

BAB II

TINJAUAN PUSTAKA

2.1 State of Art

Penulisan penelitian ini didasarkan pada penelitian sebelumnya yang memiliki kemiripan pada objek penelitian. Tabel 2.1 menunjukkan beberapa penelitian yang telah dilakukan sebelumnya dan menjadi acuan dalam penelitian ini.

Tabel 2. 1 State of Art

No	Nama Peneliti	Metode yang digunakan	Kelebihan	Kelemahan
1	Bagus Muhammad Islam, dkk	<i>Clustering</i> Fasilitas Kesehatan Berdasarkan Kecamatan Di Karawang Dengan Algoritma K-Means	Kinerja yang dihasilkan dengan algoritma k-means menghasilkan nilai Davies Index Boundaries sebesar 0,109.	Proses perhitungan <i>Clustering</i> kurang jelas
2	Wardatul Maghfiroh, dkk	Penerapan Fuzzy C-Means dalam Melompokkan Kabupaten/kota Berdasarkan Fasilitas Pelayanan	Dari validasi <i>Clustering</i> yang dilakukan, didapatkan nilai sillhoute coefficient sebesar 0,695 dan	Tidak menampilkan tahun pada data yang digunakan

		Kesehatan Di Jawa Timur	nilai purity sebesar 1.	
3	Elly Nur Fitriyani, dkk	Penerapan Analisis K-Medoids Cluster untuk Mengelompokkan Wilayah di Provinsi Jawa Barat Berdasarkan Fasilitas Kesehatan Tahun 2021	Penggunaan algoritma k-medoids membuat sensitif terhadap pencilan (outlier) karena menggunakan nilai rata-rata (mean) sebagai pusat clusternya	Tidak menampilkan data dengan jelas, langsung menampilkan hasil clusternya saja
4	Aryo Satrio Wibowo1, Indurasmi Dian Mulyastuti2, dkk	Penerapan Algoritma K-Means <i>Clustering</i> Pada Jumlah Fasilitas Kesehatan Menurut Pemerintah Provinsi DKI Jakarta	K-Means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien	Tidak menampilkan hasil cluster wilayah dengan jelas.
5	hd Gading Sadewo	Pemanfaatan Algoritma <i>Clustering</i> Dalam Mengelompokkan	Hasil data yang ditampilkan cukup lengkap	Data yang digunakan berasal dari tahun

		Jumlah Desa/Kelurahan Yang Memiliki Sarana Kesehatan Menurut Provinsi Dengan K-Means		2008,2011, dan 2014 yang yang dirata ratakan.
--	--	---	--	---

Berdasarkan State of Art tersebut dapat disimpulkan penelitian sebelumnya *Clustering* digunakan untuk membantu analisis kebutuhan fasilitas kesehatan suatu wilayah sehingga pada penelitian ini juga dapat digunakan untuk menganalisis kebutuhan fasilitas kesehatan di Jawa Barat

2.2 Data mining

Data Mining merupakan proses pengekstrasian dari sekumpulan data yang sangat besar menggunakan algoritma dan Teknik penarikan dalam bidang statistic, pembelajaran mesin dan sistem manajemen basis data. Data yang diolah dengan menggunakan data mining akan menghasilkan data baru yang bersumber dari data sebelumnya. Berdasarkan tugasnya data mining terbagi menjadi 6 metode yaitu description, estimation, prediction, classification, *Clustering*, serta association[4][5]. Tujuan utama data mining adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki[6].

2.3 Machine Learning

Machine learning adalah cabang AI yang paling umum digunakan. Machine learning memungkinkan mesin untuk belajar menggunakan algoritma dari data dan membuat keputusan berdasarkan pola-pola yang ditemukan dalam data tersebut. [7]. Machine Learning melibatkan penggunaan metode matematika dan statistik untuk memungkinkan mesin belajar dari data masa lalu. Ini dapat dikategorikan ke dalam beberapa keluarga teknik, termasuk diawasi, semi-diawasi, tidak diawasi, penguatan, dan pembelajaran mendalam [8].

