

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Jumlah kejadian kriminalitas yang beragam di tiap daerah merupakan salah satu masalah sosial yang sedang di hadapi oleh Negara Indonesia. Statistik kriminalitas sering kali digunakan sebagai indikator penting untuk menilai keamanan suatu wilayah/daerah. Berdasarkan data Badan Pusat Statistik (BPS), jumlah kasus kriminalitas di Indonesia menunjukkan kecenderungan yang bervariasi dalam beberapa tahun terakhir. Hal ini dapat dilihat dari jumlah kejadian kejahatan terhadap nyawa pada tahun 2018 sampai tahun 2022 yang cukup fluktuatif dengan kecenderungan menurun. Jumlah kasus kejahatan terhadap nyawa pada tahun 2018 sebanyak 1024 kasus, yang mengalami penurunan pada tahun 2019 dan 2020 sebanyak 965 dan 898 kasus secara berturut-turut, namun mengalami kenaikan pada tahun 2021 yaitu sebanyak 927 kasus dan pada tahun 2022 kembali mengalami penurunan sebanyak 854 kasus (Badan Pusat Statistik, 2023).

Kriminalitas adalah masalah sosial yang kompleks yang mempengaruhi keamanan dan kesejahteraan masyarakat. Kriminalitas yang tinggi juga dapat menjadi indikator buruknya kondisi sosial ekonomi suatu daerah (Yuzani dkk., 2024). Tindakan kriminalitas merupakan suatu tindakan yang melanggar nilai Pancasila yang menjadi dasar ideologi Negara Kesatuan Republik Indonesia, sehingga tindakan kriminalitas tidak akan lepas dari tindak pidana (Simanunjak dkk., 2023).

Berdasarkan laporan Global Organized Crime Index (2023), *score* kriminalitas Indonesia menduduki posisi ke-20 pada skala global, yang mana ini menandakan tingginya kejadian kriminalitas di Indonesia dibanding dengan negara lain. Hal ini pastinya disebabkan oleh berbagai faktor. Misalnya menurut Febrianti dkk (2023) tindakan kriminalitas di Indonesia dipengaruhi oleh faktor internal, seperti perilaku individu, dan faktor eksternal, seperti lingkungan masyarakat. Faktor-faktor ini berkaitan satu sama lain sehingga menciptakan

variasi pada pola kriminalitas di berbagai provinsi. Asdami dkk (2024) juga menambahkan bahwa perilaku masyarakat sangat dipengaruhi oleh nilai-nilai budaya, norma sosial, tingkat pendidikan, dan tekanan lingkungan. Selain itu, karakteristik sosial, ekonomi, dan geografis yang berbeda di setiap daerah juga mempengaruhi pola perilaku masyarakat, yang pada akhirnya menciptakan perbedaan tingkat kriminalitas di daerah yang berbeda.

Analisis berbasis data menjadi penting untuk memahami pola-pola ini, karena dapat membantu mengidentifikasi karakteristik yang berkontribusi pada kriminalitas di setiap daerah. Salah satu metode analisis yang efektif dalam mengelompokkan daerah berdasarkan tingkat kriminalitasnya adalah metode analisis *clustering*, yang memungkinkan pengelompokan wilayah berdasarkan karakteristik kriminalitas yang serupa (Rosiana dkk., 2023). Menurut Agung Triayudi dkk. (2024) metode *K-Means* adalah jenis metode *clustering* yang paling umum untuk digunakan. Metode *K-Means* menggunakan *centroid*, yaitu jarak rata-rata pusat titik data, sehingga menjadikan metode *K-Means* lebih mudah digunakan dan efektif dalam menemukan *cluster* dalam bentuk *elips* atau bola.

Meskipun demikian metode *K-Means* memiliki beberapa kekurangan yang perlu diperhatikan. Misal menurut Alfina dkk. (2012) hasil pengelompokan dengan metode *K-Means* sangat bergantung pada inisialisasi nilai *centroid* awal *cluster* yang menyebabkan solusi *cluster* hanya berupa solusi yang sifatnya *local optimum*. Kekurangan metode *K-Means* lainnya berhubungan dengan *outlier*, yang mana metode *K-Means* kurang mampu dalam menangani data yang memiliki *outlier* (Maulana Pranata dkk., 2017).

*Outlier* pada data sosial seperti data kriminalitas sangat rentan terjadi. Beberapa hal yang menjadi penyebab mengapa *outlier* pada data kriminalitas rentan terjadi adalah karena besar kemungkinan terjadinya perubahan perilaku sistem, perilaku curang, dan *human error* dalam penginputan data atau penentuan jenis kriminalitas dari tiap kasus yang terjadi (Sutresno dkk., 2018). Data *outlier* adalah data yang memiliki karakteristik unik yang berbeda dari kebanyakan data pada observasi dan muncul dalam bentuk nilai yang ekstrim (Sulaeman, 2021). *Outlier* sangat perlu untuk diperhatikan, selain dapat membuat data berdistribusi

tidak normal, *outlier* juga dapat berdampak pada pengambilan kesimpulan atau keputusan dalam penelitian (Sihombing dkk., 2023).

