

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Sebagai negara kepulauan terbesar di dunia, Indonesia memiliki lebih dari 17.000 pulau serta wilayah laut sekitar 5,9 juta km<sup>2</sup>, sehingga menempatkannya pada posisi yang sangat strategis baik secara ekonomi maupun geopolitik [1]. Dengan jumlah penduduk mencapai sekitar 284 juta jiwa, Indonesia menempati peringkat keempat sebagai negara dengan penduduk terpadat di dunia [2]. Indonesia memiliki 38 provinsi di mana setiap provinsi, kabupaten, dan kota memiliki karakteristik yang berbeda, baik dalam aspek ekonomi, sosial, pendidikan, kesehatan, maupun infrastruktur. Perbedaan ini dapat menyebabkan ketimpangan tingkat pembangunan antardaerah yang hingga saat ini masih menjadi tantangan utama dalam mewujudkan pemerataan kesejahteraan masyarakat secara menyeluruh dan berkelanjutan [3]. Hal ini dapat menjadi hambatan dan tantangan besar dalam mewujudkan visi Indonesia Emas 2045, yang menekankan pertumbuhan ekonomi dan menghasilkan pembangunan nasional yang lebih adil dan merata.

Pemahaman terhadap kondisi ketimpangan tersebut memerlukan landasan konsep pembangunan sebagai dasar dalam menilai tingkat kesejahteraan suatu wilayah. Pembangunan merupakan proses perubahan yang bertujuan meningkatkan kualitas hidup manusia pada aspek ekonomi, sosial, dan lingkungan. Keberhasilan pembangunan tidak hanya ditentukan oleh pertumbuhan ekonomi, tetapi juga oleh peningkatan kualitas sumber daya manusia dan pemerataan manfaat pembangunan [4]. Secara nasional, pembangunan bertujuan untuk memperkuat daya saing dan pertumbuhan berbasis pemerataan wilayah. Tingkat pembangunan suatu wilayah dapat direpresentasikan melalui indikator sosial-ekonomi seperti Produk Domestik Regional Bruto (PDRB) per kapita, Indeks Pembangunan Manusia (IPM), serta Persentase Kemiskinan (P0) [5],[6].

Pengelompokan tingkat pembangunan daerah menjadi langkah penting untuk memahami ketimpangan secara sistematis dan berbasis data. Melalui

pengelompokan ini, wilayah dapat diklasifikasikan ke dalam beberapa kategori sehingga memudahkan identifikasi kondisi masing-masing daerah secara objektif [7]. Hasil pengelompokan dapat digunakan sebagai dasar dalam penyusunan kebijakan yang lebih tepat sasaran, khususnya dalam menentukan prioritas alokasi sumber daya dan intervensi pembangunan pada wilayah yang membutuhkan perhatian lebih. Selain itu, pengelompokan ini memberikan gambaran distribusi kesejahteraan antardaerah sehingga mendukung upaya pemerataan pembangunan. Bagi masyarakat, informasi ini meningkatkan transparansi kondisi daerah, mendorong partisipasi, serta menjadi pertimbangan dalam pengambilan keputusan. Dengan demikian, pengelompokan pembangunan berperan strategis dalam mendukung peningkatan kesejahteraan dan terwujudnya Indonesia Emas 2045.

Pulau Jawa dan Bali dipilih sebagai fokus penelitian karena memiliki peran strategis sebagai pusat pemerintahan sekaligus penggerak pertumbuhan ekonomi nasional, di mana pembangunan infrastruktur, perkembangan industri, dan sektor pariwisata menjadi pilar utama perekonomian di wilayah ini [8]. Kedua pulau ini juga merupakan wilayah dengan jumlah penduduk dan aktivitas ekonomi terbesar di Indonesia, sehingga dinamika pembangunannya sangat berpengaruh terhadap kondisi sosial ekonomi nasional. Namun, tingginya aktivitas ekonomi tersebut tidak selalu diikuti oleh pemerataan pembangunan. Ketimpangan ini tidak hanya terlihat dari satu indikator seperti Indeks Pembangunan Manusia (IPM), tetapi juga tercermin pada indikator lain seperti Produk Domestik Regional Bruto (PDRB) per kapita, Gini Rasio, Tingkat Pengangguran Terbuka (TPT), serta Persentase Kemiskinan (P0).

