

## ABSTRAK

Pariwisata merupakan potensi yang dapat mendorong kehadiran wisatawan ke suatu daerah tertentu. Hal ini dapat direalisasikan dengan melakukan cara perluasan, pembangunan potensi wisata dan pemerataan daya tarik wisata. Untuk mendukung terlealisasinya hal tersebut perlunya dibuat sebuah pengelompokan tempat wisata yang dilihat dari sisi pemanfaatan luas potensi wisata dan jumlah kunjungan yaitu dengan tujuan untuk mengetahui tempat wisata mana yang tergolong sebagai *high priority* (prioritas tinggi dalam mengembangkan tempat wisata dengan jumlah kunjungan yang tinggi), *high potential* (tempat wisata yang memiliki potensi tinggi dalam mengolah tempat wisata dan menarik kunjungan wisatawan), dan *low potential* (tempat wisata yang masih memiliki potensi rendah dalam hal menarik pengunjung wisatawan). Hasil clustering ini menunjukkan bahwa ada beberapa cluster yang baik dikelompokkan (seperti cluster 2), tetapi ada cluster dengan kualitas clustering yang kurang optimal (seperti cluster 0). Mungkin perlu penyesuaian jumlah cluster, fitur data, atau algoritma untuk meningkatkan kualitas clustering. Berdasarkan grafik, jumlah cluster terbaik adalah  $K=6$ , karena memberikan nilai Silhouette Score tertinggi. Meskipun demikian, nilai Silhouette Score keseluruhan relatif rendah, yang mungkin menunjukkan bahwa data memiliki tumpang tindih antar cluster atau fitur yang digunakan kurang memadai untuk menghasilkan cluster yang jelas. Dan Grafik nilai eror (inertia) digunakan untuk menentukan jumlah kluster optimal dalam *K-Means*. Meskipun nilai *inertia* terus menurun, pemilihan  $K$  terbaik tidak hanya didasarkan pada *inertia* tetapi juga pada metrik lain seperti *Silhouette Score* atau metode *Elbow*. Dalam kasus ini,  $K=6$  dianggap optimal.

**Kata Kunci :** *K-Means*, Clustering, Text Mining, Silhouette Score.

## **ABSTRACT**

*Tourism is a potential that can encourage the presence of tourists to a particular area. This can be realized by expanding, developing tourism potential and distributing tourist attractions evenly. To support the realization of this, it is necessary to create a grouping of tourist attractions that is seen in terms of the wide use of tourism potential and the number of visits, namely with the aim of finding out which tourist attractions are classified as high priority (high priority in developing tourist attractions with a high number of visits), high potential (tourist attractions that have high potential in developing tourist attractions and attracting tourist visits), and low potential (tourist attractions that still have low potential in terms of attracting tourist visitors). These clustering results show that there are several clusters that are well grouped (such as cluster 2), but there are clusters with less than optimal clustering quality (such as cluster 0). It may be necessary to adjust the number of clusters, data features, or algorithms to improve clustering quality. Based on the graph, the best number of clusters is  $K=6$ , because it gives the highest Silhouette Score value. However, the overall Silhouette Score value is relatively low, which may indicate that the data has overlap between clusters or that the features used are insufficient to produce clear clusters. And the error value graph (inertia) is used to determine the optimal number of clusters in K-Means. Although the inertia value continues to decrease, the selection of the best  $K$  is not only based on inertia but also on other metrics such as the Silhouette Score or the Elbow method. In this case,  $K=6$  is considered optimal.*

**Keywords :** K-Means, Clustering, Text Mining, Silhouette Score.