

BAB I

PENDAHULUAN

1.1. Latar Belakang

Demam Berdarah *Dengue* (DBD) adalah penyakit yang sering menyerang anak-anak dan orang dewasa dengan gejala utama berupa demam dan nyeri otot serta sendi, yang cenderung memburuk setelah dua hari pertama. Keberadaan penyakit DBD menjadi salah satu permasalahan kesehatan masyarakat di Indonesia yang terus mengalami peningkatan jumlah penderitanya dan penyebaran yang semakin meluas. Hal ini seiring dengan meningkatnya mobilitas dan kepadatan penduduk di berbagai wilayah. DBD umumnya menyebar di daerah-daerah yang beriklim tropis dan sub-tropis.[1]

Penyebaran penyakit DBD ditularkan melalui gigitan nyamuk dari *genus Aedes*, terutama *Aedes aegypti* dan *Aedes albopictus*, yang menyukai kondisi lingkungan yang kualitasnya buruk dan menyukai darah manusia. *Aedes aegypti* lebih menyukai permukiman penduduk, sedangkan *Aedes albopictus* habitatnya di kebun dan hewan. Penyakit Demam Berdarah *Dengue* (DBD) tersebar di hampir seluruh wilayah Indonesia, kecuali di daerah dengan ketinggian lebih dari 1000 meter di atas permukaan laut. Penyakit ini sering terjadi di wilayah tropis dan menyebabkan kejadian luar biasa (KLB)[2].

Kasus penyakit Demam Berdarah *Dengue* (DBD) dapat ditemukan hampir diseluruh kota dan kabupaten di Indonesia, Jumlah kasus demam berdarah di Indonesia pada akhir tahun 2022 mencapai 143.000 kasus. Provinsi Jawa Barat, Jawa Timur, dan Jawa Tengah menjadi wilayah dengan tingkat kejadian *dengue* tertinggi. Memprediksi bahwa jumlah kasus Demam Berdarah *Dengue* (DBD) di Indonesia mencapai 7,590,213 kasus atau 50 kali lebih tinggi dibanding jumlah kasus yang dilaporkan di tahun 2022[3]. Provinsi Jawa Timur menempati posisi kedua dalam jumlah kasus DBD tertinggi, dengan total 8.894 kasus tercatat di wilayah tersebut. Sisi paling barat Pulau Madura tepatnya Kabupaten Bangkalan merupakan daerah yang masuk pada daftar wilayah endemik untuk penyakit DBD di Provinsi Jawa Timur[3]. Hal ini menyebabkan Kabupaten Bangkalan berstatus menjadi wilayah endemik penyakit ini. Mengutip data profil kesehatan Kabupaten

Bangkalan tahun 2022 memperhitungkan *Incidence Rate* (IR) tahun 2022 mengalami kenaikan dibanding tahun 2021, dengan kasus pada tahun 2021 140 kasus, sedangkan pada tahun 2022 terdapat 163 kasus, yang lebih banyak terjadi pada jenis kelamin laki-laki yaitu 82 orang atau 50,3%, dan perempuan 81 orang atau 49,7% dari jumlah penemuan kasus pada tahun 2022.

Masih banyak penderita DBD di wilayah Bangkalan, hal ini menjelaskan perkembangan virus dengue yang cukup signifikan menjadi faktor utama terjadi penyakit DBD. Program Pemberantasan Sarang Nyamuk (PSN) menjadi langkah awal yang dilakukan untuk mencegah penyebaran penyakit DBD dan banyaknya penduduk di suatu daerah tertentu juga dapat mempengaruhi terjadi penyakit ini, maka dari beberapa unsur tersebut terdapat ciri-ciri yang membentuk suatu pengclusteran daerah yang sering terjangkit penyakit DBD[1].

Analisis *clustering* adalah proses pengelompokkan sejumlah besar data berdasarkan ciri khasnya, algoritma clustering yang paling efektif untuk menangani data dengan kepadatan yang bervariasi adalah algoritma *density based clustering*[4]. *Clustering* merupakan salah satu teknik yang menjanjikan dalam data mining, proses pengelompokkan sekumpulan objek ke dalam kelas-kelas atau *cluster* sehingga objek-objek di dalam sebuah *cluster* memiliki kemiripan yang tinggi. Dengan *clustering*, seseorang dapat mengidentifikasi wilayah yang padat dan jarang, sehingga memungkinkan untuk menemukan pola penyebaran secara keseluruhan. Namun, menemukan *cluster* dalam data bisa menjadi tantangan, terutama ketika *cluster* memiliki ukuran, bentuk, dan kepadatan yang sangat beragam, serta ketika data mengandung noise dan outlier[5]. Sehingga diperlukan algoritma clustering yang sesuai dengan karakteristik data, seperti metode K-Means yang mengelompokkan data berdasarkan jarak ke pusat *cluster*, serta metode DBSCAN yang mampu membentuk klaster berdasarkan kepadatan dan mendeteksi outlier.

