BAB I

PENDAHULUAN

1.1 Latar Belakang

Indeks Pembangunan Manusia (IPM) digunakan sebagai tolok ukur untuk menilai keberhasilan pembangunan suatu daerah, yang didasarkan pada tiga aspek utama, yaitu kesehatan, pendidikan, dan kualitas hidup. Aspek kesehatan dinilai melalui indikator Umur Harapan Hidup (UHH), sedangkan aspek pendidikan ditinjau dari Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS). Sementara itu, standar kehidupan yang layak diukur melalui indikator Pengeluaran Per Kapita (PPK) yang telah disesuaikan. Berdasarkan *United Nations Development Programme (UNDP)*, IPM digunakan untuk menilai pencapaian pembangunan manusia melalui aspek-aspek dasar yang mencerminkan kualitas hidup [2]. Nilai IPM yang tinggi menunjukkan tingkat kesejahteraan dan kondisi sosial ekonomi yang lebih baik di suatu daerah.

Dalam menganalisis IPM, salah satu tantangan utama adalah sulitnya menentukan batas kategori karena nilai IPM antarkabupaten/kota di Jawa Timur cenderung berdekatan. Hal ini menyulitkan proses pengelompokan secara tegas. Untuk mengatasi permasalahan tersebut, dibutuhkan penerapan teknologi *data mining*. *Data mining*, atau dikenal juga sebagai *Knowledge Discovery in Database (KDD)*, merupakan proses untuk mengekstrak pengetahuan dari data berukuran besar guna menemukan pola-pola yang menghasilkan informasi bermanfaat [3]. Salah satu teknik dalam *data mining* yang relevan untuk kasus ini adalah *clustering*, yaitu metode pengelompokan data ke dalam beberapa klaster berdasarkan kemiripan karakteristik [4].

Pendekatan yang mampu menangani ketidakpastian dalam pengelompokan data adalah *clustering* berbasis *fuzzy*. Dalam *fuzzy clustering*, suatu objek dapat termasuk dalam lebih dari satu klaster dengan derajat keanggotaan yang berbeda-beda di setiap klaster, bergantung pada tingkat kesesuaiannya [5]. *Fuzzy C-Means (FCM)* dan *Possibilistic C-Means (PCM)* merupakan algoritma yang sering digunakan karena mampu menangani data dengan batas klaster yang kabur atau saling tumpang tindih.

FCM menggunakan fungsi derajat keanggotaan untuk menentukan sejauh mana suatu data termasuk dalam klaster tertentu, sehingga dapat meningkatkan akurasi pengelompokan dan mengurangi kesalahan klasterisasi [6]. Meski demikian, *FCM* memiliki sejumlah keterbatasan, seperti perlunya menentukan jumlah klaster sejak

awal dan penggunaan inisialisasi matriks partisi secara acak yang dapat menyebabkan hasil yang tidak konsisten [7]. Selain itu, *FCM* sensitif terhadap *noise* dan *outlier*, karena data yang jauh dari pusat kluster masih dapat memiliki nilai keanggotaan tinggi, sehingga menurunkan kualitas hasil klusterisasi.

Sebagai alternatif, *PCM* menawarkan pendekatan yang lebih tahan terhadap *noise* dan *outlier*, dengan membebaskan total keanggotaan data dalam satu kluster untuk tidak selalu berjumlah 1. Pendekatan ini dirancang untuk menangani *outlier* dengan lebih efektif dan memberikan hasil *clustering* yang lebih stabil dalam situasi data yang kompleks [8]. *PCM* bekerja dengan menyesuaikan parameter keanggotaan lokal untuk setiap kluster secara independen, sehingga mampu mempertahankan integritas kluster meskipun terdapat data yang berada di luar distribusi utama.

Beberapa metrik evaluasi digunakan untuk menilai kualitas klusterisasi, yaitu *Partition Coefficient (PC)*, *Partition Entropy (PE)*, dan *Modified Partition Coefficient (MPC)*. *PC* mengukur kejelasan keanggotaan data dalam kluster; nilai yang lebih tinggi menunjukkan struktur kluster yang lebih jelas. *PE* menilai ketidakpastian klusterisasi; nilai mendekati 0 menandakan kluster yang lebih baik. *MPC* dikembangkan untuk mengatasi kelemahan *PC*, dengan menilai kesamaan dalam kluster dan perbedaan antar-kluster; nilai *MPC* yang lebih tinggi menunjukkan hasil klusterisasi yang lebih baik [9].