Selain itu, hubungan antar indikator sosial-ekonomi dalam mengukur tingkat pembangunan tidak selalu bersifat linear. Peningkatan pada satu variabel tidak secara langsung diikuti oleh perubahan yang pada variabel lainnya. Sebagai ilustrasi, Kota Jakarta Barat mengalami peningkatan IPM dari 81,38 pada tahun 2020 menjadi 81,76 pada tahun 2021, namun pada saat yang sama TPT mengalami penurunan signifikan dari 12,27 menjadi 9,06, sementara PDRB per kapita meningkat dari 192.975 menjadi 201.829. Pola serupa juga terlihat pada Kota Jakarta Selatan, IPM meningkat dari 84,72 menjadi 84,90, diikuti penurunan TPT

dari 10,79 menjadi 7,33, namun Gini Rasio mengalami perubahan dari 0,43 menjadi 0,38. Variasi antar indikator tersebut menunjukkan bahwa hubungan antar variabel pembangunan bersifat kompleks dan tidak dapat dijelaskan hanya melalui satu indikator. Kondisi ini mencerminkan bahwa pembangunan bersifat multidimensional dan melibatkan interaksi antar berbagai aspek sosial-ekonomi.

Berbagai penelitian terdahulu telah mengkaji ketimpangan pembangunan dengan menggunakan indikator sosial-ekonomi yang beragam. Penelitian oleh A. Rosyid et al. [9] menunjukkan bahwa variabel pendidikan, kesehatan, dan pengangguran memiliki pengaruh terhadap Indeks Pembangunan Manusia (IPM), sehingga mencerminkan kualitas sumber daya manusia dalam suatu wilayah. Sementara itu, penelitian oleh Hulu dan Wahyuni [10] menggunakan variabel pembangunan infrastruktur dan pertumbuhan ekonomi untuk menganalisis ketimpangan pendapatan, yang berkaitan erat dengan distribusi kesejahteraan antarwilayah. Selain itu, penelitian oleh Widiyasari et al. [11] memanfaatkan variabel ketenagakerjaan untuk melihat produktivitas tenaga kerja, yang berkontribusi terhadap kondisi ekonomi dan kesejahteraan masyarakat.

Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa indikator sosial-ekonomi seperti kualitas sumber daya manusia, kondisi ekonomi, ketenagakerjaan, dan pemerataan pendapatan merupakan faktor penting dalam merepresentasikan tingkat pembangunan suatu wilayah. Oleh karena itu, penelitian ini menggunakan lima variabel utama, yaitu Indeks Pembangunan Manusia (IPM), Produk Domestik Regional Bruto (PDRB) per kapita, Gini Rasio, Tingkat Pengangguran Terbuka (TPT), dan Persentase Kemiskinan (P0). Kelima variabel tersebut dipilih karena mampu merepresentasikan dimensi utama pembangunan, yaitu kualitas manusia, pertumbuhan ekonomi, ketimpangan pendapatan, kondisi ketenagakerjaan, serta tingkat kesejahteraan masyarakat secara menyeluruh.

Indikator sosial-ekonomi tersebut mampu menggambarkan kondisi pembangunan suatu wilayah. Namun, penggunaan berbagai indikator secara bersamaan tidak selalu mudah untuk diinterpretasikan karena masing-masing indikator memiliki karakteristik dan arah pengaruh yang berbeda. Perbedaan nilai antarindikator sering kali menghasilkan gambaran yang tidak seragam, sehingga

