ALGORITMA *CLUSTERING K-MEDOIDS* PADA E-GOVERNMENT BIDANG INFORMATION AND COMMUNICATION TECHNOLOGY DALAM PENENTUAN STATUS EDGI

Zaenal Mustofa[1], Iman Saufik Suasana[2]

[1,2] Sistem Komputer STEKOM Semarang
[1] Zaenalmustofa@Stekom.ac.id, [2]saufik@stekom.ac.id

Abstract

E-Government is a tool to improve relations with the community, now the development of E-Government in various world governments are monitored directly by the UN through the United Nations E-Government Survey. This monitoring uses the framework where in this EGDI there are 3 factors considered: Online Service Index, Telecommunication Infrastructure Index and Human Capital Index. But the determination of EDGI status is less accurate because it must be based on knowledge and processing of the number of existing data so that required a calculation that applies clustering method with data mining techniques. K-medoids using clustering techniques, capable of producing optimal Bouldin Index values and this study also determines the medoid distance calculation to obtain optimal algorithm, the value obtained from Bouldin Index on the Chebyshev K-medoids method 0.593. Thus, the optimal clustering scheme with distance ratio is the minimum index value of K-medoids Chebyshev.

Keywords: ***e-government development index; k-medoids; chebyshev; clustering***

1. Pendahuluan

Informasi pada era globalisasi menjadi kebutuhan pokok bagi setiap orang, namun tidak semua informasi yang ada dapat menjadi kebutuhan. Dapat dipengaruhi oleh kemajuan teknologi internet sehingga informasi mengalami peningkatan, sementara kapasitas berita elektronik berbahasa Indonesia yang semakin meningkat merupakan sumber yang berharga, dan memungkinkan banyak pengguna informasi dapat merubah, menghasilkan informasi baru dan memperbanyak. Sehingga perlu adanya peninjauan yang lebih agar mendapatkan informasi yang akurat dan sesuai dengan apa yang diinginkan pengguna informasi, pengelompokan berita dibutuhkan untuk mempermudah pencarian informasi mengenai suatu event tertentu. (L. Noviani, Atall, 2012)

Saat ini, tantangan bagi pemerintah adalah bagaimana bergerak dari fokus pada penyampaian pelayanan ke penyediaan aplikasi yang berpusat pada obyek. Dengan kata lain, keberhasilan pemerintah bergantung pada efektifitas komunikasi dalam penyampaian informasi kepada warga negara dan membangun ikatan yang kuat dengan negara dalam memberdayakan partisipasi masyarakat untuk proses pengambilan keputusan. Pada titik ini, internet memiliki potensi secara radikal mengubah pemerintahan dengan meningkatkan komunikasi antara pejabat publik dengan warga negara. Saat ini, pemerintah di seluruh dunia telah meluncurkan rencana membangun aplikasi dan pelayanan e-government, Aplikasi dan pelayanan e-goverment bertujuan untuk membangun masyarakat informasi. Sebuah masyarakat informasi sepenuhnya memanfaatkan kemajuan terbaru dalam teknologi informasi dan komunikasi (TIK). Aspek kunci dari suatu masyarakat informasi adalah pelayanan e-government yang berfungsi penuh. Diharapkan bahwa e-goverment bermanfaat untuk menghadapi tantangan utama seperti untuk menjamin pelayanan yang lebih baik, meningkatkan peluang warga untuk mobilitas dan untuk kepentingan bisnis, menghadapi perubahan iklim atau terorisme, keamanan yang lebih baik dan demokrasi yang lebih baik.(G. Stylios, Atall, 2010)

Algoritma *K-medoids* atau dikenal pula dengan PAM *(Partitioning Around Medoids)* menggunakan metode partisi *clustering* untuk mengelompokkan sekumpulan n objek menjadi sejumlah k *cluster*. Algoritma ini menggunakan objek pada kumpulan objek untuk mewakili sebuah *cluster*. Objek yang terpilih untuk mewakili sebuah cluster disebut medoids. *Cluster* dibangun dengan menghitung kedekatan yang dimiliki antara medoids dengan objek non-medoids (Kaufman, L. and Rousseeuw, P.J, 1990).

Dalam memisahkan data yang memiliki karakteristik yang berbeda dan mengelompokkan data yang memiliki karaktersitik yang sama dipergunakan metode *clustering*. Penelitian ini akan menggunakan dua metode *clustering* yaitu algoritma *K-means* dan algoritma *K-medoids* dimana hasil dari kedua algoritma tersebut akan dibandingkan untuk menentukan algoritma yang sesuai untuk menganalisa karakteristik data perangkingan *e-goverment* berdasarkan pengelompokan EGDI (*E-Government Development Index*).

