

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sustainable Development Goals (SDGs) atau Tujuan Pembangunan Berkelanjutan (TPB) adalah pembangunan yang menjaga peningkatan kesejahteraan ekonomi masyarakat secara berkesinambungan, pembangunan yang menjaga keberlanjutan kehidupan sosial masyarakat, pembangunan yang menjaga kualitas lingkungan hidup serta pembangunan yang menjamin keadilan dan terlaksananya tata kelola yang mampu menjaga peningkatan kualitas hidup dari satu generasi ke generasi berikutnya (Badan Perencanaan Pembangunan Nasional, 2020). SDGs berisi 17 tujuan yang ingin dicapai, salah satunya yaitu air bersih dan sanitasi layak yang merupakan SDGs ke-6. Air merupakan sumber kehidupan bagi manusia. Ketersediaan air bersih dan sanitasi merupakan hal yang sangat krusial bagi masyarakat untuk memenuhi kebutuhan hidup sehari-hari. Peraturan Menteri Kesehatan RI No. 2 tahun 2023 menerangkan bahwa air bersih adalah air yang dapat digunakan untuk keperluan sehari-hari yang kualitasnya memenuhi syarat kesehatan dan dapat diminum. Air bersih dan sanitasi merupakan kebutuhan dasar meliputi air minum, *hygiene* dan sanitasi, kualitas air, efisiensi penggunaan air, dan pengelolaan sumber air. Air bersih memegang peranan penting dalam berbagai kegiatan sehari-hari seperti mandi, memasak, mencuci, dan berbagai kegiatan lain.

Pencapaian tujuan SDGs menjadi salah satu fokus utama pemerintah dan berbagai pihak di Indonesia, salah satu target dalam SDGs adalah target ke-6 mengenai air bersih dan sanitasi layak bagi masyarakat Indonesia. Publikasi Badan Pusat Statistik (2023) mengenai indikator Tujuan Pembangunan Berkelanjutan

(TPB) Indonesia menerangkan bahwa hal-hal yang mendukung tercapainya SDGs 6 mengenai sumber air bersih dan sanitasi layak beberapa di antaranya yaitu adanya akses sumber air minum yang layak, akses terhadap layanan sanitasi layak, adanya fasilitas cuci tangan menggunakan air bersih dan sabun, dan penurunan praktek Buang Air Besar Sembarangan (BABS) di tempat terbuka. Penelitian Azizah dan Syafiuddin (2022) menjelaskan bahwa akses sanitasi berupa ketersediaan Instalasi Pengolahan Air Limbah (IPAL) di beberapa provinsi masih belum memenuhi standar baku mutu yang ditetapkan. Nisa *et al* (2023) menganalisis program Penyediaan Air Minum dan Sanitasi Berbasis Masyarakat (PAMSIMAS) pada salah satu desa di Banjarmasin yaitu Desa Purwosari II. Hasil analisis program yaitu terdapat komponen yang cukup baik seperti perilaku masyarakat terhadap kebersihan dengan mempraktekkan buang air yang tidak sembarangan, namun prasarana air minum dan sanitasi yang bersih dikategorikan kurang baik. Akses air bersih dan sanitasi layak di Indonesia memiliki perbedaan dalam akses yang mendukungnya. Hal tersebut dapat mendukung adanya pengelompokan atau klasterisasi provinsi di Indonesia. Klasterisasi ini dapat membantu mengidentifikasi kelompok-kelompok provinsi dengan karakteristik serupa, sehingga intervensi dan kebijakan pemerintah dapat lebih tepat sasaran untuk meningkatkan pencapaian target SDGs ke-6 di seluruh provinsi Indonesia.

Analisis Klaster merupakan teknik pengelompokan multivariat yang tujuan utamanya untuk mengelompokkan objek berdasarkan karakteristik yang dimilikinya (Hair *et al*, 2006). Analisis *Cluster* akan mengelompokkan elemen/objek seperti orang, produk, toko atau organisasi ke dalam beberapa kelompok yang relatif homogen berdasarkan pada suatu set atribut yang dipertimbangkan untuk diteliti.