menyulitkan dalam menentukan posisi relatif suatu wilayah dalam tingkat pembangunan. Dalam kondisi tersebut, diperlukan suatu pendekatan analisis yang mampu mengelompokkan wilayah berdasarkan kemiripan karakteristik pembangunan. Analisis kluster digunakan sebagai metode untuk mengelompokkan wilayah ke dalam kategori tertentu, seperti maju, berkembang, dan tertinggal, sehingga informasi yang kompleks dapat disederhanakan menjadi lebih terstruktur dan mudah dipahami serta memberikan dasar dalam perumusan kebijakan pembangunan yang lebih tepat sasaran dan berbasis data.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Pulau Jawa dan Bali berdasarkan tingkat pembangunan sosial-ekonomi menggunakan pendekatan klusterisasi dengan memanfaatkan indikator Indeks Pembangunan Manusia (IPM), Produk Domestik Regional Bruto (PDRB) per kapita, Gini Rasio, Tingkat Pengangguran Terbuka (TPT), serta Persentase Kemiskinan (P0). Selain itu, penelitian ini juga digunakan untuk mengidentifikasi pola ketimpangan antarwilayah serta menghasilkan pemetaan kategori pembangunan yang lebih terarah. Hasil dari penelitian ini diharapkan dapat memberikan gambaran kondisi pembangunan daerah secara lebih jelas serta menjadi dasar dalam mendukung perumusan kebijakan pembangunan yang lebih efektif dan berbasis data.

Penelitian mengenai klusterisasi wilayah di Indonesia menunjukkan metode tradisional seperti K-Means dan *Hierarchical Clustering* banyak digunakan untuk menganalisis kondisi sosial-ekonomi dan pembangunan daerah. Studi sebelumnya melakukan klusterisasi pada pertumbuhan ekonomi, tenaga kerja, kemiskinan, kesetaraan gender, kesehatan, pendidikan, akses air dan kebersihan, serta perlindungan sosial dengan membandingkan AGNES (*Agglomerative Hierarchical Clustering*) dan K-Means menemukan bahwa AGNES lebih efisien dengan nilai *silhouette* lebih tinggi [12]. Penelitian lain menggunakan K-Means dalam mengelompokkan kabupaten dan kota di Papua menggunakan P0, rata-rata lama sekolah, IPM, indeks kedalaman dan keparahan kemiskinan, Tingkat Pengangguran Terbuka (TPT), serta PDRB, menghasilkan dua kluster tingkat kemiskinan tinggi dan rendah [13]. Selain itu, metode *K-Medoids* juga telah digunakan untuk

menganalisis kesenjangan pendidikan di seluruh Indonesia berdasarkan tingkat penyelesaian pendidikan di jenjang SD – SMA menemukan bahwa wilayah bagian tengah serta timur Indonesia memiliki kesenjangan pendidikan lebih tinggi [14].

Penelitian lain menunjukkan bahwa kombinasi *Autoencoder* dan K-Means dapat meningkatkan kualitas klasterisasi dan mampu menghasilkan klaster yang lebih optimal dibanding K-Means tunggal [15]. Pendekatan lain menerapkan PCA dengan *Hierarchical Clustering* untuk mengelompokkan tingkat kesejahteraan sosial-ekonomi berdasarkan persentase penduduk miskin, PDRB regional, harapan hidup, rata-rata lama sekolah, pengeluaran per kapita, dan tingkat pengangguran, dan berhasil membentuk empat klaster yang representatif dan terstruktur [16]. Hasil-hasil tersebut menunjukkan bahwa metode klasterisasi tradisional seperti K-Means, Hierarchical Clustering, dan K-Medoids dapat digunakan dalam mengidentifikasi pola pada data sosial-ekonomi. Namun, metode tersebut memiliki keterbatasan dalam menangkap hubungan non-linear antar variabel serta sensitivitas terhadap dimensi data yang kompleks. Selain itu, sebagian besar penelitian masih menggunakan variabel asli tanpa transformasi representasi yang lebih mendalam, sehingga pola tersembunyi dalam data belum sepenuhnya terungkap. Di sisi lain, metode *clustering* dalam penelitian sebelumnya hanya diterapkan pada wilayah tertentu seperti Pulau Jawa, Papua, atau konteks pendidikan, sehingga belum memberikan gambaran pembangunan daerah secara menyeluruh.

