

P-ISSN : 2337 - 8344

E-ISSN : 2623 - 1247

Jurnal InformaSI dan Komputer



**Diterbitkan Oleh :
STMIK DIAN CIPTA CENDIKIA KOTABUMI**

Volume 10 Nomor 1 Tahun 2022

Penerbit

Lembaga Penelitian STMIK Dian Cipta Cendikia Kotabumi

**Hak atas naskahh/tulisan tetap berada pada penulis, isi diluar tanggung jawab
penerbit dan Dewan Penyunting**



PENGANTAR REDAKSI

Puji syukur dipanjatkan kehadirat Tuhan Yang Maha Esa, atas karunia dan limpahan rahmatNYA jualan Jurnal Informasi dan komputer (JIK) STMIK Dian Cipta Cendikia Kotabumi ini dapat terwujud. Jurnal Informasi dan Komputer (JIK) yang terbit dua (2) kali dalam setahun ini merupakan suatu wadah untuk penyebar luasan hasil-hasil penelitian, studi pustaka, karya ilmiah yang berkaitan dengan Informasi dan Komputer khususnya bagi dosen-dosen STMIK Dian Cipta Cendikia Kotabumi serta umumnya para cendekiawan, praktisi, peneliti ilmu Informatika dan Komputer.

Harapan, dengan diterbitkannya Jurnal Informasi dan Komputer (JIK) ini sebagai salah satu bentuk sumbangan pemikiran dalam pengembangan ilmu informatika dan komputer yang berkaitan dengan kajian-kajian di bidang teknologi Informatik, Komunikasi Data dan Jaringan Komputer, perancangan dan Rekayasa Perangkat Lunak, serta ilmu-ilmu yang terkait dengan bidang Informasi dan Komputer lainnya.

Berkenaan dengan harapan tersebut, kepada para peneliti, dosen dan praktisi yang memiliki hasil-hasil penelitian, kajian pustaka, karya ilmiah dalam bidang tersebut diatas, dengan bangga redaksi Jurnal Informasi dan Komputer (JIK) menerima naskah ringkasan untuk dimuat pada jurnal Informasi dan Komputer (JIK) STMIK Dian Cipta Cendikia Kotabumi dengan berpedoman pada penulisan naskah jurnal sebagaimana dilampirkan pada halaman belakang (Bagian kulit dalam) buku jurnal ini.

Mutu dari suatu jurnal ilmiah tidak hanya ditentukan oleh para pengelolanya saja, tetapi para penulis dan pembaca jualan yang mempunyai peranan besar dalam meningkatkan mutu jurnal Informatika dan Komputer ini. Merujuk pada realita ini kamu sangat mengharapkan peran aktif dari peneliti untuk bersama-sama menjaga dan memelihara keberlangsungan dari jurnal Informasi dan Komputer STMIK Dian Cipta Cendikia Kotabumi ini. Yang juga tidak kalah pentingnya dari partisipasi tersebut diatas, adalah saran dan kritik yang membangun dari pembaca yang budiman agar kiranya dapat disampaikan langsung kepada redaksi JIK. Saran dan kritik yang membangun akan dijadikan masukan dan pertimbangan yang sangat berarti guna peningkatan mutu dan kualitas Jurnal Informasi dan Komputer STMIK Dian Cipta Cendikia Kotabumi.

Tak lupa diucapkan terima kasih yang tak terhingga atas perhatian dan kerjasama dari semua pihak yang tak dapat disebutkan satu persatu hingga dapat diterbitkan nya Jurnal Informasi dan Komputer (JIK) STMIK Dian Cipta Cendikia Kotabumi. Semoga apa yang telah diperbuat untuk kebaikan akan menjadi amal ibadah, amin.

Kotabumi, 25 April, 2022



Dewan Redaksi

JURNAL INFORMASI DAN KOMPUTER

Volume 10 Nomor 2 April 2022

Jurnal Informasi dan Komputer merupakan Sarana informasi ilmu pengetahuan, Teknologi dan Komunikasi yang berupa hasil penelitian, tulisan ilmiah, Atau pun studi pustaka. Jurnal ini terbit dua kali setahun pada bulan April dan Oktober. Berisi hasil penelitian ilmiah di bidang informatika yang bertujuan untuk menghubungkan adanya kesenjangan antar kemajuan teknologi dan hasil penelitian. Jurnal ini di terbitkan pertama kali pada tahun 2013.

Penanggung Jawab:

Ketua STMIK Dian Cipta Cendikia Kotabumi

Pembina:

Ketua STMIK Dian Cipta Cendikia Kotabumi
Ketua Lembaga Penelitian STMIK Dian Cipta Cendikia Kotabumi

Pimpinan Redaksi

Dwi Marisa Efendi, S.Kom., M.Ti

Redaksi pelaksana

Rustam, S.Kom., M.Ti (STMIK Dian Cipta Cendikia Kotabumi)
Nurmayanti M.Kom (STMIK Dian Cipta Cendikia Kotabumi)
Sukatmi, S.Kom., M.Kom (AMIK DCC Bandar Lampung)
Sampurna Dadi Riskiono, M.Kom (Universitas Teknokrat Indonesia)
Ifo Wahyu Pratama, S.Kom., M.Ti (AMIK MASTER Lampung)

Mitra Bestari

Dr. RZ. ABDUL AZIZ, ST., MT (Institut Informatika dan Bisnis Darmajaya)
Dr. Dadang Sudrajat, S.Si, M.Kom (STMIK IKMI Cirebon)
Dr. Septafiansyah Dwi Putra, S.T., M.T (Politeknik Negeri Lampung)
Dr. Evi Grativiani, S.E., M.S.I (Universitas Sebelas Maret)
Rohmat Indra Borman (Universitas Teknokrat Indonesia)
Ferry Wongso, S.KOm., M.Kom (STMIK Darma Pala Riau)
Ferly Ardhy, S.Kom., M.Ti (Universitas Aisyah Pringsewu)
Firmansyah, S.E., M.Si (STMIK Darma Pala Riau)

Amarudin (Universitas Teknokrat Indonesia)
Didi Susianto, S.T., M.Kom (AMIK Dian Cipta Cendikia Bandar Lampung)
Alhibarsyah, St., M.Kom (STMIK Tunas Bangsa Bandar Lampung)
Kemal Farouq Mauladi, S.Kom .M.Kom (Universitas Islam Lamongan)
Rima Mawarni, M.Kom (STMIK Dian Cipta Cendikia Kotabumi)
Wira Jaya Hartono, S.Pd., M.Pd (STMIK Darma Pala Riau)

Penerbit : STMIK Dian Cipta Cendikia Kotabumi Bekerja Sama Dengan LPPM STMIK Dian Cipta Cendikia Kotabumi.