2.4 Unsupervised Learning

Unsupervised learning adalah bagian dari machine learning yang tidak bergantung pada data berlabel. Sebaliknya, ia memanfaatkan struktur yang melekat dan redundansi data input untuk menemukan pola, mengompres informasi, dan mengidentifikasi struktur input yang signifikan melalui proyeksi nonlinier dan peta yang mengatur sendiri. Teknik umum dalam kategori ini termasuk analisis komponen utama (PCA), pengelompokan k-means, dan metode pengelompokan lainnya. Pendekatan ini dapat secara otomatis mengklasifikasikan data mikroskop video yang luas mengenai pergerakan kolektif agen berkerumun. Ketika diintegrasikan dengan analisis model numerik, ini membantu dalam menjelaskan perilaku berkerumun yang diamati. Drescher et al. menggunakan metode pengelompokan lingkungan t-stochastic (t-SNE) dan k-means untuk mengidentifikasi fase dinamis kawanan bakteri dari satu set lengkap pengamatan statistik yang tidak berlabel yang ditangkap oleh mikroskop adaptif, menjembatani reaksi biokimia sel in vivo dengan morfologi jaringan pada skala ruang dan waktu [9]

2.5 Fasilitas Kesehatan

Fasilitas kesehatan merupakan sarana prasarana yang digunakan untuk mendukung proses pelayanan kesehatan terhadap masyarakat. Fasilitas kesehatan dapat berupa rumah sakit, klinik, puskesmas, apotek dan lain lain.

Penyediaan Fasilitas Pelayanan Kesehatan merupakan tanggung jawab Pemerintah Pusat dan Pemerintah Daerah sesuai dengan ketentuan Undang-Undang Nomor 36 Tahun 2009 tentang Kesehatan yang menyatakan bahwa Pemerintah bertanggung jawab atas ketersediaan Fasilitas Pelayanan Kesehatan bagi masyarakat untuk mencapai derajat kesehatan yang setinggi-tingginya [1].

2.6 *Clustering*

Clustering/pengelompokan adalah proses membagi data dalam suatu kumpulan menjadi beberapa kelompok yang mempunyai data serupa dengan data dari kelompok lain. Pengelompokan adalah alat yang berguna dalam ilmu data. Ini adalah metode untuk menemukan struktur cluster dalam kumpulan data yang dicirikan oleh kesamaan terbesar dalam cluster yang sama dan ketidaksamaan terbesar di antara cluster yang berbeda [10]. Kelas yang dikelompokkan disebut cluster. Cluster merupakan kumpulan data yang mempunyai persamaan dengan satu cluster dan perbedaan dengan cluster lainnya. *Clustering* adalah metode Data mining *unsupervised* yang bertujuan untuk menemukan karakteristik umum pola data dalam database [11].

Pendekatan yang dapat digunakan untuk mengembangkan *Clustering* adalah partisi. Metode partisi mendistribusikan data dalam sebuah cluster sedemikian rupa sehingga setiap cluster memiliki setidaknya satu data dan setiap data dimasukkan ke dalam salah satu grup. Pengklasteran ini berbeda dengan klasifikasi, perbedaannya yaitu tidak adanya variabel target dalam pengklasteran. Potensi *Clustering* adalah dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola.

2.7 *K-Means Clustering*

K-Means Clustering adalah metode clusterisasi yang digunakan secara partitioning yang memisahkan objek data ke dalam *cluster* yang berbeda berdasarkan jarak kedekatannya data pada setiap *cluster*-nya. Mempartisi objek ke dalam cluster (K) yang saling terpisah dilakukan dengan cara sedemikian rupa sehingga objek-objek di dalam setiap cluster tetap sedekat mungkin sedekat mungkin satu sama lain tetapi sejauh mungkin dari objek di cluster lainnya [12].

Metode k-means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama

satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain[13].

Penggunaan algoritma ini dalam proses *Clustering* bergantung pada data yang diperoleh dan kesimpulan yang diambil diakhir proses [14]. Dalam algoritma K-means terdapat beberapa aturan yaitu:

1. Berapa jumlah cluster yang perlu dimasukkan
2. Hanya memiliki atribut berjenis numeric (angka).

Algoritma K-means pada dasarnya hanya akan mengambil sebagian dari banyaknya objek yang didapatkan untuk kemudian dijadikan sebagai pusat cluster(centroid) awalnya. Untuk menentukan pusat *cluster* akan dipilih secara acak dari populasi data yang ada[15]. Kemudian, dengan algoritma K-means akan diuji setiap objeknya dari populasi data dan akan menandai objek tersebut kedalam salah satu *cluster* pusat yang sudah ditentukan sebelumnya. Penentuan objek terhadap pusat cluster ini ditentukan berdasarkan jarak kedekatannya, semakin dekat jaraknya maka objek tersebut akan dimasukkan terhadap *cluster* tersebut. Selanjutnya posisi pusat cluster akan dihitung lagi sampai objek data tergolong kedalam setiap cluster dan akhirnya akan terbentuk sebuah *cluster* baru [14].

