

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining merupakan salah satu bidang ilmu yang digunakan untuk menangani penggalian data berupa informasi dari suatu *database* yang besar (Nida, dkk., 2024). *Database* yang besar atau *big data* didefinisikan dengan adanya komponen aspek yang terdiri dari volume, variasi dan kebenaran/keakuratan. Ketiga aspek tersebut dapat memberikan indikasi yang jelas mengenai dimensi dari suatu *big data* dalam menjadi sebuah objek atau data penelitian (Cappa, dkk., 2020). Dalam prosesnya *data mining* akan menerapkan berbagai ranah keilmuan dari bidang matematika, statistika, *machine learning* dan *artificial intelligence*. Konsep matematika dalam *data mining* akan digunakan untuk melakukan identifikasi, pengambilan keputusan dan menghasilkan penemuan serta pengetahuan yang baru. *Output data mining* dapat digunakan untuk membantu pengambilan keputusan di masa mendatang (Normah, dkk., 2020). Salah satu jenis teknik yang biasa digunakan dalam *data mining* adalah *clustering*.

Clustering adalah salah satu teknik dalam *data mining* yang digunakan untuk menentukan kelompok objek data yang serupa (Nida, dkk., 2024). Secara lebih lengkap *clustering* juga dapat didefinisikan sebagai teknik pengelompokan data yang pada dasarnya dilakukan atas suatu kesamaan atau karakteristik tertentu, seperti jarak, kemiripan dan lain sebagainya. Dalam

proses *clustering*, nantinya data yang memiliki kemiripan karakteristik akan dikumpulkan dalam satu *cluster*, sedangkan data yang memiliki perbedaan karakteristik akan dikumpulkan dalam *cluster* yang lain. Dengan kata lain, data dengan tingkat kehomogenan yang tinggi akan ditempatkan dalam *cluster* yang sama dan antar masing-masing *cluster* tersebut akan memiliki tingkat keheterogenan yang tinggi. Terdapat beberapa jenis *clustering* yang berbeda dalam *data mining* seperti *clustering* berbasis kepadatan atau *density-based*, berbasis hierarki atau *hierarchical-based*, berbasis partisi atau *partitioning-based*, berbasis kisi atau *grid-based* dan berbasis model atau *model-based* (Ghazal, dkk., 2021). Selain itu, *clustering* juga dapat dikelompokkan menjadi dua jenis, yaitu pengelompokan hierarki dan *non-hierarki* (Sanusi, dkk., 2019). Untuk jenis *non-hierarki* yang berbasis partisi atau *partitioning-based* salah satu diantaranya adalah *K-Medoids* dan *Fuzzy Possibilistic C-Means*.

K-Medoids adalah metode *clustering* yang mengelompokkan suatu data atau n objek menjadi beberapa partisi ke dalam c cluster. *K-Medoids* juga biasa dikenal sebagai *Partitioning Around Medoids* atau PAM, dimana pusat *cluster*-nya adalah berupa *medoid*. *Medoid* sendiri merupakan objek yang letaknya terpusat dalam suatu *cluster* (Nahdliyah, dkk., 2019). Penggunaan *medoid* sebagai pusat *cluster* memiliki dampak yang baik karena pembagian *cluster* didasarkan pada *medoid* yang merupakan titik-titik data aktual dalam *cluster* sehingga *cluster* yang dihasilkan memiliki makna yang lebih dimengerti dalam analisis data (Rahmawati, O., dkk., 2024). Metode *K-Medoids* merupakan pengembangan dari metode *K-Means* dan termasuk ke dalam varian *clustering* yang cukup tangguh dalam mengatasi kelemahan yang terdapat pada *K-Means* ketika mengelompokkan suatu data.

Fuzzy Possibilistic C-Means adalah metode *clustering* yang merupakan pengembangan dari salah satu metode *clustering* lain yaitu *Fuzzy C-Means* (Apsari, dkk., 2020). *Fuzzy C-Means* sendiri merupakan metode *clustering* dimana keberadaan setiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan dari data tersebut (Lestari, dkk., 2023). *Fuzzy C-Means*

juga dapat didefinisikan sebagai suatu algoritma *fuzzy clustering* yang menerapkan metode pengelompokan partisi yang sebagaimana sebelumnya telah disebutkan bahwa letak objek atau data dalam tiap-tiap *cluster* ditentukan oleh nilai derajat keanggotaan. Kehadiran *Fuzzy Possibilistic C-Means* yang merupakan pengembangan dari *Fuzzy C-Means* dapat digunakan sebagai sebuah metode dengan kemampuannya untuk menangani ketidakpastian secara lebih luas melalui konsep fungsi keanggotaan yang bersifat *possibilistic*. Pengembangan tersebut membuat *Fuzzy Possibilistic C-Means* memiliki keunikan dan lebih fleksibel dalam menangani ketidakpastian serta kemungkinan adanya objek yang terhubung dengan beberapa kelompok sedemikian sehingga menjadikannya sebagai metode *clustering* yang lebih baik dari *Fuzzy C-Means*.