2. Landasan Teori

2.1. Penelitian Yang Terkait

Menurut penelitian Hae-Sang Park dan Chi-Hyuck Jun, membahas tentang sebuah algoritma baru untuk *K-medoids* dalam pengelompokan yang berjalan seperti *K-means*s dan tes beberapa metode untuk memilih medoids awal. Algoritma yang diusulkan menghitung jarak metrik dan menggunakan untuk menemukan medoids baru di setiap iterasi langkah. Untuk mengevaluasi algoritma yang diusulkan, peneliti menggunakan beberapa data set nyata dan buatan serta dibandingkan dengan hasil algoritma laindalam hal pembuatan indeks. Hasil penelitian menunjukkan bahwa algoritma yang diusulkan membutuhkan waktu berkurang secara signifikan dalam perhitungan dengan kinerja yang sebanding terhadap partisi di sekitar medoids.

Menurut penelitian T. Velmurugan dan T, membahas algoritma yang paling representatif *K-means*s dan *K-medoids* diperiksa dan dianalisa berdasarkan pendekatan dasar. Algoritma terbaik di setiap kategori ditemukan berdasarkan kinerja. Titik input data yang dihasilkan oleh dua cara, satu dengan menggunakan distribusi normal dan menerapkan distribusi uniform. Hasil titik data terdistribusi secara acak diambil sebagai masukan untuk algoritma ini dan menemukan cluster disetiap algoritma. Algoritma dilaksanakan dengan menggunakan bahasa *Java* dan dianalisis berdasarkan kualitas pengelompokan mereka. Eksekusi waktu untuk algoritma dalam setiap kategori dibandingkan agar mendapatkan waktu yang berbeda. Keakuratan algoritma diselidiki selama eksekusi yang berbeda dari program pada titik input data. Kesimpulan waktu rata-rata yang diambil oleh algoritma *K-means* lebih besar dari waktu yang dibutuhkan oleh algoritma *K-medoids* untuk kedua kasus distribusi normal dan uniform atau waktu nilai propabilitas. Hasil terbukti memuaskan.

Menurut penelitian Niphat Claypo dan Saichon, membahas tentang perlunya review dari pelanggan untuk meningkatkan kualitas restoran dengan memanfaatkan algoritma MRF. Data review pelanggan yang berupa teks dirubah menjadi kata kunci dan mendapatkan vector input, dari vector input yang didapatkan akan dikelompokkan menggunakan *K-means* untuk dikategorikan berupa ulasan positif dan negative. Dari hasil percobaan menunjukkan bahwa penggunaan MRF sebagai seleksi fitur dapat memberikan efisiensi waktu yang signifikan dengan mengurangi jumlah fitur dalam data set. Sedangkan *K-means* dapat mencapai kinerja pengelompokan terbaik dibandingkan algoritma *clustering* yang lain yaitu Self-Organizing Map (SOM), Fuzzy C-Means, dan Hierarchical *Clustering*.

Tabel Perbandingan Algoritma *clustering Self-Organizing Map (SOM), Fuzzy C-Means*, dan *Hierarchical Clustering*.

MRF Seleksi Fitur	Metode Pembelajaran Unsupervised							
	K-means		Hierarchial		Fuzzy c-mean		SOM	
Jumlah Fitur	Akurasi (%)	Waktu (detik)	Akurasi (%)	Waktu (detik)	Akurasi (%)	Waktu (detik)	Akurasi (%)	Waktu (detik)
103	71.7	0.12	63.0	0.05	58.3	0.01	59.9	0.70
135	70.3	0.12	30.7	0.06	59.5	0.01	59.2	0.69
183	74.6	0.14	46.6	0.06	62.3	0.02	58.4	0.71
261	75.5	0.14	66.6	0.07	57.2	0.01	59.2	1.23
467	74.6	0.15	32.2	0.09	61.1	0.01	58.6	1.87
1768	69.3	0.89	77.9	0.31	59.2	0.04	59.0	5.55
Rata-rata	**71.73**	**0.26**	53	0.11	59.6	0.02	59.22	1.79

Dari hasil perbandingan pada Tabel menunjukkan bahwa algoritma *K-means* memiliki tingkat akurasi 71.73 dan waktu 0.26 sehingga dapat mencapai kinerja pengelompokan terbaik dibandingkan algoritma *clustering* yang lain.

Menurut penelitian Manish Vermaet, al. membahas tentang perbandingan berbagai jenis algoritma *clustering* dalam data mining. Dalam penelitian disebutkan bahwa melakukan pengelompokan data bukanlah hal yang mudah sehingga dibutuhkan algoritma yang dapat memberikan hasil yang relevan. Pada penelitian tersebut membandingkan enam algortma *clustering* yaitu *K-means*s, Hierarchical, DBSCAN, Density Based, Optics dan EM *algorithm* dengan analisa menggunakan tool

WEKA. Hasil pengujian menunjukkan kinerja algoritma *K-means* lebih baik dari algoritma Hierarchical, semua algoritma yang diuji memiliki ambiguitas pada beberapa noisy saat data bergerombol, kualitas algoritma *K-means* dan EM sangat baik ketika menggunakan *data-set* yang besar, algoritma DBSCAN dan Optics memiliki kualitas yang kurang baik untuk *data-set* berukuran kecil, algoritma *K-mean*s lebih cepat daripada algoritma *clustering* yang lain dan juga menghasilkan cluster yang berkualitas ketika menggunakan *data-set* berukuran besar.

