

IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA PENYEBARAN COVID-19 (STUDI KASUS DI JAWA BARAT)

Andina Dyaning Putri

(180441100084)

Dosen Pembimbing

1. Mula'ab,S.Si.,M.Kom
2. Dr.Wahyudi Setiawan ,S.Kom.,M.Kom

ABSTRAK

Corona virus atau (COVID-19) adalah varian virus jenis baru yang pertama kali ditemukan Desember 2019 di Cina lebih tepatnya provinsi Wuhan. Virus baru ini penyakit yang penularannya cepat, penyebaran virus ini akibat adanya kontak fisik antar manusia ataupun dapat melewati percikan kecil (droplet) berasal dari mulut atau hidung saat bersin ataupun berbicara. Pada permasalahan ini akan dilakukan *K-Means clustering* untuk mengetahui tingkat penyebaran virus COVID-19 pada tiap-tiap kecamatan di Jawa Barat. *K-Means Cluster* termasuk ke dalam *clustering* non-hierarki yang menggabungkan informasi ke dalam beberapa *cluster* yang mempunyai atribut serupa ke dalam beberapa *cluster* yang serupa. Untuk melihat seberapa baik kemampuan dalam pengklusteran suatu data diletakkan dalam suatu *cluster* tersebut maka perlu digunakan *Silhouette Coefficient* dan Metode *Elbow*. Pengelompokan berdasarkan parameter kasus positif terinfeksi COVID-19, pasien isolasi/dalam perawatan, pasien sembuh/selesai isolasi, pasien yang meninggal dunia dan ada atribut tambahan seperti jumlah penduduk, kepadatan penduduk. Metode yang digunakan untuk pengumpulan data yaitu metode dokumentasi berupa data sekunder yang didapatkan dari publikasi buku Badan Pusat Statistik dan dari website resmi pusat informasi dan koordinasi COVID-19 di Jawa Barat dengan alamat web pikobar.jabarprov.go.id. Proses yang dilakukan pada penelitian ini dimulai dari pengumpulan data, normalisasi data, *K-Means clustering*. Data yang dipakai ada 625 kecamatan hasil dari *clustering* tersebut menghasilkan nilai *k* terbaik yaitu 3 *cluster*.

Kata Kunci : COVID-19, Jawa Barat, *Silhouette Coefficient*, *Clustering*, *K-means*

DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
LEMBAR PERNYATAAN	iii
LEMBAR PERSEMBAHAN	v
ABSTRAKvi	
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.2.1 Permasalahan.....	2
1.2.2 Metode Usulan	2
1.2.3 Pertanyaan Penelitian	2
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat	3
1.5 Metodologi Penulisan Skripsi	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 COVID-19.....	5
2.2 Data Mining	5
2.2.1 Teknik-Teknik <i>Data mining</i>	7
2.3 Clustering	9

2.3.1	Metode Clustering	9
2.4	Algoritma Berbasis Jarak	10
2.5	Algoritma K-Means	12
2.6	Metode Elbow	13
2.7	Validasi Cluster	14
2.8	Normalisasi Data	16
2.8.1	Metode Normalisasi Data	16
2.11	Penelitian Terdahulu	17
BAB III	METODE PENELITIAN	20
3.1	Tahapan Penelitian	20
3.2	Gambaran Penelitian	20
3.3	Perangkat Penelitian	27
3.3.1	Perangkat Lunak (<i>software</i>)	27
3.3.2	Perangkat Keras (<i>hardware</i>)	27
BAB IV	HASIL DAN PEMBAHASAN	28
4.1	Deskripsi Data	28
4.2	Normalisasi Data	28
4.3	K-Means Clustering	29
BAB V	PENUTUP	38
5.1	KESIMPULAN	38
5.2	SARAN	38
	DAFTAR PUSTAKA	39

DAFTAR GAMBAR

Gambar 2. 1 Tahapan proses KDD	6
Gambar 2. 2 Teknik <i>Data Mining</i>	7
Gambar 3. 1 Blok diagram <i>Clustering</i>	21
Gambar 3. 2 <i>Flowchart K-Means Clustering</i>	22
Gambar 4. 1 <i>source code</i> import library	30
Gambar 4. 2 <i>source code</i> input data	31
Gambar 4. 3 output dari load data	32
Gambar 4. 4 output dari load data	32
Gambar 4. 5 <i>source code</i> clustering	33
Gambar 4. 6 hasil <i>clustering</i>	34
Gambar 4. 7 <i>source code</i> Metode <i>Elbow</i>	34
Gambar 4. 8 output Metode <i>Elbow</i>	36
Gambar 4. 9 <i>source code</i> evaluasi menggunakan <i>silhouette</i>	36
Gambar 4. 10 hasil output nilai <i>silhouette</i>	38
Gambar 4. 11 <i>source code</i> visualisasi <i>clustering</i>	38
Gambar 4. 12 visualisasi <i>clustering</i>	39

DAFTAR TABEL

Tabel 2. 1 Kriteria Nilai <i>Silhouette Coefficient</i>	17
Tabel 3. 1 Titik Pusat <i>Cluster</i>	23
Tabel 3. 2 Tabel Pengklusteran.....	25
Tabel 3. 3 Tabel Centroid Baru.....	26
Tabel 3. 4 Tabel <i>Clustering</i> Baru.....	26
Tabel 4. 4 Tabel Normalisasi	30