Alamat Redaksi/Penerbit:

Jl. Negara No. 3 Candimas Kotabumi Lampung Utara
No Telp/Fax 0724 23003
Email : lppm-stmik@dcc.ac.id



JURNAL INFORMASI DAN KOMPUTER VOL. 10 NO. 2 THN. 2022

DAFTAR ISI

	Halaman
Peningkatan Pengelolaan Arsip Surat Menyurat Melalui Aplikasi Berbasis Web Dengan Metode <i>First In First Out</i> Yuli Syafitri ¹ , Reni Astika ² , Lusia Septia Eka Esti Rahayu ³ , (AMIK Dian Cipta Cendikia ¹² , AMIK Lampung ³)	01-08
Pengembangan Sistem Informasi Perpustakaan Berbasis Web Pada Amik Dian Cipta Cendikia Bandar Lampung Sukatmi ¹ , Euis Mustika Prianganti ² , Astriyanti ³ (AMIK DCC Bandar Lampung ¹²³)	09-14
Klasifikasi Penyakit <i>Powdery Mildew</i> Pada Ceri Manis Dengan Menggunakan Algoritma <i>Convolutional Neural Network</i> (CNN) Iwansyah Edo Hendrawan ¹ , M. Ilhamsyah ² , Dadang Yusup ³ (Universitas Singaperbangsa Karawang ¹²³)	15-20
Penerapan Finite State Automata Pada Desain Vending Machine Masker Dan Hand Sanitizer Ridwan ¹ , Windu Gata ² , Hafifah Bella Novitasari ³ , Laela Kurniawati ⁴ , Sri Rahayu ⁵ (Universitas Nusa Mandiri ¹²).....	21-28
Analisis Perhitungan Muatan Sedimentasi Berdasarkan Kedalaman Air (<i>Chart Datum</i>) Pada Senipah Channel Di Kabupaten Kutai Kartanegara Berbasis Web Salmajah (Stmik Handayani Makasar)	29-43
Aplikasi Pembelajaran Ilmu Tajwid Berbasis Mobile Novita Lestari Anggreini ¹ , Ichsan Perdana Putra ² (Politeknik TEDC Bandung).....	44-49
Implementasi Algoritma Naïve Bayes Untuk Memprediksi Pengaruh Media Sosial Terhadap Semangat Belajar Mahasiswa Di Masa Pandemi Covid 19 Fiqih Satria ¹ , Hermanto ² (Universitas Raden Intan Lampung)	50-56
Klasifikasi Kinerja Pembayaran Angsuran Dengan Algoritma Naive Bayes (Studi Kasus : Data Nasabah Koperasi Simpan Pinjam Pembiayaan Syariah Bina Bersama) Dwi Marisa ¹ , Sigit Mintoro ² , Supriyanto ³ , Sani Hanika lubis ⁴ , Sri Lestari ⁵ (STMik Dian Cipta Cendikia Kotabumi)	57-61
Peningkatan Akurasi Prediksi Pengadaan Bahan Baku Produksi Dengan Menggunakan Metode <i>Neural Network</i> Mumtaz Muttakin ¹ , Sabar Hanadwiputra ²	

(STMIK Bani Saleh, Bekasi)	62-72
Penerapan Konsep Finite State Automata Pada Simulasi Vending Machine	
Pergantian Seragam Karyawan	
Ristyani Slamet ¹ , Windu Gata ² , Ketut Sakho Parthama ³ , Nita Merlina ⁴ , Eni Heni Hermaliani ⁵	
(Universitas Nusa Mandiri ^{1,2,4,5} , Universitas Pramita Indonesia ³)	73-79
Penerapan Metode Electre Untuk Pemilihan Pengajar Terbaik	
Muchamad Maskhur ¹ , Wiwien Hadikurniawati ²	
(Universitas Stikubank, Semarang).....	80-88
Sistem Pendukung Keputusan Kelayakan Kenaikan Pangkat Jabatan Fungsional(Asn)	
Metode Topsis	
Nurmayanti ¹ , Merri Parida ² , M. Reka Yuansyah ³	
(STMIK Dian Cipta Cendikia kotabumi)	89-96
Media Pembelajaran Interaktif Berbasis Articulate Storyline 3 Pada Mata Kuliah Pemrograman	
Berorientasi Objek	
Dikwan Moeis ¹ , Andi Harmin ²	
(STMIK Profesional Makasar ¹²)	97-106
Penentuan Penerima Beasiswa Di Stmik Bani Saleh Dengan Perbandingan	
Metode Algoritma C4.5 Dan Knearest Neighbors	
Siti Chodijah ¹ , Mohammad Iqbal ²	
(Universitas Gunadama ¹²)	107-114
Perancangan Sistem Informasi Kehadiran Pegawai Dan Skp (Sikap)	
Pada Institut Agama Islam Negeri (Iain) Metro	
Toto Andri Puspito	
(Institut Agama Islam Negeri Metro ¹)	115-120
Kualitas Layanan Sistem Informasi Akademik (Siakad)	
Terhadap Kepuasan Mahasiswa Sebagai Pengguna	
Aidah Hami ¹ , Dyah Anggraini ²	
(Stmik Bani Saleh ¹ , Universitas Gunadarma)	121-129
Implementasi Metode Bag Of Visual Words Dalam Pengenalan Citra Masker Pada Wajah	
Komang Budiarta ¹ , I Made Budi Adnyana ² , Gede Herdian Setiawan ³	
(ITB STIKOM BALI)	130-137
Sistem Tiket Helpdesk Pada Stmik Bani Saleh	
Zaenal Mutaqin Subekti ¹ , Kresno Murti Prabowo ² , Budi ³	
(STMIK Bani Salih ¹²³)	138-144
Algoritma Naive Bayes Untuk Memprediksi Jumlah Siswa Berpotensi Drop Out	
Sidik Rahmatullah ¹ , Ngajiyanto ² , Pakarti Riswanto ³ , Arief Hendriawan ⁴	
(STMIK Dian Cipta Cendikian Kotabumi ¹²³)	145-153
Pengklasteran Risiko Covid-19 Di Riau Menggunakan Teknik <i>One Hot Encoding</i>	
Dan Algoritma <i>K-Means Clustering</i>	
Silviana ¹ , Rahmad Kurniawan ² , Alwis Nazir ³ , Elvia Budianita ⁴ ,	