Menurut penelitian Galuh Indah Zatadiniet, al. membahas tentang pemodelan suatu data set *Wholesale Customers* dimana dataset tersebut mengacu pada data klien dari distributor yang mencakup pengeluaran tahunan dalam satuan moneter pada berbagai macam produk. Pemodelan dataset tersebut dilakukan dengan menggunakan metode *clustering* dengan membandingkan 3 algoritma yaitu Simple *K-means, Expectation Maximization (EM)* dan *X-Means*. Tujuanmebandingkan 3 algoritma tersebut untuk mengetahui algoritma yang paling baik dalam pengolahan *Wholesale custumer dataset*. Hasil yang didapatkan pada perbandingan 3 algoritma *clustering* tersebut menunjukkan bahwa algoritma EM menunjukkan hasil yang cukup baik dengan error rate lebih sedikit, namun untuk efisiensi waktu algoritma *X-Means* jauh lebih cepat dibandingkan dengan algoritma yang lain.Hasil pengujian keempat metode *clustering* menggunakan dataset *Wholesale Customers.*

Tabel *clustering* menggunakan dataset *Wholesale Customers*.

Metode	Error rate	Waktu
*K-means*s (Euclideanm Distances)	13.116 %	0.04
*K-means*s (Manhattan Distances)	21.67%	0.39
EM (Expectation Maximisation)	12.76 %	7.2
X-means	13.116 %	0.06

Dari hasil yang ditunjukkan pada Tabel dapat diambil kesimpulan bahwa untuk dataset yang digunakan *(Wholesale Customers)* metode *clustering*EM memberikan hasil terbaik dengan error rate12.76%. Sedangkan jika dilihat dari segi efiesiensi waktu dan hasil *clustering*, *K-meanss Euclidean Distances* memberikan hasil terbaik (0,04 detik),walaupun tidak terlalu signifikan dibandingkandengan *X –Means* (0.06). Walaupun ada banyak metode yang biasa digunakan pada analisis data mining, namun analisis dengan menggunakan metode *clusteing* cukup membantu untuk mengetahui *knowledge* yang ada pada sebuah dataset.

Menurut penelitian Safuan et, al. membahas adanya *spam email* mengurangi produktivitas karyawan karena harus meluangkan waktu untuk menghapus pesan spam, Untuk mengatasi permasalahan tersebut dibutuhkan sebuah filter email yang akan mendeteksi keberadaan spam sehingga tidak dimunculkan pada inbox mail. Banyak peneliti yang mencoba untuk membuat filter email dengan berbagai macam metode, tetapi belum ada yang menghasilkan akurasi maksimal. Pada penelitian ini akan dilakukan klasifikasi dengan menggunakan algoritma Decision Tree Iterative Dicotomizer 3 (ID3) karena ID3 merupakan algoritma yang paling banyak digunakan di pohon keputusan, terkenal dengan kecepatan tinggi dalam klasifikasi, kemampuan belajar yang kuat dan konstruksi mudah. Tetapi ID3 tidak dapat menangani fitur kontinyu sehingga proses klasifikasi tidak bisa dilakukan. Pada penelitian ini, feature discretization berbasis Expectation Maximization (EM) *Clustering* digunakan untuk merubah fitur kontinyu menjadi fitur diskrit, sehingga proses klasifikasi spam email bisa dilakukan.

Menurut Mohammad Razavi Zadegan et, al. membahas tentang partisi baru dari algoritma yang inisialisasi tidak mengarah algoritma tetapi untuk mengoptimum dalam penemuan gaussian berbentuk cluster jika memiliki jumlah yang tepat dari algoritma. Dalam algoritma ini, kesamaan antara pasang benda dihitung sekali dan memperbarui medoids di setiap biaya iterasi $\Theta\ (k\ x\ m)$ di mana kadalah jumlah cluster dan m adalah jumlah objek yang dibutuhkan untuk memperbarui medoids dari cluster.

Perbandingan antara algoritma dua algoritma partisi lain dilakukan dengan menggunakan empat *wellknown* langkah-langkah validasi eksternal lebih dari tujuh dataset standar. Hasil untuk dataset yang lebih besar menunjukkan keunggulan

algoritma yang diusulkan lebih dari dua algoritma lain dalam hal kecepatan dan knosy, algoritma *K-medoids* yang menggunakan fungsi peringkat berguna untuk mencari benda-benda yang terletak di pusat. peneliti mengklarifikasi definisi optimum lokal untuk partisi pendekatan *clustering* dan menunjukkan bahwa algoritma dapat menemukan semua cluster yang Gaussian berbentuk jika jumlah yang tepat dari mereka diberikan inisialisasi tidak mengarah algoritma ke dalam lokal. Kesamaan antara objek-objek dalam dataset dihitung sekali dan memperbarui medoids biaya Θ (k x m) per iterasi, di mana k adalah jumlah *cluster* anggota di kelompokan untuk memilih medoids berikutnya. Untuk mengevaluasi algoritma, dua dataset buatan dan lima dataset nyata yang diujikan. *K-Harmonic Means* (KHM), algoritma dan sederhana dan cepat *K-medoids* dibandingkan dengan bantuan empat terkenal eksternal langkah-langkah validasi. Hasil penelitian menunjukkan bahwa KHM dan sederhanadan cepat *K-medoids* menderita optimum lokal. algoritma menemukan cluster di dataset yang lebih besar lebih cepat dan lebih akurat dibandingkan dengan dua algoritma lain dan tampaknya bahwa peringkat *k-medoids* pengelompokan adalah algoritma cocok untuk dataset yang besar .

