Jurnal Sistem Komputer dan Informatika (JSON)
Volume 3, Nomor 4, Juni 2022
e-ISSN 2685-998X
DOI 10.30865/json.v3i4.4055

Implementasi Algoritma K-Medoids Dengan Evaluasi *Davies-Bouldin-Index* Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru

Ike Wahyu Septiani, Abd.Charis Fauzan*, Muhamat Maariful Huda

Fakultas Ilmu Eksakta, Program Studi Ilmu Komputer, Universitas Nahdlatul Ulama, Blitar, Indonesia
Email: [1]ikewahyu0509@gmail.com, [2,*]abdcharis@unublitar.ac.id, [3]muhamatmaariful@unublitar.ac.id
Email Penulis Korespondensi: abdcharis@unublitar.ac.id
Submitted: **21/04/2022**; Accepted: **30/06/2022**; Published: **30/06/2022**

Abstrak–Kanker paru-paru merupakan suatu penyakit yang mana terdapat sel-sel yang tumbuh di dalam paru-paru oleh sekumpulan karsinogen secara tidak terkendali. Penyakit kanker paru-paru dapat diatasi dengan operasi, kemoterapi dan radio terapi. Penangaanan lebih dini yang perlu dilakukkan untuk mengurangi tingkat kematian pada pasien penderita kanker paru-paru setelah melakukkan operasi toraks, dengan cara pengumpulan data-data dari setiap pasien mengenai suatu informasi ini menyebabkan suatu masalah yang baru diantaranya adalah data yang diperoleh meliputi data yang berdimensi tinggi dan mempunyai banyak atribut sehingga bisa menghasilkan informasi yang kurang akurat. Maka diperlukannya perhitungan data mining clustering. Pada umumnya metode dalam melakukkan klustering dikelompokkan menjadi empat bagian diantaranya adalah partitioning, hierarchical, grid-based and model-based. Penelitian ini menggunakan algoritma k-medoids karena mampu mengatasi data sensitif terhadap outlier dan memiliki akurasi yang tinggi dan efisien dalam memproses objek dalam jumlah besar. Hasil dari perhitungan K-Medoids dievaluasi menggunakan euclidean distance Davies Bouldin Index yang menghasilkan nilai DBI sebesar 0,93543 menunjukkan bahwa algoritma k-medoids mencapai pengelompokkan yang baik karena hasil akhir dari perhitungannya kurang dari 0. Dari hasil evaluasi menggunakan DBI menunjukkan bahwa algoritma k-medoid mempunyai akumulasi rata-rata rata-rata pada saat eksekusi cukup cepat dan kualitas cluster yang baik.

Kata Kunci: Kanker Paru-Paru; Data Mining; Klasterisasi; Algoritma K-Medoids; Indeks Davies Bouldin

Abstract–Lung Cancer is a disease in which there are cells that grow in the lungs by a collection of carcionogens uncontrollably. Lung Cancer can be treated with surgery, chemotherapy and radiotherapy. Early treatment that needs to be done to reduce the mortality rate in patients with lung cancer after performing thoracic surgery, by collecting data from each patients regarding this information causes a new problem, including the data obtained including high-dimensional data and has many attributes so that it can produce less accurate information. So it is necessary to calculate data mining clustering. In general, the methods for performing clustering are grouped into four parts, namely partitioning, hierarchial, grid-based and model-based. This study used the k-medoids algorithm because it is able to handle data sensitive to outliers and has high accuracy and efficiency in processing large numbers objects. The results of the k-medoids calculation were evaluated using the euclidean distance Davies Bouldin Index which resulted in a DBI value of 0.93543 indicating that the k-medoids algoritm achieves good grouping because the final result of the calculation is less than 0. From the results of the evaluation using DBI it shows that the k-medoids algorithm has an average accumulation average at the time of execution is quite fast and the cluster quality is good.

Keywords: Lung Cancer; Data Mining; Clustering; K-Medoids Algorithm; Davies Bouldin Index

1. PENDAHULUAN

Penyakit ialah suatu keadaan yang mana tubuh mengalami penurunan fungsi yang mengakibatkan produktifitas menurun [1]. Penyakit merupakan salah satu keadaan tidak normal yang mana tubuh maupun pikiran mengalami ketidaknyamanan terhadap orang yang dipengaruhinya. Pola hidup yang kurang sehat dan lingkungan sangat berpengaruh terhadap penyakit yang diderita pasien [2]. Menurut *Unicef* ada beberapa faktor yang mempengaruhi kesehatan tubuh diantaranya yaitu faktor lingkungan, keturunan, perilaku dan faktor pelayanan kesehatan. Dimasa sekarang ini jumlah kematian yang disebabkan operasi menjadi topik yang menarik bagi kalangan dokter, pasien serta masyarakat.

Kanker paru-paru merupakan suatu penyakit yang mana terdapat sel-sel yang tumbuh di dalam paru-paru oleh sekumpulan karsinogen secara tidak terkendali [3]. Kanker paru-paru adalah suatu jenis penyakit yang sering dialami pada anak-anak maupun orang dewasa. Menurut R. T. Prasetio and S. Susanti [4] pasien penderita kanker paru-paru sebanyak 4,9% sehingga menempati peringkat ke-6. Paparan asap rokok yang terlalu sering baik perokok aktif ataupun pasif, polusi udara, dan paparan pada lingkungan kerja merupakan salah satu faktor penyebab kanker paru-paru. Adapun gejala kanker paru-paru pada umumnya yaitu hemoptisi, suara serak, batuk, nyeri pada bagian dada, terdapat abses pada paru-paru dan sesak nafas [3]. Cara menangani penyakit kanker paru-paru yaitu bisa dilakukkan dengan kemoterapi, radioterapi serta terapi bedah. Penyakit kanker paru-paru dapat diatasi dengan operasi toraks karena hampir seluruh saluran pernafasan terletak pada toraks [3] [4]. Operasi toraks yaitu suatu tindakan operasi yang disebabkan oleh penyakit atau cidera pada kerongkongan, paru-paru atau organ tubuh yang ada disekitar dada yang mengakibatkan sulitnya bernafas. Fungsi paru-paru sangatlah penting pada kondisi pasien terutama setelah menjalani operasi atau pembedahan yang menjadi konsentrasi tersendiri dalam mendiagnosis harapan hidup mereka [5].