Dua metode pengelompokan dalam analisis kluster yaitu metode hierarki (*hierarchical methods*) dan metode tidak berhierarki (*nonhierarchical methods*). Metode hierarki merupakan metode yang menggabungkan observasi ke dalam sebuah hierarki atau struktur pohon (Hair *et al*, 2006), sedangkan metode non hierarki merupakan metode pengumpulan data berdasarkan jumlah k kluster (Johnson dan Wichern, 2007).

Algoritma *K-Means* merupakan algoritma umum yang digunakan dalam *clustering*. Algoritma *K-Means* terkenal karena kesederhanaan dan efisiensinya untuk mengkluster suatu data. *K-Means* juga dapat beradaptasi dengan baik untuk data yang terstruktur karena algoritma ini mengandalkan jarak *euclidean* untuk menentukan jarak antar data dan pusat kluster. *K-Means* termasuk ke dalam metode pengklasteran data secara *partitioning* yang memisahkan data ke dalam kelompok yang berbeda. Kelebihan metode *K-Means* yaitu waktu iterasinya efisien dan cepat serta metode pengelompokannya terkenal sederhana, namun memiliki kelemahan seperti sensitivitasnya terhadap *outlier* (Xu dan Wunsch, 2009).

Logika *Fuzzy* merupakan salah satu komponen pembentuk *soft computing*. Dasar logika *fuzzy* adalah teori himpunan *fuzzy*. Pada teori himpunan *fuzzy*, nilai keanggotaan sebuah objek atau data bukan bernilai 1 (ya) dan 0 (tidak), melainkan terletak pada rentang antara 0 sampai dengan 1 (Kusumadewi dan Purnomo, 2010). Semakin tinggi suatu nilai keanggotaan dari sebuah data, maka semakin tinggi pula nilai derajat keanggotaannya, begitu juga sebaliknya, semakin kecil nilai keanggotaan suatu data, maka semakin rendah juga derajat keanggotaannya. Metode *clustering* yang menggunakan logika *fuzzy* salah satunya adalah *Fuzzy C-Means* (FCM).

Fuzzy C-Means adalah suatu teknik pengklasteran data yang mana keberadaan tiap-tiap titik data dalam suatu klaster ditentukan oleh derajat keanggotaan (Kusumadewi dan Purnomo, 2010). *Fuzzy C-Means* pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. Konsep *Fuzzy C-Means* yaitu menentukan pusat klaster yang tepat untuk menandai lokasi rata-rata dari tiap klaster. Kelebihan *Fuzzy C-Means* adalah dapat melakukan proses *clustering* lebih dari satu variabel sekaligus dan juga memiliki kemampuan dalam mengelompokkan data yang besar. *Fuzzy C-Means* juga memiliki kemampuan untuk memberikan derajat keanggotaan yang menggambarkan seberapa kuat data tersebut termasuk dalam suatu klaster. Kekurangan *Fuzzy C-Means* adalah ketergantungan pada nilai *random* yang ditentukan di awal dan parameter awal seperti jumlah klaster (c) dan parameter *fuzziness* (w) yang dapat mempengaruhi hasil klastering karena pemilihan parameter yang tidak tepat dapat mengarah ke hasil yang kurang optimal (Xu dan Wunsch, 2009).