Penelitian sebelumnya telah menerapkan metode *FCM* dalam pengelompokan IPM, baik secara nasional maupun provinsi. Isah et al. (2022) dalam penelitian "*Perbandingan Metode Fuzzy C-Means dan Fuzzy Probabilistic C-Means dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Pembangunan Manusia Tahun 2022*" menunjukkan bahwa *FCM* dengan tujuh kluster menghasilkan kualitas klusterisasi yang baik untuk pengelompokan data wilayah berdasarkan IPM [10]. Selain itu, Hartanto et al. (2024) dalam penelitian "*Analisis Indeks Pembangunan Manusia di Jawa Timur Tahun 2022–2023 Berdasarkan Indikator Menggunakan Metode Fuzzy C-Means*" menemukan bahwa *FCM* efektif dalam menganalisis IPM di Jawa Timur, dengan nilai *validity silhouette index* 0,7759 pada 2022 dan 0,7742 pada 2023 [11].

Selain itu, penelitian oleh Zhang et al. (2022) dalam "*An Online Weighted Bayesian Fuzzy Clustering Method for Large Medical Data Sets*" menunjukkan bahwa *Bayesian Optimization* dapat digunakan untuk menentukan parameter optimal dalam

clustering fuzzy [12]. Pendekatan ini dapat membantu pemilihan jumlah kluster (c) dan parameter *fuzziness* (m) secara lebih akurat dengan menggunakan metode probabilistik berdasarkan *Gaussian Process Regression (GPR)*. Dengan cara ini, hasil klusterisasi lebih optimal dan tidak bergantung pada inisialisasi awal yang acak.

Penelitian sebelumnya telah membandingkan algoritma *FCM* dan *PCM* menggunakan variabel utama penyusun IPM, namun cakupannya masih terbatas pada skala nasional. Hingga kini, belum terdapat studi yang secara khusus menerapkan perbandingan *FCM* dan *PCM* untuk pengelompokan wilayah kabupaten/kota di Provinsi Jawa Timur. Selain itu, pendekatan optimasi parameter dalam *clustering fuzzy* pada penelitian sebelumnya umumnya menggunakan metode *Grid Search*, yang memiliki keterbatasan efisiensi karena hanya mengevaluasi kombinasi parameter dalam ruang pencarian tetap. Untuk mengatasi hal tersebut, penelitian ini menggunakan *Bayesian Optimization* sebagai pendekatan yang lebih efisien dan adaptif dalam menentukan parameter optimal, seperti jumlah kluster dan nilai *fuzziness*.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma *FCM* dan *PCM* dalam pengelompokan kabupaten/kota di Provinsi Jawa Timur berdasarkan *Indeks Pembangunan Manusia* tahun 2023. Dengan menggunakan teknik evaluasi yang sesuai, diharapkan hasil penelitian ini dapat memberikan rekomendasi metode *clustering* yang lebih akurat untuk pengelompokan daerah berdasarkan tingkat pembangunan manusia, serta membantu pemerintah dalam perumusan kebijakan pembangunan daerah yang lebih tepat sasaran.

1.2 Rumusan Masalah

Merujuk pada uraian latar belakang yang telah disampaikan, maka dapat dirumuskan beberapa permasalahan yang akan menjadi fokus pembahasan, yaitu:

1. Bagaimana penerapan metode *Fuzzy C-Means (FCM)* dan *Possibilistic C-Means (PCM)* dalam proses klusterisasi kabupaten/kota di Provinsi Jawa Timur berdasarkan Indeks Pembangunan Manusia (IPM) tahun 2023?
2. Bagaimana hasil analisis perbandingan metode *Fuzzy C-Means (FCM)* dan *Possibilistic C-Means (PCM)* dalam mengelompokkan kabupaten/kota di Provinsi

Jawa Timur berdasarkan Indeks Pembangunan Manusia (IPM), ditinjau dari metrik evaluasi klusterisasi?

1.3 Batasan Masalah

Beberapa batasan yang diterapkan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya berfokus pada kabupaten/kota di Provinsi Jawa Timur dan tidak mencakup wilayah lain di Indonesia.
2. Variabel yang digunakan dalam analisis IPM terdiri dari Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita (PPK).
3. Penelitian ini hanya membandingkan metode *Fuzzy C-Means (FCM)* dan *Possibilistic C-Means (PCM)*.
4. Data yang digunakan dalam penelitian ini merupakan data Indeks Pembangunan Manusia (IPM) tahun 2023, sehingga tidak mencakup analisis tren IPM dari tahun-tahun sebelumnya.

1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengimplementasikan metode *Fuzzy C-Means (FCM)* dan *Possibilistic C-Means (PCM)* untuk mengelompokkan kabupaten/kota di Provinsi Jawa Timur berdasarkan Indeks Pembangunan Manusia (IPM).
2. Menganalisis dan membandingkan performa metode *Fuzzy C-Means (FCM)* dan *Possibilistic C-Means (PCM)* dalam mengelompokkan kabupaten/kota di Provinsi Jawa Timur berdasarkan Indeks Pembangunan Manusia (IPM), ditinjau dari metrik evaluasi klusterisasi.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi beberapa pihak, yaitu penulis, pembaca, dan penelitian selanjutnya:

1. Bagi Penulis

Dari segi penulis, penelitian ini bertujuan untuk:

- a. Menambah pemahaman dalam penerapan metode *Fuzzy C-Means (FCM)* dan *Possibilistic C-Means (PCM)* dalam proses klusterisasi wilayah berdasarkan Indeks Pembangunan Manusia (IPM).

