

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

1. TINJAUAN PUSTAKA

Pada penelitian ini, ada beberapa penelitian terdahulu yang digunakan sebagai bahan acuan dalam melakukan penelitian ini. Juniar Hutagalung dkk (2022) dengan judul penelitian “Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma *K-Means*” mengatakan penerapan algoritma K-Means dapat mempercepat dalam menentukan pengelompokan data.

Tabel 2. 1 Tinjauan Pustaka

No	Peneliti (Tahun) & Penerbit	Judul (Topik)	Kesimpulan
1	Juniar Hutagalung, Yopi Hendro Syahputra, Zohana Pertiwi Tanjung (2022). Penerbit : Jurnal Teknik Informatika dan Sistem Informasi Vol. 9, No. 1, Maret 2022, Hal. 606-620. ISSN 2407-4322 E-ISSN 2503-2933	PEMETAAN SISWA KELAS UNGGULAN MENGGUNAKAN ALGOTIRMA K-MEANS CLUSTERING.	Penerapan algoritma k-means, mampu mempercepat dalam menentukan pengelompokan siswa kelas unggulan dengan menginputkan nilai centroid dan mencari nilai terdekat sehingga menghasilkan 3 cluster yaitu: Cluster 1: Penguasaan pemrograman android sebanyak 10 orang siswa Cluster 2: Penguasaan pemrograman web sebanyak 62 orang siswa Cluster 3: Penguasaan pemrograman desktop sebanyak 48 orang siswa 2.Aplikasi yang dibangun berbasis web dapat mempermudah admin sekolah

No	Peneliti (Tahun) & Penerbit	Judul (Topik)	Kesimpulan
			dalam pengolahan data siswa dan mempermudah dalam pemodelan sistem menggunakan UML.
2	<p>Suhandio Handoko, Fauziah, Endah Tri Esti Handayani (2020).</p> <p>Penerbit : Jurnal Ilmiah Teknologi dan Rekayasa , (2020), 76-88, 25(1).</p> <p>DOI 10.35760/tr.2020.v25i1.2677 ISSN 1410-9093</p>	<p>IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING</p>	<p>Berdasarkan hasil penelitian dapat ditarik kesimpulan bahwa hasil dari metode Algoritma K-Means Clustering data mining didapatkan daerah penjualan produk yang tinggi , sedang , dan rendah. Daerah dengan penjualan produk yang rendah akan dilakukan promosi penjualan produk dan untuk daerah penjualan yang tinggi tidak diadakan promosi.</p>
3	<p>Irmanita Nasution, Agus Perdana Windarto, M. Fauzan (2020).</p> <p>Penerbit : Building of Informatics, Technology and Science (BITS, (2020), 76-83, 2(2)).</p> <p>ISSN 2684-8910 ISSN 2685-3310</p>	<p>PENERAPAN ALGORITMA K-MEANS DALAM PENGELOMPOKKAN DATA PENDUDUK MISKIN MENURUT PROVINSI</p>	<p>Jumlah record yang digunakan sebanyak 34 provinsi dengan menghasilkan 2 cluster yakni cluster tinggi sebanyak 8 provinsi dan cluster rendah sebanyak 26 provinsi. Berdasarkan hasil pengujian k-means untuk kasus presentase data penduduk miskin menggunakan tools RapidMiner versi 5.3 diperoleh hasil yang sama dengan analisis perhitungan algoritma dimana diperoleh 8 provinsi dengan</p>