Permasalahan pertama pada penelitian ini yaitu menjelaskan bagaimana cara memprediksi harapan hidup pasien pasca operasi melalui analisa keadaan pasien sebelum dan sesudah melakukkan operasi. Adapun dataset

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 3, Nomor 4, Juni 2022
e-ISSN 2685-998X
DOI 10.30865/json.v3i4.4055

pasien pasca operasi mempunyai dua kelas diantarannya yaitu meninggal dalam kurun waktu satu tahun *(die)* dan mampu bertahan hidup *(survival)*. Namun, ada beberapa masalah yang muncul yaitu data yang diperoleh meliputi data yang berdimensi tinggi dan mempunyai banyak atribut sehingga menghasilkan informasi yang kurang akurat. Maka perlu dilakukkanya perhitungan *data mining clustering* atau pengelompokkan. *Clustering* merupakan suatu cara pengelompokkan *record* pada suatu *database* berdasarkan suatu kondisi tertentu. Hal yang paling mendasar dari *clustering* adalah mengelompokkan beberapa objek kedalam *cluster*, dikatakan *cluster* yang baik jika memiliki tingkat kemiripan yang tinggi antara objek-objek dalam *cluster* tersebut dan memiliki tingkat ketidaksamaan yang tinggi dengan objek *cluster* lainnya [6]. Metode *clustering* adalah metode untuk mengelompokkan suatu data menjadi beberapa *cluster* dan data dalam satu *cluster* memiliki tingkat kemiripan yang maksimal sedangkan data antar *cluster* mempunyai kemiripan yang minimal [7]. Secara umum metode dalam melakukkan proses *clustering* dikelompokkan menjadi empat bagian yaitu *partitioning, hierarchical, grid-based and model-based.* Ada dua metode *partitioning* yaitu algoritma K-Means dan K-Medoids [8]. Berdasarkan penelitian sebelumnya yang meliputi sampel dan data pasien untuk memprediksi harapan hidup pasca operasi toraks dilakukkan oleh beberapa peneliti dengan menggunakan algoritma yang berbeda-beda yaitu [3] Boosted Neural Network dan Smote, [4] Boosted k-Nearest Neighbor, [5] Metode Genetic dan algoritma Naive Bayes Classifier. Adapun macam-macam metode klasterisasi berbasis partisi yaitu K-Means, K-Modes, K-Medoids dan Fuzzy C-Means. Pada penelitian sebelumnya dengan kasus yang berbeda peneliti menggunakan berbagai macam algoritma klustering diantaranya yaitu menggunakan algoritma [9] K-Means, K-Medoids, dan DBSCAN, [10] Perbandingan algoritma K-Means dan K-Medoids dan [11] Komparansi *distance measure* pada algoritma K-Medoids. Algoritma K-Medoids ialah varian dari algoritma K-Means. Menurut F. Hardiyanti, H. S. Tambunan, and I. S. Saragih [12] Algoritma *Partitioning Around Medoids* (PAM) atau disebut K-Medoids yang dikembangakan oleh Leonard Kaufman dan Peter J.Rousseeuw. Algoritma K-Medoids ialah suatu algoritma yang menggunakan metode partisi *clustering* untuk mengelompokkan sekumpulan n objek menjadi sejumlah k *cluster* [13].

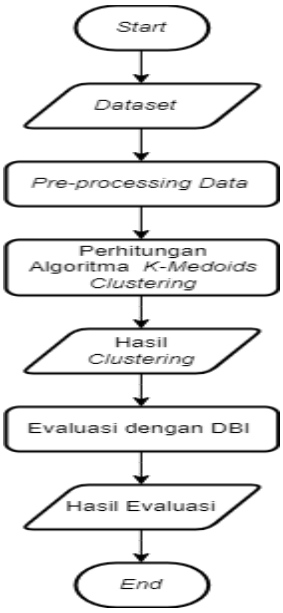
Dari hasil referensi penelitian terdahulu maka peneliti memutuskan memakai algoritma K-Medoids dikarenakan metode K-Medoids memiliki kinerja lebih baik dibandingkan dengan algoritma K-Means untuk dataset kecil dan besar. Metode ini digunakan untuk dapat memecah dataset menjadi kelompok-kelompok dan tidak sensitive terhadap outlier [2] [14] [15] dan mempunyai tingkat akurasi tinggi serta efisien dalam memproses objek dalam jumlah besar [6]. Algoritma K-Medoids cukup efisien untuk mengolah data yang sensitif terhadap *outlier* dan data dalam jumlah kecil [11]. Menurut S. Samudi, S. Widodo, and H. Brawijaya [16] algoritma k-medoids dianggap lebih baik dari algoritma K-Means dikarenakan algoritma k-medoids mampu meminimalkan jumlah ketidaksamaan pada sejumlah objek data dan mampu mencari k sebagai representasi objek. Akan tetapi, algoritma k-means memakai total atau jumlah jarak *Euclidean* pada objek data. Kinerja pengelompokan dari algoritma yang diusulkan dievaluasi dengan Davies-Bouldin Index (DBI), yang mengukur efek dari pengelompokan kohesi dan pemisah. Menurut D. A. I. C. Dewi and D. A. K. Pramita [17] hasil dari Davies Bouldin Index (DBI) dalam menentukan hasil akhir *cluster* yang lebih baik hal ini ditunjukkan dengan membandingkan metode *elbow* dan koefisien *silhoutte* dengan menggunakan metode Davies Bouldin Index (DBI) *cluster* dianggap menghasilkan *clustering* yang optimal jika nilai DBI mendekati nol akan tetapi tidak negatif. Data set yang digunakan peneliti ini diambil dari UCI *Machine Learning Repositort* yang dikumpulkan secara retrospektif di Pusat Bedah Toraks Wroclaw untuk pasien yang menjalani reseksi paru mayor untuk kanker paru-paru primer. Penelitian ini diharapkan mampu menghasilkan pengelompokan harapan hidup pasca operasi pada pasien kanker paru-paru dan menghasilkan data yang akurat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Sebelum membangun atau membuat sistem *claterisasi* harapan hidup pasca operasi pada pasien penderita kanker paru-paru maka perlu adanya desain sistem sebagai penerapan aplikasi secara rinci dan teratur. Tahapan-tahapan paling utama dalam melakukkan penelitian yaitu dimulai dari mengidentifikasi masalah, kemudian melakukkan pengumpulan berbagai macam data yang dibutuhkan pada suatu permasalahan yang ada. Data kemudian diolah dengan proses *cleaning*, transformasi serta normalisasi, setelah selesai malakukkan itu kemudian terapkan perhitungan algoritma k-medoids [1]. Menurut [18] hal yang paling utama sebelum melakukaan penelitian perlu dilakukkannya *studi literatur*, pengumpulan data dan setelah itu lakukkan *preprocessing* data yaitu *cleaning* data yang nanti akan dilakukannya proses perhitungan algoritma dan dievaluasi sistem. Apabila hasil akhir dari *cluster* telah selesai atau didapatkan maka dilakukkannya analisis dan nantinya didapatkan hasil dan kesimpulan.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 3, Nomor 4, Juni 2022
e-ISSN 2685-998X
DOI 10.30865/json.v3i4.4055



