

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pemerintah Indonesia yang memiliki tujuan untuk memajukan masyarakat umum dalam keadaan sosial yang nyata, hal ini karena sudah sesuai dengan pasal yang telah ditetapkan oleh pemerintah Indonesia dalam pasal 33 undang – undang dasar tahun 1945 [1]. Amanat konstitusional ini menjadi dasar dalam penyusunan berbagai kebijakan pembangunan di setiap wilayah agar program pembangunan menjadi lebih tepat sasaran [2]. Pembangunan yang inklusif mencakup aspek ekonomi sosial menjadi upaya nyata untuk meningkatkan kualitas hidup masyarakat [3]. Salah satu instrumen yaitu RPJMN selama periode 5 tahun masa jabatan. Hal ini diadaptasi oleh daerah dalam RPJMD untuk meningkatkan kualitas sumber daya manusia [4]. Seperti dari situs resmi yaitu Badan Pusat Statistik (BPS), Provinsi Jawa Timur menjadi wilayah yang menarik untuk diteliti karena, Provinsi Jawa Timur memiliki jumlah penduduk miskin tertinggi di Pulau Jawa sebesar 4,259 juta jiwa pada tahun 2021 [5]. Hal tersebut menjadikan tantangan untuk menanggulangi. Pemerintah Provinsi Jawa Timur yaitu dalam RPJMD menetapkan sejumlah kebijakan Upaya preventif, seperti halnya yaitu target dalam penurunan tingkat kemiskinan menjadi di bawah 9% pada tahun 2024, IPM berpengaruh dalam pembangunan infrastruktur terhadap wilayah yang tertinggal dan memajukan UMKM setempat.

Oleh karena itu, pertumbuhan ekonomi juga menjadi indikator yang paling penting dalam menilai stabilitas dalam kesejahteraan rakyat, hal itu karena kontribusi yang diberikan saat terciptanya banyak lapangan kerja, dari pendapatan yang terus meningkat, dan angka kemiskinan yang terus menurun. Laju pertumbuhan di Pulau Jawa tercatat sebesar 5,31% dan menyumbang 56,48% terhadap pertumbuhan ekonomi nasional di tahun 2022, di mana Provinsi Jawa Timur berada di posisi kedua dalam hal jumlah penduduk bekerja sebesar 21,61 juta orang setelah Jawa Barat [6]. Data yang didapatkan pada indikator yaitu PDRB pada tahun 2009 sampai 2016 menunjukkan bahwa wilayah Jawa Timur mencatat rata-rata tertinggi sebesar 1.167,76 triliun rupiah, hal ini mengungguli untuk wilayah Provinsi Jawa Barat dan Provinsi Jawa Tengah. Indikator pada PDRB ini menjadi keterkaitan erat dengan indikator IPM karena memengaruhi pertumbuhan ekonomi di Pulau Jawa. Hal ini terbukti adanya

nilai yang terus menurun hingga sebesar 0,099542 yang menunjukkan bahwa peningkatan yang terjadi pada IPM yaitu sebesar 1% dapat meningkatkan sebuah PDB dengan nilai sebesar 0,099542%, terutama melalui kontribusi yang dilakukan pada indikator AHH dan juga indikator HLS [7]. Ketidakseimbangan tersebut menjadikan sebuah kondisi pada suatu daerah yang ada di Jawa Timur baik wilayah perkotaan maupun pedesaan dan juga wilayah pinggiran menjadi kesejahteraan rakyat yang masih belum merata. Oleh karena itu, penelitian ini menjadi penting dilakukan karena untuk memahami pola pada setiap indikator dalam kesejahteraan rakyat. Pendekatan ini melalui klasterisasi dengan algoritma GMM, hasil yang akan didapatkan untuk pemetaan seluruh Kabupaten atau Kota di Provinsi Jawa Timur. Penelitian ini juga diharapkan dapat memberikan sebuah kontribusi terhadap perumusan kebijakan, serta ikut berpartisipasi dalam mendorong pembangunan disuatu daerah yang adil dan merata di seluruh wilayah Provinsi Jawa Timur [8].

