

## INTISARI

Perkembangan teknologi informasi menyebabkan banyak informasi tidak lagi disimpan dalam bentuk berkas - berkas dokumen yang diletakkan di dalam lemari melainkan informasi disimpan dalam bentuk *softcopy* dalam bentuk *database*. Seiring bertambahnya jumlah dan keanekaragaman dokumen yang tersimpan menyebabkan terjadinya penumpukan yang dapat mempersulit *user* untuk mendapatkan informasi yang diinginkan, serta proses pencarian memakan waktu apabila dokumen yang tersebar tidak terkoordinasi dengan baik. Oleh sebab itu, perlu dilakukannya proses ekstraksi informasi dan pengelompokan terhadap dokumen dengan pendekatan *data mining* untuk mempermudah kinerja *user*.

Dalam penelitian ini data yang digunakan sebanyak 822 yang merupakan judul dari dokumen tata kerja organisasi (TKO) *Refinery* Unit IV Pertamina Cilacap. Proses awal yang dilakukan terhadap data yaitu menentukan atribut menggunakan analisis *text mining* serta pembobotan TF-IDF, kemudian dilakukan pengelompokan menggunakan metode *k-means clustering*.

Hasil dari proses penentuan atribut maksimal diperoleh sebanyak 52 kata dengan minimal kemunculan 24 dan maksimal 205 dari total 822 dokumen. Penentuan klaster terbaik dilakukan dengan menghitung nilai *davies boldin index* dan *silhouette index* terhadap 5 hingga 50 klaster dokumen TKO. Hasil evaluasi klaster maksimal didapat dalam jumlah 25 klaster dengan nilai *davies boldin index* terkecil, yaitu 0,983 dan nilai *silhouette index* terbesar 0,302. Anggota klaster dengan jumlah terbanyak dipegang oleh klaster 2 yakni sejumlah 99 dokumen dan paling sedikit pada klaster 17 yakni 14 dokumen. Penentuan atribut dengan analisis *text mining* serta pengelompokan dokumen yang dilakukan diharapkan dapat menjadi acuan bagi peneliti lain untuk kasus yang sejenis.

Kata Kunci : *Data Mining, Davies Bouldin Index, K-mean, Klasterisasi, Pengelompokan Dokumen, Silhouette index, Text Mining, TF-IDF*

## ABSTRACT

The development of information technology has an effect on how information no longer saved as paper-printed files which stored in rows and rows of boxes but as a digital softcopy instead. With the growing number and varieties of stored documents causes stacking which in turn making it difficult for users to access information, also the seeking process of the information takes a long time if the documents were not well-organized. Therefore, it is important to extract information and group documents using data mining approach to ease users in their jobs.

In this thesis, the number of data used is 822 data which is the title of documents stored in TKO Refinery Unit IV Pertamina Cilacap. The first step is to decide the attribute of data using text mining analysis and TF-IDF assessment, then using k-means clustering the data is classified into groups.

The result of data attribute, there are 52 words in maximum with minimum 24 and maximum 205 occurrence from the total of 822 documents. Best clustering is achieved through calculating the davies boldin index and silhouette index on 5 to 50 TKO document clusters. The maximum clustering performance is shown by dividing data into 25 clusters with smallest davis boldin index 0.983 and biggest silhouette index 0.302. Largest size cluster is in cluster 2 with 99 documents and the smallest size cluster is in cluster 17 with 14 documents only. The assignment of attribute using text mining analysis and documents clustering is expected to be a reference for the upcoming research in this field.

**Keyword :** Clustering, Data Mining, Davies Bouldin Index, Document Clustering, K-mean, Silhouette index, Text Mining, TF-IDF