Gambar 1. Tahapan Penelitian

Gambar 1 adalah desain sistem yang diterapkan. Alur dari tahapan ini dimulai dengan proses input data, *preprocessing* data, perhitungan algoritma k-medoids dan menentukan tingkat keakuratan.
Tahapan-tahapan dalam penelitian adalah:

a) Dataset

Hal utama sebelum melakukkan perhitugan yaitu menyiapkan data terlebih dahulu. Data yang digunakan yaitu data untuk serangkaian proses *clasterisasi* harapan hidup pasca operasi toraks pada pasien penderita kanker paru-paru. Pada dataset ini terdapat 16 atribut yaitu:

Tabel 1. Atribut DataSet

Attributes	***Descriptions***	***Data Type***
DGN	Diagnosis-kombinasi spesifik kode ICD-10 untuk tumor primer dan sekunder serta lebih dari satu tumor	*Nominal*
PRE4	Jumlah udara yang bisa dihempuskan secara paksa dari paru-paru setelah mengambil nafas sedalam mungkin (FVC)	*Numerik*
PRE5	Jumlah udara yang telah dihembuskan pada akhir detik pertama dari FVC (FEV1)	*Numerik*
PRE6	Ukuran kemampuan umum pasien kanker dalam aktifitas sehari-hari *(Zubrod Scale)*	*Nominal*
PRE7	Rasa sakit sebelum operasi	*Binary*
PRE8	Hemoptysis sebelum operasi	*Binary*
PRE9	Dyspnea sebelum operasi	*Binary*
PRE10	Batuk sebelum operasi	*Binary*
PRE11	Kondisi lemah sebelum operasi	*Binary*
PRE14	Ukuran tumor (TNM)	*Nominal*
PRE17	Diabetes	*Binary*
PRE19	*Myocardial infarction* (MI) hingga 6 bulan	*Binary*
PRE25	Penyakit yang menyerang arteri/aliran darah (PAD)	*Binary*
PRE30	Merokok	*Binary*
PRE32	Asma	*Binary*
AGE	Usia saat operasi	*Numerik*

Tabel 2 adalah tabel dataset pada data harapan hidup pasca operasi pada pasien kanker paru-paru :

Tabel 2. Dataset

No	**DGN**	**PRE4**	**PRE5**	**PRE6**	**PRE7**	**PRE8**	**PRE9**	**PRE10**	**PRE11**	**PRE14**	**PRE17**	**PRE19**	**....**
1	DGN2	2.88	2.16	PRZ1	F	F	F	T	T	OC14	F	F	**....**
2	DGN3	3.4	1.88	PRZ0	F	F	F	F	F	OC12	F	F	**....**
3	DGN3	2.76	2.08	PRZ1	F	F	F	T	F	OC11	F	F	**....**

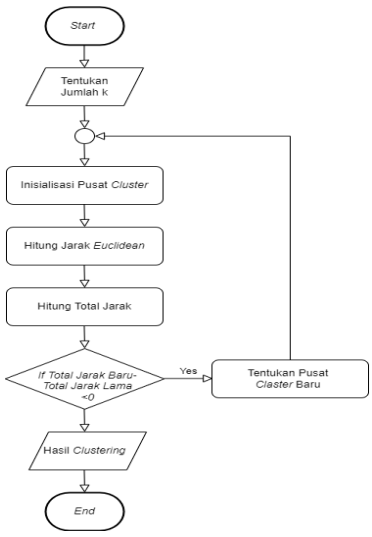
Jurnal Sistem Komputer dan Informatika (JSON)
Volume 3, Nomor 4, Juni 2022
e-ISSN 2685-998X
DOI 10.30865/json.v3i4.4055