Metode K-Means merupakan salah satu algoritma clustering yang paling populer dan banyak digunakan karena kesederhanaan dan efisiensinya. K-Means membagi data ke dalam sejumlah klaster berdasarkan jarak ke pusat klaster (centroid) dan memaksimalkan kedekatan data dalam setiap klaster[1]. Namun, K-

Means memiliki kelemahan dalam menangani data dengan bentuk cluster yang tidak bulat, kepadatan yang berbeda, serta tidak dapat mendeteksi outlier atau noise.

Metode DBSCAN merupakan algoritma clustering berbasis kepadatan yang lebih fleksibel dalam mengelompokkan data dengan berbagai bentuk dan ukuran klaster, serta memiliki kemampuan untuk mendeteksi noise. Dalam metode ini, kepadatan diukur berdasarkan jarak yang ditentukan oleh parameter Epsilon dan MinPts. Epsilon menentukan jarak maksimum yang diizinkan antara pusat titik dan anggota dalam sebuah cluster, sementara MinPts mengacu pada jumlah minimum titik yang diperlukan untuk membentuk cluster. Apabila kedua parameter tersebut telah terpenuhi, maka akan terbentuklah suatu *cluster*[6].

DBSCAN adalah algoritma pengelompokan berbasis kepadatan data. Dalam DBSCAN, setiap data dapat dikategorikan sebagai inti (*core*), batas (*border*), atau noise, tergantung pada kepadatan di sekitarnya. Algoritma ini memerlukan dua parameter: Epsilon (*Eps*), yang menentukan jarak maksimum antar data dalam kelompok, dan Minimal Points (*MinPts*), yang menunjukkan jumlah minimum data dalam radius *Eps* untuk membentuk kelompok. DBSCAN masuk kedalam pengelompokan sekatan (*partitioning*) dan eksklusif, yang artinya DBSCAN merupakan metode pengelompokan yang membagi set data ke dalam sejumlah kelompok yang tidak tumpang-tindih antara satu kelompok dengan kelompok yang lain. DBSCAN dapat mendeteksi data yang mempunyai jarak yang jauh dari suatu kelompok (*noise*) yaitu data yang tidak ikut dalam kelompok mayoritas[7].

Pada penelitian sebelumnya yang dilakukan oleh Anggara dan Rahman, terkait pengelompokan pasien Penyakit Ginjal Kronis (PGK) menjelaskan algoritma DBSCAN pada basis data pasien terdiagnosa PGK nilai *epsilon* dan nilai *MinPts* sangat berpengaruh terhadap hasil *clustering* untuk basis data pasien terdiagnosa PGK, dan didapatkan nilai *Silhouette* terbesar pada skenario kedua, yaitu skenario yang menghasilkan 2 *cluster* dengan nilai *Epsilon* = 3,5 dan *MinSample* = 5, dan didapatkan nilai *intercluster* terbaik yang didapatkan oleh skenario pertama yang mendapatkan hasil 3 *cluster* dengan nilai *Epsilon* = 3 dan *MinSample* = 5 sedangkan untuk *intraccluster* terbaik didapatkan oleh skenario kelima yang mendapatkan hasil 6 *cluster* dengan nilai *Epsilon* = 4 dan *MinSample* = 3 [8].

Penelitian yang dilakukan oleh Nurhaliza dan Mustakim melakukan pengelompokan Data Kasus Covid-19 di Dunia Menggunakan Algoritma DBSCAN, menghasilkan beberapa kesimpulan yaitu Hasil validitas *cluster* tertinggi menggunakan *Silhouette Index* (SI) terletak pada percobaan ke-21 dengan nilai *Eps* 0,2 dan *Minpts* 3. Pada percobaan tersebut diperoleh nilai SI sebesar 0,3624. Percobaan yang menghasilkan *Cluster* paling optimal pada penelitian ini menghasilkan 3 *cluster* yaitu cluster 0 yang merupakan data outlier, *cluster* 1 yang diisi oleh negara-negara yang mulai mengalami lonjakan kasus Covid-19, dan *cluster* 2 yang beranggotakan negaranegara dengan kenaikan kasus yang signifikan[9].