Penelitian ini menggunakan pendekatan berbasis *machine learning* dan *deep learning* dengan mengombinasikan *Variational Autoencoder* (VAE) dan *Improved K-Means*. Penggunaan *Improved K-Means* tunggal belum cukup optimal apabila data yang digunakan masih berada pada representasi asli, karena hubungan antar variabel sosial-ekonomi cenderung bersifat non-linear dan kompleks di mana peningkatan salah satu variabel belum tentu meningkatkan variabel lainnya. Selain itu, performa metode clustering seperti *Improved K-Means* cenderung meningkat ketika diterapkan pada representasi data yang lebih representatif. Oleh karena itu, VAE digunakan terlebih dahulu untuk membentuk representasi laten yang lebih informatif serta memperkaya struktur data melalui pembentukan representasi data

yang lebih kompleks. Selanjutnya, *Improved K-Means* diterapkan pada ruang laten tersebut untuk meningkatkan kualitas klasterisasi melalui pembobotan jarak dan optimasi centroid. Kombinasi kedua metode ini memungkinkan proses clustering tidak hanya bekerja pada data mentah, tetapi pada representasi data yang telah diperkaya, sehingga menghasilkan pengelompokan yang lebih akurat dan stabil dibandingkan penggunaan *Improved K-Means* tunggal.

Berbeda dengan beberapa pendekatan sebelumnya yang menggunakan *Principal Component Analysis* (PCA) sebagai metode reduksi dimensi berbasis linear, penelitian ini menggunakan *Variational Autoencoder* (VAE) yang mampu menangkap hubungan non-linear serta membentuk representasi laten yang lebih kompleks. Selain berfungsi sebagai metode pembelajaran representasi, VAE juga memungkinkan terjadinya ekspansi dimensi pada ruang laten sehingga pola tersembunyi dalam data dapat direpresentasikan dengan lebih baik. VAE tidak hanya mereduksi, tetapi juga mempelajari distribusi data sehingga mampu merepresentasikan struktur kompleks yang tidak dapat ditangkap oleh metode linear seperti PCA. Visualisasi hasil klasterisasi dilakukan menggunakan kombinasi *Principal Component Analysis* (PCA) dan *Multidimensional Scaling* (MDS). PCA digunakan untuk mereduksi dimensi dengan mempertahankan variansi terbesar dalam data, sedangkan MDS digunakan untuk mempertahankan struktur jarak antar data sehingga pola klaster dapat terlihat lebih jelas. Kombinasi kedua metode ini memberikan visualisasi yang lebih representatif dibandingkan penggunaan satu metode saja.

Data yang digunakan melalui tahapan pre-processing, ekspansi dimensi menggunakan VAE, serta proses klasterisasi dengan *Improved K-Means* untuk memperoleh hasil yang optimal. Hasil pengelompokan kemudian divisualisasikan melalui Graphical User Interface (GUI) yang dirancang untuk memungkinkan pengguna melakukan eksplorasi data secara interaktif, seperti memilih wilayah, membandingkan hasil klaster antar daerah, serta melihat perubahan kondisi pembangunan. Selain itu, GUI menyajikan informasi secara komprehensif dengan menampilkan hasil klasterisasi beserta nilai indikator sosial-ekonomi yang digunakan, sehingga memberikan gambaran yang lebih menyeluruh mengenai

kondisi pembangunan antarwilayah. Pembaruan utama penelitian ini terletak pada penerapan *deep clustering* berbasis VAE untuk membentuk representasi laten yang lebih informatif, kemudian dikombinasikan dengan *Improved K-Means* untuk meningkatkan ketepatan hasil klasterisasi dan ketajaman penilaian status pembangunan. Dengan demikian, penelitian ini mengisi *research gap* dari studi sebelumnya yang hanya menggunakan metode klasterisasi tradisional berbasis variabel asli, yang kurang mampu menangkap hubungan non-linear antarindikator dan memberikan wawasan yang lebih mendalam bagi perumusan kebijakan pembangunan wilayah Jawa–Bali.

## 1.2. Rumusan Masalah

Mengingat pemaparan dalam bagian latar belakang, penelitian ini dirancang untuk menjawab permasalahan utama sebagai berikut:

1. Bagaimana implementasi metode *Variational Autoencoder* (VAE) dan *Improved K-Means clustering* dalam melakukan pengelompokan status pengembangan wilayah Kabupaten/Kota maju, berkembang, dan tertinggal di Jawa Bali tahun 2020 – 2024?
2. Bagaimana kondisi ketidakseimbangan pembangunan antar kabupaten kota di wilayah Jawa dan Bali jika dianalisis menggunakan indikator sosial-ekonomi multidimensi tahun 2020 – 2024?
3. Bagaimana hasil evaluasi cluster yang dihasilkan menggunakan *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski Harunk* melihat seberapa baik kualitas cluster yang terbentuk?
4. Bagaimana implementasi hasil analisis klaster pembangunan ke dalam antarmuka *Graphical User Interface* (GUI) menggunakan *Streamlit*?