Fadhillah Syafria ⁵ , Siska Kurnia Gusti ⁶ (Universitas Riau ² , Universitas Islam Negeri Sultan Syarif Kasim Riau ^{1,3,4,5,6})	154-163
Aplikasi Pengelolaan <i>E-Document</i> Sistem Penjaminan Mutu Internal Menggunakan Metode <i>User Centered Design</i> Andi Harmin ¹ , Rosnani ² (STMIK Profesional Makassar ¹²)	164-173
Game Edukasi Mengenal Kepulauan Indonesia Menggunakan <i>Unity 3d</i> Untuk Meningkatkan Minat Belajar Siswa Sekolah Dasar Tri Aditama ¹ , Ade Irma Purnamasari ² , Tati Suprapti ³ (STMIK IKMI Cirebon)	174-179
Alat Pemantau Bilik Desinfektan Untuk Pencegahan Penularan Covid 19 Dengan Internet Of Things (I.O.T) Berbasis Microcontroller Yusup Supriadi (Universitas Panca Sakti Bekasi)	180-193
Penerapan Metode <i>Fuzzy Ahp (Analytical Hierarchy Process)</i> Sebagai Sistem Pendukung Keputusan Dosen Terbaik (Studi Kasus : Stmik Pringsewu) Afrizal Martin ¹ , Bambang Suprpto ² , Sulasminarti ³ , Akni Widiyastuti ⁴ , Deny Firmansyah Kurniawan ⁵ , Henry Simanjuntak ⁶ (STMIK Pringsewu ¹ , AMIK Dian Cipta Cendikia Pringsewu ²³⁴⁵⁶)	194-207
Game Edukasi Pembelajaran Anak Usia Dini Berbasis Android Ferly Ardhy ¹ Gusnaedi Adam ² Agustinus Eko Setiawan ³ Anti Aisyah ⁴ (unversitas aisyah pring sewu, STMIK Dian Cipta Cendikia Kotabumi)	208-213
Sistem Informasi Penjualan Alat-Alat Pancing Berbasis Web Pada Toko Yoggi Bersaudara Di Talang Padang Kabupaten Tanggamus (Studi Kasus Toko Yoggi Bersaudara) Rima Mawarni ¹ , Dewi Triyanti ² , Dodi Afriansyah ³ , Yoggi Kurniawan ⁴ (STMIK Dian Cipta Cendikia Kotabumi ¹⁴ AMIK Dian Cipta Cendikia Pringsewu ²³)..	214-219
Implementasi Algoritma <i>Winnowing</i> Dalam Mendeteksi Plagiarisme Pada Tugas Mahasiswa Ida Bagus Ketut Surya Arnawa (ITB STIKOM BALI)	220-230
Sistem Pendukung Keputusan Pemilihan Guru Teladan Menggunakan Metode <i>Composite Performance Index (Cpi)</i> Pada Smk Negeri 1 Kotabumi Rustam ¹ , Pakarti Riswanto ² , Dwi Marisa Efendi ³ , Asep Afandi ⁴ , Supriyanto ⁵ , Desri Arisandi ⁶ (STMIK Dian Cipta Cendikia Kotabumi ¹²³⁴)	231-238

PENGLASTERAN RISIKO COVID-19 DI RIAU MENGGUNAKAN TEKNIK *ONE HOT ENCODING* DAN ALGORITMA *K-MEANS CLUSTERING*

Silviana¹, Rahmad Kurniawan², Alwis Nazir³, Elvia Budianita⁴, Fadhillah Syafria⁵, Siska Kurnia Gusti⁶
Universitas Riau², Universitas Islam Negeri Sultan Syarif Kasim Riau^{1,3,4,5,6}

Jl.H.R. Soebrantas no. 155 KM 18 Simpang Baru, Pekanbaru 28293², Jl. H.R. Soebrantas KM 12,5
Simpang Baru, Kec. Tampan, Kota Pekanbaru, Riau 28293^{1,3,4,5,6}

Email: 11850120868@uin-suska.ac.id¹, rahmadkurniawan@lecturer.unri.ac.id², alwis.nazir@uin-suska.ac.id³, elvia.budianita@uin-suska.ac.id⁴, fadhilah.syafria@uin-suska.ac.id⁵, siskakurniagusti@uin-suska.ac.id⁶

ABSTRAK

Coronavirus disease 2019 (COVID-19) merupakan jenis penyakit baru yang diketahui menjangkiti manusia pada Desember 2019. Kasus COVID-19 telah menyebar di seluruh penjuru dunia termasuk di Indonesia. Salah satu provinsi dengan jumlah kasus COVID-19 yang cukup tinggi adalah Provinsi Riau. Tindakan mitigasi yang tepat diperlukan untuk mencegah wabah COVID-19. Berdasarkan studi pustaka, ahli epidemiologi telah menggunakan metode *clustering* untuk mengelompokkan daerah-daerah yang terkena Pandemi COVID-19. Oleh karena itu, penelitian ini menggunakan teknik *one hot encoding* dan algoritma *k-means clustering* untuk mengelompokkan daerah yang memiliki karakteristik data yang mirip. Penelitian ini menggunakan data 12 Kabupaten di Provinsi Riau dengan tujuh fitur. Berdasarkan eksperimen, dihasilkan tiga klaster yaitu *C1 [Bengkalis, Siak, Pelalawan, Indragiri Hilir, Indragiri Hulu, Kuantan Singingi, Meranti, Dumai, Rokan Hulu, Rokan Hilir]*, *C2 [Pekanbaru]* dan *C3 [Kampar]*. Hasil klaster tersebut telah diuji dengan skor *silhouette* sebesar 0,7. Dengan demikian, dapat disimpulkan bahwa teknik *one hot encoding* dan algoritma *k-means clustering* berpotensi digunakan untuk mengelompokkan wilayah pandemi COVID-19 berdasarkan karakteristik data yang mirip.

