

ABSTRAK

Di era digital saat ini, manusia sangat mudah bila ingin mendengarkan sebuah lagu. Kemudahan tersebut didukung dengan banyaknya *platform* yang menyediakan layanan *streaming* atau *download* dan dapat dilakukan dimana saja. *Platform online* yang dimaksud adalah *platform* yang memerlukan koneksi *internet* untuk mengaksesnya seperti, Youtube Music, Spotify, Joox, SoundCloud, dsbg. Sedangkan untuk *platform offline* merupakan *platform* yang menggunakan sinyal untuk mengaksesnya, seperti radio. Dalam hal ini terdapat salah satu lembaga yang memiliki wewenang untuk mengatur dan melakukan pengawasan konten, yaitu KPI (Komisi Penyiaran Indonesia). Beberapa lagu yang disiarkan di radio, memiliki lirik yang terkesan cabul, vulgar, mengandung unsur seksual, dan kasar terutama pada lagu berbahasa asing. *K-Means Clustering* merupakan metode yang digunakan untuk mempartisi data ke dalam sebuah kelompok. Metode ini akan mempartisi data yang memiliki ciri yang sama kedalam sebuah *cluster* atau kelompok, dan begitu juga sebaliknya. *K-Means Clustering* juga memiliki kelemahan, yaitu hasil pengelompokan dari *K-Means Clustering* sangat bergantung dengan inisialisasi awal dalam proses pengelompokannya. Untuk melihat kekuatan struktur dari *cluster* yang telah terbentuk, maka metode *Silhouette Coefficient* dipilih. Pada penelitian ini dilakukan tiga skenario pengujian yang berfokus untuk melakukan improvisasi dan melihat pengaruh inisialisasi terhadap metode *K-Means Clustering*, serta mencari inisialisasi mana yang memiliki nilai *Silhouette Coefficient* terbaik. Pengujian skenario pertama akan menggunakan penentuan inisialisasi titik centroid yang telah ditentukan. Kemudian pengujian skenario kedua akan menggunakan inisialisasi acak dengan parameter 'k-means++' dengan perhitungan jarak yang lebih cerdas. Untuk pengujian skenario ketiga akan menggunakan inisialisasi dengan parameter default dari *K-Means Clustering* 'random' yaitu secara acak. Dari 3 skenario yang dilakukan menghasilkan nilai *silhouette* sebesar 0,67 pada skenario pertama, kemudian skenario kedua 0,668, kemudian skenario ketiga menghasilkan 0,665. Ketiganya termasuk kedalam *cluster* yang memiliki kekuatan struktur yang baik, namun skenario pertama menghasilkan hasil yang paling baik.

Kata Kunci : Pengelompokan lagu, Klustering, Silhouette Coefficient, K-Means Clustering

ABSTRACT

In the current digital era, it is very easy for people to listen to a song. This convenience is supported by the many platforms that provide streaming or download services and can be done anywhere. The online platforms in question are platforms that require an internet connection to access them, such as YouTube Music, Spotify, Joox, SoundCloud, etc. Meanwhile, offline platforms are platforms that use signals to access them, such as radio. In this case, there is one institution that has the authority to regulate and supervise content, namely KPI (Indonesian Broadcasting Commission). Several songs broadcast on the radio have lyrics that seem obscene, vulgar, contain sexual elements and are rude, especially in foreign language songs. K-Means Clustering is a method that is suitable for partitioning data into one or more clusters or groups. This method will partition data that is similar and has the same characteristics into a cluster or group, and data that has different characteristics will be grouped into other groups. K-Means Clustering also has a weakness, namely that the grouping results from K-Means Clustering are very dependent on the initial initialization in the grouping process. To see the structural strength of the clusters that have been formed, the Silhouette Coefficient method was chosen. In this research, three test scenarios were carried out which focused to improve and seeing the effect of initialization on the K-Means Clustering method, as well as finding which initialization had the best Silhouette Coefficient value. The first scenario test will use the initialization of the centroid points that have been determined. Then the second scenario test will use random initialization with 'k-means++' parameters with smarter distance calculations. For testing the third scenario we will use initialization with the default parameters of K-Means Clustering 'random', namely randomly. The 3 scenarios carried out resulted in a silhouette value of 0.67 in the first scenario, then in the second scenario 0.668, then in the third scenario it resulted in 0.665. All three are included in the cluster that has good structural strength, but the first scenario produces the best results.

Keywords: Song grouping, Clustering, Silhouette Coefficient, K-Means Clustering