Clustering adalah suatu teknik matematika yang dirancang untuk mengungkapkan struktur klasifikasi dalam data yang dikumpulkan pada fenomena dunia nyata (Mirkin, 2013). Proses *clustering* akan didasarkan atas metrik matematis berupa nilai jarak yang berguna untuk menetapkan dan mengevaluasi pengelompokan data. Selain itu, *clustering* membutuhkan aljabar linear, optimasi fungsi objektif dan berbagai perhitungan matematika lainnya untuk dapat digunakan sebagai pilar dalam konsep *data mining* dan *machine learning* secara utuh (Deisenroth, dkk., 2020). Oleh karena itu, matematika yang merupakan pilar dalam proses *clustering* dapat digunakan untuk membantu berbagai proses pada ranah bidang keilmuan yang lain. Secara khusus dalam penelitian ini, *clustering* yang berfokus pada matematika akan digunakan untuk keperluan pengelompokan data dalam bidang kesehatan. Kesehatan adalah keadaan sejahtera dari badan, jiwa dan sosial yang memungkinkan orang hidup produktif secara sosial dan ekonomis (Efendi, dkk., 2022). Kesehatan juga merupakan salah satu bidang yang patut dijadikan prioritas bagi pemerintah sebagai upaya dalam melaksanakan kewajiban untuk memberikan kesejahteraan jasmani kepada masyarakat. Salah satu bagian dari masyarakat yang sangat rentan terganggu kesehatannya adalah balita. Balita adalah istilah yang umum digunakan untuk

usia anak hingga 5 tahun (Gunawan, dkk., 2018). Masa balita merupakan masa yang sangat penting untuk bertumbuh dan berkembang karena pada rentang usia tersebut anak-anak akan mengalami berbagai perkembangan fisik, mental dan perilaku.

Salah satu ancaman terbesar bagi kesehatan balita di Indonesia adalah keberadaan stunting atau malnutrisi kronik yang masih sangat mengakar kuat dalam masyarakat. Stunting adalah kondisi anak balita yang mengalami gagal tumbuh atau tinggi tubuhnya tidak sesuai dengan usianya yang disebabkan oleh kekurangan gizi kronis dalam seribu hari pertama kehidupan (HPK), terhitung mulai dari dalam kandungan hingga anak berusia dua tahun. Kondisi stunting dapat membuat balita menjadi sangat rentan terhadap penyakit dan menurunkan tingkat produktivitasnya di masa depan. Secara lebih luas, stunting juga berpotensi besar dalam menghambat pertumbuhan ekonomi, meningkatkan kemiskinan dan memperlebar ketimpangan. Menurut WHO pada tahun 2020, komponen malnutrisi yang mengganggu kesehatan balita terdiri dari gizi kurang dan gizi buruk sebesar 6.7%, serta yang paling utama adalah stunting (malnutrisi kronik) yang mencapai angka sebesar 22.2% atau sekitar 149.2 juta penderita (Kemenkes, 2022). Persentase balita dengan tingkat kesehatan yang baik atau mengalami gizi lebih di dunia hanya sebesar 5.7%. Oleh karena itu, penanganan kasus stunting menjadi sebuah pekerjaan rumah yang harus diselesaikan dengan baik.

Sebagai langkah awal menangani stunting di Indonesia, presiden menerbitkan Peraturan Presiden (PP) Nomor 72 Tahun 2021 yang mengatur tentang percepatan penurunan stunting di Indonesia. Berdasarkan PP tersebut, Indonesia akan menargetkan penurunan kasus stunting sebesar 14% pada akhir tahun 2024 (Rahman, dkk., 2023). Selain itu, untuk menurunkan kasus stunting dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) dibentuk sebuah instrumen evaluasi yang dapat mengkaji hal tersebut. Instrumen evaluasi itu disebut sebagai Indeks Khusus Penanganan Kasus Stunting atau IKPS. Keberadaan IKPS akan berguna untuk mengukur