Berikut merupakan langkah-langkah untuk mengelompokan dengan metode K-means:

1. Tentukan nilai K jumlah cluster pada data set.
2. Menentukan nilai pusat(*centroid*). Nilai pada tahap awal akan dilakukan tentukan secara random, lalu saat tahap iterasi akan digunakan rumus berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (2.1)$$

Keterangan :

V_{ij} = centroid atau rata-rata cluster ke-i dari variabel j

N_i = jumlah data yang menjadi anggota cluster ke-i

i,k = indeks dari cluster

j = indeks dari variabel

X_{kj} = nilai data ke-k yang ada pada cluster tersebut dalam variabel ke-j

3. Menghitung jarak antara titik centroid dengan titik tiap objek. Rumus yang digunakan yaitu:

$$D_e = \sqrt{\sum(x_i - y_i)^2} \quad (2.2)$$

dimana:

D_e = Euclidean distance

i = banyaknya objek

x = merupakan koordinat objek

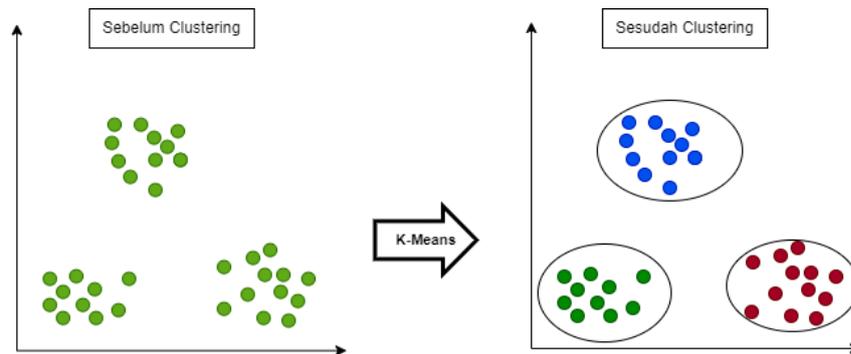
y = merupakan koordinat centroid

4. Pengelompokan objek

dalam menentukan anggota cluster akan menggunakan jarak minimum objek atau jarak ke centroid terdekat. Nilai yang diperoleh dalam keanggotaan data pada jarak matriks adalah 0 atau 1. Dimana 1 merupakan data yang masuk kedalam cluster dan nilai 0 merupakan data yang masuk kedalam cluster lain.

5. Perulangan

Perulangan akan dilakukan kembali dari tahap ke-2 hingga mendapatkan nilai centroid yang dihasilkan tetap dan anggota cluster tidak berpindah ke cluster lainnya.



Gambar 2. 1 Gambaran algoritma K-means

2.8 Validasi Cluster

Validasi *cluster* adalah metode yang digunakan untuk melakukan evaluasi hasil dari sebuah algoritma *Clustering* yang telah dilakukan dengan tujuan untuk memperoleh jumlah *cluster* yang terbaik. Validasi hasil *Clustering* dapat menggunakan beberapa indeks diantaranya yaitu:

1. Metode *Elbow*

Metode *Elbow* merupakan salah satu metode untuk menentukan jumlah *cluster* yang tepat melalui persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik.

Jika nilai *cluster* pertama dengan nilai *cluster* kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka jumlah nilai *cluster* tersebut yang tepat [16].