4	DGN 3	3.68	3.04	PRZ 0	F	F	F	F	F	OC11	F	F	....
5	DGN 3	2.44	0.96	PRZ 2	F	T	F	T	T	OC11	F	F	....
6	DGN 3	2.48	1.88	PRZ 1	F	F	F	T	F	OC11	F	F	....
7	DGN 3	4.36	3.28	PRZ 1	F	F	F	T	F	OC12	T	F	....
8	DGN 2	3.19	2.5	PRZ 1	F	F	F	T	F	OC11	F	F	....
9	DGN 3	3.16	2.64	PRZ 2	F	F	F	T	T	OC11	F	F	....
10	DGN 3	2.32	2.16	PRZ 1	F	F	F	T	F	OC11	F	F	....
....	....	....	....	....	....	....	....	....	....	....	....	....	....
463	DGN 3	2.12	1.68	PRZ 2	T	T	F	F	F	OC11	F	F	....
464	DGN 4	3.44	2.16	PRZ 1	F	F	F	T	T	OC12	T	F	....
465	DGN 3	3.08	2.16	PRZ 1	F	F	F	T	T	OC13	F	F	....
466	DGN 2	3.88	2.12	PRZ 1	F	F	F	T	F	OC13	F	F	....
467	DGN 3	3.76	3.12	PRZ 0	F	F	F	F	F	OC11	F	F	....
468	DGN 3	3.04	2.08	PRZ 1	F	F	F	T	F	OC13	F	F	....
469	DGN 3	1.96	1.68	PRZ 1	F	F	F	T	T	OC12	F	F	....
470	DGN 3	4.72	3.56	PRZ 0	F	F	F	F	F	OC12	F	F	....

Tabel 2 merupakan tabel dataset prediksi harapan hidup pasca operasi pada pasien penderita kanker paru-paru sebelum dilakukkannya *preprocessing* data.
Ada 3 tahapan dalam *preprocessing* yakni:
1) *Cleaning* Data
2) Transformasi Data
3) Normalisasi Data

b. Perhitungan Algoritma K-Medoids

Algoritma K-Medoids ialah suatu algoritma partisi *clustering* yang mengelompokkan sekumpulan n objek menjadi sejumlah k *cluster* [13]. K-Medoids bertujuan untuk memecah *dataset* menjadi kelompok-kelompok [2]. Metode ini, terdiri dari sekumpuan data atas n objek yang dipartisi menjadi *cluster* yang memiliki jumlah $k \leq n$ [17]. Medoids merupakan suatu objek pusat *cluster* dan objek yang mewakili *cluster*. Untuk menemukan k *cluster* dari n objek dalam algoritma k-medoids yaitu dengan cara mencari terlebih dahulu objek yang mewakili medoids tiap *cluster*. Ditengah setiap *cluster*, objek yang kuat terhadap outlier disebut medoids. *Cluster* dibentuk dengan menghitung jarak antara objek medoids dan non medoids[19].



Gambar 2. Flowchart Perhitungan Agoritma K-Medoids

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 3, Nomor 4, Juni 2022
e-ISSN 2685-998X
DOI 10.30865/json.v3i4.4055

Tahapan-tahapan dalam perhitungan algoritma k-medoids yaitu:

1. Tentukan k (jumlah *cluster*) yang diinginkan dari data.
2. Pilihlah secara *random* medoids pertama sebanyak k dari n data.
3. Kemudian hitunglah masing-masing jarak objek ke medoids dengan menggunakan rumus *Euclidean Distance:*

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

4. Tandailah jarak terdekat dari objek ke Medoids dan hitunglah totalnya.
5. Tentukan anggota cluster ke medoid sementara.
6. Lakukan iterasi seperti langkah pada nomer 2 sampai 4.
7. Kemudian hitunglah total s atau simpangan .
   - Jika a merupakan jumlah dari jarak terdekat objek dari medoids awal atau iterasi 1
   - Jika b merupakan jumlah dari jarak terdekat objek dari medoids baru atau iterasi 2
   - Rumus dari total simpangan yaitu s = b-a
8. Total Simpanga (S)
   - Apabila S > 0 maka proses *clustering* dihentikan dan didapatkan anggota *cluster* dari masing-masing medoids.
   - Apabila S < 0 maka tukar objek tersebut dengan data lain agar membentuk sekumpulan k baru sebagai medoids. Lakukkan iterasi hingga diperoleh hasil nilai S > 0.

c. Perhitungan Validasi Internal Sistem Menggunakan Davies-Bouldin Index (DBI)

Davies Bouldin Index (DBI) adalah salah satu validasi *cluster* yang diperkenalkan oleh D.L. Davies dan Donald W.Bouldin maka dari itu penamaan pada metode ini adalah gabungan nama antara keduanya yaitu Davies-Bouldin [20]. DBI ialah salah satu cara untuk menganalisa kualitas *cluster* pada setiap *clustering* [21]. DBI merupakan suatu fungsi rasio dari sejumlah distribusi kedalam *cluster* sebagai pemisah antar *cluster* [22]. Bentuk pendekatan pada pengujian nilai DBI meliputi nilai separasi dan kohesi. Separasi ialah jarak antara pusat *cluster* dari *cluster*. Kohesi berupa jumlah dari kemiripan data terhadap pusat *cluster* dari *cluster* tersebut. *Cluster* dikatakan optimal jika *cluster* tersebut memiliki nilai kohesi yang rendah sedangkan nilai separasi yang tinggi [17].

