

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Analisis multivariat merupakan analisis yang cocok untuk meringkas data dengan peubah yang banyak. Salah satu analisis dalam multivariat yang dapat digunakan untuk memahami dan mempermudah interpretasi data multivariat di antaranya adalah analisis kelompok (Rachmatin, 2014). Analisis kelompok (cluster analysis) merupakan salah satu teknik dalam analisis multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya. Analisis cluster mengelompokkan objek-objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam kelompok yang sama, serta mempunyai kemiripan satu dengan yang lain (Johnson & Wichern, 2002). Dalam analisis cluster dikenal dua metode pengelompokan, yaitu metode hierarki dan metode nonhierarki. Salah satu analisis cluster metode non-hierarki adalah algoritma K-Means. Algoritma K-Means termasuk dalam partitional clustering, yang merupakan suatu cara mempartisi himpunan data menjadi beberapa kelompok yang saling lepas dengan jumlah klaster yang sudah ditetapkan di awal (Nazeer & Sebastian, 2009). Beberapa kelebihan dari algoritma K-Means diantaranya adalah mudah diimplementasikan, memiliki tingkat konvergensi yang tinggi, dan menghasilkan klaster yang lebih padat jika dibandingkan dengan metode hirarki.

Cluster merupakan salah satu metode penting untuk dapat mengetahui kemiripan dari himpunan objek. Salah satu masalah yang paling penting dalam analisis cluster adalah evaluasi hasil pengelompokan untuk menemukan jumlah cluster yang paling optimum. Hal ini merupakan subjek utama dari validitas cluster. Validitas cluster

dilakukan untuk membandingkan algoritme clustering, untuk membandingkan dua set cluster, serta membandingkan dua hasil clustering untuk menentukan yang lebih baik (Dewanti, 2013). Validasi cluster akan menentukan apakah cluster yang diperoleh sudah baik dan mampu mewakili populasi yang ada.

Penentuan jumlah cluster yang optimum menggunakan metode indeks validitas. Terdapat tiga jenis kriteria indeks validitas, yaitu kriteria eksternal dan kriteria internal yang menggunakan uji statistik, dan kriteria relatif yang menggunakan kriteria yang spesifik (Khairati dkk, 2019). Pendekatan pertama didasarkan pada kriteria eksternal. Ini menyiratkan bahwa mengevaluasi hasil dari suatu algoritma clustering didasarkan pada pra-spesifikasi struktur yang diterima dari sebuah data dan mencerminkan intuisi pengguna tentang struktur pengelompokan dari data. Contoh metode dengan pendekatan eksternal yaitu Rand statistic, Jaccard coefficient, Huberts statistic, Q index, dan Folkes and Mallows index. Pendekatan kedua didasarkan pada kriteria internal. Pengguna dapat mengevaluasi hasil algoritma clustering dalam konsep kuantitatif yang didapat dari data. Contoh metode dengan pendekatan internal yaitu coefficient silhouette dan indeks X<sub>I</sub>e Beni. Pendekatan ketiga validitas cluster didasarkan pada kriteria relatif. Ide dasarnya adalah evaluasi struktur clustering dengan membandingkan struktur clustering lain yang dihasilkan dari algoritma clustering yang sama tetapi nilai-nilai parameter berbeda. Contoh metode dengan pendekatan relatif yaitu Dunn index, Davies-Bouldin index, root-mean-square standard deviation (RMSSTD) of the new cluster, semi-partial r-squared (SPR), dan r-squared (Halkidi dkk, 2001 dalam Dewanti, 2013).

Beberapa penelitian terkait dengan indeks validitas cluster adalah Dewanti (2013) menggunakan tiga indeks validitas cluster yakni indeks Dunn, Huberts statistic, dan koefisien silhouette untuk mengelompokkan tiga jenis data yaitu numeric, kategorik dan numeric serta kategorik. Hasil yang diperoleh menunjukkan bahwa indeks Huberts statistic dapat digunakan untuk ketiga jenis data tersebut. Sedangkan Ningrat,

dkk (2016) serta Haqiqi dan Robert (2015) mengaplikasikan indeks Xie Beni untuk menentukan jumlah cluster yang optimum pada Algoritma K-Means. Berdasarkan penelitian sebelumnya, diketahui bahwa sudah ada yang menggunakan perbandingan 2 indeks validitas. Akan tetapi, perbandingan indeks Xie-Beni dan indeks Huberts statistics pada cluster K-Means belum diteliti. Oleh karena itu, dalam penelitian ini akan dibandingkan kedua indeks tersebut untuk melihat indeks validitas yang mampu memberikan hasil maksimum pada cluster K-Means. Metode cluster dengan indeks Xie-Beni dan indeks Huberts statistic akan diterapkan pada data pengelompokan daerah rawan DBD di Kabupaten Gorontalo. Penyakit Demam Berdarah Dengue (DBD) adalah penyakit yang disebabkan oleh virus dengue, yang masuk ke peredaran darah manusia melalui gigitan nyamuk dari genus Aedes, misalnya Aedes aegypti atau Aedes albopictus. Penyakit DBD dapat muncul sepanjang tahun dan dapat menyerang seluruh kelompok umur. Penyakit ini berkaitan dengan kondisi lingkungan dan perilaku masyarakat (Depkes RI, 2004). Penyakit DBD sampai saat ini masih merupakan salah satu masalah kesehatan masyarakat di Indonesia dengan jumlah pasien yang cenderung meningkat serta daerah penyebaran yang semakin meluas.

