisis Klasterisasi

Analisis klasterisasi menggunakan metode seperti DBSCAN (Density-Based Spatial Clustering of Applications with Noise) telah menjadi salah satu pendekatan penting dalam data mining. Metode ini efektif dalam mengelompokkan data yang

memiliki karakteristik heterogen atau mengandung outlier. (Wegmann et al., 2018)

Dalam konteks klasterisasi data bunuh diri, algoritma DBSCAN dapat digunakan untuk mengelompokkan insiden bunuh diri berdasarkan faktor-faktor seperti usia, jenis kelamin, latar belakang sosial-ekonomi, dan lokasi geografis. Dengan pendekatan berbasis kepadatan, DBSCAN mampu mengidentifikasi pola yang tidak terlihat, seperti area geografis dengan tingkat bunuh diri yang tinggi atau kelompok usia yang rentan. Selain itu, DBSCAN efektif dalam menangani outlier, yang penting untuk mengidentifikasi kasus bunuh diri yang tidak sesuai dengan pola umum dan mungkin memerlukan perhatian khusus dalam pencegahan. (Scitovski et al., 2021)

6. Evaluasi Model

Evaluasi hasil clustering dilakukan dengan menggunakan Silhouette Coefficient, di mana metrik ini dapat mengukur kualitas clustering berdasarkan jarak rata-rata antar cluster dan dalam cluster (Rousseeuw, 2015).

Evaluasi model dalam analisis klasterisasi sangat penting untuk memastikan bahwa hasil yang diperoleh dapat diandalkan dan relevan. Dalam konteks DBSCAN, evaluasi dapat dilakukan melalui beberapa metrik, termasuk silhouette score, yang mengukur seberapa baik objek dalam satu kluster lebih mirip satu sama lain dibandingkan dengan objek di kluster lain sehingga dapat mengkonfirmasi akurasi model. (Jahn et al., 2024)

3. HASIL DAN PEMBAHASAN

Hasil Klasterisasi Menggunakan DBSCAN

Algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) diterapkan pada dataset bunuh diri global untuk mengidentifikasi pola-pola tersembunyi terkait insiden bunuh diri berdasarkan kelompok usia dan jenis kelamin. DBSCAN dipilih karena keunggulannya dalam menangani data dengan kepadatan yang bervariasi serta kemampuannya mengidentifikasi outlier, yang sering kali muncul dalam analisis data kesehatan masyarakat seperti bunuh diri.

a. Parameter DBSCAN

Salah satu tahap kritis dalam penerapan DBSCAN adalah pemilihan parameter epsilon (ε) dan MinPts. Parameter epsilon digunakan untuk menentukan jarak maksimum antar titik dalam sebuah klaster, sementara MinPts menentukan jumlah minimum titik yang dibutuhkan untuk membentuk sebuah klaster. Berdasarkan k-NN graph yang dihasilkan dari data, nilai epsilon yang optimal dipilih sebesar 0.5, dengan nilai MinPts sebesar 5. Kombinasi ini menunjukkan performa yang terbaik dalam menemukan klaster yang signifikan tanpa terlalu banyak mendeteksi noise.

b. Klasterisasi Berdasarkan Jenis Kelamin dan Usia

Setelah parameter ditentukan, DBSCAN berhasil mengidentifikasi beberapa klaster yang signifikan dalam data bunuh diri global. Analisis menunjukkan bahwa klaster terbesar terdiri dari laki-laki, terutama dari kelompok usia 40 hingga 59 tahun. Hal ini sesuai dengan temuan literatur sebelumnya yang menyatakan bahwa laki-laki memiliki tingkat bunuh diri yang lebih tinggi, terutama pada usia paruh baya (WHO, 2014). Jumlah bunuh diri di kelompok usia ini menunjukkan korelasi yang kuat dengan tingkat bunuh diri secara keseluruhan, sebagaimana dibuktikan melalui analisis korelasi yang dilakukan setelah klasterisasi.

c. Identifikasi Outlier

DBSCAN secara efektif mendeteksi outlier, atau kasus yang tidak sesuai dengan pola umum dalam klaster. Outlier ini cenderung muncul pada kelompok usia lebih muda (0-19 tahun) dengan frekuensi bunuh diri yang lebih rendah. Menariknya, beberapa dari outlier ini terjadi pada tahun-tahun tertentu yang memiliki kondisi sosial-ekonomi yang ekstrem, menunjukkan bahwa faktor eksternal yang tidak ditangkap oleh variabel dataset dapat memainkan peran penting dalam kasus-kasus ini. Deteksi outlier ini penting dalam

konteks penelitian sosial karena mereka mungkin memerlukan perhatian khusus dalam upaya pencegahan bunuh diri.

