

ABSTRAK

Pengangguran menjadi salah satu masalah yang ada di Indonesia sebagai negara berkembang. Sesuai data Badan Pusat Statistik (BPS), pengangguran merupakan salah satu permasalahan yang ada di Provinsi Jawa Barat dengan salah satu persentase pengangguran yang tinggi berada di atas rata-rata nasional di antara provinsi lain yang ada di Indonesia. Sehingga menjadi salah satu permasalahan yang harus diselesaikan di Provinsi Jawa Barat. Oleh karena itu, perlu dilakukan analisis pada kabupaten dan kota di Jawa Barat untuk dapat mengetahui daerah yang memiliki tingkat pengangguran yang tinggi dengan menerapkan teknik *clustering*. *Clustering* memiliki banyak jenis, terutama dalam pendekatan *supervised* atau *unsupervised*. Salah satu teknik *clustering unsupervised* yaitu *K-Means*. *K-Means* populer digunakan dalam masalah sosial dan memiliki kelebihan seperti sederhana, konvergensinya cepat, komputasinya efisien, dan mudah diimplementasikan. Penelitian ini bertujuan untuk menerapkan standarisasi *Z-Score* sebelum menerapkan algoritma *K-Means* dalam pengelompokan dan bagaimana pengaruh *Z-Score* dalam *clustering* itu sendiri. Pada penelitian ini pengujian dilakukan dengan menguji *hyperparameter* k 2-10 menggunakan *Silhouette Score* dan *Elbow Method*. Berdasarkan pengujian hasil terbaik berada pada $k=3$ dengan nilai 4,305 untuk *silhouette score* dan 46,4512 untuk *elbow method*. Setelah dilakukan analisis visualisasi dengan PCA dan t-SNE diketahui bahwa hasil *K-Means* dengan penerapan standarisasi memberikan hasil *clustering* yang tanpa tumpang tindih antar data pada masing-masing anggota *cluster*. Setelah penerapan standarisasi *Z-Score* dan *K-Means*, didapatkan tiga *cluster* pada penelitian yang terbagi menjadi *cluster 0*, *cluster 1*, dan *cluster 2*, serta penerapan standarisasi *Z-Score* membantu data untuk menghindari dominasi data pada data bersifat heterogen sehingga memberikan hasil *cluster* yang tidak bias karena perbedaan skala nilai tiap variabel.

Kata Kunci : Pengangguran, Jawa Barat, *Clustering*, *K-Means*, *Z-Score*, *Standardization*, *Silhouette Score*

ABSTRACT

Unemployment is one of the issues faced by Indonesia as a developing country. According to data from the Badan Pusat Statistik (BPS), unemployment is a significant problem in West Java Province, with one of the highest unemployment rates above the national average compared to other provinces in Indonesia. Therefore, addressing unemployment in West Java Province becomes a critical issue. To tackle this, an analysis of districts and cities in West Java is necessary to identify areas with high unemployment rates by applying clustering techniques. Clustering has various types, particularly within the approaches of supervised and unsupervised learning. One of the unsupervised clustering techniques is K-Means. K-Means is widely used in social issues and has advantages such as simplicity, fast convergence, computational efficiency, and ease of implementation. This study aims to apply Z-Score standardization before implementing the K-Means algorithm in clustering and to examine the effect of Z-Score standardization on clustering results. In this research, testing was conducted by evaluating the hyperparameter k ranging from 2 to 10 using the Silhouette Score and Elbow Method. Based on the tests, the best results were obtained at $k=3$, with a Silhouette Score of 4.305 and an Elbow Method score of 46.4512. After conducting visualization analysis using PCA and t-SNE, it was observed that the clustering results with K-Means, combined with standardization, produced clusters without overlap between the data points within each cluster. Following the implementation of Z-Score standardization and K-Means, three clusters were identified, designated as Cluster 0, Cluster 1, and Cluster 2. The application of Z-Score standardization helped the data avoid dominance issues in heterogeneous datasets, ensuring unbiased clustering results due to the varying scales of the variables.

Keywords : *Unemployment, West Java, Clustering, K-Means, Z-Score, Standardization, Silhouette Score*