

# BAB I PENDAHULUAN

## 1.1. Latar Belakang

Angkatan kerja merupakan penduduk berusia 15 tahun ke atas yang aktif dalam kegiatan ekonomi, baik yang sedang bekerja, memiliki pekerjaan tetapi untuk sementara tidak bekerja, maupun yang tidak memiliki pekerjaan tetapi aktif mencari kerja [1]. Angkatan kerja dapat dikategorikan berdasarkan usia yang memiliki keterkaitan erat dengan permasalahan kependudukan dan ketenagakerjaan [1]. Kelompok usia produktif yaitu usia 15-64 tahun berperan penting dalam pertumbuhan ekonomi yang menggambarkan potensi sejumlah orang yang siap dan mampu bekerja di suatu wilayah. Dalam proses pembangunan nasional, Indonesia menghadapi tantangan meningkatnya jumlah angkatan kerja yang tidak sebanding dengan ketersediaan lapangan kerja [2].

Selain itu, tingkat pendidikan berperan penting bagi angkatan kerja dalam menentukan kompetensi dan daya saing untuk memperoleh pekerjaan yang sesuai [3]. Pendidikan yang lebih tinggi cenderung membuka akses ke pekerjaan dengan upah lebih baik dan prospek karier yang lebih stabil, sedangkan rendahnya tingkat pendidikan sering dikaitkan dengan pekerjaan informal, upah rendah, dan risiko pengangguran yang lebih tinggi. Ketidaksesuaian antara bidang pendidikan dan pekerjaan (*job mismatch*) juga menjadi tantangan dalam dunia kerja, yang dapat berdampak pada produktivitas tenaga kerja dan pertumbuhan ekonomi secara keseluruhan [4].

Ketersediaan lapangan pekerjaan di setiap provinsi Indonesia adalah faktor penting yang mempengaruhi tingkat ketenagakerjaan dan pola migrasi angkatan kerja, selain usia dan tingkat pendidikan [5]. Wilayah dengan sektor industri dan jasa yang berkembang cenderung menyerap tenaga kerja lebih banyak dibandingkan daerah dengan dominasi sektor pertanian atau yang memiliki keterbatasan investasi sehingga membuat banyaknya angkatan kerja di Indonesia merantau untuk mendapatkan lapangan pekerjaan serta gaji yang lebih sesuai dan lebih baik [5].

Permasalahan utama dalam angkatan kerja adalah ketidaksesuaian kompetensi tenaga kerja dengan kebutuhan pasar, yang mengakibatkan ketimpangan distribusi angkatan kerja di setiap provinsi [1]. Kondisi ketenagakerjaan yang berbeda-beda mengharuskan kebijakan tenaga kerja yang diterapkan disesuaikan dengan kebutuhan spesifik tiap daerah. Selain itu, banyaknya pencari kerja juga mengalami kendala dalam memenuhi persyaratan pendidikan yang ditetapkan oleh industri, sementara di sisi lain, daerah tertentu belum mampu menyediakan lapangan kerja yang cukup bagi tenaga kerja yang tersedia [6].

Daya saing tenaga kerja juga menjadi salah satu tantangan dalam pembangunan nasional. Salah satu fokus utama dalam Rencana Pembangunan Jangka Menengah 2025-2029 yaitu peningkatan daya saing sumber daya manusia. Hal ini juga ditekankan dalam Peraturan Presiden Nomor 12 Tahun 2025 tentang rencana pembangunan yang menyebutkan bahwa rendahnya kualifikasi pendidikan tenaga kerja menjadi tantangan dalam meningkatkan daya saing sumber daya manusia [7]. Rendahnya kualitas tersebut juga tercermin dari capaian rata-rata nilai *Programme for International Student Assessment (PISA)* siswa Indonesia untuk semua aspek yaitu membaca, matematika, dan sains pada tahun 2022. Kondisi ini berdampak langsung pada rendahnya kesiapan tenaga kerja yang kompeten di pasar kerja. Selain itu, berdasarkan hasil *The IMD World Talent Ranking*, Indonesia berada di peringkat 50 pada tahun 2021 dan menurun ke peringkat 52 pada tahun 2022 dari 64 negara yang disurvei terkait kesiapan tenaga kerja dalam menghadapi persaingan ekonomi dengan negara lain [8]. Kondisi tersebut menunjukkan bahwa Indonesia memiliki tantangan dalam menghadapi persaingan ekonomi dan kompetensi tenaga kerja yang menjadi perhatian bagi para pemangku kepentingan dan pemerintah untuk terus meningkatkan kesiapan tenaga kerja agar dapat bersaing di pasar global.