Ada empat tahapan dalam menghitung DBI diantaranya yaitu:

a) Menghitung *Sum Of Square Within Cluster* (SSW) ialah keterikatan anggota satu *cluster* atau seberapa mirip antara anggota satu dan dua dan semakin kecil semakin bagus dikarenakan semakin mirip. SSW dihitung untuk mengetahui matrik/kohesi/homogenitas. Kohesi merupakan keterikatan anggota *cluster* dalam satu *cluster.*

$$\boldsymbol{SSW_i = \frac{1}{mi}\sum_{j=i}^{mi} d(x_j, c_i)} \quad \textbf{(2)}$$

Keterangan :

mi = jumlah data dalam *cluster* ke-i
x = data dalam *cluster*
$d(x,c)$ = jarak data ke centroid
x_j = data pada *cluster* tersebut
c_i = centroid *cluster* ke-i

b) Menghitung *Sum Of Square between cluster* (SBB) merupakan jarak antar *cluster* cukup besar sehingga terpisah ke dalam kelompok lain. SSB bertujuan untuk mengetahui separasi/heterogenitas. Separasi merupakan perbedaan antara satu *cluster* dengan *cluster* lainnya.

$$\boldsymbol{SSB_{i,j} = d\,(c_i, c_j)} \quad \textbf{(3)}$$

Keterangan :

C_i = *Cluster* 1
C_j = *Cluster* lainnya
$d\,(c_i, c_j)$ = Jarak antara centroid satu dengan lainnya

c) Menghitung Rasio berfungsi untuk dapat mengetahui seberapa bagus nilai perbandingan *cluster* satu dengan *cluster* lainnya. Jumlah kohesi harus kecil sedangkan jumlah separasi harus lebih besar.

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

Keterangan :

$R_{i,j}$ = Rasio antar *cluster*
SSW_i = *cluster* 1
SSW_j = *cluster* 2
SSB_{ij} = separasi dari *cluster* 1 dan 2

d) Menghitung DBI *(Davis Bouldin Index)*

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 3, Nomor 4, Juni 2022
e-ISSN 2685-998X
DOI 10.30865/json.v3i4.4055

Faktanya jika hasil dari perhitungan DBI yang diperoleh semakin kecil mendekati nol akan tetapi tidak negatif (*non-negatif >= 0)* , maka nilai hasil *clustering* semakin baik [23].

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

Keterangan :
k = kluster yang ada
$R_{i,j}$ = rasio antara kluster i dan j
Max = dicari rasio antar kluster yang terbesar

3. HASIL DAN PEMBAHASAN

3.1 *Pre-processing*

Ada 3 tahapan dalam *preprocessing* yaitu :

1) *Cleaning* Data
Cleaning Data atau pembersihan data dilakukkan apabila terdapat data yang kosong atau menghapus suatu nilai yang salah dan memeriksa serta memperbaiki data yang mengalami kesalahan dalam hal penulisan ataupun data tidak konsisten. Jika tidak ada data yang tidak valid, tidak relevan maupun data yang kosong maka tidak perlu di *cleaning* data melainkan langsung ke langkah perhitungan transformasi data.
2) Transformasi Data

Tabel 3. Data Transformasi

No	DGN	PRE4	PRE5	PRE6	PRE7	PRE8	PRE9	PRE10	PRE11	PRE14	PRE17	PRE19	….
1	2	2.88	2.16	1	2	2	2	1	1	14	2	2	….
2	3	3.4	1.88	0	2	2	2	2	2	12	2	2	….
3	3	2.76	2.08	1	2	2	2	1	2	11	2	2	….
4	3	3.68	3.04	0	2	2	2	2	2	11	2	2	….
5	3	2.44	0.96	2	2	1	2	1	1	11	2	2	….
6	3	2.48	1.88	1	2	2	2	1	2	11	2	2	….
7	3	4.36	3.28	1	2	2	2	1	2	12	1	2	….
….	….	….	….	….	….	….	….	….	….	….	….	….	….
465	3	3.08	2.16	1	2	2	2	1	1	13	2	2	….
466	2	3.88	2.12	1	2	2	2	1	2	13	2	2	….
467	3	3.76	3.12	0	2	2	2	2	2	11	2	2	….
468	3	3.04	2.08	1	2	2	2	1	2	13	2	2	….
469	3	1.96	1.68	1	2	2	2	1	1	12	2	2	….
470	3	4.72	3.56	0	2	2	2	2	2	12	2	2	….
Min	2	1.96	0.96	0	1	1	2	1	1	11	1	2	….
Max	4	4.72	3.56	2	2	2	2	2	2	14	2	2	….

Transformasi data merupakan tahap dimana data yang telah dibersihkan tadi di ubah menjadi bentuk yang sesuai agar dapat memudahkan proses perhitungan karena jika tidak ditransformasi maka akan mempersulit perhitungan apalagi jika data tersebut tercampur antara data kategorikal dan numerikal maka bisa dirubah menjadi numerikal. Namun, jika semua data sudah berbentuk numerikal atau kategorikal saja maka tidak perlu dilakukkannya transformasi data. Pada proses pengubahan ini dilakukkan pada *Microsoft Excel.* Berikut adalah tabel yang telah diinisialisasi bisa dilihat pada tabel 3.

3. Normalisasi Data

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 3, Nomor 4, Juni 2022
e-ISSN 2685-998X
DOI 10.30865/json.v3i4.4055

Normalisasi data dapat dihitung dengan rumus normalisasi yaitu *minimal - maximal* dari data yang mempunyai *range* nilai antara 0 sampai dengan 1 dan tidak mempunyai nilai yang terlalu jauh. Rumus normalisasi dapat dilihat pada rumus nomer 6 yang mana x^i ialah hasil nilai normalisasi, x ialah nilai dari data yang telah dinormalisasi, x_{max} ialah nilai maksimum pada data sedangkan x_{min} ialah nilai minimum data aktual.

$$x^i = \frac{x - x_{min}}{x_{max} - x_{min}} \quad \textbf{(6)}$$

Berdasarkan rumus (6) maka akan diperoleh hasil perhitungan normalisasi pada tabel 4.

