

## ABSTRAK

### PENGGUNAAN METODE *K-MEANS* DAN *K-MEANS++* SEBAGAI *CLUSTERING* DATA COVID-19 DI PULAU JAWA

Oleh  
Nursatio Nugroho  
18102208

*Virus Corono(covid-19) merupakan penyakit menular antara hewan dan manusia. Pada akhir desember 2019, virus itu teridentifikasi di Provinsi Wuhan, China. Data penyebaran yang disajikan oleh covid19.go.id hanya berupa agregat data setiap provinsi tetapi tidak ada informasi sebaran kasus covid perkabupaten/kota. Informasi data covid perkabupaten dan kota sangat diperlukan untuk mengetahui cluster penyebaran kasus covid. Penelitian ini bertujuan untuk mengklaster data penyebaran covid-19 di setiap kabupaten di Pulau Jawa sehingga menghasilkan klaster zona yang harus dilaksanakan PPKM berdasarkan kasus positif, vaksin dosis pertama, dan dosis kedua. vaksin. Penelitian ini akan menggunakan algoritma *K-Means* dan *K-Means++* untuk menentukan tingkat penyebaran covid-19 di pulau jawa. Berdasarkan jumlah kasus positif, vaksin pertama, dan vaksin kedua, kasus tersebut dikategorikan. Setelah dilakukan pengelompokan dan mendapatkan cluster pada masing-masing kelompok, setiap cluster akan dievaluasi kualitasnya menggunakan koefisien siluet untuk memilih yang terbaik. Hasil kajian tersebut diharapkan dapat mengungkap sejauh mana penyebaran virus covid-19 di setiap kabupaten/kota di Pulau Jawa, serta cluster dengan skor *Silhouette Coefficient* tertinggi. Untuk hasil pengujian cluster menggunakan *Silhouette Coefficient*, metode *K-Means*  $K=3$  menghasilkan 0,825,  $K=4$  menghasilkan 0,873,  $K=5$  menghasilkan 0,862, dan  $K=6$  menghasilkan 0.841, untuk metode *K-Means++*,  $k=3$  menghasilkan 0,822,  $K=4$  menghasilkan 0,865,  $K=5$  menghasilkan 0,882, dan  $K=6$  menghasilkan 0.858. Hasil penelitian menunjukkan bahwa Kota Jakarta Barat, Kota Jakarta Selatan, Kota Jakarta Utara, Kota Jakarta Pusat termasuk daerah yang sangat rawan kasus covid-19. Berdasarkan hasil pengujian menggunakan *Silhouette Coefficient* akan lebih baik metode *K-Means* dalam pembentukan cluster dengan nilai  $k$  yang lebih rendah sedangkan *K-Means++* lebih unggul untuk pembentukan cluster yang lebih tinggi.*

Kata Kunci: *K-Means*, *K-Means++*, *Clustering*, *Covid-19*, *Silhouette Coefficient*