

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada era modern ini, pemahaman yang mendalam terhadap data populasi suatu wilayah menjadi sangat penting untuk pengembangan kebijakan dan perencanaan pembangunan suatu wilayah (Marpaung dan Siahaan, 2021). Provinsi Jawa Timur, yang terletak di bagian timur Pulau Jawa, Indonesia, adalah provinsi dengan jumlah penduduk terbesar kedua di Indonesia (Wahyuni dkk, 2021). Provinsi Jawa Timur merupakan salah satu wilayah yang terus mengalami perkembangan penduduk yang signifikan dari tahun 2021 hingga 2023. Fenomena ini menjadi sorotan karena adanya kebutuhan untuk memahami dan mengelola distribusi penduduk secara efektif (Asyfani dkk., 2024). Dengan lebih dari 40 juta jiwa, Jawa Timur menjadi rumah bagi berbagai kelompok etnis, budaya, dan ekonomi.

Pertumbuhan penduduk di Jawa Timur menunjukkan pola yang unik dan beragam, di mana beberapa wilayah mengalami pertumbuhan yang pesat, sementara yang lain mengalami pertumbuhan yang lambat atau bahkan penurunan jumlah penduduk. Ketidakmerataan ini menciptakan tantangan signifikan dalam pengelolaan sumber daya dan pembangunan infrastruktur. Faktor-faktor seperti urbanisasi yang cepat, migrasi dari daerah pedesaan ke perkotaan, serta perubahan sosial-ekonomi berkontribusi terhadap dinamika ini. Urbanisasi yang pesat di kota-kota besar menyebabkan tekanan pada infrastruktur dan layanan publik, seperti transportasi, perumahan, dan kesehatan, yang sering kali tidak dapat memenuhi kebutuhan penduduk yang terus meningkat. Di sisi lain, daerah yang mengalami penurunan jumlah penduduk mungkin menghadapi masalah ekonomi dan kurangnya investasi, yang dapat memperburuk ketidakmerataan pembangunan. (Sinuhaji dan Ginting, 2023). Oleh karena itu, diperlukan suatu pendekatan analisis yang dapat mengidentifikasi pola dan karakteristik dalam distribusi jumlah penduduk.

Meskipun data jumlah penduduk tersedia, masih terdapat kendala dalam memahami pola sebaran penduduk pada tingkat kota di Provinsi Jawa Timur. Ketidakmampuan dalam mengidentifikasi kelompok atau pola tertentu dapat menghambat upaya perencanaan pembangunan yang efisien. Oleh karena itu, perlu dilakukan analisis lebih lanjut untuk mengatasi masalah ini. Maka digunakan algoritma *K-Means* sebagai metode analisis untuk

mengklasternisasi jumlah penduduk pada setiap kota di Provinsi Jawa Timur. *K-Means* merupakan algoritma klasternisasi yang dapat mengelompokkan data ke dalam beberapa kelompok berdasarkan kesamaan karakteristik tertentu (Nofiar dan Defit, 2019). Penggunaan algoritma ini diharapkan dapat memberikan gambaran yang jelas mengenai pola sebaran penduduk di wilayah tersebut.

K-Means bekerja dengan mengelompokkan data ke dalam k kelompok (klaster) yang ditentukan sebelumnya, dengan meminimalkan variasi atau jarak antara titik data dalam satu klaster. Pendekatan ini sangat relevan dalam konteks analisis sebaran penduduk, karena dapat membantu mengidentifikasi pola sebaran yang mungkin tidak terlihat secara langsung (Talukua dkk., 2017). Dengan menggunakan *K-Means*, dapat mengelompokkan kota-kota berdasarkan tingkat persebaran jumlah penduduk dari rentang tahun 2021 hingga tahun 2023. Selain itu, untuk menentukan jumlah klaster yang optimal, penelitian ini akan menerapkan metode siku (*Elbow Method*). Metode ini melibatkan pengujian berbagai jumlah klaster dan mengamati tingkat variasi yang dijelaskan oleh model terhadap jumlah klaster tersebut.