Kata kunci : Covid-19, *One Hot Encoding*, Algoritma *K-means*

ABSTRACTS

Coronavirus disease 2019 (COVID-19) is a new type known to infect humans in December 2019. COVID-19 cases have spread throughout the world, including in Indonesia. Riau Province is one of the provinces with a fairly high number of COVID-19 cases. Appropriate mitigation measures are needed to prevent the COVID-19 outbreak. Based on a literature review, COVID-19 outbreaks are infected based on the closest distance. Epidemiologists have also used the clustering method to group the areas affected by the COVID-19 pandemic. Therefore, this study applied the one-hot encoding technique and the k-means clustering algorithm to cluster regions with similar data characteristics. Twelve districts in Riau with seven features were obtained for clustering. Based on the experimental testing results, three clusters were obtained, namely *C1 [Bengkalis, Siak, Pelalawan, Indragiri Hilir, Indragiri Hulu, Kuantan Singingi, Meranti, Dumai, Rokan Hulu, Rokan Hilir]*, *C2 [Pekanbar]*, and *C3 [Kampar]*. The results of the cluster were tested with a silhouette score of 0.7. Thus, it can be concluded that the one-hot encoding technique and the k-means clustering algorithm have the potential to be used to cluster areas of the COVID-19 pandemic based on similar data characteristics.

Keywords: Covid-19, *One Hot Encoding*, *K-means Algorithm*

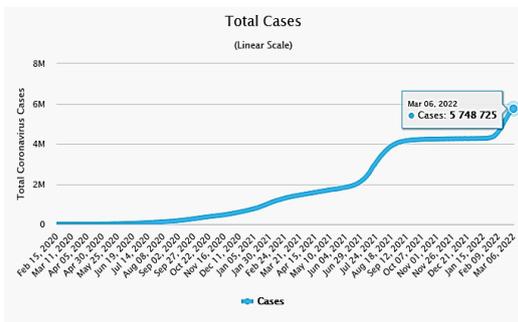
1. PENDAHULUAN

Wabah pneumonia yang belum pernah terjadi sebelumnya muncul di Kota Wuhan, provinsi Hubei di Cina pada Desember 2019. Virus corona menjangkiti jutaan orang dan menyebabkan banyak kematian. Virus ini disebut COVID-19 oleh Organisasi Kesehatan Dunia (WHO) [1] [2]

[3]. WHO resmi menetapkan virus tersebut dengan nama Novel Coronavirus pada manusia dengan sebutan *Coronavirus Diseases 2019*. Pandemi COVID-19 adalah wabah terburuk yang menyebabkan lebih dari 445.096.684 juta kasus terkonfirmasi secara global dan 5.998.301 juta kematian [4] [5].

Terlepas dari upaya pencegahan dan karantina global yang ketat, kasus COVID-19 terus meningkat di Indonesia. Kasus yang terkonfirmasi hingga 20 Maret 2022 sebanyak 5.956.561 orang terinfeksi dan 153.599 total kematian [5]. COVID-19 pertama dilaporkan di Indonesia pada 2 maret 2020. Virus ini memiliki angka kematian mencapai 8,9%. Ini merupakan yang tertinggi di Asia Tenggara [6]. Provinsi Riau merupakan salah satu provinsi di Indonesia dengan kasus terkonfirmasi hingga 20 Maret mencapai 34.012 orang terinfeksi. Total kasus kematian tercatat mencapai 4.107 orang meninggal dunia [7].

Saat ini, para peneliti dengan sungguh-sungguh berupaya untuk memprediksi penyebaran dan risiko penularan virus mematikan ini. Data jumlah kasus terkonfirmasi COVID-19 sesuai data di WHO [5] dapat dilihat pada Gambar 1 berikut:



Gambar 1. Total kasus COVID-19 di Indonesia

Virus COVID-19 bisa dicegah dan dianalisis penyebarannya. Para peneliti telah berupaya untuk memprediksi penyebaran dan risiko dari penularan virus tersebut. Para ahli epidemiologi

menggunakan pembelajaran mesin untuk memprediksi penyebaran virus menular baik pada epidemi ataupun pandemi. Dimulai dengan memprediksi dari cara penularan melalui faktor lingkungan sekitar. Sampai saat ini, informasi data yang ada belum dipublikasikan untuk bisa diteliti lebih lanjut. Oleh karena itu, prediksi penyebaran virus COVID-19 melalui kecerdasan buatan, yaitu pembelajaran mesin sangat penting untuk mempercepat studi saat ini [1].

Kumpulan data digunakan untuk menganalisis penyebaran COVID-19. Analisis risiko penyebaran COVID-19 sangat dibutuhkan terutama di Kabupaten/Kota yang terkena dampak. Penelitian ini akan menerapkan metode pembelajaran mesin tanpa pengawasan yang populer, yaitu *clustering*. Metode *clustering* mampu mengelompokkan data yang tidak berlabel berapapun jumlah fiturnya [8]. Metode ini juga sesuai digunakan pada data yang sedikit [1].

Penelitian terkait penggunaan *machine learning* untuk pengklasteran 200 negara di dunia yang terinfeksi COVID-19 pernah dilakukan oleh Rahmad Kurniawan pada tahun 2020. Pada penelitian tersebut algoritma *k-means* secara otomatis mencari *cluster* yang tidak diketahui dari banyak negara di dunia yang terinfeksi COVID-19. Hasil penelitian dengan cepat menunjukkan bahwa sekelompok $m=5$ dapat menghasilkan akurasi sebesar 97% [1].

Berdasarkan tinjauan pustaka, algoritma *k-means clustering* mampu mengelompokkan daerah yang terkonfirmasi COVID-19 berdasarkan karakteristik data yang mirip. Akan tetapi, belum ada penelitian yang menggunakan teknik *one hot encoding* dalam mentransformasi fitur non numerik untuk pengklasteran suatu daerah. Penelitian yang secara signifikan melakukan pengklasteran risiko COVID-19 sangat terbatas.

Untuk memangkas penyebaran virus COVID-19 di Provinsi Riau, salah satunya dengan mendapatkan data penyebaran COVID-19 yang sudah dikelompokkan. Oleh karena itu, pada penelitian ini akan mengelompokkan 12 Kabupaten/Kota di Provinsi Riau menggunakan teknik *one hot encoding* dan algoritma *k-means clustering*. Sehingga dapat menjadi rujukan bagi pihak terkait dan pembuat keputusan untuk pencegahan dan mitigasi risiko penularan COVID-19 yang mematikan ini untuk setiap Kabupaten/Kota yang terkonfirmasi COVID-19.

