

# Penerapan Algoritma K-Medoids untuk Pengelompokan Penyakit di Pekanbaru Riau

Tri Juninda<sup>1</sup>, Mustakim<sup>2</sup>, Elvia Andri<sup>3</sup>

<sup>1,2,3</sup> Puzzle Research Data Technology (Predatech) Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau,

<sup>1,2,3</sup> Program Studi Sistem Informasi Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl. HR. Soebrantas No. 155 Simpang Baru, Tampan, Pekanbaru, Riau – Indonesia 28293

e-mail: <sup>1</sup>11653200095@students.uin-suska.ac.id,

<sup>2</sup>mustakim@uin-suska.ac.id, <sup>3</sup>11653200497@students.uin-suska.ac.id

## Abstrak

Penyakit merupakan suatu kondisi yang membuat seseorang mengalami penurunan fungsi tubuh dan membuat produktifitas aktivitas menurun. Berdasarkan Riset Kesehatan Dasar tahun 2013 mencatat terdapat beberapa penyakit dengan prevalensi tertinggi di Indonesia yaitu hipertensi, penyakit sendi, dan hepatitis B. Masih rendahnya kebiasaan hidup sehat dan menjaga lingkungan masyarakat membuat kondisi kesehatan di Indonesia cukup mengkhawatirkan. Lingkungan dapat mempengaruhi timbulnya penyakit dominan pada suatu daerah. Untuk itu perlu adanya pengelompokan penyakit berdasarkan daerah tertentu untuk mengetahui penyakit dominan yang ada pada lingkungan tersebut. Pengelompokan penyakit menggunakan algoritma K-Medoids dapat menghasilkan kelompok terbaik dengan penyakit dominan dalam cluster tersebut. Berdasarkan penelitian yang dilakukan terhadap data penyakit yang sering diderita di Pekanbaru Riau didapatkan hasil 4 cluster sebagai pengelompokan terbaik dengan nilai Davies Bouildien Indeks sebesar 0,043. Pada cluster 1 didapatkan 420 record dengan penyakit dominan adalah Maag sebesar 44,39%, cluster 2 didapatkan 349 record dengan penyakit dominan adalah Diare dan Sakit Perut sebesar 16,98%, pada cluster 3 didapatkan 794 record dengan penyakit dominan adalah Batuk dan Pilek sebesar 65,21% dan pada cluster 4 didapatkan 1248 record dengan penyakit dominan adalah Batuk dan Pilek sebesar 54,10%. Hasil dari penelitian ini menunjukkan bahwa algoritma K-Medoids mampu melakukan pengelompokan terhadap penyakit di Pekanbaru Riau.

**Kata kunci:** Pengelompokan, K-Medoids, Penyakit, Pekanbaru Riau

## Abstract

Disease is a condition in which a person suffers from bodily function and makes productivity activity fall. According to basic health research in 2013, there have been some diseases with the highest prevalence in Indonesia is hypertension, joint disease, and hepatitis B. Poor track of health habits and keeping the environment making health conditions in Indonesia worrisome. The environment can affect an area's dominant disease. It requires a clustering of disease-based areas to determine the dominant environmental diseases. Disease clustering using K-Medoids algorithms can produce the best cluster with dominant diseases in the cluster. Based on research done on the disease data suffered frequently in new weeks we've got a four-cluster of best clustering with Davies Bouildien Indeks of 0.043. In the first cluster earned 420 records with a dominant illness is a Ulcer of 44.39%, the second cluster got a 349 record with dominant diseases is Diarrhea and Abdominal Pain by 16.98%, the third cluster had a 794 record with the dominant illness was Cough and Colds at 65.21% and the fourth cluster it's 1248 record with the dominant disease is Cough and a Common Cold by 54.10%. Results from this study suggest that K-Medoids algorithms are capable of clustering against disease in the Pekanbaru Riau.