No	Peneliti (Tahun) & Penerbit	Judul (Topik)	Kesimpulan
			cluster tinggi yang menjadi pusat perhatian bagi pemerintah dalam melakukan sosialisasi dan pemetaan dalam pemberian bantuan pada provinsi tersebut.
4	<p>Castaka Agus Sugianto, Ayu Hendrati Rahayu, Aditia Gusman (2020).</p> <p>Penerbit : JOINT (Jurnal of Information Technology) Vol. 02 No. 02, Agustus (2020), Hal 39-44.</p> <p>ISSN 2527-9467 E-ISSN 2656-7539</p>	<p>ALGORITMA K-MEANS UNTUK PENGELOMPOKKAN PENYAKIT PASIEN PADA PUSKESMAS CIGUGUR TENGAH</p>	<p>Berdasarkan hasil pengujian dari algoritma k-means dan algoritma k-medoids, didapat cluster model untuk algoritma k-means sebanyak 241 items pada cluster_0 atau penyakit akut dan 9 items pada cluster_1 atau penyakit tidak akut, sedangkan untuk algoritma k-medoids sebanyak 224 items pada cluster_0 atau penyakit akut dan 26 items pada cluster_1 atau penyakit tidak akut, dengan nilai Davies Bouldin untuk algoritma k-means sebesar -0.453 dan algoritma k-medoids sebesar -1.276.</p>
5	<p>Nita Mirantika, Annisa Tsamratul'Ain, Futry Diviana Agnia (2021).</p> <p>Penerbit : JURNAL NUANSA INFORMATIKA Vol. 15 No. 2, Juli 2021.</p>	<p>PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKKAN</p>	<p>Untuk klaster 1 sebagai klaster dengan jumlah penyebaran covid-19 yang paling tinggi diperoleh 2 kota/kabupaten yaitu kota Depok dan kota Bekasi. Untuk klaster 2</p>

No	Peneliti (Tahun) & Penerbit	Judul (Topik)	Kesimpulan
	<p>ISSN 1858-3911 E-ISSN 2614-5405</p>	<p>PENYEBARAN COVID-19 DI PROVINSI JAWA BARAT</p>	<p>sebagai klaster dengan jumlah penyebaran covid-19 menengah diperoleh 5 kota/kabupaten yaitu Kota Bandung, Kabupaten Bandung, Kabupaten Bekasi, Kabupaten Bogordan Kabupaten Karawang. Sisanya sebanyak 20 kota/kabupaten masuk dalam klaster 3 dengan jumlah penyebaran covid-19 yang lebih sedikit. Dengan pengelompokan penyebaran covid-19 bisamenjadibahan pertimbangan pemerintah provinsi Jawa Barat dalam mengambil keputusan untuk strategi penanganan covid-19 di Jawa Barat pada setiap klasternya. Klaster 1 dan klaster 2 perlu mendapatkan perhatian serius dan prioritas dalam penanganan covid-19 karena mempunyai jumlah penyebaran yang tinggi. Dengan membuat strategi penanganan yang tepat pada setiap klasternya diharapkan dapat</p>

No	Peneliti (Tahun) & Penerbit	Judul (Topik)	Kesimpulan
			mengakhirirantai penyebaran covid-19 di provinsi Jawa Barat

2. LANDASAN TEORI

2.1 Pendidikan

Pendidikan merupakan wadah atau tempat untuk membentuk suatu citra baik dari seorang manusia agar bisa berkembang dengan seluruh potensi yang ada di dalam dirinya. Menurut Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan juga menjelaskan bahwa Pendidikan adalah tempat atau wadah untuk mengembangkan seluruh potensi diri yang ada pada diri manusia. Pendidikan di Indonesia dibagi menjadi 2 (dua) bagian yaitu Pendidikan Formal dan Pendidikan Non-Formal.

Kualitas / Mutu Pendidikan merupakan kualitas atau ukuran baik atau buruknya proses perubahan sikap dan tata laku seseorang atau kelompok dalam usaha mendewasakan manusia melalui proses bimbingan pengajaran dan pelatihan [7]. Mutu Pendidikan dapat tercipta dengan adanya aspek pendukung yaitu peserta didik, guru, pegawai, dan sekolah.

1. Peserta Didik

Peserta didik merupakan seorang individu yang masuk ke dalam sebuah sistem pendidikan untuk mengembangkan sebuah potensi yang ada dalam dirinya melalui pengajaran sehingga menjadi individu yang berkualitas dan dapat berguna untuk orang-orang disekitarnya.

2. Guru

Guru dalam arti yang luas merupakan seorang individu yang memiliki keterampilan dan pengetahuan untuk membantu atau memberikan pengajaran kepada peserta didik agar dapat mengembangkan bakat dan potensi yang ada dalam dirinya.

3. Pegawai Sekolah

Pegawai sekolah merupakan individu yang membantu guru dalam sistem pendidikan di sekolah sehingga segala urusan diluar tugas seorang guru dapat berjalan lancar seperti administrasi sekolah dan keuangan sekolah.

4. Sekolah

Sekolah merupakan wadah atau rumah kedua bagi para siswa dimana mereka belajar melalui pengajaran yang dilakukan orang guru untuk mengembangkan potensi atau bakat yang ada di dalam dirinya.