1.1 COVID-19

Coronavirus diseases 2019 atau yang dikenal dengan COVID-19 adalah virus yang menginfeksi saluran pernapasan manusia ditandai dengan gejala pada bagian pernapasan akut. Virus ini merupakan keluarga besar Coronavirus yang dapat menyerang hewan. Penularan pada virus ini terjadi jika adanya kontak antara sesama manusia. Untuk memutuskan rantai penyebaran COVID-19, pemerintah dan masyarakat menerapkan *physical distancing* dan pembatasan kegiatan masyarakat. Analisa dan pengetahuan digunakan untuk membantu pemerintah dalam menangani Kabupaten/Kota yang memiliki risiko persebaran virus yang lebih tinggi [3]. Oleh sebab itu, pengetahuan tentang data serta analisis yang tepat mengenai persebaran COVID-19 di Riau sangat diperlukan.

1.2 Teknik One Hot Encoding

One hot encoding adalah salah satu teknik untuk menyeragamkan data kategorik dan numerik. Cara kerja *one hot encoding* adalah dengan membuat sebuah *array* 1 dimensi dengan panjang sebanyak jenis fitur yang ada dan mempunyai isi biner antara 0 dan 1. *One hot encoding* dapat merepresentasikan data bertipe kategorik menjadi lebih ekspresif [9]. Pada algoritma pembelajaran *k-means* tidak dapat bekerja dengan data kategorik.

Oleh karena itu, data kategorik harus diubah menjadi data numerik yang bernilai *integer* 0 dan 1. Selanjutnya mengubah setiap nilai di dalam kolom menjadi kolom baru dan diisi dengan nilai numerik 0 dan 1 untuk fitur yang memiliki nilai kategorik tersebut. Salah satu contoh penggunaan teknik *one-hot encoding* pada fitur data status pasien dapat dilihat pada Tabel 1 berikut:

Tabel 1. Transformasi fitur data status pasien

No.	Fitur	Sebelum	Sesudah		
			JP_RawatInap	...	JP_IGD
1.	Jenis Pasien	RawatInap	1	..	0
		Rawat Jalan	0	..	0
		IGD	0	..	1

1.3 K-Means Clustering

Pada dasarnya algoritma pengelompokan *k-means* termasuk ke dalam metode *clustering* berbasis jarak yang membagi data-data ke dalam beberapa *cluster*. Metode ini hanya akan bekerja pada fitur yang bertipe numerik [10]. *K-means* mengelompokkan data berdasarkan titik pusat *cluster* terdekat dengan data melalui penentuan pusat *cluster* [11].

Pengelompokan dilakukan dengan memilih dari dataset yang digunakan secara acak. Setelah itu *k-means* akan memeriksa tiap komponen dari dataset. Mengelompokkan data yang telah dihitung jaraknya berdasarkan jarak antara dataset dan titik pusat yang menjadi sebuah *cluster* [12]. Selanjutnya akan dihitung kembali untuk menentukan pusat *cluster* hingga terbentuk titik pusat *cluster* baru. Ulangi proses tersebut sampai hasil pusat *cluster* tidak mengalami perubahan dengan hasil pusat *cluster* sebelumnya.

Menentukan jarak terdekat antara dataset dengan pusat *cluster* menggunakan persamaan *Euclidean Distance* berikut:

$$d_e = \sqrt{(x_i - s_i)^2 + (x_i - t_i)^2} \quad (1)$$

Keterangan:

- d_e : *Euclidean Distance*
- i : banyaknya objek
- (x,y) : nilai objek
- (s,t) : nilai *centroid* (titik pusat *cluster*)

Memperbaharui titik pusat *cluster* dengan persamaan berikut:

$$\bar{v}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

Keterangan:

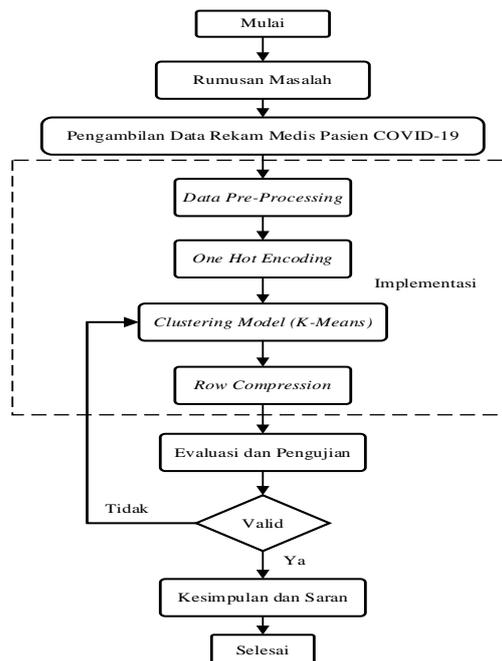
- \bar{v}_{ij} : rata-rata *cluster* ke- i untuk variabel ke- j
- N_i : jumlah data yang menjadi anggota *cluster* ke- i
- i,k : indeks dari *cluster*
- j : indeks dari variabel

Tabel 2. Contoh data rekam medis pasien COVID-19

No.	Inisial	Kabupaten	Jenis Kelamin	Usia	Jenis Pasien	...	Status Keluar
1.	SO	Pekanbaru	Laki-laki	60	IGD	...	Meninggal
2.	PA	Bengkalis	Laki-laki	70	IGD	...	Meninggal
3.	AH	Siak	Laki-laki	10	IGD	...	Meninggal
4.	NR	Siak	Perempuan	60	IGD	...	Meninggal
5.	GH	Kuansing	Laki-laki	40	IGD	...	Meninggal
6.	NL	Pekanbaru	Perempuan	70	IGD	...	Meninggal
7.	HER	Rokan Hulu	Laki-laki	60	Rawat inap	...	Meninggal
8.	KS	Siak	Laki-laki	70	Rawat inap	...	Di rawat
9.	KS	Pekanbaru	Perempuan	40	Rawat inap	...	Di rawat
...
...
36 47	Tn.M	Pekanbaru	Laki-laki	70	Rawat inap	...	Sembuh

2. METODOLOGI PENELITIAN

Metodologi penelitian dilakukan untuk mendapatkan hasil sesuai dengan yang diinginkan. Berikut tahapan-tahapan yang digunakan dalam penelitian dapat dilihat pada Gambar 2 berikut:



Gambar 2. Metodologi pengklasteran risiko COVID-19 di Provinsi Riau menggunakan teknik *One Hot Encoding* dan algoritma *K-means*

2.1 Pengumpulan Data

Metode yang dilakukan dalam mengumpulkan data dan informasi terkait penelitian diperoleh dari studi literatur dan Rumah Sakit Umum Daerah Arifin Ahmad Provinsi Riau. Data yang digunakan yaitu data rekam medis pasien terkonfirmasi COVID-19 tahun 2020-2021. Data dapat dilihat pada Tabel 2 berikut:

2.2 Implementasi

Implementasi pada penelitian ini menggunakan bahasa pemrograman *python*, dengan 3 tahap diantaranya:

1. Data pre-processing

Data pre-processing merupakan proses pembersihan untuk menangani *missing value*. Proses pembersihan dilakukan agar data bisa digunakan pada saat pemodelan. Sehingga data lebih mudah dikenali dan dapat meningkatkan proses pembelajaran dari algoritma *machine learning* yang digunakan.