**Keywords:** Clustering, K-Medoids, Disease, Pekanbaru Riau

## 1. Pendahuluan

Penyakit merupakan suatu kondisi dimana tubuh ataupun pikiran mengalami ketidaknyamanan atau disfungsi terhadap orang yang dipengaruhinya. Menurut Unicef beberapa faktor yang mempengaruhi kesehatan adalah faktor keturunan, faktor lingkungan, faktor pelayanan kesehatan, dan faktor perilaku. Penyakit dapat menjangkit seseorang baik tua maupun muda, laki-laki maupun perempuan. Berdasarkan Riset Kesehatan Dasar (Riskesdas) tahun 2013 mencatat terdapat beberapa penyakit dengan prevalensi tertinggi di Indonesia yang menyebar di provinsi-provinsi di Indonesia. Penyakit yang berprevalensi tertinggi tersebut adalah hipertensi sebesar 25,8%, penyakit sendi sebesar 24,7%, penyakit hepatitis B sebesar 21,8%, stroke sebesar 12,1%, dan kurang gizi sebesar 19,6% [1].

Kondisi kesehatan di Indonesia yang mengkhawatirkan disebabkan karena pola hidup sehat masyarakat yang masih buruk, karena rendahnya kesadaran menjaga lingkungan dan mengatur pola makan [1]. Lingkungan dapat menjadi pengaruh besar dalam penyakit yang diderita

setiap individu, dimana setiap daerah juga dapat memiliki penyakit dominan tergantung demografi wilayah.

Pengetahuan tentang *clustering* penyakit sangat bermanfaat dalam penanggulangan penyakit karena dapat memberikan informasi wilayah dengan risiko suatu penyakit yang tinggi [2]. Dengan adanya informasi mengenai daerah pemberantasan penyakit dan potensi adanya penyakit membuat pihak terkait dapat mengarahkan paramedis melakukan usaha pemberantasan penyakit di daerah tersebut [3].

Penelitian sebelumnya tentang pengelompokan penyakit menggunakan algoritma *K-Means* menjelaskan bahwa metode *clustering* dapat digunakan untuk mengelompokkan penderita penyakit di kota Pekanbaru, dari hasil pengelompokan tersebut akan didapatkan pola pengelompokannya yang dapat memberikan informasi kepada pihak pelayanan kesehatan untuk dapat mengantisipasi penyakit pada daerah tersebut [2].

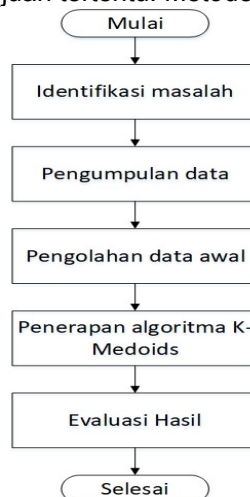
Pengelompokan (*clustering*) merupakan suatu teknik yang digunakan untuk *statistical data analysis*, dalam berbagai bidang salah satunya *data mining* [4] dan bertujuan untuk membagi data menjadi *cluster* [5][23]. Salah satu metode *clustering* adalah *K-Medoids* yang merupakan suatu algoritma untuk menemukan *medoids* didalam sebuah kelompok (*cluster*) [6] dan dinilai memiliki karakteristik penting dibanding *K-Means* karena dapat menghitung *medoids* menggunakan frekuensi kemunculannya [7]. *K-Medoids* juga memberikan banyak informasi pada proses pembuatan *cluster* [8].

Terkait penggunaan algoritma *K-Medoids* terdapat penelitian tentang Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau menunjukkan hasil bahwa kedua algoritma tersebut tidak menunjukkan perbedaan yang signifikan dalam pengelompokan data, tetapi algoritma *K-Medoids* dapat menanggulangi kelemahan algoritma *K-Means* yang sensitif terhadap *outlier* [9]. Penelitian lainnya tentang *K-Medoids Clustering Based on MapReduce and Optimal Search of Medoids* menyimpulkan bahwa pengelompokan menggunakan *K-Medoids* berdasarkan *MapReduce* dapat digunakan untuk menemukan *medoid* dari *dataset* secara efisien dan efektif [10]. Selanjutnya penelitian yang dilakukan oleh Marlina, dkk (2018) tentang Implementasi Algoritma *K-Medoids* dan *K-Means* untuk Pengelompokan Wilayah Sebaran Cacat pada Anak menghasilkan nilai validitas *Silhouette Coefficient K-Medoids* sebesar 0.5009, sedangkan nilai validitas *K-Means* sebesar 0.1443. Hal tersebut menunjukkan bahwa algoritma *K-Medoids* lebih baik dalam melakukan pengelompokan data sebaran anak cacat [11].