2.2. Data Mining

Data mining merupakan salah satu bidang paling penting dalam penelitian yang bertujuan untuk memperoleh informasi dari data set. Data mining mulai ada sejak 1990-an sebagai cara yang efektif untuk mengambil pola dan informasi yang sebelumnya tidak diketahui dari suatu data set. Teknik data mining digunakan untuk menemukan hubungan antara data untuk melakukan pengklasifikasian yang digunakan dalam klasifikasi nilai-nilai dari beberapa variabel, membagi data yang diketahui menjadi kelompok-kelompok yang mempunyai kesamaan karakteristik (*clustering*). Data mining merupakan bagian dari proses penemuan pengetahuan dari basis data (*Knowledge Discovery in Databases*), yang mana tahapan dari *Knowledge Discovery in Databases*

2.3. Clustering

Clustering adalah bagian pembelajaran *unsupervised* yang digunakan sebagai data mining yang efektif (S. Ding, F. Wu, Dkk, 2013). Metode c*lustering* dapat mengungkapkan hubungan dan struktrur yang sebelumnya tidak jelas dari *data-set*. Dikarenakan *Clustering* tidak memiliki label, sehingga seluruh atribut yang dimiliki dianggap sama. Tujuan *clustering* adalah untuk mengelompokkan data yang memiliki kesamaan karakteristik kedalam kelompok yang sama dan data yang berbeda karakteristik kedalam kelompok yang lain. Algoritma *clustering* dapat dikategorikan kedalam empat metode yaitu *Partitional-based*, *Hierarchical-based*, *Density-based* dan *Grid-based*.

2.4. Manhattan *Distance*

Manhattan *distance* atau dikenal juga dengan *City block distance* digunakan untuk menghitung jarak dengan tujuan untuk mendapatkan jarak dari satu titik data ke titik data yang lain. Manhattan *distance* mencerminkan jarak antar titik di jalan perkotaan dalam 1 blok (P. Grabust, 2011). Persamaan matematik dari manhattan *distance* yaitu:

$$D(x,y) = \sum_{i=1}^{n} |x_i - y_j| \dots\dots\dots\dots\dots\dots\dots (2.1)$$

Dari persamaan 1, x = ($x_1, x_2, x_3, \dots, x_n$) dan y = ($y_1, y_2, y_3, \dots, y_n$). Perhitungan manhattan *distance* adalah dengan menjumlahkan hasil absolut dari pengurangan antar titik.

2.5. Euclidean *Distance*

Dalam matematika. Euclidean *distance* digunakan untuk megukur antara dua titik dalam satu dimensi yang memberikan hasil seperti rumus *Pythagoras* (H. K. Sagar and V. Sharma, 2014). Persamaan Euclidean *distance* yaitu (P. Grabust, 2011):

$$D(x,y) = \sqrt{\sum_{k-1}^{n} (x_{ik} + y_{jk})^2} \dots\dots\dots\dots\dots\dots\dots (2.2)$$

Dari persamaan 2, x = ($x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}$) dan y = ($y_{j1}, y_{j2}, y_{j3}, \dots, y_{jn}$). Euclidean *distance* diperoleh dari jumlah kuadrat antar titik yang diakar kuadratkan.

2.6. Chebyshev *Distance*

Nilai jarak maksimum atau disebut juga Chebychev *distance* merupakan perhitungan jarak yang menghitung besarnya hasil absolut dari perbedaan antara sepasang objek (P. Grabust, 2011). Chebychev *distance* dapat dihitung menggunakan persamaan:

$$D(x,y) = max_k |X_{ik} - Y_{jk}| \dots\dots\dots\dots\dots\dots\dots (2.3)$$

Metrik dalam *chebychev distance*, didefinisikan didalam ruang vektor yang mana jarak antara dua vektor yang memiliki perbedaan terbesar disepanjang dimensi koordinatnya.