Tabel 4. Normalisasi Data

No	DGN	PRE4	PRE5	PRE6	PRE7	PRE8	PRE9	PRE10	PRE11	PRE14	PRE17	PRE19	….
1	0.143	0.296	0.014	0.500	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000	….
2	0.286	0.403	0.011	0.000	1.000	1.000	1.000	1.000	1.000	0.333	1.000	1.000	….
3	0.286	0.272	0.013	0.500	1.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000	….
4	0.286	0.461	0.024	0.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	….
5	0.286	0.206	0.000	1.000	1.000	0.000	1.000	0.000	0.000	0.000	1.000	1.000	….
6	0.286	0.214	0.011	0.500	1.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000	….
7	0.286	0.601	0.027	0.500	1.000	1.000	1.000	0.000	1.000	0.333	0.000	1.000	….
….	….	….	….	….	….	….	….	….	….	….	….	….	….
465	0.286	0.337	0.014	0.500	1.000	1.000	1.000	0.000	0.000	0.667	1.000	1.000	….
466	0.143	0.502	0.014	0.500	1.000	1.000	1.000	0.000	1.000	0.667	1.000	1.000	….
467	0.286	0.477	0.025	0.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	….
468	0.286	0.329	0.013	0.500	1.000	1.000	1.000	0.000	1.000	0.667	1.000	1.000	….
469	0.286	0.107	0.008	0.500	1.000	1.000	1.000	0.000	0.000	0.333	1.000	1.000	….
470	0.286	0.675	0.030	0.000	1.000	1.000	1.000	1.000	1.000	0.333	1.000	1.000	….

3.2 Perhitungan Algoritma K-Medoids

a) Pilih Secara Random Medoid Pertama Sebanyak k Dari n Data

Tabel 5. Data Acak

No	DGN	PRE4	PRE5	PRE6	PRE7	PRE8	PRE9	PRE10	PRE11	PRE14	PRE17	PRE19	….
469	0.286	0.107	0.008	0.500	1.000	1.000	1.000	0.000	0.000	0.333	1.000	1.000	….
470	0.286	0.675	0.030	0.000	1.000	1.000	1.000	1.000	1.000	0.333	1.000	1.000	….

Tabel 5 merupakan data acak dari data yang telah di normalisasi tadi, data acak ini akan dijadikan sebagai objek ke medoids sementara.

b. Menghitung jarak pada objek masing-masing ke medoid sementara menggunakan rumus jarak *euclidiean.*

$$d(x,y) = \sqrt{(0.286 - 0.143)^2 + (0.107 - 2.96)^2 + (0.008 - 0.014)^2 + (0.500 - 0.500)^2} + dst \ldots$$

c. Beri tanda jarak terdekat dari objek ke medoid serta hitung jumlah nilainya.

Untuk menghitung nilai kedekatan yaitu dengan cara mencari nilai minimum dari cost 1 dan cost 2 kemudian hitung total kedekatannya. Untuk menghitung jumlah kedekatan yaitu jumlah kedekatan pertama dijumlahkan sampai ke akhir seperti 0.763932645 + 0.272317424 + 1.109079161 dst….

d. Penentuan anggota cluster ke medoid sementara

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 3, Nomor 4, Juni 2022
e-ISSN 2685-998X
DOI 10.30865/json.v3i4.4055

Rumus untuk perhitungan cluster yaitu jika cost 1 = Minimal (Cost 1 dan Cost 2) maka bernilai (1,2).

Tabel 6. Hasil Perhitungan dari Jarak, Jumlah Kedekatan dan *Cluster*

Cost 1	Cost 2	Kedekatan	Cluster
0.763932645	1.696188479	0.763932645	1
1.586750964	0.272317424	0.272317424	2
1.109079161	1.240463152	1.109079161	1
1.905281408	1.076571796	1.076571796	2
1.174394042	2.107899149	1.174394042	1
1.517414366	1.604347964	1.517414366	1
1.528412293	1.506715007	1.506715007	2
….	….	….	….
0.405279141	1.629488693	0.405279141	1
1.160340097	1.20197144	1.160340097	1
1.604041835	0.416068477	0.416068477	2
1.525736271	1.575162029	1.525736271	1
0	1.659210278	0	1
1.659210278	0	0	2
Jumlah Kedekatan		510.9421298	

e. Lakukan iterasi medoid

Untuk mencari iterasi medoid langkah-langkah yang dilakukan yaitu memilih medoid sementara dan ikuti langkah-langkah diatas mulai dari langkah a, b dan c.

Tabel 7. Medoids Sementara

No	**DGN**	**PRE4**	**PRE5**	**PRE6**	**PRE7**	**PRE8**	**PRE9**	**PRE10**	**PRE11**	**PRE14**	**PRE17**	**PRE19**	**….**
4	0.286	0.461	0.024	0.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	….
5	0.286	0.206	0.000	1.000	1.000	0.000	1.000	0.000	0.000	0.000	1.000	1.000	….

Tabel 7 merupakan tabel medoids sementara. Data pada tabel ini dipilih secara acak dari data yang telah dinormalisasi yang nantinya akan dijadikan sebagai medoids sementara.

Tabel 8. Hasil Perhitungan

Cost 1	**Cost 2**	**Kedekatan**	**Cluster**
2.075060264	1.522366424	1.522366424	2
1.056731406	2.064305424	1.056731406	1
1.513836318	1.516411234	1.513836318	1
0	2.269045333	0	1
2.269045333	0	0	2
1.14595689	1.833383503	1.14595689	1
1.84022694	1.88756969	1.84022694	1
….	….	….	….
1.962975449	1.311582312	1.311582312	2
1.653861353	1.681007348	1.653861353	1
1.00574385	2.026688855	1.00574385	1
1.308751615	1.952203353	1.308751615	1
1.905281408	1.174394042	1.174394042	2
1.076571796	2.107899149	1.076571796	1
Jumlah Kedekatan		631.1123182	

Tabel 8 merupakan tabel hasil perhitungan dari medoids sementara. Urutan pada perhitungan ini sama seperti pada langkah a, b dan c.

f. Hitung Total Simpangan

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 3, Nomor 4, Juni 2022
e-ISSN 2685-998X
DOI 10.30865/json.v3i4.4055

Rumus dari total simpangan yaitu s = b - a
S = 631.1123182 – 510.9421298 = 120.1701884
Total Simpangan (S)
-Apabila S > 0 maka proses *clustering* dihentikan dan didapatkan anggota cluster dari masing-masing medoids.
-Apabila S < 0 maka lakukkan iterasi lagi sampai memperoleh nilai S > 0 dengan cara menukar objek dengan data agar membentuk sekumpulan k baru sebagai medoids.