Dalam penelitian ini menerapkan algoritma *K-Medoids* untuk pengelompokan penyakit di Pekanbaru Riau, dengan memiliki empat atribut yaitu kecamatan, sakit yang sering diderita, usia dan jenis kelamin. Dari data tersebut akan ditentukan *cluster* terbaik menggunakan validitas *Devies Bouildien Indeks (DBI)*. Penelitian ini diharapkan mampu membantu pemerintah serta instansi terkait tentang pola penyakit yang sering diderita masyarakat Pekanbaru berdasarkan *cluster* terbaik menggunakan algoritma *K-Medoids*.

## 2. Metode Penelitian

Metode penelitian adalah cara ilmiah untuk memperoleh dan mengumpulkan data-data yang dibutuhkan dengan fungsi dan tujuan tertentu. Metode penelitian dapat dilihat pada gambar 1.



Gambar 1. Metode Penelitian

Metode penelitian dimulai dengan identifikasi masalah, lalu melakukan pengumpulan data-data yang dibutuhkan berdasarkan permasalahan yang ada. Selanjutnya data tersebut diolah dengan proses *cleaning*, transformasi, dan normalisasi, setelah itu dilakukan penerapan algoritma *K-Medoids* sesuai dengan langkah-langkah yang ada pada algoritma tersebut sehingga didapatkan hasil dan analisis dari data yang telah diolah.

## 2.1. Data Mining

*Data mining* merupakan istilah yang digunakan dalam penemuan tentang pengetahuan di *database*. *Data mining* menggunakan teknik statistik, kecerdasan buatan, matematika, dan *machine learning* untuk mengidentifikasi informasi yang berguna dan pengetahuan yang berhubungan dari berbagai *database* besar [12]. Banyak orang mengartikan data mining sebagai persamaan dari istilah yang populer seperti *Knowledge Discovery Data* (KDD), dan yang lain mengartikan *data mining* hanya sebagai langkah penting dalam proses penemuan pengetahuan [13]. *Data mining* memiliki karakteristik sebagai berikut [14]

1. *Data mining* berkaitan dengan suatu penemuan tersembunyi dan pola data yang belum diketahui sebelumnya.
2. *Data mining* biasanya menggunakan data yang besar. Data yang besar akan membuat hasil lebih akurat dan terpercaya.
3. *Data mining* berguna untuk membuat keputusan kritis, terutama dalam strategi.

## 2.2. Clustering

*Clustering* merupakan teknik dalam suatu pengelompokan data yang dibagi dalam beberapa *cluster* sesuai dengan tingkat kemiripannya [15]. *Clustering* berfungsi untuk mempartisi objek yang ada kedalam sebuah kelompok (*cluster*) yang memiliki kemiripan satu sama lain dibandingkan dengan kelompok (*cluster*) lain [16]. Dengan menggunakan *clustering* dapat mengidentifikasi daerah yang padat, menemukan pola-pola distribusi secara keseluruhan, serta menemukan keterkaitan antara atribut-atribut data [17].

## 2.3. K-Medoids

*K-Medoids* merupakan suatu algoritma yang digunakan untuk menemukan *medoids* didalam sebuah kelompok (*cluster*) yang merupakan titik pusat dari suatu kelompok (*cluster*). Algoritma *K-Medoids* lebih baik dibandingkan dengan *K-Means* karena pada *K-Medoids* kita menemukan *k* sebagai objek yang representatif untuk meminimalkan jumlah ketidaksamaan objek data, sedangkan pada *K-Means* menggunakan jumlah jarak *euclidean distances* untuk objek data [6] [18]. Langkah-langkah algoritma *K-Medoids* sebagai berikut [19]:

1. Inisialisasi pusat *cluster* sebanyak *k* (jumlah *cluster*).
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan ukuran jarak *Euclidean Distance* dengan persamaan:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; 1, 2, 3, \dots, n \quad (1)$$

3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat medoid baru.
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoid baru.
5. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika  $S < 0$ , maka ganti objek dengan data *cluster* untuk memperoleh sekelompok *k* objek yang baru sebagai medoid.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

Untuk mendapatkan jumlah *k* yang ada pada *clustering K-Medoids*, bisa menggunakan beberapa jenis validitas yang ada, seperti *Davies Bouldin Index* (DBI) [18]. Dalam perhitungan menggunakan DBI, terdapat *Sum of Square Within Cluster* dan *Sum of Square Between Cluster*. Matrik kohesi dalam sebuah kelompok (*cluster*) ke-*i* disebut *Sum of Square Within Cluster* (SSW). Untuk menghitung *Sum of Square Within Cluster* (SSW) dapat dilihat pada persamaan 2.

$$d(x_j, y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$SSW_i = \frac{1}{n} \sum_{j=1}^n d(i, j)^2 \quad (2)$$

Persamaan tersebut adalah jumlah yang terdapat pada *cluster* ke-*i*. selain *Sum of Square Within Cluster*, terdapat *Sum of Square Between Cluster* (SSB) yang merupakan metrik separasi antara dua buah *cluster* (misalnya *cluster i* dan *j*). Untuk menghitung SSB dapat dilihat pada persamaan 3.

$$SSB_{i,j} = d(y_i, y_j) \quad (3)$$

Selanjutnya untuk rasio perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j* didefinisikan dengan  $R_{i,j}$ . Nilai tersebut diperoleh dari komponen kohesi dan seperasi. Untuk menghitung  $R_{i,j}$  dapat dilihat pada persamaan 4.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (4)$$

Kemudian untuk menghitung DBI yang dilakukan terhadap nilai *k*. *k* merupakan jumlah *cluster* yang digunakan. Untuk menghitung DBI dapat dilihat pada persamaan 5.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} (R_{i,j}) \quad (5)$$

Untuk menentukan jumlah *k* yang baik dan cocok dalam proses *clustering* data menggunakan *K-Medoids* dapat dipilih berdasarkan nilai DBI terkecil.

#### 2.4. Penyakit di Pekanbaru Riau

Tingginya morbiditas dan mortalitas di Indonesia salah satunya di Provinsi Riau membuat masyarakat mudah terserang penyakit [20]. Untuk menangani hal tersebut perlu untuk mengetahui pola penyakit agar dapat menganalisa besaran permasalahan kesehatan yang dihadapi agar nantinya dapat dilakukan rencana kebutuhan obat, upaya promotif dan preventif.

#### 3. Hasil dan Analisis

Pengelompokan dilakukan dengan membagi kecamatan yang ada di Pekanbaru berdasarkan sakit yang sering diderita, usia, dan jenis kelamin. Pada penelitian ini diketahui data sebanyak 2811 *record*. Tujuan dilakukan penelitian ini adalah untuk mengelompokkan kecamatan berdasarkan sakit yang sering di derita, usia, dan jenis kelamin. Teknik pengumpulan data dilakukan dengan menggunakan *kuesioner online* yang diisi oleh masyarakat Pekanbaru. Terdapat 4 atribut dan 2811 *record*. Adapun data awal dapat dilihat pada tabel 1.

Tabel 1. Data Awal

No	Kecamatan	Sakit yang Sering di Derita	Usia	Jenis Kelamin
1	Kecamatan Tampan	Demam/panas	18	Laki-laki
2	Kecamatan Tampan	Lain-lain	20	Laki-laki
3	Kecamatan Bukit Raya	Batuk, pilek	18	Laki-laki
4	Kecamatan Tampan	Batuk, pilek	19	Laki-laki

....	....	....	....	....
....	....	....	....	....
281	Kecamatan	Diare	18	Laki-laki
0	Tampan			
2811	Kecamatan	Diare	19	Perempuan
	Tampan			