Menurut Rahma & Muhajir (2024) untuk data yang memiliki *outlier*, metode *clustering K-Medoids* lebih unggul dari pada metode *K-Means*, karena metode *K-Medoids* diciptakan untuk menangani kekurangan *K-Means* yang tidak peka terhadap *outlier*. Selain itu *output clustering* pada metode *K-Medoids* akan lebih konstan karena tidak bergantung pada penentuan inisialisasi nilai *centroid* awal *cluster*. Salah satu metode pengidentifikasian *outlier* untuk data *multivariat* adalah menggunakan metode *Mahalanobis Distance*. Pengidentifikasian *outlier* bertujuan agar hasil pengelompokan dapat lebih representatif.

Proses pengelompokan dengan menggunakan metode *K-Means* dan *K-Medoids* memerlukan metode pengukuran jarak. Hal ini diperlukan karena dalam mengelompokkan data, jarak antara suatu titik data dengan *centroid cluster* akan menjadi patokan dalam menentukan *cluster* dari titik data. Beberapa metode pengukuran jarak yang umum digunakan yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Minkowski Distance* (Nishom, 2019).

Selain itu evaluasi kualitas pengelompokan sangat dibutuhkan dalam melihat apakah suatu pengelompokan telah dikelompokkan secara baik atau tidak. Menurut Hasan (2024), *Silhouette Score* dan *Davies-Bouldin Index* memberikan informasi yang berharga dalam menunjukan kualitas dan kehomogenan *cluster* dalam suatu pengelompokan.

Berdasarkan uraian diatas, dilakukanlah penelitian dengan menggunakan metode analisis *clustering* untuk mengelompokkan jumlah kasus kriminalitas di Indonesia. Metode analisis *clustering* yang digunakan adalah metode *K-Medoids*, dengan judul penelitian yaitu **“Pengelompokan Provinsi di Indonesia Berdasarkan Jumlah Kasus Kriminalitas dengan Menggunakan Metode *Clustering K-Medoids*”**.

## 1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah:

1. Bagaimana pemilihan jumlah *cluster* optimum pada pengelompokan dengan menggunakan metode *Dunn Index*?
2. Bagaimana perbandingan metode pengukuran jarak *Euclidean Distance*, *Manhattan Distance* dan *Minkowski Distance* dalam penerapannya pada pengelompokan dengan menggunakan metode *clustering K-Medoids*?
3. Bagaimana kualitas *clustering* yang ditunjukkan oleh *Silhouette Score* dan *Davies Bouldin Index* untuk tiap-tiap pengelompokan yang didapat?
4. Bagaimana pengelompokan provinsi di Indonesia berdasarkan jumlah kasus kriminalitas menggunakan metode *clustering K-Medoids*?

## 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Penelitian ini hanya mencakup data jumlah kasus kriminalitas dari seluruh provinsi di Indonesia pada periode 2022.
2. Jumlah *cluster* yang digunakan dalam penelitian ditentukan dengan menggunakan metode *Dunn Index*.
3. Pengukuran jarak yang digunakan dalam penelitian adalah *Euclidean Distance*, *Manhattan Distance* dan *Minkowski Distance*.
4. Pendeteksian outlier dalam penelitian dilakukan menggunakan Metode pengukuran jarak *Mahalanobis Distance* sebelum proses *clustering*.
5. Proses evaluasi pengelompokan menggunakan pengukuran jarak *Euclidean Distance*.

## 1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Menganalisis jumlah *cluster* optimum pada pengelompokan yang didapat dari metode *Dunn Index*.
2. Menganalisis perbandingan metode pengukuran jarak *Euclidean Distance*, *Manhattan Distance*, dan *Minkowski Distance* dalam penerapannya pada pengelompokan dengan menggunakan metode *clustering K-Medoids*.

3. Menganalisis kualitas *clustering* yang ditunjukkan oleh *Silhouette Score* dan *Davies Bouldin Index* untuk setiap pengelompokan yang didapat.
4. Menganalisis pengelompokan provinsi di Indonesia berdasarkan jumlah kasus kriminalitas menggunakan metode clustering *K-Medoids*.

### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Mempermudah identifikasi wilayah-wilayah provinsi di Indonesia yang memerlukan perhatian lebih besar dalam hal keamanan.
2. Sebagai referensi untuk penelitian lebih lanjut dalam bidang analisis data, khususnya yang berkaitan dengan kriminalitas dan pengelompokan data.
3. Meningkatkan kesadaran masyarakat tentang pola dan tren kriminalitas di lingkungan mereka.