3.3 Perhitungan Validasi Sistem Menggunakan Davies-Bouldin Index (DBI)

Ada 4 tahapan yang harus dilakukkan sebelum menghitung validasi menggunakan DBI *(Davies-Bouldin Index)* yaitu:

a) Menghitung *Sum Of Square Within Cluster* (SSW)
Sebelum menghitung SSW maka dilakukkannya pengelompokan *cluster* terlebih dahulu seperti pada tabel 9:

Tabel 9. Pengelompokan *Claster*

Data ke-	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	…	Cluster yang diikuti
1	0.143	0.296	0.014	0.5	1	1	1	0	0.00	1.00	1.00	1.00	…	1
3	0.286	0.272	0.013	0.5	1	1	1	0	1.00	0.00	1.00	1.00	…	1
5	0.286	0.206	0	1	1	0	1	0	0.00	0.00	1.00	1.00	…	1
6	0.286	0.214	0.011	0.5	1	1	1	0	1.00	0.00	1.00	1.00	…	1
8	0.143	0.36	0.018	0.5	1	1	1	0	1.00	0.00	1.00	1.00	…	1
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
460	0.286	0.584	0.026	0	1	1	1	0	1.00	0.333	1.00	1.00	….	2
461	0.429	0.66	0.033	0.5	1	1	1	0	1.00	0.333	1.00	1.00	…	2
463	0.286	0.14	0.008	1	0	0	1	1	1.00	0.00	1.00	1.00	…	2
467	0.286	0.477	0.025	0	1	1	1	1	1.00	0.00	1.00	1.00	…	2
470	0.286	0.675	0.03	0	1	1	1	1	1.00	0.333	1.00	1.00	…	2

Hitung rata-rata data pada setiap atribut pada 1 cluster. Kemudian hitung jarak data yang diikuti (euclidean).

Tabel 10. Rata-rata setiap cluster dan jarak data yang diikuti

Data ke-	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	Cluster yang diikuti	Jarak data yang diikuti *(euclidean)*
1														10.789
3														0.4554
5													1	12.491
6													1	10.135
8														10.844
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
460														10.768
461														11.273
463													2	16.291
467														0.5162

Nilai SSW pada *cluster* 1 adalah 0.8082 diperoleh dari nilai rata-rata jarak data pada *cluster* 1 sedangkan nilai SSW dari *cluster* 2 adalah 0.9134 yang diperoleh dari nilai rata-rata jarak data pada *cluster* 2.

b) Menghitung *Sum Of Square between cluster* (SBB)

Tabel 11. Hasil Perhitungan SSB

	Centroid	
SBB	**1**	**2**
1	0	0.9202
2	0.9202	0

c) Menghitung Rasio
Hasil Rasio pada perhitungan ini adalah 1.8708 yang diperoleh dari nilai SSW pada *cluster 1* ditambah nilai SSW pada *cluster 2* kemudian dibagi hasil dari SSB.

d) Menghitung *Davis Bouldin Index* (DBI)

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 3, Nomor 4, Juni 2022
e-ISSN 2685-998X
DOI 10.30865/json.v3i4.4055

Hasil perhitungan DBI adalah 0.9354 yang diperoleh dari hasil Rasio dikali $\frac{1}{2}$. Hasil DBI yang dihasilkan dari perhitungan menggunakan *Euclidean Distance* menghasilkan nilai akhir kurang dari 0 yang berarti bahwa hasil *clustering* ini mempunyai tingkat kemiripan yang lumayan tinggi pada satu kelompoknya.

4. KESIMPULAN

Algoritma K-Medoids merupakan suatu algoritma metode partisi *clustering* yang bertujuan agar bisa mengelompokkan sekumpulan n objek menjadi sejumlah k *cluster*. Berdasarkan pengelompokkan menggunakan algoritma k-medoids untuk memprediksi harapan hidup pasca operasi mengenai analisa keadaan pasien sebelum dan setelah melakukkan operasi yang terdiri atas dua kelas yaitu meninggal dalam kurun waktu satu tahun dan mampu bertahan hidup. Dataset ini terdiri atas 16 atribut dan memiliki data sebanyak 470, data sebanyak ini akan diolah menggunakan algoritma k-medoid karena memiliki kinerja yang baik. Hasil dari penelitian ini dapat ditarik kesimpulkan bahwa algoritma K-Medoids menghasilkan *cluster* yang lebih optimal dibandingkan dengan algoritma *k-means*. K-medoids digunakan untuk memecah dataset menjadi sekumpulan kelompok serta tidak sensitive terhadap outlier dan mempunyai tingkat akurasi tinggi dan efisien ketika memproses objek dalam jumlah besar. Hal ini dapat dilihat dari hasil evaluasi *Davies Bouldin Index* menggunakan perhitungan *Euclidean Distance* yaitu sebesar 0,93543 menujukkan bahwa algoritma k-medoids mencapai efek pengelompokkan yang baik karena hasil akhir dari perhitungannya kurang dari 0. Dari hasil evaluasi atau pengujian menggunakan DBI menunjukkan bahwa algoritma k-medoids mempunyai akumulasi rata-rata pada saat eksekusi cukup cepat serta mempunyai kualitas *cluster* yang baik.