2. Teknik *One Hot Encoding*

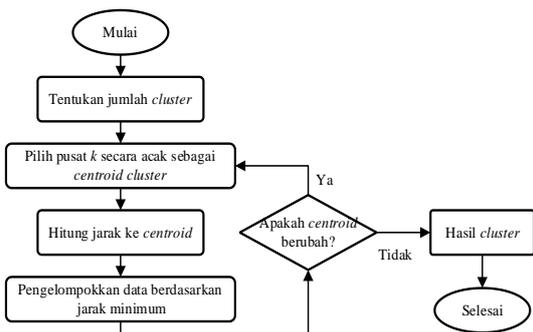
Pada tahapan teknik *one hot encoding* data yang

bertipe kategorik akan ditransformasikan atau diseragamkan menjadi data bertipe numerik yaitu bobot nilai 0 dan 1. Fitur yang memiliki beberapa parameter diubah menjadi fitur baru dengan tipe data antara 0 dan 1. Fitur baru menghasilkan 1 jika sesuai dengan nilai pada data kategorik dan menghasilkan 0 jika sebaliknya.

Proses transformasi sangat perlu dilakukan karena algoritma *k-means clustering* hanya akan bekerja pada atribut dengan nilai data bertipe numerik. Hasil *one hot encoding* pada fitur non numerik terdapat pada Tabel 3 berikut:

3. Proses Perhitungan K-means

Penerapan algoritma *k-means clustering* dilakukan pada tahap implementasi atau proses pemodelan data. Tahap implementasi *clustering* risiko COVID-19 menggunakan algoritma *k-means* dapat dilihat pada *flowchart* Gambar 3 berikut:



Gambar 3. Flowchart algoritma *k-means clustering*

Penjelasan proses perhitungan model *clustering* menggunakan *k-means* berdasarkan Gambar 3 sebagai berikut:

- Menentukan jumlah *cluster* (*k*) : Pada tahap awal menentukan jumlah *cluster* yang diinginkan, pada penelitian ini menggunakan 3 *cluster* terbaik.
- Menentukan titik pusat atau *centroid* : Menentukan secara acak pusat data sebanyak *cluster* yang dipilih.
- Menghitung jarak antar titik objek ke titik *centroid* : menghitung jarak antar titik objek ke titik pusat menggunakan *Euclidean Distance* pada persamaan 1.
- Mengelompokkan data berdasarkan jarak minimum: Mengelompokkan data yang telah dihitung jaraknya berdasarkan jarak

minimum antara data tersebut dengan *centroid* dan menjadikan sebuah kelompok *cluster* baru.

- Memperhatikan perubahan hasil *cluster*: Ulangi lagi setiap objek memakai pusat *cluster* yang baru. Jika hasil pusat *cluster* sudah tidak berubah, maka proses iterasi selesai. Kembali ke tahap 2 sampai hasil pusat *cluster* tidak mengalami perubahan.

2.3 Pengujian menggunakan Silhouette Coefficient

Tabel 3. Hasil akhir fitur kategori setelah proses *One hot encoding*

No.	Kabupaten	Jk_Pr	Jk_Laki	...	SK_Sembuh
1.	Pekanbaru	0	1	...	0
2.	Bengkalis	0	1	...	0
3.	Siak	0	1	...	0
4.	Siak	1	0	...	0
5.	Kuansing	0	1	...	0
...
...
3646.	Kampar	0	0	...	1
3647.	Pekanbaru	0	1	...	1

Pada tahap ini, akan dilakukan evaluasi dan pengujian terhadap kualitas dan efektivitas dari

model yang dibuat menggunakan teknik *one hot encoding* dan *algoritma k-means*. Untuk pengujian menggunakan *silhouette coefficient*. Tujuan dilakukannya pengujian yaitu untuk melihat kekuatan dan kualitas dari *cluster* yang terbentuk dari proses pengklasteran analisis risiko COVID-19 Provinsi Riau. Tahapan perhitungan menggunakan *silhouette coefficient* adalah sebagai berikut:

- Menghitung rata-rata jarak dari suatu objek, misalkan data ke-*i* dengan semua data berada dalam satu *cluster*. Misalkan data ke-*i* berada di *cluster* *a*, maka:

$$a(i) = \frac{1}{|A_i|-1} \sum_{j \in A_i, j \neq i} d(i,j) \quad (3)$$

Dimana a adalah jarak rata-rata pasien dan Kabupaten/Kota ke masing-masing data lain dalam *cluster* yang sama.

- b. Menghitung rata-rata jarak dari data ke- i dengan semua data di *cluster* lain dan diambil nilai terkecil.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \tag{4}$$

Dimana d adalah jarak rata-rata pasien dan Kabupaten/Kota ke *cluster* berbeda.

- c. Hitung nilai *silhouette coefficient* dengan persamaan:

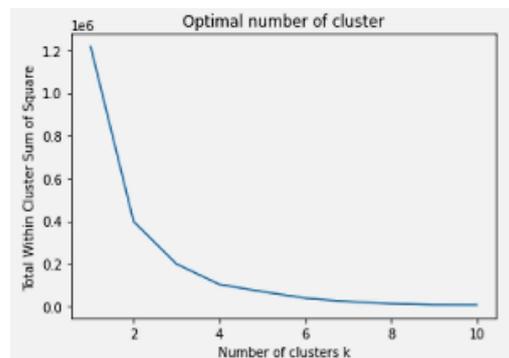
$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{5}$$

Nilai hasil *silhouette coefficient* terletak pada kisaran nilai -1 hingga 1. Jika nilai *silhouette coefficient* mendekati 1 maka semakin baik pengelompokkan data dalam satu *cluster*. Sebaliknya, jika nilai *silhouette coefficient* mendekati -1 maka kualitas pengelompokkan data didalam suatu *cluster* tidak cukup baik.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan dua eksperimen menggunakan data yang diperoleh dengan menggunakan teknik *one hot encoding* dan algoritma *k-means*.