Nilai dari metode *elbow* ini menggunakan perhitungan *Sum Squared Error (SSE)*. SSE merupakan kriteria termudah untuk mengukur *Clustering*. Adapun perhitungan SSE [17] sebagai berikut:

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - c_k\|^2 \quad (2.3)$$

Keterangan

K = banyak *cluster*

x_i = data ke- i

S_k = himpunan dari elemen *cluster k*

c_k = rata-rata (pusat) dalam *cluster k*

2. *Davies Bouldin Index*

Davies Bouldin Index (DBI) Merupakan salah satu uji validitas internal pada metode pengelompokan berbasis partisi berdasarkan jumlah dari kedekatan data terhadap centroid dari cluster yang diikutinya dan diantara dua cluster diukur dengan kedekatan dua centroid cluster. DBI bertujuan untuk memaksimalkan jarak inter-cluster, Dengan menggunakan Davies Bouldin Index suatu cluster akan dianggap memiliki skema *Clustering* yang optimal semakin kecil nilai DBI maka semakin baik pula hasil dari cluster tersebut [18]. Berikut merupakan beberapa langkah DBI:

Sum of Square Within cluster (SSW) Sum of Square Within cluster (SSW) merupakan persamaan yang digunakan untuk mengetahui matriks kohesi dalam sebuah cluster ke-i yang dirumuskan sebagai berikut :

$$SSWi = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (2,10)$$

Sum of Square Between cluster (SSB) Sum of Square Between cluster (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar cluster yang dihitung menggunakan persamaan sebagai berikut:

$$SSBi, j = d(c_i, c_j) \quad (2,11)$$

Setelah nilai SSW dan SSB diperoleh, kemudian dilakukan pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara cluster ke-i dan cluster ke-j. Nilai rasio dihitung menggunakan persamaan sebagai berikut :

$$R_{ij} = \frac{SSWi + SSWj}{SSBij} \quad (2,12)$$

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai Davies-Bouldin Index (DBI) dengan menggunakan persamaan berikut

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (2.13)$$

2.9 Python

Python merupakan salah satu bahasa pemrograman tingkat tinggi dan bersifat bersifat bertujuan umum (*general-purpose*). Bahasa pemrograman ini dirancang oleh seseorang bernama Guido van Rossum pada tahun 1980[19]. bahasa ini tidak memiliki dukungan khusus untuk struktur data atau algoritma ilmiah, tidak seperti banyak platform komputasi mapan lainnya pada saat itu. Namun terdapat keunggulan pada bahasa ini, seperti kemampuannya untuk membungkus pustaka C dan Fortran, dan kemudian menggerakkan perpustakaan tersebut secara interaktif. Dengan demikian, dari bahasa python dapat memperoleh akses ke berbagai macam komputasi yang ada tanpa harus memikirkan konsep pemrograman tingkat rendah seperti manajemen memori.

2.10 Laravel

Laravel adalah sebuah framework untuk membuat aplikasi berbasis web dengan bahasa pemrograman PHP. Laravel merupakan salah satu framework PHP yang paling diminati dalam pengembangan aplikasi back end dengan menggunakan bahasa pemrograman PHP. Laravel memiliki banyak fitur yang sangat mendukung dalam mempercepat pembuatan aplikasi atau sistem informasi berbasis web, dimana laravel dapat digunakan baik sebagai back end dan front end atau hanya digunakan sebagai back end saja[20]. Laravel dikembangkan oleh Taylor Otwell dan digunakan untuk mengembangkan aplikasi berbasis web yang menerapkan sebuah pola yaitu MVC. Struktur MVC yang diterapkan laravel ini agak berbeda

dari MVC yang pada umumnya. Pada Laravel memiliki fitur routing yang digunakan untuk menghubungkan antara request user dan sebuah controller yang menerimanya. Sehingga controller tidak bisa langsung dapat menerima sebuah request tertentu[21].

2.11 MYSQL

MySQL merupakan salah satu aplikasi yang digunakan untuk mengelola database dengan menggunakan sebuah sintak Structured Query Language atau SQL sebagai dasar untuk mengelola akses data dari databasenya. MySQL merupakan aplikasi yang bebas dipakai karena memiliki lisensi gratis atau GPL (General Public License). MySQL juga termasuk dalam jenis manajemen database yang RDBMS atau Relational Database Management System yang membuat istilah tabel, baris dan kolom sering dipakai pada penggunaan MySQL. Database dalam MySQL mengandung beberapa tabel yang memiliki masing-masing atribut spesial yang menjadi pembeda antara satu tabel dengan tabel yang lainnya.[21]