

**FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS STIKUBANK (UNISBANK) SEMARANG**

Program Studi: Sistem Informasi

Tugas Akhir Sarjana Komputer

Semester Genap tahun 2020/2021

***PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK
MENGELOMPOKKAN DATA OPINI PENGGUNA TWITTER
MENGENAI PANDEMI COVID-19***

Dikogigih Prasetyantoro

17.01.55.0073

ABSTRAK

Pandemi COVID-19 membuat masyarakat sering menggunakan media sosial untuk berkomunikasi jarak jauh saat melakukan karantina mandiri. Melalui media sosial daring twitter didapatkan suatu opini yang menjadi tren pembahasan. Banyaknya data opini membuat data menjadi tidak terstruktur dan tidak berkelompok. Hal ini menimbulkan permasalahan berupa sulitnya mengetahui opini mana saja yang dapat masuk ke kelompok tren pembahasan tertentu.

Clustering dapat berguna dalam mengelompokkan data opini dan menjadi solusi masalah yang telah dipaparkan sebelumnya, dengan begitu kita dapat mengetahui tren pembahasan apa saja yang dibahas masyarakat saat pandemi COVID-19. Pada penelitian ini algoritma yang digunakan adalah algoritma *K-Means* dan *davies bouldin index* dalam menentukan jumlah *cluster*.

Metode *davies bouldin index* dalam menentukan jumlah *cluster* terbaik menunjukkan 3 *cluster* merupakan jumlah yang optimal, dengan DBI *score* sebesar 7.6050425313519225. Hasil yang didapatkan dari 1350 data opini didalam 3 *cluster* ditemukan: 1070 opini di *cluster* 0, sebanyak 74 opini di *cluster* 1 dan 206 opini didalam *cluster* 2.

Kata Kunci: Clustering, K-Means, Twitter, Pandemi COVID-19, Text Mining

ABSTRACT

The COVID-19 pandemic has made people often use social media to communicate remotely when self-quarantining. Through the online social media, twitter, an opinion has become a trend of discussion. The large number of opinion data makes the data unstructured and ungrouped. This raises problems in the form of difficulty in knowing which opinions can fit into certain discussion trend groups.

Clustering can be useful in classifying opinion data and being a solution to the problems that have been described previously, that way we can find out what trends are discussed by the community during the COVID-19 pandemic. In this study, the algorithm used is the K-Means algorithm and the Davies Bouldin Index in determining the number of clusters.

The davies bouldin index method in determining the best number of clusters shows that 3 clusters is the optimal number, with a DBI score of 7.6050425313519225. The results obtained from 1350 opinion data in 3 clusters were found: 1070 opinions in cluster 0, as many as 74 opinions in cluster 1 and 206 opinions in cluster 2.

Key Words: Clustering, K-Means, Twitter, Pandemic COVID-19, Text Mining

Pembimbing

5 Juni 2021

Herny Februariyanti, ST, M. Cs