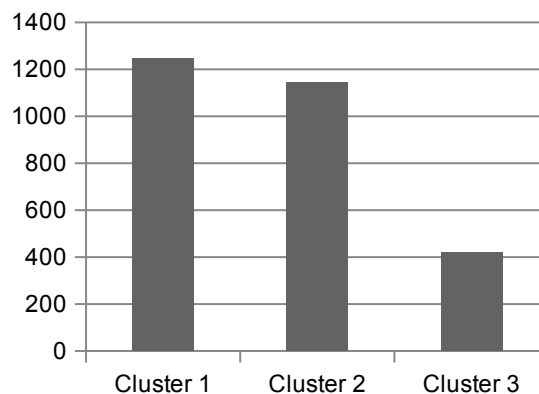
### 3.1. Pengolahan Data Awal

Data yang sudah dikumpulkan akan diolah dalam beberapa tahap sehingga menjadi sebuah *dataset* yang akan diimplementasikan terhadap metode-metode *data mining*. Semua *record* yang ada akan digunakan dalam tahap pengolahan data. Pengolahan data yang dilakukan yaitu *data preprocessing*. *Data preprocessing* yaitu pembersihan dan persiapan data untuk menghilangkan inkonsistensi pada data, data tidak lengkap dan *redundant* yang terdapat pada data awal. *Data preprocessing* juga melakukan perubahan dimana data berbentuk *string* diubah menjadi *numeric*.

### 3.2. Penerapan K-Medoids

Algoritma *K-Medoids* biasa disebut sebagai *partitioning around medoids*, yang merupakan varian dari metode *K-Means*. Hal ini didasarkan pada penggunaan *medoids* bukan dari pengamatan *mean* yang dimiliki setiap *cluster*, yang bertujuan untuk mengurangi sensitivitas dari partisi yang dihasilkan sehubungan dengan nilai-nilai ekstrim yang ada pada *dataset* [21].

Algoritma *K-Medoids* merupakan suatu algoritma yang mengatasi kelemahan Algoritma *K-Means* yang sensitif terhadap *outlier* karena objek dengan suatu nilai yang besar mungkin menyimpang dari distribusi data [22]. Untuk perhitungan menggunakan algoritma *K-Medoids* dapat mengikuti langkah 1 sampai 6 seperti yang sudah dijelaskan sebelumnya. Dengan melakukan percobaan tiga *cluster* ( $k=3$ ), sehingga didapatkan grafik perbandingan jumlah item 3 *cluster* seperti pada gambar 2.



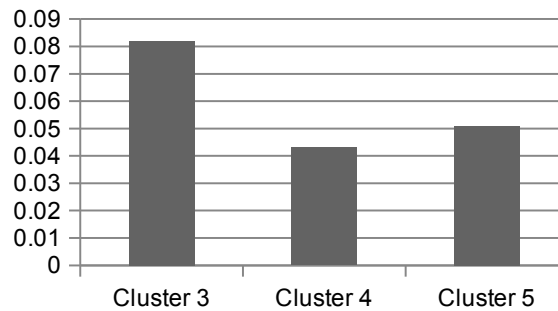
Gambar 2. Perbandingan jumlah item 3 cluster

Gambar 2 merupakan perbandingan jumlah *item* percobaan 3 *cluster*, didapatkan pada *cluster* 1 sebanyak 1248 *item*, pada *cluster* 2 sebanyak 1143 *item* dan pada *cluster* 3 sebanyak 420 *item*. Setelah melakukan langkah-langkah penyelesaian dalam algoritma *K-Medoids*, selanjutnya melakukan validasi dengan *Davies Bouldien Indeks* (DBI) untuk menentukan jumlah  $k$  terbaik dalam pengelompokan data tersebut. Didapatkan nilai DBI seperti pada tabel 2.