Pendekatan clustering serupa juga diterapkan dalam penelitian yang dilakukan oleh Muhammad Iqbal, Syaripuddin, dan Moh. Nurul Huda dengan judul "Implementasi Algoritma K-Means Clustering dengan Jarak Euclidean dalam Pengelompokan Daerah Penyebaran Covid-19 di Kabupaten Bogor." Berdasarkan hasil penelitian tersebut, diperoleh dua klaster tingkat risiko penyebaran COVID-19, yaitu klaster C1 dengan tingkat risiko penyebaran sedang yang terdiri atas 36 kecamatan, dan klaster C2 dengan tingkat risiko penyebaran tinggi yang terdiri atas 4 kecamatan. Nilai validitas yang diperoleh sebesar 0,71676 yang menunjukkan bahwa struktur hasil klaster tersebut kuat[10].

Penelitian ini bertujuan untuk mengidentifikasi cluster yang terbentuk berdasarkan kepadatan spasial data penyakit Demam Berdarah Dengue (DBD) serta mengelompokkan daerah-daerah penyebaran penyakit DBD berdasarkan kemiripan karakteristik data menggunakan metode K-Means dan DBSCAN. Oleh karena itu, penelitian ini mengaplikasikan dua pendekatan clustering, yaitu metode K-Means sebagai algoritma non-hierarki berbasis centroid, dan DBSCAN sebagai algoritma berbasis density (kepadatan).

Metode DBSCAN digunakan untuk mendeteksi pola penyebaran DBD yang tidak beraturan dan untuk mengelompokkan wilayah berdasarkan kepadatan data. Sementara itu, metode K-Means digunakan untuk membentuk kelompok berdasarkan jarak centroid dan homogenitas data. Validasi hasil clustering dilakukan menggunakan nilai *Silhouette Coefficient*, yang berguna untuk mengevaluasi kualitas dan jumlah cluster yang terbentuk serta membantu dalam

memilih metode clustering yang paling optimal. Hasil clustering ini diharapkan dapat membantu dan mempermudah Dinas Kesehatan Kabupaten Bangkalan dalam melakukan tindakan pencegahan dan pengendalian penyebaran penyakit DBD secara lebih tepat sasaran dan efektif.

1.2. Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya maka didapatkan permasalahan, berikut ini adalah permasalahan, metode usulan dan pertanyaan penelitian:

1.2.1. Permasalahan

Penyebaran penyakit Demam Berdarah Dengue (DBD) dipengaruhi oleh kondisi lingkungan yang mendukung perkembangan nyamuk *Aedes aegypti* dan *Aedes albopictus* sebagai vektor virus dengue. Salah satu faktor utama adalah tingginya curah hujan di Kabupaten Bangkalan yang menyebabkan terbentuknya genangan air, sebagai tempat berkembang biak nyamuk penyebab DBD. Banyaknya faktor yang memengaruhi penyebaran DBD tersebut menunjukkan perlunya analisis spasial untuk mengelompokkan wilayah berdasarkan karakteristik lingkungan dan kasus DBD yang terjadi. Dengan adanya pengelompokan ini, dapat diidentifikasi wilayah yang memiliki potensi penyebaran DBD yang tinggi maupun rendah. Informasi ini dapat menjadi dasar dalam mendukung pengambilan keputusan yang lebih terarah, seperti penerapan program Pemberantasan Sarang Nyamuk (PSN) atau fogging di wilayah yang paling berisiko.

1.2.2. Metode Usulan

Penelitian ini menerapkan pendekatan data mining untuk menganalisis penyebaran penyakit DBD di Kabupaten Bangkalan. Metode yang digunakan adalah Density-Based Spatial Clustering of Applications with Noise (DBSCAN), yang mampu mengelompokkan wilayah berdasarkan kepadatan data. Penggunaan metode ini diharapkan dapat menghasilkan pengelompokan (klaster) wilayah dengan tingkat akurasi yang baik, sehingga dapat menjadi acuan informasi bagi masyarakat dan pihak terkait dalam upaya meminimalisir penyebaran penyakit DBD di Kabupaten Bangkalan.

1.2.3. Pertanyaan Penelitian

1. Berapa Nilai Parameter Epsilon Dan Minpts Metode Dbscan Dan K Optimal Metode K Means Yang Menghasilkan Silhouette Coefficient Tertinggi Dalam Proses Clustering Penyebaran Penyakit Dbd Di Kabupaten Bangkalan?
2. Bagaimana hasil visualisasi pengelompokan wilayah penyebaran penyakit DBD di Kabupaten Bangkalan dalam bentuk peta Sistem Informasi Geografis (SIG)?

1.3. Tujuan dan Manfaat

Dari rumusan masalah tersebut maka dapat diperoleh tujuan dan manfaat sebagai berikut:

1.3.1. Tujuan

1. Menentukan nilai parameter Epsilon dan MinPts yang menghasilkan performa clustering terbaik berdasarkan nilai Silhouette Coefficient tertinggi pada penyebaran penyakit DBD di Kabupaten Bangkalan.
2. Menganalisis dan memvisualisasikan pola penyebaran penyakit DBD berdasarkan hasil clustering menggunakan metode DBSCAN dalam bentuk peta Sistem Informasi Geografis (SIG).