2.7. Davies Bouldin Index

Davies Bouldin Index (DBI) merupakan cara validasi *cluster* yang dibuat oleh D.L. Davies. DBI adalah fungsi rasio dari jumlah distribusi didalam *cluster* untuk pemisahan antar *cluster*. Pengukuran menggunakan DBI bertujuan untuk memaksimalkan jarak *inter-cluster*. Dalam penelitian ini, DBI digunakan untuk melakukan validasi data pada setiap *cluster*. DBI dapat dihitung menggunakan persamaan:

$$R_i = \max_{j=1\ldots k, i\neq j} R_{ij} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.4)$$

$$var(x) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.5)$$

$$R_{\substack{ij \\ i\neq j}} = \frac{var(C_i) + var(C_j)}{\|c_i - c_j\|} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.6)$$

$$DB = \frac{1}{k}\sum_{i=1}^{k} R_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.7)$$

Keterangan:

- R : jarak antar *cluster*
- Var : *variance* dari dataT6
- x : data ke-i
- $\bar{x}$: rata-rata dari tiap *cluster*
- DB : validasi Davies Bouldin

Dengan menggunakan *Davies Bouldin Index* suatu *cluster*akan dianggap memiliki skema *clustering* yang optimal adalah yang memiliki *Index Davies Bouldin* minimal.

2.8. *K-medoids*

K-medoids menggunakan k sebagai jumlah pusat cluster awal yang dihasilkan secara acak diawal proses clustering. Setiap obyek yang lebih dekat dengan pusat cluster akan dikelompokkan dan membentuk cluster baru. Algoritma kemudian secara acak menentukan cluster center baru darisetiap cluster yang terbentuk sebelumnya dan menghitung ulang jarak antara obyek dan pusat cluster baru yang dihasilkan. Jarak antar obyek i dan j dihitung dengan menggunakan dissimilarity measurement function, dimana salah satunya adalah Euclidean Distance Function yang ditunjukkan dalam persamaan berikut:

$$dii = \sqrt{\sum_{a=1}^{p}(x_{iu} - x_{iu})^2}, i = 1, \ldots\ldots n; j = 1, \ldots . n \quad (2.8)$$

dimana Xia adalah variabel ke-a dari obyek i (i=1, ..., n; a=1, ..., p) dan dij adalah nilai Euclidean Distance. Algoritma juga menghitung probabilitaspenukaran setiap obyek dengan pusat cluster yang lain menggunakan fungsi kriteria. Salah satu fungsi kriteria yang digunakan adalah absolute-error (Kaufman, L., and Rousseeuw, P. J, 1990), seperti pada persamaan berikut:

$$E = \sum_{j=1}^{k}\sum_{\rho\epsilon cj}^{n} |p - 0j| \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.9)$$

di mana E adalah jumlah dari absolut error untuk semua objek dalam dataset; p adalah titik dalam ruang yang mewakili suatu objek dalam kluster Cj, dan oj adalah obyek didalam cluster Cj.

2.9. Keranga Pemikiran

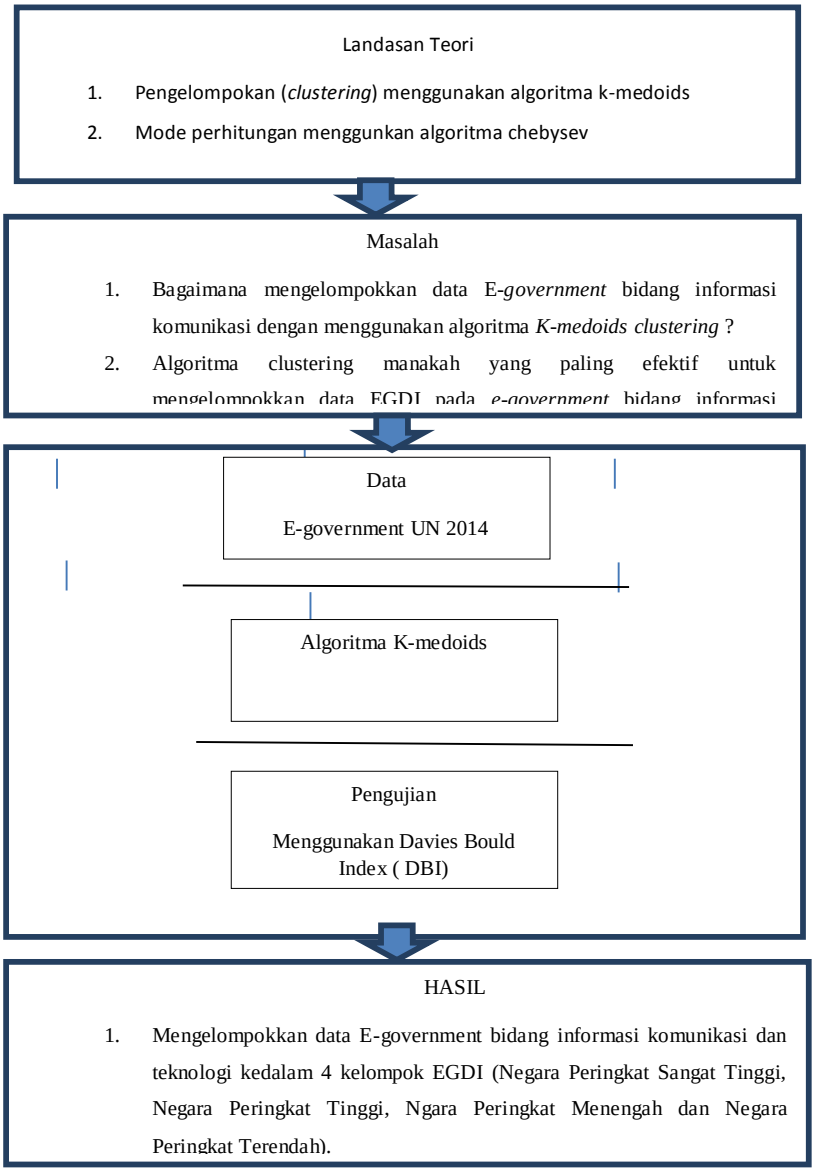
Kerangka pemikiran adalah cara memecahkan masalah yang telah diidentifikasi. Dalam kerangka pemikiran akan dipaparkan tentang perpaduan antara asumsi teoretis dan asumsi logika didalam menjelaskan dan memunculkan variabel yang akan diteliti, selain itu juga berisi tentang bagaimana kaitan di antara variabel-variabel ketika digunakan untuk mengungkapkan fenomena atau masalah yang akan diteliti. Kerangka pemikiran yang disusun dalam tugas akhir ini akan menggambarkan konsep pemecahan masalah yang dihadapi, meliputi:

a. Alur atau jalan pikiran secara logis dalam menjawab dan meyelesaikan masalah yang didasarkan pada landasan teoretik dan hasil penelitian yang relevan.

b. Kerangka logika yang mampu menunjukkan dan menjelaskan masalah yang telah dirumuskan kedalam kerangka teori.

c. Model penelitian yang dapat disajikan secara skematis dalam bentuk gambar atau model matematis yang menyatakan hubungan-hubungan variabel penelitian yang merupakan rangkuman dari pemikiran yang digambarkan dalam suatu model.

Adapun kerangka pemikiran dalam penelitian ini. Secara lengkap dapat dilihat pada Gambar 2.2.



Gambar 2. 1. *Kerangka Pemikiran*

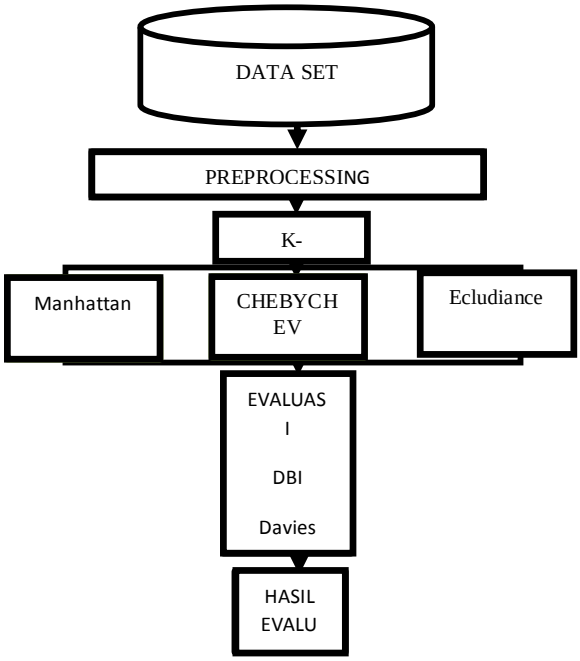
Pada Gambar 2.2 menunjukkan bahwa kerangka pemikiran dari penelitian ini berisi tentang masalah penelitian yaitu mengelompokkan data E-government bidang informasi komunikasi dan teknologi kedalam 4 (empat) kelompok EGDI (Negara Peringkat Sangat Tinggi, Negara Peringkat Tinggi, Ngara Peringkat Menengah dan Negara Peringkat Terendah). Teori yang akan digunakan adalah menggunakan data mining dengangan metode *K-meansclustering* dan *K-medoidsclustering*. Eksperimen akan dilakukan dengan mengelompokkan dengan membandingkan dua algoritma.

3. Metode Penelitian

Dalam penelitian ini akan mengusulkan metode data mining yaitu *clustering* menggunakan algoritma *K-means* dan *K-medoids* dengan Kemudian dari metode tersebut akan digunakan untuk

mengelompokkan dibagi kedalam 4 *cluster* yaitu *cluster* 0, *cluster* 1, *cluster* 2, *cluster* 3.

.



Gambar 3. 1. *Metode yang diusulkan*

Pada penelitian ini, data EDGI *E-Government* yang ada terlebih dahulu dilakukan pengolah awal, di dapatkan data *cluster* yang akan diuji serta dibagi menurut kelompok. Kelompok yang dimaksud adalah pengkategorian dari tiap jenis status EDGI yang disesuaikan dengan pengkategorian EDGI oleh *K-means* dan *K-medoids*, sehingga bisa dibedakan menjadi 4 (empat) kelas status EDGI, yaitu : Negara Peringkat Sangat Tinggi, Negara Peringkat Tinggi, Negara Peringkat Menengah dan Negara Peringkat Terendah.

Dari keseluruhan data EDGI diambil analisa dengan menggunakan metode *Chebyshev* distance dengan mencari jarak cendtroid . Hasil dari ekstraksi dilakukan *cluster* menggunalan *K-medoids dengan Davies Bouldin Index*untuk mendapatkan table *Bouldin index* serta mendapatkan hasil *inter-cluster*.