Kelebihan dan kekurangan dari metode *K-Means* dan *Fuzzy C-Means* dapat menjadi alasan dilakukannya penggabungan antara kedua metode. Metode *hybrid K-Means* dan *Fuzzy C-Means* diharapkan dapat menghasilkan nilai validasi yang lebih baik sehingga hasil klaster pun dapat lebih valid. Metode *K-Means* yang dilanjutkan *Fuzzy C-Means* dapat memfungsikan hasil *K-Means* untuk membentuk nilai matriks keanggotaan *fuzzy*. Hal tersebut dapat memberikan nilai awal yang lebih terarah untuk mencapai titik konvergen pada pemrosesan *Fuzzy C-Means*. Penelitian yang dilakukan oleh Duggirala (2019) mengenai segmentasi gambar menggunakan hibridisasi algoritma *K-Means* dan *Fuzzy C-Means* yang bertujuan untuk mencapai hasil pengelompokan yang lebih baik. Percobaan yang dilakukan menghasilkan

bahwa penggunaan algoritma *hybrid K-Means* dan *Fuzzy C-Means* dapat membentuk nilai performa yang lebih tinggi dibandingkan penggunaan metode *K-Means* dan *Fuzzy C-Means* secara terpisah. Hal tersebut dapat dilihat dari hasil yang diteliti berupa nilai pada waktu *Central Processing Unit* (CPU), kecocokan pengelompokan, dan *Sum of Squared* (SSE) dari berbagai gambar yang diteliti. Atiyah *et al* (2018) pun melakukan penelitian mengenai *hybrid clustering KC-Means* (*K-Means* dan *Fuzzy C-Means*). Penelitian tersebut menjelaskan mengenai waktu pemrosesan dan akurasi dalam penggunaan metode *hybrid K-Means Fuzzy C-Means* yang digunakan pada beberapa dataset yaitu iris, lens, dan wine. Penelitian ini menunjukkan bahwa terdapat nilai akurasi yang lebih unggul pada penggunaan metode *hybrid K-Means Fuzzy C-Means* dibanding metode penggunaan yang terpisah.

Berdasarkan latar belakang yang telah diuraikan sebelumnya, penelitian ini membahas mengenai penggunaan metode *Hybrid K-Means* dan *Fuzzy C-Means* berdasarkan akses terhadap sumber air dan sanitasi layak di Indonesia. Penelitian ini mengelompokkan 34 provinsi di Indonesia berdasarkan sumber air dan sanitasi yang divalidasi menggunakan validasi *Silhouette Coefficients*. Data penelitian menggunakan sumber data yang diperoleh dari Badan Pusat Statistik (BPS) mengenai akses sumber air dan sanitasi layak di Indonesia pada tahun 2023.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana hasil pengelompokan provinsi di Indonesia terhadap akses sumber air dan sanitasi layak dengan menerapkan metode *hybrid K-Means* dan *Fuzzy C-Means*?

2. Bagaimana nilai validitas *Silhouette Coefficient* pada jumlah klaster yang telah terbentuk?
3. Bagaimana hasil profilisasi 34 provinsi di Indonesia berdasarkan klaster yang terbentuk?

1.3 Batasan Masalah

Batasan-batasan masalah pada penelitian ini sebagai berikut:

1. Data yang digunakan yaitu data indikator akses terhadap sumber air dan sanitasi layak di Indonesia pada tahun 2023 berdasarkan indikator Tujuan Pembangunan Berkelanjutan (TPB) Indonesia. Data diambil dari publikasi *website* resmi Badan Pusat Statistik (BPS) Indonesia.
2. Penentuan klaster terbaik dilakukan menggunakan validasi *Silhouette Coefficient*.
3. Pengolahan data penelitian menggunakan bantuan R Studio dan *Microsoft Excel*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan dari penelitian ini sebagai berikut:

1. Menentukan pengelompokan provinsi di Indonesia terhadap akses sumber air dan sanitasi layak dengan menerapkan metode *hybrid K-Means* dan *Fuzzy C-Means*.
2. Menerapkan validasi *Silhouette Coefficient* untuk menentukan jumlah klaster terbaik yang terbentuk.
3. Memperoleh hasil profilisasi 34 provinsi di Indonesia pada klaster yang telah terbentuk.