Oleh karena itu, analisis terhadap variabel usia, pendidikan, dan lapangan pekerjaan menjadi penting dalam memahami pola ketenagakerjaan yang memiliki keterkaitan erat dengan indikator ketenagakerjaan. Indikator ketenagakerjaan ini digunakan sebagai dasar dalam mengukur kinerja pasar tenaga kerja dan membantu pemerintah dalam merumuskan kebijakan strategis untuk mengoptimalkan angkatan kerja di setiap provinsi di Indonesia [9].

Berdasarkan variabel yang digunakan, usia berperan dalam menentukan tingkat partisipasi kerja dengan kelompok usia produktif (15–64 tahun) yang cenderung lebih aktif berpartisipasi dibanding kelompok usia non-produktif [10]. Tingkat pendidikan memengaruhi status pekerjaan dimana individu dengan pendidikan tinggi umumnya lebih berpeluang bekerja di sektor formal dengan penghasilan yang lebih baik dibandingkan mereka yang berpendidikan rendah. Di sisi lain, ketersediaan lapangan kerja di suatu wilayah turut menentukan tingkat pengangguran serta mobilitas tenaga kerja, khususnya di daerah yang minim industri dan investasi. Dalam menghadapi tantangan ini, analisis kluster dapat berperan penting dalam memahami karakteristik angkatan kerja berdasarkan kelompok usia, pendidikan, dan sektor pekerjaan yang dapat membantu pemangku kebijakan memperoleh wawasan yang lebih mendalam untuk memahami kesiapan angkatan kerja setiap wilayah provinsi dalam menghadapi persaingan bebas di pasar tenaga kerja dengan mengidentifikasi dominasi tingkat pendidikan dan sektor pekerjaan dalam masing-masing kelompok wilayah.

Metode analisis kluster digunakan untuk mengelompokkan data ke dalam kluster-kluster yang terbentuk dari kesamaan karakteristik [11]. Ciri-ciri pengelompokan yang baik adalah memiliki tingkat kesamaan (homogenitas) yang tinggi di antara anggota dalam satu kluster, serta memiliki tingkat perbedaan (heterogenitas) yang tinggi antar kluster [12]. *Clustering* dapat dibagi menjadi dua kategori berdasarkan tingkat kemiripan, yaitu metode hierarki (*hierarchical clustering methods*) dan metode non-hierarki (*non-hierarchical clustering methods*). Hasil dari clustering dengan metode hierarki umumnya menghasilkan diagram pohon (*tree diagram*) atau dendogram, yang menggambarkan pengelompokan objek berdasarkan jarak antar objek [13].

Metode hierarki yang paling umum digunakan adalah metode agglomeratif, atau metode penggabungan. Dalam prosesnya, metode agglomeratif dimulai dengan menganggap setiap objek sebagai kluster terpisah, kemudian secara bertahap mengelompokkan objek-objek tersebut ke dalam kluster yang semakin besar. Dalam metode agglomeratif, terdapat beberapa metode yang digunakan, yaitu metode *linkage*, *variance*, dan *centroid* [14]. Metode *linkage* digunakan untuk menentukan jarak antar kluster menggunakan beberapa teknik, seperti *single*

*linkage*, *complete linkage*, dan *average linkage*. Metode *variance* bertujuan untuk membentuk kluster dengan meminimalkan varians dalam kluster tersebut dan salah satu teknik yang sering digunakan adalah metode Ward. Metode *centroid* digunakan untuk mengukur jarak antara dua kluster berdasarkan rata-rata dari seluruh variabel yang ada dalam masing-masing kluster. Setiap metode aglomeratif melakukan proses klusterisasi yang berbeda sehingga setiap metode memiliki kelebihan dan kekurangan masing-masing [15]. Namun, metode aglomeratif memiliki kelemahan dalam menangani pencilan yang dapat memengaruhi hasil klusterisasi [16].