Terdapat beberapa algoritma yang digunakan untuk klasterisasi dengan berbagai studi terkait isu-isu yang terjadi, salah satunya penelitian yang dilakukan oleh Karli Eka Setiawan & Afdhal Kurniawan pada tahun 2023 berjudul “Pengelompokan Rumah Sakit di Jakarta Menggunakan Model DBSCAN, Gaussian Mixture, Dan Hierarchical Clustering” [9]. Pada penelitian ini bertujuan untuk membandingkan performa dari tiga algoritma yaitu DBSCAN, *Gaussian Mixture Model* (GMM), dan *Agglomerative Hierarchical Clustering* dalam melakukan pengelompokan data rumah sakit di wilayah DKI Jakarta Dengan menggunakan data yang digunakan dari situs Kementerian Kesehatan Republik Indonesia yang mencakup tenaga medis serta fasilitas tempat tidur dari 191 rumah sakit. Hasil dari penelitian ini yang telah ditunjukkan tersebut menjelaskan bahwa model Gaussian Mixture mampu menghasilkan performa terbaik dengan nilai evaluasi berupa *Davies Bouldin Score* dengan nilai sebesar 0,6457, dibandingkan dengan algoritma lain seperti *Agglomerative Hierarchical Clustering* yaitu dengan nilai sebesar 0,6680, dan juga pada DBSCAN hanya sebesar 1,2258. Dengan selisih nilai evaluasi yang hanya sebesar 0,5791 terhadap algoritma seperti DBSCAN dan untuk nilai 0,0223 terhadap algoritma *Agglomerative Hierarchical*. Algoritma GMM memberikan sebuah hasil yang optimal pada data dari rumah sakit berdasarkan atribut yang meliputi tenaga medis dan fasilitas. Sehingga GMM mampu mengelompokkan secara efektif untuk distribusi fasilitas kesehatan di wilayah Jakarta.

Penelitian sebelumnya algoritma GMM menghasilkan performa terbaik melalui hasil *davies bouldin score* terendah melalui proses normalisasi dan reduksi dimensi seperti pada penelitian yang dilakukan oleh Zhang dan Ning pada tahun 2021 berjudul “*A Comparison for Dimensionality Reduction Methods of Single-Cell RNA-seq Data*” [10]. Penelitian ini dengan melalui evaluasi berbagai metode reduksi dimensi terhadap data single-cell RNA sequencing (scRNA-seq), dan juga termasuk metode yang digunakan yaitu ada PCA, ada juga t-SNE, UMAP, ICA, ZINB-WaVE, dan terakhir yaitu GLMPCA. Hasilnya metode PCA unggul dalam menjaga struktur global data serta efisien secara komputasi dan hasil reduksi pada data PBMC menjadi 50 komponen utama dalam waktu kurang dari 1 detik, hal ini menunjukkan bahwa hasilnya jauh lebih cepat dibandingkan metode reduksi yang lain seperti t-SNE hanya 14 detik, begitu juga dengan UMAP hanya 11 detik, dan yang terakhir yaitu reduksi menggunakan ZINB-WaVE melibih dari yang lain sekitaran 30 detik lebih. Akan tetapi, PCA efektif tetap mempertahankan data informasi yang bersifat linier itu yang menjadi kekurangan pada data bersifat non-linear terhadap variabel seperti data sosial ekonomi.

Untuk mengatasi masalah keterbatas tersebut, dalam penelitian ini melibatkan sebuah kombinasi reduksi antara principal component analysis dengan Uniform Manifold Approximation and Projection (UMAP) dengan judul “Initialization is critical for preserving global data structure in both t-SNE and UMAP” [11] menunjukkan penggunaan reduksi PCA memberi UMAP gambaran dalam mempertahankan struktur global data tanpa merubah bentuk dari keadaan aslinya yang sering muncul pada random *initialization*. PCA sendiri mampu menstabilkan untuk tahapan UMAP dan memastikan hasil *embedding* lebih konsisten dengan pola asli data. Dengan adanya penelitian terdahulu tersebut yang menunjukkan bahwa kombinasi PCA dan UMAP menjaga keseimbangan antara struktur global serta lokal, dan mengurangi efek over-smoothing sehingga meningkatkan stabilitas dan merepresentasi pada data. Selain itu juga pada penelitian terbaru dengan judul “Diffusion on PCA–UMAP Manifold: The Impact of Data Structure” [12] menunjukkan bahwa kombinasi PCA dan UMAP menghasilkan performa struktur lokal dan global dibandingkan metode reduksi tunggal seperti PCA, t-SNE, maupun UMAP saja. Hasil dari penelitian berhasil memproyeksikan data yang lebih stabil dan menghasilkan pemisahan kluster yang lebih jelas antar variable.