2.2 Clustering

Clustering adalah proses mengklasifikasikan data menjadi sejumlah kelompok (*cluster*) dari data multidimensi berdasarkan ukuran kesamaan atau kemiripan[8]. Setiap himpunan bagian adalah *cluster*, sehingga objek yang didalam *cluster* mirip satu sama dengan yang lainnya dan memiliki perbedaan dengan objek dari *cluster* yang lain [9]. *Clustering* juga merupakan suatu proses untuk mengetahui kelas-kelas taksonomi atau batryologi, atau analisis topologi berdasarkan data-data yang ada. Dilihat dari sudut pandang *data mining*, *clustering* bukanlah proses klarifikasi. Karena dalam suatu proses klarifikasi, data yang ada dikelompokkan ke dalam kelas-kelas yang telah diketahui sebelumnya[10]. Ada beberapa jenis metode yang dipakai pada *Clustering* yaitu *Centroid-based Clustering*, *Density-based Clustering*, *Distribution-based Clustering*, dan *Hierarchical Clustering*.

2.3 K-Means

K-Means merupakan metode yang paling populer dan banyak digunakan pada proses pengelompokan data. Metode ini membagi data ke dalam beberapa kelompok sehingga data yang memiliki karakteristik yang sama akan dimasukkan ke dalam satu kelompok yang sama, sedangkan untuk data yang memiliki karakteristik yang berbeda akan dimasukkan ke dalam kelompok yang lain. *K-Means* merupakan salah satu metode *clustering* non-hirarki yang melakukan pengelompokan data ke dalam satu atau lebih *cluster*/kelompok Data-data yang memiliki karakteristik yang

sama akan dimasukkan ke dalam satu cluster/kelompok yang sama dan data yang memiliki karakteristik yang berbeda dimasukkan ke dalam cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil [11].

Langkah-langkah algoritma K-Means sebagai berikut:

- Menentukan jumlah *cluster* (k), lalu tetapkan pusat *cluster* sembarang.
- Menghitung jarak terdekat antara setiap data dalam *cluster* menggunakan rumus Euclidean Distance dengan persamaan:

$$Di(x_1, x_2) = \| x_2 - x_1 \| = \sqrt{\sum_{j=1}^P \{x_{2j} - x_{1j}\}^2} \quad (2,1)$$

Dimana $Di(x_1, x_2)$ merupakan jarak diantara data ke- i dan data ke- j .

P adalah dimensi data

x_{2j} adalah kordinat data x_2 pada dimensi j

x_{1j} adalah kordinat data x_1 pada dimensi j

- Kelompokkan data ke dalam suatu *cluster* dengan jarak terpendek dengan menggunakan rumus pada persamaan:

$$V_{ij} = \frac{\sum_{k=1}^{N_i} X_{kj}}{N_i} \quad (2,2)$$

V_{ij} adalah data *cluster* ke- i kolom j

X_{kj} data ke- k kolom ke- j

N_i banyaknya anggota *cluster* ke- i

- Hitung pusat *cluster* yang baru menggunakan persamaan.
- Ulangi langkah 2 sampai dengan 4 hingga sudah tidak ada lagi data yang berpindah ke *cluster* yang lain.

2.4 Purity

Purity merupakan suatu ukuran nilai kemurnian sebuah *cluster* yang telah diperoleh dari sebuah proses *clustering*. Nilai kemurnian (*purity*) sebuah *cluster* berkisaran antara 0 dan 1. *Clustering* dinilai buruk jika nilai kemurnian (*purity*) mendekati 0, dan dinilai baik jika nilai kemurnian

(*purity*) mendekati angka 1 [12]. Nilai kemurnian (*purity*) sebuah cluster dapat ditentukan dengan rumus:

$$\sum_{i=1}^k \frac{n_j}{n} Purity(j) \quad (2,3)$$

2.5 *Davies-builden Index*

Davies-builden Index merupakan metode yang digunakan untuk mengevaluasi cluster secara umum berdasarkan kuantitas dan kedekatan antar anggota *cluster* [13]. Semakin kecil nilai DBI yang didapatkan, maka semakin baik *cluster* yang didapatkan. Perhitungan DBI dapat ditentukan dengan rumus:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (2,4)$$