4. Pembahasan

4.1. Pembahasan

Pada perhitungan algoritma *k-medoids* kali ini menggunakan dataset data E-goverment UN 2014, dimana pada dataset tersebut dipilih 30 sampel data yang dikelompokkan ke dalam 4 klaster seperti ditunjukkan pada Tabel 4.39 dan iterasi yang digunakan sebanyak 10 akan tetapi yang ditunjukkan proses perhitungannya di sini hanya satu buah iterasi.

Langkah-langkah perhitungan:

4.2. Persiapkan sampel dataset berjumlah 30 data seperti ditunjukkan pada Tabel 4.39

Tabel 4. 1. sampel dataset E-government UN 2014

No	Human Capital index (HCI)	Online Service index (OSI)	Infra Struktur (TII)
1	0,2418	0,1811	0,1472

2	0,7100	0,4488	0,3548
3	0,6543	0,0787	0,1989
4	0,7277	0,4331	0,7671
5	0,4941	0,2992	0,0978

Pemilihan medoids dilakukan secara random, diambil empat data secara acak dari sampel dataset.

Dari pemilihan secara acak didapatkan urutan data 2, 7, 9, 12 sehingga medoids adalah seperti di bawah ini.

Tabel 4. 2. Daftar medoids

Medoids	Atribut1	Atribut2	Atribut3
medoids 1	0,7100	0,4488	0,3548
medoids 2	0,8571	0,5512	0,4835
medoids 3	0,9978	0,9291	0,8041
medoids 4	0,7138	0,3386	0,4176

Tabel 4.40 berisi empat buah medoids yang diambil dari sampel dataset yang terdapat pada urutan data ke 2,7, 9, 12 pada dataset E-goverment UN 2014.

Menghitung jarak antara medoids dengan seluruh data pada sampel dataset dengan menggunakan metode pencarian jarak *chebyshev*. Perhitungaan jarak antara data pertama (x1) dengan medoids 1 (y1).

$$x1 = [0{,}2418 \quad 0{,}1811 \;\; 0{,}1472]$$

$$y1 = [0{,}71000{,}44880{,}3548]$$

Sehingga perhitungan jarak Antara x1 dengan medoids 1 (y1) dengan menggunakan *chebyshev* adalah seperti berikut.

$$D(x1, y1) = \max(|0{,}2418 - 0{,}7100|, |0{,}1811 - 0{,}4488|, |0{,}1472 - 0{,}3548|)$$

$$D(x1, y1) = max(0{,}4682, 0{,}2677, 0{,}2076)$$

$$D(x1, y1) = 0{,}4682$$

Sehingga proses perhitungan jarak antara ke tiga puluh data dengan medoids 1 (y1) adalah dapat ditunjukkan pada Tabel 4.41.

Tabel 4. 3. Perhitungan jarak antara data dengan medoids 1

Indeks urutan data	$\|At1 - y1{,}1\|$	$\|At2 - y1{,}2\|$	$\|At3 - y1{,}3\|$	$D(xi, y1)$
1	$\|0{,}2418 - 0{,}7100\|$=0,4682	$\|0{,}1811 - 0{,}4488\| =$ 0,2677	$\|0{,}1472 - 0{,}3548\| =$ 0,2076	0,4682
3	$\|0{,}6543 - 0{,}7100\|$=0,0557	$\|0{,}0787 - 0{,}4488\| = 0{,}3701$	$\|0{,}1989 - 0{,}3548\| =$ 0,1559	0,3701

Tabel 4.41 menunjukkan proses perhitungan pencarian jarak dengan metode chebisev dimana

$y1{,}1$ menunjukkan atribut pertama dari medoids pertama, $y1{,}2$ menunjukkan atribut kedua dari medoids pertama, $y1{,}3$ menunjukkan atribut ke tiga pada medoids pertama, atribut1 menunjukkan atribut pertama, atribut2 menunjukkan atribut kedua, atribut3 menunjukkan atribut ke tiga dari data (x).

Selanjutnya dilakukan perhitungan nilai *cost* pada setiap cluster. Untuk itu dilakukan pengumpulan nilai *cost* pada setiap data sesuai dengan klasternya

$$Total\ cost\ baru = jumlah\ cost\ kluster 1 + jumlah\ cost\ kluster\ 2 + jumlah\ cost\ kluster\ 3 + jumlah\ cost\ kluster\ 4$$

$$Total\ cost\ baru = 1{,}4697 + 4{,}4821 + 0 + 0{,}3695 = 6{,}3213$$

Karena total *cost* baru lebih besar dibandingkan total *cost* sebelumnya yaitu $5{,}9651$ maka medoids sebelumnya tetap dipertahankan sedangkan kandidat medoids dibatalkan dan harus dilakukan pemilihan kandidat medoids yang baru kembali.Setelah melalui iterasi ke 10 maka didapatkan medoids seperti Tabel 4.56 dan klaster data seperti Tabel 4.57.