kinerja pelaksanaan program percepatan penurunan stunting di tingkat nasional, provinsi, hingga di tingkat kabupaten/kota. Dalam IKPS akan terdapat beberapa dimensi penilaian yang mengukur sejauh mana penanganan stunting di Indonesia. Dimensi-dimensi tersebut terdiri dari dimensi kesehatan, gizi, perumahan, pangan, pendidikan dan perlindungan sosial (BPS, 2023). IKPS apabila ditinjau dari tiga komponen aspek yang terdiri dari volume, variasi dan kebenaran/keakuratan termasuk ke dalam *big data* (data yang besar). Hal tersebut karena untuk aspek volume, IKPS berasal dari banyak provinsi dengan masing-masing indikator yang berbeda sehingga memerlukan analisis dan proses perhitungan yang cukup mendalam. Hal tersebut secara langsung berkontribusi pada besarnya volume data IKPS secara keseluruhan. Untuk aspek variasi, IKPS juga memiliki cukup banyak indikator yang termuat dalam beberapa dimensi, indikator-indikator kategorikal yang bersifat numerik tersebut menunjukkan suatu keberagaman yang memenuhi aspek variasi dari *big data*. Terakhir, untuk aspek kebenaran/keakuratan, sumber IKPS berasal dari BPS yang kualitas kebenaran dari datanya bernilai valid dan dapat digunakan untuk proses penelitian lebih lanjut.

Clustering dalam penelitian ini akan digunakan untuk mengetahui tingkat kesehatan balita dari seluruh provinsi di Indonesia yang dinilai berdasarkan pada penanganan kasus stunting. Menurut Peraturan Menteri Kesehatan Republik Indonesia Nomor 39 Tahun 2016, tingkat atau derajat kesehatan dapat diukur salah satunya dengan melihat prevalensi gizi kurang/stunting pada balita, sedemikian sehingga penanganan kasus stunting dapat dijadikan tolak ukur dalam melihat tingkat kesehatan balita pada setiap provinsi di Indonesia. Dengan kata lain, semakin baik penanganan kasus stunting pada suatu daerah, maka tingkat kesehatan balita pada daerah tersebut juga akan semakin baik, begitu juga sebaliknya.

Terdapat beberapa penelitian terdahulu yang relevan dengan penelitian ini. Kurniasari, dkk. (2024) menggunakan data kemiskinan dari provinsi di Indonesia yang terdiri dari beberapa variabel, kemudian diolah de-

ngan menggunakan metode *Fuzzy C-Means* dan *Fuzzy Possibilistic C-Means*. Pada penelitian tersebut diketahui bahwa metode terbaik adalah dengan menggunakan *Fuzzy Possibilistic C-Means* dimana hasil pengelompokan terdiri atas 2 *cluster*. Putriana, dkk. (2021) menggunakan data titik gempa pulau Sumatera tahun 2013-2018, kemudian diolah dengan menggunakan *Fuzzy Possibilistic C-Means*. Pada penelitian tersebut diketahui bahwa hasil pengelompokan terdiri atas 3 *cluster*. Yosia, dkk. (2024) menggunakan data *customer segment* dan *product segment*, kemudian diolah dengan menggunakan metode *K-Means* dan *K-Medoids*. Pada penelitian tersebut diketahui bahwa metode terbaik adalah dengan menggunakan *K-Medoids* dengan hasil pengelompokan untuk *customer segment* terdiri atas 3 *cluster* dan untuk *product segment* terdiri atas 4 *cluster*.

Hasil *clustering* dari *K-Medoids* dan *Fuzzy Possibilistic C-Means* terhadap suatu data memiliki nilai informasi yang penting untuk diketahui, karena baik itu *K-Medoids* ataupun *Fuzzy Possibilistic C-Means* merupakan metode *clustering* yang tercipta akibat dari pengembangan metode-metode *clustering* yang terdahulu. Kemampuan dan keunggulan masing-masing dari setiap metode dalam mengelompokkan suatu data akan memberikan informasi berupa kesimpulan untuk menentukan mana metode terbaik yang dapat dipilih ke depannya. Dalam menentukan metode *clustering* terbaik digunakan suatu nilai yang disebut sebagai *davies-bouldin index* (DBI). Dengan menggunakan data IKPS serta berdasarkan pada metode *K-Medoids* dan metode *Fuzzy Possibilistic C-Means* akan juga ditentukan jumlah *cluster* (c) optimal dalam mengelompokkan tingkat kesehatan balita di Indonesia. Untuk menentukan jumlah *cluster* (c) optimal pada *K-Medoids* akan digunakan uji penilaian dengan metode *silhouette* dan metode *elbow* (Rahmawati, T., dkk., 2024). Untuk menentukan jumlah *cluster* (c) optimal pada *Fuzzy Possibilistic C-Means* akan digunakan suatu konsep *cluster validity index* (CVI) yang terdiri dari *partition entropy* (PE), *partition coefficient* (PC) dan *modified partition coefficient* (MPC) (Lestari, dkk., 2023).