Dan pencilan menjadi salah satu tantangan utama dalam pengelompokan data ketenagakerjaan, seperti perbedaan jumlah angkatan kerja yang sangat tinggi di Pulau Jawa dibandingkan dengan daerah berpenduduk kecil seperti Kalimantan. Kepadatan penduduk yang tinggi ada di Pulau Jawa menyebabkan pertumbuhan angkatan kerja juga semakin besar. Pertumbuhan angkatan kerja ini perlu diimbangi dengan tersedianya lapangan pekerjaan yang memadai agar tidak meningkatkan jumlah pengangguran. Hingga saat ini, Pulau Jawa masih mendominasi ketenagakerjaan di Indonesia, yaitu sekitar 63,38 persen, terutama di provinsi Jawa Timur, Jawa Tengah, dan Jawa Barat. Provinsi-provinsi ini dikenal memiliki jumlah penduduk yang besar, kondisi wilayah yang menyediakan banyak lapangan pekerjaan, serta tingkat pendidikan yang relatif tinggi [17].

Temuan ini sejalan dengan studi yang dilakukan oleh Amalia dan rekan-rekan pada tahun 2022 yang meneliti pengaruh variabel ketenagakerjaan terhadap produktivitas tenaga kerja Indonesia. Penelitian tersebut menggunakan variabel produktivitas tenaga kerja, tingkat pengangguran terbuka, persentase angkatan kerja dengan pendidikan dasar ke bawah, dan rata-rata upah buruh (rupiah). Berdasarkan hasil penelitian, variabel rata-rata upah buruh menunjukkan bahwa DKI Jakarta dan Banten termasuk pencilan. Hal ini disebabkan oleh keduanya yang berperan sebagai pusat perekonomian Indonesia, sehingga menarik tenaga kerja dari berbagai wilayah dengan kompensasi upah lebih tinggi namun juga beban kerja yang lebih berat [18].

Perbedaan yang dianggap sebagai pencilan ini dapat menyebabkan ketidakseimbangan dalam metode klusterisasi tradisional yang tidak memiliki

mekanisme deteksi pencilan yang efektif. Oleh karena itu, metode HDBSCAN menjadi pilihan yang lebih tepat karena kemampuannya dalam mengelompokkan data berdasarkan kepadatan serta secara otomatis mengidentifikasi dan menangani pencilan sebagai *noise*, sehingga menghasilkan kluster yang lebih akurat.

Pencilan terjadi ketika terdapat nilai ekstrem yang berjarak jauh dari sebagian besar data dalam sampel atau populasi [19]. *Outlier* atau pencilan dalam klusterisasi dapat menimbulkan ketidaknormalan karena berbeda dari pola umum pada kluster utama, sehingga diperlukan penanganan tersendiri untuk menjaga validitas hasil kluster [20]. Dalam upaya mengatasi masalah pencilan pada data, metode HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) hadir sebagai pengembangan dari klusterisasi agglomeratif, dengan keunggulan mampu membentuk kluster berbasis kepadatan dan secara otomatis mengidentifikasi pencilan (*outlier*) sebagai *noise* [21].

HDBSCAN merupakan modifikasi dari algoritma OPTICS yang hanya menggunakan parameter input tunggal yaitu *Minpts* (*Minimum Points*). Algoritma ini tidak hanya mengelompokkan data berdasarkan kepadatan tinggi, tetapi juga secara otomatis mengidentifikasi pencilan sebagai data yang tidak masuk dalam kluster. HDBSCAN menggunakan metrik *mutual reachability distance* untuk menentukan titik data berdasarkan kepadatan dan membentuk struktur kluster hierarki. Titik yang tidak memenuhi ambang batas kepadatan akan diklasifikasikan sebagai pencilan, sementara kluster yang terpecah dengan jumlah titik di bawah batas minimum akan dianggap sebagai *noise* [22].