Tabel 2. Nilai DBI

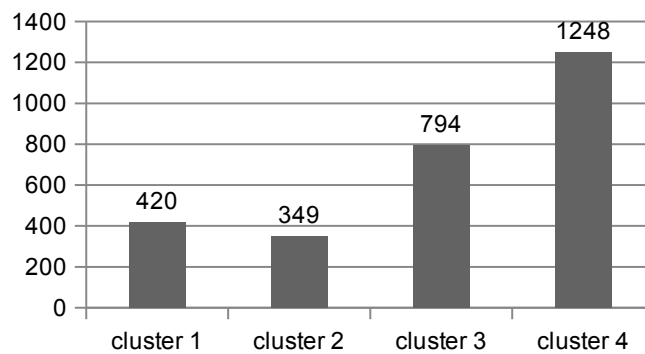
Cluster	DBI
Cluster 3	0,082
Cluster 4	0,043
Cluster 5	0,051

Tabel 2 merupakan nilai DBI yang didapatkan dari perhitungan *K-Medoids* dengan  $k=3$ ,  $k=4$ , dan  $k=5$ . Sehingga didapatkan  $k=4$  sebagai *cluster* terbaik karena memiliki nilai terendah. Untuk perbandingan nilai *Devies Bouldien Indeks* terlihat pada gambar 3.



Gambar 3. Perbandingan nilai DBI

Berdasarkan hasil perbandingan nilai DBI, diperoleh nilai terbaik menggunakan 4 *cluster* karena memiliki nilai DBI terendah. Hasil pengelompokan menggunakan 4 *cluster* dapat dilihat pada gambar 4.



Gambar 4. Pengelompokan 4 cluster

Berdasarkan hasil dan analisis, pada *cluster 1* diperoleh sebanyak 420 *record*. Kecamatan Tampan mendominasi *cluster 1* dengan 48,81%. Penyakit yang sering di derita pada Kecamatan tersebut adalah Maag sebesar 44,39%,. Dengan rentang Usia dominan adalah 17 – 25 tahun sebesar 94,15%. Dengan Jenis Kelamin Perempuan sebesar 100%. Pada *cluster 2* diperoleh sebanyak 349 *record*. Kecamatan Tampan mendominasi *cluster 2* dengan 45,56%. Penyakit yang sering di derita pada Kecamatan tersebut adalah Diare dan sakit perut sebesar 16,98%. Dengan rentang Usia dominan adalah 17 – 25 tahun sebesar 91,19%. Dengan Jenis Kelamin Laki-laki sebesar 100%. Pada *cluster 3* diperoleh sebanyak 794 *record*. Kecamatan Tampan mendominasi *cluster 3* dengan 48,87%. Penyakit yang sering di derita pada kecamatan tersebut adalah Batuk dan Pilek sebesar 65,21%. Dengan rentang Usia dominan adalah 17 – 25 tahun sebesar 94,33%. Dengan Jenis Kelamin Laki-laki sebesar 100%. Pada *cluster 4* diperoleh sebanyak 1248 *record*. Kecamatan Tampan mendominasi *cluster 4* dengan 46,96%. Penyakit yang sering di derita pada kecamatan tersebut adalah Batuk dan Pilek sebesar 54,10%. Dengan rentang Usia dominan adalah 17 – 25 tahun sebesar 95,22%. Dengan jenis kelamin Laki-laki sebesar 100%.

#### 4. Kesimpulan

Berdasarkan perhitungan yang telah dilakukan, algoritma *K-Medoids* dapat melakukan pengelompokan penyakit yang ada di Pekanbaru Riau dengan pengklasteran terbaik dilakukan dengan 4 *cluster*. Dari 2811 *record* diperoleh 420 *record* pada *cluster* pertama, 349 *record* pada *cluster* kedua, 794 *record* pada *cluster* ketiga, dan 1248 *record* pada *cluster* keempat. Dengan demikian untuk melakukan proses implementasi pada sistem dan analisis dapat menerapkan percobaan tersebut. Hasil dari percobaan juga dapat diimplementasikan dengan data yang besar dan atribut yang kompleks.