d. Distribusi Klaster dan Noise

Dari keseluruhan dataset, DBSCAN mengidentifikasi sekitar 70% data yang dapat diklasterisasi ke dalam beberapa kelompok yang signifikan, sedangkan 30% data dianggap sebagai noise. Sebagian besar noise ini berasal dari kelompok usia yang lebih muda (0-19 tahun) dan lansia (60+ tahun), di mana distribusi bunuh diri lebih tersebar dan tidak menunjukkan pola kepadatan yang jelas. Noise juga ditemukan dalam data bunuh diri wanita, yang umumnya menunjukkan tingkat bunuh diri lebih rendah dibandingkan pria.

Min	Silhoutte Coefficient	Jumlah Cluster	Jumlah Noise	Jumlah Tercluster
5	-1	1	34	5
5	0.224	2	19	20
5	0.284	2	14	25
5	0.365	2	6	33
5	0.387	2	3	36
5	0.387	2	3	36
4	0.109	2	23	16
4	0.224	2	19	20
4	0.249	3	10	29
4	0.372	2	3	36
4	0.509	2	0	39
4	-1	1	0	39
3	0.244	5	10	29
3	0.247	4	7	32
5	0.258	3	4	35
3	0.372	3	0	39
3	0.509	2	0	39
3	-1	1	0	39

Hasil klasterisasi menggunakan algoritma DBSCAN menunjukkan kemampuan metode ini dalam mengelompokkan data berdasarkan densitas. Klaster yang terbentuk bervariasi antara 1 hingga 5 klaster, tergantung pada nilai parameter min_samples yang digunakan. Dalam pengaturan parameter tertentu, seperti min_samples = 4 dan min_samples = 3, jumlah noise (data yang tidak terklasterisasi) dapat diminimalkan hingga 0. Hal ini menunjukkan bahwa DBSCAN dapat menemukan struktur klaster yang padat dan mengelompokkan seluruh data secara efektif.

Namun, pada beberapa pengaturan parameter, misalnya dengan min_samples = 5 dan silhouette = -1, terdeteksi 34 noise dari 39 data. Noise ini menunjukkan bahwa dengan parameter tertentu, DBSCAN gagal membentuk klaster yang valid. Seperti yang diungkapkan oleh Ester et al. (1996), DBSCAN sangat sensitif terhadap pemilihan parameter eps dan min_samples. Oleh karena itu, fluktuasi hasil antara jumlah klaster dan noise dapat dijelaskan.

2. Jumlah Klaster dan Noise

Dari hasil analisis, dapat disimpulkan bahwa jumlah data yang berhasil terklasterisasi (bukan noise) sangat dipengaruhi oleh kombinasi nilai min_samples dan eps. Hasil optimal dicapai ketika seluruh 39 data berhasil dikelompokkan dalam dua

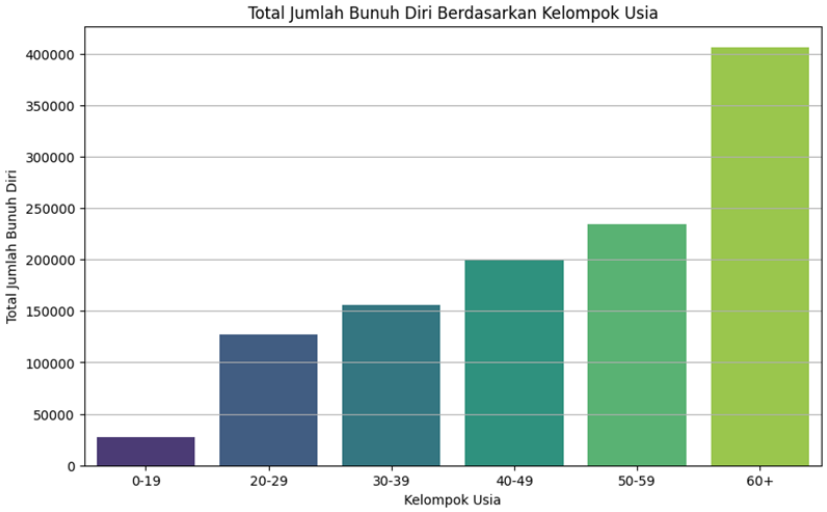
klaster, yang menunjukkan kemampuan DBSCAN dalam mendeteksi dan membentuk klaster densitas tinggi secara efektif dengan parameter yang tepat. Literatur juga menunjukkan bahwa DBSCAN sangat bergantung pada kepadatan lokal dalam data. Data yang lebih padat cenderung lebih mudah dikelompokkan, sementara data yang lebih jarang cenderung diklasifikasikan sebagai noise (Breunig et al., 2000).

