Penerapan Algoritma K-MEANS Clustering untuk Pemetaan Data Tingkat Pengangguran Terbuka di Kota/Kabupaten di Indonesia

La Ode Muhammad Yudhy Prayitno#1, Rama Qubra Putra#2, Fiqrah Idhul Dwi Anugerah,b3

*#Teknik Informatika Universitas Halu Oleo  
Kambu, Kendari, Sulawesi Tenggara 93231*

1yudhyprayitno567@gmail.com

2ramaqubra.15@gmail.com

3emailnyafiqrah10@gmail.com

**Abstrak**

Tingkat pengangguran terbuka merupakan indikator penting dalam menilai kesehatan ekonomi suatu daerah. Penelitian ini bertujuan memetakan tingkat pengangguran terbuka di kota/kabupaten di Indonesia menggunakan algoritma K-MEANS clustering. Metode ini dipilih karena kemampuannya mengelompokkan data berdasarkan karakteristik serupa, sehingga memudahkan analisis pola pengangguran di berbagai wilayah. Data berasal dari Badan Pusat Statistik (BPS), mencakup tingkat pengangguran terbuka di 514 kota/kabupaten di Indonesia pada tahun 2023. Data dipreproses melalui pembersihan dan normalisasi untuk memastikan kualitas dan konsistensi. Selanjutnya, algoritma K-MEANS diterapkan untuk mengelompokkan data menjadi beberapa klaster berdasarkan kemiripan tingkat pengangguran. dan Silhouette. Hasil klastering menunjukkan tiga klaster utama yang mengelompokkan wilayah dengan tingkat pengangguran berbeda. Klaster pertama mencakup wilayah dengan tingkat pengangguran rendah, klaster kedua mencakup wilayah dengan tingkat pengangguran sedang, dan klaster ketiga mencakup wilayah dengan tingkat pengangguran tinggi. Analisis hasil klastering memberikan wawasan berharga bagi pembuat kebijakan dalam merancang strategi pengentasan pengangguran yang tepat sasaran. Kesimpulan penelitian menunjukkan algoritma K-MEANS efektif dalam memetakan tingkat pengangguran terbuka di Indonesia dan dapat digunakan sebagai alat bantu dalam pengambilan keputusan terkait kebijakan ekonomi dan sosial.

**Kata kunci**: Analisis Data, K-MEANS Clustering, Kota/Kabupaten, Pemetaan Data, Pengangguran Terbuka, Indonesia

The Application of K-MEANS Clustering Algorithm for Mapping Open Unemployment Data in Cities/Regencies in Indonesia

**Abstract**

The open unemployment rate is an important indicator in assessing the economic health of a region. This study aims to map the open unemployment rate in cities/regencies in Indonesia using the K-MEANS clustering algorithm. This method was chosen because of its ability to group data based on similar characteristics, making it easier to analyze unemployment patterns in various regions. The data comes from the Central Bureau of Statistics (BPS), covering open unemployment rates in 514 cities/districts in Indonesia in 2023. The data was preprocessed through cleaning and normalization to ensure quality and consistency. Next, the K-MEANS algorithm was applied to group the data into clusters based on the similarity of unemployment rates. Selection of the optimal number of clusters is done using the Elbow and Silhouette methods. The clustering results show three main clusters that group regions with different unemployment rates. The first cluster includes regions with low unemployment rates, the second cluster includes regions with medium unemployment rates, and the third cluster includes regions with high unemployment rates. Analysis of the clustering results provides valuable insights for policy makers in designing targeted unemployment alleviation strategies. The conclusion shows that the K-MEANS algorithm is effective in mapping the open unemployment rate in Indonesia and can be used as a decision-making tool related to economic and social policies.