REFERENCES

[1] T. Juninda, Mustasim, and E. Andri, "Penerapan Algoritma K-Medoids untuk Pengelompokan Penyakit di Pekanbaru Riau," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, vol. 11, no. 1, pp. 42–49, 2019.
[2] A. D. Andini and T. Arifin, "Implementasi Algoritma K-Medoids Untuk Klasterisasi Data Penyakit Pasien Di Rsud Kota Bandung," *J. RESPONSIF Ris. Sains ...*, vol. 2, no. 2, pp. 128–138, 2020, [Online]. Available: http://ejurnal.ars.ac.id/index.php/jti/article/view/247.
[3] I. F. Anshori and D. Riana, "Prediksi Harapan Hidup Pasien Kanker Paru-Paru Pasca Operasi Bedah Thoraks Menggunakan Boosted Neural Network Dan Smote," vol. 6, no. 1, pp. 9–15, 2021.
[4] R. T. Prasetio and S. Susanti, "Prediksi Harapan Hidup Pasien Kanker Paru Pasca Operasi Bedah Toraks Menggunakan Boosted k-Nearest Neighbor," *J. Responsif*, vol. 1, no. 1, pp. 64–69, 2019, [Online]. Available: http://ejurnal.univbsi.id/index.php/jti.
[5] Y. A. Setyadi, I. Asror, Y. Firdaus, and A. Wibowo, "Prediksi Harapan Hidup Pasca Operasi Toraks pada Pasien Penderita Kanker Paru-paru Menggunakan Metode Genetic Algorithm untuk Feature Selection dan Naïve Bayes Classifier," *e-Proceeding Eng.*, vol. 7, no. 2, pp. 8349–8360, 2020.
[6] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient," vol. 5, no. 2, pp. 109–116, 2021.
[7] D. Kurmiati, M. Z. Fauzi, Ripangi, A. Falegas, and Indria, "Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm," *Malcolm Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 47–57, 2021, doi: 10.21108/indojc.2019.4.3.359.
[8] I. Syukra, A. Hidayat, and M. Z. Fauzi, "Implementation of K-Medoids and FP-Growth Algorithms for Grouping and Product Offering Recommendations," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, p. 107, 2019, doi: 10.24014/ijaidm.v2i2.8326.
[9] R. W. Sembiring Brahmana, F. A. Mohammed, and K. Chairuang, "Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, p. 32, 2020, doi: 10.24843/lkjiti.2020.v11.i01.p04.
[10] A. Supriyadi, A. Triayudi, and ..., "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *JIPI (Jurnal ...*, vol. 06, pp. 229–240, 2021, [Online]. Available: https://www.jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/2008.
[11] M. N. P. Pamulang, M. N. Aini, and U. Enri3, "Komparasi Distance Measure Pada K-Medoids Clustering untuk Pengelompokkan Penyakit ISPA," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 99–107, 2021, doi: 10.29408/edumatic.v5i1.3359.
[12] F. Hardiyanti, H. S. Tambunan, and I. S. Saragih, "Penerapan Metode K-Medoids Clustering Pada Penanganan Kasus Diare Di Indonesia," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 598–603, 2019, doi: 10.30865/komik.v3i1.1666.
[13] P. E. Prakasawati, Y. H. Chrisnanto, and A. I. Hadiana, "Segmentasi Pelanggan Berdasarkan Produk Menggunakan Metode K- Medoids," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 335–339, 2019, doi: 10.30865/komik.v3i1.1610.
[14] S. R. Ningsih, I. S. Damanik, A. P. Windarto, H. S. Tambunan, J. Jalaluddin, and A. Wanto, "Analisis K-Medoids Dalam Pengelompokkan Penduduk Buta Huruf Menurut Provinsi," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 721, 2019, doi: 10.30645/senaris.v1i0.78.
[15] A. Hermawati, S. Jumini, M. Astuti, F. Ismail, and R. Rahim, "Unsupervised Data Mining with K-Medoids Method in Mapping Areas of Student and Teacher Ratio in Indonesia," *TEM J.*, vol. 9, no. 4, pp. 1614–1618, 2020, doi: 10.18421/TEM94-37.
[16] S. Samudi, S. Widodo, and H. Brawijaya, "The K-Medoids Clustering Method for Learning Applications during the

COVID-19 Pandemic," *SinkrOn*, vol. 5, no. 1, p. 116, 2020, doi: 10.33395/sinkron.v5i1.10649.
[17] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
[18] M. H. Herviany, S. P. Delima, T. Nurhidayah, and K. Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat: Comparison of K-Means and K-Medoids Algorithms for Grouping Landslide Prone Areas in West Java Province," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021.
[19] R. K. Dinata, S. Retno, and N. Hasdyna, "Minimization of the Number of Iterations in K-Medoids Clustering with Purity Algorithm," *Rev. d'Intelligence Artif.*, vol. 35, no. 3, pp. 193–199, 2021, doi: 10.18280/ria.350302.
[20] M. Herviany, S. P. Delima, T. Nurhidayah, and Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021.
[21] V. V Arganata, "Algoritma Partitioning Around Medoids (Pam) Dengan Metode Davies Bouldin Index Untuk Mengelompokkan Provinsi Di Indonesia …," 2021, [Online]. Available: http://repository.unmuhjember.ac.id/id/eprint/8381.
[22] Y. Religia and R. T. B. Jaya, "Pengelompokan Menggunakan Algoritma K- Medoid Untuk Evaluasi Performa Siswa," *J. Pelita Teknol.*, vol. 15, no. 1, pp. 49–55, 2020.
[23] F. Farahdinna, I. Nurdiansyah, A. Suryani, and A. Wibowo, "Perbandingan Algoritma K-Means Dan K-Medoids Dalam Klasterisasi Produk Asuransi Perusahaan Nasional," *J. Ilm. FIFO*, vol. 11, no. 2, p. 208, 2019, doi: 10.22441/fifo.2019.v11i2.010.
[24] A. A. D. Sulistyawati and M. Sadikin, "Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan," *Sistemasi*, vol. 10, no. 3, p. 516, 2021, doi: 10.32520/stmsi.v10i3.1332.