Metode *Elbow* atau *Elbow Method* merupakan suatu teknik yang dimanfaatkan untuk mengidentifikasi jumlah klaster yang optimal dalam proses klasternisasi, terutama diterapkan pada algoritma *K-Means*. Konsep inti dari metode ini adalah dengan mengamati penurunan variansi di dalam setiap klaster seiring dengan variasi jumlah klaster yang berbeda (Syahfitri dkk., 2023). Metode *Elbow* bekerja dengan mengukur variansi atau dispersi data dalam setiap klaster saat jumlah klaster berubah-ubah. Proses ini melibatkan iterasi dengan berbagai jumlah klaster, dan pada setiap iterasi, variansi dihitung untuk setiap klaster. Selanjutnya dianalisis untuk mengidentifikasi titik dimana penurunan variansi menjadi kurang signifikan, menandakan bahwa penambahan klaster tidak lagi memberikan keuntungan substansial dalam mengurangi dispersi data (Fahrozi dkk., 2023). Dengan menggunakan metode *Elbow*, dapat diambil keputusan yang lebih informasional dan berbasis data dalam menentukan jumlah klaster optimal untuk suatu *dataset* tertentu. Teknik ini menjadi kunci dalam memahami struktur data dan mengoptimalkan kinerja algoritma klasternisasi, khususnya algoritma *K-Means*, untuk mencapai segmentasi yang optimal dan lebih bermakna. Penelitian terkait dengan penggunaan *K-Means* antara lain “Penggunaan *K-Means* untuk klasternisasi Penetapan Instruktur Diklat Pada PT PLN (Persero) Udiklat Jakarta” (Budiana dkk., 2019), “penggunaan *K-Means* Klasternisasi untuk klasternisasi tingkat kehadiran dosen” (Virgo dkk., 2020), dan “penggunaan *K-Means* untuk

klasterisasi sayuran unggulan” (Harahap dkk., 2022). Namun pada penelitian ini tidak digunakan metode lain untuk menentukan jumlah klaster atau k sehingga klaster yang dihasilkan kurang begitu optimal.

Penelitian lainnya yang berkaitan dengan penggunaan *K-Means* digunakan untuk klasterisasi siswa yang Berprestasi (Dewi dkk., 2022). Pada penelitian ini, jumlah klaster ditentukan menggunakan metode *Davies Bouldin Index* yang memungkinkan penentuan jumlah klaster yang optimal. Penelitian lainnya terkait penggunaan metode *Elbow* yaitu metode ini digunakan untuk klasterisasi pemerataan bantuan sosial di Kabupaten Bojonegoro (Fitriyah dkk., 2023) dan algoritma *K-Means* Klasterisasi Metode *Elbow* digunakan untuk menganalisa Motivasi Pengunjung Festival Halal JHF (Wicaksana., 2023).

Tujuan utama dari penelitian ini adalah untuk menganalisis dan mengklasterisasi sebaran jumlah penduduk pada setiap kota di Provinsi Jawa Timur menggunakan algoritma *K-Means*. Metode *K-Means* akan diaplikasikan dengan menggunakan metode *Elbow* untuk menentukan jumlah klaster yang optimal. Maka dari itu, penelitian ini bertujuan untuk mendapatkan informasi yang lebih rinci dan akurat mengenai pola sebaran penduduk di tingkat kota. Diharapkan penelitian ini bisa berkontribusi terhadap pengembangan bidang analisis demografis dan pengelolaan populasi serta menjadi landasan yang kuat bagi perumusan kebijakan pembangunan yang lebih efektif di Provinsi Jawa Timur.

1.2 Rumusan Masalah

1. Bagaimana mengklasterisasi jumlah penduduk di Provinsi Jawa Timur Tahun 2021 – 2023 menggunakan algoritma *K-Means*?
2. Bagaimana hasil interpretasi karakteristik pertumbuhan penduduk pada setiap kota yang terdapat dalam setiap klaster yang terbentuk dari klasterisasi oleh algoritma *K-Means*?

1.3 Tujuan Penelitian

1. Mengklasterisasi jumlah penduduk di Provinsi Jawa Timur menggunakan algoritma *K-Means* untuk mendapatkan informasi yang lebih rinci dan akurat mengenai pola sebaran penduduk.
2. Menentukan profil pertumbuhan penduduk pada setiap kota yang terdapat dalam setiap klaster yang terbentuk dari klasterisasi menggunakan algoritma *K-Means*.

1.4 Batasan Penelitian

1. Penelitian ini berfokus pada analisis laju sebaran jumlah penduduk di kota-kota Provinsi Jawa Timur yang diambil dari website <https://kedirikota.bps.go.id/indicator/12/358/1/jumlah-penduduk-provinsi-jawa-timur.html>
2. Menggunakan data populasi kota di Provinsi Jawa Timur dari tahun 2021 hingga 2023.
3. *Tools* yang digunakan untuk analisa yaitu RapidMiner, dan *Python*

1.5 Manfaat Penelitian

1. Memberikan informasi yang lebih rinci dan akurat mengenai pola sebaran penduduk di tingkat kota.
2. Kontribusi pada bidang analisis demografis dan pengelolaan populasi.