Eksperimen pertama yaitu mengelompokkan data kasus COVID-19 di 12 Kabupaten/Kota berdasarkan karakteristik data yang mirip. Eksperimen kedua yaitu mengakumulasi menggunakan teknik *Row Compression*. Pengklasteran risiko COVID-19 menjadi tiga kategori termasuk jumlah *cluster* (k) terbaik yang ditentukan menggunakan metode *elbow* yang dapat dilihat pada Gambar 4 berikut:



Gambar 4. Penentuan jumlah *cluster* menggunakan metode *Elbow*

Berdasarkan metode *elbow* untuk menentukan jumlah *cluster* terbaik, pada sumbu x terdapat jumlah *cluster* k sedangkan pada sumbu y terdapat nilai total WCSS (*within cluster sum of square*) sebagai penentu k optimal. Hasilnya, terlihat garis mengalami penurunan yang membentuk siku di sumbu x pada saat $k=3$. Hal ini dapat menunjukkan bahwa jumlah *cluster* optimal untuk pembagian risiko COVID-19 berdasarkan 3 kategori yang merupakan nilai *cluster* terbaik.

3.1 Pengelompokkan Berdasarkan Kabupaten/Kota pada Tahun 2020-2021

Berdasarkan hasil *clustering* 12 Kabupaten di Provinsi Riau menggunakan data pada Tabel 2 diperoleh hasil *cluster* setelah dilakukan pra pemrosesan data dan teknik *one hot encoding* pada fitur yang digunakan, selanjutnya data dikelompokkan menggunakan algoritma *k-means*. Tahapan dalam pengelompokkan data diantaranya:

- a. Menentukan jumlah *cluster* menjadi tiga *cluster*.
- b. Menentukan titik pusat *cluster* awal seperti pada Tabel 4 berikut:

Tabel 4. Contoh data pusat *centroid* awal *cluster*

No.	Kabupaten	Jk_Pr	Jk_Laki	SP_konfir	...	SK_Sembuh
1.	Siak	0	1	0	...	0
2.	Pekanbaru	1	0	1	...	0
3.	Rokan Hulu	0	1	0	...	0

- c. Menghitung jarak menggunakan rumus *Euclidean distance* seperti berikut:

$$KI = \sqrt{(0-1)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-1)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (1-0)^2 + (1-1)^2 + (0-0)^2} = 2,4494$$

Selanjutnya, menghitung jarak antar data dengan tiap titik pusat sampai dengan titik pusat 3.

Tabel 5. Contoh perhitungan jarak antar data dengan ke-3 *centroid*

No.	Kabupaten	C1	C2	C3
1.	Pekanbaru	2,4494	2,2360	3,3166
2.	Bengkalis	1,4142	2,2360	3
3.	Siak	2,4494	1,7320	3
4.	Siak	1,4142	2,6457	2,6457
5.	Kuansing	2,2360	2,4494	3,1622
...
3547	Pekanbaru	3,162	2,645	2,645

Hasil dari perhitungan jarak antar *cluster* seperti

Tabel 6. Hasil pengklasteran kasus COVID-19 di Provinsi Riau

No.	Inisial	Kabupaten	Usia	Jk_Pr	Jk_Laki	SP_suspect	...	Cluster
1.	SO	Pekanbaru	60	0	1	1	...	1
2.	PA	Bengkalis	70	0	1	1	...	1
3.	AH	Siak	40	0	1	1	...	3
4.	NR	Siak	50	1	0	1	...	3
5.	GH	Kuansing	40	0	1	1	...	3
6.	NL	Pekanbaru	80	1	0	1	...	1
7.	HER	Rokan Hulu	60	0	1	1	...	1
8.	KS	Siak	40	0	1	0	...	1
9.	KS	Pekanbaru	70	1	0	0	...	3
...
...
3546	Tn.Am	Kampar	80	0	1	1	...	1
3547	Tn.M	Pekanbaru	70	0	1	0	...	1

pada Tabel 5 berikut:

- d. Mengelompokkan data yang telah dihitung jaraknya berdasarkan jarak minimum.
- e. Menghitung rata-rata tiap kelompok *cluster*.

Ulangi langkah b sampai langkah e menggunakan titik pusat baru sampai objek tidak mengalami perubahan hasil *cluster*. Provinsi Riau setelah proses teknik *one hot encoding* menggunakan algoritma *k-means*

clustering dapat dilihat pada Tabel 6 berikut:

Tabel 7. Hasil akumulasi data pada setiap Kabupaten/Kota menggunakan teknik *Row Compression*

Kabupaten	Laki-laki	Pr	Rawat Inap	Rawat Jalan	IGD	...	Sembuh
Bengkalis	57	49	81	6	25	...	47
Pekanbaru	1079	1274	1534	102	819	...	1143
Siak	79	70	111	5	38	...	15
Kampar	202	227	324	18	105	...	221
Pelalawan	48	27	63	7	12	...	42
Indragiri Hilir	25	23	31	48	17	...	23
Indragiri Hulu	15	15	26	1	4	...	0
Kuantan Singingi	47	39	66	8	20	...	41
Meranti	10	10	19	5	1	...	15
Dumai	19	23	34	3	8	...	15
Rokan Hilir	43	35	38	4	20	...	34
Rokan Hulu	46	56	83	5	14	...	52

3.2 Pengelompokan dengan Mengakumulasi Menggunakan *Row Compression*

Penelitian ini juga melakukan eksperimen dengan mengakumulasi data rekam medis pasien COVID-19 per Kabupaten/Kota menggunakan teknik *row compression*. *Row compression* atau pemampatan data dilakukan agar mendapatkan jumlah total semua pada baris data.

Selanjutnya menggabungkan dua atau lebih baris pada data menjadi satu baris data untuk melihat risiko total persebaran dari masing-masing Kabupaten/Kota. Hasil akumulasi menggunakan teknik *row compression* dapat dilihat pada Tabel

7 berikut:

Berdasarkan Tabel 7 menunjukkan Kabupaten/Kota dengan total karakteristik kasus yang tinggi dan rendah setelah dilakukan akumulasi menggunakan *row compression*. Kabupaten/Kota yang memiliki tingkat akumulasi kasus yang tinggi yaitu Kota Pekanbaru, Kabupaten Kampar dan Kabupaten Siak yang masing-masing Kabupaten/Kota termasuk kedalam kluster yang berbeda.