#### Daftar Pustaka

- [1] Kementerian Kesehatan Republik Indonesia. *Data Riset Kesehatan Dasar (Riskesdas)*, 2013.
- [2] Wardhani AK. Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan. *Jurnal Transformatika*. 2016; 14(1): 30-37
- [3] Kurniawan H, Fujiati, Saleh A. Analisa Pola Data Penyakit Rumah Sakit dengan Menerapkan Metode Association Rule menggunakan Algoritma Apriori. *Seminar Nasional Informatika*. 2014; 1(1): 195-201
- [4] Madhulatha TS. An Overview on Clustering Methods. *IOSR Journal of Engineer*. 2012; 2(4): 719-725.
- [5] Sabzi A, Farjami Y, Zihayat M. *An Improved Fuzzy K-Medoids Clustering Algorithm with Optimized Number of Clusters*. IEEE International Conference on Hybrid Intelligent System. Malacca. 2011: 206-210.
- [6] Aurora P, Deepali, Varshney S. Analysis of K-Means and K-Medoids Algorithm for Big Data. *ScienceDirect Procedia Computer Science*. 2016; 78: 507-512.
- [7] Arbin N, Suhaimi NS, Mokhtar NZ, Othman Z. *Comperative Analysis Between K-Means and K-Medoids form Statistical Clustering*. IEEE International Conference on Artificial Intelligence, Modelling and Simulation. Kinabalu. 2015: 117-121.
- [8] Tamura Y, Miyamoto S. *Two-Stage Clustering Using One-Pass K-Medoids and Medoid-Based Agglomerative Hierarchical Algorithms*. IEEE International Conference on Soft Computing and Intelligent Systems and 15<sup>th</sup> International Symposium on Advanced Intelligent System. Kitakyushu. 2014: 484-488.
- [9] Kamila I, Khairunnisa U, Mustakim M. Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*. 2019; 5(1): 119-125.
- [10] Zhu Y, Wang F, Shan X, Lv X. *K-Medoids Clustering Based on MapReduce and Optimal Search of Medoids*. IEEE International Conference on Computer Science & Education. Canada. 2014: 573-577.
- [11] Marlina D, Putri NF, Fernando A, Ramadhan A. Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT*. 2018; 4(2): 64-71.
- [12] Pane DK. Implementasi Data Mining pada Penjualan Produk Elektronik Algoritma Apriori (Studi Kasus: Kreditplus). *Jurnal Pelita Informatika*. 2013; 4(3): 25-29.
- [13] Han J, Pei J, Kamber M. *Data Mining: Concept and Techniques*. Third Edition. Waltham: Elsevier. 2012.
- [14] Beynon-Davies P. *Database System*. USA: Palgrave Macmillan. 2004
- [15] Irwansyah E, Faisal M. *Advanced Clustering: Teori dan Aplikasi*. Deepublish. 2015: 5-6.
- [16] Noordiansyah J, Nhita F, Murdiansyah DT. Prediksi Penyakit Menggunakan Algoritma K-Means dan GA untuk Reduksi Dimensi dengan Mengintegrasikan SVM pada Data Berdimensi Tinggi. *E-Proceeding of Engineering*. 2016: 3(2): 3755-3762.
- [17] Defiyanti S, Jajuli M, Rohmawati N. Optimasi K-Medoid dalam Pengklasteran Mahasiswa Pelamar Beasiswa dengan Cubic Clustering Criterion. *Jurnal Nasional Teknologi & Sistem Informasi*. 2017: 3(1): 211-218.
- [18] Syahbana YA. Perancangan dan Implementasi Aplikasi Android Penentu Salient Area pada Vidio dengan Algoritma K-Medoids. *Annual Research Seminar*. 2016: 2(1): 96-101
- [19] Pramesti DF, Furqon M, Dewi C. Implementasi Metode K-Medoids Clustering untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 2017: 1(9): 723-732.
- [20] Dinas Kesehatan Provinsi Riau. *Profil Kesehatan Provinsi Riau*. 2016.
- [21] Vercellis C. *Business Intelligence Data Mining and Optimization for Decision Making*. Milan: Wiley. 2009.
- [22] Han J, Kamber M, Pei J. *Data Mining: Concept and Techniques*. Morgan Kaufmanh. 2006.
- [23] Mustakim. Effectiveness of K-means clustering to distribute training data and testing data on K-nearest neighbor classification. *Journal of Theoretical and Applied Information Technology*. 95 (21), 5693-5700. 2017.