Provinsi Gorontalo merupakan salah satu Provinsi Endemis DBD yang artinya setiap tahun dilaporkan kasus DBD dengan CFR  $\geq$  1%. Berdasarkan data dari Dinas Kesehatan Provinsi Gorontalo tahun 2016-2017, Kabupaten Gorontalo merupakan salah satu Kabupaten yang memiliki angka insiden rate DBD terbesar dibandingkan daerah lainnya. Tercatat sebanyak 301 kasus pada tahun 2016, pada tahun 2017 menurun menjadi 151 kasus (DINKES Prov. Gorontalo, 2017) dan tahun 2018 naik menjadi 224 kasus (BPS, 2019). Tingginya kasus DBD setiap tahunnya di Kabupaten Gorontalo membuat kasus DBD perlu mendapat perhatian yang khusus.

Menurut Kemenkes (2014) ada beberapa faktor yang mempengaruhi peningkatan dan penyebaran kasus DBD, diantaranya adalah (1) pertumbuhan penduduk yang tinggi, (2) urbanisasi yang tidak terencana dan tidak terkendali, (3) tidak adanya kon-

trol vektor nyamuk yang efektif di daerah endemis, dan (4) peningkatan sarana transportasi. Sedangkan menurut WHO (2002) dalam Sanggara dkk (2017) kejadian DBD dipengaruhi oleh faktor nyamuk sebagai vektor (penular penyakit) yang bertelur pada genangan air yang ditimbulkan akibat musim hujan, faktor lingkungan yaitu lingkungan yang tidak bersih, faktor iklim dan musim (seperti suhu, kelembaban, curah hujan, serta kecepatan angin), dan faktor manusia yang kemungkinan dapat tertular atau terjangkiti dan kemudian menjadi penderita Demam Berdarah Dengue (DBD).

Untuk memaksimalkan usaha menurunkan angka incident rate di Kabupaten Gorontalo maka perlu dilakukan pengelompokan dan pemetaan daerah-daerah yang rawan DBD berdasarkan faktor-faktor yang mempengaruhi kejadian DBD. Pengelompokan dan pemetaan daerah rawan DBD dapat membantu pemerintah untuk mengambil kebijakan yang tepat dalam menanggulangi kejadian DBD di Kabupaten Gorontalo. Pengelompokan dan pemetaan daerah dilakukan dengan metode cluster. Penelitian ini akan membandingkan validasi hasil cluster akan menggunakan indeks validitas Xie Beni dan indeks Huberts statistic untuk mengetahui indeks validitas mana yang memberikan jumlah cluster optimum pada pengelompokan daerah rawan DBD di Kabupaten Gorontalo.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang diatas, rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana perbandingan indeks validitas Xie Beni dan indeks Huberts statistic pada pengelompokan daerah rawan DBD di Kabupaten Gorontalo?
2. Bagaimana pengelompokan dan peta sebaran daerah rawan DBD di Kabupaten Gorontalo berdasarkan angka kejadian DBD dan faktor-faktor yang mempengaruhinya?

### **1.3 Tujuan Penelitian**

Berdasarkan latar belakang diatas, rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Mengetahui perbandingan indeks validitas Xie Beni dan indeks Huberts statistic pada pengelompokkan daerah rawan DBD di Kabupaten Gorontalo.
2. Mengetahui pengelompokkan dan peta sebaran daerah rawan DBD di Kabupaten Gorontalo berdasarkan angka kejadian DBD dan faktor-faktor yang mempengaruhinya.

### **1.4 Manfaat Penelitian**

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut.

1. Manfaat teoritis
  - Bagi akademisi penelitian ini diharapkan dapat menambah referensi, informasi dan wawasan teoritis mengenai metode validitas cluster.
  - Sebagai sumber informasi tentang pola kerawanan wabah penyakit DBD di Kabupaten Gorontalo
  - Menjadi referensi bagi penelitian selanjutnya.
2. Manfaat Praktis
  - Memberi informasi mengenai daerah-daerah yang rawan DBD di Kabupaten Gorontalo sehingga dapat membantu pemerintah dalam menurunkan angka incident rate DBD di masa mendatang.
  - Memberikan informasi kepada masyarakat untuk lebih memperhatikan lingkungannya sehingga dapat meminimalisir faktor-faktor yang menyebabkan kejadian DBD.