Kabupaten dengan akumulasi kasus yang cukup rendah yaitu Kabupaten Bengkalis, Kabupaten Indragiri Hilir, Kabupaten Indragiri Hulu, Kabupaten Kuantan Singingi, Kabupaten Meranti, Kota Dumai, Kabupaten Rokan Hilir, dan Kabupaten Rokan Hulu.

Selanjutnya dilakukan pengklasteran secara otomatis menggunakan algoritma *k-means* untuk semua fitur yang ada. Berikut hasil pengklasteran pada 12 Kabupaten/Kota seperti pada Tabel 8 berikut:

Tabel 8. Hasil pengklasteran menggunakan algoritma *k-means* pada data yang telah diproses dengan teknik *Row Compression*

No.	Kabupaten/Kota	Cluster
1.	Bengkalis	C1
2.	Pekanbaru	C2
3.	Siak	C1
4.	Kampar	C3
5.	Pelalawan	C1
6.	Indragiri Hilir	C1
7.	Indragiri Hulu	C1
8.	Kuantan Singingi	C1
9.	Meranti	C1
10.	Dumai	C1
11.	Rokan Hilir	C1
12.	Rokan Hulu	C1

4. KESIMPULAN

Penelitian ini telah berhasil mengelompokkan data kasus COVID-19 di Provinsi Riau. Ada dua eksperimen utama yang dilakukan. Pertama, eksperimen pada data rekam medis pasien dengan teknik *one-hot encoding* dan algoritma *k-means clustering*. Hasil *cluster* tersebut telah diuji dengan skor *silhouette* sebesar 0,58.

Kedua, eksperimen dengan mengelompokkan data kasus COVID-19 pada 12 Kabupaten/Kota di Provinsi Riau dengan teknik *row compression*, *one hot encoding* dan algoritma *k-means clustering*. Terdapat tiga *cluster* yang dihasilkan yaitu C1 [Bengkalis, Siak, Pelalawan, Indragiri Hilir, Indragiri Hulu, Kuantan Singingi, Meranti, Dumai, Rokan Hulu, Rokan Hilir], C2 [Pekanbaru] dan C3 [Kampar]. Hasil eksperimen kedua juga telah diuji dengan skor *silhouette* sebesar 0,7. Berdasarkan hasil *clustering* tersebut, maka daerah yang terkelompok dalam satu *cluster* dianggap memiliki karakteristik risiko kasus yang mirip. Penelitian ini diharapkan sebagai rujukan bagi pihak terkait dalam mitigasi, penanganan yang tepat dan efisien untuk kasus COVID-19 di Provinsi Riau terutama pada daerah yang terkelompok pada *cluster* yang sama.

UCAPAN TERIMA KASIH

Terima kasih kepada Rumah Sakit Umum Daerah Arifin Ahmad Provinsi Riau, lembaga riset CIDSCI (*Center of Islamic Data Science and Continuous Improvement*) Universitas Islam Negeri Sultan Syarif Kasim Riau dan Universitas Riau yang telah mendukung penelitian ini.

DAFTAR PUSTAKA

[1] R. Kurniawan, S. N. H. S. Abdullah, F. Lestari, M. Z. A. Nazri, A. Mujahidin, and N. Adnan, "Clustering and Correlation Methods for Predicting Coronavirus COVID-19 Risk Analysis in Pandemic Countries," *2020 8th Int. Conf. Cyber IT Serv. Manag. CITSM 2020*, 2020, doi:

10.1109/CITSM50537.2020.9268920.

- [2] S. R. Vadyala, S. N. Betgeri, E. A. Sherer, and A. Amritphale, "Prediction of the number of COVID-19 confirmed cases based on K-means-LSTM," *Array*, vol. 11, p. 100085, 2021, doi: 10.1016/j.array.2021.100085.
- [3] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSl>.
- [4] WHO, "WHO Coronavirus Disease (COVID-19) Dashboard," 2020. <https://covid19.who.int/> (accessed Jan. 31, 2022).
- [5] Worldometers.info, "COVID-19 Coronavirus Outbreak," *Dadax*, 2020. <https://www.worldometers.info/coronavirus/> (accessed May 09, 2022).
- [6] Henderi, M. Maulana, H. L. H. S. Warnars, D. Setiyadi, and T. Qurrohman, "Model Decision Support System for Diagnosis COVID-19 Using Forward Chaining: A Case in Indonesia," *2020 8th Int. Conf. Cyber IT Serv. Manag. CITSM 2020*, pp. 6–9, 2020, doi: 10.1109/CITSM50537.2020.9268853.
- [7] "Total Kasus COVID-19 di Riau," *riau24.com*, 2021. <https://www.riau24.com/berita/baca/1611110272-total-kasus-pasien-covid-19-di-riau-sudah-27592-pekanbaru-nyaris-13-ribu-kasus> (accessed Mar. 09, 2022).
- [8] R. Baruri, A. Ghosh, R. Banerjee, A. Das, A. Mandal, and T. Halder, "An Empirical Evaluation of k-Means Clustering Technique and Comparison," *Proc. Int. Conf.*

Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019, pp. 470–475, 2019, doi: 10.1109/COMITCon.2019.8862215.

Means untuk Pengelompokan Usia Calon Penerima Vaksin di Kab. Ngawi,” no. 2, p. 6, 2021.

- [9] F. S. Silfia, Rahmad Kurniawan, Nazruddin Safaat, Elvia Budianita, “Jurnal Teknik Informatika Atmaluhur,” *J. Tek. Inform. Atmaluhur*, vol. 6, no. 1, p. 40, 2018.
- [10] M. W. Talakua, Z. A. Leleury, and A. W. Talluta, “Cluster Analysis by Using K-Means Method for Grouping of District/City in Maluku Provinse Industrial Based on Indicators of Maluku Development Index in 2014,” *J. Ilmu Mat. dan Terap.*, vol. 11, pp. 119–128, 2017.
- [11] A. Nur, R. Kurniawan, M. Z. A. Nazri, K. Rajab, P. Papilo, and A. Mas’ari, “Solution to Traveling Freelance Teacher Problem using the Simple K-Means Clustering,” *Proc. - 2021 4th Int. Conf. Comput. Informatics Eng. IT-Based Digit. Ind. Innov. Welf. Soc. IC2IE 2021*, pp. 112–116, 2021, doi: 10.1109/IC2IE53219.2021.9649086.
- [12] S. Irna Yuniarfi, “Penerapan Algoritma K-