## 1.3. Batasan Masalah

Agar batas penelitian menjadi lebih jelas dan pembahasan lebih mendalam, ditetapkan beberapa batasan masalah sebagai berikut:

1. Fokus analisis adalah data pembangunan sosial-ekonomi di semua kabupaten kota yang berada di wilayah Pulau Jawa dan Bali.

2. Sumber data utama yang digunakan merupakan data publikasi resmi BPS selama 5 tahun dari tahun 2020 hingga tahun 2024.
3. Variabel yang digunakan dalam penelitian ini mencakup indikator sosial dan ekonomi, yakni PDRB per kapita, rasio gini, tingkat pengangguran terbuka (TPT), indeks pembangunan manusia (IPM), serta persentase penduduk miskin (P0)
4. Metode yang digunakan adalah metode gabungan antara *Variational Autoencoder* (VAE) untuk melakukan proses ekspansi dimensi dan *Improved K-Means clustering* untuk melakukan pengelompokan data.
5. Kategori status pembangunan (Maju, Berkembang, dan Tertinggal) dalam penelitian ini merupakan hasil klasifikasi berbasis metode clustering dan skoring yang dikembangkan penulis, serta tidak merepresentasikan penetapan resmi dari pemerintah atau lembaga tertentu.

#### **1.4. Tujuan Penelitian**

Merujuk pada rumusan masalah yang telah dipaparkan sebelumnya, tujuan penelitian ini dapat dijabarkan sebagai berikut:

1. Menganalisis kondisi ketimpangan pembangunan antar Kabupaten/Kota di Pulau Jawa dan Bali berdasarkan indikator sosial-ekonomi multidimensi yang meliputi PDRB per kapita, Gini Rasio, IPM, TPT, dan P0.
2. Menerapkan dan mengevaluasi kinerja metode *Variational Autoencoder* (VAE) untuk mengekspansi dimensi data serta *Improved K-Means clustering* untuk melakukan pengelompokan Kabupaten/Kota ke dalam kategori maju dan berkembang berdasarkan tingkat pembangunannya.
3. Menghasilkan klaster pembangunan wilayah Jawa dan Bali yang lebih objektif dan adaptif terhadap pola non-linear antar variabel, serta memberikan informasi yang dapat dijadikan dasar dalam melakukan perumusan strategi dan program Pembangunan yang fokus pada kebutuhan dan karakteristik daerah.

### 1.5. Manfaat Penelitian

Penelitian ini diharapkan mampu berkontribusi serta memberikan manfaat yang dapat dijelaskan sebagai berikut:

#### 1. Manfaat Teoretis

Penelitian ini diharapkan bisa menjadi literatur mengenai penggunaan kombinasi metode VAE dan *Improved K-Means clustering* dalam analisis sosial-ekonomi, khususnya dalam pemetaan pembangunan antarwilayah di Indonesia, serta memberikan kontribusi dalam pengembangan metode analisis data multidimensi yang dapat menangkap pola non-linear antar variabel secara lebih baik.

#### 2. Manfaat Praktis

- a. Bagi Pemerintah, temuan penelitian ini diharapkan dapat memberikan Gambaran kondisi wilayah, informasi tambahan dan masukan dalam merumuskan kebijakan pembangunan yang tepat sasaran, terutama dalam menentukan wilayah yang menjadi prioritas pembangunan.
- b. Bagi Masyarakat, penelitian ini diharapkan dapat menginformasikan secara lebih jelas tentang posisi daerah dalam konteks pembangunan nasional, yang dapat mendorong partisipasi dalam membangun dan meningkatkan kualitas hidup.
- c. Bagi Penulis, penelitian ini menjadi sarana untuk mengasah kemampuan dalam menggunakan metode *Deep Learning* untuk analisis sosial-ekonomi dan menambah pengetahuan mengenai isu ketimpangan pembangunan di Indonesia.

*Halaman ini sengaja dikosongkan*