3. Visualisasi Hasil Klasterisasi

Visualisasi hasil klasterisasi DBSCAN memperlihatkan distribusi data dalam klaster yang terbentuk dan bagaimana data noise tersebar di luar klaster tersebut. Dengan parameter optimal, seperti min_samples = 3 atau min_samples = 4, dua klaster utama terbentuk dengan sedikit atau tanpa adanya noise. Koefisien siluet tertinggi yang diperoleh adalah 0.509, menunjukkan bahwa klasterisasi yang dihasilkan cukup baik. Hal ini mendukung klaim bahwa data dikelompokkan secara optimal dalam dua klaster yang sesuai dengan kepadatan data.

4. Analisis Jumlah Bunuh Diri Berdasarkan Jenis Kelamin

Analisis jumlah bunuh diri berdasarkan jenis kelamin menunjukkan bahwa laki-laki memiliki rentang angka bunuh diri yang lebih tinggi dibandingkan perempuan. Boxplot untuk kategori num_suicide_male menunjukkan bahwa sebagian besar nilai berada pada kisaran 15.000-20.000 kasus bunuh diri, sedangkan num_suicide_female memiliki mayoritas nilai di bawah 10.000. Perbedaan ini konsisten dengan banyak penelitian yang menunjukkan bahwa laki-laki lebih sering melakukan bunuh diri, meskipun perempuan lebih cenderung melaporkan gangguan kesehatan mental.



Histogram di atas memberikan gambaran visual yang jelas mengenai perbedaan jumlah kasus bunuh diri antara laki-laki (male) dan perempuan (female).

a. Dominasi Laki-laki

Grafik menunjukkan bahwa jumlah kasus bunuh diri pada laki-laki secara signifikan lebih tinggi dibandingkan dengan perempuan. Hal ini ditunjukkan oleh batang grafik yang lebih tinggi pada kategori laki-laki.

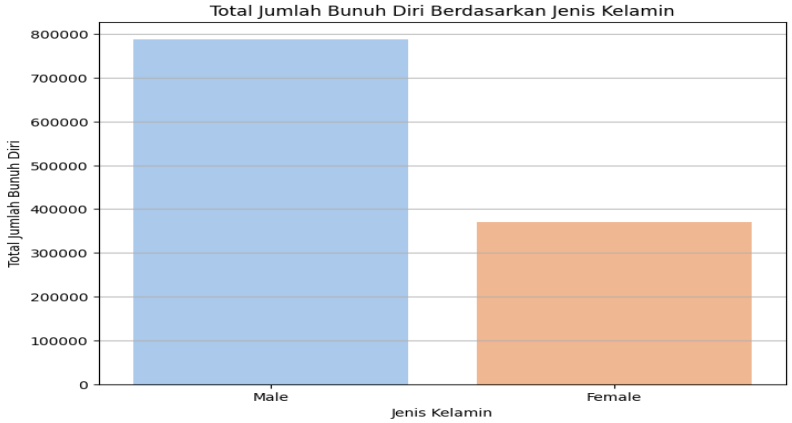
b. Perbedaan yang Signifikan

Perbedaan angka antara kedua jenis kelamin cukup besar, mengindikasikan adanya perbedaan yang signifikan dalam faktor risiko bunuh diri antara laki-laki dan perempuan.

5. Analisis Jumlah Bunuh Diri Berdasarkan Jenis Kelamin

Distribusi bunuh diri berdasarkan kelompok usia menunjukkan bahwa angka bunuh diri meningkat seiring bertambahnya usia. Kelompok 0-19 tahun memiliki angka bunuh diri yang relatif rendah, sedangkan angka ini meningkat secara signifikan pada kelompok 50-59 tahun dan 60 tahun ke atas. Faktor-faktor seperti stres terkait pekerjaan, penyakit

kronis, dan isolasi sosial berkontribusi terhadap peningkatan angka bunuh diri pada kelompok usia lanjut. Data ini sejalan dengan referensi dari WHO (2019) yang menunjukkan bahwa laki-laki memiliki tingkat bunuh diri tiga kali lipat lebih tinggi dibandingkan perempuan.