- b. Memperdalam wawasan tentang evaluasi hasil klasterisasi menggunakan metrik seperti *Partition Coefficient (PC)*, *Partition Entropy (PE)*, dan *Modified Partition Coefficient (MPC)*.

2. Bagi Pembaca

Dari segi pembaca, penelitian ini bertujuan untuk:

- a. Memberikan referensi bagi akademisi, mahasiswa, dan praktisi yang tertarik dalam bidang *data mining* dan *fuzzy clustering*, khususnya dalam analisis pembangunan manusia.
- b. Menyediakan wawasan mengenai perbandingan metode *FCM* dan *PCM* dalam mengelompokkan daerah berdasarkan IPM.

3. Bagi Penelitian Selanjutnya

Untuk penelitian selanjutnya, penelitian ini bertujuan sebagai berikut:

- a. Dapat menjadi dasar untuk penelitian lanjutan, misalnya dengan menggunakan metode *clustering* lain atau menambahkan variabel tambahan dalam analisis IPM.
- b. Menjadi acuan dalam pengembangan model analisis pembangunan daerah berbasis *data mining*, baik dalam konteks Jawa Timur maupun wilayah lainnya.

4. Bagi Pemerintah

Untuk pemerintah sekitar, penelitian ini bertujuan sebagai berikut:

- a. Memberikan gambaran yang lebih jelas tentang karakteristik wilayah di Provinsi Jawa Timur berdasarkan IPM, yang dapat membantu dalam penentuan prioritas pembangunan daerah.
- b. Menyediakan informasi yang berguna dalam merumuskan kebijakan pembangunan yang lebih terarah dan efisien berdasarkan hasil klasterisasi, meskipun bukan menjadi dasar keputusan utama dalam penyusunan kebijakan.

Halaman ini sengaja dikosongkan

BAB II TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Dalam menyusun sebuah penelitian, diperlukan referensi dasar yang digunakan sebagai acuan. Referensi ini diperoleh dari penelitian sebelumnya yang telah dilakukan. Dengan mempelajari penelitian terdahulu, dapat diketahui keterkaitan antara penelitian yang sedang dikerjakan dengan penelitian sebelumnya, sehingga dapat menghindari terjadinya duplikasi penelitian.

Pada tahun 2022, Ghina Nabila Saputro Putri dkk. melakukan sebuah penelitian dengan judul "*Implementasi Algoritma Fuzzy C-Means dan Fuzzy Possibilistics C-Means untuk Klasterisasi Data Tweets pada Akun Twitter Tokopedia*". Penelitian ini bertujuan untuk mengelompokkan *tweet* dari akun Twitter *@tokopedia* berdasarkan jenis konten yang mendapatkan banyak *retweet* dan *like*. Metode yang digunakan dalam penelitian ini adalah *Fuzzy C-Means (FCM)* dan *Fuzzy Possibilistics C-Means (FPCM)*, dengan evaluasi performa menggunakan *Modified Partition Coefficient (MPC)*. Proses *klasterisasi* dilakukan sebanyak lima kali percobaan dengan jumlah *klaster* tiga hingga tujuh. Hasil penelitian menunjukkan bahwa *Fuzzy C-Means (FCM)* menghasilkan performa yang lebih baik dibandingkan *Fuzzy Possibilistics C-Means (FPCM)* dalam *klasterisasi* data *tweet*, dengan jumlah *klaster* optimal sebanyak empat *klaster*. *Klaster* yang terbentuk mencakup konten mengenai *promo*, *diskon*, *cashback*, *kuis* berhadiah, dan promosi acara Tokopedia. Dari hasil analisis, konten yang paling banyak mendapatkan *retweet* dan *like* adalah *tweet* mengenai *promo* barang otomotif, alat olahraga, dan *merchandise* artis ternama, sehingga dapat dijadikan sebagai strategi pemasaran yang lebih efektif bagi PT Tokopedia [13].

Pada tahun 2022, Laily Rizqi Amaliyah dkk. melakukan sebuah penelitian dengan judul "*Comparison of Fuzzy C-Means, Fuzzy Possibilistic C-Means, and Possibilistic Fuzzy C-Means Algorithms on the Distribution of Contraceptive Users in NTB Province*". Penelitian ini membandingkan tiga metode *clustering*, yaitu *Fuzzy C-Means (FCM)*, *Fuzzy Possibilistic C-Means (FPCM)*, dan *Possibilistic Fuzzy C-Means (PFCM)* dalam mengelompokkan distribusi pengguna kontrasepsi di Provinsi NTB. Hasil penelitian menunjukkan bahwa metode *FPCM* lebih unggul dalam jumlah iterasi dan efisiensi waktu komputasi, sedangkan metode *FCM* lebih baik berdasarkan nilai