

Abstrak

Pada penelitian ini akan dilakukan pencarian nilai K terbaik dari 2 metode pendekatan *clustering K-Means* yaitu *Elbow* dan *Silhouette*. Berdasarkan penelitian sebelumnya oleh Purnamasari (2020) ditemukan pendekatan *Elbow* menghasilkan hasil yang lebih ideal. Maka dari itulah penulis berpikir untuk melakukan pencarian nilai K terbaik pada *clustering* aksara Jawa.

Terdapat banyak tahapan yang dilakukan pada penelitian ini. Yang pertama adalah pengumpulan data yang berupa buku hamong tani halaman 2 dan 59. Kemudian *preprocessing* untuk mendapatkan data yang bersih. Setelah itu akan disegmentasi untuk mendapatkan tiap aksaranya. Lalu akan diambil ciri untuk tiap aksaranya dengan menggunakan *Intensity of Character*. Data ciri kemudian dikelompokkan menggunakan *K-Means clustering* yang mana akan menghasilkan jumlah *cluster*, *idx*, dan *sumd*. Kemudian *idx* akan digunakan untuk evaluasi *silhouette* dan *sumd* akan digunakan untuk evaluasi *elbow*. Hasil kedua metode tersebut akan dibandingkan untuk mencari K terbaiknya.

Dari pengujian yang dilakukan disimpulkan jika K nya lebih besar maka akan menghasilkan hasil pengelompokan yang bagus sehingga metode *elbow* dianggap sebagai metode yang ideal untuk menentukan nilai K.

Kata kunci: Aksara Jawa, *Preprocessing*, Segmentasi, Ekstraksi Ciri, *K-Means*, *Silhouette*, *Elbow*.

Abstract

In this research, the best K value will be searched from 2 methods of K-Means clustering approach, namely Elbow and Silhouette. Based on previous research by Purnamasari (2020) it was found that the Elbow approach produced more ideal results. That's why the writer thought to search for the best K value in Javanese script clustering.

There are many stages carried out in this research. The first is collecting data in the form of the Hamong Tani book page 2 and 59. Then preprocessing to get clean data. After that it will be segmented to get each character. Then the characteristics for each character will be taken using the Intensity of Character. The feature data are then grouped using K-Means clustering which will produce the number of clusters, idx, and sumd. Then idx will be used for silhouette evaluation and sumd will be used for elbow evaluation. The results of the two methods will be compared to find the best K.

From the tests carried out, it is concluded that if the K is greater then it will produce good grouping results so that the elbow method is considered an ideal method to determine the K value.

Keywords: Javanese script, Preprocessing, Segmentation, Feature Extraction, K-Means, Silhouette, Elbow.