Penelitian sebelumnya oleh Ghaly Ghiffary, dkk bertujuan membandingkan algoritma HDBSCAN dengan *Agglomerative Hierarchical Clustering* dalam klusterisasi data yang mengandung pencilan [21]. Penelitian ini menggunakan data Produk Domestik Regional Bruto Atas Dasar Harga Konstan (PDRB ADHK) Provinsi Jawa Timur tahun 2023 dengan 17 variabel dan 234 observasi. Metode agglomeratif yang digunakan mencakup *Ward*, *Average Linkage*, dan *Complete Linkage* untuk dibandingkan dengan HDBSCAN dalam menangani pencilan. Hasil evaluasi menunjukkan bahwa HDBSCAN memiliki performa terbaik dalam mengelompokkan data PDRB yang kompleks, terutama dalam wilayah dengan distribusi yang tidak normal berdasarkan *Silhouette Coefficient* sebesar 0.624 serta

mendeteksi dan mengelola pencilan secara efektif. Sebaliknya, metode *Ward*, *Average Linkage*, dan *Complete Linkage* hanya mencapai *Silhouette Coefficient* sebesar 0.4 yang menandakan bahwa metode agglomeratif tersebut belum optimal dalam menangani pencilan.

Penelitian sebelumnya mengenai penerapan metode HDBSCAN telah dilakukan oleh Wahyuni, dkk dalam mengelompokkan desa tertinggal atau perkotaan di Kabupaten Kutai Kartanegara [13]. Penelitian ini menggunakan data PODES tahun 2018 dari 84 desa/kelurahan tertinggal di Kabupaten Kutai Kartanegara dengan 14 variabel terkait infrastruktur dan kependudukan. Hasil evaluasi menunjukkan bahwa HDBSCAN memperoleh nilai *Silhouette Coefficient* sebesar 0.184. Meskipun nilai evaluasi masih tergolong rendah, penelitian ini menyimpulkan bahwa HDBSCAN tetap efektif dalam mengelompokkan wilayah yang kompleks tanpa memerlukan penentuan jumlah kluster di awal.

Berdasarkan keunggulan metode HDBSCAN dalam menangani pencilan serta mengelompokkan data dengan kepadatan tinggi dan kompleks, penelitian ini akan menerapkan metode HDBSCAN dalam mengelompokkan provinsi di Indonesia berdasarkan indikator ketenagakerjaan dalam upaya optimalisasi angkatan kerja.

Dengan demikian, penelitian ini bertujuan untuk menerapkan metode HDBSCAN dalam mengelompokkan provinsi di Indonesia berdasarkan indikator ketenagakerjaan dengan fokus pada kelompok usia, tingkat pendidikan, dan lapangan pekerjaan. Evaluasi dilakukan menggunakan *Silhouette Score* untuk menilai kualitas kluster yang dihasilkan.

Melalui penelitian ini, hasil pengelompokkan diharapkan dapat membantu pemerintah dan pemangku kepentingan dalam memahami kesiapan angkatan kerja berdasarkan kelompok provinsi dengan karakteristik serupa dalam menghadapi persaingan bebas di pasar tenaga kerja. Dengan demikian, pemerintah dapat merancang program ketenagakerjaan yang lebih spesifik dan sesuai dengan karakteristik tiap kelompok provinsi, seperti program pelatihan tenaga kerja berbasis daerah atau kebijakan insentif bagi sektor industri tertentu di provinsi dengan kelompok yang sama. Penelitian ini juga menawarkan metode klusterisasi yang lebih efektif dalam menangani data yang mengandung pencilan, sehingga dapat menjadi acuan dalam penelitian dan perumusan kebijakan ketenagakerjaan ke

depannya. Sebagai tambahan, penelitian ini juga akan mengembangkan aplikasi berbasis web menggunakan *Streamlit* untuk menampilkan hasil pengelompokan secara visual dan interaktif. Aplikasi ini bertujuan untuk mempermudah analisis serta memberikan wawasan yang lebih mendalam bagi pembuat kebijakan dan masyarakat dalam memahami distribusi tenaga kerja di Indonesia untuk mengoptimalkan angkatan kerja.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang penelitian yang telah dijelaskan di atas, rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana algoritma HDBSCAN dapat diterapkan untuk mengelompokkan provinsi di Indonesia berdasarkan indikator ketenagakerjaan yang ada?
2. Bagaimana performa algoritma HDBSCAN dalam menghasilkan kluster yang optimal pada data ketenagakerjaan yang mengandung pencilan (*outlier*)?
3. Bagaimana hasil klusterisasi dapat digunakan untuk menghasilkan analisis terhadap karakteristik ketenagakerjaan antar kelompok provinsi?
4. Bagaimana *Graphical User Interface* yang interaktif dapat dirancang untuk menampilkan hasil klusterisasi untuk mempermudah analisis terhadap distribusi ketenagakerjaan di Indonesia?