1.2 Perumusan Masalah

Berdasarkan penjelasan pada latar belakang, dapat diambil beberapa rumusan masalah yang diteliti sebagai berikut:

1. Bagaimana penerapan metode *K-Medoids* dan *Fuzzy Possibilistic C-Means* dalam mengelompokkan tingkat kesehatan balita berdasarkan penanganan kasus stunting di Indonesia?
2. Bagaimana interpretasi hasil pengelompokan tingkat kesehatan balita berdasarkan penanganan kasus stunting di Indonesia dengan menggunakan metode *K-Medoids* dan *Fuzzy Possibilistic C-Means*?
3. Metode *clustering* manakah yang lebih baik dalam mengelompokkan tingkat kesehatan balita berdasarkan penanganan kasus stunting di Indonesia?
4. *Cluster* mana yang harus diprioritaskan dalam meningkatkan kesehatan balita berdasarkan penanganan kasus stunting di Indonesia menurut metode dan hasil *clustering* terbaik?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui penerapan metode *K-Medoids* dan *Fuzzy Possibilistic C-Means* dalam mengelompokkan tingkat kesehatan balita berdasarkan penanganan kasus stunting di Indonesia.
2. Untuk mengetahui interpretasi hasil pengelompokan tingkat kesehatan balita berdasarkan penanganan kasus stunting di Indonesia dengan menggunakan metode *K-Medoids* dan *Fuzzy Possibilistic C-Means*.

3. Untuk mengetahui metode yang terbaik dalam mengelompokkan tingkat kesehatan balita berdasarkan penanganan kasus stunting di Indonesia.
4. Untuk mengetahui *cluster* mana yang harus diprioritaskan dalam meningkatkan kesehatan balita berdasarkan penanganan kasus stunting di Indonesia menurut metode dan hasil *clustering* terbaik.

1.4 Batasan Penelitian

Dalam suatu penelitian diperlukan batasan-batasan tertentu yang bertujuan untuk menghindari pembahasan di luar lingkup rumusan masalah. Batasan masalah dalam hal ini adalah sebagai berikut:

1. Data yang digunakan adalah data Indeks Khusus Penanganan Stunting atau IKPS tahun 2022 dari seluruh provinsi di Indonesia yang diperoleh dari Badan Pusat Statistik (BPS).
2. Pada metode *K-Medoids distance metric* yang digunakan adalah *euclidean*, sedangkan pada metode *Fuzzy Possibilistic C-Means distance metric* yang digunakan adalah *square euclidean*.
3. Dalam menentukan jumlah *cluster* (c) optimal pada *K-Medoids* akan memakai metode *elbow* dan metode *silhouette*, sedangkan pada *Fuzzy Possibilistic C-Means* akan memakai konsep *cluster validity index* (CVI) yang terdiri dari nilai *partition entropy* (PE), *partition coefficient* (PC) dan *modified partition coefficient* (MPC).
4. Perbandingan dalam menentukan metode terbaik antara *K-Medoids* dan *Fuzzy Possibilistic C-Means* akan berdasarkan pada nilai *davies-bouldin index* (DBI).
5. Jumlah *cluster* (c) yang dipilih adalah $c = 2, 3, \dots, 6$. Jumlah *cluster* (c) tersebut merupakan jumlah yang baik karena merepresentasikan nilai c yang ideal dalam proses *clustering*. Hal tersebut karena rentang

nilai $c = 2, 3, \dots, 6$ dapat dianggap cukup untuk mewakili variasi data tanpa menambah kerumitan yang tidak perlu dalam proses analisis (Kurniasari, dkk., 2020).

1.5 Manfaat Penelitian

Berdasarkan tujuan penelitian, manfaat dari penelitian ini adalah sebagai berikut:

1. Manfaat teoritis

(a) Bagi peneliti

Sebagai sarana untuk menambah pengetahuan sekaligus melatih diri dalam melaksanakan penelitian untuk mempraktikkan keilmuan matematika yang telah diperoleh selama masa perkuliahan.

(b) Bagi pembaca

Hasil dari penelitian dapat menjadi sumber informasi dan referensi untuk mengetahui lebih dalam mengenai metode *clustering* khususnya metode *K-Medoids* dan *Fuzzy Possibilistic C-Means*.

(c) Bagi universitas

Sebagai bahan evaluasi dan tolak ukur dalam memantau mutu perkembangan akademik sekaligus menambah khasanah keilmuan terkait persoalan matematika terapan di Universitas Negeri Jakarta khususnya program studi matematika.

2. Manfaat praktis

(a) Bagi pemerintah

Sebagai bahan informasi dan pertimbangan dalam menentukan kebijakan untuk meningkatkan kualitas kesehatan balita pada seluruh provinsi di Indonesia berdasarkan penanganan kasus stunting atas hasil pengelompokan serta interpretasinya.