

ABSTRAK

Revolusi industri 4.0 memberikan dampak yang signifikan terhadap kehidupan masyarakat sehari-hari, dengan digitalisasi sebagai penggerak utamanya. Era ini ditandai dengan pesatnya perkembangan teknologi dan informasi, menjadikan konektivitas perangkat serta digitalisasi layanan menjadi bagian tak terpisahkan dalam kehidupan. Khususnya menggunakan algoritma K-Means, diaplikasikan untuk mengelompokkan keluhan masyarakat (Rizki M, 2021) menjadi berbagai kategori yang telah ditentukan sebelumnya. Permasalahan yang diangkat bagaimana membangun suatu model klusterisasi menggunakan metode K-Means untuk melakukan klusterisasi teks aduan.

Metode Penelitian ini menggunakan metode Intrinsic evaluation dengan menggabungkan metode recall (R), precision (P), dan F-Measure pada ringkasan manual diilustrasikan dengan huruf T dengan hasil ringkasan yang dibuat oleh mesin diilustrasikan dengan huruf S

Dalam uji K-Means terhadap data “Keluhan Masyarakat” yang dipublikasikan mulai 14 Februari 2018 hingga 20 Februari 2019, berbagai ukuran digunakan untuk menilai kualitas analisis. Metrik ini mencakup inersia, Silhouette Score, akurasi, Recall, dan F Measure. Metrik inersia digunakan untuk menentukan jarak antara titik-titik dalam cluster dan pusat cluster. Selain itu, metrik Skor Silhouette mengukur kualitas setiap titik data dalam suatu kluster dibandingkan dengan kluster lainnya. Nilai Silhouette Score yang mendekati 0 menunjukkan tumpang tindih antar kelompok kluster, sedangkan nilai yang mendekati 1 menunjukkan pemisahan antar klaim yang baik. Hasilnya menunjukkan bahwa K-Means clustering menghasilkan nilai akurasi, recall dan F-measure sebesar 100%, menunjukkan akurasi yang sangat tinggi dalam mengidentifikasi dan mengklasifikasikan keluhan berdasarkan karakteristiknya. Kluster K-Means mempunyai kinerja yang sangat baik dalam mengklasifikasikan pengaduan masyarakat, sehingga menghasilkan kluster yang berkualitas tinggi dalam hal keterpisahan dan akurasi.

Kata kunci : Klusterisasi, Algoritma K-Means, Keluhan masyarakat.

ABSTRAK

The industrial revolution 4.0 has a significant impact on people's daily lives, with digitalization as the main driver. This era is marked by the rapid development of technology and information, making device connectivity and digitalization of services an inseparable part of life. Specifically using the K-Means algorithm, applied to group community complaints (Rizki M, 2021) into various predetermined categories. The problem raised is how to build a clustering model using the K-Means method to cluster complaint text.

This research method uses the Intrinsic evaluation method by combining the recall (R), precision (P), and F-Measure methods in a manual summary illustrated with the letter T with the summary results made by the machine illustrated with the letter S

In the K-Means test of "Community Complaints" data published from February 14, 2018 to February 20, 2019, various measures were used to assess the quality of the analysis. These metrics include inertia, Silhouette Score, accuracy, Recall, and F Measure. Inertia metrics are used to determine the distance between points in the cluster and the center of the cluster. In addition, the Silhouette Score metric measures the quality of each data point in a cluster compared to other clusters. Silhouette Scores close to 0 indicate overlap between cluster groups, while values close to 1 indicate good separation between claims. The results showed that K-Means clustering produced accuracy, recall and F-measure values of 100%, showing very high accuracy in identifying and classifying complaints based on their characteristics. K-Means clusters performed very well in classifying community complaints, resulting in clusters that are of high quality in terms of separability and akurasi.

Keywords: Clustering, K-Means Algorithm, Public complaints.