1.3.2. Manfaat

Besar harapan penelitian ini dapat memberikan beberapa manfaat diantaranya sebagai berikut:

1. Memberikan rekomendasi nilai Epsilon dan MinPts yang optimal dalam metode DBSCAN untuk menghasilkan clustering yang akurat dalam analisis penyebaran penyakit DBD di Kabupaten Bangkalan.
2. Memberikan solusi berbasis data dalam upaya pencegahan dan pengendalian penyebaran penyakit DBD di Kabupaten Bangkalan dengan mengidentifikasi daerah yang berisiko tinggi, sehingga dapat meminimalkan peningkatan jumlah kasus DBD.

1.4. Batasan – batasan

Adapun agar penelitian ini tetap berjalan sesuai dengan jalur yang ditentukan dan tidak terjadi penyimpangan, maka diberlakukan batasan-batasan masalah sebagai berikut:

1. Data yang digunakan untuk penelitian ini adalah data penyebaran penyakit DBD dari Dinas Kesehatan Kabupaten Bangkalan dan data pendukung dari website resmi Badan Pusat Statistik (BPS) dari tahun 2016 hingga 2023.
2. Data – data tersebut terdiri dari kepadatan penduduk, jumlah hari hujan, curah hujan, laki-laki, perempuan, larvasidasi dan Angka Bebas Jentik (ABJ).
 - a. Kepadatan penduduk merupakan berisi data total populasi yang tinggal dalam satu kilometer persegi wilayah Bangkalan .
 - b. Jumlah hari hujan merupakan informasi mengenai berapa hari hujan terjadi dalam satu tahun di wilayah Bangkalan.
 - c. Curah hujan merupakan rata-rata curah hujan tahunan di masing-masing wilayah Bangkalan (mm/th).
 - d. Laki – laki merupakan jumlah penderita penyakit Demam Berdarah Dengue (DBD) dari jenis kelamin laki-laki di wilayah Bangkalan.
 - e. Perempuan merupakan jumlah penderita penyakit Demam Berdarah Dengue (DBD) dari jenis kelamin Perempuan di wilayah Bangkalan.
 - f. Larvasidasi merupakan data kegiatan untuk mencegah demam berdarah dengan cara membasmi jentik di wilayah Bangkalan.
 - g. Angka Bebas Jentik (ABJ) merupakan persentase wilayah yang bebas dari jentik nyamuk.
3. Pada penelitian ini menggunakan metode DBSCAN dengan perhitungan jarak *Euclidean Distance*, metode uji validasi menggunakan *Silhouette Coefficient* dan divisualisasikan dalam peta melalui Sistem Informasi Geografis (SIG) untuk menunjukkan penyebaran DBD di Kabupaten Bangkalan.

1.5. Sistematika Skripsi

Sistematika penulisan skripsi ini dibagi menjadi lima bab, yang masing-masing bab telah dirancang untuk tujuan tertentu, berikut penjelasan masing-masing bab:

BAB I PENDAHULUAN

Pada bab ini merupakan bab pertama yang akan menjelaskan tentang latar belakang penelitian, rumusan masalah, tujuan dan manfaat.

BAB II LANDASAN TEORI

Pada bab ini memaparkan teori penunjang yang berkaitan dengan DBD, data mining, *Clustering* DBSCAN, *Clustering K-Means*, SC

(*Silhouette Coefficient*), serta penelitian-penelitian terdahulu yang berhubungan dengan permasalahan yang diteliti dan menjadi acuan penelitian ini.

BAB III METODE PENELITIAN

Pada bab ini menjelaskan tentang tahapan – tahapan Penelitian, berawal dari pengumpulan data, proses preprocessing data, menjabarkan perhitungan implementasi metode DBSCAN, dan K-Means dengan pengujian *clustering* menggunakan *Silhouette Coefficient*.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan mengenai hasil dan pembahasan dari proses penelitian yang telah dilakukan berupa source code, penjelasan source code, uji scenario, hasil analisis, dan implementasi antar muka sistem.

BAB V PENUTUP

Bab penutup merupakan bab terakhir yang berisi tentang kesimpulan dan saran pada penelitian yang telah dibuat.

DAFTAR PUSTAKA

Pada bab ini berisi tentang keterangan mengenai sumber rujukan yang digunakan dalam penyusunan skripsi. Pada penelitian ini daftar pustaka disusun secara otomatis dengan menggunakan aplikasi mendeley.