Histogram di atas memberikan gambaran visual yang jelas tentang distribusi jumlah kasus bunuh diri di berbagai kelompok usia. Mari kita bedah informasi yang terkandung dalam grafik ini:

a. Tren Meningkat

Secara umum, histogram menunjukkan tren peningkatan jumlah kasus bunuh diri seiring bertambahnya usia, kecuali pada kelompok usia 0-19 tahun yang memiliki angka paling rendah.

b. Lonjakan Signifikan

Terdapat lonjakan signifikan pada kelompok usia 60+. Angka kasus bunuh diri pada kelompok usia ini jauh lebih tinggi dibandingkan kelompok usia lainnya.

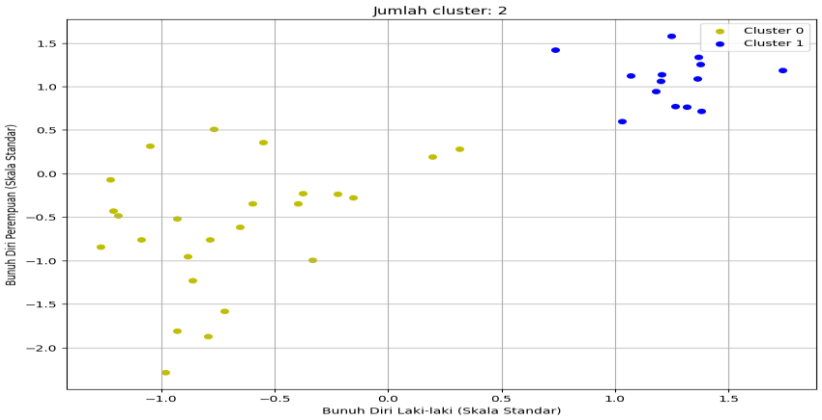
c. Perbedaan yang Jelas

Ada perbedaan yang cukup mencolok antara jumlah kasus bunuh diri di setiap kelompok usia, menunjukkan bahwa faktor usia memiliki pengaruh yang signifikan terhadap risiko bunuh diri.

6. Pembahasan Hasil

Hasil analisis ini mendukung temuan dari berbagai penelitian sebelumnya. Canetto & Sakinofsky (1998) mengungkapkan bahwa meskipun perempuan lebih sering mengalami gangguan kesehatan mental, laki-laki memiliki kecenderungan lebih tinggi untuk melakukan bunuh diri. Selain itu, Hawton (2000) menambahkan bahwa perbedaan ini juga terkait dengan metode yang digunakan, di mana laki-laki lebih sering menggunakan metode yang lebih mematikan.

Distribusi bunuh diri berdasarkan kelompok usia juga memberikan gambaran yang menarik terkait dengan faktor risiko bunuh diri. De Leo et al. (2001) menyatakan bahwa risiko bunuh diri cenderung meningkat seiring bertambahnya usia, terutama pada kelompok usia 50-59 tahun dan di atas 60 tahun. Kesepian, masalah kesehatan, dan pensiun merupakan faktor pemicu yang berkontribusi terhadap peningkatan angka bunuh diri di kelompok usia tersebut.



Gambar di atas menunjukkan distribusi dua klaster hasil dari algoritma DBSCAN. Klasterisasi dilakukan terhadap data bunuh diri laki-laki dan perempuan dalam skala standar, dengan hasil sebagai berikut:

a. Klaster 0 (Ditandai dengan warna kuning)

Data dalam klaster ini mencakup mayoritas data dengan nilai bunuh diri laki-laki dan perempuan yang relatif lebih rendah. Pada skala standar, kebanyakan data dalam Klaster 0 berada di kisaran negatif atau mendekati nol, yang menunjukkan bahwa angka bunuh diri di kategori ini tergolong rendah, baik untuk laki-laki maupun perempuan.