**Keywords**: Cities/Districts, Data Analysis, Data Mapping, Indonesia, K-MEANS Clustering, Open Unemployment,

1. Pendahuluan

Tingkat pengangguran terbuka adalah salah satu indikator utama dalam menilai kesehatan ekonomi suatu daerah. Pengangguran terbuka didefinisikan ketika seseorang yang termasuk dalam usia kerja sedang mencari pekerjaan, sedang mempersiapkan usaha, atau tidak mencari pekerjaan karena merasa tidak mungkin mendapatkan pekerjaan [1]. Sebagai negara berkembang Indonesia tidak akan lepas dari masalah pengangguran [2]. Pengangguran terbuka biasanya terjadi di kalangan generasi muda yang baru saja menyelesaikan pendidikan menengah dan tinggi. Ada kecenderungan bagi mereka yang baru saja menyelesaikan pendidikannya untuk mencari pekerjaan sesuai dengan keinginannya [3]. Badan Pusat statistik (BPS) merilis data Tingkat Pengangguran Terbuka (TPT) pada bulan Februari dan bulan Agustus tiap tahun. Pengangguran tidak hanya berdampak pada individu dan keluarga yang mengalaminya tetapi juga pada stabilitas ekonomi dan sosial suatu negara. Pemetaan tingkat pengangguran di berbagai wilayah sangat penting untuk memahami distribusi pengangguran dan mengidentifikasi daerah-daerah yang membutuhkan perhatian lebih.

Clustering merupakan salah satu metode data mining terpercaya dan menjadi instrumen yang valid memecahkan masalah komplek ilmu komputer dan statistik. Clustering bekerja dengan mengelompokkan titik-titik data dalam dua kelompok atau lebih, dimana titik-titik data didalam kelompok yang sama lebih mirip satu sama lain dibanding dengan kelompok data lainnya [4]. Penggunaan algoritma K-MEANS clustering [5] dalam penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas tentang pola pengangguran di berbagai kota dan kabupaten di Indonesia. Dengan mengelompokkan wilayah-wilayah berdasarkan karakteristik tingkat penganggurannya, pemerintah dan pemangku kepentingan dapat merancang strategi yang lebih efektif dan tepat sasaran untuk mengatasi masalah pengangguran.

Penelitian terdahulu telah menunjukkan bahwa pengelompokan wilayah berdasarkan tingkat pengangguran dapat memberikan wawasan yang berharga. Misalnya, penelitian di Kalimantan Timur mengungkapkan bahwa meskipun provinsi ini memiliki Pendapatan Domestik Bruto Regional (PDRB) tinggi, tingkat penganggurannya masih signifikan, menunjukkan adanya dua klaster utama dengan karakteristik yang berbeda. Dalam penelitian ini, algoritma K-MEANS diterapkan untuk mengelompokkan data tingkat pengangguran di berbagai kabupaten/kota di Kalimantan Timur. Hasil klastering menunjukkan nilai koefisien Silhouette sebesar 0.45 dan nilai Davies-Bouldin Index (DBI) sebesar 1.25, yang menunjukkan bahwa klaster yang terbentuk cukup baik [6]. Penelitian serupa di Jawa Barat menggunakan data pengangguran dari tahun 2013-2021 untuk mengelompokkan kabupaten/kota berdasarkan tingkat pengangguran, yang terbagi menjadi klaster dengan tingkat pengangguran tinggi dan rendah​​ . Penelitian ini menggunakan metode Elbow untuk menentukan jumlah klaster optimal, yang menunjukkan bahwa tiga klaster adalah yang paling optimal. Nilai koefisien Silhouette yang diperoleh adalah 0.51 dan DBI sebesar 1.10, yang menunjukkan bahwa klaster yang terbentuk cukup valid [7]. Selain itu, penelitian di Indonesia menunjukkan bahwa pengelompokan provinsi berdasarkan tingkat pengangguran dapat membantu pemerintah dalam memperluas lapangan pekerjaan untuk mengembangkan dan meningkatkan ekonomi di setiap provinsi. Penelitian ini menemukan bahwa klaster dengan tingkat pengangguran tertinggi memiliki potensi untuk perbaikan ekonomi melalui peningkatan lapangan pekerjaan dan pelatihan kerja. Hasil klastering menunjukkan nilai koefisien Silhouette sebesar 0.58 dan DBI sebesar 0.95, yang menunjukkan klaster yang sangat baik [8]. Hal ini didukung oleh penelitian lain yang mengidentifikasi faktor-faktor yang mempengaruhi tingkat pengangguran terbuka di Indonesia, termasuk pertumbuhan ekonomi, tingkat pendidikan, dan pertumbuhan penduduk. Penelitian ini menggunakan algoritma K-MEANS untuk mengelompokkan data tingkat pengangguran berdasarkan provinsi, dengan hasil menunjukkan bahwa nilai koefisien Silhouette sebesar 0.53 dan DBI sebesar 1.05 [9]. Klasifikasi wilayah berdasarkan tingkat pengangguran juga membantu dalam merancang kebijakan yang lebih tepat sasaran untuk mengatasi pengangguran.