Penjelasan sebelumnya menjadikan GMM sebagai algoritma dalam klusterisasi dengan judul “Optimasi Algoritma Gaussian Mixture Model Menggunakan PCA – UMAP untuk Klusterisasi Tingkat Kesejahteraan Rakyat”. Oleh karena itu tujuan dari penelitian ini yaitu melakukan identifikasi terhadap tingkat kesejahteraan rakyat di berbagai wilayah baik kabupaten maupun kota di Provinsi Jawa Timur. Sehingga untuk mencapai tersebut dilakukan dengan melibatkan kombinasi antara PCA dan UMAP sebagai metode reduksi dimensi berlapis PCA untuk menyederhanakan data yang berdimensi global dan mengurangi *noise*, serta reduksi UMAP untuk mampu menjaga struktur data lokal dan data yang non-linear antar wilayah. Dengan pendekatan diharapkan mampu menghasilkan gambaran pada masing-masing wilayah baik Kabupaten atau Kota sesuai dengan kelompok yang sudah sesuai dengan karakteristik dari tiap wilayah.

1.2 Rumusan Masalah

Berdasarkan latar belakang merumuskan berbagai masalah yang meliputi sebagai berikut.

1. Berapa jumlah klaster yang optimal menggunakan GMM untuk klusterisasi tingkat kesejahteraan rakyat?
2. Apa yang menjadi pembeda antara metode reduksi dimensi dari kombinasi PCA dan UMAP dengan reduksi PCA sendiri pada algoritma GMM dalam mengklaster Tingkat kesejahteraan rakyat?
3. Apakah dengan melakukan optimasi GMM dengan melibatkan kombinasi reduksi dimensi PCA - UMAP menghasilkan nilai *silhouette score* tertinggi untuk mengklaster tingkat kesejahteraan rakyat?

1.3 Tujuan Penelitian

Dari rumusan masalah tersebut memberikan tujuan penelitian ini yang meliputi sebagai berikut.

1. Mengetahui jumlah klaster yang optimal dari hasil klusterisasi tingkat kesejahteraan rakyat menggunakan GMM
2. Analisis mendalam terhadap GMM dalam membandingkan hasil menggunakan PCA dan kombinasi PCA dan UMAP untuk mengklaster kesejahteraan rakyat

3. Analisis pada hasil optimasi GMM dengan melibatkan kombinasi reduksi dimensi PCA dan UMAP melalui evaluasi *silhouette score*

1.4 Manfaat Penelitian

Penelitian yang dilakukan diharapkan bisa memberikan manfaat sebagai berikut:

1. Bagi Penulis, penelitian ini memberikan informasi kombinasi reduksi PCA dan UMAP dengan melibatkan GMM dalam menganalisis data tingkat kesejahteraan rakyat melalui evaluasi *silhouette score*.
2. Bagi Masyarakat untuk memberikan informasi wilayah mana saja yang masih perlu dilakukan pemerataan dari segi indikator sosial ekonomi
3. Bagi Pembaca sebagai referensi informasi terhadap kombinasi PCA dan UMAP sebagai metode reduksi dimensi dengan algoritma GMM untuk klasterisasi tingkat kesejahteraan rakyat dalam pembangunan wilayah yang masih berada di kesenjangan.

1.5 Batasan Penelitian

Batasan masalah pada penelitian ini sebagai berikut:

1. Menggunakan data dari BPS Provinsi Jawa Timur pada rentang tahun 2010 hingga 2024 pada semua Kabupaten atau Kota pada indikator yang digunakan mencakup aspek sosial ekonomi.
2. Wilayah yang akan dilakukan untuk penelitian hanya melibatkan Kabupaten/Kota yang berada di Provinsi Jawa Timur. Hal ini bertujuan untuk menghasilkan klasterisasi tingkat kesejahteraan rakyat di Provinsi Jawa Timur saja
3. Algoritma *Gaussian Mixture Model* (GMM) dipilih untuk penelitian ini dalam mengklaster tingkat kesejahteraan rakyat
4. Menggunakan reduksi dimensi kombinasi PCA dan UMAP untuk meningkatkan kualitas dari algoritma GMM
5. Evaluasi yang dilakukan untuk mengetahui pemisahan klaster dengan menggunakan metrik *silhouette score* dan juga untuk menilai kualitas serta konsistensi klaster yang terbentuk.
6. Label klaster berdasarkan buku panduan “Perhitungan dan Analisis Kemiskinan Makro Indonesia 2016”