Tabel 4. 4. Nilai medoids final

Medoids	At1	At2	At3
medoids 1	0,5189	0,1732	0,2075
medoids 2	0,7372	0,5984	0,4668
medoids 3	0,796	0,2362	0,5941
medoids 4	0,8932	0,6772	0,6988

Tabel 4.56 menunjukkan hasil akhir medoids setelah mencapai iterasi ke 10.

Tabel 4. 5. *Claster* setelah algoritma *k-medoids* berakhir

Klaster1	Klaster 2	Klaster 3	Klaster 4
1	2	4	9
3	7	6	10
5	8	12	13
14	11	15	17
18	21	16	
19	24	22	
20		25	
23		26	
27			
28			
29			

Tabel 4.57 menunjukkan pengelompokkan berdasarkan ke empat klaster, setiap item pada

tabel tersebut menunjukkan kedudukan data pada sampel dataset E-government UN 2014 yang terdapat pada Tabel 4.15.

4.3. Pengujian Davies Bouldin Index

Dalam penelitian ini, Davies Bouldin Index (DBI) digunakan untuk melakukan validasi data pada setiap cluster. Pengukuran menggunakan DBI bertujuan untuk memaksimalkan jarak inter-cluster. Dengan menggunakan DBI suatu cluster akan dianggap memiliki skema clustering yang optimal jika yang memiliki Index Davies minimal. Adapun dari pengujian yang sudah dilakukan diperoleh nilai Index Davies dari *K-medoids Chebyshev* didapat bahwa metode tersebut lebih optimal dalam penentuan pengelompokan EDI *E-govermnet Surve* 2014 ke dalam 4 status EDGI, dengan nilai DBI 0.593

5. Kesimpulan Dan Saran

5.1. Kesimpulan

Dalam penelitian ini dilakukan pengujian model dengan menggunakan *K-medoids Chebyshev* dalam pengelompokan data EDGI E-government Surve 2014 kedalam 4 status EDGI. Model yang dihasilkan diuji untuk mendapatkan nilai Bouldin Index dari setiap algoritma sehingga didapat pengujian dengan menggunakan data dan setelah dilakukan pengujian dengan tools rapidminer didapat nilai Bouldin Index adalah 0.593 pada metode *K-medoids Chebyshev*. suatu *cluster* akan dianggap memiliki skema *clustering* yang optimal adalah yang memiliki *Index Davies Bouldin* minimal maka dapat disimpulkan pengujian data EDGI E-goverment Surve 2014 tersebut lebih optomal dalam penentuan pengelompokan EDI E-govermnet Surve 2014 ke dalam 4 status EDGI di bandingkan menggunakan mahattan atau ecludian.

5.2. Saran

Dari hasil pengujian yang telah dilakukan dan hasil kesimpulan yang diberikan maka ada saran atau usul yang di berikan antara lain:

a. Dari penelitian ini di ketahui bahwa setiap cluster yang dihasilkan memiliki jarak dari cluster yang lain, dengan hasil cluster yang diperoleh diharapkan dapat diukur jarak setiap atribut antar cluster agar dapat ditentukan atribut mana yang perlu diperhatikan sebagai prioritas perkembangan E-govermnet agar status EDGI dapat meningkat.

b. Untuk meningkatkan hasil jarak centroid dapat dilakukan metode pemilihan jarak dengan metode *Euclidean Distance, Manhattan Distance,* dan lain-lain

c. Mencoba menerapkan metode optimasi sebagai bahan perbandingan seperti EM *(Expectation Maximisation).*

6. Daftar Pustaka

L. Noviani, A. A. Suryani and A. P. Kurniati, "Pengklasifikasian Dokumen Berita Berbahasa Indonesia Menggunakan Latent Semantic Indexing (LSI) dan Support Vector Machine (SVM)," *ISSN:1979-911X*, 2012.

G. Stylios, D. Christodoulakis, J. Besharat, M.-A. Vonitsanou, I. Kotrotsos, A. Koumpouri, S. P. U. Stamou and G. , "Public Opinion Mining for Governmental Decisions," *Electronic Journal of e-Government*, vol. 8, no. 2, 2010.

Kaufman, L. and Rousseeuw, P.J., Finding Groups in Data: An Introduction to Cluster Analysis, vol. 39, New York: Wiley, 1990, pp. 1-38.

S. Ding, F. Wu, Q. Jun, H. Jia and F. Jin, "Research on data stream clustering algorithms," *Artificial Intelligence Review*, vol. 43, no. 4, pp. 593-600, 2013.

P. Grabust, "The Choice of Metrics for Clustering Algorithms," in *Proceedings of the 8th International Scientific and Practical Conference*, Augstskola, 2011.

H. K. Sagar and V. Sharma, "Error Evaluation on K-Means and Hierarchical Clustering with Effect of Distance Functions for Iris Dataset," *International Journal of Computer Applications*, vol. 86, no. 18, pp. 1-5, 2014.

Kaufman, L., and Rousseeuw, P. J.,, "Finding groups in data: An introduction to cluster analysis," Wiley, New York, 1990.