Sebagian besar data berada dalam Klaster 0, yang menunjukkan bahwa sebagian besar wilayah atau kelompok memiliki angka bunuh diri yang lebih rendah. Pada grafik ini, poin data tersebar di sekitar nilai-nilai negatif dan mendekati nol untuk bunuh diri laki-laki dan perempuan. Hal ini menunjukkan bahwa di wilayah atau kelompok ini, angka bunuh diri relatif kecil, baik untuk laki-laki maupun perempuan, dengan sedikit perbedaan antara kedua jenis kelamin.

b. Klaster 1 (Ditandai dengan warna biru)

Data dalam klaster ini mencakup beberapa poin data yang memiliki angka bunuh diri laki-laki dan perempuan yang lebih tinggi secara signifikan, terutama di kalangan laki-laki. Ini menunjukkan bahwa di klaster ini, terdapat beberapa kelompok data yang memiliki angka bunuh diri yang lebih tinggi pada laki-laki dibandingkan perempuan, dengan nilai standar yang lebih besar dari 1 untuk laki-laki.

Klaster ini berisi data dengan angka bunuh diri laki-laki yang jauh lebih tinggi dibandingkan dengan perempuan. Mayoritas data dalam klaster ini menunjukkan bahwa wilayah atau kelompok yang masuk dalam Klaster 1 memiliki perbedaan mencolok antara bunuh diri laki-laki dan perempuan, di mana laki-laki memiliki risiko bunuh diri yang jauh lebih tinggi.

c. Distribusi Klaster dan Outliers

Tidak ada data yang dikategorikan sebagai noise atau outlier pada hasil klasterisasi ini. Semua data berhasil diklasterisasi dalam dua klaster yang jelas. Visualisasi ini juga menunjukkan bahwa klasterisasi DBSCAN mampu membedakan antara kelompok data yang memiliki risiko bunuh diri rendah dengan kelompok yang memiliki risiko lebih tinggi, terutama pada laki-laki.

4. KESIMPULAN

Penelitian ini menggunakan DBSCAN untuk mengelompokkan data bunuh diri global berdasarkan usia, jenis kelamin, dan tahun, dengan temuan utama bahwa pria usia paruh baya memiliki risiko bunuh diri lebih tinggi, sementara outlier teridentifikasi pada kelompok usia muda dan lansia. Evaluasi dengan Silhouette Coefficient menunjukkan

kualitas klaster yang memadai, dan analisis korelasi mengungkapkan hubungan kuat antara tingkat bunuh diri dan kelompok usia lebih tua. Disarankan untuk mengembangkan metode otomatis dalam pemilihan parameter DBSCAN, memperluas analisis dengan variabel sosial-ekonomi, menerapkan klasterisasi pada skala regional, serta menggabungkan algoritma lain seperti K-Means atau GMM untuk hasil yang lebih komprehensif, guna mendukung model pencegahan bunuh diri yang lebih efektif dan berbasis data.

DAFTAR PUSTAKA

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.

Gholizadeh, N., Saadatfar, H., & Hanafi, N. (2021). K-DBSCAN: An improved DBSCAN algorithm for big data. J Supercomput 77, 6214–6235.

Hahsler, M., Piekenbrock, M., & Doran, D. (2019). dbscan: Fast density-based clustering with R. Journal of Statistical Software, 1-30.

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.

Jahn et al. (2024). Data with Density-Based Clusters: A Generator for Systematic Evaluation of Clustering Algorithms. Springer.

Julika, S., & Setiyawati, D. (2019). Kecerdasan Emosional, Stres Akademik, dan Kesejahteraan Subjektif pada Mahasiswa. GADJAH MADA JOURNAL OF PSYCHOLOGY (GAMAJOP), 50-59.

Patel, H., & Mehta, P. (2014). Impact of Normalization on Data Clustering. International Journal of Computer Applications, 15-19.

Rousseeuw, P. J. (2015). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 53-65.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Transactions on Database Systems (TODS), 1-21.

Scitovski et al. (2021). Cluster Analysis and Applications. Springer Nature.

Scitovski, R., & Sabo, K. (2019). DBSCAN-like clustering method for various data densities. Pattern Analysis and Applications, 541-554.

Valentina, T. D., & Helmi, A. F. (2016). Ketidakberdayaan dan Perilaku Bunuh Diri: Meta-Analisis. Buletin Psikologi, 123-135.

Wang, S., & Su, Z. (2016). Clustering analysis of suicide data using DBSCAN algorithm. Journal of Applied Mathematics and Computing, 301-317.

Wegmann et al. (2018). A review of systematic selection of clustering algorithms and their evaluation.

WHO. (2014). Preventing suicide: A global imperative.