## 1.3. Batasan Masalah

Batasan masalah dalam penelitian ini difokuskan pada:

1. Penelitian ini hanya menggunakan data sekunder tahun 2022 dari publikasi Badan Pusat Statistika (BPS), karena tahun tersebut diproyeksikan sebagai puncak era bonus demografi, sehingga dapat dijadikan panduan atau referensi untuk tahun-tahun berikutnya.
2. Indikator ketenagakerjaan yang digunakan pada penelitian ini terbatas pada usia produktif, pendidikan terakhir yang ditamatkan, dan jenis lapangan kerja tiap provinsi di Indonesia.
3. Penelitian ini menghasilkan analisis kondisi angkatan kerja tiap provinsi di Indonesia berdasarkan hasil klusterisasi dan tidak mencakup perencanaan kebijakan operasional secara teknis.

4. Hasil klasterisasi disajikan melalui *Graphical User Interface* (GUI) berupa visualisasi peta interaktif menggunakan *framework* Streamlit.

#### 1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan sebelumnya, penelitian ini bertujuan untuk mencapai hal-hal berikut:

1. Menerapkan algoritma HDBSCAN untuk mengelompokkan provinsi di Indonesia berdasarkan indikator ketenagakerjaan.
2. Mengevaluasi performa algoritma HDBSCAN dalam menghasilkan klaster yang optimal pada data ketenagakerjaan yang mengandung pencilan (*outlier*).
3. Menganalisis hasil klasterisasi untuk memberikan analisis terhadap karakteristik ketenagakerjaan antar kelompok provinsi.
4. Menampilkan hasil klasterisasi melalui *Graphical User Interface* (GUI) berupa visualisasi peta interaktif menggunakan *framework* Streamlit.

#### 1.5. Manfaat Penelitian

Berdasarkan tujuan penelitian yang telah disebutkan di atas, terdapat beberapa hal yang dapat dianggap sebagai manfaat dari penelitian. Manfaat tersebut dapat dijabarkan secara akademis atau teoritis maupun praktis, yakni:

1. Manfaat Teoritis:

Penelitian ini berkontribusi dalam memperluas eksplorasi dan penerapan metode dalam bidang *data mining*, khususnya pada penggunaan algoritma klasterisasi HDBSCAN untuk mengelompokkan provinsi di Indonesia berdasarkan karakteristik ketenagakerjaan. Dengan menyesuaikan parameter HDBSCAN selama proses klasterisasi, penelitian ini memberikan pemahaman yang lebih mendalam mengenai efektivitas algoritma tersebut dalam menangani data yang mengandung pencilan dan memiliki distribusi yang tidak merata. Selain itu, penelitian ini juga berkontribusi pada pengembangan *Graphical User Interface* (GUI) sebagai sarana visualisasi hasil klasterisasi berupa peta interaktif.

## 2. Manfaat Praktis:

### a. Bagi Penulis

Penelitian ini memberikan pengalaman langsung dalam pengolahan data, deteksi pencilan, penerapan algoritma klasterisasi HDBSCAN, serta pengembangan visualisasi interaktif menggunakan GUI berbasis Streamlit. Melalui proses ini, penulis memperoleh pengalaman dalam mengimplementasikan solusi klasterisasi berbasis data untuk mengungkap pola tersembunyi antar wilayah dan menyajikannya dalam bentuk visualisasi yang informatif dan mudah dipahami.

### b. Bagi Peneliti Selanjutnya

Penelitian ini dapat menjadi referensi dalam pengembangan studi lanjutan yang berkaitan dengan klasterisasi data sosial, khususnya yang mengandung pencilan dan distribusi tidak merata. Metode HDBSCAN dan pendekatan visualisasi GUI yang digunakan dapat dijadikan acuan serta dikembangkan lebih lanjut dalam konteks data lain yang sejenis.

### c. Bagi Pemerintah dan Pembuat Kebijakan

Hasil klasterisasi yang diperoleh dapat memberikan gambaran mengenai karakteristik ketenagakerjaan antar provinsi, sehingga dapat menjadi masukan awal dalam merancang kebijakan ketenagakerjaan yang lebih spesifik dan sesuai dengan kebutuhan tiap kelompok provinsi. Visualisasi interaktif yang disajikan juga dapat mempermudah interpretasi data oleh pemangku kepentingan.