Bagaimana algoritma K-MEANS clustering dapat diterapkan untuk memetakan data tingkat pengangguran terbuka di kota/kabupaten di Indonesia? Tujuan penelitian ini adalah untuk: (1) menerapkan algoritma K-MEANS clustering pada data tingkat pengangguran terbuka di kota/kabupaten di Indonesia, (2) mengidentifikasi klaster-klaster utama yang menggambarkan tingkat pengangguran di berbagai wilayah, dan (3) memberikan rekomendasi kebijakan berdasarkan hasil klastering.

1. METODOLOGI PENELITIAN
2. *Pengumpulan dan Pra-Pemrosesan Data*

Pengumpulan data dilakukan dalam format CSV yang kemudian akan dibersihkan untuk memastikan kualitas analisis. Pembersihan ini mencakup penghapusan kolom-kolom yang tidak relevan serta penanganan nilai-nilai yang hilang. Data yang diekstraksi difokuskan pada tahun terbaru, yaitu 2021, 2022, dan 2023. Untuk memastikan analisis yang lebih fokus, seleksi fitur dilakukan dengan mempertahankan kolom 'Kabupaten/Kota' dan 'Tahun Terbaru', yang penting untuk klasterisasi.

1. *Penentuan Jumlah Klaster*

Untuk menentukan jumlah klaster optimal, digunakan metode Elbow. Metode ini melibatkan perhitungan Within-Cluster Sum of Squares (WCSS) untuk berbagai jumlah klaster, mulai dari satu hingga sepuluh. WCSS dihitung menggunakan rumus berikut:



Keterangan:

 = adalah penjumlahan dari seluruh data point, di mana n adalah total jumlah data point.

 = adalah penjumlahan dari seluruh klaster, di mana k adalah jumlah klaster.

 = adalah data point ke-i dalam klaster ke-j.

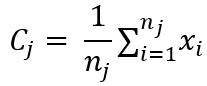
 = adalah centroid dari klaster ke-j.

 = adalah kuadrat jarak antara data point ke-i dengan centroid klaster ke-j.

Di mana xij​ adalah data point ke-i dalam klaster ke-j, dan cj​ adalah centroid dari klaster ke-j. Grafik Elbow Point kemudian digunakan untuk mengidentifikasi titik di mana penambahan klaster tidak lagi memberikan penurunan signifikan pada WCSS, menunjukkan jumlah klaster optimal.

1. *Klasterisasi Menggunakan K-Means*

Untuk menentukan jumlah klaster optimal, digunakan metode Elbow. Metode ini melibatkan perhitungan Within-Cluster Sum of Squares (WCSS) untuk berbagai jumlah klaster, mulai dari satu hingga sepuluh. WCSS dihitung menggunakan rumus berikut:



Di mana cj adalah centroid klaster ke-j, nj adalah jumlah data point dalam klaster tersebut. Setiap data point, yaitu kabupaten atau kota, diberi label klaster yang sesuai. Selain itu, dilakukan perhitungan dan analisis terhadap sentroid masing-masing klaster, yang merupakan pusat dari klaster tersebut. Analisis ini membantu dalam memahami karakteristik utama dari setiap klaster.

1. *Penetapan Kategori Klaster*

Hasil klasterisasi kemudian dikategorikan berdasarkan tingkat pengangguran: rendah, menengah, dan tinggi. Kategori ini ditentukan berdasarkan urutan nilai sentroid dari masing-masing klaster. Dengan demikian, setiap klaster dapat diidentifikasi sebagai representasi dari tingkat pengangguran tertentu, memudahkan interpretasi dan tindakan selanjutnya.

1. *Penyesuaian Nama Wilayah*

Pada tahap ini, nama wilayah antara shapefile dan data pengangguran diperiksa dan disesuaikan. Ini dilakukan untuk mengatasi perbedaan nama wilayah yang mungkin ada. Sebuah dictionary khusus digunakan untuk menyelaraskan perbedaan ini, memastikan integritas data dan akurasi analisis.

1. *Penyimpanan Hasil*

Hasil akhir dari analisis disimpan dalam file CSV. Penyimpanan ini bertujuan untuk keperluan lebih lanjut, seperti pelaporan atau analisis tambahan. File CSV yang berisi hasil analisis klasterisasi pengangguran di tingkat kabupaten/kota akan memudahkan peneliti atau pihak terkait dalam melakukan evaluasi dan pengambilan keputusan berdasarkan data yang telah dianalisis.

1. HASIL DAN PEMBAHASAN